A W H Phillips 50th Anniversary Symposium

Economica, Volume 78, Issue 309, 2011

Economica LSE

Economica (2011) **78**, 1–9
doi:10.1111/j.1468-0335.2009.00813.x

A. W. H. PHILLIPS 50th ANNIVERSARY SYMPOSIUM

In 1958, A. W. H. Phillips published in *Economica* what was to become one of the most widely cited articles ever written in economics. To mark the 50th anniversary of the paper, the New Zealand Association of Economists and the Econometric Society hosted the conference "Markets and Models: Policy Frontiers in the A. W. H. Phillips Tradition" in July 2008. The four articles that follow were originally presented at that conference.

Man, Money and Machines: The Contributions of A. W. Phillips

By Alan E. Bollard

Reserve Bank of New Zealand

Final version received 2 February 2009.

A. W. Phillips is widely known for the Phillips curve, from an article in *Economica* fifty years ago. Less well known is that he was an unusually talented electrical engineer who retrained as an economist and made pioneering, if unconventional, contributions across a wide range of macroeconomics: physical modelling, dynamic stabilization, continuous-time modelling and applied econometrics. In making these contributions, he was a child of his times. This paper focuses on the environment that helped make Phillips, what motivated his questions, how he formed his ideas, how he solved problems, and what difference that has made to the study of economics.

Introduction

A sophisticated hydraulic model, pioneering dynamic stabilization models, a famous relationship between prices and activity, little known continuous-time modelling work, unfinished work in Chinese development: Bill Phillips was an extraordinary person living in an extraordinary time.

The life of Bill Phillips is well documented in Barr (1973), Bergstrom *et al.* (1978), Blyth (1975) and Leeson (2000). This paper focuses on the environment that made Bill Phillips: where he found his ideas, how he formed them and put them into practice, and how he made a difference to economics.

I. Upbringing

Bill Phillips was born in 1914 into a pioneering, hard-working and unusually innovative farming family in Te Rehunga in the Hawke's Bay region of New Zealand. His father had won a block of land in a ballot at the turn of the century, which by hard labour he converted from rough bush to a productive dairy farm. The family members were committed 'rationalist' Christians (the father had chosen Anglicanism only after studying world religions and philosophy). They lived an ethos of self-help and community-help in a small tight-knit community of farmers. Compared with the British culture of many of their ancestors, this was a different place: its culture valued hard work, innovation and equality.

Phillips' father was particularly important to him. On the surface he was a pillar of the community, a lay preacher and chairman of many committees. At the same time he was an unusual inventor: he built the farmhouse and outbuildings himself. He followed this with a water wheel which was used to generate electricity from a stream. The Phillips farm was the first in the district to have mechanical milking machines. With electricity in the house, the family could have light at night time, and they read widely as well as playing a variety of instruments.

School was a long bike, train and walking trip away. Despite arriving home late each day, the young Bill Phillips was expected to help with milking and other farm work. Bill soon showed his own technical precocity: building a book rack onto his bike so he could read while cycling, and later rebuilding a neighbour's broken-down truck so that he could (illegally) drive to school. Weekends meant more farm work, but Bill and his brother also roved the farm, playing, fishing and in particular building things: among these were a zoetrope, a mechanical shooting gallery, a magic lantern and a crystal radio.

In the 1920s, Phillips' father took advantage of strong dairy prices to buy more land and build up a dairy cow stud herd, using the modern principles of dairy breeding that New Zealand was developing. But in 1929, world depression hit the remote hamlet of Te Rehunga. Dairy prices fell by a third.

The young Bill was clearly very talented, but his parents reluctantly decided that they could not afford to keep him at school, and any dreams of a university education were abandoned. Aged 15, Bill signed up as an apprentice electrician with the government's Public Works Department, which at that time was building infrastructure around rural New Zealand. He spent the next few years roughing it at working men's camps in remote rural sites, helping to build hydroelectric dams to generate electricity for the national grid. Bill had played with photography and seen early movies, and he was fascinated by the idea of 'talkies'. He hired a hall in the Tuai camp and set up the first talkies cinema. Recreation involved playing his violin, riding an acquired motorbike and reading his treasured encyclopaedia of world religions.

But rural New Zealand was not enough. Phillips wanted to sample the world. In 1935, still aged only 21, he packed his swag and his fiddle, and shipped to Australia. Here he spent a couple of years travelling the outback, hitching rides on freight trains and working in mining camps. Money came from a range of jobs: picking bananas, working on building sites, mining gold, running a cinema, and even crocodile hunting. These were tough jobs in a rough country, but at the same time Phillips had set his intellectual sights higher. He enrolled in a correspondence course in electrical engineering and remembers learning his first differential equations under a harsh Australian sun at an outback mining camp.

Phillips had a lifelong fascination with Eastern cultures. In 1937, despite the worsening international situation, he boarded a Japanese ship to travel to Shanghai. While he was at sea, the Japanese invaded Manchuria, and the ship was diverted to Yokohama. Phillips took advantage of this by travelling around the newly militarized Japan; at one point he was detained by the authorities, who suspected that he might be a spy. Eventually he made his way out through Korea, Manchuria and Harbin, and crossed Russia on the Trans-Siberian railway. With Antipodean optimism, he looked for casual jobs across Soviet Russia, only to find them all taken by political prisoners. From Stalin's Moscow he travelled on through threatened Poland and Nazi Germany during the fragile last years of peace. He settled in London, where he found work as an electrical engineer. Having continued his correspondence course, Phillips now graduated from the Institute of Electrical Engineers, gaining his first formal qualifications. He also took classes in several languages.

When war broke out, Phillips enlisted in the RAF. He was sent as a Flight Lieutenant to Singapore, where his job was to maintain the RAF Brewster Buffalos stationed there. These were old aeroplanes, but they were the only real Allied fighters in the East. Phillips found a way to equip them with new American heavy machine guns that were synchronized to fire through the propellers.

In 1942, the Japanese forces swept down the Malayan Peninsula and took Singapore. Phillips was aboard the *Empire Star*, the last merchant ship to leave Singapore, overcrowded with soldiers, women and children. Steaming out of Singapore, the ship came under attack from Japanese fighters and bombers, and was severely damaged. Phillips managed to improvise some heavy mountings on the boat deck, allowing troops to operate machine guns for anti-aircraft fire. For this gallantry he was later awarded the MBE.

The ship limped into Java, where Phillips and several comrades evaded the Japanese and trekked to the south coast. Here they set up a hidden camp, acquired an old bus body, and under Phillips' inventive supervision started to rebuild it into a boat with the intention of sailing to Australia. However, they were discovered and captured.

The next three years of Phillips' life were spent as a Japanese prisoner of war at Bandura in Java. Life was very harsh, discipline brutal, food and medicine completely inadequate. Many men died in these terrible conditions. Despite the environment, Phillips was never idle. He helped to organize language classes teaching Mandarin and Russian. He secretly built electrical immersion heaters to help the troops make cups of tea. He was involved in a dangerous mission to steal parts, and build several clandestine radios, one set into the laundry floor, another contained in a homemade wooden clog. It was Phillips' radio that brought the eventful news of Hiroshima and eventual rescue.

Phillips was repatriated to his family in New Zealand in 1945 but, like many ex-prisoners of war, he was in very poor shape. He had lost considerable weight, was addicted to nicotine, and was deeply scarred by his experiences. He was given a New Zealand ex-serviceman's grant, and he chose to return to London and enrol at the London School of Economics (LSE) where he decided to study sociology. We may surmise that he was desperate to rationalize or comprehend the experiences of the war years. However, he did not enjoy sociology, finding his lectures boring and describing the discipline as 'a combination of ethics, social statistics and pseudo-science'.

He was awarded a bare pass degree.

II. Hydraulic Model

It was only at this stage, in his thirties, that Phillips had his first experience of some basic classes in economics. The LSE in this political period was a hotbed of competing classical and Keynesian arguments. Phillips' insight from limited macroeconomics teaching was that he could demonstrate how a macroeconomy worked in a very visual and insightful way, using a Keynesian/classical IS-LM model. He drew up plans for a hydraulic model representing the UK economy, with a flow of liquid representing money.

The head of department, Sir Lionel Robbins, records how he was confronted by Phillips in the lobby of the LSE. Barr (1973) tells the story of 'a wild man from New Zealand waving blue-prints in one hand and queer shaped pieces of perspex in the other'. Robbins had sufficient insight not to squash this 'madman' completely, instead handing him over to a junior colleague, James Meade, who also had a sympathetic, even eccentric, interest in things mechanical. Phillips and Meade became lifelong friends and colleagues.

With Meade's help, Phillips secured a small grant from the LSE to build his prototype model.

Economic modelling had grown significantly during the First World War. The Soviets had developed cybernetics for central planning, the British had developed national quality accounting techniques to monitor the war effort, and the US defence establishment was using early operations research methods. Leontief's input/output economic mapping had provided a basis for allied bombing to identify the key sectors of an economy; Nicholas Kaldor had been involved in appraising the effects of this strategic bombing.

Now the war was over, but the problems of allocating resources in a peacetime economy designed for wartime conditions were very acute. Britain was enduring particular problems: harsh rationing, even malnutrition, a historically bad winter in 1947, continued sterling crises, the financial burden of repaying the harsh US lend-lease loans, and a disruptive workforce led by militant unions flirting with communism, strikes and riots. The newly elected Labour government had nationalized key sectors of the economy, but was struggling with the principles of peacetime macro management.

After a tough year 'living on air', as he put it, Phillips, assisted by colleague Walter Newlyn, built his prototype machine in a garage belonging to his landlord in Croydon. In 1949, he transported it to the LSE and demonstrated it in a seminar for a distinguished audience that included Robbins, Kaldor and Meade, and also Hayek, Coase and Sen. Some of the audience no doubt came to laugh, but by all accounts they were bowled over by the cleverness of this hydraulic model.

Most unusually, the department agreed to grant Phillips £700 to build a full model of the UK economy, and this he proceeded to do. What emerged was a wonderful machine made largely of war surplus materials: it had transparent tubes and tanks made of perspex from Lancaster Bomber windscreens, clockwork mechanisms, and floats and pulleys connected by fishing line and linked to mechanical graphs driven by a motor from an aeroplane windscreen wiper. Cochineal-dyed water was used to represent money, pumped (by aeroplane landing gear pumps) around a system representing the identity $Y = C + I + G + X - M$. Propensities to consume, save, invest, etc. were cut as slits into perspex sheets. Each variable was driven by a combination of water flows and indirect signals from the interactions of other variables, all linked by an ingenious network of floats and pulleys.

The whole represented a system of nine differential equations. Ordinarily, at the time, such a system would be too complicated for anyone to solve. However, Phillips had ingeniously found an analogue solution. Not only that, but he calibrated the model for the UK economy, going as far as to estimate confidence intervals for the accuracy of results (which were primarily driven by surface tension and engineering tolerances). Though 'hard-wired' for 1949 conditions, the machine is extraordinarily adaptable for Keynesian interpretations, allowing for either flexible or regulated capital and trade flows with the world, and permitting variation in the behavioural parameters.

The machine was set up to allow experiments in fiscal policy through tax and government spending, monetary policy and exchange rate adjustments. Remarkably, this was all designed at a time when such policy was still immature. The machine could show a transitional adjustment to a policy experiment, and eventually a new steady state, all printed out automatically on one of the overhead graphs. Despite its calibration, Phillips did not see the machine as a practical computer; rather, it was designed for its expositional value, and it soon became popular for demonstrations.

The second machine that Phillips built was a mere mirror image of the first. This one was calibrated to represent the economy of the rest of the world. It could be linked with the first machine so that capital and international trade could flow between them. James Meade is remembered for lectures in the 1950s where he would use the two machines linked together to demonstrate theories of trade and international finance. He would designate students to control fiscal and monetary policy on each model, aiming to optimize economic growth. A number of British and American policy-makers, including Paul Volcker (later Chairman of the Federal Reserve) and Richard Cooper (Assistant Secretary of the US Treasury), learned some practicalities of policy harmonization in this way.

Over the next few years, Phillips received orders for a dozen such machines. He engaged model-makers to produce them, and they sold for around £1500 each (approximately £40,000 today) to universities, central banks and companies around the world. Approximately half of them still exist in some form today.

American economist Abba Learner was a special enthusiast for the machine and acted as Phillips' US agent. He christened the machine the MONIAC (Monetary National Income Automatic Computer), at the time of IBM's ENIAC and Von Neumann's MANIAC. Engaging as they were, the MONIACs had significant limitations: they were big, clumsy and damp. Phillips mentions being called on by desperate LSE lecturers to help mop up after experiments went disastrously wrong. Economic relationships were hard-wired. And the first real electronic calculators were starting to emerge in the USA. For Phillips it was time to move on.

The LSE faced an anomalous situation: on their hands they had a unique economic model maker, with the potential for greatness but only a bare pass in sociology. Proving flexible, the school persuaded the New Zealand government to waive the bond on Phillips' scholarship, and then appointed him to a junior staff position. Phillips wrote an article about his machine, as an economic model and as a mechanical device, which was published in *Economica* in 1950, receiving intense interest. He was then encouraged to commence work on his PhD, and was appointed Assistant Lecturer in 1951.

III. Dynamic Modelling

The early 1950s were a turbulent time; the Cold War had begun in earnest with the H-bomb stand-off between Russia and the USA, the Korean War and the McCarthy witch hunts. British attention was focused on the decline of the prewar British Empire and the growing independence movements.

At home, the British economy was still very fragile: wartime debt, the transition from a command economy, price shocks and currency instability joined together to make for economic instability. Presented with a destabilizing shock, the MONIAC economy soon re-equilibrated, but this hardly seemed realistic. In his PhD work, Phillips turned his mind to more complex questions about restabilizing an economy. He brought to this problem his electrical engineering background, initially taking an optimal control theory approach. His 1954 *Economic Journal* paper represents one of the first applications of dynamic control theory to macroeconomics. In it he constructed a very simple mathematical macro model and applied various shocks. He then constructed a simple policy response and, using manual calculation, showed that in principle it was possible to dampen the cycles. Economists since have carried out tens of thousands of such experiments, but Phillips' paper, with its elegant simplicity, was pioneering.

This confirmed for the LSE their conviction about Phillips' talents, and he was appointed Reader in Economics. His new position and salary gave Phillips some personal security for the first time in his life, and that year he decided to marry Valda, a New Zealand woman living in London.

But if his personal life was now more stable, the international scene was not. In 1956, France and Britain disastrously invaded Suez, sparking huge unrest in the region and another sterling crisis at home. In the same year, the Soviets invaded Hungary. With Hungarian economists Nicky Kaldor, Thomas Balogh and Peter Bauer all associated with the LSE (and many others, such as John von Neumann who was influential in the USA), the intellectual impact of this invasion should not be underestimated.

Against this background, Phillips continued his work on dynamic stabilization. He keenly felt the need to design a more realistic model for policy simulation. His new model involved lags, more realistic dynamics, variable prices and interest rates. But it was clear that such a model would quickly become too complex for manual calculation.

Big economic model design had been led by the Cowles Commission in the USA in the early 1950s. They estimated model parameters using groups of human 'computers', typically middle-aged women calculating parameters in the models using electromechanical calculating machines and statistical shortcuts. The effort was immense. Fresh from his hydraulic analogue machine experience, Phillips had been speculating that an electronic analogue machine could rapidly approximate trials to test the responses stimulated by various wattages, to be shown on an oscilloscope. That was essentially the description of what electronic analogue computers might deliver in the future.

In the USA, electronic computers were becoming available from two sources: US defence work (especially the Manhattan Project) and civilian development of business machines. In the UK, much of the clever wartime computing developments were now locked up in British government security establishments. The most promising university computing work was at Manchester, where the ACE machine was being developed in the early to mid-1950s. A descendent of this machine, named the DEUCE, was housed in the National Physical Laboratories (NPL) in London near the LSE. This was a massive machine, consisting of stacks of boxes of valves, dials and switches connected by tangles of wire. Phillips had a friend working at the NPL, Richard Tizard, a very clever physicist, and together they discussed the problems of stabilizing models. One weekend, Tizard let Phillips into the building and they proceeded to programme and successfully run Phillips' new, more complex mathematical model on the DEUCE machine.

Once again, this was Phillips the pioneer at work. In the USA, economic models had been run on large electrical mechanical computers before (Leontief's input/output of the US economy run on the Harvard Mark I computer is the best known example). However, Phillips' 1957 work arguably marked the first model-based computer policy simulations of the type that are so common today. Such work was not generally known or accepted at the time, as is shown in a later report by the Department of Scientific and Industrial Research on the use of the DEUCE model: it speculated that it was still an open question as to whether: 'Phillips' work on the analogue computer is useful or a menace to society'.

The results of this work may be seen in the 1957 *Economic Journal* article. Using his more realistic model, Phillips was able to shock the system and then examine various policy responses. As a result, he was able to list what he saw as conditions for optimal stabilization. He argued that design of policies could be extremely intricate, that it was necessary to minimize response delays, that it could be necessary to adjust policy continually and incrementally, and that attention needed to be paid to lags. Phillips speculated that economic agents' expectations could exacerbate anticipated responses to policy—a precursor to later work that pre-dated the Lucas critique—and that wrong policy responses could materially worsen a disequilibrium.

Unburdened by any classical economics education, once again Phillips was well ahead of his time. His experiments had involved fiscal policy, but he noted that, in principle, monetary policy could be a preferable stabilizing instrument. He further noted that he could obtain the best outcome with a combination of interventions based on levels, momentum and accumulation rules. Today, monetary policy at many central banks is informed by Taylor and momentum rules that use this approach.

IV. The Curve

The 1950s had unleashed some years of price and wage inflation. Phillips knew this was an important element in economic stability, and knew also that his models had incorporated prices changes in an unsatisfactory way. In 1958, Phelps-Brown, an LSE colleague, drew Phillips' attention to an interesting but little known historical dataset on UK wages and inflation from 1861 to 1957. Phillips had not generally found his insights in such sources, and he took the data home for what he described as 'a wet weekend's work'. He listed the data on tabular paper and played around with it using slide-rule and basic statistics.

He found that by grouping the data into sub-periods he could detect an interesting, apparently robust, nonlinear relationship between the rates of change of money, wages and unemployment. He wrote a note on his results and some possible interpretations of them. The paper was handed to a colleague, who passed it on to Basil Yamey, Acting Editor of *Economica*. Within a day, Yamey had received positive comments from colleagues and had accepted the paper for publication that year.

Phillips, in the meantime, set off on leave to the University of Melbourne. Cut off from day-to-day communication with the LSE, he was very surprised to come back to confront an outpouring of interest in his work. Academics were fascinated by the wider implications of his results. They saw his paper as important because it pointed to an apparently robust relationship between proxies for nominal and real economic variables, in a way not previously articulated. More than this, acceptance or non-acceptance of the relationship added arguments to the newly emerging Keynesian/monetarist debates.

The Phillips curve became the topic of an intense industry of economic work. Today, there are over half a million Google hits on the subject. The story of its academic development, from Lipsey's further work confirming and extending the results to Friedman and Phelps' incorporation of expectations, is well known and discussed elsewhere (Leeson 2000). Some academic work was very loose in its interpretation of Phillips' original results. Phillips was often given the opportunity to comment on these interpretations, and generally refused. One gets the impression that he knew that the relationship and its implications were more complicated than his note suggested. Indeed, he concluded his article by saying: 'there is a need for much more detailed research into the relationship between unemployment, wage rates, prices and productivity'. He was uncomfortable with the branding of the work as 'the Phillips curve', and did not think that this was part of his more important work.

The article on the Phillips curve stimulated others besides academics. In Britain, wartime planners were now being replaced by new breeds of policy economists and macroeconomic forecasters, alert to using new economic tools to advise their political masters. Some of them saw the Phillips curve as offering a menu of policy options. At its most basic, this was interpreted as a choice between operating at the bottom of the Phillips curve—a 'preserve the pound' strategy that looked attractive to Tories—versus operating at the top of the Phillips curve—a 'go for growth' strategy more attractive to Labour. Clearly, Phillips' original work had never envisaged or suggested such interpretation. But this did not hold back some officials who interpreted the curve as allowing fine-tuning of the economy. In 1964, Ted Callahan, Chancellor of the Exchequer, mentioned the Phillips curve in support of his expansionary budget of that year.

V. Econometric Techniques

As was his way, Phillips was moving on again. The next period in his professional life was spent teaching and researching increasingly technical areas of macroeconomic modelling and econometrics. He was interested in the problems of dynamic disequilibria that required tools and a modelling approach that was then unavailable. It was to this end that he made his contributions to continuous-time modelling, to other econometric problems and to the design and estimation of models (Phillips, P. C. B. 2000; Hendry and Mizon 2000). Laidler (2001) notes:

> The overarching scientific apparatus that he envisaged as the end product of his work would have been an estimated model with control mechanisms fully integrated into its structure, which would have been continuously re-estimated as new data accumulated. Such a model has proved impossible to build, but more of the components that Phillips thought essential to its construction have turned out to be extremely useful, not to say durable, in their own right.

Phillips himself appears to have viewed some of this work as more important than his earlier more famous pieces, and many economists would agree. Interestingly, he published less and less frequently in this period, his last major publication being the 1962 *Economica* article based on his inaugural professorial lecture of some years earlier. He was accused of leaving papers on his desk rather than publishing them. He remained active in his thinking, simulating debate among his colleagues. However, increasingly he avoided the policy limelight and looked to a quieter academic life.

VI. Chinese Development

Phillips was also becoming increasingly unhappy with his personal life in London. How much this was the farmer boy yearning for the open spaces, and how much it was the ex-prisoner of war feeling vulnerable in a stressed world, is debatable. At any rate, the 1960s did not bring the peace of mind that Phillips might have wished for. By 1967 he had published nothing major for five years. That year, Britain was lurching into yet another sterling crisis, and in the precursor to the famous 1968 year of student activism, there was increasing unrest on the LSE campus. This culminated in a major student occupation of the university, sparking eviction by the police, violence and protests.

Phillips reacted strongly to this unrest; he wanted none of it for himself or his family (he now had two young daughters). He burned his papers and left the LSE, taking up a professorship offered at the Australian National University, a high-quality establishment but one outside the bustle of northern academia. A condition that he insisted on was that he could spend half his time working on the Chinese economy.

This seemed an abrupt turnaround in his interests. Development economics in those days was a quite different discipline, with undeveloped theories and little empirical work. China was a mysterious, closed, controlled economy. It had just gone through the (largely unsuccessful) 'Great Leap Forward', and was now entering a period of massive socioeconomic disruption known as the Cultural Revolution. Data were almost unavailable. With very few exceptions, serious economists did not waste their time there, leaving the field to activist historians.

Once again, Phillips was a pioneer. He spoke some Mandarin from his prisoner of war days, and he had always had a fascination for Eastern cultures. Phillips never completed his China work, and it is fascinating to speculate where it might have gone if he had. We know from his surviving notes and from colleagues' reminiscences only that he spent considerable time (together with an old Russian colleague) translating into English the few official reports on Chinese agriculture that he could access, and making detailed notes on Chinese work teams, with occasional speculations about the potential of this economy to one day grow, given the right conditions. He was almost alone in working on this subject.

Phillips was small and lightly built, prematurely aged by his wartime experience and his nicotine addiction. In 1969 he suffered the first of a series of strokes that affected him badly. He returned to New Zealand a sick and fatigued man. Several years later, in 1975, against medical advice he offered to teach a course on the Chinese economy at Auckland University. After only a couple of weeks, he suffered another debilitating stroke and died aged only 61.

Silverstone (2000) records Phillips' natural quietness and modesty about his achievements on the presentation of a festschrift volume commemorating his 60th birthday: 'He was moved greatly and expressed his heartfelt thanks with the comment that he had not really done very much, just "put out a few hares for people to chase".' Some hares! Some chase!

ACKNOWLEDGMENTS

I appreciate the comments of a number of colleagues at the Reserve Bank of New Zealand.

REFERENCES

BARR, N. A. (1973). The Phillips machine returns to life. *LSE Magazine*, no. 46, November.

BERGSTROM, A. R., CATT, A. J. L., PESTON, M. H. and SILVERSTONE, B. (1978). *Stability and Inflation: A Volume of Essays to Honour the Memory of A. W. H. Phillips*. New York: John Wiley.

BLYTH, C. A. (1975). A. W. H. Phillips, M.B.E.: 1914–1975. *Economic Record*, **51**, 303–7.

HENDRY, D. F. and MIZON, G. E. (2000). The influence of A. W. Phillips on econometrics. In R. Leeson (ed.), *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press, pp. 353–64.

LAIDLER, D. (2001). *Phillips in retrospect*. Mimeo: University of Western Ontario Department of Economics.

LEESON, R. (ed.) (2000). *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press.

PHILLIPS, A. W. (1950). Mechanical models in economic dynamics. *Economica*, **17**, 283–305.

——— (1954). Stabilisation policy in a closed economy. *Economic Journal*, **64**(254), 290–323.

——— (1957). Stabilisation policy and the time-forms of lagged responses. *Economic Journal*, **67**, 266–77.

——— (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, **25**, 283–99.

——— (1962). Employment, inflation and growth. *Economica*, **29**, 1–16.

PHILLIPS, P. C. B. (2000). The Bill Phillips legacy of continuous time modelling and econometric model design. In R. Leeson (ed.), *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press, pp. 342–7.

SILVERSTONE, B. D. (2000). The festschrift. In R. Leeson (ed.), *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press, pp. 26–8.

Economica

Economica (2011) **78**, 10–50
doi:10.1111/j.1468-0335.2009.00815.x

The History of the Phillips Curve: Consensus and Bifurcation

By Robert J. Gordon

Northwestern University, NBER and CEPR

Final version received 15 December 2008.

While the early history of the Phillips curve up to 1975 is well known, less well understood is the post-1975 fork in the road. The left fork developed a theory of policy responses to supply shocks in the context of price stickiness in the non-shocked sector. Its econometric implementation interacts shocks with backward-looking inertia. The right fork approach emphasizes forward-looking expectations that can jump in response to anticipated policy changes. The left fork approach is better suited to explaining the postwar US inflation process, while the right fork approach is essential for understanding behaviour in economies with unstable macroeconomic environments.

Introduction

The history of the Phillips curve (PC) has evolved in two phases, before and after 1975, with a widespread consensus about the pre-1975 evolution, which is well understood. Bifurcation begins in 1975, when the PC literature split down two forks of the road, with little communication or interaction between the two forks. The major contribution of this paper, and hence the source of 'bifurcation' in its subtitle, is to examine, contrast and test the contributions of the two post-1975 forks.

The pre-1975 history is straightforward and is covered in Section I. The initial discovery of the negative inflation–unemployment relation by Phillips, popularized by Samuelson and Solow, was followed by a brief period in which policy-makers assumed that they could exploit the trade-off to reduce unemployment at a small cost of additional inflation. Then the natural rate revolution of Friedman, Phelps and Lucas overturned the policy-exploitable trade-off in favour of long-run monetary neutrality. Those who had implemented the econometric version of the trade-off PC in the 1960s reeled in disbelief when Sargent demonstrated the logical failure of their test of neutrality, and finally were condemned to the 'wreckage' of Keynesian economics by Lucas and Sargent following the twist of the inflation–unemployment correlation from negative in the 1960s to positive in the 1970s. The architects of neutrality and the opponents of the Keynesian trade-off emerged triumphant, with two major *caveats* that their own models based on information barriers were unconvincing, and that their core result, that business cycles were driven by monetary or price surprises, floundered without supporting evidence.

After 1975 the evolution of the PC literature split in two directions, each of which has largely failed to recognize the other's contributions. Section II reviews the 'left fork of the road', the revival of the PC trade-off in a coherent and integrated dynamic aggregate supply and demand framework that emerged in the late 1970s in econometric tests, in theoretical contributions, and in intermediate macro textbooks. This approach, which I have called 'mainstream', is resolutely Keynesian, because the inflation rate is dominated by persistence and inertia in the form of long lags on past inflation. An important difference between the mainstream approach and other post-1975 developments is that the role of past inflation is not limited to the formation of expectations, but also includes a pure persistence effect due to fixed-duration wage and price contracts, and lags between changes in crude materials and final product prices. Inflation is dislodged from its past inertial values by demand and supply shocks.

The econometric implementation of this approach is sometimes called the 'triangle' model, reflecting its three-cornered dependence on demand, supply and inertia. Demand is proxied by the unemployment or output gap, and explicit supply shock variables include changes in the relative prices of food, energy and imports, changes in the trend growth of productivity, and the effect of Nixon-era price controls. The triangle approach explains the twin peaks of inflation and unemployment in the 1970s and early 1980s as the result of supply shocks, and provides a symmetric analysis of the 'valley' of low inflation and unemployment in the late 1990s. It emphasizes that inflation and unemployment can be either positively or negatively correlated, depending on the source of the shocks, the policy response and the length of lagged responses.

The right fork in the road is represented by models in which expectations are not anchored in backward-looking behaviour but can jump in response to current and anticipated changes in policy. Reviewed in Section III, important elements in this second literature include policy credibility, models of the game played by policy-makers and private agents forming expectations, and the new Keynesian Phillips curve (NKPC), which derives a forward-looking PC from alternative theories of price stickiness. The common feature of these theories is the absence of inertia, the exclusion of any explicit treatment of supply shock variables, the ability of expected inflation to jump in response to new information, and alternative barriers to accurate expectation formation due to such frictions as 'rational inattention'.

Which post-1975 approach is right? Models in which expectations can jump in response to policy are essential to understanding Sargent's (1982) ends of four big inflations and other relatively rapid inflations in nations with a history of monetary instability, e.g. Argentina. But the mainstream/triangle approach is unambiguously the right econometric framework in which to understand the evolution of postwar US inflation, and the NKPC alternative has been an empirical failure as it has been applied to US data.

Section IV develops and tests the triangle econometric specification alongside one recently published version of the NKPC approach. The latter can be shown to be nested in the former model and to differ by excluding particular variables and lags, and these differences are all rejected by tests of exclusion restrictions. The triangle model outperforms the NKPC variant by orders of magnitude, not only in standard goodness-of-fit statistics, but also in post-sample dynamic simulations.

The scope of this paper is limited to the American theoretical and empirical literature, with the exception of Phillips' (1958) article itself. There are three main interrelated themes in this paper that have not previously received enough attention. First, two quite legitimate responses occurred after 1975 to the chaotic state of the PC. Second, each response is important and helps us to understand how inflation behaves, albeit in different environments. Third, the two approaches need to pay more attention to each other, and this paper represents a start toward that reconciliation.

I. Changing Interpretations of the Phillips Curve, 1958–75

We begin by reviewing the evolution of the PC from Phillips' 1958 article through the development of the Friedman and Phelps natural rate hypothesis and Lucas' introduction of rational expectations. Beyond the scope of this paper are developments before 1958, in particular the many references ably surveyed by Humphrey (1991) dating back to Hume in the mid-eighteenth century regarding the long-run neutrality and short-run non-neutrality of money. The only exception to the 1958 starting cut-off in this paper is Fisher's 1926 article, which anticipates Phillips' relation, albeit interpreting it with the reverse direction of causation.

The Phillips curve is born: Phillips and Samuelson–Solow

The acceptance of new ideas and doctrines is often facilitated if they help to elucidate an outstanding empirical puzzle. Thus the acceptance in the late 1960s of Friedman's natural rate hypothesis occurred rapidly, because it helped to explain the ongoing acceleration of the US inflation rate far beyond the rate forecast by previous research. Likewise, the acceptance of the negative PC a decade earlier was almost immediate, since the PC appeared to resolve an ongoing puzzle about the interpretation of American inflation in the 1950s.

Implicit in pre-Phillips views of US inflation was a 'reverse L' aggregate supply curve, with the joint of the reverse 'L' at a level of economic activity often called 'full employment'. Sustained increases of 'demand-pull' inflation would occur when the economy was operating at a higher level of activity than full employment. But below full employment the inflation rate would be near zero or, at very low levels of activity, even negative as occurred between 1929 and 1933. The early history of the postwar era was reassuring, in that during the recession of 1949 the inflation rate was negative (– 2.0% at an annual rate for the GDP deflator between 1948(IV) and 1950(I)). Then inflation returned during the low-unemployment Korean War years 1950–53 to an extent that had to be suppressed by price controls.

Doubts emerged beginning with the failure of the inflation rate to decline for a single quarter during the 1953–54 recession, followed by its inexorable rise during 1955–57, 'despite growing overcapacity, slack labor markets, slow real growth, and no apparent great buoyancy in over-all demand' (Samuelson and Solow 1960, p. 177). No consensus emerged on the right combination of demand-pull with alternative supply-driven explanations, variously named 'cost-push', 'wage-push' and 'demand-shift'. Into this fractured intellectual atmosphere, the remarkable Phillips (1958) article replaced discontinuous and qualitative descriptions by a quantitative hypothesis based on an unusually long history of evidence. Since 1861 there had been a regular negative relationship in Britain between the unemployment rate and the growth rate of the nominal wage rate. By implication, since the inflation rate would be expected to equal the growth rate of wages minus the long-term growth rate of productivity, there was a regular negative relationship between the unemployment rate and the inflation rate.

Before examining the data, Phillips makes two important theoretical observations. First, the negative relationship between the unemployment rate and the rate of nominal wage change should be 'highly non-linear' due to downward wage rigidity that reflects in turn the reluctance of workers 'to offer their services at less than the prevailing rates when the demand for labor is low and unemployment is high' (1958, p. 283). Second, the rate of change of wages may depend not just on the level of unemployment but also on its rate of change, and subsequently we will discuss the role of this 'rate of change' effect in the context of US postwar models and of the interpretation of the Great Depression.

However, Phillips surprisingly debunks a third possible correlation, that between the rate of change of wages and the retail inflation rate ('working through cost of living adjustments'). He was thinking of a world in which wage rates represented four-fifths of factor costs and import prices the other one-fifth, and normally wage rates and import prices would rise at the same rate. Only when import prices rise five times as fast as productivity growth would retail prices influence wage rates. An interesting note is that Phillips was already thinking of a world in which demand shocks (the level and change of unemployment) and supply shocks (the rate of change of import prices relative to final goods prices) both mattered in determining wage and price changes. However, the role of supply shocks was not fully integrated into PC analysis until the late 1970s.

Most of Phillips' article consists of a set of 11 graphs displaying the rate of change of the nominal wage on the vertical axis and the unemployment rate on the horizontal axis. Graphs are shown for the major sub-periods (1861–1913, 1913–48 and 1948–57) and for each business cycle within the first sub-period. The accompanying text provides an explanation for each point that lies off the fitted regression line, which for 1861–1913 is

$$(1) \qquad w_t = -0.90 + 9.64U_t^{-1.39},$$

where, as in the rest of this paper, upper-case letters are levels, lower-case letters are rates of change, w_t is the rate of change of the nominal wage rate, and U_t is the unemployment rate. Points above the line are identified as years of declining unemployment or rapidly rising import prices, and vice versa. Note the nonlinear formulation and the fact that neither the rate of change effect nor the import price effect is explicitly incorporated into the equation. An econometric representation that included both the level and rate of change effect was soon provided by Lipsey (1960).

Recall that equation (1) is estimated for data only from 1861–1913, and the remaining post-1913 data are plotted against this curve in order to locate episodes when the actual data lie away from the curve. The change in wage rates is remarkably close to the prediction of the 1861–1913 curve except for the two years 1951–52, which were influenced by rapid increases in import prices in 1950–51 resulting from the 1949 devaluation of sterling.

Phillips concludes by translating the fitted curve for wage change into an unemployment–inflation relationship by subtracting long-term productivity growth; it appears that stable prices require an unemployment rate of roughly 2.5%. Notably, Phillips does not conjecture about circumstances in which the apparently stable 1861–1913 curve might shift up or down in the long run. Also, Phillips does not mention policy implications at all, and this provides the setting in which Samuelson and Solow (1960) christen the relationship as the 'Phillips' curve and explore its policy implications.

So widely read and discussed was the Samuelson–Solow article that the term 'PC' entered the language of macroeconomics almost immediately and soon became a lynchpin of the large-scale macroeconometric models which were the focus of research activity in the 1960s. Much of the Samuelson–Solow article provides a critique of the pre-Phillips hypotheses and the difficulty of identifying them.

Then, turning to the Phillips evidence, Samuelson and Solow lament the absence of a similar study for the USA and extract some observations from a scatter plot of US data. First, the US relationship does not work for the 1930s and the two world wars. Second, the implied zero-inflation rate of unemployment is about 3% for the remaining prewar years, similar to Phillips' estimate of 2.5%. Third, there is a clear upward shift in the relationship from the prewar years to the 1950s, and the zero-inflation unemployment rate for the 1950s had risen from 3% to '5 to 6 percent'.

They struggle to explain the postwar upward shift by invoking powerful trade unions that are less 'responsible' than their UK counterparts, and/or the expectation of permanent full employment in the USA. Another conjecture is that the compact size of the UK compared to the USA makes labour markets in the former more flexible. One policy conclusion is that anything that makes US labour markets more flexible will help to shift the PC downwards.

Samuelson and Solow have rightly been criticized for posing a long-run inflation–unemployment trade-off available for exploitation by policy-makers. As the authors conclude: 'We rather expect that the tug of war of politics will end us up in the next few years somewhere in between their selected points. We shall probably have some price rise and some excess unemployment' (Samuelson and Solow 1960, p. 193).

While Samuelson and Solow conclude by warning that the PC relationship could shift over the longer run, their example involves a 'low-pressure' (i.e. high-unemployment) economy in which expectations of low inflation could shift the PC down or could aggravate structural unemployment, thus shifting the PC up. They regard either outcome as possible and notably fail to reason through the long-run implications of a high-pressure economy with its implications of a steady increase in inflation expectations and an associated steady upward shift in the PC. That inference had to wait another eight years for the contributions of Friedman and Phelps.

An interesting side issue is the antecedent of Phillips' article published by Irving Fisher in an obscure journal in 1926, reprinted and brought to a wider audience in 1973.[1] Recall that Samuelson and Solow lament the availability of a detailed statistical study of the USA analogous to Phillips' UK research, yet Fisher had already provided such research more than 30 years earlier.[2] A notable difference with Phillips is that Fisher reverses the direction of causation, so that changes in the rate of inflation cause changes in the level of the unemployment rate. Fisher explains the mechanism in modern textbook terms—because costs of production (including interest, rent, salaries and wages) are fixed in the short run 'by contract or by custom', a faster rate of inflation raises business profits and provides an incentive to raise output. 'Employment is then stimulated—for a time at least' (1973 version, p. 498). Because of the lag of costs behind prices, Fisher emphasizes that the relationship is between unemployment and the inflation rate, not the price level, and that the price level has 'nothing to do with employment'. He uses the analogy of driving, in which it takes more fuel per mile to climb a hill than descend it, but exactly the same amount to navigate a 'high plateau as on the lowlands'.

Fisher's statistical study is limited to monthly data for the years 1915–25. When the influence of inflation is represented by a short distributed lag over five months, the correlation coefficient is 90% between the unemployment rate and the short distributed lag of inflation. An important weakness of the Fisher study is evident in his Chart II but is not discussed by the author. The 90% correlation applies to 1915–25 but his chart extends back to 1903. During the period 1903–15, unemployment is almost as volatile as during 1915–25 but the variance of inflation is much lower, implying that the relationship is not stable and that Fisher's main result may be picking up the special features of the First World War and its aftermath (just as Phillips' UK correlation is strong during the First World War).

Aspects of Phillips curve economics in the 1960s

During the early to mid-1960s, at least three aspects of the PC emerged that would have subsequent consequences. First, the PC trade-off appeared to provide policy-makers with a menu of options. The policy advisors of the Kennedy and Johnson administrations, led by Walter Heller with support roles by Robert Solow and James Tobin, argued that the previous Republican administration had chosen a point too far south-east along the PC trade-off, and that it was time to 'get the country moving again' by moving to the north-

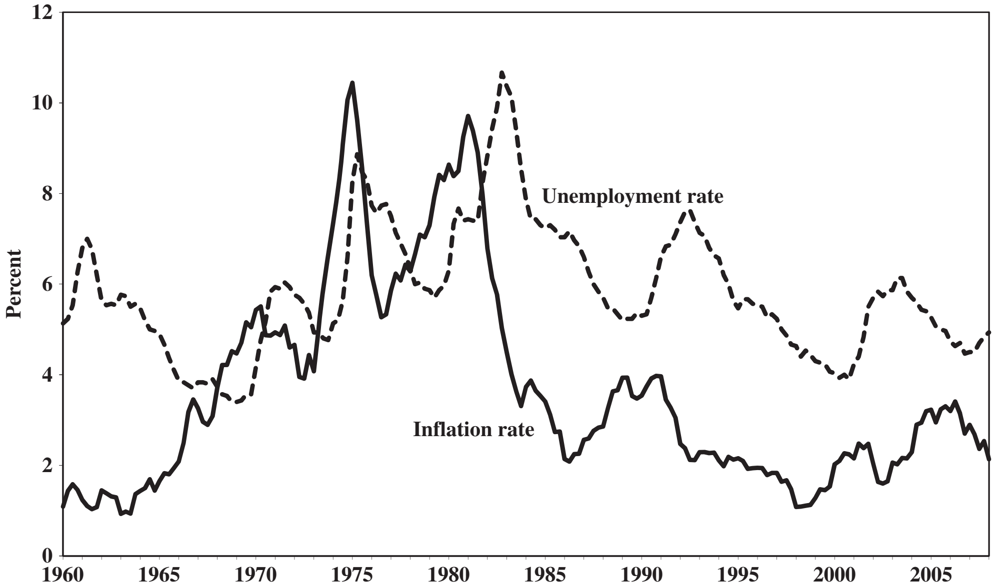

FIGURE 1. The unemployment and inflation rates, quarterly data, 1960–2007. (Source: US Bureau of Labor Statistics (www.bls.gov) and US Bureau of Economic Analysis (www.bea.gov).)

west. Heller's group convinced President Kennedy to recommend major cuts in Federal income taxes, and these were implemented after his death by the Johnson administration in two phases during 1964 and 1965. However, in late 1963 the economy was already operating at an unemployment rate of 5.5% that Samuelson and Solow had calculated was consistent with zero inflation, and so the expansionary Kennedy–Johnson fiscal policy would have implied an acceleration of inflation even without the further loosening of the fiscal floodgates due to the Vietnam War.

Figure 1 plots the US inflation and unemployment rates in quarterly data since 1960, and we shall refer to it here to examine the period 1960–71 and then return to the same graph below to link the evolution of PC debates to the post-1971 behaviour of inflation and unemployment.[3] The unemployment rate fell below 5.5% in 1964 and remained below 4% between 1966 and 1970. The sharp acceleration of inflation from less than 2% in 1963 to 5.5% in 1970 is consistent with current econometric estimates of the 1963 natural rate of unemployment (the rate that is consistent with steady inflation rather than zero inflation) in the range of 5.5% to 6.0% (see Figure 5 below).

A second aspect of this period was the development of mainframe electronic computers that made it practical for the first time to specify and estimate large-scale econometric models (a book-length policy analysis using the Brookings model is contained in Fromm and Taubman 1968). The specification of the inflation process in these models always consisted of at least two equations. The PC was embodied in an equation for the rate of change of the nominal wage in which the main explanatory variables were the unemployment rate, sometimes its rate of change, some measure of expected inflation based on a backward-looking set of lags, and perhaps various tax rates.

Then the estimated change in wages was typically translated into the inflation rate in an equation that related the price *level* to the wage *level* adjusted for the level of trend productivity, the so-called 'trend unit labour cost'. The price–labour cost ratio or 'mark-up' was allowed to respond to a measure of demand, usually not the unemployment rate but rather a measure more directly related to the product market, such as the ratio of unfilled orders to shipments. The reduced form of this approach implied that the inflation rate depended on the level and rate of change of unemployment, perhaps other measures of demand, and lagged inflation. We return below to the problems encountered by these models in confronting the data of the late 1960s and in dealing with the challenge of the Friedman–Phelps natural rate hypothesis.

A third, albeit peripheral, feature of this era was the rivalry between the economics departments at the University of Chicago and MIT in general, and between Milton Friedman and Franco Modigliani in particular. In 1965 more than 100 pages in the *American Economic Review* were devoted to a debate between them and their co-authors over the issue of whether 'only monetary policy mattered' or 'only fiscal policy mattered', a debate that seemed bizarre when the consensus view based on the IS-LM model showed that both monetary and fiscal policy mattered except in certain extreme cases.[4]

The natural rate revolution

Prior to the publication of the Friedman and Phelps articles, theoretical questions had been raised about the PC framework. Why did the nominal wage adjust slowly, particularly in a downward direction, and what determined the speed with which it responded to inadequate demand? Why did the PC lie so far to the right, that is, why did nominal wages rise so fast at a low unemployment rate, and why was such a high unemployment rate required to maintain zero inflation? Perhaps most relevant in anticipation of Friedman and Phelps, how could the PC be so stable over history when there were so many episodes of hyperinflations fuelled by permissive monetary and fiscal policy? I have always thought that the development of the natural rate hypothesis at Chicago, rather than at Harvard or MIT, reflected the deep involvement of several Chicago economists as advisers to several countries in Latin America, where the lack of correlation between inflation and unemployment was obvious.

Friedman's (1968) Presidential Address contained two sections that each had a main point, closely interrelated. First, the central bank could not control the nominal interest rate if that implied faster inflation, because the implied reduction in the real interest rate would add fuel to the inflationary fire. The second section was most important for the PC debate, his then-startling conclusion that policy-makers had no ability to choose any unemployment rate in the long run other than the natural rate of unemployment, the rate that would be 'ground out' by the microeconomic structure of labour and product markets. A more practical interpretation of the natural rate was the unemployment rate consistent with accurate inflation expectations, which implied a steady rate of inflation.

Conventional analysis based on a policy trade-off ignored the adjustment of expectations. Consider an economy operating at the natural rate of unemployment and with an initial inflation rate of 1% that was accurately anticipated. Any policy-maker attempting to reduce the actual unemployment rate below the natural rate would move the economy north-west along the short-run PC, pushing the unemployment rate lower but the actual inflation rate higher. Once agents notice that the actual inflation rate is higher than the initially anticipated rate of 1%, expectations will adjust upward and shift the entire short-run PC higher. This process will continue until the unemployment rate rises back to the natural rate of unemployment.

The timing of Friedman's address was impeccable and even uncanny. The Kennedy–Johnson fiscal expansion, including both the tax cuts and Vietnam War spending, accompanied by monetary accommodation, had pushed the unemployment rate down from 5.5% to 3.5%, and each year between 1963 and 1969 the inflation rate accelerated, just as Friedman's verbal model would have predicted. The acceleration of inflation bewildered the large-scale econometricians, who had previously estimated a 'full employment' unemployment rate of 4% and whose forecasts of inflation had been exceeded by the actual outcome year after year.

Well aware of their own failure to forecast the late 1960s acceleration of US inflation, Friedman's detractors attacked the verbal model that Friedman used to motivate the natural rate hypothesis. In what later became known as the 'fooling' model, Friedman postulates employers with expectations of the price level that are always accurate, but workers with an expected price level that does not respond until after a substantial lag to a higher actual price level. In a business expansion, firms raise the wage but raise the price level by more, thus reducing the real wage as needed to provide the incentive to hire additional workers. But workers see the higher nominal wage and interpret it as a higher actual real wage, because they fail to adjust their expectation of the price level. Friedman's model was attacked as grossly implausible, because workers have access to monthly announcements of the Consumer Price Index (CPI) and indeed observe actual prices as they shop almost every day. In Friedman's world, there could be no business cycle.

Phelps (1967, 1968) is credited with co-discovering the natural rate hypothesis. In contrast to Friedman's distinction between smart firms and dumb workers, in Phelps's world everyone is dumb, i.e. equally fooled. Both firms and workers see the price rise in their industry and produce more, not realizing that the general price level has risen in the rest of the economy. Phelps developed one model in which workers are isolated from information about the rest of the economy. Normally there is frictional unemployment, as workers regularly quit one firm to go look for more highly paid work at other firms. But in a situation in which their own firm raises the wage, they stay with that firm instead of quitting. Thus the unemployment rate decreases even though, without their knowledge, all other firms in the economy have raised the wage by the same amount at the same time. The workers are fooled into a reduction in frictional unemployment, and the macroeconomic data register a decline in the unemployment rate. Hence there is a short-term correlation between the rate of wage change and the unemployment rate, but this lasts only as long as expectations are incorrect.

Whether firms or workers or both are fooled, the criticisms directed against the Friedman fooling model apply to Phelps as well. Workers and their employers buy many goods and services frequently; they obtain news on the CPI every month; and perhaps most important if periods of high real GDP and low unemployment had always been accompanied by an increase in the aggregate price level, workers and firms learn from these past episodes and use their experience to form expectations accurately.

Rational expectations and the 'policy ineffectiveness proposition'

Both the Friedman and Phelps models were based on the twin assumptions of continuous market clearing and imperfect information. Soon thereafter, in two influential articles, Lucas (1972, 1973) extended their model by adding a third component: rational expectations. Workers and firms use their knowledge of past history to work out the implications of an observed fall or rise in wages on the overall wage level. Rational expectations imply that erroneous expectations errors are not repeated.

Lucas collapsed the distinction between firms and workers and treated all economic agents as 'yeoman farmers' who face both idiosyncratic shocks to their own *relative* price and macro shocks caused by fluctuations in monetary growth and other factors. The agents use rational expectations to deduce from past history how much of an observed change in the local price represents an idiosyncratic shock and how much represents a macro shock. When local price shocks have a high correlation with macro shocks, agents do not adjust production, knowing that no change in relative prices has occurred. Lucas used this insight to explain why the PC in a country like Argentina with high macro volatility would be much steeper than in a country like the US with low macro volatility.

The concept of rational expectations led Lucas and his followers to make a startling prediction. He argued that *anticipated monetary policy cannot change real GDP in a regular or predictable way*, a result soon known as the 'policy ineffectiveness proposition'. In common with Friedman and Phelps, the Lucas approach implied that movements of output away from the natural level require a price surprise, so that the central bank can alter output not by carrying out a predictable change in monetary policy but only by creating a surprise. (The formal development of the proposition was carried out in Sargent and Wallace 1975.)

By the end of the 1970s the Lucas approach was widely criticized. The problem was not Lucas' introduction of rational expectations, but rather the twin assumptions inherited from Friedman and Phelps, namely continuous market clearing and imperfect information. Deviations of the current actual price level from the expected price were the *only* allowable source of business cycle movements in real GDP. Thus, despite the widespread appeal of the Friedman–Phelps–Lucas approach, it ran aground on the shoals of an inadequate theory of business cycles. With monthly information available on the aggregate price level, the business cycle could last no more than one month.[5] In the recent evaluation of Sims (2008, p. 4), the microeconomic underpinnings of the Lucas supply curve were 'highly abstract and unrealistic—for example models of "island economies" in which people had to infer the value of the economy-wide interest rate or money stock from the price level on their own island'. Even Lucas later confessed that: 'Monetary shocks just aren't that important. That's the view I've been driven to. There's no question that's a retreat in my views' (Cassidy 1996, p. 53).

Rejection of the empirical case against monetary neutrality

Whatever the model used to explain the business cycle, the natural rate hypothesis and long-run monetary neutrality are intact if empirical coefficients imply that a reduction of the unemployment rate below a certain level (whether it is called the natural rate or the full employment rate) leads to continuously accelerating inflation. In the first few years after the Friedman and Phelps articles, those who had developed econometric models supporting a permanent long-run trade-off claimed that the validity of long-run neutrality could be tested by estimating whether the sum of coefficients on the lagged dependent variable in an inflation equation was equal to unity or was significantly below unity. Here we ignore the distinction between wage and price changes and examine the relationship between the inflation rate (p_t), its lagged value (p_{t-1}), and the unemployment rate (U_t):

(2) $$p_t = \alpha p_{t-1} + \beta U_t + e_t.$$

Here the response of inflation to unemployment is negative ($\beta < 1$). If the sum of coefficients on lagged inflation is significantly below unity, then in the long run when $p_t = p_{t-1}$ there is a long-run trade-off between inflation and unemployment:

(3) $$p_t = \beta U_t / (1 - \alpha).$$

Numerous research papers written in the late 1960s and early 1970s placed major emphasis on the finding that the α coefficient was significantly below unity, implying a permanent trade-off as in equation (3). However, these results were ephemeral and quickly abandoned for two reasons. First, as the sample period extended over more of the period of accelerating wage and price change in the late 1960s, the α coefficient kept creeping up and by 1972 had reached unity, particularly when the coefficient was allowed to vary over time.[6]

The second and more important reason to abandon this test of the long-run trade-off was Sargent's simple but devastating econometric point. Here we simplify Sargent's exposition by suppressing the difference between wages and prices, and by making expected inflation depend on only a single lag of inflation rather than a distributed lag. The original specification is not (2) but rather

$$(4) \qquad p_t = \alpha E p_t + \beta U_t + e_t,$$

where Ep_t is the expected rate of inflation. An observable proxy for expected inflation must be obtained, and this requirement is satisfied by backward-looking or adaptive expectations:

$$(5) \qquad Ep_t = v p_{t-1}.$$

When (5) is substituted into (4), we obtain

$$(6) \qquad p_t = \alpha v p_{t-1} + \beta U_t + e_t.$$

Now Sargent's point becomes clear: the single equation (6) cannot be used to estimate both α and v. The only way that α can be interpreted as the coefficient on expected inflation is for an extraneous assumption to be introduced, in particular that $v = 1$.

Yet, Sargent argues, there is no reason for v to be unity, and rather if the inflation rate can be approximated as a covariance-stationary stochastic process, v must be less than unity. For v to be unity, the inflation rate would display extremely strong serial correlation or 'drift', but during much of US history before 1950 the inflation rate displayed relatively little serial correlation. Thus it is quite possible that α was equal to unity throughout the postwar era but that v gradually increased with the serial correlation of inflation in the 1960s that was higher than in the 1940s and 1950s. In short, Sargent made a convincing case that the previous econometric estimates of α in the context of equation (2) had no relevance to the validity of the natural rate hypothesis. Not surprisingly, such econometric exercises ceased quite abruptly after 1972.

Sargent's observation that the v coefficient should be smaller in periods with less serial correlation of the inflation rate was subsequently validated. Gordon developed quarterly data back to 1892 and showed that the sum of coefficients on lagged inflation rose from 0.40 in 1892–1929 to 0.60 in 1929–53 and then to 1.06 in 1954–80 (Gordon 1982a, Table 3). We return to his results below, because they directly address the shifting form of the PC relationship during the two world wars and during the Great Depression that was originally noticed by Samuelson and Solow (1960).[7]

II. The Post-1975 Left Fork in the Road: the Dynamic Demand–Supply Model with Inertia

The 1960s were the glory years of the PC's interpretation as a negative correlation between inflation and unemployment, initially as incorporating a permanent negative trade-off, and subsequently as a significant short-run trade-off subject to the longer-run

FIGURE 2. Scatter plot of the unemployment and inflation rates, quarterly data, 1960–80.

adjustment of expectations in the natural rate PC. But almost from the beginning, the decade of the 1970s seemed to overturn any thought that the negative PC trade-off was intact or stable. The nature of the problem is evident when we look again at Figure 1, which plots the inflation and unemployment rates in quarterly data since 1960, with the four-quarter change in the deflator for personal consumption expenditures (PCE) used to represent inflation. For the 1970s as a whole, the inflation–unemployment correlation is strongly positive, not negative, and in Figure 1, sharp changes in the inflation rate appear to *lead* by about one year changes in the same direction of the unemployment rate.

When plotted in Figure 2 on a scatter plot from 1960 to 1980, the inflation and unemployment rates are uncorrelated, with a combination of negative and positive correlations that range all over the map. The negative PC trade-off appeared to be utterly defunct. In flowery language that amounted to a simultaneous declaration of war and announcement of victory, Lucas and Sargent (1978, pp. 49–50) described 'the task which faces contemporary students of the business cycle [is] that of sorting through the wreckage . . . of that remarkable intellectual event called the Keynesian Revolution'.

The year 1975 marks a clear break in the history of the PC. Surveys written at the time focus on the demise of the short-run trade-off and the emergence of the consensus expectational natural rate PC (see, for instance Laidler and Parkin 1975). Two complementary reasons lead us to mark 1975 as the transition year for PC doctrine. First, it was the year of the publication of the policy ineffectiveness proposition summarized above, which was the beginning of the end of business cycle theory based on expectation errors. Second, 1975 was a year in which both the US inflation and unemployment rates experienced the maximum impact (at least up to that time) of supply shocks, calling for a revised PC theory that explicitly incorporated supply shocks.

The demise of the Friedman–Phelps–Lucas information barriers model occurred in two stages. First, the theory was flawed by its inability to reconcile multi-year business cycles with one-month lags faced by agents in obtaining complete information about the aggregate price level. Second, the attempt to develop an empirical counterpart of the policy-ineffectiveness proposition was a research failure. It floundered on the inability to develop a symmetric explanation of output and price behaviour. Barro (1977) showed that output was not related to anticipated monetary changes but could not demonstrate the required corollary—the full and prompt responsiveness of price changes to anticipated nominal disturbances. This failure reflected the fundamental conflict between the fully flexible prices required by the information barriers model and the inflation inertia deeply embedded into the US inflation process, a conflict that has returned to haunt the application of the NKPC approach in the past decade. Soon Mishkin (1982) and Gordon (1982a) showed that anticipated monetary changes had a strong effect on output in the short run and on inflation in the long run, preserving long-run but not short-run neutrality.

Since 1975 the development of PC doctrine has bifurcated into two divergent paths, called here the 'left fork' and 'right fork' of the road, with no sign of convergence. The left fork in the road, treated in this section, is the resurrection of Keynesian economics in the form of what I call the 'mainstream' PC model that incorporates long-run neutrality, that incorporates explicitly the role of supply shocks in shifting the PC up or down, and that interprets the influence of past inflation as reflecting generalized inertia rather than expected inflation. The right fork in the road of the post-1975 evolution, examined in Section III, features an approach developed by Kydland, Prescott and Sargent, and more recently by Galí, Gertler and others. Inflation depends on forward-looking expectations, and expectations respond rationally to actual and expected changes in monetary and fiscal policy. This two-way game between policy and expectation formation leaves no room for supply shocks or inertia.

The resurrection of the PC

Several years *before* the famous 'wreckage' pronouncement by Lucas and Sargent, the resurrection of the PC began. The first and perhaps most important element was the new theory of policy responses to supply shocks, developed independently by Gordon (1975) and Phelps (1978) in two slightly different models that were later merged by Gordon (1984). The 'Gordon–Phelps' model starts from the proposition that the price elasticity of demand of the commodity experiencing the adverse supply shock, e.g. oil, is less than unity, so that following an increase in the relative price of oil, the expenditure share of that commodity must increase and the expenditure share of all other components of spending must decrease. For instance, energy's share of nominal US GDP tripled between 1972 and 1981.[8]

The required condition for continued full employment is the opening of a gap between the growth rate of nominal GDP and the growth rate of the nominal wage (Gordon 1984, p. 40) to make room for the increased nominal spending on oil. If nominal wages are flexible, one option is for the growth rate of wages to become negative, allowing the growth rate of nominal GDP to remain fixed. At the alternative extreme with rigid wages, to avoid a decline in non-energy output, an accommodating monetary policy must boost nominal GDP growth by the amount needed to 'pay for' the extra spending on oil, but this will lead to an inflationary spiral if expectations respond to the observed increase in the inflation rate. A third alternative, and the one that actually occurred in the 1970s, was a combination of wage rigidity with a partial response of nominal GDP growth, pushing down both real non-energy spending and employment.

By 1976 this model had made its way into the popular press when a *New York Times* headline announced, 'A new theory: inflation triggers recession' (18 July 1976, p. F13). Indeed, we can see in Figure 1 that throughout the period 1974–81, there was a time *lead* of roughly one year of inflation relative to unemployment. This real-world result, that an adverse supply shock can depress real output and employment in a world of sticky non-oil prices, had been christened by Okun in 1974 conversations as a 'macroeconomic externality'.[9]

Sometimes the output effect of the supply shock in the Gordon–Phelps framework is likened to an 'oil tax' that reduces non-oil real consumption by more, the smaller is the price elasticity of demand for oil. The extent of the resulting decline in real output depends not only on that elasticity but on the response of nominal demand, which in turn depends not just on the response of monetary policy but also on additional factors listed by Blinder and Rudd (2008)—bracket creep in a non-indexed tax system, negative wealth effects, scrappage of obsolete capital, and the effect of uncertainty in dampening demand.

By 1977 supply shocks had been incorporated into the natural rate expectational Phillips curve. This theoretical formulation (Gordon 1977a, equation 13), except for the absence of explicit lagged terms, is identical to the econometric 'mainstream' model developed subsequently and described below:

$$(7) \qquad p_t = Ep_t + b\left(U_t - U_t^N\right) + z_t + e_t,$$

where the notation is the same as above with the addition of U_t^N to represent the natural rate of unemployment and z_t to represent 'cost-push pressure by unions, oil sheiks, or bauxite barons' (Gordon 1977a, p. 133). Other types of supply shocks include the imposition and termination of price controls (as in the US in 1971–74), changes in the relative price of imports, and changes in the trend growth of productivity. Episodes in which political events cause sharp changes in wages, such as the French general strike of 1968, also qualify as adverse supply shocks. A detailed narrative of the role of food, oil and price-control shocks in the inflation of the 1970s is provided by Blinder (1979, 1982).

The process of integrating supply shocks into macroeconomics took place during 1975–78 simultaneously on three fronts: theoretical as described above, empirical as described below, and in an unusual development, through a new generation of intermediate macroeconomic textbooks. An explanation was needed to reconcile the dominant role of demand shocks as the explanation of the Great Contraction of 1929–33 in the same model as would explain the positive correlation of inflation and unemployment in 1974–75. Once recognized, that explanation became obvious. Just as the output and price of corn or wheat could be positively or negatively correlated depending on the importance of micro demand or supply shocks, so aggregate output and the rate of inflation could be positively or negatively correlated, depending on the relative importance of aggregate demand or supply shocks.

The textbooks appeared simultaneously in 1978, and both used alternative versions of a simple diagram that can be traced back to a classroom handout used by Dornbusch at the Chicago Business School in early 1975.[10] The diagram, which has the inflation rate on the vertical axis and either the unemployment or output gap on the horizontal axis, combines three elements—the expectational PC, shifts in that PC caused by supply shocks, and an identity that decomposes nominal GDP growth into inflation and output growth. The textbook version shows that the dynamic aggregate demand–supply model implies a simple first-order difference equation. Following a permanent upward or downward shift in nominal GDP growth, any lags in the formation of expected inflation cause the economy to cycle through loops to its new long-run equilibrium at a zero value of the unemployment or output gap and a permanently higher or lower rate of inflation.

Econometric implementation of the mainstream model

As in equation (7) above, the mainstream specification of the inflation process contains three sets of explanatory variables representing inertia, demand and supply, leading me to call it the 'triangle' model.[11] Replacing the expected inflation term is a set of long lags on past inflation, reflecting the view that the influence of past inflation reflects generalized backward-looking inertia, not just the formation of expectations. Important sources of inertia include the set of explicit and implicit contracts that dampen short-term changes in prices and wages (as recognized explicitly by Fisher 1926), and the input–output supply chain that creates thousands of links of unknown magnitude and duration between changes in crude and intermediate goods prices and the prices of final goods, as emphasized by Blanchard (1987). All of these channels interact to create the 'inertia' effect, the first leg of the triangle.

This approach is Keynesian because the role of inertia is to make the inflation rate slow to adjust to changes in nominal demand, and as a result real GDP emerges as a residual, not as an object of choice as in the Friedman–Phelps–Lucas model. A vast theoretical literature under the rubric of 'new Keynesian economics' (NKE), starting in the late 1970s with Fischer (1977) and Taylor (1980), provided numerous models to motivate the inertia mechanism by explaining real and nominal rigidity of wages and/or prices, and many of these explicitly incorporated rational expectations.[12] In our discussion in Section III, we will be careful to distinguish between the theoretical models of the NKE and the empirical application of the NKPC.

In the triangle model, the speed of price adjustment and the speed of expectation formation are two totally different issues. Price adjustment can be delayed by wage and price contracts, and by the time needed for cost increases to percolate through the input–output table, and yet everyone can form expectations promptly and rationally based on full information about the historical response of prices to its own lagged values, to demand shocks and to supply shocks.

The demand side of the model is represented by the level and change of the output gap or alternatively the unemployment gap. As we have seen above, Phillips recognized the role of the 'rate of change' effect that at any given unemployment rate makes the inflation rate higher when the unemployment rate is falling than when it was rising. Because the unemployment gap is always entered in the triangle model with both the current value and with additional lags, the zig-zag of the estimated lagged coefficients between negative and positive incorporates the rate of change effect.

The supply side of the model is represented by a set of explicit supply shock variables, establishing a contrast between the mainstream approach and the recent NKPC literature where the supply shock effects are always hidden in the error term. The explicit supply shock variables are all defined so that the absence of supply shocks is represented in (7) as $z_t = 0$. Such variables, for instance, include changes in the *relative* price of oil and changes in the *relative* price of non-oil imports; when these relative prices exhibit zero change, there is no upward or downward pressure on the inflation rate from supply shocks. The list of supply shock variables includes dummy variables which measure the impact of the 1971–74 Nixon-era price controls in holding down inflation in 1971–73 and then adding to the supply shock impact on inflation in 1974–75.[13] Earlier versions (Gordon 1982b; Gordon and King 1982) included changes in the real exchange rate in place of real import prices.

Unfortunately, the essential role of sticky wages and/or inflexible non-oil prices was missed by many analysts who later tried to model the impact of oil prices on real output. For instance, Bruno and Sachs (1985) use a neoclassical production function to show that the elasticity of output with respect to the real energy price is the energy share in gross output, thus missing the macroeconomic externality. Hamilton (1983), in a much-cited paper, showed that oil prices Granger-cause changes in real output in all but one of the recessions that occurred between 1948 and 1980. Hamilton's results cannot be compared to those of Bruno and Sachs, or those implied by the triangle inflation equation, because he provides no elasticity estimates and no analysis of the extent to which the oil price effect works through overall inflation, as in the Gordon–Phelps model, or through a direct impact of oil on output via the production function, as in Bruno–Sachs.

Since the 1970s the literature on supply shocks has come increasingly to focus on oil and to neglect shocks related to changes in the relative price of food, of non-oil non-food imports, and the effects of the Nixon-era price controls. In fact, food rather than oil was the example used in the initial development of the macroexternality theory in Gordon (1975). Bosworth and Lawrence (1982) provide ample background on the reasons for sustained increases in the real price of food in 1973–74 and in 1978–79 (another such episode occurred in 2007–08)

Blinder and Rudd (2008) revisit the supply shock explanation of the 'Great Stagflation' of the 1970s and early 1980s, and confirm its central role. They summarize a set of arguments against the supply shock explanation, and refute each. To those (including Barsky and Kilian 2002) who cannot understand why a change in a *relative* price would be relevant for overall inflation, they point to the rigidity of prices in the non-shocked sector. They also assess arguments (like those of Bernanke *et al.* 1997) claiming that the impact of oil shocks on the economy actually represents the effects of the central bank response rather than the oil shock itself. They provide new evidence from a structural vector autoregressive regression (VAR) model reaffirming an independent role for oil shocks. In fact, given ample empirical evidence of long lags in the response of output and unemployment to monetary policy actions, the Blinder–Rudd results make perfect sense—an adverse supply shock causes an initial spike of unemployment, and the monetary policy response then determines by how much unemployment declines in the subsequent years after the shock.

Since the original Phillips (1958) article was about wage changes, not price changes, it is noteworthy that the triangle model is a single reduced-form equation for the inflation rate, with no mention of wage changes. Starting from separate wage change and price mark-up equations, as had been standard in the PC econometrics literature up to that time, Gordon (1982b) merged the two and discussed the simplifying assumptions needed to perform the merger, particularly the absence of wage–wage inertia.

The usual assumption that inflation is equal to nominal wage changes minus productivity growth assumes a fixed value of labour's share in national income. But labour's income share rose sharply in the late 1960s and has drifted down since then. The goal of the central bank is to control inflation, not wage changes, so changes in labour's income share across business cycles imply a loose relation between inflation and wage changes that is fruitfully ignored. An important contribution to the demise of the wage equation was made by Sims (1987), who argued that wage and price equations have no separate structural interpretations, and that a price equation is a wage equation stood on its head, and vice versa.

Empirical results: strengths and weaknesses

The current econometric version of the mainstream or triangle model was originally developed in the late 1970s (Gordon 1977b) and published in its current form (as a single reduced-form price-on-price equation with no wages) in Gordon (1982b). It has been maintained essentially intact since then, with the same set of explanatory variables and lag lengths, in order to allow post-sample simulations to identify forecasting errors that may call for rethinking the specification. The first challenge to the model arrived almost immediately in the form of the Volcker disinflation of 1979–86. As shown in Figure 1, the inflation rate collapsed from nearly 10% in 1981 to only 3% in 1983–84, much faster than had been forecast by commentators using an expectational PC with a heavy emphasis on wage rigidity.

The ‘sacrifice ratio’ is a convenient summary measure of the speed of inflation adjustment in response to high unemployment and low output. This ratio is defined as the cumulative years of output gap during the disinflation divided by the permanent reduction of inflation expressed as an absolute value. Some *ex ante* forecasts of the sacrifice ratio made in 1980–81 were as high as 10, but the actual sacrifice ratio in retrospect turned out to be between 3.5 and 4.5.[14] The key to the surprisingly low sacrifice ratio turned out to be the role of supply shocks, and in particular the 1981–86 decline in the relative price of energy, and the 1980–85 appreciation of the dollar that reduced the relative price of imports. In a remarkable forecasting success achieved in the middle of the disinflation, Gordon and King (1982) estimated a six-equation VAR model that combined the triangle inflation equation with equations that allowed monetary policy to influence endogenous oil prices and the exchange rate, and their main result was a sacrifice ratio in the range of 3.0 to 3.5, much below the prevailing wisdom of the time.[15]

The Gordon–King result is consistent with the Kydland–Prescott–Sargent interpretation—reviewed in the next section—that makes no mention of supply shocks but rather emphasizes the interplay between the credibility of the central bank and the expectations of the public. No doubt a major role in the speed of the disinflation, and the resulting relatively small sacrifice ratio, was the widespread perception that the Fed’s monetary policy changed after 1979 and its anti-inflation stance became more credible than before. The advantage of the Gordon–King method is that the channels of monetary policy are explicitly traced, not just through high unemployment but also through the effect of the monetary–fiscal policy mix in causing an appreciation of the dollar in 1980–85, with an accompanying decline in the relative price of oil and of non-oil imports.

Returning now to Figure 1, we see that the Volcker disinflation was followed in the late 1980s by a repeat of the negative inflation–unemployment trade-off already experienced in the 1960s, albeit with a smaller acceleration of inflation and a higher level of unemployment. Similarly, the negative trade-off is evident in the slowdown of inflation in 1990–93 in response to a marked increase in the unemployment rate during the same period.

At first glance, the behaviour of the PC in the 1990s appears to be puzzling. Unemployment in the late 1990s fell to the lowest rate since the 1960s, but there was no parallel acceleration of inflation. Instead, inflation was lower in 2000 than in 1993. As shown by Gordon (1998), low inflation in the late 1990s can be explained by beneficial supply shocks that pushed the PC down in contrast to the adverse supply shocks of the 1970s; the beneficial shocks of the 1996–99 period included lower real energy prices, lower relative import prices, and faster trend productivity growth. As shown in Figure 1, the ‘twin peaks’ of inflation and unemployment were joined by the ‘valley’ of inflation and unemployment during 1997–2000.

Despite these research successes, the evolution of the data required one change in the 1982 specification of the triangle model. Post-sample simulations in 1994–95 revealed that the model’s predictions had started to drift in the direction of predicting too much inflation, given actual values of the unemployment gap. These errors turned out to be due not to a flaw in the model but to a data choice, that is, the false assumption that the natural rate of unemployment was fixed, allowed to change only in response to the demographic composition of the unemployment rate.[16]

For several decades the natural rate of unemployment has been called by its nickname, the 'NAIRU', standing for 'non-accelerating inflation rate of unemployment'. The time-varying NAIRU (or TV-NAIRU) combined an econometric method introduced by Staiger *et al.* (1997) that was applied to a version of my triangle model, and simultaneously I published a paper which used their method applied to my model (Gordon 1997). The estimated TV-NAIRU exhibited a pronounced downward drift after 1990 that explained in a mechanical way why the inflation rate was lower in the 1990s than had previously been predicted with a fixed NAIRU. An initial set of substantive explanations of this decline in the NAIRU was provided by Katz and Krueger (1999).

III. The Post-1975 Right Fork in the Road: Jumping and Forward-looking Policy-responsive Expectations

The alternative post-1975 research approach, the right fork of the road, emphasizes jumps in expectations in response to policy actions and implicitly incorporates price flexibility, market clearing and an absence of backward-looking inflation inertia. The central idea that expected inflation can jump in response to actual or anticipated policy changes is crucial to understanding the ends of hyperinflations (Sargent 1982). It begins with the basic proposition, already embedded in the Friedman–Phelps natural rate hypothesis, that the choice by a policy-maker of a particular short-run combination of inflation and unemployment rates can alter expectations, causing the trade-off to change.

The policy game

Kydland and Prescott (1977) distinguished between policy discretion and rules, contrasting discretionary policy-makers who reassess the desired response to alternative inflation rates in each successive time period, with rule-following policy-makers who adhere to a rule which is fixed for all future time periods. They show, not surprisingly, that the long-run inflation rate is higher under a discretionary policy than under a rules-based policy. How does this approach explain the positive correlation of inflation and unemployment in the 1970s without mention of supply shocks? Papers written by Sargent (1999), Cogley and Sargent (2005) and Sargent *et al.* (2006) begin with the standard presumption that choices by discretionary policy-makers will cause the PC to shift and policy options to change. The attempt to conduct policy without knowledge of the current position of the Phillips curve can lead a policy-maker to make choices that yield persistently high inflation outcomes.

'Credibility' is an important concept in the game involving policy-makers and private agents. Because expectations can jump in response to changes in policy-makers' actions and perceived intentions, the outcome of actual inflation is higher if agents infer that the policy-maker is trying to manipulate unemployment along the short-run PC trade-off. A credible policy is one which promises to maintain a low inflation rate in the long run; agents are convinced that a policy is credible if the policy-makers pursue an inflation target and regularly raise the interest rate when inflation exceeds its target but do not lower interest rates in response to an increase in unemployment. Doubts by agents that the policy-maker is committed to low inflation in the long run can raise the unemployment cost of reducing inflation, i.e. the sacrifice ratio.

One problem with this line of research is that it ignores additional information available to policy-makers—that oil or farm prices have risen, that the dollar has been devalued, that price controls have been imposed or ended, or that trend productivity growth has slowed or revived. Indeed, it is striking that Sargent *et al.* (2006) claim to be able to explain the entire upsurge of inflation in the 1970s and early 1980s without any mention of supply shocks, despite the fact that the word 'shocks' appears in the title of their paper: 'allow the model to reverse engineer a sequence of government beliefs about the Phillips curve which, through the intermediation of the Phelps problem, capture both the acceleration of U.S. inflation in the 1970s and its rapid decline in the early 1980s'.

Another problem with the policy game approach is that it ignores the policy-maker's fundamental dilemma in the face of an adverse supply shock. As shown in Gordon (1975, 1984) and Phelps (1978), unless wages are perfectly flexible, the policy-maker cannot escape a choice between holding inflation constant at the cost of substantial extra unemployment, or holding unemployment constant at a cost of higher and accelerating inflation, or something in between. In fact, because of long lags in the impact of monetary policy on unemployment and inflation, in reality the policy-maker is incapable of holding either inflation or unemployment constant following a supply shock.

Related work by Primiceri (2006) includes the government's underestimate of the NAIRU in the 1970s as a cause of high inflation, but he does not provide any explicit analysis of supply shocks as the cause of this underestimate. Sims (2008) has suggested that Primiceri is guilty of an asymmetry, because he allows only for uncertainty about coefficient values in a model that policy-makers assume is correct, instead of allowing for the fact that the model may be wrong. In fact, the discussion above of the mainstream model suggests that Primiceri and others working on policy–expectations interactions may indeed have chosen the wrong model, at least for the USA, by assuming that expectations can jump in response to policy announcements and ignoring the role of backward-looking inertia and supply shocks.

The new Keynesian Phillips curve

The NKPC model has emerged in the past decade as the centrepiece of macro conference and journal discussions of inflation dynamics and as what Blanchard and others have called the 'workhorse' of the evaluation of monetary policy.[17] The point of the NKPC is to derive an empirical description of inflation dynamics that is 'derived from first principles in an environment of dynamically optimizing agents' (Bårdsen *et al.* 2002).

The theoretical background is that monopolistically competitive firms have control over their own prices due to product differentiation. They are constrained by a friction in the setting of prices, of which there are many possible justifications that are inherited from the theoretical NKE literature. For instance, we have already cited Taylor's (1980) model which merges rational expectations with fixed-duration contracts. More frequently cited, as in Mankiw's (2001) exposition, is Calvo's (1983) model of random price adjustment, in which prices are fixed for random periods. The firm's desired price depends on the overall price level and the unemployment gap. Firms change their price only infrequently, but when they do, they set their price equal to the average desired price until the time of the next price adjustment. The actual price level, in turn, is equal to a weighted average of all prices that firms have set in the past. The first-order conditions for optimization then imply that expected future market conditions matter for today's pricing decision. The model can be solved to yield the standard NKPC specification in which the inflation rate (p_t) depends on expected future inflation ($E_t p_{t+1}$) and the unemployment (or output) gap:

$$(8) \qquad p_t = \alpha E_t p_{t+1} + \beta\left(U_t - U_t^*\right) + e_t.$$

The constant term is suppressed, so the NKPC has the interpretation that if $\alpha = 1$, then U_t^* represents the NAIRU.

Notice that the NKPC in equation (8) is identical to the post-1975 mainstream PC written in (7) above, with two differences. First, there is no explicit treatment of supply shocks; these are suppressed into the error term. Second, expectations are explicitly forward-looking in equation (8), whereas in (7) expectations could be either forward-looking or backward-looking, or both. Because of frictions of the Taylor or Calvo type, policy changes that raise or lower the inflation rate have short-run effects on the unemployment or output gap. The Taylor NKE framework assumes fixed contract lengths of pricing intervals, while the Calvo model makes price changes dependent on a fixed gap between the actual and desired price levels. But Sims (2008) points out that this 'theory has simply moved the non-neutrality from agent behavior itself into the constraints the agent faces, the frictions'. In real-world situations in which macro shocks create Argentina-like instability, contract lengths would surely change in response to the expected inflation rate.

NKPC models vary in their inclusion of the single variable that supplements future expected inflation. This is modelled sometimes as the unemployment gap, as in (8), and sometimes as the closely related output gap. Mankiw's (2001) exposition, followed below, uses the unemployment gap.

Another version of the NKPC replaces either gap with real marginal cost. This is always proxied by real average cost, that is, the real wage divided by the average product of labour $(W/P)/(Y/N)$, which is by definition equal to labour's share in national income (WN/PY). Some papers in the NKPC literature treat real marginal cost as exogenous, but this is unacceptable because labour's income share is inherently endogenous and requires a multi-equation model with separate equations for the level of the wage rate, the price level and the level of labour productivity. Thus far empirical implementation of the marginal cost version of the NKPC has swept under the rug the endogeneity of labour's income share. In contrast, Dew-Becker and Gordon (2005) have examined joint feedback between prices and wages by endogenizing changes in labour's share.[18]

This section treats the 'right fork', with its absence of inertia and expectations that can jump in response to anticipated policy changes, as a fruitful development in macroeconomics when applied to rapid inflation episodes, whether in Germany in 1922–23 or in Brazil or Argentina more recently. Unfortunately, the empirical implementation of the NKPC has been almost entirely to data for the postwar USA, where it is the wrong model. This can be easily seen for both the 'gap' and 'real marginal cost' versions of the NKPC. The 'gap' version as written above in equation (8) drives changes in the inflation rate only with changes in the unemployment gap, or equivalently with the output gap. Its prediction is that the coefficient (β) on the unemployment gap is negative. But we have already seen in Figure 1 that the correlation between inflation and the unemployment rate is both negative and positive, with a positive correlation between 1971 and 1982 when the variance of inflation was greatest. The NKPC contains no element to capture the switch from negative to positive correlation and back again.

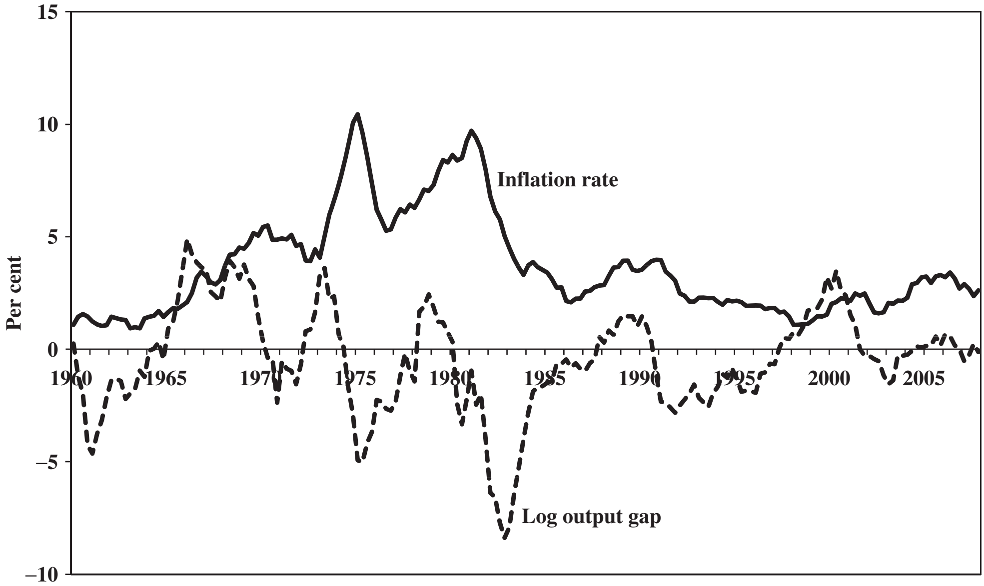

FIGURE 3. The inflation rate and the per cent log output gap, quarterly data, 1960–2007.

Correspondingly, the model's prediction is that when the unemployment gap is replaced by the output gap, the correlation should be positive. But as shown in Figure 3, the correlation between inflation and the output gap is strongly negative between 1971 and 1982. Thus it is not surprising that Rudd and Whelan (2005b, Table 1) show that the estimated coefficient on the output gap is significantly *negative*. Their result is obtained by the usual procedure of replacing the unknown expectation of future inflation with two-stage least squares estimation in which the first stage regresses the actual inflation rate on a set of instruments.[19] This apparent conundrum is resolved in the mainstream triangle model in which the output gap leads inflation positively (as in 1965–69 and 1986–89) while inflation leads the output gap negatively (as in 1971–82) due to the influence of supply shocks.

A look at the data also predicts a failure of the version of the NKPC that uses real marginal cost as the variable that drives inflation. As noted above, real marginal cost is always proxied by real average cost, which is the same as labour's income share, and this share is plotted against the inflation rate in Figure 4.[20] Labour's share exhibits one big upward jump in 1967–70, at least four years too early to explain the first inflation peak in 1974–75. After 1970, labour's share is essentially trendless, varying only between 70% and 75%, with no movements that would help to explain the second inflation peak in 1979–81 nor the Volcker disinflation of 1981–84. Accordingly, it is not surprising that Rudd and Whelan (2005b, Table 1) estimate an insignificant coefficient in equation (8) when real marginal cost replaces the unemployment gap.

The NKPC literature seems to be just as confused by the behaviour of real marginal cost as by the negative correlation of inflation with the output gap. As Woodford (2003) has pointed out, the standard model predicts that increases in output tend to be accompanied by higher real marginal cost as workers move out of a positively sloped labour supply curve, as overtime premia rise, and as input materials costs respond positively. However, Figure 4 indicates that, at least before 1990, labour's income share peaked in recessions and appears to be countercyclical. This has an easy explanation that

TABLE 1

ESTIMATED EQUATIONS FOR QUARTERLY CHANGES IN THE PCE DEFLATOR, 1962(I) TO 2007(IV)

		Roberts	
Variable	Lags	NKPC	Triangle
Constant		1.16**	
Lagged dependent variable	1–24[a]	1.01**	
	1–4	0.95**	
Unemployment gap	0–4		– 0.56**
Unemployment rate	0	– 1.17*	
Relative price of imports	1–4		0.06**
Food–energy effect	0–4		0.89**
Productivity trend change	1–5		– 0.95**
Nixon controls 'on'	0		– 1.56**
Nixon controls 'off'	0		1.78**
R^2		0.78	0.93
SEE		1.17	0.64
SSR		244.0	64.6
Dynamic simulation 1998(I) to 2007(IV)	Note b		
Mean error		– 2.75	0.29
Root mean-square error		3.20	0.70

Notes

**Indicates significance at 1%; *indicates significance at 5%.

[a]Lagged dependent variable is entered as the four-quarter moving average for lags 1, 5, 9, 13, 17 and 21, respectively.

[b]Dynamic simulations are based on regressions for the sample period 1962(I) to 1997(IV) in which the coefficients on the lagged dependent variable are constrained to sum to unity.

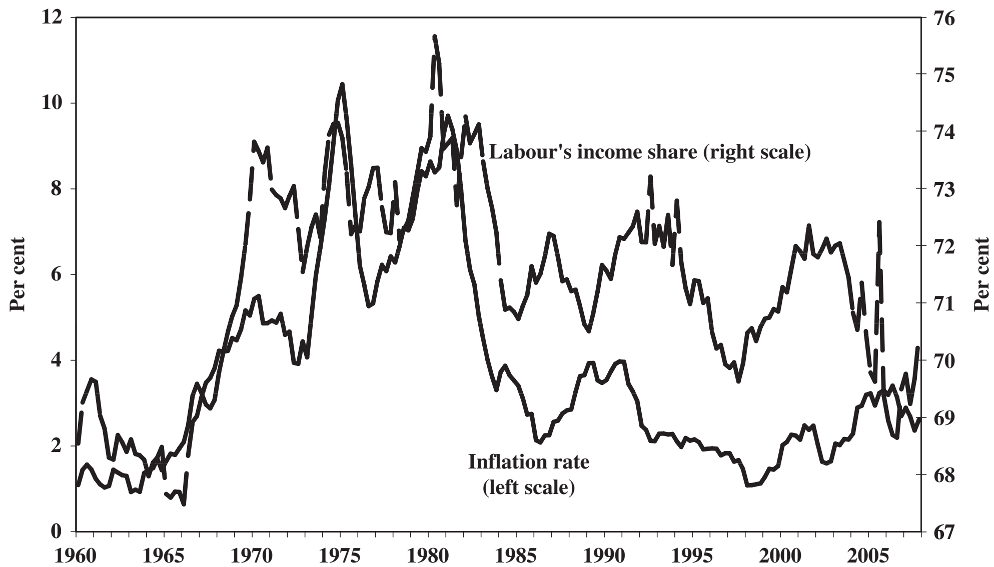

FIGURE 4. The inflation rate and labour's share in domestic net factor income, quarterly data, 1960–2007.

has apparently been neglected in many NKPC discussions: the procyclicality of labour productivity, which appears in the denominator of real average cost. Rudd and Whelan (2005b) also discuss the problem that labour's share, which equals real average cost, may be a poor proxy for real marginal cost.

The challenge of persistence

On the surface, the NKPC as written in equation (8) appears similar, except for the omission of explicit supply shock variables, to the mainstream PC as written in equation (7). But its policy implications are radically different from the mainstream model, with its costly disinflation and significant sacrifice ratio. This occurs because in the NKPC model there is no backward-looking inertia, that is, no structural dependence of inflation on its own lagged values. Instead, inflation is entirely driven by forward-looking expectations, and equation (8) can be solved forward to set the inflation rate equal to an infinite sum of expected future output gaps. Inflation can be costlessly controlled by a credible commitment to follow policies that minimize the output gap forever into the future.

However, as we shall see in Section IV, inflation persistence in the form of long lags on past inflation rates is a central feature of postwar US inflation behaviour. As a result, in the US environment expectations are unlikely to jump except in response to widely recognized supply shocks, such as the surge of oil prices in 1973–75, 1979–81 or 2006–08. The recognition that, in the absence of supply shocks, the inflation rate is dominated by persistence creates a challenge for policy-makers to reduce inflation by altering public expectations *directly*. How can policy-makers convince the public that inflation will spontaneously decrease, without any cost of higher unemployment or lost output, when the public knows that inflation behaviour is dominated by persistence?

As we show in the first subsection of Section IV, in practice the NKPC is simply a regression of the inflation rate on a few lags of inflation and the unemployment gap. As pointed out by Fuhrer (1997), the only sense in which models including future expectations differ from purely backward-looking models is that they place restrictions on the coefficients of the backward-looking variables that are used in the first stage of two-stage least squares estimation as proxies for the unobservable future expectations. In Fuhrer's words:

> Of course, some restrictions are necessary in order to separately identify the effects of expected future variables. If the model is specified with unconstrained leads and lags, it will be difficult for the data to distinguish between the leads, which solve out as restricted combinations of lag variables, and unrestricted lags. (Fuhrer 1997, p. 338)

Subtleties in the interpretation of the 'hybrid' NKPC

Galí and Gertler (1999), two of the inventors of the NKPC approach, have introduced a 'hybrid' NKPC model in which the public consists of both forward-looking and backward-looking agents, and in their empirical version current inflation depends on both expected future inflation and past inflation. However, since future inflation is always proxied by some transformation of past inflation, there is little difference in practice between the 'pure' forward-looking NKPC and the hybrid version, except for the form of the restrictions that emerge. Further, if there are enough backward-looking members of the population, then forward-looking members cannot ignore the persistence introduced by backward-looking agents. This dependence of future contract outcomes on the inheritance of ongoing contracts with staggered expiration dates has been explicit in the theoretical NKE literature since its introduction by Taylor (1980).

The hybrid NKPC model is the same as equation (8) above, except that the influence of inflation is divided between future expected inflation and lagged inflation rather than being channelled exclusively through future expected inflation:

$$(9) \qquad p_t = \alpha_f E_t p_{t+1} + \alpha_b p_{t-1} + \beta(U_t - U_t^*) + e_t.$$

The central issue is the relative size of the forward-looking and backward-looking coefficients (α_f and α_b). Galí and Gertler (1999) and Galí *et al.* (2005) have reported estimates of equation (9) with the unemployment gap replaced by labour's income share and conclude that 'forward-looking behavior is dominant', i.e. α_f is estimated to be much larger than α_b. This conclusion is important, since it appears to justify the original formulation of the NKPC (equation (8) above) and makes the role of lagged inflation appear to be a minor distraction of little empirical importance.

However, as pointed out by Rudd and Whelan (2005a, b), these estimates do not actually distinguish between forward-looking and backward-looking behaviour due to the nature of the two-stage least squares exercise. The second-stage equation (9) omits variables that belong in the true model of inflation—e.g. additional lags on inflation itself as well as explicit supply shock variables like commodity prices—but includes them in the first stage as a proxy for expected future inflation. Indeed, anything that is correlated with current inflation but not included in the second stage will serve as a good instrument for future expected inflation and thus raise α_f relative to α_b. These omitted variables boost the coefficient on expected future inflation even if expected future inflation has no influence at all on inflation itself, as occurs when Rudd and Whelan estimate a pure backward-looking model that includes some of the additional variables included as instruments in the two-stage procedure. Overall, the NKPC approach to date has delivered no evidence that expectations are forward-looking. The instruments used in the first stage are incompatible with the theory posited in the second stage. If lagged inflation and commodity prices matter for inflation, then why are they omitted from the NKPC inflation equation?

Constraints on the formation of expectations

Recent research has revived the discussion of barriers to the formation of expectations. As we have seen, the original formulations of Friedman, Phelps and Lucas were based on information barriers that prevented one set of agents (Friedman's workers) or all agents (Phelps' desert island residents) from having access to government data on income, output, money and prices that are released frequently at zero cost to all agents in the economy. That literature was flawed because it placed the information barriers in the wrong place, in an inability to perceive costless macro information, instead of where the information barriers really exist, at the micro level of costs and supplier–producer relationships.

Producers of final goods are unable to perceive cost increases of crude and intermediate materials that may be in the pipeline, and they have no choice but to wait until they receive notification of actual cost changes (with the exception of crude materials like oil where prices are determined in public auction markets). This approach, based on supplier–producer arrangements, was introduced in Blanchard (1987) and was christened the 'input–output' approach in Gordon (1990), who suggests a four-cell matrix of information barriers of supply and demand shocks at both the macro and micro level. A fundamental source of persistence is not just explicit wage contracts as analysed by Taylor, but also explicit or implicit price contracts between suppliers and producers of final goods. Even without contracts, persistence and inertia are introduced by lags between price changes of crude materials, intermediate goods and final goods. For some goods, e.g. cars or aircraft, there are literally thousands of separate intermediate goods, and most of these are made up of further layers of intermediate goods.

The recent literature has largely ignored the *micro* uncertainty embodied in the input–output approach and instead has attempted to find credible explanations for imperfect *macro* information. One approach is that agents take time to learn about the structure of the economy (see Orphanides and Williams 2005). This information barrier is consistent with the triangle approach, in which changes in the TV-NAIRU are observed not in real time but only after the fact, as are changes in coefficients on the PC slope or the coefficients on such supply shocks as oil prices.

A second barrier may be imperfect information regarding the goals of the central bank (see, among others, Kiley 2007). Clearly, in the US context the Fed has changed goals several times, and this became evident only after the fact. The Volcker policy shift in 1979–80 was widely noticed at the time, but there was no historical antecedent to allow predictions of its consequences. Likewise, studies of the Taylor rule indicate that the Fed shifted around 1990 from a policy that mainly responded to inflation to a policy that mainly responded to the output gap, and no empirical Taylor rule can explain why the Federal Funds rate was so low in 2001–04.

The third barrier consists of costs or constraints on information acquisition and processing. One version of this approach emphasizes costs of obtaining information that lead to infrequent adjustments in expectations (Reis 2006). Another approach (see Sims 2006) is called 'rational inattention' and also emphasizes constraints on information processing capabilities. However, all of these barriers concern constraints on the ability of private agents to adjust their expectations accurately to reflect the current stance of monetary policy and anticipated future changes in policy, and none reflects any of the sources of persistence and inertia, particularly lack of information at the micro level of the input–output table. Rational inattention makes sense at the micro level, when translated into the minimization of managerial cost by avoiding daily deliberations about price changes required by changes in supplier costs and instead making decisions infrequently.

Which model applies to which episodes?

This survey has contrasted the inertia-bound triangle approach to explaining US inflation with alternative frameworks in which the expectations of private agents can jump in response to perceived changes in monetary policy. Which model best describes which historical situations? As indicated above, the mutual interplay between policy decisions and expectations formation is essential to understanding episodes of rapid inflation, including Sargent's (1982) ends of four big inflations. This approach is in fact essential to an understanding of the inflation process in any nation that has experienced high inflation volatility in the past due to shifts in policy (as contrasted with the influence of supply shocks). A prime example would be a country like Argentina in which private agents know that the government's ability to restrain monetary growth depends on fiscal decisions made at the level of states and localities.

Relatively little research has been done to establish a dividing line between situations suitable for analysis with the policy–expectations game approach vs the inertia-bound triangle approach. The convergence of inflation rates within the European Monetary Union between 1980 and 1998 provides another example in which an inertia-dominated

PC is inadequate, as countries with similar unemployment rates experienced very different time paths of inflation. After experiencing inflation rates of over 20% to 25% in the mid-1970s, Italy and the UK converged to low single digits of inflation, and an explanation of this convergence requires a model in which agents formed expectations based in part on the monetary policy of the Bundesbank, not just that of their own national central bank.[21]

Another issue in extending the PC framework to fit the postwar European experience is the question of whether the standard PC relation between the inflation rate and the level of unemployment needs to be supplemented by a hysteresis mechanism between the inflation rate and the change of unemployment. Recently Ball (2008) has suggested that the hysteresis idea, after languishing since its invention by Blanchard and Summers (1986), should be revived. Some versions of hysteresis imply that inflation depends not on the level of unemployment but on its rate of change, that ancient idea supported in the results of both Fisher (1926) and Phillips (1958) and incorporated in the triangle model specification.

The empirical results presented in Section IV suggest that the triangle model, which combines demand and supply shocks with inertia, does a much better job in explaining postwar US inflation than does the NKPC approach that omits lags and supply shocks. However, how far back can the triangle-type PC specification be pushed in US data? Samuelson and Solow (1960) had already noticed that the American PC relationship does not work in the Great Depression and during the two world wars. A quantitative answer to this question was provided by Gordon (1982a) in a unique set of interpolated quarterly data extending back to 1892. His results (Table 3) estimate PC equations for 1892–1929, 1929–53 and 1954–80, in a framework which allows the inflation rate to depend not just on the level of the output gap, but on changes in expected and unexpected nominal GDP, lagged inflation and a series of dummy variables.

Gordon's results suggest that prior to 1954 there were substantial shifts in the American PC process that are consistent with a role for an interplay between the expectations of private agents and perceived changes in policy and the macroeconomic environment more generally. The PC relation, in the form of the coefficient on the output gap, has roughly the same coefficient before 1929 as after 1954, but is zero during the middle period. In all three periods the anticipated and unanticipated change in nominal GDP is highly significant, and this 'rate of change effect' dominates the explanation of inflation in the middle period when the PC relationship is absent (see also Romer 1996, 1999 for an analysis of price changes in the 1930s).

The role of policy in shifting the inflation rate (whether this is perceived as working through expectations or not) is evident in large shift coefficients on the impact of nominal GDP changes on inflation in the First World War, and in large negative dummy variables for price controls in the First and Second World Wars, the Korean War and the 1971–74 Nixon episode, as well as a large positive coefficient for the New Deal's National Recovery Administration.[22]

IV. The New Keynesian and Triangle Phillips Curves: Specification, Estimates and Simulations

What difference does it make if we explain US inflation using the mainstream triangle model or the NKPC? We now turn to the detailed specification of the NKPC and triangle alternatives. Then we examine their performance in US data spanning 1962–2007.

The NKPC model

A central challenge to the NKPC approach is to find a proxy for the forward-looking expectations term in equation (8) above ($E_t p_{t+1}$). Surprisingly, there is little discussion in the literature of this aspect, or the implications of the usual solution, which is to use instrumental variables or two-stage least squares (2SLS) to estimate (8). The following treatment is consistent by including in the first stage only the variables that are part of the basic theory in the second stage. The first-stage equation to be included in the 2SLS estimation explains expected future inflation by recent lags of inflation and by the current unemployment gap:

$$(10) \qquad E_t p_{t+1} = \sum_{i=1}^{4} \lambda_i p_{t-i} + \phi(U_t - U_t^*).$$

When the first-stage equation (10) is substituted into the second-stage equation (8), we obtain the reduced form

$$(11) \qquad p_t = \alpha \sum_{i=1}^{4} \lambda_i p_{t-i} + (\alpha N + \beta)\left(U_t - U_t^*\right) + e_t.$$

Thus in practice the NKPC is simply a regression of the inflation rate on a few lags of inflation and the unemployment gap. We have already cited Fuhrer (1997) as pointing out that the only sense in which models including future expectations differ from purely backward-looking models is that they place restrictions on the coefficients of the backward-looking variables, as in (11). The procedure of Galí and Gertler (1999) and many others of adding additional variables like commodity prices and wage changes to the first-stage equation is entirely *ad hoc*, as pointed out by Rudd and Whelan (2005a, b) because any relevance of these variables to the forecasting of future inflation is inconsistent with the basic second-stage NKPC inflation model of equation (8), which omits these additional variables.

The Roberts (2006) version of the NKPC is of particular interest here, because of his finding that the slope of the PC has declined by more than half since the mid-1980s. Roberts describes his equation as a 'reduced form' NKPC, and indeed it is identical to equation (11) above with two differences: the NAIRU is assumed to be constant, and the sum of coefficients on lagged inflation is assumed to be unity. Thus the Roberts (2006, equation 2, p. 199) version of (11) is

$$(12) \qquad p_t = \sum_{i=1}^{4} \lambda_i p_{t-i} + \gamma + \beta U_t + e_t,$$

where the implied constant NAIRU is $-\gamma/\beta$.

The triangle model of inflation and the role of demand and supply shocks

The inflation equation used in this paper is almost identical to that developed 25 years ago (Gordon 1982b). When the influence of demand is proxied by the unemployment gap, the triangle model can be written as (13), which is identical to (7) above except for the introduction of lags. This general framework can be written as

$$(13) \qquad p_t = a(L)p_{t-1} + b(L)(U_t - U_t^N) + c(L)z_t + e_t.$$

As before, lower-case letters designate first differences of logarithms, upper-case letters designate logarithms of levels, and L is a polynomial in the lag operator.[23]

As in the NKPC and Roberts approaches, the dependent variable p_t is the inflation rate. Inertia is conveyed by a series of lags on the inflation rate (p_{t-1}). The term z_t is a vector of supply shock variables (normalized so that $z_t = 0$ indicates an absence of supply shocks), and e_t is a serially uncorrelated error term. Distinguishing features in the implementation of this model include unusually long lags on the dependent variable, and a set of supply shock variables that are uniformly defined so that a zero value indicates no upward or downward pressure on inflation. Current and lagged values of the unemployment gap serve as a proxy for the influence of demand, where the unemployment gap is defined as the difference between the actual rate of unemployment and the NAIRU, and the NAIRU is allowed to vary over time. This specification predicts steady inflation when the unemployment gap and the supply shock terms are all zero, and hence it is always estimated without a constant term.

The estimation of the time-varying (TV) NAIRU combines the above inflation equation (12) with a second equation that explicitly allows the NAIRU to vary with time:

$$(14) \qquad U_t^N = U_{t-1}^N + \eta_t, \quad E\eta_t = 0, \quad \text{var}(\eta_t) = \tau^2.$$

In this formulation, the disturbance term η_t in the second equation is serially uncorrelated and is uncorrelated with e_t. When its standard deviation $\tau_\eta = 0$, the natural rate is constant, and when τ_η is positive, the model allows the NAIRU to vary by a limited amount each quarter. If no limit were placed on the variance of the NAIRU, then the TV-NAIRU would jump up and down and soak up all the residual variation in the inflation equation (13).[24] In practice, the smoothness criterion is chosen to avoid negatively correlated zig-zags in the estimated NAIRU, to be consistent with Friedman's original (1968) idea of the NAIRU as slowly changing in response to underlying microeconomic structural factors.

The triangle approach differs from the NKPC and Roberts approaches by including long lags on the dependent variable, additional lags on the unemployment gap that incorporate a rate-of-change effect, and explicit variables to represent the supply shocks (the z_t variables in (13) above), namely the change in the relative price of non-food non-oil imports, the effect on inflation of changes in the relative price of food and energy, the change in the trend rate of productivity growth, and dummy variables for the effect of the 1971–74 Nixon-era price controls.[25]

Estimating the TV-NAIRU

The time-varying NAIRU is estimated simultaneously with the inflation equation (12) above. For each set of dependent variables and explanatory variables, there is a different TV-NAIRU. For instance, when supply shock variables are omitted, the TV-NAIRU soars to 8% and above in the mid-1970s, since this is the only way the inflation equation can 'explain' why inflation was so high in the 1970s. However, when the full set of supply shocks is included in the inflation equation, the TV-NAIRU is quite stable, shown by the dashed line plotted in Figure 5, remaining within the range 5.7% to 6.5% over the period between 1962 and 1988.

Beginning in the late 1980s, the TV-NAIRU drifts downwards until it reaches 5.3% in 1998, and then it displays a further dip in 2004–06 to 4.8%. One hypothesis to be explored below is that the Roberts NKPC implementation reaches the conclusion that

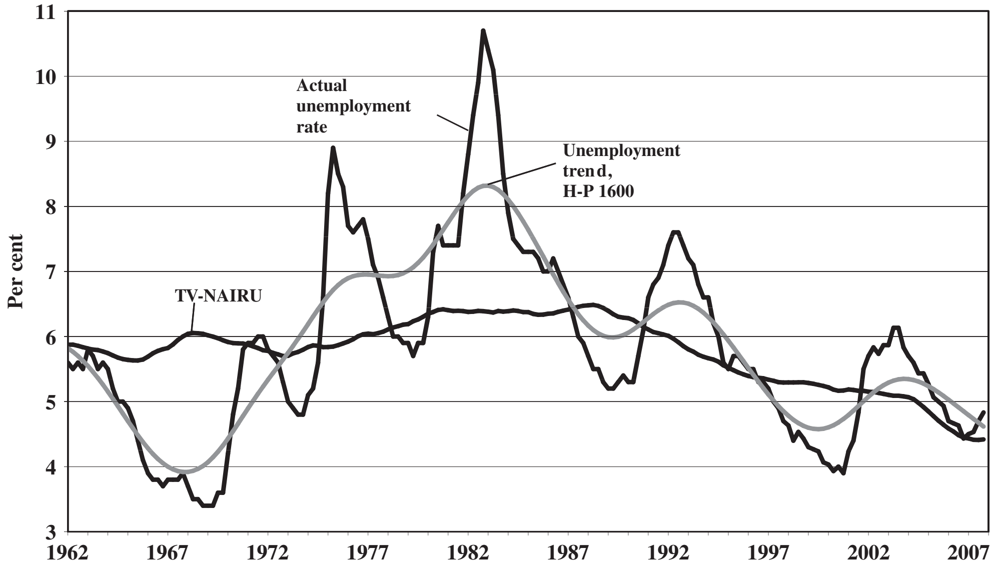

FIGURE 5. Actual unemployment rate vs time-varying NAIRU, 1962–2007.

the Phillips curve has flattened because the NAIRU is forced to be constant, and that a decline in the TV-NAIRU is an alternative to a flatter PC in explaining why inflation has been relatively well behaved in the past 20 years.

Some of the NKPC literature estimates the TV-NAIRU by directly applying a Hodrick–Prescott (H–P) filter to the time series of the unemployment rate.[26] As shown in Figure 5 using the traditional H–P parameter of 1600, this 'direct' approach to estimating the TV-NAIRU results in an unexplained increase in the TV-NAIRU from 3.9% in 1968 to 8.3% in 1985, whereas the triangle approach has no such unexplained increase because of its introduction of explicit supply shock variables.[27] In contrast, the Roberts implementation of the NKPC forces the NAIRU to be constant at an estimated 7.0%.

How much difference do the explicit supply shock variables make in the predictions of the triangle specification? Figure 6 displays predictions made with the actual supply shock variables and with the supply shock variables zeroed out (but with the other variables, including the unemployment gap, taking their historical values and estimated coefficients). Evidently, the supply shock variables explain *all* of the twin peaks of inflation in the 1973–81 period, and in addition they explain more than half of the Volcker disinflation (predicted inflation drops by 7 percentage points between 1980(I) and 1985(I) when supply shocks are included, but by only 3 percentage points when supply shocks are excluded). Notice also that without (beneficial) supply shocks, the influence of low unemployment would have caused inflation to rise by 1.3 percentage points between 1994 and 2001, whereas with supply shocks inflation is predicted to be roughly constant.

Roberts NKPC vs triangle: coefficients and simulation performance

We next turn to the estimated coefficients, goodness of fit and simulation performance of the Roberts NKPC and triangle PC specifications. Table 1 displays the estimated sums of coefficients and their significance levels for both the Roberts NKPC and triangle

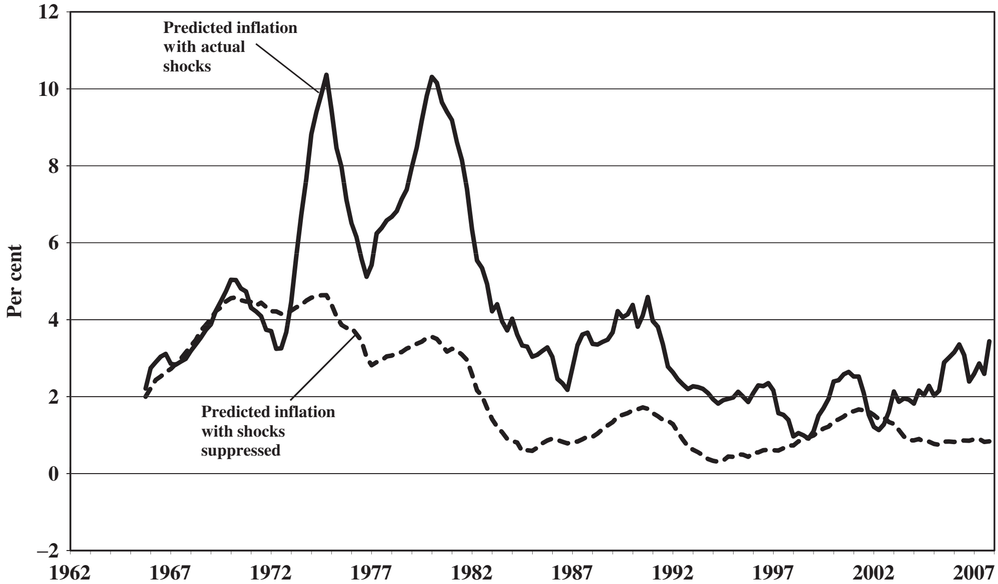

FIGURE 6. Predicted inflation in triangle mode with and without supply shocks, 1962(I) to 2007(IV).

specifications for equations in which the dependent variable is the quarterly change in the headline PCE deflator. In both specifications the sum of coefficients on the lagged inflation terms is very close to unity, as in previous research.[28] The sum of the unemployment gap variables in the triangle approach is around − 0.6, which is consistent with a stylized fact first noticed by Samuelson and Solow (1960) that the slope of the short-run Phillips curve is roughly minus one-half. Why is the Roberts NKPC unemployment coefficient lower than in the triangle specification? The excluded supply shock variables are positively correlated with inflation and positively correlated with the unemployment gap, so the omission of these supply shock variables causes the negative coefficient on the unemployment gap to be biased towards zero. We note that the sum of squared residuals (SSR) for the triangle model is barely one-quarter that of the Roberts NKPC specification.

The explicit supply shock variables in the triangle model are all highly significant and have the correct signs; except for the productivity trend variable, all of these enter exactly as specified in 1982 and thus their significance has not been diminished by an extra 25 years of data. The change in the relative import price effect has a highly significant coefficient of 0.06.[29] The food–energy effect has a coefficient of 0.89, close to the expected value of 1.0. The productivity trend variable has the expected negative coefficient and helps to explain why inflation accelerated in 1965–80 and was so well behaved in 1995–2000. The Nixon-era control coefficients, as in previous research, indicate a significant impact of the controls in holding down inflation by a cumulative − 1.6% in 1971–72 and boosting inflation by 1.8% in 1974–75.

Rather than relying on the usual statistical measures of goodness of fit, a dynamic model heavily dependent on the contribution of the lagged dependent variable is best tested by the technique of dynamic simulations. These generate the predictions of the equation with the lagged dependent variable generated endogenously rather than taking the actual values of lagged inflation. To run such simulations, the sample period is

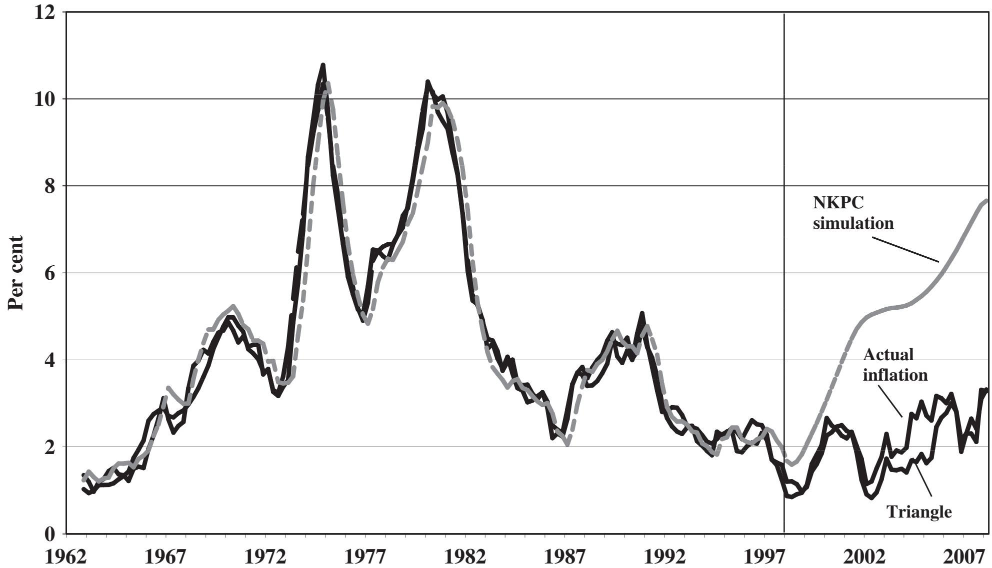

FIGURE 7. Predicted and simulated values of inflation from triangle and NKPC equations 1962(I) to 2007(IV).

truncated ten years before the end of the data interval, and the estimated coefficients through 1997(IV) are used to simulate the performance of the equation for 1998–2007, generating the lagged dependent variables endogenously. Since the simulation has no information on the actual value of the inflation rate, there is nothing to keep the simulated inflation rate from drifting away from the actual rate. These simulations have been criticized because they use the actual values of the explanatory variables other than lagged inflation, but this is an exercise not in forecasting but rather in determining whether a particular set of variables and lags adds to the explanatory power of the equation.

The bottom section of Table 1 displays results of a dynamic simulation for 1998(I) to 2007(IV) based on a sample period that ends in 1997(IV). Two statistics on simulation errors are provided: the mean error (ME) and the root mean-squared error (RMSE). The simulated values of inflation in the triangle model are close to the actual values, with a mean error over 40 quarters of only 0.29, meaning that over the forty quarters the actual inflation rate on average is 0.29 percentage points higher than the predicted value. However, the mean error for the Roberts NKPC model is a huge -2.75, and as displayed in Figure 7, this error reflects that model's wild overprediction that the inflation rate should have reached nearly 8% by late 2007. The RMSE of the triangle simulation is a bit above the standard error of estimate (SEE) for the 1962–1997 sample period, whereas for the Roberts NKPC model the RMSE is almost three times as large.

The Roberts NKPC and triangle results agree on only one aspect of the inflation process, that the sum of coefficients on the lagged inflation terms is always very close to unity. However, the Roberts coefficients on the unemployment rate are much lower than the triangle coefficients on the unemployment gap. This is an artifact of the exclusion restrictions in the Roberts approach which are statistically rejected in the triangle approach.

The triangle model outperforms the Roberts NKPC model by several orders of magnitude, as displayed in Table 1 and Figure 7. This raises a question central to future research on the US Phillips curve: what are the crucial differences that contribute to the superior performance of the triangle model? The three key differences are the inclusion in the triangle model of longer lags on both inflation and the unemployment gap, the inclusion of explicit supply shock variables, and the allowance for a time-varying (TV) NAIRU in place of the Roberts NKPC assumption of a fixed NAIRU. In the Appendix we quantify the role of these differences, taking advantage of the fact that the Roberts NKPC model is fully nested in a version of the triangle model that assumes a constant NAIRU. Each exclusion restriction in the Roberts model can be tested by standard statistical exclusion criteria, and every one of the Roberts NKPC exclusion criteria is rejected at high levels of statistical significance.[30]

Has the PC slope become flatter?

The NKPC research of Roberts and others has concluded that the Phillips curve has become flatter over the past several decades. Yet we have seen that every aspect of the Roberts NKPC specification is rejected at high levels of statistical significance.

Has the Phillips curve flattened? The Roberts NKPC specification says 'yes' and the triangle specification says 'no'. Figure 8 evaluates changes in coefficients by Roberts' own preferred method (2006, Figure 2, p. 202): rolling regressions that shift the sample period of the regression through time in order to reveal changes in coefficients. The number of quarters in our basic results in Table 1 is 184 (1962(I) to 2007(IV)), and we cut this roughly in half to 90 quarters and run rolling 90-quarter regressions which alternatively start in each quarter from 1962(I) to 1985(III).

As shown in Figure 8, the Roberts NKPC unemployment coefficient rises from – 0.17 in 1962 to a peak value of – 0.41 in 1974, and then declines back to roughly zero in 1982–84. This appears to support his basic conclusion that the Phillips curve has

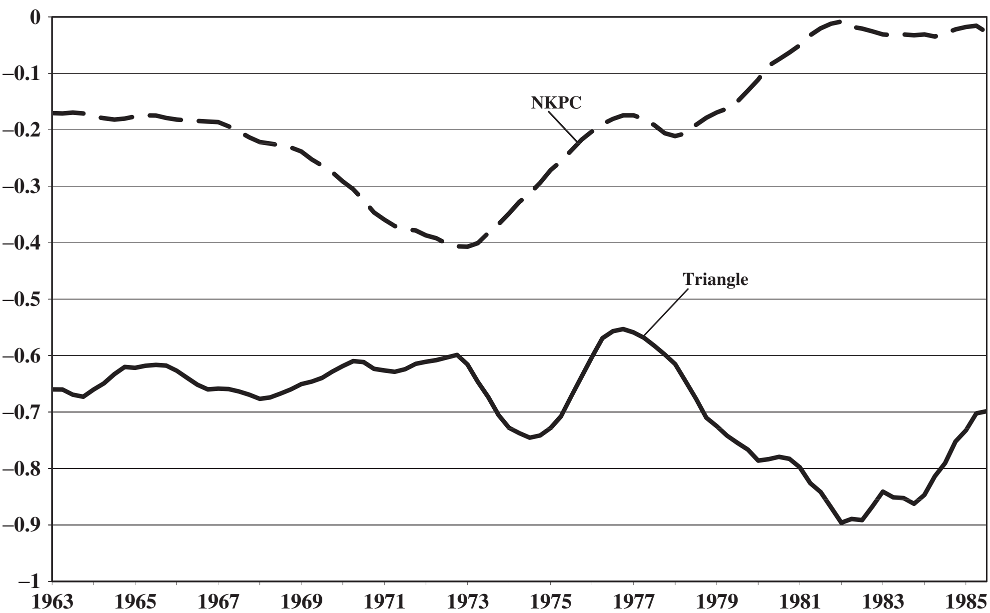

FIGURE 8. NKPC vs triangle unemployment coefficients in 90-quarter rolling regressions from 1962(I) to 1985(III).

flattened. Yet the triangle model reveals no evidence of a decline in the slope of the Phillips curve. The Phillips curve based on the triangle model has a roughly stable PC slope of about -0.6 to -0.7 from 1963 to 1977, and then the slope *rises* towards about -0.7 to about -0.9 in the simulations starting in 1982, then drifts back to -0.7 in the final year. As indicated above, the NKPC slope estimate is biased towards zero by differing amounts in each period, due to the omission of supply shocks.

Has the impact of supply shocks become less important?

Hooker (1996) was among the first to notice that the macroeconomic impact of oil shocks became smaller at some point between the mid-1970s and early 1980s. Since his work, a substantial literature has arisen to debate the sources of the decline in the impact of oil prices on real output and/or inflation. The most obvious cause of the decreased macroeconomic impact of oil originates in the shrinking input of energy in GDP, down by half since 1969.[31] Kilian (2008) provides a set of reasons that go beyond the declining input share of energy. Part of the answer lies in the role of global demand in causing much of the recent 2004–08 rise in oil prices; a demand-driven increase in oil prices may raise rather than reduce real GDP as would occur in the case of a supply disruption.

Blanchard and Galí (2007) go more deeply into the sources of the declining macroeconomic importance of oil beyond its shrinking input share. Their first reason is that the 1970s oil shocks had a big impact because they were accompanied by other significant shocks that all had the effect of raising the rate of inflation and reducing output. These other shocks, embedded in the triangle inflation equation since the beginning, include adverse food price shocks, the depreciation of the dollar after the breakdown of Bretton Woods, and the unwinding of the Nixon-era price controls. The 2004–08 rise in oil prices had a smaller impact because it was not accompanied as in the 1970s by these other adverse shocks. There was an increase in the real price of food only in 2007–08, and the post-2002 decline in the dollar was not sufficient to cause any sustained increase in the relative price of imports.

Blanchard and Galí also consider two additional factors that may have reduced oil's impact: a decrease in the extent of real wage rigidity and the increased credibility of monetary policy. Using survey data on expected inflation, they show a sharply diminished response of expectations to a given size of oil shock after the mid-1980s, which they attribute to increased credibility.

Does the triangle model confirm a reduced macroeconomic role for oil shocks? Part of oil's impact is disguised in the triangle specification, which enters not the change in the real price of oil, but rather the 'food–energy effect' defined as the difference between headline and core inflation. Thus the declining share of energy input in GDP would cause a reduced response of the food–energy effect to any given change in the real price of oil, not a change in the coefficient on the food–energy effect in the triangle inflation equation itself.

However, it does appear that the coefficient on the food–energy effect has declined, as suggested by the other literature surveyed in this section. Figure 9 plots the sum of coefficients on the food–energy effect in rolling regressions computed by the same technique and for the same time period as for the unemployment coefficient already displayed in Figure 8. Two lines are shown. The first shows changes in the food–energy sum of coefficients in the equation for headline inflation, the same as Table 1. In contrast to a coefficient sum of 0.89 when that sum is held constant over the full 1962–2007 period as in Table 1, Figure 9 shows that the sum of coefficients is at or above 1.0 in 90-quarter regressions from 1962 to 1980, and then the sum declines to about 0.6 in the final 90-quarter

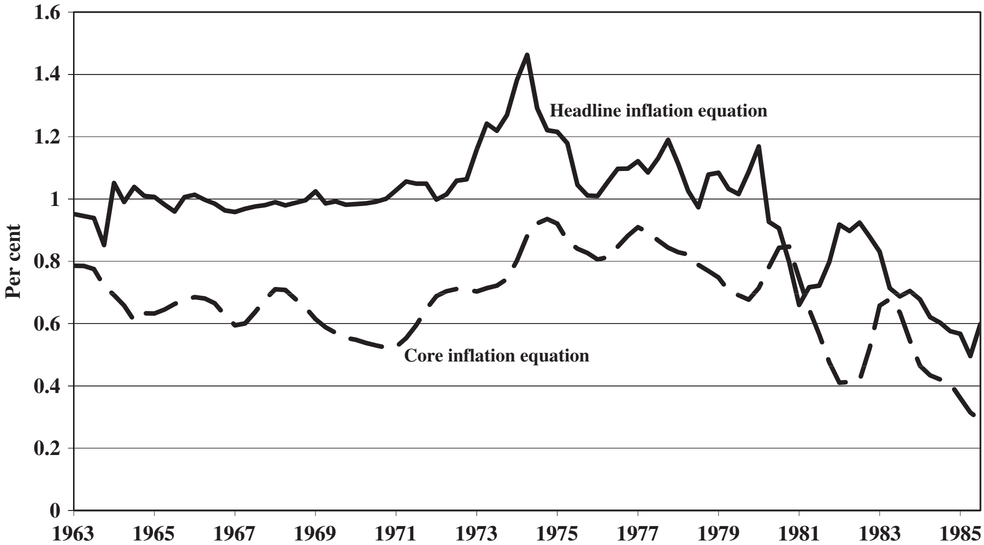

FIGURE 9. Sum of coefficients on food–energy effect in headline and core triangle inflation equations, 90-quarter rolling regressions from 1962(I) to 1985(III).

regression estimated for 1985(III) to 2007(IV). The second line shows a similar decline in the food–energy effect in the core inflation equation, from an average of 0.72 for 90-quarter regressions from 1962 to 1980 down to only 0.3 for the final 90-quarter regression.[32]

There is likely to be an interplay between shifts in coefficients and the estimated TV-NAIRU displayed in Figure 5. For instance, the sudden decline in the TV-NAIRU from 5.0% to 4.5% in 2006–07 may be an artifact of the assumed fixity of the food–energy coefficient. If the sum of coefficients on the food–energy effect in 2006–07 were allowed to be smaller, there would be less of a puzzle as to why inflation was so low during that period and hence no reason for the estimation to 'force' a decline in the NAIRU. In current research I am looking more closely at changes over time in all the sets of coefficients in the specification of Table 1 and their implication for post-sample simulation performance and the time series behaviour of the TV-NAIRU.

V. CONCLUSION

This paper makes several unique contributions. It contrasts the consensus history of the PC before 1975 with the bifurcated split in PC research since 1975. The evolution of PC doctrine before 1975 is widely accepted and no longer elicits much debate. The discovery by Phillips and his disciples Samuelson and Solow of an inverse relationship between inflation and unemployment briefly suggested an exploitable policy trade-off that was destroyed by the Friedman–Phelps natural rate hypothesis of the late 1960s. Exploitable trade-offs were out, and long-run neutrality was in (it had never disappeared in many environments, including Latin America and the University of Chicago). The econometric models developed in the 1960s to support the policy trade-off were rejected both empirically and logically by Sargent's trenchant identification argument.

Debates in the early 1970s centred on the models in which Friedman and Phelps had embedded the natural rate hypothesis, and particularly the assumption of arbitrary barriers that prevented individual workers or agents in general from learning the values of macro data—output, money and prices—provided costlessly by the government. There was also controversy about the implications of the further development of the Friedman–Phelps paradigm by Lucas, who introduced rational expectations into macroeconomics. The Lucas model implied the policy ineffectiveness proposition, which held that anticipated changes in money had no effect on output and were entirely reflected in price changes. Empirical work rejected this framework, showing that monetary surprises had little effect on output, were incapable of explaining the serial correlation of output, and were inconsistent with the persistence of inflation.

After 1975 the PC literature bifurcated into two lines of research, which since then have communicated little with each other. Along the 'left fork in the road', the PC was revived by importing micro demand and supply analysis into macroeconomics. There was no assumption that unemployment and inflation are negatively correlated. Demand shocks create an initial and temporary negative correlation, and supply shocks create an initial positive correlation that then evolves according to the policy response. As early as 1975, the theoretical literature on policy responses to supply shocks was developed and showed that adverse supply shocks force policy-makers to choose between higher inflation, lower output, or a combination. By the early 1980s an econometric specification of this AD–AS framework was available that joined demand and supply shocks with long-run neutrality and a strong role for persistence and inertia.

An important difference between the mainstream approach and other post-1975 developments is that the role of past inflation is not limited to the formation of expectations, but also includes pure persistence due to fixed-duration wage and price contracts, and lags between changes in intermediate goods and final product prices. Inflation is dislodged from its past inertial values by demand shocks proxied by the unemployment or output gap, and explicit supply shock variables including changes in the relative prices of food, energy and imports, and the role of changes in trend growth of productivity. The econometric implementation of this approach is sometimes called the 'triangle' model, reflecting its three-cornered dependence on demand, supply and inertia.

After 1975, the 'right fork in the road' built models in which expectations are not anchored in backward-looking behaviour but can jump in response to current and anticipated changes in policy. Important elements of this literature include policy credibility, models of the game played by policy-makers and private agents forming expectations, and the new Keynesian Phillips curve (NKPC), which derives a forward-looking PC from alternative theories of price stickiness. The common feature of these theories is the absence of inertia, the exclusion of any explicit supply shock variables, the ability of expected inflation to jump in response to new information, and alternative barriers to accurate expectation formation due to such frictions as 'rational inattention'.

Which post-1975 approach is right? Models in which expectations can jump in response to policy are essential to understanding Sargent's (1982) ends of four big inflations and other relatively rapid inflations in nations with a history of monetary instability, e.g. Argentina. But the mainstream/triangle approach is the right econometric framework in which to understand the evolution of postwar US inflation. The paper tests the triangle econometric specification alongside one recently published version of the NKPC approach. The latter can be shown to be nested in the former model and to differ by excluding particular variables and lags, and these differences are all rejected by tests of exclusion restrictions. The triangle model outperforms the NKPC variant by orders of magnitude, not only in standard goodness-of-fit statistics, but also in post-sample dynamic simulations. The triangle estimates show that the slope of the Phillips curve has not become appreciably flatter in the past two decades, a conclusion reached in the NKPC framework due to the specification error of omitting explicit supply shock variables. The triangle estimates do, however, confirm other work indicating that the impact of oil shocks on inflation has diminished.

Thus there are three main interrelated themes in this paper that have not previously received sufficient attention. First, two quite legitimate responses, the left and right forks, occurred after 1975 to the chaotic state of the PC literature at that time. Second, each response is important and helps us to understand how inflation behaves, albeit in different environments. Third, the two approaches need to pay more attention to each other and to engage in a dialogue about which models apply to which episodes, and what factors would motivate a shift in relevance between the alternative models. This paper represents a start towards that reconciliation.

APPENDIX: THE 'TRANSLATION MATRIX' BETWEEN THE ROBERTS NKPC AND TRIANGLE SPECIFICATIONS

Which differences matter in explaining the poor performance of the Roberts NKPC specification in Table 1 and Figure 7? In this appendix we start with the Roberts NKPC specification and gradually change, step by step, to the triangle specification, allowing the NAIRU alternatively to be constant and to vary over time. In everything that follows, the sample period is 1962(I) to 2007(IV).

Table A1 provides the 'translation matrix' that guides us between the Roberts NKPC specification and the triangle specification. There are 24 rows that allow us to trace the role of each specification difference, and the individual rows of alternative specification are evaluated based not just on the SSR measure of goodness of fit, but also on the post-sample simulation performance in 1998–2007 based on coefficient estimates for 1962–97.

We have already seen in Table 1 that the performance of the Roberts NKPC specification for the PCE deflator is inferior to that of the triangle specification by both the criterion of goodness of fit (SSR) and also the less conventional criterion of dynamic simulation performance (ME and RMSE). In Table A1 the basic Roberts variant is in row 1 and the basic triangle variant is in row 21. Roberts' row 1 and the triangle row 21 have SSRs of 244.0 and 64.6, exactly the same as in Table 1.

Table A1 allows the three main differences between the Roberts NKPC and triangle specifications to be evaluated, step by step. Is the crucial difference the longer lags, the supply shocks, the TV-NAIRU, or an interaction of these differences?

In the 24 rows of Table A1, the first 12 rows exclude supply shock variables, and rows 13–24 include the supply shock variables. Scanning down the column for 'SSR', we find that the variants in rows 13–24 including supply shocks all have SSRs below 100, while most of the SSRs that exclude supply shocks have values above 200. Thus our first conclusion is that the exclusion of explicit supply shocks in the Roberts NKPC research is the central reason for its empirical failure either to explain postwar inflation or to track the evolution of inflation in post-sample 1998–2007 simulations. This finding applies to all previous NKPC research, all of which excludes explicit SS variables.

What difference is made by long lags and by the TV-NAIRU? When supply shocks are omitted as in rows 1–12 of Table A1, there is little difference among the alternative variants, which yield SSRs ranging from 183.8 to 244.0. Simulation mean errors (MEs) range from -2.04 to -2.75 when the NAIRU is fixed. Much lower MEs are obtained when the NAIRU is allowed to vary over time.

The results that include supply shocks are displayed in rows 13–24 of Table A1. When supply shocks are included but lag lengths are short, as in rows 13–14, 17–19 and 22, the post-sample simulation errors are very large. When supply shocks are included, the best results are in rows 15–16 with a fixed NAIRU and in rows 20–21 with a TV-NAIRU. Long lags on the dependent variable (inflation) matter in the specification of a PC including supply shocks.

The right-hand section of Table A1 contains a large number of significance tests on the exclusion of variables that are omitted in the Roberts NKPC specification and included in the triangle specification. Starting in row 3, even without supply shock variables, the significance value of excluding lags 9–24 on the lagged dependent variable is 0.01, and on lags 1–4 of the

Table A1

Transformation of the Philips Curve from Roberts NKPC (Row 1) to Triangle (Row 21), 1962(I) to 2007(IV)

	Inf lag	U lag length	Fixed or TV NAIRU	Supply shocks	Regression statistics			Simulation errors		Exclude inf lags 5–8		Exclude inf lags 9–24		Exclude U lags 1–4		Exclude supply shocks	
					R2	SEE	SSR	ME	RMSE	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig. Level	F-stat.	Sig.level
1	1 to 4	0	Fixed	No	0.78	1.17	244.0	− 2.75	3.20								
2	1 to 8	0	Fixed	No	0.78	1.17	237.9	− 2.73	3.15	1.11	0.35						
3	1 to 24	0	Fixed	No	0.80	1.11	195.0	− 2.04	2.41			2.18	0.01				
4	1 5 9 13 17 21	0	Fixed	No	0.78	1.16	238.5	− 2.19	2.53								
5	1 to 4	0 to 4	Fixed	No	0.79	1.15	229.2	− 2.27	2.68					2.81	0.03		
6	1 to 4	0 to 4	TV	No	0.79	1.13	224.8	0.26	1.43								
7	1 to 8	0 to 4	TV	No	0.79	1.14	222.0	0.36	1.46	0.55	0.70						
8	1 to 24	0 to4	TV	No	0.80	1.10	188.7	0.15	1.33			1.71	0.05				
9	1 5 9 13 17 21	0 to 4	TV	No	0.79	1.14	225.6	0.03	1.33								
10	1 to 8	0 to 4	Fixed	No	0.78	1.15	226.0	− 2.11	2.53	0.61	0.66						
11	1 to 24	0 to 4	Fixed	No	0.81	1.09	183.8	− 2.18	2.54			2.21	0.01				
12	1 5 9 13 17 21	0 to 4	Fixed	No	0.80	1.12	216.7	− 2.23	2.56								
13	1 to 4	0	Fixed	Yes	0.90	0.78	100.3	− 2.45	3.06							18.20	0.00
14	1 to 8	0	Fixed	Yes	0.90	0.78	99.0	− 2.60	3.15	0.52	0.72					17.39	0.00
15	1 to 24	0	Fixed	Yes	0.92	0.71	73.8	− 0.18	1.00			3.09	0.00			18.31	0.00
16	1 5 9 13 17 21	0	Fixed	Yes	0.92	0.71	82.2	− 0.24	0.97							23.85	0.00
17	1 to 4	0 to 4	Fixed	Yes	0.90	0.78	97.3	− 2.11	2.72					1.22	0.31	16.79	0.00
18	1 to 4	0 to 4	TV	Yes	0.91	0.76	94.1	− 1.14	1.47							17.31	0.00
19	1 to 8	0 to 4	TV	Yes	0.91	0.76	92.1	− 1.15	1.40	0.84	0.50					17.12	0.00
20	1 to 24	0 to 4	TV	Yes	0.93	0.65	59.6	0.30	0.74			4.84	0.00			23.66	0.00
21	1 5 9 13 17 21	0 to 4	TV	Yes	0.93	0.64	64.6	0.29	0.70							30.70	0.00
22	1 to 8	0 to 4	Fixed	Yes	0.90	0.78	96.6	− 2.23	2.77	0.30	0.88					16.18	0.00
23	1 to 24	0 to 4	Fixed	Yes	0.92	0.69	67.2	− 0.50	1.08			3.85	0.00			18.81	0.00
24	1 5 9 13 17 21	0 to 4	Fixed	Yes	0.93	0.68	73.4	− 0.56	1.06							23.88	0.00

unemployment gap is 0.03. Throughout rows 1–12 of Table A1, we learn that excluding short lags (e.g. excluding lags 5–8 from equations containing inflation lags 1–4) is insignificant, whereas excluding lags 9–24 yields highly significant exclusion tests.

Rows 13–24, which all include the full set of supply shock variables, differ only in the length of lags included on the lagged dependent (inflation) variable and on lagged unemployment, and also in whether the NAIRU is forced to be fixed or is allowed to vary over time. We can interpret the bottom half of Table A1 by looking at blocks of four rows.

The first group of four rows, 13–16, share in common the inclusion of supply shocks, the assumption of a fixed NAIRU, and alternative lags on the dependent variable. The mean error in the dynamic simulations falls by 80% when lags up to 24 are included, and the exclusion of lags 9–24 is rejected at a 0.00 significance value. The same result occurs in rows 22–25 when with a time-varying NAIRU the significance of long lags on the dependent variable is strongly supported at a significance level of 0.00.

ACKNOWLEDGMENTS

This paper was written on the occasion of the 50th anniversary of Phillips' original article. An earlier version was given as the A. W. H. Phillips Invited Lecture at the Australasian Meetings of the Econometric Society in Wellington, New Zealand, on 9 July 2008. Neil Sarkar was the exemplary research assistant on the empirical section of this paper, and Lucas Zalduendo helped to prepare the manuscript for publication. In writing the early history of the Phillips curve, I found a useful source to be Lacker and Weinberg (2007). I am grateful to Bob Buckle for suggesting the topic and to David Hendry, Giorgio Primiceri and participants in the Northwestern macro workshop for helpful comments.

NOTES

1. The article was unearthed by Jacob Mincer, and as co-Editor of the *Journal of Political Economy* during 1971–73, I was responsible for the decision to reprint it and give it the dramatic title shown in the reference list of this paper.
2. An amusing commentary on the research technology of the 1920s is Fisher's comment that: 'During the last three years in particular I have had at least one computer in my office almost constantly at work on this problem' (1926, p. 786).
3. In Figure 1 the inflation rate is the four-quarter change in the deflator for personal consumption expenditures.
4. Spectators at the time called the *American Economic Review* debate the 'battle of the radio stations', after the AM–FM initials of the protagonists, Ando and Modigliani vs Friedman and Meiselman.
5. Lucas (1973, equation (3), p. 327) created a serially correlated business cycle by introducing a lagged value of cyclical output into the equation explaining cyclical output by the price surprise. This lagged term is gratuitous and neither called for by the theory nor consistent with it.
6. Eckstein and Brinner (1972) produced the first paper to emerge with a specification in which the α coefficient was unity. Gordon (1972) concurred, based on a different parameterization of a time-varying coefficient, and provided comparisons of his approach with those of Eckstein–Brinner and Perry (1970).
7. See also Pakko (2000) for a study of differences in the shape of frequency distributions of inflation and output over the pre- and post-1929 period.
8. While the Gordon and Phelps papers were the first to develop the theory of policy responses to supply shocks, Pierce and Enzler (1974) had previously run simulations with a large-scale econometric model that showed large macroeconomic impacts of commodity price shocks, with a trade-off between the output and inflation effects as the policy response varied.
9. Blinder (1981) subsequently extended the theoretical analysis to allow for rational expectations in the formation of wages and for a distinction between anticipated and unanticipated shocks. The topic was more recently revisited by Ball and Mankiw (1995).
10. The 1978 rival textbooks were by Dornbusch and Fischer and by myself. The dynamic version of the demand–supply model in the form of a first-order difference equation was confined to intermediate level textbooks, with many imitators published soon after. Elementary macro principles textbooks limited themselves to the display of static aggregate demand and supply curves, starting with Baumol and Blinder in 1979.
11. The 'triangle' nomenclature has been picked up by a several authors, including Rudd and Whelan (2005b) and Fitzenberger *et al.* (2009).
12. The theoretical NKE research literature is surveyed and placed in the perspective of historical puzzles of price and wage behaviour in Gordon (1990).

13. The study of wage and price control effects began with Gordon (1972). Nixon-era control dummies were included in all the econometric tests of the Gordon mainstream model, beginning with Gordon (1977b). The current version of the triangle model (see Table 1) estimates that the imposition of controls reduced the price level by a cumulative 1.6%, and the removal of the controls raised the price level by 1.8%. Using CPI rather than PCE data, and with a slightly different method, Blinder and Newton (1981, Table 2, Model 1) estimated that the controls held down the price level by a cumulative 3.1% through early 1974, followed by a 3.2% bounce-back in 1974–75.
14. The concept of the sacrifice ratio was developed by Okun (1978), and his preferred estimate of the sacrifice ratio was 10 (p. 348).
15. In a paper written in early 1981, Gordon (1982b, Table 10) integrated an endogenous flexible exchange rate into the effects of a hypothetical Volcker disinflation and simulated a sacrifice ratio of 4.8, a relatively low number that resulted in part from a 33% exchange rate appreciation.
16. The pre-1994 estimate of the NAIRU incorporated changes in the difference between the official unemployment rate and the demographically-adjusted unemployment rate, originally introduced by Perry (1970). The Perry-weighted NAIRU was assumed to be fixed.
17. Blanchard (2008, p. 8) uses the 'workhorse' label for a small three-equation macro model of which the NKPC is one equation, even though he calls the NKPC equation 'patently false' (p. 9). We return in Section IV to quantify the extent to which the NKPC is 'patently false' and leave to the reader the puzzle as to how this model could have become a 'workhorse'.
18. The insight that feedback between wage and price equations is transmitted through labour's share dates back to Franz and Gordon (1993).
19. They use the same list of instruments as in Galí and Gertler (1999), with two lags instead of four lags on each.
20. Labour's share in Figure 4 is defined as employee compensation divided by domestic net factor income (www.bea.gov, NIPA Table 1.10: GDP minus consumption of fixed capital minus taxes on production and imports less subsidies).
21. See Del Boca *et al.* (2008) for a wide-ranging econometric assessment of the inflation rate of the Italian lira from 1861 to 1998. They conclude that Italy has had a 'conventional inflation–output trade-off only during times of low inflation and stable aggregate supply'.
22. All four price control dummy variables, as well as the NRA dummy, are entered in the form of 'on effects' summing to 1.0 followed by 'off effects' summing to -1.0, implying no permanent impact of the controls. The coefficients on these dummy variables measure the cumulative impact of the controls in displacing the price level prior to the reverse snap-back effect.
23. This triangle or three-term PC equation, with each term including lags and a lag operator, was introduced in Gordon and King (1982, equation 13). The notation has been used by many authors in the mainstream 'left-fork' tradition, most recently by Blinder and Rudd (2008, unnumbered equation, p. 73)
24. This method of estimating the TV-NAIRU was introduced in simultaneous papers by Gordon (1997) and Staiger *et al.* (1997).
25. The relative import price variable is defined as the rate of change of the non-food non-oil import deflator minus the rate of change of the core PCE deflator. The relative food–energy variable is defined as the difference between the rates of change of the overall PCE deflator and the core PCE deflator. The Nixon-era control variables and the lag lengths (shown explicitly in Table 1) remain the same as originally specified in Gordon (1982b). The productivity variable is the eight-quarter change in a Hodrick–Prescott filtered trend of the change in non-farm private business output per hour (using 6400 as the Hodrick–Prescott smoothness parameter).
26. In subsequent work, Staiger *et al.* (2001) and Stock and Watson (2007) abandoned the attempt to estimate the TV-NAIRU as a byproduct of the inflation equation, and Blinder and Rudd (2008) follow their lead. Now they estimate the NAIRU as a trend on the actual unemployment rate, taking no account of the role of supply shocks in making this trend unusually high in 1974–75 and 1981–82, or unusually low in 1997–2000. As a result, their version of the unemployment gap ($U - U^N$) greatly understates the size of that gap and its influence on inflation during the key supply shock episodes.
27. Basistha and Nelson (2007) are among those authors who exclude explicit supply shock variables from their equations and derive estimates of the TV-NAIRU that are extremely high, e.g. 8% in 1975 and 10% in 1981 (2007, Figure 6, p. 509).
28. The inclusion of lags 13–24 (years four to six) in an exclusion test is strongly significant at the 0.00 confidence level. As stated in the notes to Table 1, we conserve on degrees of freedom by including six successive four-quarter moving averages of the lagged dependent variable at lags 1, 5, 9, 13, 17 and 21, rather than including all 24 lags separately.
29. This can be compared to an import share in nominal GDP of 10% at the midpoint of the sample period in 1985.
30. Dew-Becker (2006) has previously traced the statistical significance of stripped-down Phillips curves and reached conclusions that are similar to those arrayed in Appendix Table A1.
31. Blinder and Rudd (2008, Figure 19) plot a ratio of British Thermal Units (BTUs) to real GDP that declines from 18 in 1969 to 8.7 in 2007.

32. The core inflation equation is specified exactly as the headline equation in Table 1, with the exception of the food–energy effect. To allow for longer lags in the impact of energy on core inflation, the food–energy effect is measured as the four-quarter moving average of the difference between headline and core inflation, and is entered at lags 1, 5 and 9.

REFERENCES

BALL, L. (2008). Hysteresis in unemployment. Paper presented at Federal Reserve Bank of Boston Conference, 10–11 June.

——— and MANKIW, G. N. (1995). Relative-price changes as aggregate supply shocks. *Quarterly Journal of Economics*, **110**, 161–93.

BÅRDSEN, G., JANSEN, E. S. and NYMOEN, R. (2002). Testing the new Keynesian Phillips curve. Paper presented at Workshop on the Phillips curve: new theory and evidence, Facföreningsrörelsens Institut För Ekonomisk Forskning (Trade Union Institute For Economic Research), Stockholm, 25–6 May.

BARRO, R. J. (1977). Unanticipated money growth and unemployment in the United States. *American Economic Review*, **67**, 101–15.

BARSKY, R. B. and KILIAN, L. (2002). Do we really know that oil caused the great stagflation? A monetary alternative. *NBER Macroeconomics Annual 2001*, 137–83.

BASISTHA, A. and NELSON, C. R. (2007). New measures of the output gap based on the forward-looking new Keynesian Phillips curve. *Journal of Monetary Economics*, **54**, 498–511.

BERNANKE, B. S., GERTLER, M. and WATSON, M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity*, **28**(1), 91–142.

BLANCHARD, O. J. (1987). Aggregate and individual price adjustment. *Brookings Papers on Economic Activity*, **18**(1), 57–109.

——— (2008). The state of macro. NBER Working Paper no. 14259.

——— and GALÍ, J. (2007). The macroeconomic effects of oil shocks: why are the 2000s so different from the 1970s? NBER Working Paper no. 13368.

——— and SUMMERS, L. (1986). Hysteresis and the European unemployment problem. *NBER Macro Annual 1985*, 15–77.

BLINDER, A. S. (1979). *Economic Policy and the Great Stagflation*. New York: Academic Press.

——— (1981). Monetary accommodation of supply shocks under rational expectations. *Journal of Money, Credit, and Banking*, **13**, 425–38.

——— (1982). The anatomy of double-digit inflation in the 1970s. In R. E. Hall (ed.), *Inflation: Causes and Effects*. Chicago: University of Chicago Press for NBER, pp. 261–82.

——— and NEWTON, W. J. (1981). The 1971–74 controls program and the price level: an econometric post-mortem. *Journal of Monetary Economics*, **8**, 1–23.

——— and RUDD, J. B. (2008). The supply-shock explanation of the great stagflation revisited. NBER Working Paper no. 14563.

BOSWORTH, B. P. and LAWRENCE, R. Z. (1982). *Commodity Prices and the New Inflation*. Washington DC: Brookings Institution.

BRUNO, M. and SACHS, J. D. (1985). *Economics of Worldwide Stagflation*. Cambridge, MA: Harvard University Press.

CALVO, G. A. (1983). Staggered price in a utility maximizing framework. *Journal of Monetary Economics*, **12**, 383–98.

CASSIDY, J. (1996). The decline of economics. *New Yorker*, 2 December, pp. 54–7.

COGLEY, T. and SARGENT, T. J. (2005). The conquest of U.S. inflation: learning and robustness to model uncertainty. *Review of Economic Dynamics*, **8**, 528–63.

DEL BOCA, A.,, FRATIANNI, M., SPINELLI, F. and TRECROCI, C. (2008). The Phillips curve and the Italian lira, 1861–1998. University of Brescia MoFiR working paper no. 8.

DEW-BECKER, I. (2006). Was Lucas right? Evidence on structural shift in the Phillips curve. Senior honours thesis submitted to Northwestern Economics Department, May.

——— and GORDON, R. J. (2005). Where did the productivity growth go? Inflation dynamics and the distribution of income. *Brookings Papers on Economic Activity*, **36**, 67–127.

ECKSTEIN, O. and BRINNER, R. (1972). The inflation process in the United States. A study prepared for the use of the Joint Economic Committee, 92 Congress 2 session.

FISCHER, S. (1977). Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy*, **85**, 191–205.

FISHER, I. (1926). A statistical relation between unemployment and price changes. *International Labor Review*, **13**, 785–92. (Reprinted as Fisher, I. (1973). I discovered the Phillips curve. *Journal of Political Economy*, **81**, 496–502.)

FITZENBERGER, B., FRANZ, W. and BODE, O. (2009). The Phillips curve and NAIRU revisited: new estimates for Germany. *Journal of Economics and Statistics*, forthcoming.

FRANZ, W. and GORDON, R. J. (1993). Wage and price dynamics in Germany and America: differences and common themes. *European Economic Review*, **37**, 719–54.

FRIEDMAN, M. (1968). The role of monetary policy. *American Economic Review*, **58**, 1–17.

FROMM, G. and TAUBMAN, P. (1968). *Policy Simulations with an Econometric Model*. Washington: Brookings Institution.

FUHRER, J. C. (1997). The (un)importance of forward-looking behavior in price specifications. *Journal of Money, Credit, and Banking*, **29**, 338–50.

GALÍ, J. and GERTLER, M. (1999). Inflation dynamics: a structural econometric analaysis. *Journal of Monetary Economics*, **44**, 195–222.

———, ——— and LOPEZ-SALIDO, D. (2005). Robustness of estimates of the hybrid new Keynesian Phillips curve. *Journal of Monetary Economics*, **52**, 1107–18.

GORDON, R. J. (1972). Wage–price controls and the shifting Phillips curve. *Brookings Papers on Economic Activity*, **3**(2), 385–421.

——— (1975). Alternative responses of policy to external supply shocks. *Brookings Papers on Economic Activity*, **6**(1), 183–206.

——— (1977a). The theory of domestic inflation. *American Economic Review, Papers and Proceedings*, **67**, 128–34.

——— (1977b). Can the inflation of the 1970s be explained? *Brookings Papers on Economic Activity*, **8**(1), 765–78.

——— (1982a). Price inertia and policy ineffectiveness in the United States, 1890–1980. *Journal of Political Economy*, **90**, 1087–117.

——— (1982b). Inflation, flexible exchange rates, and the natural rate of unemployment. In M. N. Baily (ed.), *Workers, Jobs, and Inflation*. Washington: Brookings, pp. 89–158.

——— (1984). Supply shocks and monetary policy revisited. *American Economic Review, Papers and Proceedings*, **74**, 38–43.

——— (1990). What is new-Keynesian economics? *Journal of Economic Literature*, **28**, 1115–71.

——— (1997). The time-varying NAIRU and its implications for economic policy. *Journal of Economic Perspectives*, **11**, 11–32.

——— (1998). Foundations of the Goldilocks economy: supply shocks and the time-varying NAIRU. *Brookings Papers on Economic Activity*, **29**, 297–333.

——— and KING, S. R. (1982). The output cost of disinflation in traditional and vector autoregressive models. *Brookings Papers on Economic Activity*, **13**, 205–42.

HAMILTON, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, **91**, 228–48.

HOOKER, M. A. 1996. What happened to the oil price macroeconomy relationship? *Journal of Monetary Economics*, **38**, 195–213.

HUMPHREY, T. M. (1991). Nonneutrality of money in classical monetary thought. *Federal Reserve Bank of Richmond Economic Review*, **77**, 3–15.

KATZ, L. F. and KRUEGER, A. B. (1999). The high-pressure U.S. labor market of the 1990s. *Brookings Papers on Economic Activity*, **30**, 1–65.

KILEY, M. T. (2007). An empirical comparison of sticky-price and sticky-information models of price setting. *Journal of Money, Credit, and Banking*, **39**, 101–25.

KILIAN, L. (2008). The economic effects of energy price shocks. *Journal of Economic Literature*, **46**, 871–909.

KYDLAND, F. E. and PRESCOTT, E. C. (1977). Rules rather than discretion: the inconsistency of optimal plans. *Journal of Political Economy*, **85**, 473–91.

LACKER, J. M. and WEINBERG, J. A. (2007). Inflation and unemployment: a layperson's guide to the Phillips curve. *Federal Reserve Bank of Richmond Economic Quarterly*, **93**, 201–27.

LAIDLER, D. and PARKIN, M. J. (1975). Inflation: a survey. *Economic Journal*, **85**, 741–809.

LIPSEY, R. G. (1960). The relation between unemployment and the rate of change of money wage rates: a further analysis. *Economica*, **27**, 1–31.

LUCAS, R. E. Jr. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4**, 103–24.

——— (1973). Some international evidence on output–inflation tradeoffs. *American Economic Review*, **63**, 326–34.

——— and SARGENT, T. J. (1978). After Keynesian macroeconomics. In *After the Phillips Curve: Persistence of High Inflation and High Unemployment*. Boston: Federal Reserve Bank of Boston, pp. 49–72.

MANKIW, G. N. (2001). The inexorable and mysterious tradeoff between inflation and unemployment. *Economic Journal*, **111**, 45–61.

MISHKIN, F. S. (1982). Does anticipated monetary policy matter? An econometric investigation. *Journal of Political Economy*, **90**, 22–51.

OKUN, A. M. (1978). Efficient disinflationary policies. *American Economic Review, Papers and Proceedings*, **68**, 348–52.

ORPHANIDES, A. and WILLIAMS, J. C. (2005). Inflation scares and forecast-based monetary policy. *Review of Economic Dynamics*, **8**, 498–527.

PAKKO, M. R. (2000). The cyclical relationship between output and prices: an analysis in the frequency domain. *Journal of Money, Credit, and Banking*, **32**, 382–99.

PERRY, G. L. (1970). Changing labor markets and inflation. *Brookings Papers on Economic Activity*, **1**, 411–41.

PHELPS, E. S. (1967). Phillips curves, expectations of inflation and optimal unemployment over time. *Economica*, **34**, 254–81.

——— (1968). Money–wage dynamics and labor–market equilibrium. *Journal of Political Economy*, **76**, 678–711.

——— (1978). Commodity-supply shock and full-employment monetary policy. *Journal of Money, Credit, and Banking*, **10**, 206–21.

PHILLIPS, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, **25**, 283–99.

PIERCE, J. L. and ENZLER, J. J. (1974). The effects of external inflationary shocks. *Brookings Papers on Economic Activity*, **1**, 13–54.

PRIMICERI, G. (2006). Why inflation rose and fell: policymakers' beliefs and US postwar stabilization policy. *Quarterly Journal of Economics*, **121**, 867–901.

REIS, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, **53**, 1761–800.

ROBERTS, J. M. (2006). Monetary policy and inflation dynamics. *International Journal of Central Banking*, **2**, 193–230.

ROMER, C. D. (1996). Inflation and the growth rate of output. NBER Working Paper no. 5575.

——— (1999). Why did prices rise in the 1930s? *Journal of Economic History*, **59**, 167–99.

RUDD, J. and WHELAN, K. (2005a). New tests of the new-Keynesian Phillips curve. *Journal of Monetary Economics*, **52**, 1167–81.

——— and ——— (2005b). Modelling inflation dynamics: a critical review of recent research. Finance and Economics Discussion Series 2005–19, Board of Governors of the Federal Reserve System.

SAMUELSON, P. A. and SOLOW, R. M. (1960). Analytical aspects of anti-inflation policy. *American Economic Review, Papers and Proceedings*, **50**, 177–94.

SARGENT, T. J. (1982). The ends of four big inflations. In R. E. Hall (ed.), *Inflation*. Chicago: University of Chicago Press, pp. 41–97.

——— (1999). *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.

——— and WALLACE, N. (1975). 'Rational' expectations, the optimal monetary instrument, and the optimal money supply rule. *Journal of Political Economy*, **83**, 241–54.

———, WILLIAMS, N. and ZHA, T. (2006). Shocks and government beliefs: the rise and fall of American inflation. *American Economic Review*, **96**, 1193–224.

SIMS, C. A. (1987). Comment. *Brookings Papers on Economic Activity*, **1**, 117–20.

——— (2006). Rational inattention: beyond the linear-quadratic case. *American Economic Review*, **97**, 586–606.

——— (2008). Inflation expectations, uncertainty, the Phillips curve, and monetary policy. Paper presented at Federal Reserve Bank of Boston Conference, 10–11 June.

STAIGER, D., STOCK, J. H. and WATSON, M. W. (1997). The NAIRU, unemployment, and monetary policy. *Journal of Economic Perspectives*, **11**, 33–49.

———, ——— and ——— (2001). Prices, wages, and the U.S. NAIRU in the 1990s. In A. B. Krueger and R. M. Solow (eds), *The Roaring Nineties: Can Full Employment Be Sustained?* New York: Russell Sage Foundation and Century Foundation Press, pp. 3–60.

STOCK, J. H. and WATSON, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking*, **39**, 3–34.

TAYLOR, J. (1980). Aggregate dynamics and staggered contracts. *Journal of Political Economy*, **88**, 1–24.

WOODFORD, M. (2003). *Interest and Prices*. Princeton, NJ: Princeton University Press.

Economica

Economica (2011) **78**, 51–66
doi:10.1111/j.1468-0335.2009.00839.x

The Evolution of the Phillips Curve: A Modern Time Series Viewpoint

By Clive W. J. Granger*† and Yongil Jeon‡

†*University of California, San Diego* ‡*Sungkyunkwan University (SKKU), South Korea*

Final version received 6 January 2009.

Phillips' (1958) original curve involves a nonlinear relationship between inflation and unemployment. We consider how his original results change due to updated theoretic and empirical studies, increased computer power, enlarged datasets, increases in data frequency and developed time series econometric models. In the linear models, there was weak causation from unemployment to inflation. Rather than using any of the many nonlinear models that are now available, we adopt a time-varying parameter linear model as their convenient proxy, which empirically supports Phillips' use of nonlinear model form and causation, but the strength of this result is much weaker in recent periods.

I. The Beginnings

In his original paper (1958), Bill Phillips used British data to consider the relationship between wage inflation and unemployment, and in the following half century his findings have been extended and re-examined by many authors and in many ways. Our intention is to consider how his original results have changed due to the results of new empirical studies, as computer power has increased, as the dataset has enlarged and increased in frequency, and as econometric models and tastes have developed. We have attempted to follow the route that Phillips might well have taken, although it is very difficult to appreciate his tastes and preferences as to model specification. We have taken a fairly 'mainline' route using time series models and have attempted to avoid controversy although we realize that this is virtually impossible. We have ignored the macro-theory based developments, including the rational expectations sidetrack, as we would prefer to rely just on the message that is available in the data when they are viewed carefully.

Over the full fifty years we have focused our attention on the two main variables: inflation, here denoted by I_t, and unemployment, denoted by U_t. Originally I_t consisted of wage inflation, but later it moved to price inflation as the economy evolved and trade unions declined in importance. Inflation and unemployment are probably widely considered to be the two most important economic variables by the majority of the workforce of a country, and consequently considerable attention is paid to them by the state government and by the Central Bank. Only interest rates and some financial variables are moving towards a similar level and width of interest. Some other major economic variables, such as production and trade balance, are certainly important but are mainly of concern only to economists.

The original Phillips' paper (1958) initially discusses the joint, contemporaneous relationship between I and U, and the paper's title also suggests a joint relationship. It is not until the end of the second page of the paper that it is suggested that the flow of the impact is from U to I. The final paragraph of the first section clearly states that the purpose of the study 'is to see whether statistical evidence supports the hypothesis that the rate of change of money wage rates in the UK can be explained by the level of

*Professor Sir Clive Granger passed away on 27 May 2009.

employment and the rate of change of employment'. Thus the relationship to be studied is that of employment explaining inflation.

As was usual at that time, when most macroeconomic data were available only annually, no lags were used in the basic formulation, but only contemporaneous terms were used; however, in other early studies by Phillips, lagged terms were used. Apart from import prices, no other economic variables were mentioned as possible explanatory variables. To be precise, Phillips (1958) writes on page 283 (our emphasis):

> The purpose of the present study is to see whether statistical evidence supports the hypothesis that the rate of change of money wage rates in the UK can be explained by the level of unemployment and the rate of change of unemployment *except in, or immediately after, those years in which there was a very rapid rise in import prices.*

In virtually all later studies the last part of this statement seems to have been ignored! We have also not investigated the relevance of import prices.

Phillips used British data, which could be hoped to be of fairly high quality, and his series contained almost one hundred terms, which at that time was exceptionally long for a macro series. However, the time period used, 1861–1957, was tempestuous, containing at least three major wars as well as several large business cycle swings. On the other hand, this high activity level does allow a model to show its ability to be relevant in a wide variety of circumstances.

Phillips also faced a clear shortage of computing power. We understand that the London School of Economics did not have an electronic computer in 1957 when the work was started, and believe that all the calculations would have been carried out on electric calculators.[1]

Phillips' paper (1958) is a very nice piece of empirical work, particularly given the computing shortcomings of those times. He obtained a simple curve relating the two variables of interest and found that this curve forecasts fairly well into future decades. The specification was a mixture of the sophisticated, with nonlinear explanatory terms being considered, but also with rather simple statements about the quality of the model. No t-values or R-squared statistics are provided, probably because of the computing limitations. The equations include no lag variables, and evaluation was undertaken by forming the model on an early part of the data and comparing the curve so obtained by superimposing it onto later segments of data. The fit was often seen to be surprisingly good by eye, but no numerical comparison measure was used.

Just two years later, Lipsey (1960) produced a follow-up paper on the same topic and with similar data but using more modern-looking specifications. Lipsey adopted a simpler and more convenient form of nonlinearity that proved to be equally as successful as that used by Phillips, and R-squared values were now given. However, neither Phillips nor Lipsey provided Durbin–Watson statistics, even though during this period Jim Durbin was their colleague at the LSE. This probably reflects the lack of interest in dynamics used in the specification of econometric models in this era.

During the late 1950s economic models generally were inclined not to use lagged variables in their specifications, so the lack of their use by Phillips is not surprising even though he was writing about continuous time error-correction models in his papers on control theory in the same period. We feel that Phillips would have probably been building 'feedback' discrete time models once sufficient computer power became available and as data became more plentiful. This belief is supported by the specification used in Phillips (1959), as discussed below.

II. Consideration of the Work by Phillips and Lipsey from a More Modern Viewpoint

The various models considered by Phillips and Lipsey can be summarized in the form

$$[\text{M}] \quad I_t = a + b_1F_1[U_t] + b_2F_2[U_{t-1}] + b_3F_3[U_{t-2}] + k_0F_4[Z_t] + e_t,$$

where I_t is the rate of change of wage rates, U_t is percentage unemployment, and Z_t are some extra explanatory series. Usually $k_0 = 0$. Here the $F[\cdot]$ are various functions, possibly linear, and the b_i are coefficients. It should be noted that this equation is both dynamic and nonlinear, and if its specification is correct and b_2 or b_3 is non-zero, it also suggests, but does not prove, Granger-causality (later here called just causality, as in Granger 1988) from unemployment to changes in wage rates.

Phillips' model A

In Phillips' original paper (1958) the curve he considered had a single explanatory term U_t^c with no lag so that effectively b_2 and b_3 were taken to be zero. Initially using annual data for the period 1861–1913, he obtained estimates $b_1 = 10^{0.984} = 9.638$ and $a = -0.9$. The estimated value of c was -1.394, and Table 1 provides detailed summary statistics.[2]

The curve fitted for this 1861–1913 period was compared diagrammatically with data from other decades and visually was found to fit adequately well. This included the latest period, 1948–57, but it was shown diagrammatically that if unemployment had been lagged by seven months, then the fit would have been excellent.

It should be noted that model A is not strictly balanced as the variable c can take negative as well as positive values, whereas U_t^c is a 'limited variable' as U_t is necessarily positive. As the constant a is found to be negative and b_1 is positive, the right-hand side of the equation can give a negative estimate for c but cannot go below a.

Phillips' model B

In his Melbourne (or Australian) paper,[3] Phillips (1959) used just U to powers 2 and 3, both with lag 3, and the Z_t consisted of the rate of change of export prices (X) with lags 1 and 2, and the rate of change of import prices (IM) with lag 3. The regression found was

$$I_t = 1.46/U_{t-3}^2 + 0.415/U_{t-3}^3 + 0.15[[X_{t-1} + X_{t-2}]/2] + 0.134IM_{t-3} + 2.11.$$

This equation is balanced as the two terms involving X and IM can be negative.

Phillips' model C

On the final page of the 1959 article, Phillips provides a model of the form

$$I_t = 0.57I_{t-1} + 0.93/(U_{t-2} - 0.26) + 2.44e^{0.02X_{t-1}} + 0.022IM_{t-1} + 0.295,$$

where X and IM are respectively the rates of change of export and import prices. He reports that this model fits well, the best of all the models considered, but still not well enough according to his tastes. Although a very small sample is used, annual data over just twelve years, a complicated model is fitted with two explanatory variables as well as lagged U. This model is also balanced. Values of R-squared are not provided for either equation. It should be noted that Phillips spends a great deal of time and effort in his papers discussing data sources and problems.

TABLE 1

PHILLIPS CURVE ESTIMATION USING FOUR AGGREGATED OBSERVATIONS

	Equations	Coefficients (*t*-values) [*p*-values]			R^2	Log-likelihood	Akaike information criterion	Bayesian information criterion	Durbin–Watson
Case 1	$y + a = b \cdot x^c$	– 0.252	16.031	– 2.895	0.997	3.404	– 0.202	– 0.662	3.012
		(0.855)	(4.544)	(– 4.668)					
		[0.550]	[0.138]	[0.134]					
Case 2	$y + 0.9 = b \cdot x^c$	**0.9**	11.466	– 1.617	0.978	– 0.532	1.266	0.959	2.382
			(7.575)	(– 7.658)					
			[0.017]	[0.017]					
Case 3	$\log_{10}(y + a) = \log_{10} b + c \cdot \log_{10} x$	0.319	1.031	– 1.850	na	6.731	– 1.865	– 2.326	3.193
		(0.422)	(6.086)	(– 2.075)					
		[0.746]	[0.104]	[0.286]					
Case 4	$\log_{10}(y + 0.9) = \log_{10} b + c \cdot \log_{10} x$	**0.9**	0.984	– 1.394	0.966	6.607	– 2.304	– 2.610	2.705
			(11.347)	(– 7.579)					
			[0.008]	[0.017]					

Notes

The six aggregated observations for unemployment and wage in the UK are (1.5167, 5.0585), (2.3500, 1.5472), (3.4833, 0.8482), (4.4900, 0.3466), (5.9545, – 0.1817) and (8.3722, – 0.3539), the first four observations of which are used for estimation.

Phillips' papers 1954–1959

In the period 1954–1959, Phillips published five papers (1954, 1956, 1957, 1958, 1959); these are republished in Leeson (2000). Only the 1958 paper was empirical in nature; the 'Australian' paper was empirical but not published until much later. The other papers involved economics and control theory and did mention error-correction models. A useful, brief discussion of the control papers has been provided by Pagan (2000).

The Lipsey models

In his well-known comment on Phillips' earlier work, Lipsey (1960) considered a model of form [M] but with just the first two powers *c*, one and two. He found that this model provided a very good approximation to Phillips' 1958 model. As the Lipsey model is easier to use, especially in a regression framework, it is clearly preferable.

Two simple forms of the achieved Lipsey model are

$$\text{[L1]} \quad I_t = -1.42 + 7.06U_t^{-1} + 2.31U_t^{-2}, \quad R^2 = 0.64,$$

and

$$\text{[L2]} \quad I_t = -1.52 + 7.60U_t^{-1} + 1.61U_t^{-2} - 0.023\dot{U}_t, \quad R^2 = 0.78,$$

where $\dot{U}_t$ is the rate of change of unemployment. These models use data for the period 1862–1913, with the Bowley data being used for the years 1881–85. Model L1 is not balanced, but model L2 is balanced as it contains a variable that is not limited at zero.

As a final experiment, Lipsey includes $\dot{P}_t$, the rate of change of the cost of living index, in the equation and obtains

$$\text{[L3]} \quad I_t = -1.21 + 6.45U_t^{-1} + 2.26U_t^{-2} - 0.019\dot{U}_t + 0.21\dot{P}_t, \quad R^2 = 0.85.$$

This equation is balanced.

A difficulty with using these 'dot variables' from a modern viewpoint is that they are defined as, for example, $\dot{U}_t = (U_{t+1} - U_{t-1})/2U_t$, so that the past and future get mixed up. Lipsey states that he did try a form like $\dot{P}_t = (P_t - P_{t-1})/((P_t + P_{t-1})/2)$ and found that 'the results were broadly similar but the correlations slightly lower'.

For the years 1923–39 and 1948–57, Lipsey fitted the model

$$\text{[L4]} \quad I_t = 0.74 + 0.43U_t^{-1} + 11.18U_t^{-4} + 0.038\dot{U}_t + 0.69\dot{P}_t, \quad R^2 = 0.91.$$

It is not surprising to find that this last term has the greatest explanatory power. It should be noted that a quite different specification is used and the parameter values have changed. This can be interpreted as suggesting the necessity of using a time-varying parameter form of model.

The models often contain functions of variables. Although in very general terms the positive function of a linear non-stationary series will have essentially the same major properties as such a series, it is probably best to apply a test. At the end of his paper, on page 31, Lipsey (1960) mentions the direction of causation between price changes and wage changes, stating: 'The analysis so far conducted is not inconsistent with the hypothesis that there is a strong feed-back from price changes to wages. We should test!'

Evaluation

The attitude towards the evaluation of models has evolved over the years. At the time of the appearance of the Phillips' models, the emphasis was on how well the model fitted the data, but more recently the emphasis has been on how well the model performs in its planned task, such as forecasting or controlling variables of interest. Phillips' original nonlinear curve was estimated using some early data and evaluated by showing diagrammatically that the same curve fitted remarkably well to later periods, although no numerically formed measures were employed.

We superimposed the same curve for later decades using the same coefficients; visually, they were not found to fit well. If the coefficients of the curve are re-estimated using more recent data, the curve fits later data fairly well. In the present study we consider the purpose of the model to provide forecasts, and thus evaluations will be made in terms of the relative forecasting abilities of alternative models.

III. Taking a More Recent Time Series Viewpoint

Current time series analysis will usually start by asking if each individual series is stationary, denoted I(0), or linear 'non-stationary', denoted I(1), where the change of an I(1) series is I(0). It is important to determine the appropriate labels for a pair of series to ensure that an equation in a model is 'balanced', so that the variables on both sides of the equation have the same major properties. The designation is also important when one is trying to avoid spurious regressions, which can occur with a pair of I(1) series. Several tests for I(0) exist. However, as these tests may not produce correct results, we prefer to build models using both the levels and the differences of all the series that are involved. This will produce several alternative models that can then be compared and evaluated. The models that are considered are as follows.

- The standard autoregressive model with a preselected number of lagged terms.
- A standard bivariate autoregressive model, with the variables to be explained being inflation and unemployment. If cointegrated variables are involved, an error-correction model is used. However, no cases of cointegration were discovered.
- Models could be of the standard 'linear' form with constant coefficients or of the 'time-varying parameter' (TVP) form. As discussed below, TVP models are equivalent to nonlinear models, including those of Terasvirta (2006).
- TVP and nonlinear models. The various Phillips' models discussed above are all nonlinear, so it should be expected that the present relevant models would also be nonlinear but the relevant type of nonlinearity could also be expected to change with time. It is thus difficult to specify a relevant form of nonlinear model for our explanatory analysis. Fortunately, recently Halbert White proved that any nonlinear time series model could be well approximated by a TVP model; a proof is given in Granger (2007). The TVP can be determined using a standard Kalman-filter program, which is available on several econometric computer packages.

In recent years there has been a change in strategy towards model building. Originally a 'best' model was determined using several criteria, fitted to the data and then applied to various practical problems. Currently the alternative strategy being considered is to fit to the data several alternative high quality models, and then to evaluate them jointly, particularly by building combined forecasts.

If a test for I(0) is indecisive, then two alternative models should be considered, one with I(0) variables and the other with I(1) variables. There is no reason to consider just a single model for purposes of decision making. We call this 'thick modelling' in Granger and Jeon (2004), and this is proving to be a popular approach.

IV. Questions of Interest

The questions of interest to Phillips and Lipsey fifty years ago initially can be considered within the context of a linear bivariate vector autoregressive model or an error-correction model but possibly with time-varying parameters and extended by the addition of other explanatory variables. This model is appropriate because of the recent theorem by Halbert White, as noted above, where it is shown that any nonlinear time series model can be well approximated by a TVP linear model. Usually the variables of interest, inflation and unemployment, will be tested as being either I(1) or I(0). If found to be I(1), then they are considered as linear and TVP autoregressive models and linear and TVP error-correction models.

The specific questions considered here will include the following.

(a) Is unemployment a useful 'causal' variable for inflation? This is investigated in two stages: the first models inflation in terms of lagged inflation, and in the second stage one then adds lagged unemployment to the equation. Evaluation is conducted by noting if the extra information leads to improved forecasts. We use linear autoregressive models, vector autoregressive models and error-correcting models to study what may be called the 'causality' of inflation by unemployment. The main question asks: 'What is the appropriate specification of the models linking the variables of interest?' Using standard tests there is some rather weak evidence that inflation is I(1) but unemployment is usually I(0).

The causal question is first asked using linear models and then reconsidered using TVP linear models as a convenient proxy for nonlinear models, using the result that any nonlinear model can be well approximated by a TVP model.

(b) Phillips' original curve involved a nonlinear relationship, of course. He built his model using data from a few early years and then visually compare this curve to data for later periods, often getting remarkably good fits. We have continued this process by using this original curve (Phillips' model A) in much later periods but visually we found a poor fit on every occasion (see Figure 1).

Initially annual data were used for both the UK and the USA over several long data spans, starting with Phillips' data and ending in 2006. The nonlinear relationship that Phillips (1958) estimated did not consider the time dimension. The causality and nonlinearity are tested through the time period, as indicated in Figure 2. Later, some of the results were replicated using monthly data starting after the Second World War. Annual data for Australia and Turkey are also analysed.

(c) Using the linear models, basically it was found that there was little or very weak causation from unemployment to inflation. However, there is rather clearer evidence for causation of inflation to unemployment, using linear models.

(d) As many alternative forecasting models are being considered, evaluation in terms of the success of combinations of forecasts becomes particularly relevant. A recent useful discussion is provided by Timmermann (2006).

V. Outline of Results

In this section we use the abbreviations U_t and I_t for unemployment and inflation, and U_{t-1} for unemployment lagged one period.

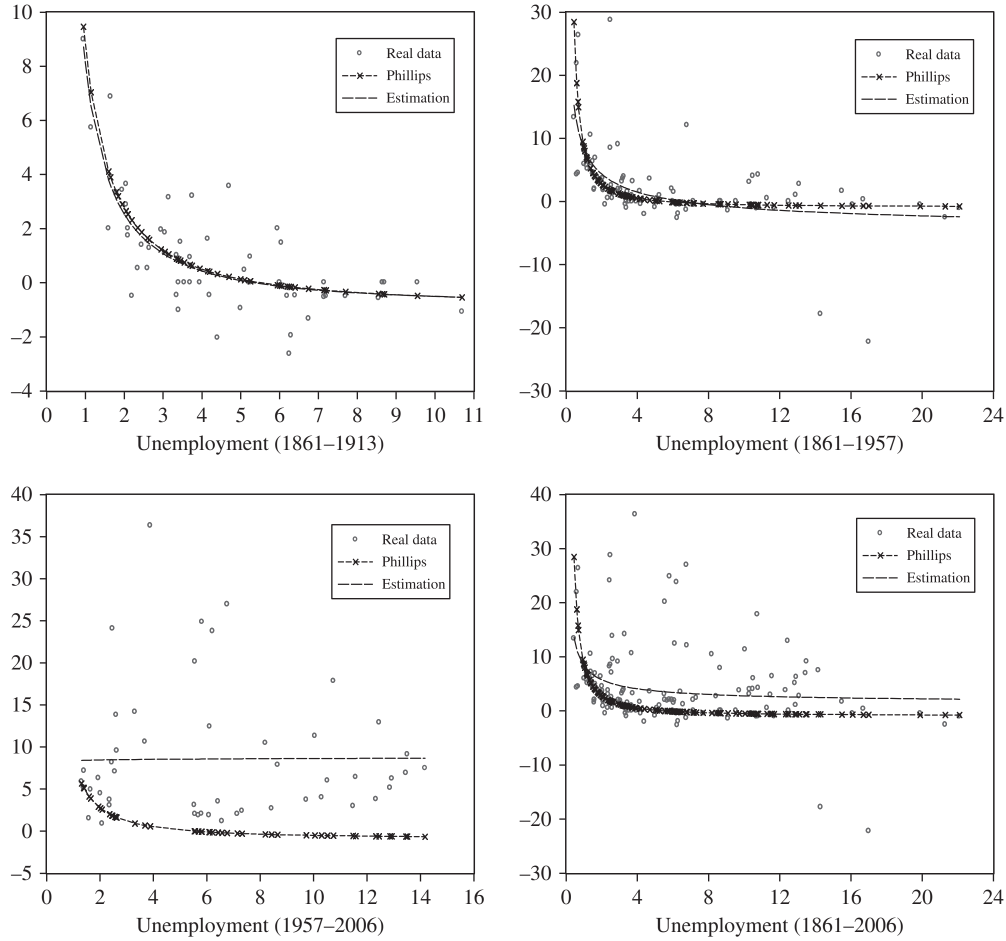

FIGURE 1. Phillips curve fitting—nonlinear relationship using UK annual data. (Phillips: Phillips' model specification with his coefficients. Estimation: Phillips curve with his model specification but estimating coefficients.)

UK Annual data—re-analysis of Phillips' data

The original UK data are annual and go from 1861 to 1957. Data before 1914 were used to form a model, choosing the 'best' in sample among the variables considered. Thus a 'univariate model' would consider a few lags of I_t as explanatory variables, whereas a 'bivariate' model considers lags of both I_t and U_t. If there is causality from unemployment to inflation, then the second equation should forecast better.

(a) We plotted the original Phillips curve on later data (years 1861–2006) using his estimated coefficients but obtained visibly poor fits, as shown in Figure 1.

(b) Figure 1 also used the same specification but with re-estimated coefficients and superimposed on data, again producing a visibly poor fit. It can be concluded that the specific form of the model proposed by Phillips is no longer as successful as he found on early British data.

(c) It was next asked if U_t can be seen to cause I_t within a set of standard linear and TVP models. The results are summarized in Tables 2 and 3.[4] Variables are considered in

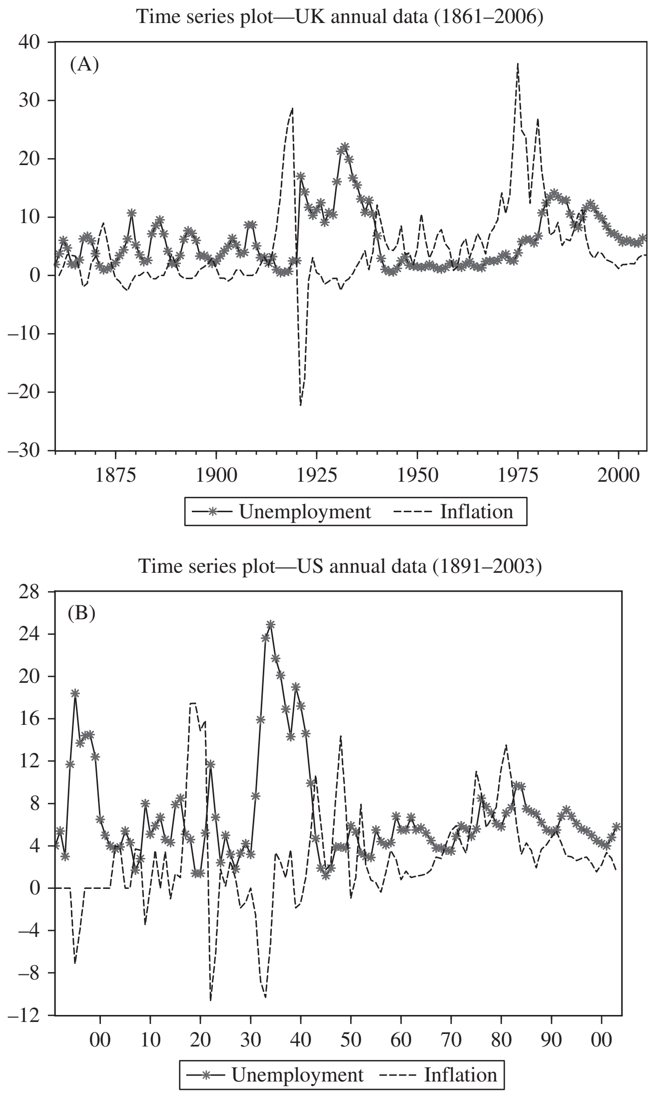

FIGURE 2. Nonlinear relationship in time series dimension.

undifferenced form, where denoted 'level', or differenced. It is seen that in all periods the addition of the unemployment variable does not improve the forecasts, in terms of a reduced mean squared forecast error (MSFE), except for a very slight change in the last period considered. Thus there is no evidence of unemployment causing inflation in these linear models.

(d) Comparing MSFE values between the linear and TVP models, it is seen that the 'nonlinear proxies' are always better, as shown in Table 3. Thus Phillips' use of a nonlinear form is justified in other periods but not his actual formulation.

(e) Within the group of TVP models there is clear superiority for the bivariate models using differences in unemployment, so there is evidence for unemployment causing inflation within this class of models.

TABLE 2

MEAN SQUARED FORECASTING ERROR (MSFE)—LINEAR AR(2) MODELS USING THE UK ANNUAL DATA

Forecasting series	Out of sample period	Independent variables	MSFE under expanded window	MSFE under rolling window	Out of sample observations
Inflation	1914–2006	Univariate	28.41	27.51	93
		Bivariate–Levels	43.10	46.26	
		Bivariate–Difference	68.66	67.92	
	1914–1957	Univariate	31.60	32.28	44
		Bivariate–Levels	64.36	66.27	
		Bivariate–Difference	117.78	118.31	
	1958–2006	Univariate	25.55	23.22	49
		Bivariate–Levels	24.01	28.28	
		Bivariate–Difference	24.56	22.68	
Unemployment	1914–2006	Univariate	5.30	5.32	93
		Bivariate–Levels	8.90	14.84	
		Bivariate–Difference	3.78	4.35	
	1914–1957	Univariate	10.10	10.23	44
		Bivariate–Levels	17.83	17.57	
		Bivariate–Difference	6.12	6.24	
	1958–2006	Univariate	1.00	0.92	49
		Bivariate–Levels	0.88	12.38	
		Bivariate–Difference	1.68	2.65	

Notes

Each forecasting series is first-differenced. The 'expanded window' means that the equations are iteratively estimated. That is, we estimate the equation for 1861–1913 and then forecast for 1914. Then we estimate it on 1861–1914 with forecast for 1915. We proceed like this until estimating 1861–2005 with forecasting for 2006. The 'rolling window' means that we initially estimate the equation with 1861–1913, then with 1862–1914, and then with 1863–1915. That is, we keep the data length of Phillips' original estimation periods.

We can conclude from UK data the original Phillips proposition that there is a nonlinear causal relationship from unemployment to inflation is correct as lagged unemployment is included in the estimated model. However, for the sub-period ending in 1957 the reverse is found.

There is no point in asking if one model is 'significantly worse' than another because this is not a question that now arises in practice. Rather, what is important is what weight does the method get in the best combination? Some exploratory results are shown in Table 4,

TABLE 3
MSFE UNDER TIME-VARYING MODELS USING THE UK ANNUAL DATA

				Forecasting inflation			Forecasting unemployment		
				Out of sample period			Out of sample period		
		Forecasting series	Independent variable	1914–2006	1914–1957	1958–2006	1914–2006	1914–1957	1958–2006
TVP	Bivariate	Level	Level	22.7	26.5	19.4	3.6	5.9	1.6
		Difference	Level	24.1	29.7	19.1	3.7	6.3	1.5
		Level	Difference	22.4	25.8	19.4	3.7	6	1.7
		Difference	Difference	23.4	28.3	19.1	3.9	6.1	1.9
	Univariate	Level		23.3	25.6	21.2	4.5	8.3	1
		Difference		23.8	27.6	20.3	4.5	8.5	1
Linear (expand window)	Bivariate	Level	Difference	144.5	236.8	61.5	35.6	37.8	33.6
		Difference	Difference	68.7	117.8	24.6	3.8	6.1	1.7
	Univariate	Level		35.5	43.9	27.9	24.5	27.2	22.1
		Difference		28.4	31.6	25.6	5.3	10.1	1
Number of observations				93	44	49	93	44	49

just for the period 1958–2006. The left column is for inflation forecasts, with the top half summarizing the MSFE for individual methods and the lower half showing these values for various combinations. It is seen that the simple equal weighted combination of the TVP and bivariate model forecasts performs somewhat better than all the alternatives.

(f) Does inflation cause unemployment? To further consider the methodology being employed here, we also investigated a 'reverse Phillips curve' by asking if inflation causes unemployment. The results are shown in Tables 2 and 3. It is found that in each of the 'out-of-sample' periods 1914–2006 and 1914–57, bivariate models are the best but could be linear or TVP. However, in the more recent period 1958–2006 the univariate models were best, either TVP or linear. Thus there seems to have been evidence of causality from inflation to unemployment in the earlier periods but not the recent one, and there is no evidence for nonlinearity (or TVP) being required. However, using a combination of TVP and linear does seem to do slightly better. It is worse than for the linear models, but the combination of all TVP and linear models is superior to all the alternatives. The combination of all the models is slightly better, which is just the well-known 'portfolio effect' showing up.

Annual US data

Figure 3 exhibits the nonlinear relationship that Phillips discussed, while Table 5 shows the corresponding results for the US annual data over the period 1914–2003. On the left, concerning inflation, there is little, clear and consistent evidence of causality from any of linear or TVP models. The TVP models forecast better than the linear in most cases.

TABLE 4
MSFE OF FORECASTING COMBINATIONS USING UK ANNUAL DATA

Model			Forecasting series	Independent variable	Forecasting inflation 1958–2006	Forecasting unemployment 1958–2006
Individual models	TVP	Bivariate	Level	Level	19.4	1.6
			Difference	Level	19.1	1.5
			Level	Difference	19.4	1.7
			Difference	Difference	19.1	1.9
		Univariate	Level		21.2	1
			Difference		20.3	1
	Linear	Univariate	Level		64.6	29.8
		Bivariate	Level	Difference	68.5	83.1
Forecasting combination	Combination equal weights		TVP–bivariate		18.5	1.5
			TVP–univariate		19.8	0.9
			Linear		57.2	50.8
			All in TVP + linear		22.2	6.7
	Combination regression weights		TVP–bivariate		25.7	2.8
			TVP–univariate		22.8	1.1
			Linear		33.7	4.1
			All in TVP + linear		30.6	3.9
Number of observations					49	

For unemployment, TVP models are usually superior to the linear models, and the bivariate forms generally have lower MSFE values than the univariate, suggesting that there is clearer causality from inflation to unemployment in all periods.

Monthly US data (results are not shown)

A few experiments were performed with monthly US data. For two longer periods (January 1971 to December 2007, and January 1971 to December 1989) and for inflation the TVP bivariate model was usually the best, suggesting again that there could be a nonlinear causality of inflation by unemployment in these periods. However, for the later period January 1990 to December 2007, a univariate (no causal) TVP model was slightly better than the best causal TVP model, which was better than all alternatives.

Annual Australian data (results are not shown)

Annual data were considered from Australia for the years 1956–2006, to expand the group of economies considered. The results for forecasting inflation showed that when using a linear model the bivariate forms did not perform better than the univariate model, but the TVP models were always superior to the linear models and the bivariate forms were always the best, suggesting that inflation can be explained from

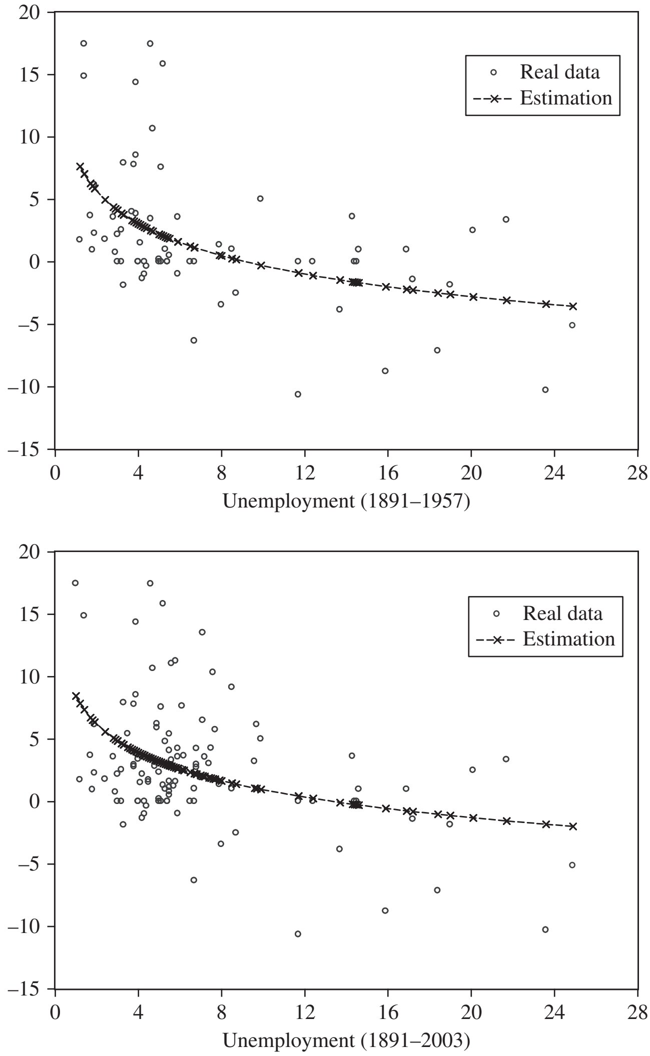

FIGURE 3. Phillips curve fitting using US annual data. (Estimation: Phillips curve with his model specification but estimating coefficients.)

unemployment with a nonlinear model. To forecast unemployment, the best model was linear and involved just past unemployment.

Annual Turkish data (results are not shown)

Our final example uses annual data for the period 1956–2006. Turkey provides an interesting example as it is a fairly advanced economy with an interesting inflation history, being around 10% in the 1970s but up in the 40% range from 1970 to 2000, with occasional peaks up to 80%. However, the inflation rate has declined to 6% or so in recent times. During this period unemployment rate was mostly around 7–8% but with the occasional peak around 10%. When trying to forecast inflation, the TVP causal model appears to be clearly superior. The results suggest that unemployment can be very

TABLE 5
MSFE USING US ANNUAL DATA

				Forecasting inflation			Forecasting unemployment		
				Out of sample period			Out of sample period		
		Inflation	Unemployment	1914–2003	1914–1957	1958–2003	1914–2003	1914–1957	1958–2003
TVP	Bivariate	Level	Level	13.47	24.47	2.94	4.04	7.43	0.80
		Difference	Level	14.12	25.77	2.97	3.98	7.26	0.84
		Level	Difference	13.44	24.42	2.94	4.36	8.02	0.85
		Difference	Difference	14.31	26.32	2.82	4.41	8.14	0.84
	Univariate	Level		14.22	26.03	2.93	4.17	7.63	0.86
		Difference		15.34	28.47	2.78	4.57	8.33	0.97
Linear (rolling window)	Univariate	Difference		19.73	37.56	2.69	5.91	10.90	1.14
	Bivariate	Difference	Level	111.33	212.88	14.19	12.12	20.69	3.93
	Bivariate	Difference	Difference	26.97	50.23	4.73	8.97	16.49	1.78
Number of observations				90	44	46	90	44	46

well forecast using a univariate TVP model that is just lagged unemployment with no need for lagged inflation.

Overall summary of results

In all cases it is found that the TVP formulation is superior to the linear, supporting Phillips' use of nonlinear model forms but not the particular type used by him and by Lipsey.

For the TVP models it is usual to find 'causation from unemployment to inflation', so that models that used unemployment generally forecast inflation better than those that did not. However, the strength of this result declined in the more recent periods.

Concerning the 'reverse Phillips curve' there was no evidence found for inflation causing unemployment in the Australian or Turkish data. But, there was evidence of inflation causing unemployment in early periods for both the USA and Britain but not in recent periods. This does not seem to be a very reliable relationship.

VI. Conclusions

We have discovered many features of the Phillips curve surviving in four economies and through many changes in the economies. It seems that the basic relationship considered by Phillips continues in a nonlinear form but with changing coefficients that we have captured using time varying parameters. The causation is basically one way, from unemployment to inflation and not in the reverse direction.

It would be natural to consider other explanatory series such as import prices. It would have been fascinating to know what Bill Phillips would have produced using the data that is now available, modern computers and more modern techniques.

ACKNOWLEDGMENT

We are grateful to Robert Leeson for providing the data set used in Phillips (1958).

NOTES

1. (By Clive W. J. Granger.) I clearly remember those days of painful computing, as I completed my PhD at Nottingham in that period. On a personal note, I did meet Bill Phillips, but did not know him well. I met him a couple of times at the LSE and acted as an external examiner for him for a masters exam. We got on fine, although he seemed a little stern. Unfortunately he assumed that I knew a lot more about control theory than I actually did, which made our conversation rather difficult.
2. From a comment on page 290 of the paper it seems that Phillips actually fitted model A using least squares, except that the constant a is chosen by trial and error.
3. Chapter 28 of Leeson (2000).
4. Values are shown for three periods: 1914–57, 1914–2006 and 1958–2006.

REFERENCES

Granger, C. W. J. (1988). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control*, **2**, 329–52.

——— (2007). *Nonlinear models—where do we go next: time varying parameter models?* University of California, San Diego, monograph.

——— and Jeon, Y. (2004). Thick modeling. *Economic Modelling*, **21**, 323–43.

Leeson, R. (ed.) (2000). *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press.

LIPSEY, R. G. (1960). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862–1957: a further analysis. *Economica*, **27**, 1–31.

PAGAN, A. (2000). The optimal control articles. In R. Leeson (ed.), *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press, pp. 130–3.

PHILLIPS, A. W. (1954). Stabilisation in a closed economy. *Economic Journal*, **64**, 290–323.

——— (1956). Some notes on the estimation of time forms of reactions in inter dynamic systems. *Economica*, **23**, 99–113.

——— (1957). Stabilisation policy and time forms of lagged responses. *Economic Journal*, **67**, 265–77.

——— (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, **25**, 283–99.

——— (1959). The estimation of parameters in systems of stochastic differential equations. *Biometrika*, **46**, 67–76.

TERASVIRTA, T. (2006). Forecasting economic variables with nonlinear models. In G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*. Amsterdam: North Holland, pp. 413–57.

TIMMERMANN, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*. Amsterdam: North Holland, pp. 135–96.

Economica

Economica (2011) **78**, 67–88
doi:10.1111/j.1468-0335.2009.00807.x

Stabilization Theory and Policy: 50 Years after the Phillips Curve

By Stephen J. Turnovsky

University of Washington, Seattle

Final version received 15 December 2008.

This paper discusses the impact of A. W. H. Phillips' seminal contributions to stabilization policy on subsequent developments in that field. We review the policy rules introduced by Phillips and show how these relate to the optimal rules emerging from linear–quadratic optimization problems. The consequences of rational expectations for the design of optimal stabilization policy are discussed. These challenges arose from the role of 'forward-looking' variables, through issues such as the 'Lucas Critique' and the 'time consistency' of policy. The paper also discusses some of the contemporary aspects of stabilization policy, thereby highlighting the durability of Phillips' contributions.

Introduction

A key objective of macroeconomic policy is to maintain economic stability. Interest in the topic originated with Tinbergen (1952). Employing a static linear framework, he proved the now classic proposition stating that under certainty the policy-maker need use only as many policy instruments as there are independent target variables in order to achieve any desired values for these target variables.[1] Extra policy variables are redundant, while with insufficient instruments not all objectives can be achieved simultaneously. In practice, with the economy being inherently dynamic, it is clearly important to cast the theory of stabilization in a dynamic context, thereby enabling us to consider the stability of the economy as it evolves over time as well as in response to unforeseen stochastic disturbances that may occur at any point in time.[2]

Few would dispute the proposition that Bill Phillips was a pioneer in the development of dynamic stabilization policy. His contributions were manifest in two seemingly different, but as it turns out highly interrelated, areas. The first, and more direct, contribution was contained in a pair of papers published in the *Economic Journal* in the 1950s (Phillips 1954, 1957). These papers draw on his background as an engineer and are the first papers to apply feedback control methods to the stabilization of a macroeconomy. Today that is a burgeoning field, and despite challenges stemming from the subsequent development of rational expectations, the application of control methods is now an integral part of the analysis of dynamic economic systems.

Like many fundamental contributions, Phillips' initial contribution was simple. Previously, Samuelson (1939) and Hicks (1950) had shown how, if one combines the multiplier in consumption with the accelerator in investment, one can derive a dynamic equation determining the evolution of national income (output). The precise nature of this relationship depends on how the lags generating the dynamics are introduced and many ways exist to accomplish this. The dynamics can be expressed in discrete time, as employed by Samuelson and Hicks, or in continuous time, as used by Phillips himself. Phillips took these simple aggregate models and showed how, if one introduces a government that instead of remaining passive follows some active policy intervention rule, then it will be able to influence the dynamic time path of the economy.

The second contribution relates to the celebrated Phillips curve (1958). While this was originally proposed as an empirical relationship between (wage) inflation and unemployment and has spawned generations of empirical research in this area, its introduction into the macroeconomic system, and precisely how this is done, turns out to have profound consequences for the efficacy of stabilization policy.

Our objective in this paper is to review and evaluate the impact of Phillips' seminal work of half a century ago on subsequent developments in macroeconomic stabilization policy. We begin by first briefly reviewing the formulation of linear stabilization rules adopted by Phillips and show how these relate to the optimal stabilization rules that emerge from conventional linear–quadratic optimization problems. These originated in the engineering literature, but turned out to be most convenient for the formal analysis of optimal stabilization policy. Most of the early stabilization literature assumed fixed prices, or in any event was associated with 'backward-looking' or 'sluggish' variables. However, the development of rational expectations in the 1960s and 1970s posed a challenge for stabilization policy, and this is discussed in Section III. This arose from the role of 'forward-looking' inflationary expectations in the Phillips curve, and the effect this had on the design of optimal stabilization rules, through issues such as the 'Lucas Critique' and the 'time consistency' of policy.

Section IV briefly comments on a longstanding debate, the merits of fixed policy rules versus discretionary or optimal policy. Section V discusses some of the more contemporary aspects of stabilization policy, serving to illustrate the durability of Phillips' contributions. The 'expectations-augmented Phillips curve' and 'new-classical Phillips curves' of the 1970s are now replaced by the 'new-Keynesian Phillips curve'. The method of optimal linear–quadratic stabilization theory of the 1970s is now applied as an approximation to more general utility functions. In addition, the linear feedback control rules initially proposed by Phillips have now been introduced into multi-agent dynamic games, while issues of learning are receiving increasing attention.

I. Phillips' Policy Rules

A simple textbook macrodynamic model

The Phillips analysis was based on the dynamic multiplier-accelerator model, previously developed by Samuelson (1939) and Hicks (1950). There are numerous versions of this model and we shall introduce the simplest formulation employed by Phillips. He expressed it using continuous time, which is more convenient for the purpose of characterizing the implied dynamic behaviour, but essentially the same conclusions can be reached using discrete time, as did Samuelson and Hicks.[3]

Aggregate demand of the economy at time t, $Z(t)$, is defined by

$$(1) \qquad Z(t) = C(t) + I(t) + G(t),$$

where $C(t)$ denotes consumption, $I(t)$ denotes investment, and $G(t)$ denotes government expenditure. Dynamics can be introduced in various ways. Whereas Samuelson and Hicks did so by introducing lags into consumption and investment behaviour, Phillips assumed instead gradual product market clearance. This is specified by

$$(2) \qquad \dot{Y}(t) = \alpha[Z(t) - Y(t)], \quad \alpha > 0$$

where $Y(t)$ denotes aggregate supply at time t, and the dot denotes time derivative. If aggregate demand exceeds output, supply is increased at a rate proportional to excess demand, and vice versa.

To complete the model, behavioural hypotheses must be introduced for consumption and investment. The simplest of these is to specify that consumption is proportional to current output,

$$(3) \qquad C(t) = cY(t), \quad 0 < c < 1,$$

and to assume a constant rate of investment, $I(t) = I$. If, further, we assume that government spending is constant as well, then combining these equations we see that equilibrium output evolves in accordance with the simple equation, specifying the textbook dynamic multiplier model

$$(4) \qquad \dot{Y}(t) = \alpha[(c-1)Y(t) + I + G].$$

Phillips' contribution was to introduce various policy rules for $G(t)$. Much of this was developed and can be discussed using this simple model. Most of the subsequent literature, as well as much of Phillips' own contributions, endogenized investment by employing some form of the accelerator theory. The effect of this is to increase the order of the equilibrium dynamics, thereby generating a richer array of time paths for output and other relevant variables. But for present purposes, the simpler model suffices. With I and G fixed, and provided that $0 < c < 1$, the time path of Y described by (4) is stable and output will converge monotonically to the stationary equilibrium level

$$(5) \qquad \bar{Y} = \frac{I+G}{1-c}.$$

This will be immediately recognized as being the equilibrium level of income in the simplest static linear macroeconomic model.

Policy rules

Within this framework, Phillips introduced government expenditure as an active policy variable that is continuously adjusted to meet certain specified objectives. In doing so he emphasized the lags associated with adjusting the policy instrument itself. These are often referred to as *policy lags*, and reflect delays in implementing decisions due to, for example, the political process and appropriation of the required resources. They are quite distinct from lags from the underlying economic structure, such as those embodied in the market disequilibrium relationship (2), which may be appropriately characterized as being *system lags*.[4]

Phillips assumed that the policy actually implemented at any point in time adjusts only gradually to past policy decisions. Thus, if $G^d(t)$ is the policy choice made at time t, the actual value of the policy variable at that time, $G(t)$, is adjusted in accordance with[5]

$$(6) \qquad \dot{G}(t) = \beta(G^d(t) - G(t)), \quad \beta > 0.$$

The desired value of the policy variable, $G^d(t)$, is related by some rule to the ultimate target objective that Phillips took to be the stabilization of national income. He proposed three such policy rules, which he called (i) proportional policy, (ii) integral policy and (iii) derivative policy, all of which influenced the dynamics of the economy in different ways, having both desirable and undesirable effects on its evolution. These terms did not originate with Phillips. Rather, they were part of the tradition of classical control, where engineers referred to them as 'PID feedback rules'. We shall briefly discuss each in turn.

Proportional policy This was specified by Phillips to be

(7a) $G^d(t) = -\gamma_p(Y - Y^*), \quad \gamma_p > 0,$

where Y^* is the (desired) target level of output. The parameter γ_p represents the intensity of the policy-maker's desired policy response when output deviates from its target. According to (7a), if $Y(t) < Y^*$, then $G^d(t) > 0$, and if $Y(t) > Y^*$, then $G^d(t) < 0$. Since the rule may require $G^d(t) < 0$, Phillips interprets it as 'net fiscal stimulus' rather than pure government spending, which by its nature is non-negative. Thus (7a) asserts that the desired net fiscal stimulus is proportional, but opposite to, the deviation between current and desired level of output. Combining equations (4), (6) and (7a), the dynamic evolution of the economy is described by the following pair of equations:

(8) $$\begin{pmatrix} \dot{Y} \\ \dot{G} \end{pmatrix} = \begin{pmatrix} \alpha(c-1) & \alpha \\ -\beta\gamma_p & -\beta \end{pmatrix} \begin{pmatrix} Y \\ G \end{pmatrix} + \begin{pmatrix} \alpha I \\ \beta\gamma_p Y^* \end{pmatrix}.$$

Three observations about this system are worth noting. First, the necessary and sufficient conditions for stability are (i) $1 - c + \gamma_p > 0$, (ii) $1 - c + \beta/\alpha > 0$, so that it is clear that γ_p, β, which characterize the stabilization policy and implementation, will influence the dynamics. While (i) and (ii) will certainly be met if $0 < c < 1$, it may be possible to stabilize the system even in the implausible event where $c > 1$. Second, the eigenvalues to (8) will be complex if and only if $4\alpha\beta\gamma_p > [\beta - \alpha(1-c)]^2$, implying that policy lags may induce cycles into the adjustment. This is presumably undesirable, but hardly surprising, since with policies taking time to implement, by the time $G^d(t)$ is yielding its desired effect, the conditions leading to that decision may have changed, causing the system to over-adjust during the transition.[6]

Third, output in (8) converges to the stationary level

(9) $$\bar{Y} = \frac{1 + \gamma_p Y^*}{1 - c + \gamma_p} \neq Y^*.$$

That is, in general, the level of output will fail to converge to its desired target level. This was viewed by Phillips as being an undesirable feature of the proportional policy rule, but in fact it can be regarded as reflecting an inadequate specification of the rule, as given in (7a). This formulation ignores the fact that given the behaviour of the private sector as reflected by C, I, the government must also choose an appropriate target level of expenditure, G^*, if it wishes to attain Y^* in the long run. This appropriate level is determined by the stationary relationship

$$G^* = (1 - c)Y^* - I.$$

Once this fact is recognized, it becomes natural to express (7a) in deviation form

(7a′) $G^d(t) - G^* = -\gamma_p(Y - Y^*),$

in which case the stability conditions remain unchanged, and if satisfied, ensure that output converges to its target, Y^*.

Integral policy As an alternative policy, Phillips proposed that $G^d(t)$ is determined by the integral (sum) of past deviations in output from its target, rather than only the current deviation. This is specified by

(7b) $$G^d(t) = -\gamma_i \int_0^t [Y(s) - Y^*]\mathrm{d}s, \quad \gamma_i > 0.$$

Differentiating with respect to t enables the policy to be written in the equivalent form

$$(7\text{b}') \qquad \dot{G}^d(t) = -\gamma_i[Y(t) - Y^*].$$

Expressed in this way, the rule asserts that the policy variable should be increased if output is above its target, and decreased otherwise. It is the form of policy adjustment rule specified by Mundell (1962) and others in their analysis of 'the assignment problem', relating the appropriate adjustment of policy instruments to targets.

Combining, equations (4), (6) and (7b′) leads to a system of three dynamic equations in $Y(t)$, $G^d(t)$ and $G(t)$:

$$(10) \qquad \begin{pmatrix} \dot{Y} \\ \dot{G}^d \\ \dot{G} \end{pmatrix} = \begin{pmatrix} \alpha(c-1) & 0 & \alpha \\ -\gamma_i & 0 & 0 \\ 0 & \beta & -\beta \end{pmatrix} \begin{pmatrix} Y \\ G^d \\ G \end{pmatrix} + \begin{pmatrix} \alpha I \\ \gamma_i Y^* \\ 0 \end{pmatrix}.$$

This yields several differences from the proportional rule. Assuming $0<c<1$,

$$[\beta + \alpha(1-c)](1-c) > \gamma_i$$

is necessary and sufficient for (10) to be stable. This indicates a trade-off between the intensity of the stabilization policy and lags in policy for stability to prevail. If the policy lags are sufficiently long (β small), overly intensive policy may generate instability. Indeed, this was one of the concerns originally expressed by Friedman (1948). In the absence of policy lags ($\beta \to \infty$), the integral policy will always ensure stability, although it may be associated with cyclical adjustment if the adjustment is too intensive ($4\gamma > \alpha(1-c)^2$). In any event, if stable, Y will converge to Y^*, thereby avoiding one of the undesirable features associated with Phillips' specification of the proportionate rule.

Derivative policy The third policy rule introduced by Phillips, the derivative policy, is of the form

$$(7\text{c}) \qquad G^d(t) = -\gamma_d \dot{Y}(t), \quad \gamma_d > 0.$$

That is, fiscal stimulus depends on the current rate of change of output, behaving like a 'negative accelerator'. For an appropriately chosen γ_d this can stabilize an otherwise unstable system, although it will not succeed in driving income to its target level.

Phillips also proposed combining these three policy rules, by postulating, for example,

$$(11) \qquad G^d(t) = -\gamma_p(Y(t) - Y^*) - \gamma_i \int_{-\infty}^{t} [Y(s) - Y^*]\mathrm{d}s - \gamma_d \dot{Y}(t),$$

and showing how by the judicious choice of weights, γ_j, the policy-maker can take advantage of the various desirable features of the individual policies, while reducing their unattractive aspects. For example, the presence of the integral component ensures that income converges to its target, while undesired cyclical adjustments associated with this policy can be reduced by simultaneously using the proportional and derivative policies. In this respect it is intriguing to observe that combining the policies as in (11) is a step in the direction of choosing the optimal stabilization policy. Finally, we again emphasize that Phillips also introduced these policies into more complex models that include an accelerator-determined investment demand, leading to higher-order dynamic systems.

While Phillips developed these rules in the context of fiscal stabilization policy, early applications of stabilization theory also applied them to monetary stabilization issues (see, for example, Lovell and Prescott 1968; Sargent 1971). They also formed the basis for simulation studies involving both monetary and fiscal policies in larger macro models (see, for example, Cooper and Fischer 1974). Most contemporary research on stabilization policy has focused on monetary policy, with fiscal policy being directed more toward longer-run issues such as economic growth and capital accumulation.[7]

II. Linear–Quadratic Optimal Stabilization

The policy rules introduced by Phillips, while plausible, are not in general optimal, although, as we shall now see, they appear as important components of optimal policies.

General approach

Beginning in the 1960s, interest developed in the question of optimal stabilization policy. The framework employed to address this issue was the linear–quadratic system, an adaptation of the 'state-regulator problem' developed by control engineers (see, for example, Kalman 1960; Athans and Falb 1966; Bryson and Ho 1969). In general, this can be outlined as follows.

Consider an economy summarized by n state (target) variables, x, and m control (policy) variables, u. Assume that the structure of the economy can be expressed by the linear vector system

$$\text{(12a)} \quad \dot{x}(t) = A(t)x(t) + B(t)u(t),$$

where $A(t)$ is an $n \times n$ matrix and $B(t)$ is an $n \times m$ matrix.[8] Assume further that the objective is to minimize the quadratic cost function

$$\text{(12b)} \quad J \equiv \frac{1}{2}x'(T)Fx(T) + \frac{1}{2}\int_0^T [x'(t)M(t)x(t) + u'(t)N(t)u(t)]\mathrm{d}t,$$

where F and $M(t)$ are positive semi-definite matrices, $N(t)$ is positive definite, T is the planning horizon, and primes denote the vector transpose operator.

In economic terms, the policy-maker wishes to keep a set of target variables, and the corresponding values of the policy variables, as close as possible to their desired target values, with failure to achieve these objectives being penalized by quadratic costs. The optimal (cost-minimizing) value of the control vector, $\hat{u}(t)$, is a linear feedback rule of the form

$$\text{(13a)} \quad \hat{u}(t) = -N^{-1}(t)B'(t)P(t)x(t),$$

where $P(t)$ is the unique positive semi-definite solution to the Riccati equation

$$\text{(13b)} \quad \dot{P}(t) = -M(t) - A'(t)P(t) - P(t)A(t) + P(t)B(t)N^{-1}(t)B'(t)P(t),$$

and satisfies the boundary condition

$$\text{(13c)} \quad P(T) = F.$$

The critical thing to note about this solution is that the optimal policy is a time-varying linear feedback control law, in which, in general, all of the control variables are linear functions of all of the current state variables.

Several observations about this form of solution can be noted. First, by simple redefinition of variables it can be easily adapted to incorporate exponential time discounting, as economic applications typically employ. Second, while the quadratic function has the convenience of yielding linear optimal control laws, it implies that positive and negative deviations from targets are weighted equally, which may or may not be appropriate. Third, if the time horizon is infinite and the matrices A, B, M, N defining the optimal stabilization problem are constant over time, then the optimal policy summarized by (13) simplifies to the stationary rule

$$\hat{u}(t) = -N^{-1}B'Px(t), \tag{14a}$$

where P is the unique positive semi-definite solution to the matrix equation

$$M + A'P + PA - PBN^{-1}B'P = 0. \tag{14b}$$

By substitution, it then follows from (13a) that when policy is set optimally, the economy evolves in accordance with

$$\dot{x}(t) = [A - BN^{-1}B'P]x(t). \tag{15}$$

The policy rules (13a) (or (14a)) can be characterized as being a kind of 'generalized proportional' policy of the type proposed by Phillips, in the sense that the current policy variables are related proportionately to the current state of the economy relative to its long-run target. The elements of the feedback rule, as described by (13a) or (14a), are given by specific values, which may be required to follow specific time paths, depending on whether or not the economic structure is constant or time-varying. By construction, (15) will ensure that the economy converges to its desired target value; problems of instability, which we saw could be associated with the inappropriate setting of the integral policy rule, do not arise.

One final point is that much of the interest in stabilization policy relates to stochastic systems. Among the earliest treatments was Howrey (1967), who extended an earlier discrete-time multiplier-accelerator model formulated by Baumol (1961) to allow for additive stochastic disturbances. But what if the parameters themselves in the basic structural equations such as (12a) are stochastic, giving rise to multiplicative stochastic shocks? Important work by Wonham (1968, 1969) showed that the optimal policy rules of the form summarized by (13) and (14) above extend to this case, where the feedback rules are shown to depend on the variance–covariance matrix of the underlying stochastic parameters.[9] Early work by Turnovsky (1973, 1977) then applied these results to issues in macroeconomic stabilization policy, of the type pioneered by Phillips.

Recently, Kendrick (2005) has provided an excellent overview of the applications of stochastic control methods to the class of linear–quadratic economic models outlined in this section. With the introduction of stochastic elements, the availability of information at the time a policy is to be implemented becomes important. In particular, it becomes necessary to distinguish between open-loop control, when the entire time path for policy is solved at the outset, and feedback rules, when current policy is updated as new information becomes available. In the case of deterministic systems that are fully known, the two solutions coincide. There is no gain from feedback control, although the open-loop solution can be expressed in feedback form as in (13), and as Phillips himself did. An important aspect of this distinction involves learning, an issue to which we return later.

Some special cases

To give a sense of how these early applications of optimal stabilization policy relate to Phillips' contributions, we consider several special cases.

First, suppose that the economy is purely one-dimensional, such as in equation (4) above, being expressed by

$$\text{(16a)} \quad \dot{y}(t) = \alpha[-sy(t) + g(t)],$$

where $y(t)$ and $g(t)$ denote output and government expenditure, both measured about their respective target values, and $s = 1 - c$ is the marginal propensity to save. Investment is constant, equal to its desired value, and there are no policy lags. The policy-maker's objective is to choose the fiscal instrument, $g(t)$, so as to minimize the quadratic cost function

$$\text{(16b)} \quad J \equiv \frac{1}{2}\int_0^\infty [my(t)^2 + ng(t)^2]\mathrm{d}t$$

subject to the evolution (16a). The optimal policy for this problem is the linear feedback rule

$$\text{(17a)} \quad g(t) = -\frac{\alpha}{n}py(t),$$

where p is the positive solution to the quadratic equation

$$\text{(17b)} \quad \frac{\alpha^2}{n}p^2 + 2\alpha sp - m = 0.$$

The optimal policy summarized by (17) is a purely proportional one, as proposed by Phillips, though of the modified (7a′) form. In terms of his notation, $\gamma_p > 0$, assuming the specific values implied by (17a) and (17b), which in turn depend on the underlying structural parameters.

Turnovsky (1973) extended this to the case where s or α was stochastic. In the former case, for example, (17b) is modified to

$$\text{(17b)} \quad \frac{\alpha^2}{n}p^2 + (2\alpha\bar{s} + \alpha^2\sigma_s^2)p - m = 0,$$

where $\bar{s}, \sigma_s^2$ denote the mean and variance of the stochastic marginal propensity to save. More stochastic variation in the savings propensity implies more intensive fiscal intervention.

More generally, suppose now that the economy is described by the following multiplier-accelerator model, with all variables expressed in deviation form about their steady-state values:

$$\text{(18a)} \quad \dot{y}(t) = \alpha[z(t) - y(t)],$$

$$\text{(18b)} \quad z(t) = cy(t) + i(t) + g(t),$$

$$\text{(18c)} \quad \frac{\mathrm{d}i(t)}{\mathrm{d}t} = \zeta[v\dot{y}(t) - i(t)],$$

where $z(t)$ is aggregate demand and now investment, $i(t)$, is expressed as a lagged accelerator, where the desired stock of capital is proportional to output.[10] The system is now driven by two state variables, $y(t)$, $i(t)$, and so optimal fiscal policy will be of the generic form

$$g(t) = \theta_1 y(t) + \theta_2 i(t), \tag{19}$$

where the feedback components θ_1, θ_2 are computed from (14a) and (14b) (see Turnovsky 1973). Solving (18c) for $i(t)$ and substituting, we can express (19) in the form

$$g(t) = (\theta_1 + \theta_2 \zeta v) y(t) - \theta_2 \zeta^2 v \int_{-\infty}^{t} y(t) \mathrm{e}^{-\zeta(t-s)} \mathrm{d}s. \tag{19$'$}$$

This can be seen to be the sum of Phillips' proportional policy and a form of integral policy, where past outputs (or their deviations) have exponentially declining weights. Furthermore, if the dynamics can be represented by a second-order differential equation (as can easily be done in some variants of the Samuelson–Hicks model of the business cycle), the optimal policy can be written as the sum of a purely proportional component plus a derivative component (see Turnovsky 1977).

As a final example, we go beyond Phillips' early work and introduce the Phillips curve, augmented by 'backward-looking' inflationary expectations.[11] Consider the simple monetary model

$$\dot{y}(t) = \alpha[z(t) - y(t)], \tag{20a}$$

$$z(t) = d_1 y(t) - d_2[r(t) - \pi(t)], \tag{20b}$$

$$p(t) = \kappa y(t) + \pi(t), \quad \kappa > 0, \tag{20c}$$

$$\dot{\pi}(t) = \rho[p(t) - \pi(t)], \quad \rho > 0. \tag{20d}$$

For this modified structure, (20b) specifies aggregate demand to depend positively on output and negatively on the real interest rate, $r(t) - \pi(t)$. Equation (20c) is an expectations-augmented Phillips curve, where the current rate of inflation, $p(t)$, increases with output and expected inflation, where the coefficient on expected inflation, $\pi(t)$, is unity.[12] Inflationary expectations evolve in accordance with the backward-looking adaptive expectations scheme, (20d). By substitution, this economy can be reduced to the pair of dynamic equations

$$\dot{y}(t) = \alpha((d_1 - 1)y(t) - d_2[r(t) - \pi(t)]), \tag{21a}$$

$$\dot{\pi}(t) = \rho\kappa y(t). \tag{21b}$$

Assume that the policy-maker sets the nominal interest rate, $r(t)$, to minimize quadratic costs associated with deviations of output and inflation from their respective target values. The optimal monetary policy will be of the generalized proportional form

$$r(t) = \varphi_1 y(t) + \varphi_2 \pi(t), \tag{22}$$

which is essentially a form of the widely-employed Taylor (1993) rule.[13]

Other examples can be found, but we have surely made the point that the form of policy rules proposed by Phillips (1954) played a central role in the early applications of optimal control theory to stabilization policy.

III. THE CHALLENGE OF RATIONAL EXPECTATIONS

The dynamic system considered by Phillips, as well as the early applications of dynamic control theory that we have been discussing—including the last example of the expectations-augmented Phillips curve—are of the classical type, in that all variables are assumed to evolve continuously from some given initial condition. In the jargon of contemporary macrodynamics, all variables are 'backward-looking' or 'sluggish'. This reflects the fact that economists were using the traditional techniques of differential equations as developed by applied mathematicians and control engineers, which of course was consistent with Phillips' own academic background.

However, the development of rational expectations and its applications to macrodynamics in the 1970s introduced the notion that some economic variables, most notably financial variables, are 'forward-looking', incorporating agents' expectations of the future. It is clearly more realistic to permit these variables to respond instantaneously to new information as it impinges on the economy, instead of forcing them to evolve gradually from the past. This was first illustrated in a simple monetary model by Sargent and Wallace (1973). They showed how, given the inherent instability of the underlying differential equation driving the dynamics in this model, plausible economic behaviour requires that the forward-looking variable (in their case the price level) jump so as to ensure that the economy follows a bounded (stable) adjustment path. Most economic dynamic systems consist of a combination of sluggish variables, such as physical capital, which by their nature can be accumulated only gradually, and forward-looking jump variables, such as exchange rates or financial variables, that are not so constrained. As a consequence, the standard dynamic macroeconomic system embodying rational expectations has a combination of stable and unstable dynamics, with the case of a unique convergent saddlepath arising when the number of unstable roots equals the number of jump variables (see Blanchard and Kahn 1980).

This represents a fundamentally different approach to macroeconomic dynamics from that in the earlier literature, and the introduction of rational expectations has had a profound effect on the application of control methods to stabilization policy. Several issues arise, and we shall discuss these in turn.

Computation of optimal policy rules under rational expectations

The first issue is the task of solving for rational expectations equilibrium, even in the absence of any active stabilization policy. While the equilibrium economic structure may be conceptually straightforward, its solution is likely to be computationally challenging, depending on the dating of the forward-looking expectations variables, their forecast horizons, and the dimensionality of the system. Solutions procedures have been proposed by several authors, including Blanchard and Kahn (1980), Fair and Taylor (1983), and more recently Sims (2001).

Currie and Levine (1985) provide a lucid description of the computation of optimal feedback rules for a continuous-time formulation containing both sluggish and forward-looking variables. They show how, in a system embodying rational expectations, one can partition the dynamic variables into predetermined (sluggish) variables and non-predetermined (forward-looking) variables, while taking account of the saddlepoint structure associated with the rational expectations equilibrium. For the usual quadratic loss function, the resulting optimal policy rule can be expressed in several alternative but equivalent forms. One form is as a linear function of both the predetermined and non-predetermined variables. Alternatively, it can be expressed as a linear function of the predetermined state variables and the predetermined co-state variables (those associated with the non-predetermined variables). However, since the latter can be expressed as an integral of the vector of the underlying state variables, the optimal policy rule can be expressed as a generalized linear feedback rule on the state variables combined with an integral feedback rule on the state variables. To this extent the form of the Phillips' policies still prevails.

The Lucas Critique

As we have been discussing, the objective of stabilization policy is to influence the time paths of a set of target variables, such as output, inflation, etc. Being forward-looking, a key feature of rational expectations, in contrast to the traditional adaptive expectations scheme such as (20d), is that it incorporates the agent's information regarding the structure of the economy. In particular, rational expectations include the agent's perception of policy as part of the economic environment. Lucas (1976) made the profound observation that in these circumstances it is not rational for policy-makers to conduct policy under the assumption that the coefficients describing the evolution of the economy remain fixed and invariant with respect to its chosen policy. In the dynamic system (8), for example, the behavioural parameters α, c will in general vary with the chosen policy parameter γ_p. This dependence needs to be taken into account in determining the effects of policy rules, as well as for the determination of optimal stabilization policy.[14]

The Lucas Critique is a general proposition having far-ranging implications for analysing economic policy. Its main message is that if we want to predict the effects of policy changes, we must model the 'underlying parameters' such as technology and preferences that govern individual behaviour. To the extent that modern macroeconomics is based on intertemporal optimization of utility subject to production constraints, the macroeconomic equilibrium so derived is immune from the Lucas Critique in that it is conditional on government policy. We can then model policy-making as a game, whereby the government, acting as leader, makes its stabilization (policy) decisions taking into account the reactions of the private sector. This approach is at the core of the voluminous optimal tax literature, for example.

However, solving for optimal policy in this way may be difficult, particularly over time, and, furthermore, it may be unrealistic to assume that the policy-maker knows precisely the private sector's response to its decisions. Amman and Kendrick (2003) propose an approximation based on the use of the Kalman filter. The idea is that the policy-maker does not need to be able to predict exactly how private agents will respond to its policies. Rather, it can simply use the Kalman filter and update parameter estimates each period. While this means that the policy-maker will always be one period behind in his perception of the private sector's behavioural responses, they argue that this may be good enough for most applications of macroeconomic policy. Monte Carlo runs that they run provide some support for this procedure.

Policy neutrality

One area where the Lucas Critique is particularly potent is in the role of the Phillips curve in determining the trade-offs between inflation and unemployment. In this context the Lucas Critique implies that the nature of the trade-off depends on government intervention policy. The issue of policy neutrality is an extreme form of this, and asserts that, because of the Lucas Critique, the trade-off breaks down completely.

In an influential article, Sargent and Wallace (1976) provided an example to show that under rational expectations only unanticipated policy changes can have real effects, so that any feedback policy rule, such as the Phillips rules we have been discussing, will have no effect on output. In our example, the time path of output would become independent of the policy parameters such as γ_p, so that there is no trade-off between output and inflation. It turns out that this policy neutrality proposition, as it is known, which is potentially devastating to the use of control theory as a tool of stabilization policy, is sensitive to model specification, and in particular to the timing of expectations. This can be illustrated by comparing two simple examples.

First, consider an economy represented by the pair of stochastic difference equations

$$\text{(23a)} \quad y_t = -e[r_t - (E_{t-1}(P_{t+1}) - E_{t-1}(P_t))] + u_t,$$

$$\text{(23b)} \quad P_t - P_{t-1} = \theta y_t + E_{t-1}(P_t) - P_{t-1} + v_t,$$

where y_t denotes output (in deviation form) in logarithms, P_t denotes the price level in logarithms, $E_{t-1}(.)$ denotes expectations formed at time $t-1$ and assumed to be rational, r_t is the nominal interest rate, and u_t, v_t are white noise random disturbances in demand and supply, respectively. In keeping with the contemporary literature, we employ discrete time. Equation (23a) is a standard IS curve, relating output negatively to the real interest rate, where the expected rate of inflation over the period $(t, t+1)$ is based on information at time $t-1$. The second equation is the new-classical Phillips curve.[15] Assuming that the monetary authority treats the nominal interest rate r_t as the policy variable, the equilibrium value of output y_t can be shown to be

$$\text{(24)} \quad y_t = -e[r_t - E_{t-1}(r_t)] + u_t.$$

The point about (24) is that (the deviation of) output depends only on the unanticipated component of monetary policy. Any feedback policy rule based on past observed data that the monetary authority follows is fully incorporated into private agents' expectations and thus is fully negated in terms of its effects on current output.

Things change dramatically, however, if we now modify (23a) to

$$\text{(23a}'\text{)} \quad y_t = -e[r_t - (E_t(P_{t+1}) - P_t)] + u_t.$$

The only difference is that we modify expected inflation for period $(t, t+1)$ to be conditional on information at time t, when the actual price level is observed. The rational expectations solution for output under this seemingly modest reformulation is

$$\text{(24}'\text{)} \quad y_t = -\frac{e}{(1+e\theta)}[r_t - E_{t-1}(r_t)] - \frac{e}{(1+e\theta)}\sum_{i=1}^{\infty}[E_t(r_{t+i}) - E_{t-1}(r_{t+i})] + \frac{(u_t - ev_t)}{(1+e\theta)}.$$

In addition to the current unanticipated component of monetary policy, given by the first term of (24′), current output now depends on the sum of all revisions to future monetary policy between time $t-1$ and t, which takes account of new information acquired at time t. By impacting the forecast of inflation, a feedback policy rule will now exert an impact on current output.

As an example, suppose that the monetary authority adopts the interest rate rule

$$r_t = \lambda\left(\frac{u_{t-1} - ev_{t-1}}{1+e\theta}\right).$$

This is a feedback rule, whereby the monetary authority adjusts the interest rate in response to the previous period's stochastic shocks, which are known at time t. With u_t, v_t being white noise, taking expected values over relevant periods (24′) reduces to

$$y_t = \left(1 - \frac{\lambda e}{1 + \theta e}\right)\frac{u_t - e v_t}{1 + e\theta},$$

which clearly is influenced by the policy rule.[16] Indeed, setting $\lambda = \theta + 1/e$ succeeds in stabilizing output completely. The presence of policy neutrality thus depends critically on the available information, an issue that is taken up at greater length by Canzoneri *et al.* (1983).

Time consistency of optimal policy

A fourth issue to arise when the system contains forward-looking jump variables is that of 'time consistency'. This concept was first introduced by Strotz (1955–56). It describes a situation where an agent's preferences change over time, so that what is preferred at one instant in time is no longer preferred at some other later point in time. This issue has ramifications for various aspects of economics and in particular for stabilization policy. In this context the issue is whether or not a future policy decision that forms part of an optimal plan formulated at some initial date remains optimal when considered at some later date, even though no relevant information has changed in the meantime. If it does not, the optimal plan is said to be time-inconsistent.

This problem was emphasized by Kydland and Prescott (1977), who argued that the problem of time inconsistency has grave implications for the application of control theory methods to problems of economic stabilization. In the abstract to their paper they write: 'We conclude that there is no way control theory can be made applicable to economic planning when expectations are rational.' In the conclusions they argue: 'active stabilization may very well be dangerous and it is best that it not be attempted. Reliance on policies such as a constant growth in the money supply and constant tax rates constitute a safer course of action.' These are strong statements and many people in the community of control theorists view this assessment of the role of control theory to stabilization policy as overly pessimistic.

But the question of time consistency (or inconsistency) is important, and attempts to resolve it have generated a lot of research. The pursuit of time-inconsistent policies will eventually cause the government to lose credibility, and issues such as commitment and reputational equilibria have been analysed by a number of authors (see, for example, Barro and Gordon 1983).

One simple solution, very much within the spirit of the linear–quadratic framework, is the following. As noted, the attainment of a rational expectations equilibrium involves an initial jump in the forward-looking variable. These initial jumps presumably impose real dislocational costs on the economy, and these should be taken into account in the design of the optimal policy system. Stemp and Turnovsky (1987) show how, if these initial costs are large enough, it may cease to be optimal for the policy-maker to re-optimize along a transitional path.

To see this, assume that the policy-maker's objective function is to

$$\text{minimize} \int_0^\infty [ay^2 + (1-a)p^2]\mathrm{e}^{-\beta t}\mathrm{d}t + k|P(0) - P_0|^q.$$

This cost function now has two components. The first is the standard quadratic loss function, asserting that the policy-maker's target is to achieve a zero inflation rate ($p = 0$) at a full employment level of output ($y = 0$). One objective is to minimize the discounted intertemporal deviations from these targets, with $0 \leqslant a \leqslant 1$ reflecting the relative importance assigned to these two objectives.

In a world of rational expectations, a change in monetary policy will cause an initial unanticipated discontinuous jump in the price level, $P(0)$, from its previously inherited level, P_0. Given sluggishness in the economy, this causes jumps in real magnitudes, which impose structural adjustment costs (e.g. labour reallocation) on the economy, and these should be taken into account in assessing the overall benefits of the optimal stabilization policy. These initial adjustment costs are not reflected in the conventional term, but are incorporated in the second term. Stemp and Turnovsky (1987) show that the time consistency, or otherwise, depends critically on q, being time-consistent if $q \leqslant 1$ and time-inconsistent otherwise.

IV. Rules Versus Optimal Policy

Despite the fact that the generic form of the optimal policy rule is the generalized proportional policy as set out in (13), from a practical point of view the policy may turn out to be extremely complicated to compute, especially for a large system, and even more so if it includes forward-looking variables. This leads to the question of the gains from applying optimal control over using some simple, but reasonable, policy such as the three rules proposed by Phillips, or the Taylor rule, or perhaps even doing nothing at all.

This is an old question, predating Phillips, going back to Friedman (1948) and early discussions of policy rules versus discretionary policy. At that time Friedman advanced the proposition that due to the length and variability of lags in the effects of monetary policy, the monetary authority should abandon discretionary monetary management and simply should allow the money supply to grow at a fixed rate. Indeed, our discussion of the Phillips rules provides some support for this view. We have seen that the presence of policy lags can introduce unwanted cycles in the economy, and even instability that otherwise would not exist. But the debate of rules versus discretion raises several issues, most important of which relate to the information that the policy-maker possesses. Here we briefly note some of them.

First, suppose that the world is deterministic and the policy-maker knows the true structure. Then by its nature the optimal policy dominates and so the question is whether the gains in economic stability are sufficient to justify the effort (and cost) involved in computing the optimal rule over something much simpler and mechanistic. Some years ago Feldstein and Stock (1994) addressed this question in an analysis where the objective is to target nominal income. They reached the conclusion that there is little difference between a very simple adaptive rule and an optimal policy. If this kind of proposition is robust, then simple policy rules of the type originally proposed by Phillips will continue to play an important role in the stabilization of dynamic economic systems.

Second, what if there is uncertainty? This introduces different levels of complications. The optimal policy model introduced in Section II assumes that everything is known except for the specific values taken on by the stochastic parameters. As we have seen, this will influence the optimal setting of the associated policy instrument, just as it did in Brainard's (1967) early analysis. But one of the important results obtained by Wonham (1969) is that feedback control in a system with stochastic parameters, whereby the effects of policy become stochastic, is feasible if and only if the noise is not too great. In this case, it is possible for the policy instrument used for stabilization to introduce too much randomness into the system, implying that the economy will actually be more stable if the policy-maker does not intervene.

Third, and most fundamentally, all optimal policies that one derives are specific to an assumed economic structure, rather than the true economic environment that policy-makers do not and cannot know. What are the merits of employing the optimal policy to the wrong model, rather than adopting some arbitrary alternative rule? This issue is addressed in detail by Brock *et al.* (2007). Their approach is to construct a model space that includes all candidate models for the economy, evaluate the policies for each of the candidate models, and then determine how to draw policy inferences given the fact that the true model is unknown. In contrast to the usual robustness analysis that measures misspecification relative to some baseline model, they acknowledge the global nature of model uncertainty. They focus on the sensitivity of the rules to model uncertainty, rather than on the derivation of optimal rules in the presence of model uncertainty. The other issue that they address concerns the trade-offs of policy on variances of different frequencies; policies that may reduce the variance of high-frequency fluctuations may come at the expense of enhanced longer-term fluctuations.

V. Recent Developments

In this section we briefly note some of the more recent aspects of stabilization policy that pertain to Phillips' contribution.

New-Keynesian Phillips curve

The new-Keynesian Phillips curve is based on a model of optimal pricing in imperfect competition and a theory of price stickiness (see, for example, Roberts 1995; Woodford 2003). It is of the generic form

$$P_t - P_{t-1} = \theta y_t + \beta(E_t(P_{t+1}) - P_t), \quad 0<\beta<1, \tag{25}$$

and differs from the new-classical Phillips curve in that the expected inflation to which the current inflation is reacting extends for the next period $(t, t+1)$, rather than for the previous period $(t-1, t)$. This has important consequences for stabilization policy. To see this, we shall combine (25) with (23a), for which the new-classical Phillips curve yields policy neutrality.

The form of the rational expectations solution depends on the magnitude of $\beta + \theta e$. We consider the case $\beta + \theta e < 1$, when the unique stable solution for y_t is

$$y_t = -er_t - e^2\theta \sum_{j=1}^{\infty} (\beta + \theta e)^{j-1} E_{t-1}(r_{t+j}) + u_t. \tag{26}$$

It is clear that interest rate rules based on past information will influence current output. For example, if $r_t = \mu u_{t-1}$, then the solution to (26) is

$$y_t = -e\mu u_{t-1} + u_t,$$

which is clearly dependent on the policy parameter μ. The case $\beta + \theta e > 1$ is associated with non-uniqueness issues of the type identified in rational expectations models by

Taylor (1977), but depending on how the non-uniqueness is resolved, policy rules will have real effects.

Multi-agent stabilization

Thus far we have focused on a single decision-maker, acting in isolation. In reality, many economic situations are characterized by multiple decision-makers operating in an interactive environment. Decisions made by one agent influence the other, and vice versa, giving rise to strategic behaviour that we can analyse as a dynamic game. As is well known, crucial factors determining the equilibrium outcome include: (i) the availability of information at the time decisions are made; (ii) the sequencing of the decisions by agents; and (iii) whether they behave non-cooperatively to maximize their own welfare, or cooperatively to maximize their joint wellbeing.

Insofar as stabilization policy is concerned, there are two main areas where strategic interaction is particularly important. The first is the interaction between monetary and fiscal policy, allowing for the fact that the central bank may have different objectives from the treasury. This raises issues relating to credibility of policy, reputation and political aspects that are somewhat removed from the approach to stabilization that we are discussing here.

The second application is in the area of international economic policy coordination, and in particular, monetary and exchange rate policy. Miller and Salmon (1985) and Oudiz and Sachs (1985) have analysed two-country dynamic games of monetary policy that are direct generalizations of the class of optimal policy model summarized in Section II. To give a flavour of this, suppose the policy-maker in country 1 wishes to solve the following dynamic optimization problem:

$$(27a) \qquad \min \int_0^\infty w'(t)Q_1 w(t)\mathrm{d}t, \quad \text{where} \quad w'(t) \equiv [x(t)\; u_1(t)\; u_2(t)]',$$

where, subject to the dynamic evolution

$$(27b) \qquad \dot{x}(t) = Ax(t) + B_1u_1(t) + B_2u_2(t)$$

for given $x(0)$. The policy-maker in country 2 solves an analogous problem.

As in previous examples, the objective is to minimize a quadratic loss function, which depends on country 1's controls, $u_1(t)$, which of course this policy-maker sets, country 2's controls, $u_2(t)$, which he may react to, and a common set of target variables, $x(t)$. The latter may be more relevant to one country than to the other and may also include non-predetermined variables as well as pure state variables. Equation (27b) describes the evolution of the state variables, which depend in part on the choices each policy-maker makes.

For this setup Miller and Salmon discuss open-loop and feedback Nash and Stackelberg solutions. For feedback Nash, for example, each policy-maker sets his controls in accordance with the feedback rule

$$(28) \qquad u_1(t) = R_1x(t), \; u_2(t) = R_2x(t),$$

taking the other's actions as given when making his decision. The components of the feedback are determined by a generalized Riccati type equation that involves the structural parameters of both economies. It also depends on the specific rule defining the dynamic game. But the point we wish to make is that equations (28) are generalized proportional policy rules directly analogous to (14a) discussed earlier, and hence the relationship to Phillips' early work extends to this type of analysis.[17]

Utility maximization

The optimal stabilization rules that we have derived have been chosen so as to minimize quadratic costs involving the deviations of the state variables and the control (policy) variables about some stationary level. Many variants of this criterion can be found, differing in such aspects as to whether the deviations in output are measured relative to the full employment level, the frictionless level of output, etc.

Apart from the limitation noted earlier that the quadratic function is weighting positive and negative deviations equally, it suffers from the more serious criticism that it may or may not be an appropriate representation of welfare, which presumably is the issue of ultimate concern. Indeed, for almost three decades now, the 'representative agent model' has been the standard macroeconomics paradigm, although it too has been the source of criticisms.[18] With macroeconomic equilibrium being derived through utility maximization, this framework is much more oriented towards analysing welfare issues and therefore addressing issues pertaining to optimal policy-making.

Recently, several authors have sought to examine the relationship between utility maximization and the conventional stabilization criteria that we have been adopting; see in particular Woodford (2003) where this is discussed in great detail. He establishes conditions under which the quadratic loss function, so widely employed in stabilization policy, can be viewed as a second-order approximation to the expected value of a more general utility function. Here we informally sketch the relationship in a simple example.

Suppose that welfare is represented by a utility function of the form $U(c,g)$, where c denotes consumption and g denotes government expenditure (the control variable). Suppose further that through stabilization, c and g are restricted to stochastic fluctuations about their respective mean levels $(\tilde{c},\tilde{g})$. Employing a second-order approximation to $U(c,g)$ about $(\tilde{c},\tilde{g})$, and taking expected values, we may write

$$EU(c,g) \cong U(\tilde{c},\tilde{g}) + \tfrac{1}{2}U_{cc}E(c-\tilde{c})^2 + U_{cg}E(c-\tilde{c})(g-\tilde{g}) + \tfrac{1}{2}U_{gg}E(g-\tilde{g})^2. \tag{29}$$

Assume that output is produced by the production function $y=f(k)$, where k denotes capital stock. If the agent maximizes intertemporal utility, it is well known that equilibrium consumption along an evolving stable adjustment path is of the form $c=c(k,g)$, which may be linearly approximated by

$$c-\tilde{c} \cong c_k(k-\tilde{k}) + c_g(g-\tilde{g}).$$

Substituting this linear approximation into (29) yields a second-order approximation to expected utility of the form

$$EU(c,g) \cong U(\tilde{c},\tilde{g}) + X,$$

where X is a quadratic loss term involving the state variable k and the control variable g. For the simple production function, the state variable can be immediately transformed to y, as in the stabilization literature. In order for the quadratic loss function to give the correct welfare rankings of different stabilization policies, it must be the case that $U(\tilde{c},\tilde{g})$ is independent of policy, or at least is only weakly sensitive to it. One case where it is independent is if the utility function is of the form $U(c+g)$. With capital stock constant in steady state, product market equilibrium implies $\tilde{c}+\tilde{g}=f(\tilde{k})$, where the steady-state stock of capital is fixed and determined by the marginal product condition $f'(\tilde{k}) = \rho$, the rate of time discount.

Learning

Throughout this discussion we have assumed that the policy-maker has complete knowledge of the true underlying economic structure. In the case of deterministic systems, all parameters are known, as is their evolution if they are time-varying. In the case of stochastic systems, all characteristics of relevant probability distributions are also known; only the specific random outcomes are unknown until they occur. In reality, of course, policy-makers do not know the true system. Even if they know the broad structure of the economy, such as the general qualitative relationship among the variables, they will at best have only some estimate of the relevant parameters, and worse still, they are unlikely to even know the general structure of the economy, as Brock *et al.* (2007) have emphasized. At best, policy-makers and agents in general may learn about the structure of the economy as it evolves over time.

The qualitative information about the economic structure becomes particularly important in dynamic models involving rational expectations, the key characteristic of which is that they incorporate agents' perceived structure of the economy. As a result of this, their beliefs about the economy will influence its actual evolution. The fact that applications of rational expectations assume complete knowledge of the economy's structure (apart from pure stochastic shocks) has been a source of its criticism. While this is a reasonable objection, we view the traditional rational expectations specification as a useful benchmark, with its underlying characteristic of forward-looking behaviour providing significant insights into macroeconomic dynamics in general, and stabilization policy in particular.

To incorporate learning is challenging and raises many issues. By its nature, learning is a gradual process that takes place over time. The interaction of the dynamics of this process with that of the system itself is important, and not all learning processes need be stable. The most comprehensive general study of learning as an element of macroeconomic dynamics is Evans and Honkapohja (2001). They emphasize the method of expectational stability. The key element of this concept is that it involves a mapping from the perceived law of motion (dynamic structure), which in general is incorrect, to the corresponding actual law of motion, which incorporates this incorrect information. If the system is expectationally stable, the learning process for the unknown parameters will converge to the true values, and the agent will ultimately learn the true economic process. It is possible, however, for the learning process to diverge, and cause the dynamics of the overall system to diverge as well.

Several issues in this process arise and should be mentioned. First, the time period involved in updating information on parameters need not coincide with the time interval that characterizes the system dynamics. Second, it is possible for updating of information to involve nonlinear relationships, leading to a multiplicity of solutions, and for learning not to converge to any of them (see Blanchard and Fischer 1989). Third, information and learning are almost certainly not uniform across the economy; different agents have different degrees of information and varying capacities to learn. Evans and Honkapohja focus primarily on learning by private agents, but the same issues apply to policy-makers engaged in optimal policy-making. Fourth, learning may take different forms, the two most common being least squares learning and Bayesian learning.

The learning procedures that we have been outlining can be characterized as being 'passive', in the sense that the agent learns about the relevant parameters over time as the system evolves and information is updated. Kendrick (2005) contrasts this with 'active' learning, sometimes referred to as 'dual control'. In the stabilization policies that we have been considering in previous sections, the policy instrument is used for a single purpose, namely to help move the economy toward its target. In contrast, in dual control the policy variables are used for two purposes. In addition to the usual stabilization objective, the second purpose is to perturb the system so as to enhance learning of the relevant parameters and thereby improve control performance at later stages. This form of learning was introduced originally by Kendrick (1982) and later by Amman and Kendrick (1994), using techniques previously developed in the control literature by Tse and Bar-Shalom (1973).

VI. Conclusions

It is evident that Bill Phillips, through his initial contributions to dynamic stabilization policy in conjunction with the Phillips curve, has had a profound impact on the theory of economic policy. First, the policy rules that he proposed frequently lie in the class of optimal policies and thus serve as useful benchmarks, thereby assisting in the interpretation of more complex optimal policy rules. Indeed, the relationship of the Phillips policy rules to the optimal rules applies not only to traditional optimal policy-making based on sluggish backward-looking systems, but also to systems involving forward-looking expectations, as well as multi-agent strategic policy-making problems.

The Phillips curve has been a remarkably resilient concept and has remained a key component of the output–inflation trade-offs that may characterize stabilization policy. Beginning with the original negative inflation–unemployment relationship, through the (backward-looking) expectations-augmented Phillips curve of the 1960s, to the (forward-looking) new-classical Phillips curve of the 1970s, and most recently the new-Keynesian Phillips curves of the 1990s, it has been a central component of short-run macrodynamic models for 50 years.

Indeed, the implementation of rational expectations presented a serious challenge to the use of control theory as an instrument of macroeconomic stabilization policy. But it is fair to say that macroeconomists have accepted the challenge and that the methods of control theory are being applied more successfully than ever to dynamic macro models involving rational expectations. The economics profession owes a great debt to Bill Phillips for introducing these analytical tools over half a century ago.

ACKNOWLEDGMENTS

This is a revised version of a plenary address presented at ESAM08 Conference 'Markets and Models: Policy Frontiers in the A. W. H. Phillips Tradition', 9–11 July 2008, Wellington, New Zealand. I am grateful to David Kendrick for discussions on these issues over the years, as well as his for his comments on this paper. Support was provided in part by the Castor Endowment at the University of Washington.

NOTES

1. Brainard (1967) re-examined the Tinbergen proposition in a simple stochastic model and showed how it ceases to apply once multiplicative stochastic disturbances are introduced. Henderson and Turnovsky (1972) showed how adjustment costs associated with policy instruments also lead to its breakdown.
2. Preston (1974) referred to Tinbergen's theorem as one of 'static controllability', and using results from control engineering developed an analogous condition for the controllability of a linear dynamic system. Preston and Pagan (1982) provide an excellent treatment of earlier developments in both the static and the dynamic theory of economic policy.

3. The contemporary literature on stabilization policy almost always employs discrete time (see, for example, Woodford 2003). Discrete time is in fact much more convenient for capturing some of the recent theoretical developments, which sometimes depend on subtle issues of timing. For example, the difference between the 'new-classical' and 'new-Keynesian' Phillips curve is one of timing, a difference that can be best captured using discrete time. In our exposition we shall introduce time in whichever way is more convenient.
4. These two kinds of lags are also sometimes referred to as being 'inside lags' and 'outside lags', respectively.
5. Solving equation (6), the actual policy at time t is $G(t) = \beta \int_{-\infty}^{t} G^d(s)e^{-\beta(t-s)}ds$, which is an exponentially declining weighted average of past policy decisions. As $\beta \to \infty$, the desired policy is fully implemented immediately.
6. The policy parameters β, γ_p also affect the *speed* of convergence. While this was not an aspect that concerned Phillips, speeds of convergence have assumed an important role in contemporary macrodynamics, particularly in the dynamics of growth.
7. For example, the 'endogenous growth' literature pioneered by Romer (1986), and its extensions, emphasize the impact of tax rates and the role of public capital on growth. There is much less focus on monetary policy.
8. A system of an arbitrary order can always be reduced to a first-order system by redefining the higher-order derivatives as new state variables, so in this respect (12a) is a general representation of a linear system.
9. There are, however, constraints on the variance–covariance matrix of the underlying stochastic parameters that in effect assert that control is possible only if the stochastic components are not too large. For a discussion of this stochastic stabilizability condition in the context of the conventional aggregate macroeconomic model, see Turnovsky (1973).
10. Equation (18c) may be derived as follows. Suppose that the desired stock of capital is proportional to output, $k^d(t) = vy(t)$, and that the actual capital stock adjusts gradually to its desired value in accordance with $\dot{k}(t) = \zeta[k^d(t) - k(t)]$. Combining these two equations with the relationship $\dot{k}(t) = i(t)$ yields (18c).
11. The so-called 'expectations-augmented Phillips curve' was associated with the seminal contributions of Phelps (1968) and Friedman (1968). While Phelps developed a formal technical derivation of this relationship, Friedman provided an informal version in his 1968 American Economic Association presidential address. This formulation is also closely related to the 'Lucas supply function' that we shall discuss in Section III below.
12. Much of the early empirical work on the expectations-augmented Phillips curve was concerned with whether or not this coefficient is unity, an issue that has bearing on the existence or otherwise of a long-run unemployment–inflation trade-off; see Turnovsky (1977) for a discussion of this issue. Despite this early debate, (20c) is a consensus canonical specification of the expectations-augmented Phillips curve.
13. Taylor rules are feedback rules that tie the current interest rate to deviations in expected inflation and output, about their desired target levels. Taylor proposed the specific coefficients of 0.5 on the output variable and 1.5 on the inflation deviation. Turnovsky (1981) has analysed in detail the optimal trade-offs between unemployment and inflation in an expanded version of this model.
14. As a related observation, the Lucas Critique calls into question the practice of econometrically estimating the parameters of a reduced form equation such as (8). This is because as the policy varies, so do the reduced form parameters, and thus the assumption that they remain fixed over a sample period is inappropriate. Note that the Lucas Critique does not apply to dynamic systems such as the original Phillips models, which are entirely backward-looking.
15. Lucas adopted what has become known as the 'Lucas supply function', which replaces (23b) by $y(t) = \lambda[P_t - E_{t-1}(P)] + v'_t$, so output deviations depend on unanticipated price movements. The same results obtain.
16. We should point out that setting the nominal interest rate, as in this example, leads to an indeterminate price level, an issue that has generated some debate, particularly in the context of the so-called monetary instrument problem; see Poole (1970), Parkin (1978), Turnovsky (1980) and McCallum (1981). This aspect does not invalidate the point that we are making, and can be easily circumvented by introducing real money balances into the aggregate demand function.
17. For an overview of the literature on linear quadratic differential games, see Engwerda (2005).
18. See Colander (2006).

REFERENCES

AMMAN, H. and KENDRICK, D. A. (1994). Active learning: Monte Carlo results. *Journal of Economic Dynamics and Control*, **18**, 119–24.

——— and ——— (2003). Mitigation of the Lucas Critique with stochastic control methods. *Journal of Economic Dynamics and Control*, **27**, 2035–57.

ATHANS, M. and FALB, P. (1966). *Optimal Control*. New York: McGraw-Hill.

BARRO, R. J. and GORDON, D. B. (1983). Rules, discretion, and reputation in a model of monetary policy. *Journal of Monetary Economics*, **12**, 101–21.

BAUMOL, W. (1961). Pitfalls in contracyclical policies: some tools and results. *Review of Economics and Statistics*, **43**, 545–56.

BLANCHARD, O. J. and FISCHER, S. (1989). *Lectures on Macroeconomics*. Cambridge, MA: MIT Press.

——— and KAHN, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica*, **48**, 1305–11.

BRAINARD, W. (1967). Uncertainty and the effectiveness of policy. *American Economic Review, Papers and Proceedings*, **57**, 411–25.

BROCK, W. A., DURLAUF, S. N., NASON, J. M. and RONDINA, G. (2007). Simple versus optimal rules as guides to policy. *Journal of Monetary Economics*, **54**, 1372–96.

BRYSON, A. E. and HO, Y. C. (1969). *Applied Optimal Control*. Waltham, MA: Blaisdell.

CANZONERI, M. B., HENDERSON, D. W. and ROGOFF, K. (1983). The information component of the interest rate and optimal monetary policy. *Quarterly Journal of Economics*, **98**, 545–65.

COLANDER, D. (ed.) (2006). *Post-Walrasian Macroeconomics*. Cambridge: Cambridge University Press.

COOPER, J. P. and FISCHER, S. (1974). Monetary and fiscal policy in the fully stochastic St Louis econometric model. *Journal of Money, Credit, and Banking*, **6**, 1–22.

CURRIE, D. and LEVINE, P. (1985). Macroeconomic policy design in an interdependent world. In W. H. Buiter and R. C. Marston (eds), *International Economic Policy Coordination*. Cambridge: Cambridge University Press.

ENGWERDA, J. C. (2005). *LQ Dynamic Optimization and Differential Games*. Chichester: Wiley.

EVANS, G. W. and HONKAPOHJA, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

FAIR, R. C. and TAYLOR, J. B. (1983). Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. *Econometrica*, **51**, 1169–85.

FELDSTEIN, M. and STOCK, J. H. (1994). The use of monetary aggregate to target nominal GDP. In N. G. Mankiw (ed.), *Studies in Business Cycles*, Vol. 29, Chicago: University of Chicago Press.

FRIEDMAN, M. (1948). A monetary and fiscal framework for economic stability. *American Economic Review*, **38**, 245–64.

——— (1968). The role of monetary policy. *American Economic Review*, **58**, 1–17.

HENDERSON, D. W. and TURNOVSKY, S. J. (1972). Optimal macroeconomic adjustment under conditions of risk. *Journal of Economic Theory*, **4**, 58–71.

HICKS, J. R. (1950). *A Contribution to the Theory of the Trade Cycle*. Oxford: Oxford University Press.

HOWREY, P. (1967). Stabilization policy in linear stochastic models. *Review of Economics and Statistics*, **49**, 404–11.

KALMAN, R. E. (1960). New approach to linear filtering and prediction problems. *Transactions of the ASME Series D—Journal of Basic Engineering*, **82**, 35–45.

KENDRICK, D. A. (1982). Caution and probing in a macroeconomic model. *Journal of Economic Dynamics and Control*, **4**, 149–70.

——— (2005). Stochastic control for economic models: past, present, and the paths ahead. *Journal of Economic Dynamics and Control*, **29**, 3–30.

KYDLAND, F. E. and PRESCOTT, E. C. (1977). Rules rather than discretion: the inconsistency of optimal plans. *Journal of Political Economy*, **85**, 473–91.

LOVELL, M. C. and PRESCOTT, E. C. (1968). Money multiplier accelerator interaction. *Southern Economic Journal*, **35**, 60–72.

LUCAS, R. E. (1976). Econometric policy evaluation: a critique. In K. Brunner and A. H. Meltzer (eds), *The Phillips Curve and Labor Markets*, Carnegie Rochester Conference Series on Public Policy. Amsterdam: North-Holland.

MCCALLUM, B. T. (1981). Price level determinacy with an interest rate policy rule and rational expectations. *Journal of Monetary Economics*, **8**, 319–29.

MILLER, M. H. and SALMON, M. (1985). Policy coordination and dynamic games. In W. H. Buiter and R. C. Marston (eds), *International Economic Policy Coordination*. Cambridge: Cambridge University Press.

MUNDELL, R. A. (1962). The appropriate use of monetary and fiscal policy for internal and external stability. *IMF Staff Papers*, **9**, 70–9.

OUDIZ, G. and SACHS, J. (1985). International policy coordination in dynamic macroeconomic models. In W. H. Buiter and R. C. Marston (eds), *International Economic Policy Coordination*. Cambridge: Cambridge University Press.

PARKIN, J. M. (1978). A comparison of alternative techniques of monetary control under rational expectations. *Manchester School*, **46**, 252–87.

PHELPS, E. S. (1968). Money-wage dynamics and labor-market equilibrium. *Journal of Political Economy*, **76**, 678–711.
PHILLIPS, A. W. (1954). Stabilisation policy in a closed economy. *Economic Journal*, **64**, 290–323.
——— (1957). Stabilisation policy and the time form of lagged responses. *Economic Journal*, **67**, 265–77.
——— (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, **25**, 283–99.
POOLE, W. (1970). Optimal choice of monetary policy instruments in a simple stochastic macro model. *Quarterly Journal of Economics*, **84**, 197–216.
PRESTON, A. J. (1974). A dynamic generalization of Tinbergen's theory of policy. *Review of Economic Studies*, **41**, 65–74.
——— and PAGAN, A. R. (1982). *The Theory of Economic Policy: Statics and Dynamics*. Cambridge: Cambridge University Press.
ROBERTS, J. M. (1995). New Keynesian economics and the Phillips curve. *Journal of Money, Credit, and Banking*, **27**, 975–84.
ROMER, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, **94**, 1002–37.
SAMUELSON, P. A. (1939). Interaction between the multiplier analysis and the principle of acceleration. *Review of Economic Statistics*, **21**, 75–8.
SARGENT, T. J. (1971). The optimum monetary instrument variable in a linear economic model. *Canadian Journal of Economics*, **4**, 50–60.
——— and WALLACE, N. (1973). The stability of models of money and growth with perfect foresight. *Econometrica*, **41**, 1043–8.
——— and ——— (1976). Rational expectations and the theory of economic policy. *Journal of Monetary Economics*, **2**, 169–83.
SIMS, C. A. (2001). Solving linear rational expectations models. *Computational Economics*, **20**, 1–20.
STEMP, P. J. and TURNOVSKY, S. J. (1987). Optimal monetary policy in an open economy. *European Economic Review*, **31**, 1113–35.
STROTZ, R. H. (1955-56). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, **23**, 165–80.
TAYLOR, J. B. (1977). Conditions for unique solutions in stochastic macroeconomic models with rational expectations. *Econometrica*, **45**, 1377–85.
——— (1993). Discretion versus policy rules in practice. *Carnegie Rochester Conference Series on Public Policy*, **39**, 195–214.
TINBERGEN, J. (1952). *On the Theory of Economic Policy*. Amsterdam: North-Holland.
TSE, E. and BAR-SHALOM, Y. (1973). On actively adaptive control for linear systems with random parameters. *IEEE Transactions on Automatic Control*, **AC-18**, 98–108.
TURNOVSKY, S. J. (1973). Optimal stabilization policies for deterministic and stochastic linear economic systems. *Review of Economic Studies*, **40**, 79–96.
——— (1977). *Macroeconomic Analysis and Stabilization Policy*. Cambridge: Cambridge University Press.
——— (1980). The choice of monetary instrument under alternative forms of price expectations. *Manchester School*, **48**, 39–63.
——— (1981). The optimal intertemporal choice of inflation and unemployment: an analysis of the steady state and transitional dynamics. *Journal of Economic Dynamics and Control*, **3**, 357–84.
WONHAM, W. M. (1968). On a matrix Riccati equation of stochastic control. *SIAM Journal of Control*, **6**, 681–97.
——— (1969). Random differential equations in control theory. In A. T. Barucha-Reid (ed.), *Probabilistic Methods in Applied Mathematics II*. New York: Academic Press.
WOODFORD, W. M. (2003). *Interest and Prices*. Princeton, NJ: Princeton University Press.

Economica

Economica (2011) **78**, 89–107
doi:10.1111/j.1468-0335.2009.00777.x

Policy Regime Changes, Judgment and Taylor rules in the Greenspan Era

By CINZIA ALCIDI†, ALESSANDRO FLAMINI‡ and ANDREA FRACASSO††

†*The Graduate Institute, Geneva* ‡*University of Sheffield* ††*University of Trento*

Final version received 11 October 2008.

This paper investigates policy deviations from linear Taylor rules motivated by the risk management approach followed by the Fed during the Greenspan era. We estimate a nonlinear monetary policy rule via a logistic smoothing transition regression model where policy-makers' judgment, proxied by economically meaningful variables, drives the transition across policy regimes. We find that ignoring judgment-induced nonlinearities while estimating Taylor rules has remarkable costs in terms of fit: above 250 bps in 10 quarters. Although linear Taylor rules describe well the broad contours of monetary policy, they fail to detect relevant policy decisions driven by policy-makers' judgment.

INTRODUCTION

Arguably, the main problem that monetary policy has to cope with is uncertainty. Alan Greenspan, former Chairman of the Federal Reserve, maintained that 'general uncertainty', consisting of risk and Knightian uncertainty (Knight 1921), leads to a risk management approach to policy, that is, the management of the 'continuum ranging from well-defined risk to the truly unknown' (Greenspan 2004, p. 37).

We believe that uncertainty-related concerns drive important changes in the way policy-makers make decisions and that such changes amount to policy regime switches. Specifically, if a set of events and/or contingencies significantly modifies, even for only a short period of time, the systematic way policy decisions are made, then a monetary policy regime switch occurs. This interpretation of monetary policy regime is more refined than the usual one, according to which the procedure for setting the policy instrument holds for a relative long period of time. In our view, a monetary regime is characterized not by its time length but instead by the impact of events and/or contingencies on the behaviour of the policy-makers.

Now, in an environment strongly characterized by risk and Knightian uncertainty, to what extent can a linear Taylor rule describe *ex post* the behaviour of the central bank (CB)? To what extent do finer regimes matter without being detected by a linear Taylor rule?

A 'narrative answer' to these questions is provided by Greenspan himself:

> Indeed rules that relate the setting of the federal funds rate to the deviations of output and inflation from their respective targets, in some configurations, do seem capture the broad contours of what we did over the past decade and a half. And the prescriptions of formal rules can, in fact, serve as helpful adjuncts to policy . . . But at crucial points, like those in our recent policy history—the stock market crash of 1987, the crises of 1997–98, and the events that followed September 2001—simple rules will be inadequate as either descriptions and prescriptions of policy. Moreover, such rules suffer much of the same fixed-coefficient difficulties we have with our large-scale models.
>
> (Greenspan 2004, pp. 38–9)

This answer points to important limits for simple linear rules *à la* Taylor. Yet we do not know to what extent, in practice, these limits matter because there is little evidence in the literature on Greenspan's account of the US monetary policy. This paper aims to fill such a gap and provide a quantitative answer to the questions mentioned above by focusing on special circumstances for which narrative and anecdotal evidence is available. From an empirical point of view, the issue is about the difficulty of ascertaining the existence of finer monetary policy regimes given that, in practice, they are unobservable. Yet, since under each regime the policy instrument is linked to the CB's targets and to its determinants by a peculiar and systematic relationship, i.e. a rule, a policy regime switch should correspond to a significant change in such rule. Building on this, when we empirically identify a change in the policy rule occurring for more than one period and corresponding to an event and/or a contingency, we also identify a policy regime switch.

We address this issue by investigating what happens to Taylor-type rules once linearity is not imposed on the specification. In particular, we use a logistic smooth transition regression (LSTR) model, as developed by Terasvirta (1994) and improved by Franses and van Dijk (2000), to detect endogenously deviations from the simple instrument rule and, when possible, find the specific rules that characterize the finer regimes.

The contribution of this paper is to show empirically that CB judgment may map into finer monetary policy regimes where interest rate decisions purposefully deviate from what would be recommended by linear Taylor-type rules. We find that during the Greenspan era these deviations matter and are related to precise Fed's concerns matching the narrative evidence. Even though linear policy rules describe well the broad contours of monetary policy, ignoring judgment-induced nonlinearities would lead to remarkable costs in terms of empirical fit. These costs amount to errors of 260 basis points over the short period 2001Q1 to 2003Q2, 60 of which were observed in 2001Q4 alone.

Our findings also suggest that central bankers' judgment represents an important source of nonlinearity in the policy conduct. This is a novelty with respect to the recent literature on nonlinear monetary policy rules which has instead focused on asymmetric preferences and nonlinear aggregate supply.[1] By identifying proxies of policy concerns and applying a flexible estimation technique, our analysis contributes to understanding 'the nonmechanistic flexibility [standing out in the Greenspan era] that allowed a forward-looking policy to anticipate what, in retrospect, plainly turned out to be different economic circumstances' (Friedman 2006, p. 176). It also helps to assess when and to what extent linear rules 'are suspended when necessary', as argued by Poole (2006, p. 8), because it associates the policy deviations with the central bankers' judgment about special events and contingencies.

Finally, this work contributes to the current debate about the existence of changes in US monetary policy by focusing on the 18 years of Greenspan's tenure, a period which has not been fully explored in previous studies.[2] The latter, in fact, have mainly focused on policy switches related to changes in the chairmanship and have not covered the period after 2000. This period turns out to be active in terms of critical policy decisions which led to deviations from what would be indicated by linear Taylor-type rules.

Our paper is closely related to Rabanal (2004) and Assenmacher-Wesche (2006) and, all in all, our results are consistent with theirs. Our analysis, however, differs from these works along several dimensions. First, it focuses in an exclusive fashion on the Greenspan era and covers it entirely. Second, the nonlinearity in policy does not relate to a certain state of the economy (recession/expansion in Rabanal (2004) or high/low inflation in Assenmacher-Wesche (2006)), but relates to the CB's concerns about less predictable future situations. Third, while we adopt economically meaningful and observable variables to describe the transition across regimes, they resort to probabilities estimated over the whole sample.

The paper is structured as follows. In Section I, we briefly introduce the main features of Taylor-type rules and the results of the linear estimations. Section II is devoted to a description of the LSTR specification and of the rationale behind the choice of this particular nonlinear estimation technique. The tests for nonlinearity close the section. In Section III we present the estimates of the nonlinear Taylor rules. Section IV concludes. As a robustness check, the Appendix briefly discusses the results for an alternative Taylor-type rule, in which current inflation is replaced by inflation forecasts.

I. The Linear Estimation and the Data

The monetary policy rule proposed by Taylor (1993) is $i_t = r^* + \pi^* + \beta_\pi(\pi_t - \pi^*) + \beta_y y_t$, where i_t is the Federal Funds rate, r^* is the equilibrium real Federal Funds rate, π_t is the average inflation rate, π^* is the target inflation rate, and y_t is the output gap. After its first formulation, the Taylor rule has become an important benchmark both for the design of policy rules and for the *ex post* empirical investigations of monetary policy decisions.[3]

In the Taylor rule literature, a common approach is to assume that the CB sets the interest rate as a weighted average of the target rate and the last period(s) rate(s), thereby adjusting the interest rate with certain inertia. Accordingly, we estimate the linear rule in the form

$$(1) \qquad i_t = a + b_\pi \bar{\pi}_t + b_y y_t + \rho_1 i_{t-1} + \rho_2 i_{t-2} + \xi_t,$$

where the choice of adding two lags is dictated by the need to remove serial correlation in the residuals,[4] and the long-run (LR) coefficients β_π and β_y are implicitly defined. Following Taylor (1993), we use the Federal Funds rates as measures of the interest rates, and the average inflation is computed as $\bar{\pi}_t = \sum_{i=0}^{3} \pi_{t-i}/4$, with π_t being the quarter inflation rate based on the GDP chain-type price index. The output gap y_t is defined as the difference between the log of the real GDP level and the log of the real potential GDP for the USA, as estimated by the Congressional Budget Office.[5] All series are quarterly data covering the period from 1987Q3 to 2005Q4 and come from the FRED database.

The estimates of the coefficients of the baseline rule (1), reported in Table 1, are not far from Taylor's predictions. The degree of interest rate smoothing, equal to the sum of the ρ terms, is 0.915 and satisfies the condition for stationarity of the Federal Funds rate series. The residuals from the regression are plotted in Figure 1.

At first sight, the rule seems to capture well the behaviour of the authorities. However, it is noteworthy that consecutively large (negative and positive) residuals mark more than one quarter in 1988–89, 1991–92, 1994 and 2000–02. Negative (positive) residuals correspond to periods in which the estimated rule is conducive to fitted Federal Funds rates higher (lower) than the actual ones. Accordingly, in such periods US monetary policy seems to have been relaxed (tightened) beyond what was suggested by the inflation and the output gap.

As maintained by Yellen (2004) and Rudebusch (2006), such a pattern of the residuals can be explained by the fact that a simple Taylor rule is not able to perfectly catch the actual stance of US monetary authorities in response to unusual economic conditions. Arguably, this could be due to the omission of significant variables and/or to

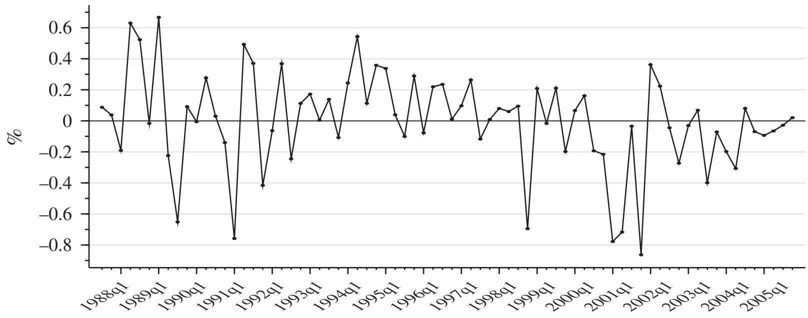

FIGURE 1. Residuals from the baseline Taylor rule (1).

the imposition of a linear specification of the model. We investigate both these possibilities, respectively in this and in the next section.

Following economic intuition and Castelnuovo (2003) and Gerlach-Kristen (2004), we augment the baseline specification (1) by adding a common measure of credit risk, specifically the spread between the Moody's BAA corporate bond index and the 10-year US Treasury note yield.[6] Accordingly, we estimate the linear augmented Taylor rule

$$(2) \qquad i_t = a + b_\pi \overline{\pi}_t + b_y y_t + \omega_z z_t + \rho_1 i_{t-1} + \rho_2 i_{t-2} + \xi_t,$$

where z_t is the BAA spread.

Augmenting the Taylor rule by means of z_t reduces the degree of interest rate smoothing from 0.915 to 0.879, in line with Rudebusch's (2002) hypothesis that at least a part of the serial autocorrelation of the residuals from the non-augmented specification is due to omitted variables.[7] The LR coefficient of inflation remains greater than 1, whereas the LR output gap coefficient falls from 1.12 to 0.93. The additional variable is statistically significant and has the expected negative sign. The overall fit of the model improves as signalled by all the summary statistics reported at the bottom of Table 1.

Potentially, this additional variable could be endogenous and its estimated coefficient therefore biased and inconsistent. To address this important technical issue we have carried out a battery of Granger causality and Hausman tests for the endogeneity of the regressor. The test statistics reject the hypothesis that the credit spreads are endogenous. We conclude, in line with Castelnuovo (2003) and Gerlach-Kristen (2004), that the endogeneity problem for z_t is negligible.[8]

Figure 2 plots the residuals of the linear augmented rule (2). Apparently, the addition of the explanatory variable contributes to reducing the correlation in the residuals, the positive spikes in the 1994–96 period and the largest negative spikes in 2001. However, the fitted interest rates remain, over certain time intervals, correlated and remarkably different (even for 60 bps) from the actual values. With the benefit of hindsight and on the basis of anecdotal evidence, such errors can possibly be associated with the Fed's reaction to the stock market crisis and the money market instability after 2001. To investigate to what extent concerns related to these issues might have affected the monetary policy stance, we adopt an LSTR model that can detect a nonlinear behaviour in the policy conduct.

TABLE 1
ESTIMATION OF LINEAR (BASELINE AND AUGMENTED) TAYLOR RULES

	Baseline		Augmented	
	Estimate	St. error	Estimate	St. error
Coefficients				
a	0.138	0.124	1.035	0.21***
b_π	0.122	0.059**	0.132	0.051***
b_y	0.095	0.032***	0.113	0.028***
w_z			– 0.363	0.074***
ρ_1	1.520	0.092***	1.336	0.088***
ρ_2	– 0.605	0.085***	– 0.457	0.080***
LR coefficients				
α	1.632		8.553	
β_π	1.436		1.090	
β_y	1.123		0.933	
β_z			3.000	
Summary statistics	SSR	AIC	BIC	HQ
Baseline	7.245	– 2.188	– 2.032	– 2.126
Augmented	5.477	– 2.441	– 2.254	– 2.366

Notes
This table reports the estimates and standard errors of the coefficients of the baseline and augmented Taylor rules, respectively equations (1) and (2) in the main text. The sample period is 1987Q3 to 2005Q4.
** and *** denote significance at the 5% and 1% levels, respectively.
The LR coefficients correspond to those in the original Taylor (1993) specification without autoregressive terms. The table also reports the sum of squared residuals (SSR) and the Akaike (AIC), Bayesian (BIC) and Hanna–Quinn (HQ) information criteria.

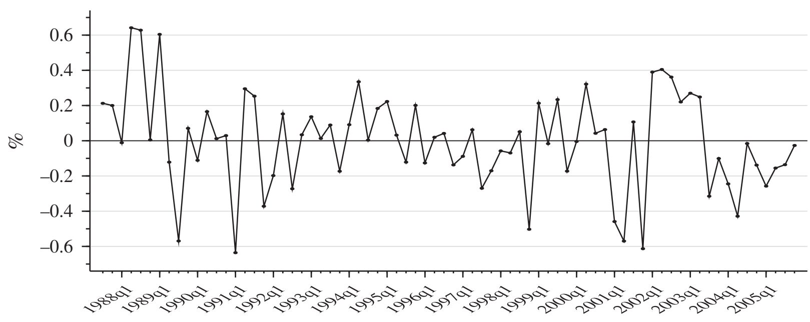

FIGURE 2. Residuals from the augmented Taylor rule (2).

II. THE NONLINEAR LSTR MODEL

Following closely the work of Franses and van Dijk (2000) and van Dijk *et al.* (2000), a two-regime smooth transition (STR) model can be represented as

$$y_t = \phi_1' x_t(1 - G(l_t; \gamma, c)) + \phi_2' x_t G(l_t; \gamma, c) + \varepsilon_t, \tag{3}$$

where y_t is a univariate series, x_t is a vector of regressors including exogenous and, possibly, lagged endogenous variables, and Φ_1 and Φ_2 are two sets of coefficients that identify two different regimes. $G(l_t;\, \gamma,\, c)$ is the transition function allowing for gradual changes between the two regimes; it is a continuous function, bounded between 0 and 1, and it depends on (i) the transition variable l_t, which determines the movements across the regimes, (ii) the speed parameter γ, which determines how fast the transition between the regimes occurs, and (iii) the threshold c that corresponds to the value of the transition variable splitting one regime from the other. Both c and γ are estimated by the model. Among the different functional forms of $G(l_t; \gamma, c)$ employed in the literature, we choose the logistic function as it does not presume symmetric regimes. This function can be written as $G(l_t; \gamma, c) = (1 + \exp\{-\gamma(l_t - c)\})^{-1}$, with $\gamma > 0$. Accordingly, the model is a logistic STR (LSTR) model.

Various scholars, among others Martin and Milas (2004), Osborn *et al.* (2005), Kesriyeli *et al.* (2006) and Haug and Siklos (2006), have already employed the STR technique in monetary policy analysis. Indeed, this technique allows for the identification of possible nonlinear behaviour without *ex ante* imposing either the existence of multiple regimes or the critical thresholds above/below which the different regimes take place.

The existence of multiple policy regimes can be tackled in alternative fashions. Many researchers, for instance, are familiar with the threshold autoregressive (TAR) model. The latter assumes that the regime occurring at any point in time can be determined by the state of an observable variable relative to a threshold value, and that the switch from one regime to another is discrete. By replacing the indicator function with a continuous one, a gradual transition between two regimes is obtained. The result is an STR model. It follows that the TAR model is a special case of the STR model when the speed of transition is infinite. In our application we prefer to allow for smooth changes because, as suggested by Boivin (2006), there is no guarantee that policy changes occur so fast as to be well identified by discrete jumps in the parameters.

A second class of regime-switching models builds on the assumption that the current regime cannot be observed, as it is determined by an unobservable process. One famous member of this class is the Markov-switching (MSW) model, which is characterized by the assumption that the current regime depends on the regime in the previous period and on defined transition probabilities of moving from one state to the other.[9] The MSW and STR methodologies differ in that the latter identifies the regimes by means of coefficients that change in a deterministic but smooth fashion according to the movements of a transition variable, whereas the former requires the *a priori* specification of the alternative regimes and a stochastic transition. Since the STR model requires that the regime changes be associated with the behaviour of an observable variable with respect to an estimated threshold, it follows that it fits our estimation problem particularly well. Indeed, we adopt as candidate transition variables those that are closely linked to the concerns that might have induced switches in the Fed's monetary policy stance.

A third approach to deal with regime switches has been followed, among others, by Cogley and Sargent (2005) and Boivin (2006), who estimated Taylor rules with drifting coefficients. The difference between the STR model and those allowing for unconstrained time-varying coefficients is in spirit and motivates our choice in favour of the STR form. A time-varying coefficient estimation does not impose restrictions on the way parameters vary over time, but it is up to the researchers to *ex post* match the estimated changes with the occurrence of some known circumstances. In our work, on the contrary, we first single out those events and contingencies that are likely to have affected monetary policy and, after having found appropriate indicators, we test whether monetary policy changed in response to these events.[10]

More traditional methods of testing for the relevance of special circumstances consist in looking at sample splits. Yet this solution has major shortcomings. In particular, the different sample periods must be identified on the basis of *a priori* knowledge and characterized by parameter constancy, the specification requires on/off discrete regime switches, and the coefficients of the variables relevant only in extreme events have to be not significantly different from zero in 'normal times'. The STR technique does not share these limitations.

Finally, it is worth stressing that since the STR model nests a linear model, an STR specification is less restrictive than the latter. If a linear functional form were correct, the STR would exclude nonlinear effects. It is possible, for instance, that Knightian uncertainty about the future path of the economy leads policy-makers not to change their stance even when an extreme event occurs. In such a case, since there is no policy regime switch to detect, the choice of the indicator is immaterial and the linear specification outperforms the nonlinear one. Instead, if some nonlinearity existed, the STR model would allow the regime changes to be either smooth or sharp, long-lasting or short-lived, without entailing restrictions on the speed, the intensity or the persistence of the changes.

These considerations have led us to use an STR technique to allow the parameters in the model to vary across regimes. Before proceeding to the estimation of the model, we need to test for the assumption of linearity against an LSTR specification. Accordingly, we first need to identify transition variables which act as indicators of the two main concerns of the Fed after 2001, suggested by the narrative evidence discussed in the Introduction as well as by the pattern of the residuals plotted in Figure 2. The first is a general concern about the health of the financial system and the implications of a financial crisis. The second refers to the Fed's concern that interest rates may hit the zero lower bound (ZLB) and the US economy may end up in deflation.

For reasons that will be spelled out in the next section, we employ the BAA spreads (z_t) as an indicator of the first concern, while for the ZLB concern we adopt the one-period lagged interest rate (i_{t-1}). We run the tests of linearity for the current values for z, for the first lags of z and i, and, as a test of parameter constancy, a time trend. An LM-type test with F-distribution is used to test the null hypothesis of linearity. Based on the fact that the power of the test is maximum if the alternative model is correctly specified, i.e. the transition is correct, the selected transition variables are those with the lowest p-values of the LM3 statistic. The results of the tests for the above-mentioned candidates are reported in Table 2.

On consideration of these results, we conclude that there is strong evidence against the linear specification of the Taylor rule in equation (2) and that the behaviour of credit spreads and lagged interest rates is likely to be related to the nonlinear behaviour of the Fed.

TABLE 2

STATISTICS AND *p*-VALUES OF TESTS FOR NONLINEARITY

Candidate transition	$z(t)$	$z(t-1)$	$i(t-1)$	Time
F-statistic	3.262	1.994	3.671	2.631
p-value	0.001	0.028	0.000	0.004

Notes

This table reports the statistics and the p-values of the tests for nonlinearity. The linearity of the augmented Taylor rule is tested against alternative LSTR specifications embedding one candidate transition variable at the time.

III. Nonlinear Estimation: Unveiling Finer Policy Regimes

In this section we report the estimates[11] of the LSTR Taylor-type rules which include as transition variables the contemporaneous credit spread and the lagged interest rate.

General concerns on financial stability

In an ideal world, where all information is freely and immediately accessible, indicators of the CB's actual concerns about the stability of the financial system would be available. Alas, this is not the case and some proxy for them must be found. We believe that a plausible solution is to encompass an indicator of market sentiment which reflects investors' expectations and worries, such as the BAA credit spreads. Traditionally, central bankers tend to put more emphasis on inflation than output control, while private investors are more worried about the real performance of the companies they lend money to. In periods of financial distress, however, a long-lasting and/or intense contraction has negative implications both for the solvency of the borrowing firms and the risk of falling into deflation. Since, according to the narrative evidence provided in the Introduction, the Fed switched its policy stance in reaction to situations of financial distress (e.g. the stock market crash started in mid-2000 and aggravated by the terrorist attacks on 11 September 2001),[12] we submit that BAA credit spreads may be a good proxy for central bankers' concerns. The tests of linearity described in Section II support such conviction.

In Table 3 we show the estimates of the LSTR version of equation (2) in which the transition variable is the BAA spread. We adopt two specifications for the 'finer' regime. Both estimations provide evidence in favour of the existence of two policy regimes, linked to low and high values of z_t. In both specifications, one regime is very close to the linear augmented rule reported in Table 1, while the other (which we call 'high-spread regime') is considerably different because inflation and output gap are not included among the determinants.

In particular, Specification 1 is minimal, and only the autoregressive component enters significantly in the functional form of the high-spread regime.[13] Although this finding seems poor, it should not be surprising that current output gap and inflation say little (if anything) about the prospective behaviour of the economy in periods characterized by financial instability. Reasonably, variables which more closely reflect the actual expectations of the central bankers can contribute to get a better description of the Fed's reaction function in such a particular situation. Given the Fed's concerns about the risk of recession as consequence of financial instability, Specification 2 embeds as additional regressor the probability of a decline in real GDP in the next quarter, i.e. rec_t.[14] The estimates of Specification 2, where rec_t enters significantly and with the expected sign, support the narrative evidence.

The estimated thresholds are similar, equal to 2.75 for Specification 1 and 2.77 for Specification 2. As can be noted in Figure 3, these values split the sample into two distinct periods of time. In particular, the high-spread regime includes the observations from late 2000 to mid-2003.[15] This is in perfect accordance with Greenspan's account of the Fed policy conduct.

To assess whether the LSTR specification performs better than the linear one in terms of the goodness of fit of the model, we resort both to some synthetic econometric indicators reported in Table 3 and to the analysis of the pattern of the residuals. Starting with the former, we notice that allowing the functional form to deviate from a constant-

TABLE 3

ESTIMATES OF LSTR TAYLOR RULES—TRANSITION VARIABLEZ z_t AS PROXY OF THE FED'S GENERAL CONCERNS

	Specification 1				Specification 2			
	No high-spread regime		High-spread regime		No high-spread regime		High-spread regime	
	Estimate	St. error	Estimate	St. error	Estimate	St. error	Estimate	St. error
Coefficients								
a	1.138	0.259***			1.157	0.235***	0.831	0.286***
b_π	0.166	0.050***			0.166	0.046***		
b_y	0.146	0.028***			0.146	0.026***		
w_z	− 0.474	0.117***			− 0.484	0.105***		
rec							− 0.031	0.010***
ρ_1	1.285	0.097***	0.802	0.032***	1.283	0.088***	0.817	0.095***
ρ_2	− 0.400	0.088***			− 0.398	0.080***		
Parameters of the transition function								
Speed (γ)	10.446	0.884***				7.043	1.123***	
Threshold (c)	2.755	0.081***				2.776	0.112***	
Summary	SSR	AIC	BIC	HQ	SSR	AIC	BIC	HQ
statistics	4.623	− 2.529	− 2.249	− 2.417	3.872	− 2.652	− 2.310	− 2.516
Diagnostic tests	Serial	Remaining	Parameter	Residual	Serial	Remaining	Parameter	Residual
(p-*values*)	correlation	nonlinearity	constancy	normality	correlation	nonlinearity	constancy	normality
	order 1	z_t	*LMc*1		order 1	z_t	*LMc*1	
	0.222	0.494	0.010	0.435	0.170	0.387	0.011	0.286
	order 2		*LMc*3		order 2		*LMc*3	
	0.282		0.004		0.182		0.007	

Notes

This table reports the estimates and standard errors of the coefficients of two specifications of LSTR Taylor rules and of the parameters of the transition function $G(z_t;\gamma, c) = (1 + \exp\{-\gamma(z_t - c)\})^{-1}$. The sample period is 1987Q3 to 2005Q4.

** and *** denote significance at the 5% and 1% levels, respectively.

The table also reports the sum of squared residuals (SSR) and the Akaike (AIC), Bayesian (BIC) and Hanna–Quinn (HQ) information criteria. The diagnostic section includes the p-values for the following test statistics: two generalized LM tests for the hypothesis of no serial correlation (order 1 and 2) in the residuals, an LM test for the null of no remaining nonlinearity linked to z_t, the LMc1 and LMc3 tests for the null of no remaining nonlinearity linked to time, and the Jarque–Bera test for the null of residual normality.

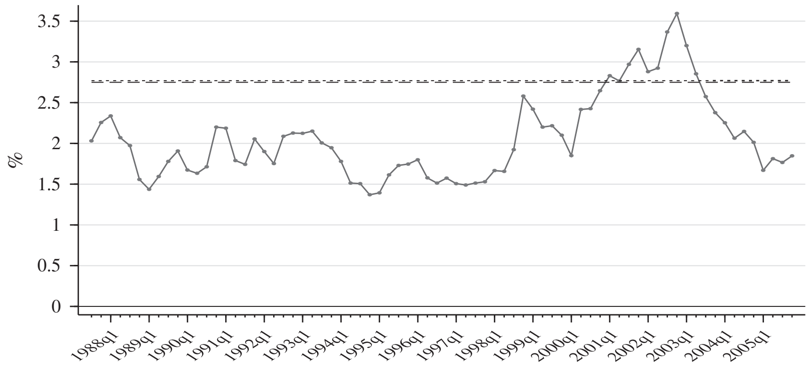

FIGURE 3. Transition variable z_t, with estimated thresholds for general concerns marked.

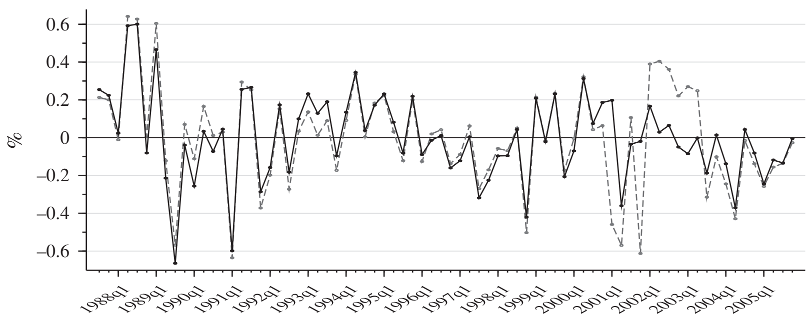

FIGURE 4. Residuals from the linear augmented (dashed) rule and Specification (2) of the nonlinear (solid) rule; z_t as transition variable.

parameter Taylor rule improves the fit of the model. This is confirmed by the diagnostic tests.[16]

In Specification 1, the residuals are almost identical to those of the linear rule over the first 13 years of the sample, yet the nonlinear estimation performs better after 2000 when it yields smaller and less autocorrelated residuals. Specification 2 performs even better, as shown in Figure 4, where the associated residuals are plotted with those from the linear augmented rule.

In order to better appreciate the gain in terms of fit obtained by moving from the linear rule to this nonlinear specification, in Figure 5 we plot the difference between the residuals (in absolute value) from the former and the latter. In the first and second quarters of 2001 the improvement is of 26 and 21 bps and, more remarkably, in the fourth quarter of 2001 the difference is of 60 bps, given that the nonlinear rule produces a residual almost equal to zero. Such an improvement is larger than half a point, which is a monetary policy decision of certain importance. Interestingly, the nonlinear specification accounts for the particularly aggressive stance of the Fed in the initial stage of the recession in 2001, when the interest rates were lowered quickly even if inflation and inflation expectations were low and stable. Besides this large one-off gain in 2001Q4, the nonlinear rule produces smaller residuals for 14 quarters in a row starting from 2001Q1. Overall, between 2001Q1 and 2003Q2, Specification 2 outperforms the linear rule of about 260 bps, i.e. an average of 25 bps a quarter over 10 quarters. This implies that the

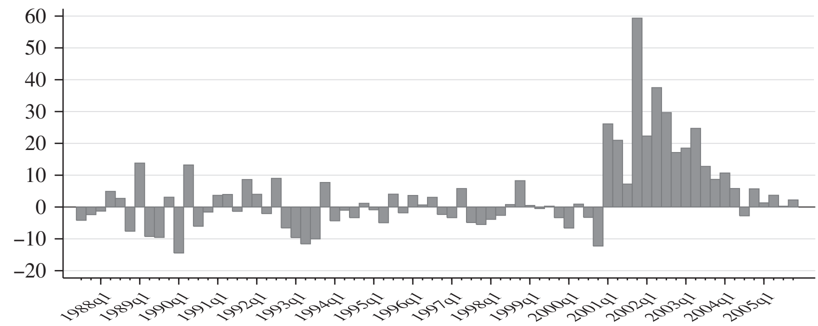

FIGURE 5. Difference in basis points between the residuals (in absolute value) from the linear augmented rule and Specification (2) of the nonlinear rule; z_t as transition variable.

nonlinear rule also captures the flattening out of the rates since 2002, while the linear rule first overestimates and then underestimates the actual rates. Notably, such improvements in the period after 2001 do not come at the cost of a worsened fit in any other part of the sample.

In this section we focused on the Fed's concerns linked to financial market distress. In the next section we intend to test for the Fed's worries about hitting the ZLB.

The zero lower bound concern

In the USA the risk of deflation materialized after the bursting of the asset bubble in 2000. Indeed, the aggressive monetary easing that mitigated the fallout of the bubble drew the interest rates close to zero. Even though the Fed considered deflation a low probability occurrence, the concern about this contingency led to unusually low interest rates because, as argued by Bernanke (2002) and Greenspan (2005), it could have had dire implications for the economy.

These considerations motivated us to investigate if, and to what extent, data could identify a source of nonlinearity in the monetary policy generated by this concern. In order to do so, we estimate the augmented Taylor rule by means of an LSTR model (as formulated in equation (3)) in which the transition variable is the lagged interest rate i_{t-1}. The estimates are reported in Table 4.

Table 4 confirms our conjectures: the worry about hitting the ZLB is mapped in a finer monetary regime (that we call the ZLB regime) where the Fed did not react to current output and inflation. While the estimates of the no-ZLB regime do not differ much from the linear ones, the rule for the ZLB regime is an autoregressive process. As mentioned in the previous case, a simple autoregressive specification might conceal an omitted variable problem and this is likely to be the case in this application. Indeed, it is impossible to know and embed those exact variables that actually informed policy-makers in this special occurrence.[17]

Nevertheless, the estimates clearly reveal the existence of a finer regime that is at odds with the classical Taylor rule. The estimated threshold for the lagged interest rate is roughly equal to 3%. This means that when the interest rate falls below such a value, monetary policy enters the ZLB regime where the policy instrument does not respond in the usual way to its determinants. By looking at Figure 6, which plots the transition variable and the estimated threshold, we can identify the timing of the policy switch:

TABLE 4
ESTIMATES OF LSTR TAYLOR RULE—TRANSITION VARIABLE i_{t-1} AS PROXY OF THE FED'S ZLB CONCERNS

	No-ZLB regime		ZLB regime	
	Estimate	St. error	Estimate	St. error
Coefficients				
a	3.627	1.090***		
b_π	0.510	0.094***		
b_y	0.246	0.040***		
w_z	− 0.788	0.138***		
ρ_1	1.048	0.138***	0.603	0.191***
ρ_2	− 0.535	0.096***		
Parameters of the transition function				
Speed (γ)	1.354	0.151***		
Threshold (c)	3.156	0.081***		
Summary statistics	SSR	AIC	BIC	HQ
	3.701	− 2.752	− 2.471	− 2.640
Diagnostic tests (p-*values*)	Serial correlation order 1	Remaining nonlinearity i_{t-1}	Parameter constancy *LMc*1	Residual normality
	0.709	0.078	0.023	0.411
	order 2		*LMc*3	
	0.630		0.075	

Notes
This table reports the estimates and the standard errors of the coefficients of the LSTR Taylor rule and of the parameters of the transition function $G(i_{t-1};\gamma, c) = (1 + \exp\{-\gamma(i_{t-1} - c)\})^{-1}$. The sample period is 1987Q3 to 2005Q4.
** and *** denote significance at the 5% and 1% levels, respectively.
The table also reports the sum of squared residuals (SSR) and the Akaike (AIC), Bayesian (BIC) and Hanna–Quinn (HQ) information criteria. The diagnostic section includes the *p*-values for the following test statistics: two generalized LM tests for the hypothesis of no serial correlation (order 1 and 2) in the residuals, an LM test for the null of no remaining nonlinearity linked to i_{t-1}, the LMc1 and LMc3 tests for the null of no remaining nonlinearity linked to time, and the Jarque–Bera test for the null of residual normality.

notably, the time span of the ZLB regime coincides with the anecdotal evidence about the period when the Fed faced the risk of hitting the ZLB.

Assessing the goodness of fit of the model, the information criteria in Table 4 show an improvement with respect to the linear rule. Figure 7 reveals that the linear and nonlinear models yield different residuals both in the very beginning of the sample and from 2001 onwards. In the first part of the sample, up to 1991, the linear and nonlinear rule produce very different predicted values without one of the two rules outperforming the other. After 2001, instead, the residuals from the nonlinear rule fluctuate more closely around zero. In 2002Q1 the improvement is about 33 bps and the overall gain from 2002Q1 to 2003Q3 is about 100 bps.

Discussion and interpretation of the results

The results of the estimations indicate that if there are situations that required central bankers to apply a special dose of judgment, then the period after 2000 is likely to be one

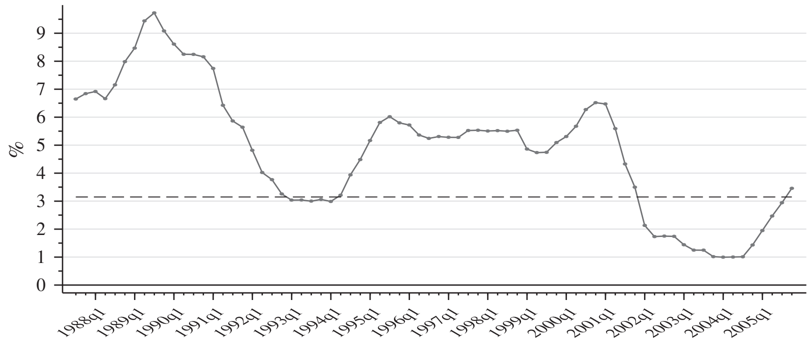

FIGURE 6. Transition variable i_{t-1}, with the estimated threshold for ZLB concerns marked.

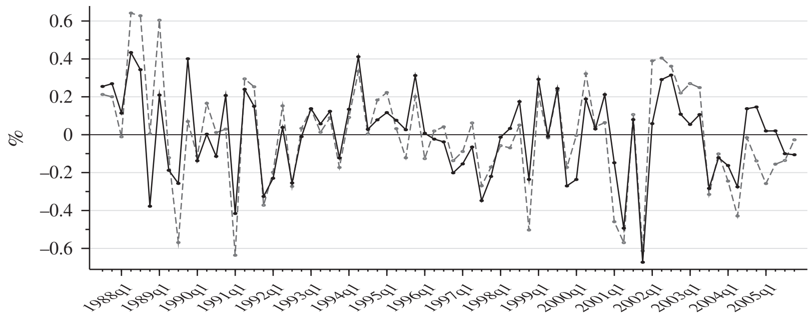

FIGURE 7. Residuals from the linear augmented (dashed) rule and the nonlinear (solid) rule; i_{t-1} as transition variables.

of them. Our findings confirm that, in such critical circumstances, the usual indicators that inform monetary policy are temporarily set aside. This is in line with the stress Svensson (2005) puts on the role of judgment in policy-making.

Given the available data and the estimation technique, we do not manage to disentangle all the finer regimes at the same time. Thus in each estimation two regimes of different length cover the whole sample under scrutiny. The shorter regime can be seen as a 'special regime' that applies only to unusual economic conditions in response to which central bankers change their usual policy conduct. On the contrary, the other one can be interpreted as either the 'general regime' or the 'remaining regime' because it is likely to include some finer regimes associated with concerns different from those proxied by the transition variable.

Since most of the observations in the sample belong to the 'general regime', it follows that its estimated coefficients are similar to those of the linear specification. This result reconciles the good *ex post* descriptive properties of simple linear Taylor rules with the judgmental way monetary policy decisions are in fact made. The linear specification, by imposing a unique regime over the entire sample, does catch the broad contours of monetary policy, but fails to take the 'special regimes' into account. Such 'special regimes', however, contain relevant information about the monetary policy conduct. Indeed, each of them refers to some special circumstances in which policy-makers disconnect the automatic pilot and extensively use their judgment to make decisions.

It could be argued that what we call a regime switch is, instead, a prolonged deviation of monetary policy from a 'normal' behaviour, which is represented by the linear Taylor rule. If several regimes exist, however, a linear rule is an average of several finer rules and it does not describe the policy stance associated with 'normal' economic conditions: an 'average' rule is not necessarily 'the normal time' rule. We do not exclude the existence of such 'normal time' behaviour, yet we warn that this is not necessarily described by an estimated linear Taylor rule.

At this point, a natural question to ask is whether inflation targeting is consistent with the sequence of regimes we find during the Greenspan era. The reason to wonder is twofold. First, according to Bernanke and Gertler (1999), the Fed has conducted an implicit inflation targeting in recent years. Second, inflation targeting policies may involve significant nonlinearities in the relationship between the interest rate and the variables that describe the state of the economy. As Svensson (2003) points out, targeting rules in general, and inflation targeting in particular, permit the rational exploitation of all the information that a CB has access to, even if outside the scope of the model used to describe the economy. Interestingly, this information is distilled in what can be interpreted as the judgment of the CB, which was a crucial ingredient in the risk management approach proposed by Greenspan. In this respect, we believe that the finer regimes that our nonlinear analysis unveils are consistent with the Fed having adopted a (*quasi*) inflation targeting approach.

Before concluding this section, we would like to bring up a final issue. One could conjecture that those same events that drive monetary regime switches also (nonlinearly) affect the CB's expectations and the transmission mechanism of the economy. Testing this hypothesis would importantly contribute to shedding light on expectations formation and on the economic transmission mechanism. Yet this would require the disclosure of the actual expectations the Fed has been using in recent years and these, at the moment, are still not publicly available.

IV. Closing Remarks

In view of some existing narrative evidence on the risk management approach advocated by Greenspan, we investigate whether during his tenure the Fed's monetary policy stance nonlinearly changed due to the bank's judgment on financial instability and the risk of hitting the ZLB. To do so, we employ an LSTR estimation model, which allows us to detect endogenously finer monetary policy regimes by exploiting observable indicators which intercept these particular CB concerns. These finer regimes are sufficiently short to be averaged out by a linear Taylor rule, but sufficiently important to be captured by an LSTR model.

Using this framework we find that while a linear Taylor rule describes well the broad contours of the Fed's behaviour during Greenspan's tenure, it neglects to detect important changes in the policy stance in response to the fallout of the financial turmoil in 2001 and the risk of hitting the ZLB in 2002–03. According to our findings, the adoption of an appropriate nonlinear specification instead of a linear one leads to a reduction in the errors of 60 bps in 2001Q4 and of 260 bps concentrated over the period 2001Q1 to 2003Q2. This implies that for 10 consecutive quarters a linear Taylor rule misses the actual interest rate by an amount as large as a typical policy decision of 25 bps.

Finally, the paper contributes to the recent literature on the existence of changes in the conduct of US monetary policy by identifying a novel source of nonlinearity in the CB judgment over special events and/or contingencies.

APPENDIX: INFLATION FORECAST-BASED NONLINEAR TAYLOR RULES

The baseline Taylor rule adopted in this paper employs contemporaneous output gap and inflation. In the literature a different specification, where current inflation is replaced by its forecast, has sometimes been estimated so as to (better) grasp the forward-looking behaviour of the monetary authorities. In this appendix, we show that our results about the existence of finer regimes carry over to this alternative specification. Since the exact inflation forecasts that the Fed used are not available, we resort to the one-year-ahead private sector's inflation forecasts (from the Survey of Professional Forecasters) to estimate linear and nonlinear Taylor rules. As in the main text, we consider z_t and i_{t-1} as transition variables. The results of the estimations are reported in Table A1.

As indicated by the summary statistics, the relaxation of the linearity constraint improves the goodness of fit of the forecast-based rule in both cases. We obtain the largest improvement when the nonlinearity is associated with z_t. As can be seen from Table A1 and from the residuals

Table A1

Linear and LSTR Estimates of Forecast-based Taylor Rules—Transition Variables z_t and i_{t-1}

	Linear		General concerns, z_t transition				ZLB concerns, i_{t-1} transition			
			No high-spread regime		High-spread regime		No-ZLB regime		ZLB regime	
	Estimate	St. error	Estimate	St. error	Estimate	St. error	Estimate	St. error	Estimate	St. error
Coefficients										
a	0.488	0.198**	0.787	0.202***	0.812	0.217***	0.812	0.204***		
$b_{\pi t+4}$	0.415	0.062***	0.386	0.057***			0.404	0.058***	0.416	0.208***
b_y	0.186	0.026***	0.196	0.024***			0.203	0.024***		
w_z	− 0.286	0.062***	− 0.445	0.087***			− 0.461	0.074***	− 0.411	0.120***
rec					− 0.030	0.008***				
ρ_1	1.124	0.081***	1.131	0.080***	0.803	0.089***	1.095	0.081***	1.181	0.292***
ρ_2	− 0.358	0.067***	− 0.348	0.068***			− 0.321	0.069***		
Parameters of the transition function										
Speed (γ)			67.802	1.146***			18.923	2.704***		
Threshold (c)			2.843	0.012***			2.115	0.145***		
Summary statistics										
SSR	3.738		2.849				3.039			
AIC	− 2.823		− 2.959				− 2.895			
BIC	− 2.636		− 2.617				− 2.552			
HQ	− 2.748		− 2.822				− 2.758			

Notes

This table reports the estimates and the standard errors of the coefficients of linear and nonlinear LSTR specifications of a forecast-based Taylor rule. The sample period is 1987Q3 to 2005Q4. The table includes the estimates of the parameters of the transition function $G(l_t;\gamma, c) = (1 + \exp\{-\gamma(l_t - c)\})^{-1}$, where l_t is either z_t or i_{t-1}.

** and *** denote significance at the 5% and 1% levels, respectively.

The table also reports the sum of squared residuals (SSR) and the Akaike (AIC), Bayesian (BIC) and Hanna–Quinn (HQ) information criteria.

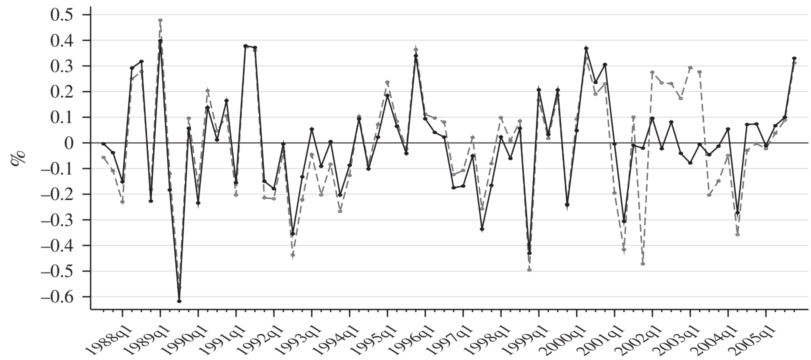

FIGURE A1. Residuals from the linear augmented (dashed) and the nonlinear (solid) forecast-based rules; z_t as transition variable.

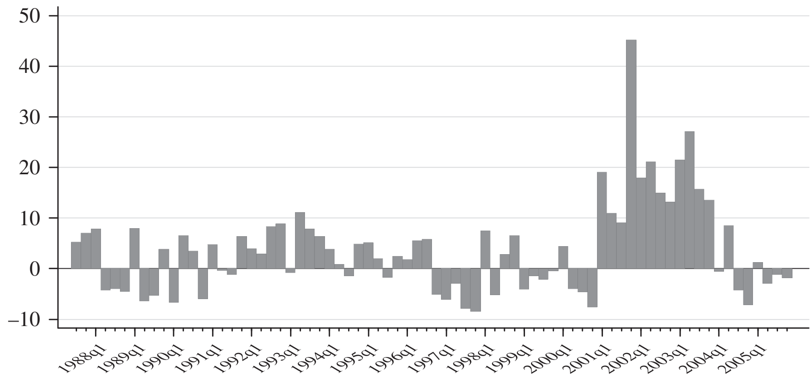

FIGURE A2. Difference in basis points between the residuals (in absolute) from the linear augmented and the nonlinear forecast-based rules; z_t as transition variables.

displayed in Figure A1, the finer (high-spread) regime resembles the one shown in the text and generated by a contemporaneous rule.

Accounting for nonlinearity leads to a 230 bps improvement in the fit from 2001Q1 to 2003Q4, without any significant difference over the remaining part of the sample. As shown in Figure A2, the largest single improvement occurs in 2001Q4 and amounts to 45 bps. Similarly, adopting a forecast-based rule to investigate the ZLB concern, we find an improvement of 150 bps over 9 quarters, starting in 2001Q4.

ACKNOWLEDGMENTS

We would like to thank an anonymous referee for important and detailed comments. We also thank Jean Boivin, Claudio Borio, Martina Bozzola, Efrem Castelnuovo, Emanuela Ceva, John Cuddy, Aaron Drew, Hans Genberg, Petra Gerlach-Kristen, Özer Karagedikli, Kirdan Lees, Ashley Lienert, Rosen Marinov, Costas Milas, Anne Peguin-Feissolle, Stefano Schiavo, Rishab Sethi, Ulf Soderstrom, Alexander Swoboda, Roberto Tamborini, Cedric Tille, Charles Wyplosz, and the participants in seminars at the New Zealand Reserve Bank and the Graduate Institute in Geneva, and at the T2M Conference in Toulouse, the Cifrem Seminar in Trento, Keele University, the 17th $(EC)^2$ Conference in Rotterdam and the SSES meeting in St Gallen. Alcidi gratefully acknowledges the financial support of the NCCR-FINRISK research programme. It goes without saying that all mistakes are ours.

NOTES

1. See Borio and Lowe (2004), Dolado *et al.* (2004), Karagedikli and Lees (2004), Taylor and Davradakis (2006) and Surico (2007). Martin and Milas (2005), instead, focus on nonlinearities associated with economic volatility.
2. See Clarida *et al.* (2000), Owyang and Ramey (2004), Osborn *et al.* (2005), Cogley and Sargent (2005), Duffy and Engle-Warnick (2006), Boivin (2006), Sims and Zha (2006) and Haug and Siklos (2006).
3. Allegedly, Taylor-type rules tend to be incomplete and not robust to changes in the details of their specification and to alternative measures of their determinants. On these and other criticisms see Kozicki (1999), Orphanides (1998), Siklos and Wohar (2005), Carare and Tchaidze (2005), Svensson (2003) and Woodford (2001).
4. The LM tests, the ACF and the PACF suggest the use of two lags, as also done in Judd and Rudebusch (1998) and Woodford (2003, p. 41). An accurate specification of the linear model is crucial to avoid over-rejection of the correct hypothesis of linearity.
5. An alternative forecast-based specification has been used in the literature to take into account the forward-looking behaviour of the CB. As in Assenmacher-Wesche (2006) and according to what is suggested by Svensson (1999), however, current values of inflation and output can be used as substitutes for future expected inflation and output. As shown in the Appendix, our results also hold using a forecast-based specification.
6. The BAA spreads are determined by the investors' risk aversion and the solvency risk of the companies issuing the assets encompassed in the index. When the spreads widen, an aggregate negative shock is likely to have hit (or is expected to hit) the economy. Since CBs tend to react to such kinds of events by relaxing monetary policy, we expect the variable to enter with a negative sign in the rule. This is in line with the role played by credit-related variables found by Borio and Lowe (2004).
7. It is still debated whether the serial correlation in the errors is due to persistent omitted variables (as maintained by Rudebusch 2002) or to authentic interest rate smoothing (as claimed by English *et al.* 2003, Castelnuovo 2003, and Gerlach-Kristen 2004).
8. We thank an anonymous referee for having raised this point. The results are available upon request.
9. For the application of the MSW model to detect the existence of multiple regimes in the US monetary policy, see, for instance, Rabanal (2004), Owyang and Ramey (2004) and Assenmacher-Wesche (2006). See Franses and van Dijk (2000) for a discussion and comparison of TAR, STR and MSW techniques.
10. Notably, even though it is possible to fail to detect a regime switch if a wrong indicator is adopted, this would be only a Type II error. An error of Type I, that is, finding a policy change where there is none, should not occur because the STR model nests the linear one.
11. For the estimation we modified the codes on 'Regime-switching models for returns' by Dick van Dijk.
12. Other researches have investigated whether stock market crashes affected the US monetary policy. Among them see Gerlach-Kristen (2004) and Surico *et al.* (2005).
13. Following the approach developed by Franses and van Dijk (2000), we drop insignificant coefficients and select the appropriate specification of the two regimes according to the minimum AIC.
14. This measure comes from the Survey of Professional Forecasters, released quarterly by the Federal Reserve Bank of Philadelphia.
15. These results are consistent with those produced by Assenmacher-Wesche (2006).
16. The tests fail to reject the hypotheses of no residual correlation, no residual nonlinearity in z_t, and residual normality. The diagnostic test for parameter constancy, which considers *time* as alleged transition, suggests the possible presence of additional unspecified nonlinearity.
17. The estimation of the rule in the special regime is also made difficult by the limited number of observations belonging to it. Tackling this problem by using monthly data would have important drawbacks. First, OLS estimations of Taylor rules at monthly frequency produce estimates with hard economic interpretation. Second, specification problems with monthly data make the tests for nonlinearity unreliable. Finally, Taylor-type rules and their stability conditions are meant for quarterly frequency. On these grounds we have decided to stick to quarterly data.

REFERENCES

Assenmacher-Wesche, K. (2006). Estimating central banks' preferences from a time-varying empirical reaction function. *European Economic Review*, **50**(8), 1951–74.

Bernanke, B. S. (2002). Deflation: making sure 'it' doesn't happen here. Speech, Federal Reserve Board, 21 November.

——— and Gertler, M. (1999). Monetary policy and asset price volatility. Paper presented at the symposium *New Challenges for Monetary Policy*, Federal Reserve Bank of Kansas City.

Boivin, J. (2006). Has US monetary policy changed? Evidence from drifting coefficients and real-time data. *Journal of Money, Credit and Banking*, **38**(5), 1149–73.

BORIO, C. and LOWE, P. (2004). Securing sustainable price stability: should credit come back from the wilderness? BIS Working Papers no. 157, Bank for International Settlements.

CARARE, A. and TCHAIDZE, R. (2005). The use and abuse of Taylor rules: how precisely can we estimate them? IMF Working Papers no. 05/148, International Monetary Fund.

CASTELNUOVO, E. (2003). Taylor rules, omitted variables, and interest rate smoothing in the U.S. *Economics Letters*, **81**(1), 55–9.

CLARIDA, R., GALI, J. and GERTLER, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *Quarterly Journal of Economics*, **115**(1), 147–80.

COGLEY, T. and SARGENT, T. S. (2005). Drift and volatilities: monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics*, **8**(2), 262–302.

DOLADO, J., PEDRERO, R. and RUGE-MURCIA, F. (2004). Nonlinear monetary policy rules: some new evidence for the U.S. *Studies in Nonlinear Dynamics & Econometrics*, **8**(3), article 2.

DUFFY, J. and ENGLE-WARNICK, J. (2006). Multiple regimes in US monetary policy? A nonparametric approach. *Journal of Money, Credit and Banking*, **38**(5), 1363–77.

ENGLISH, W., NELSON, W. and SACK, B. (2003). Interpreting the significance of the lagged interest rate in estimated monetary policy rules. *Contributions to Macroeconomics*, **3**(1), article 5.

FRANSES, P. H. and VAN DIJK, D. (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.

FRIEDMAN, B. M. (2006). The Greenspan era: discretion, rather than rules. *American Economic Review*, **96**(2), 174–7.

GERLACH-KRISTEN, P. (2004). Interest-rate smoothing: monetary policy inertia or unobserved variables? *Contributions to Macroeconomics*, **4**(1), 1169.

GREENSPAN, A. (2004). Risk and uncertainty in monetary policy. *American Economic Review*, **94**(2), 33–40.

——— (2005). Reflections on central banking. Speech at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, WY.

HAUG, A. and SIKLOS, P. (2006). The behavior of short-term interest rates: international evidence of non-linear adjustment. *Studies in Nonlinear Dynamics & Econometrics*, **10**(4), 1276.

JUDD, J. P. and RUDEBUSCH, G. D. (1998). Taylor's rule and the Fed, 1970–1997. *FRBSF Economic Review*, **3**, 3–16.

KARAGEDIKLI, O. and LEES, K. (2004). Do inflation targeting central banks behave asymmetrically? Evidence from Australia and New Zealand. Discussion Paper Series no. DP 2004/02, Reserve Bank of New Zealand.

KESRIYELI, M., OSBORN, D. R. and SENSIER, M. (2006). Nonlinearity and structural changes in interest rate reaction functions of US, UK and Germany. In C. Milas, P. Rothman and D. van Dijk (eds), *Nonlinear Time Series Analysis of Business Cycles*. Amsterdam: Elsevier.

KNIGHT, F. H. (1921). *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.

KOZICKI, S. (1999). How useful are Taylor rules for monetary policy? *FRBSF Economic Review*, **84**, 5–33.

MARTIN, C. and MILAS, C. (2004). Modelling monetary policy: inflation targeting in practice. *Economica*, **71**(281), 209–21.

MARTIN, C. and MILAS, C. (2005). Uncertainty and monetary policy rules in the United States. Keele Economics Research Papers KERP 2005/10, Centre for Economic Research, Keele University.

ORPHANIDES, A. (1998). Monetary policy evaluation with noisy information. Finance and Economics Discussion Series 1998–50, Board of Governors of the Federal Reserve System.

OSBORN, D. R., KIM, D. H. and SENSIER, M. (2005). Nonlinearity in the Fed's monetary policy rule. *Journal of Applied Econometrics*, **20**(5), 621–39.

OWYANG, M. T. and RAMEY, G. (2004). Regime switching and monetary policy measurement. *Journal of Monetary Economics*, **51**(8), 1577–97.

POOLE, W. (2006). The Fed's monetary policy rule. *Federal Reserve Bank of St Louis Review*, **88**(1), 1–12.

RABANAL, P. (2004). Monetary policy rules and the US business cycle: evidence and implications. IMF Working Papers no. 04/164, International Monetary Fund.

RUDEBUSCH, G. D. (2002). Term structure evidence on interest rate smoothing and monetary policy inertia. *Journal of Monetary Economics*, **49**(6), 1161–87.

——— (2006). Monetary policy inertia: fact or fiction? *International Journal of Central Banking*, **2**(4), 85–135.

SIKLOS, P. L. and WOHAR, M. E. (2005). Estimating Taylor-type rules: an unbalanced regression? In T. Fomby and D. Terrell (eds), *Advances in Econometrics*. Amsterdam: Elsevier.

SIMS, C. A. and ZHA, T. (2006). Were there regime switches in US monetary policy? *American Economic Review*, **96**(1), 54–81.

SURICO, P. (2007). The Fed's monetary policy rule and US inflation: the case of asymmetric preferences. *Journal of Economic Dynamics and Control*, **31**(1), 305–24.

———, D'Agostino, A. and Sala, L. (2005). The Fed and the stock market. Paper presented at the conference *Computing in Economics and Finance 2005*, Society for Computational Economics, Washington DC.
Svensson, L. E. O. (1999). Inflation targeting as a monetary policy rule. *Journal of Monetary Economics*, **43**(3), 607–54.
——— (2003). What is wrong with Taylor rules? Using judgment in monetary policy through targeting rules. *Journal of Economic Literature*, **41**(2), 426–77.
——— (2005). Monetary policy with judgment: forecast targeting. *International Journal of Central Banking*, **1**(1), 1–54.
Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie–Rochester Conference Series on Public Policy*, **39**, 195–214.
Taylor, M. and Davradakis, E. (2006). Interest rate setting and inflation targeting: evidence of a nonlinear Taylor rule for the United Kingdom. *Studies in Nonlinear Dynamics & Econometrics*, **10**(4), 1359.
Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208–18.
van Dijk, D., Terasvirta, T. and Franses, P. (2000). Smooth transition autoregressive models—a survey of recent developments. Econometric Institute Report 200, Erasmus University, Rotterdam.
Woodford, M. (2001). The Taylor rule and optimal monetary policy. *American Economic Review*, **91**(2), 232–7.
——— (2003). *Interest and Prices*. Princeton, NJ: Princeton University Press.
Yellen, J. L. (2004). Innovations and issues in monetary policy: panel discussion. *American Economic Review*, **94**(2), 45–7.

Economica

Economica (2011) **78**, 108–132
doi:10.1111/j.1468-0335.2009.00794.x

Schooling and Public Capital in a Model of Endogenous Growth

By Pierre-Richard Agénor

Centre for Growth and Business Cycle Research, School of Social Sciences, University of Manchester

Final version received 22 September 2008.

This paper studies the allocation of public spending between education services and infrastructure investment in an endogenous growth model of a developing economy where public capital in infrastructure affects human capital accumulation. The balanced growth path is derived and the possibility of local indeterminacy is discussed. Dynamics associated with a budget-neutral reallocation of spending from education to infrastructure are studied through numerical simulations. The growth-maximizing share of investment in infrastructure is shown to depend on the goods production technology and the 'productiveness' of infrastructure in the schooling technology. Properties of the welfare-maximizing solution are also discussed.

Introduction

Much of the current international debate on ways to spur growth, reduce poverty, and improve the quality of human life in low-income developing countries has focused on the need for a large increase in public investment in infrastructure (see, for instance, United Nations Millennium Project 2005). In contrast to the early literature on the design of adjustment programs—which often viewed public expenditure only through its impact on fiscal deficits and mostly, therefore, as an instrument of short-run macroeconomic adjustment—the current perspective has emphasized the supply-side effects of public capital and its implications for private capital accumulation and economic growth.

The growth effects of public spending have also received much attention in the analytical literature on endogenous growth. As shown in an influential early contribution by Barro (1990) and much of the subsequent literature spawned by it, public services and capital in infrastructure may promote growth through their effect on the productivity of factors and the rate of return on capital, and the growth-maximizing rates of taxation and public investment are in general positive. Moreover, this literature has also clarified some of the potential trade-offs that may arise, in designing growth-maximizing policies, between investment in infrastructure and other components of public spending—such as subsidies to private capital accumulation, maintenance expenditure and the provision of education or health services. A key implication of this line of work is the importance of distinguishing between productive public spending, which exerts direct supply-side effects, and unproductive spending, which does not contribute (at least, not directly) to growth. Productive spending includes therefore not only investment in infrastructure *per se*, but also possibly current expenditure on education and health—to the extent that they affect directly the stock of human capital and the productivity of labour. Put differently, from a growth perspective, the critical distinction in the government budget is not between current (consumption) and capital (investment) expenditure, but between productive and unproductive spending. In practice, of course, drawing the line between these two components may not be straightforward. For instance, wages and salaries of public servants can be viewed as (partially) productive if a better functioning bureaucracy serves to facilitate private activity. In the same vein, spending on defence and security, or the environment, could be viewed as (partially) productive, because feeling safer or breathing air of better quality may increase productivity. Conceptually, however, the distinction remains fundamental for understanding how public expenditure affects growth.

This paper contributes to the existing literature on productive public spending and growth in several ways. It develops a Lucas-type endogenous growth model of a developing economy with human capital accumulation and external effects associated with public capital in infrastructure, and examines the dynamics of spending shifts as well as the optimal determination of the tax rate and the shares of tax revenue allocated to public infrastructure investment and education services. As, for instance, in Futagami *et al.* (1993), Glomm and Ravikumar (1994), Fisher and Turnovsky (1998), Baier and Glomm (2001), Turnovsky (1997, 2000), Gómez (2004), Yakita (2004) and Chen (2007), public infrastructure is treated as a stock.[1]

A crucial feature of the model is that the production of human capital (an activity in which only the public sector is engaged, as is often the case in poor countries) requires not only public spending on education services but also access to infrastructure capital. As discussed by Brenneman and Kerf (2002), a number of microeconomic studies have documented a positive impact of infrastructure services on educational attainment, possibly through an indirect improvement in health indicators. A better transportation system and a safer road network (particularly in rural areas) help to raise school attendance. Greater access to safe water and sanitation enhance the health of individuals, increasing their ability to learn. Electricity allows more time to study and more opportunities to use electronic equipment and other devices that may improve the learning process. In quantitative terms, the difference that access to infrastructure makes can be sizeable. For instance, in the late 1990s in Nicaragua, 72% of children living in a household with electricity were attending school, compared to only 50% of those living in a household without electricity (see Saghir 2005). As far as I know, this paper is the first to account explicitly for these effects in a model where public infrastructure is treated as a stock.[2]

The paper is organized as follows. The model is presented in Section I. Its balanced growth path, and conditions for saddlepath stability (or local indeterminacy), are derived in Section II. Given the complexity of the model, Section III uses numerical techniques to examine the transitional and long-run effects of a switch in government spending from education to infrastructure, for different values of the parameters characterizing the human capital technology. The key issue here is whether reallocating funds from education to infrastructure can increase the growth rate, given that public infrastructure capital affects the production of human capital. Section IV determines the growth-maximizing shares of government spending on infrastructure investment and education services, and characterizes some features of the welfare-maximizing solution. As shown by Barro (1990), if public infrastructure services derive from *flow* expenditures, the growth-maximizing rate of spending (or, equivalently, taxation in his setting) is equal to the elasticity of output with respect to these services. A similar result obtains when the flow of services is produced by the *stock* of public capital (see Futagami *et al.* 1993), or when it is produced by a stock–flow combination (see Tsoukis and Miller 2003), in the absence of maintenance costs.[3] Here I examine how the growth-maximizing allocation differs from the Barro rule, and more specifically how it depends on the fact that public capital in infrastructure affects the human capital technology. The final section summarizes the main results of the paper and offers some concluding remarks.

I. The Economy

Consider an economy populated by a representative, infinitely-lived household. It produces and consumes a homogenous good, which can be used for consumption or investment. The price of the good is fixed and normalized to unity for simplicity. Population is constant and also normalized to unity. Schooling is mandatory, implying that households cannot choose the allocation of time between education on the one hand, and work and leisure on the other.[4] The government invests in infrastructure (namely, roads, water and sanitation, telecommunications and electricity) and provides education services (such as books and other training materials). It balances its budget continuously, by levying a flat tax on output.

Production

Aggregate output Y is produced with private physical capital K_P, public infrastructure capital K_G, and human capital H, using a Cobb–Douglas technology:[5]

$$(1) \qquad Y = K_G^{\alpha} H^{\beta} K_P^{1-\alpha-\beta},$$

where $\beta \in (0, 1)$. Thus production exhibits constant returns to scale in all factors (which are all augmentable), with diminishing returns with respect to each of them. Evidence supporting the assumption of constant returns to scale in labour, public capital and private capital is provided in a variety of studies, including Otto and Voss (1998) and Song (2002).[6] Note that the aggregate function (1) does *not* imply diminishing returns to scale at the level of the individual firm. Indeed, as discussed by Glomm and Ravikumar (1994) and Agénor (2009), it can be derived by assuming constant returns to scale with respect to private factors at the firm level, and an externality measured in terms of the ratio of the public capital stock to the aggregate private capital stock. In that case, therefore, public capital is non-excludable but partially rival.

Note also that public services in infrastructure are for simplicity taken to be directly proportional to the stock of public capital. Extending the analysis to account for both public capital and public services as separate production inputs (as in Ghosh and Roy 2004, for instance) is straightforward. Similarly, K_P denotes both the stock of private capital and the flow of services that this capital produces.

Household optimization

Abstracting from labour–learning–leisure choices, and assuming no government-provided utility-enhancing services, the representative household maximizes the discounted stream of future utility

$$(2) \qquad \max_C U = \int_0^{\infty} \ln C \exp(-\rho t)\, \mathrm{d}t,$$

where C is aggregate consumption and ρ is the discount rate. As argued by García-Peñalosa and Turnovsky (2005, p. 1052), abstracting from labour–leisure choices is a reasonable assumption in a model designed for low-income countries. Given the low levels of goods consumption in these countries, it is unlikely that much leisure is consumed to begin with.

Assuming that capital does not depreciate, the household's budget constraint is

$$C + \dot{K}_P = (1 - \tau)Y, \tag{3}$$

where $\tau \in (0, 1)$ is the tax rate on output. Implicit in the analysis is the assumption that the rent associated with public capital (an external input in the model) is collected by the household-producer. This could be made explicit by introducing in the household's budget constraint some fixed factor that 'soaks up' the extra profits. An alternative approach, as for instance in Palivos *et al.* (2003), would be to assume in (1) constant returns to scale with respect to H and K_P, and to let the exponents of K_P and K_G add up to unity to obtain balanced growth paths, that is, $Y = K_G^{\alpha} H^{\alpha} K_P^{1-\alpha}$, with $\alpha \in (0, 1)$. This, however, would imply implausibly high elasticities with respect to K_G (of the order of 0.6), given the empirical evidence discussed later.

The household takes public policies as given when solving its optimization problem. Let $s \equiv (1 - \tau)(1 - \alpha - \beta)$, so that $s \in (0, 1)$, and let λ denote the shadow price of private capital; maximizing (2) subject to (1) and (3) yields the familiar first-order condition

$$\frac{\dot{C}}{C} = s\left(\frac{K_G}{K_P}\right)^{\alpha}\left(\frac{H}{K_P}\right)^{\beta} - \rho, \tag{4}$$

together with (3) and the transversality condition

$$\lim_{t \to \infty} \lambda K_P(t) \exp(-\rho t) = 0. \tag{5}$$

Production of human capital

Human capital is produced only by the public sector. Production is specified as a Cobb–Douglas function of government spending on education G_E, public capital in infrastructure, and the existing stock of human capital. Thus infrastructure is essential to the production of human capital; the view here is that schools cannot operate without electricity and without water and sanitation. This should not, of course, be taken too literally; schools may (and actually do, especially in rural areas) operate with sunlight only, and students may be required to fulfil their own drinking water needs. But the efficiency loss is also significant. Making infrastructure an essential input for the production of knowledge helps to highlight the importance of infrastructure for education outcomes, as documented by the microeconomic studies referred to in the Introduction, and its implications for long-run growth.

Assuming no depreciation of skills, and abstracting from the (fixed) amount of time that individuals must devote to attending school, the accumulation of human capital is determined by

$$\dot{H} = G_E^{\kappa_1} K_G^{\kappa_2} H^{1-\kappa_1-\kappa_2}, \tag{6}$$

where $\kappa_1, \kappa_2 \in (0, 1)$ and $\kappa_1 + \kappa_2 < 1$. The education technology exhibits therefore constant returns to scale in all three inputs. Note that here public capital is also taken to be non-excludable; roads, for instance, can be used equally by individuals to get to school, or by firms to transport raw materials and final goods.

Equation (6) can be rewritten as

$$\frac{\dot{H}}{H} = \left(\frac{G_E}{H}\right)^{\kappa_1}\left(\frac{K_G}{H}\right)^{\kappa_2}. \tag{7}$$

Government

The government collects a proportional tax on output. It invests in infrastructure capital, G_I, and provides education services that are used in the production of human capital. It also spends on unproductive items (such as defence and protection of the environment), in quantity G_U. As noted in the Introduction, in practice the classification of public expenditure into 'productive' and 'unproductive' is somewhat arbitrary; education services, for instance, may include not only spending on goods (books, pencils, etc.) but also teachers' salaries—which may have a positive effect on their level of effort, and therefore on the quality of human capital that they contribute to create. The main goal of the present study, however, is to examine analytically the potential trade-offs in the allocation of productive components of public spending only (a key policy issue for low- and high-income countries alike), so a precise, practical definition is not needed.

The government budget constraint is thus given by

(8) $$G_E + G_I + G_U = \tau Y.$$

Both components of public spending are specified as constant fractions of tax revenues, υ_E and υ_I, with $\upsilon_E, \upsilon_I \in (0,1)$:

(9) $$G_h = \upsilon_h \tau Y, \quad h = E, I, U.$$

Using these definitions, the government budget constraint can be rewritten as

(10) $$\upsilon_E + \upsilon_I + \upsilon_U = 1.$$

Assuming also no physical depreciation, the stock of public capital in infrastructure evolves over time according to[7]

(11) $$\dot{K}_G = G_I.$$

Goods market equilibrium

Finally, note that from (3), (8) and (11), the economy's consolidated budget constraint (or, equivalently, the goods market equilibrium condition) is

(12) $$C + \dot{K}_P + \dot{K}_G + G_E + G_U = Y.$$

II. Balanced Growth Path

The condensed dynamic form of the model is derived in the Appendix. Let $c = C/K_P$, $k_G = K_G/K_P$ and $h = H/K_P$; the dynamic system boils down to

(13) $$\frac{\dot{c}}{c} = \Omega k_G^{\alpha} h^{\beta} + c - \rho,$$

(14) $$\frac{\dot{h}}{h} = (\tau \upsilon_E)^{\kappa_1} k_G^{\theta_1} h^{-\theta_2} - (1-\tau) k_G^{\alpha} h^{\beta} + c,$$

(15) $$\frac{\dot{k}_G}{k_G} = \left[\frac{\tau \upsilon_I}{k_G} - (1-\tau)\right] k_G^{\alpha} h^{\beta} + c,$$

where $\theta_1 \equiv \alpha\kappa_1 + \kappa_2$, $\theta_2 \equiv (1-\beta)\kappa_1 + \kappa_2$ and $\Omega \equiv s - (1-\tau) = -(1-\tau)(\alpha+\beta) < 0$. The restriction $\kappa_1 + \kappa_2 < 1$ implies that $\theta_1, \theta_2 \in (0,1)$.

These equations, together with the initial conditions $h_0 = H_0/K_{G,0} > 0$ and $k_{G,0} = K_{G,0}/K_{P,0} > 0$, and the transversality condition (5), rewritten (using $\lambda = 1/C$) as

$$\lim_{t\to\infty} c_t^{-1}\exp(-\rho t) = 0, \tag{16}$$

characterize the dynamics of the economy. This yields therefore the following definition:

Definition 1. A competitive equilibrium corresponds to a set of functions $\{c, h, k_G\}_{t=0}^{\infty}$ and a constant tax rate τ such that individuals maximize utility, firms maximize profits, markets clear, and the government budget is balanced. Thus equations (13), (14) and (15), the transversality condition (16) and the budget constraint (10) must all be satisfied at all times.

Based on this, a balanced growth equilibrium can also be defined:

Definition 2. The balanced growth equilibrium is a competitive equilibrium in which consumption and the stock of human capital, as well the stocks of public and private capital, all grow at the same constant rate, that is, $\dot{C}/C = \dot{H}/H = \dot{K}_P/K_P = \dot{K}_G/K_G = \gamma$.

From equations (A2), (A4) and (A5) in the Appendix, together with (10), the constant steady-state growth rate γ—which is also the rate of growth of output, given the assumption of constant returns to scale—is given by the equivalent forms[8]

$$\gamma = s\tilde{k}_G^{\alpha}\tilde{h}^{\beta} - \rho, \tag{17}$$

$$\gamma = [\tau(1-\upsilon_I-\upsilon_U)]^{\kappa_1}\tilde{k}_G^{\theta_1}\tilde{h}^{-\theta_2}, \tag{18}$$

$$\gamma = \tau\upsilon_I\tilde{k}_G^{\alpha-1}\tilde{h}^{\beta}, \tag{19}$$

where $\tilde{x}$ denotes the stationary value of x.

Consider now the following definition of a non-degenerate balanced growth equilibrium path:

Definition 3. A non-degenerate balanced growth equilibrium must satisfy the following conditions.

(a) The growth rate is strictly positive.
(b) The utility integral in (2) converges, $\int_0^{\infty} \ln C\exp(-\rho t)\,\mathrm{d}t < \infty$.
(c) The transversality condition (16) is satisfied.
(d) The steady-state value of the consumption–private capital ratio is strictly positive.

From (17), it can be verified that condition (a) holds as long as the rate of time preference is not too high. It can also be readily verified that, given the specific functional form adopted here, the utility integral converges, so that condition (b) is satisfied. Because consumption and the stock of private capital grow at the same constant rate along any equilibrium path with $\gamma > 0$, the ratio $c = C/K_P$ is constant, which implies that condition (c) is satisfied. Finally, setting $\dot{c} = 0$ in equation (13) yields

$$(20) \qquad \tilde{c} = \rho - \Omega \tilde{k}_G^{\alpha} \tilde{h}^{\beta},$$

which implies, because $\Omega < 0$, that the steady-state value of the consumption–private capital ratio is strictly positive. These results can therefore be summarized in the following proposition:

Proposition 1. For a given tax rate and spending shares, a unique, non-degenerate equilibrium with a strictly positive balanced growth rate and a stationary consumption–capital ratio $\tilde{c} > 0$ exists if the discount rate is not too high.

To investigate the dynamics in the vicinity of the steady state, equations (17) to (19) can be linearized to give

$$(21) \qquad \begin{bmatrix} \dot{c} \\ \dot{h} \\ \dot{k}_G \end{bmatrix} = \begin{bmatrix} \tilde{c} & a_{12} & a_{13} \\ \tilde{h} & a_{22} & a_{23} \\ \tilde{k}_G & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} c - \tilde{c} \\ h - \tilde{h} \\ k_G - \tilde{k}_G \end{bmatrix},$$

where the a_{ij} are given by

$$a_{12} = \beta \tilde{c} \Omega \tilde{k}_G^{\alpha} \tilde{h}^{\beta - 1} < 0,$$

$$a_{13} = \alpha \tilde{c} \Omega \tilde{k}_G^{\alpha - 1} \tilde{h}^{\beta} < 0,$$

$$a_{22} = -\theta_2 (\tau \upsilon_E)^{\kappa_1} \tilde{k}_G^{\theta_1} \tilde{h}^{-\theta_2} - \beta (1 - \tau) \tilde{k}_G^{\alpha} \tilde{h}^{\beta} < 0,$$

$$a_{23} = \theta_1 (\tau \upsilon_E)^{\kappa_1} \tilde{k}_G^{\theta_1 - 1} \tilde{h}^{1 - \theta_2} - \alpha (1 - \tau) \tilde{k}_G^{\alpha - 1} \tilde{h}^{1 + \beta},$$

$$a_{32} = \beta \tilde{k}_G [\tau \upsilon_I \tilde{k}_G^{-1} - (1 - \tau)] \tilde{k}_G^{\alpha} \tilde{h}^{\beta - 1},$$

$$a_{33} = \alpha [\tau \upsilon_I \tilde{k}_G^{-1} - (1 - \tau)] \tilde{k}_G^{\alpha} \tilde{h}^{\beta} - \tau \upsilon_I \tilde{k}_G^{\alpha - 1} \tilde{h}^{\beta}.$$

To establish the sign of a_{23}, a_{32} and a_{33}, note that from the alternative definitions of the steady-state growth rate γ given in the text, as well as the fact that, from (A1) in the Appendix, $\gamma = (1 - \tau) \tilde{k}_G^{\alpha} \tilde{h}^{\beta} - \tilde{c}$, we have

$$a_{23} = \theta_1 \tilde{h} \gamma / \tilde{k}_G - \alpha \tilde{h} (\gamma + \tilde{c}) / \tilde{k}_G,$$

$$a_{32} = \beta \tilde{k}_G \gamma / \tilde{h} - \beta \tilde{k}_G (\gamma + \tilde{c}) / \tilde{h} = -\beta \tilde{k}_G \tilde{c} / \tilde{h} < 0,$$

$$a_{33} = \alpha [\gamma - (\gamma + \tilde{c})] - \gamma = -\alpha \tilde{c} - \gamma < 0.$$

The sign of a_{23} remains in general ambiguous. Because $\tilde{c} > 0$, a necessary (although not sufficient) condition for $a_{23} > 0$ is $\theta_1 > \alpha$. Given that $\theta_1 \equiv \alpha \kappa_1 + \kappa_2$, this condition is equivalent to $\alpha(1 - \kappa_1) < \kappa_2$, which implies that the effect of infrastructure in the human capital accumulation technology must be sufficiently strong.[9]

Among the three variables whose dynamics drive the system, c can adjust instantaneously whereas h and k_G are predetermined. Therefore the equilibrium path in the neighbourhood of the unique steady-growth equilibrium is locally determinate if the Jacobian matrix in (21), denoted **J**, has only two eigenvalues with negative real part, that is, one unstable (positive) root. If all three eigenvalues have negative real parts, then the equilibrium is locally indeterminate.

Let $\Pi(\lambda) = -\lambda^3 + b_2\lambda^2 - b_1\lambda + b_0$ denote the characteristic polynomial of $\mathbf{J}$, where $b_2 = \text{tr}\,\mathbf{J}$, $b_0 = \det\,\mathbf{J}$ and b_1 is the sum of the determinants of the leading principal minors of order 2 of matrix $\mathbf{J}$:

$$b_1 = \begin{vmatrix} \tilde{c} & a_{12} \\ \tilde{h} & a_{22} \end{vmatrix} + \begin{vmatrix} \tilde{c} & a_{13} \\ \tilde{k}_G & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}.$$

The Routh–Hurwitz conditions for stability imply the following restrictions:

Conditions 1–3. $b_1 < 0$, $b_2 < 0$, $0 < b_0 < b_2 b_1$.

The condition $b_0 > 0$excludes one or three negative roots (but not two negative roots), whereas $b_2 < 0$ ensures at least one negative root. However, it cannot be guaranteed that these conditions always hold.[10] Consider, for instance, the condition $b_2 < 0$. From the definition of a_{22}, it can be established that $a_{22} = -\theta_2\gamma - \beta(\gamma + \tilde{c})$. Combining this result with the above definition of a_{33} yields

$$\text{tr}\,\mathbf{J} = (1 - \alpha - \beta)\tilde{c} - (1 + \beta + \theta_2)\gamma,$$

which is in general ambiguous in sign.[11] On the basis of these results, the following proposition can be stated:

Proposition 2. Under conditions 1–3, the economy exhibits saddlepath stability in the neighbourhood of the balanced growth path. In general, however, the equilibrium path may be locally indeterminate.

Local indeterminacy of the equilibrium means that expectations determine the equilibrium path, given that in that case the initial level of consumption can be freely chosen. Although it is difficult to prove it analytically through direct calculations, it is intuitively clear that the source of local indeterminacy in the model is fundamentally related to the sector externalities associated with public infrastructure, as measured by α and κ_2.[12] That this must be so can be inferred from the fact that if $\alpha = \kappa_2 = 0$, then $a_{13} = a_{23} = 0$ and the system becomes recursive; the dynamics of c and h become independent of k_G. Saddlepath stability then requires $\tilde{c}a_{22} - \tilde{h}a_{12} < 0$. Because now $\Omega = -(1 - \tau)\beta$, the expression on the left-hand side is $-\tilde{c}\theta_2(\tau v_E)^{\kappa_1}\tilde{k}_G^{\theta_1}\tilde{h}^{-\theta_2}$, which is indeed negative. Indeterminacy (which requires two negative roots) cannot occur, because the eigenvalues are of opposite sign.

With $\alpha > 0$ and $\kappa_2 > 0$, however, it is possible for the the balanced growth equilibrium to be locally indeterminate—even with $\kappa_1 = 0$. Intuitively, if agents expect the government to provide additional infrastructure capital, to be used in the production of goods either directly or indirectly (to the extent that it increases the rate of accumulation, and availability, of human capital), they will save and invest more because they would then expect the future return on private capital to increase. As a result, output will rise, and so will tax revenues—which leads indeed to higher public investment in infrastructure. Agents' initial expectations are thus fulfilled, leading thereby to indeterminacy of equilibrium paths.

Moreover, in the present setting, the possibility of local indeterminacy depends critically on the fact that the production of human capital requires also public infrastructure (that is, $\kappa_2 > 0$). With $\kappa_2 = 0$, local indeterminacy would require either implausibly high values of α and κ_1 or the existence of congestion effects, as for instance in Chen and Lee (2007). Nevertheless, in the numerical simulations performed in the next section, the analysis is restricted to a range of (empirically plausible) parameter values for which the system is indeed saddlepath stable.

III. Calibration and Policy Experiments

Because the complexity of the model precludes an analytical exploration of its transitional dynamics, in this section I resort to numerical techniques to examine the short- and long-run effects of a budget-neutral switch in government spending from education to infrastructure, for different values of the parameters characterizing the human capital technology.

I will focus in what follows on the parameters κ_1 and κ_2, and address the following question: given a set of plausible alternative values for the parameters κ_1 and κ_2, is a budget-neutral reallocation of government expenditure from education services to public investment in infrastructure conducive to higher growth? As noted in the Introduction, this is an important issue from the practical perspective of a low-income country which must decide on how best to allocate scarce resources to maximize their impact on growth and reduce poverty.

Calibration

Given that consumption is a forward-looking variable, the numerical solution procedure that I use is the 'extended path' method of Fair and Taylor (1983). This procedure is quite convenient (once a discrete-time approximation of the model is written) because it allows one to solve perfect foresight models in their nonlinear form, through an iterative process.[13] This is important, given recent evidence suggesting that linearization of growth models with productive public spending may entail significant errors (see Atolia *et al.* 2008). The terminal condition imposed on consumption (the only forward-looking variable in the model) is that its growth rate at the terminal horizon ($t + 40$ periods here) must be equal to the growth of the private capital stock, given the condition that $c = C/K_P$ must be constant along the balanced growth path. I discuss next the calibration procedure and the baseline solution, and then examine the simulation results themselves.

The numerical values assigned to the variables and parameters of the system dwell as much as possible on the existing empirical literature and are chosen to roughly match some well-documented facts about low-income developing countries.

I consider an economy with a relatively low stock of public capital to begin with. Specifically, the public capital stock is set so that the initial public capital–output ratio is 0.6. This ratio is quite low by industrial-country standards but it is consistent with the average estimate of the net public capital stock obtained by Arestoff and Hurlin (2005b, Table 3) for a group of developing countries.[14] The private capital stock is set so that the initial private capital–output ratio is 1.4, implying that the private–public capital ratio is about 2.3, or equivalently an initial value of k_G equal to 0.4. Thus, of the two components of physical capital, public capital is the relatively scarce factor; this is consistent with the view (shared by many observers) that lack of public infrastructure is a major impediment to growth (and private capital accumulation) in poor countries. More specifically it is consistent with the evidence for sub-Saharan Africa, which suggests that the region has some of the lowest rates of access to infrastructure, measured in terms of paved roads, safe water and improved sanitation, telecommunications and electricity (see World Bank 2006). The initial value of k_G chosen here is also similar to the value selected by Atolia *et al.* (2008, p. 18) when the elasticity of intertemporal substitution in their model is set equal to unity.

The elasticities of production of goods with respect to public capital and human capital, α and β respectively, are set equal to 0.15 and 0.45. The value of α used here corresponds to the one estimated by Easterly and Rebelo (1993) and used by Rioja (2005). Cerra *et al.* (2008) also use a value of 0.15 for the elasticity of non-traded output with respect to government spending in their simulations. In the same vein, Bose *et al.* (2007, Table 3), in a growth regression for 30 developing countries that explicitly accounts for the government budget constraint, found a coefficient of 0.15 for public capital expenditure. Baier and Glomm (2001), Rioja and Glomm (2003) and Chen (2003, 2007) use instead an estimate of α of 0.1, which is close to the figure of 0.11 estimated by Hulten (1996). By comparison, Esfahani and Ramírez (2003, Table 4) found estimates of the elasticities of per capita GDP growth ranging from 0.08 to 0.16, when infrastructure capital is measured as the number of telephone lines or power generation capacity, whereas Canning (1999) estimates an elasticity of output per worker with respect to infrastructure (as measured by the number of telephone lines) equal on average to 0.14 for his full sample, and close to 0.26 for higher-income countries. For their part, Arestoff and Hurlin (2005b, Tables 2 and 7) found elasticities of output per worker ranging from 0.05 to 0.19 when infrastructure stocks are used, and from 0.04 to 0.22 when estimates of public capital stocks are used, in the absence of threshold effects. Thus the estimate used here for α is consistent with the upper range of the values estimated by Esfahani and Ramírez, and Arestoff and Hurlin, as well as the lower range of Canning's results.

The estimates of α and β used here imply a share of private capital in output equal to 0.4. Ortigueira (1998, p. 337) and Rivas (2003, Table 1) also use a share of 0.4, whereas Rioja and Glomm (2003, Table 2) use an estimate of 0.45. In their empirical estimates for a panel of 52 countries covering the period 1965–95, Cole and Neumayer (2006, p. 925) find a value of 0.37, which is close to the value used here.

Consider now the human capital technology. The elasticities with respect to government spending on education services and public capital in infrastructure, κ_1 and κ_2 respectively, are set equal to 0.2 and 0.1 in the base case. The first estimate is twice the size of the estimate used by Rioja (2005) and the econometric estimate obtained by Blankenau *et al.* (2007) for their full sample. However, it is consistent with the parameter value used by Chen (2005). The higher estimate used here is probably quite relevant for low-income countries, where education (at least at the primary and secondary levels) is to a very large extent publicly provided.[15] An appropriate value of κ_2 is more difficult to pin down, because the empirical evidence is mostly microeconomic in nature (see Brenneman and Kerf 2002). At the same time, as noted earlier, assessing the impact of infrastructure on education and growth is a key purpose of the model. Accordingly, I chose a relatively low initial value of $\kappa_2 = 0.1$ and performed sensitivity analysis along the lines discussed below.

The initial stock of human capital, H, is calibrated in such a way that it gives an initial private physical capital–human capital ratio of 0.5 (that is, an initial value of h equal to 2), a public capital–human capital ratio of 0.21, and thus an overall physical capital–human capital ratio of 0.71. These ratios (together with the capital–output ratios mentioned earlier) capture fairly well the view that the country considered is poor and endowed with a relatively abundant supply of labour, while facing at the same time a relative scarcity of physical (particularly public) capital.

The rate of time preference, ρ, is set at 4%, a fairly conventional choice in this literature. This leads to a discount factor of approximately 0.96 (see, for instance, Canton 2001, Table 1, and Ghosh and Roy 2004, Tables 1 and 2). Private consumption, C, is set so that it represents about 85% of initial output. This value is quite sensible for many low-income countries, where the low level of income, to begin with, constrains the ability of households to allocate resources to savings and investment (see Agénor and Montiel 2008). This is also consistent with current estimates of private savings rates for sub-Saharan Africa, which are of the order of 9–10%.

I assume that neither the tax rate nor the spending shares are chosen optimally in the initial equilibrium. This is a reasonable assumption for numerical exercises that are meant to capture the reality of fiscal policy-making in developing countries—especially the low-income ones, where tax systems are subject to large inefficiencies and expenditure is often poorly allocated, due to pervasive corruption and a shortage of qualified personnel in public administration. Specifically, regarding fiscal variables, the tax rate on output (which is also the share of total government spending in output), τ, is set at 0.2. This value is in line with actual ratios for many low-income countries, where taxation—which is essentially indirect in nature—provides a more limited source of revenue than in higher-income countries (see Bird and Zolt 2005). It is also similar to the value used by Chen and Lee (2007, p. 2501). The initial shares of government spending on infrastructure services and education services, υ_I and υ_E respectively, are set at 0.3 and 0.2, implying from the budget constraint (10) that the share of unproductive spending is given by $\upsilon_U = 1 - \upsilon_I - \upsilon_E = 0.5$. This value is quite reasonable, given that it represents in practice the share of a variety of government outlays in total tax revenue—not only wages and salaries but also other current spending (including transfers to households and subsidies, military and police expenditure, etc.), which were not explicitly accounted for in the model. Multiplying these shares by the tax rate implies therefore that spending on infrastructure investment represents 6.0% of output in the base period, whereas spending on education services amounts to 4.0% of output. By comparison, Rioja (2005, p. 6) reports averages for υ_I and υ_E of 7.3% and 2.9% of GDP, respectively, for a group of 9 upper-income Latin American countries, whereas Rioja and Glomm (2003, Table 1) report shares of 3.05% and 3.13% for a larger group of 17 Latin American countries. However, these studies also indicate that there is considerable heterogeneity across countries. This is also what one would expect for low-income countries, based on the data on capital expenditure compiled by Arestoff and Hurlin (2005a). The estimates used here can be viewed as representing the 'intermediate' case of a government committed to allocating half of its resources to productive uses, physical and human capital accumulation.

Calibration of the model around these initial values and parameters (which involves also determining appropriate multiplicative constants in the production functions for goods and human capital) produces the baseline solution. Given the values described above, the initial steady-state growth rate is equal to 3.4%. Given a typical rate of population growth in the range 2.7–3.0% per annum in low-income countries, this gives an empirically plausible per capita growth rate in the range 0.4–0.7%. Persistent per capita growth at relatively low rates is one of the key feature of the 'poverty trap' that many poor countries are caught in.

In the results reported next, three alternative values are chosen for the parameters characterizing the human capital technology, κ_1 and κ_2: the base case referred to earlier, with $\kappa_1 = 0.2$ and $\kappa_2 = 0.1$, as well as two alternative scenarios—a 'high infrastructure' case, where $\kappa_1 = 0.2$ and $\kappa_2 = 0.2$, and a 'high education services' case, where $\kappa_1 = 0.3$ and $\kappa_2 = 0.1$. A new baseline is, of course, calculated for each set of parameters.

Steady-state effects

Consider first the long-run effects of a budget-neutral switch in government spending from education to infrastructure. Intuitive reasoning suggests that such a policy shift would yield ambiguous effects on the economy's growth rate. The more 'productive' public capital in infrastructure is in the production of goods (relative to how productive the stock of human capital is) and the production of human capital (relative to how productive government spending on education services is), the more likely it is that the growth rate will increase. Put differently, the growth effect should depend positively on the ratios α/β and κ_2/κ_1.

More formally, equation (17) yields

$$\frac{d\gamma}{d\upsilon_I} = \gamma\left[\alpha\left(\frac{d\tilde{k}_G}{d\upsilon_I}\right)\tilde{k}_G^{-1} + \beta\left(\frac{d\tilde{h}}{d\upsilon_I}\right)\tilde{h}^{-1}\right], \tag{22}$$

whereas equation (18), with $d\upsilon_I = -d\upsilon_E$ (given that $d\tau = 0$), gives an alternative expression that must also hold:

$$\frac{d\gamma}{d\upsilon_I} = \gamma\left[-\frac{\kappa_1}{\upsilon_E} + \frac{\theta_1}{\tilde{k}_G}\left(\frac{d\tilde{k}_G}{d\upsilon_I}\right) - \frac{\theta_2}{\tilde{h}}\left(\frac{d\tilde{h}}{d\upsilon_I}\right)\right]. \tag{23}$$

These two results show that the effect on the growth rate is in general ambiguous; as discussed in the Appendix, it depends on the effect of the shock on $d\tilde{k}_G/d\upsilon_I$ and $d\tilde{h}/d\upsilon_I$, which in turn depends on κ_1 and κ_2. Specifically, as shown in the Appendix, $d\tilde{k}_G/d\upsilon_I > 0$, whereas $d\tilde{h}/d\upsilon_I \lesseqgtr 0$; an increase in υ_I always raises the steady-state public–private capital ratio, but has an ambiguous effect on the human capital–private capital ratio. Moreover, the ambiguous nature of the latter effect persists even with $\kappa_2 = 0$, in which case $\theta_2 \equiv (1-\beta)\kappa_1$. The intuitive reason is that regardless of whether or not public capital affects the human capital technology, a trade-off exists between spending components as long as both types of outlays are productive. Indeed, as also shown in the Appendix, when $\kappa_1 = 0$, in which case $\theta_1 = \theta_2 \equiv \kappa_2$, $d\tilde{h}/d\upsilon_I > 0$. Expression (22) yields therefore $d\gamma/d\upsilon_I > 0$. At the same time, however, it can also be verified that in the general case a higher value of κ_2 makes it more likely (everything else equal) that $d\tilde{h}/d\upsilon_I > 0$. From (22) and the results in the Appendix, it is clear that the higher the value of α relative to β, the larger the positive effect of $d\tilde{k}_G/d\upsilon_I$, and the more likely it is that the spending shift will increase growth. The numerical results presented next corroborate these findings.

Transitional dynamics

Consider now the transitional dynamics associated with a budget-neutral shift in the composition of government spending that takes the form of an increase in the share of investment in infrastructure, from 0.3 to 0.4, coupled with a reduction in the share of spending on education services, from 0.2 to 0.1. In proportion of output, this shift represents an increase in spending on infrastructure investment from 6% to 8%, with a concomitant reduction in spending on education from 4% to 2%. The same experiment is run for the three cases described earlier regarding the values of the parameters characterizing the human capital technology. By contrasting these three cases, the simulations provide a sense of the importance of the inclusion of infrastructure capital in the education technology for the transitional dynamics and the steady-state behaviour of the economy.

The effects of the expenditure switch on the consumption–private capital ratio, the ratio of human capital to private physical capital, and the ratio of public to private capital are shown in Figure 1. The three lines correspond to the three sets of values for κ_1 and κ_2 given earlier, with the continuous plot corresponding to the base case of $\kappa_1 = 0.2$ and $\kappa_2 = 0.1$.

On impact, neither stock of physical capital (public or private) can change, and neither can the stock of human capital. As a result, output and tax revenues are also constant. But in all three cases, consumption (and thus the ratio of consumption to private capital) jumps upward. In subsequent periods, the shift in government spending translates into a lower stock of human capital, whereas the stock of public capital in infrastructure increases. The latter effect raises the marginal productivity of private capital, whereas the former reduces it. Although the elasticity of output with respect to public infrastructure is substantially lower than its elasticity with respect to human capital (0.15 against 0.45, respectively), the net effect is an increase in the marginal productivity of capital, thereby raising the incentive to save and invest. At the same time, however, the positive income effect (associated with the higher capital stock and output) tends to increase consumption.

The combination of increasing stocks of private and public capital translates also, during a first phase, into higher output, despite a falling stock of human capital. The adjustment process, for all three sets of values for (κ_1, κ_2), is nonmonotonic: at first, the consumption–private capital ratio and the ratio of public to private capital increase, whereas the ratio of human capital to private physical capital falls. As could be expected, the drop in the stock of human capital is lower in the 'high infrastructure' case (where $\kappa_1 = 0.2$ and $\kappa_2 = 0.2$), whereas it is more pronounced in the 'high education services' case (where $\kappa_1 = 0.3$ and $\kappa_2 = 0.1$). The initial increase in the consumption–private capital ratio is also considerably more significant in the 'high education services' case. The reason is that the future increase in the marginal product of capital is less pronounced in that case; the incentive to shift consumption forward (and save more today) is therefore mitigated and the income effect largely dominates. After reaching a peak, the consumption–private capital and public–private capital ratios start falling, whereas the human capital–private physical capital ratio starts increasing. Essentially, because of an initial phase of higher private capital accumulation, and lower accumulation of human capital, the marginal product of capital begins to fall—and so do output and consumption. In fact, in the 'base' and 'high infrastructure' scenarios, adjustment of the consumption–private capital ratio is oscillatory: after the third period, the ratio drops below its baseline value and recovers only gradually. This drop is more pronounced in the 'high infrastructure' case, because the adverse effect of the spending shift on the rate of accumulation of human capital is less dramatic, implying that the marginal product of capital falls by less, and therefore that private capital accumulation is higher.

In the long run, the shock has no discernible effect on the consumption–capital ratio in the 'base' and 'high infrastructure' cases, and only a small positive effect in the 'high education services' case. By contrast, the human capital–private capital ratio converges to a permanently lower value relative to the baseline scenario (by about -0.6 percentage points in the 'base' and 'high infrastructure' cases, and -1.1 percentage points in the

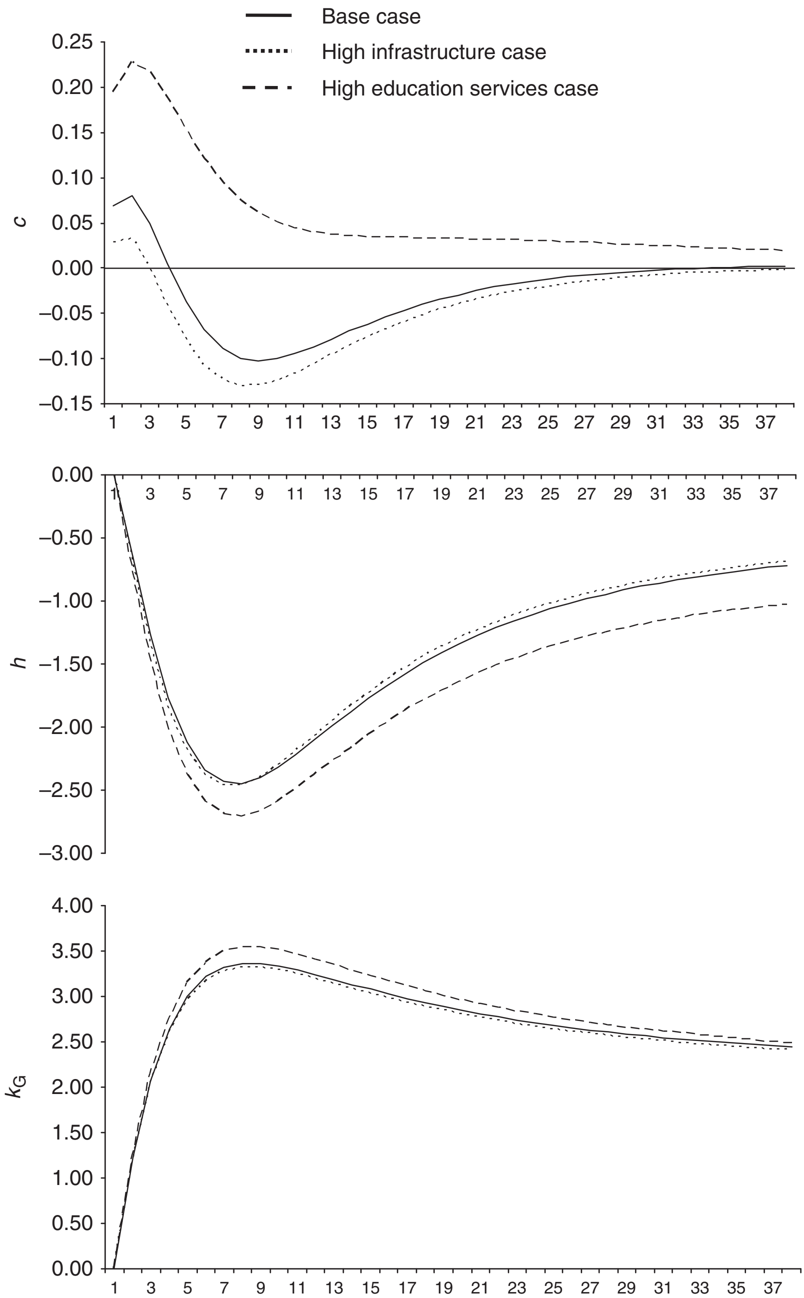

FIGURE 1. Shift in spending from education to infrastructure (absolute deviations from baseline, in per cent).

'high education services' case), whereas the public–private capital ratio converges to a permanently higher value (about 2.4 percentage points in all three cases). For the public–private capital ratio, the long-run change is not significantly affected by the choice of the parameters (κ_1, κ_2), in contrast to the other ratios. But in all three scenarios, the effect on the economy's growth rate is negligible in the long run, with a slightly higher effect in the 'high infrastructure' case relative to the other two cases. The key reason of course is that, as shown in the figure, k_G and h evolve in opposite directions; their impacts on the growth rate therefore offset each other to some extent. Qualitatively, the behaviour of the growth rate tends to follow the same inverted U-shape pattern as the public–private capital ratio shown in the bottom panel of Figure 1.

The foregoing discussion suggests therefore that, given a plausible calibration, a budget-neutral shift in spending from education to infrastructure has limited effects on long-run growth. In a sense, increasing spending on one component while decreasing the other simultaneously tends to have largely offsetting effects, given the overall structure of the model and the calibrated parameters. At the same time, however, simulation results do support the intuition that long-run effects on growth of a reallocation of spending toward infrastructure investment are higher when infrastructure has a larger impact on the human capital technology. This implies that relatively large shifts in spending would be required for growth to fall. This is in contrast with the results of Rioja (2005), who showed—using an OLG model and a schooling technology that does not depend on infrastructure—that a relatively small shift in spending of the type considered here unambiguously lowers steady-state growth.[16]

IV. Optimal Policies

The potential trade-off between spending on infrastructure investment and spending on education services has been well recognized in the recent literature on 'flow' models of government expenditure, in which infrastructure services affect the economy's ability to produce human capital (see Agénor 2008a, b). This section discusses how the presence of the stock of public infrastructure assets in the human capital technology affects the determination of growth- and welfare-maximizing fiscal policies.

Growth-maximizing policy

I first examine the determination of the optimal tax rate, holding expenditure shares (that is, the composition of tax revenues) constant. Let $\varepsilon_{\tilde{y}/z}$ denote the elasticity of the steady-state value of γ with respect to z; setting $\partial\gamma/\partial\tau = 0$ in equations (17), (18) and (19) yields the following system of equations in τ, $\varepsilon_{\tilde{k}_G/\tau}$ and $\varepsilon_{\tilde{h}/\tau}$:

$$(24) \qquad \tau^*/(1-\tau^*) = \alpha\varepsilon_{\tilde{k}_G/\tau} + \beta\varepsilon_{\tilde{h}/\tau},$$

$$(25) \qquad \theta_1\varepsilon_{\tilde{k}_G/\tau} - \theta_2\varepsilon_{\tilde{h}/\tau} = -\kappa_1,$$

$$(26) \qquad (1-\alpha)\varepsilon_{\tilde{k}_G/\tau} - \beta\varepsilon_{\tilde{h}/\tau} = 1.$$

The last two equations can be solved simultaneously for $\varepsilon_{\tilde{k}_G/\tau}$ and $\varepsilon_{\tilde{h}/\tau}$, yielding:

$$\varepsilon_{\tilde{k}_G/\tau} = \frac{\beta\kappa_1 + \theta_2}{\Delta}, \quad \varepsilon_{\tilde{h}/\tau} = \frac{\theta_1 + (1-\alpha)\kappa_1}{\Delta},$$

where $\Delta \equiv -\beta\theta_1 + (1-\alpha)\theta_2 = (1-\alpha-\beta)(\kappa_1+\kappa_2) > 0$.

Simplifying these expressions yields

$$\varepsilon_{\tilde{k}_G/\tau} = \varepsilon_{\tilde{h}/\tau} = \frac{1}{1-\alpha-\beta},$$

which indicates that the elasticities of the steady-state values of the public–private capital ratio and the human capital–capital ratio with respect to the tax rate must be the same at the optimum. These results can be substituted in equation (24) to lead to the following proposition:

Proposition 3. With spending shares on education services and infrastructure investment held constant, the growth-maximizing value of the tax rate is $\tau^* = \alpha + \beta \in (0, 1)$.

This result confirms the one derived in a previous contribution (see Agénor 2008a), in which public infrastructure was treated as a flow. In a sense, therefore, both results confirm those obtained in Barro-type models where human capital is absent. In the present setting, where human capital is also productive, the optimal tax rate is higher than α by a factor β.[17]

Next, consider the optimal allocation of spending between education services υ_E and infrastructure investment υ_I, for a given tax rate τ, and for $\upsilon_U = 0$ for simplicity. As a result of the budget constraint (10), only one of these shares can be chosen independently. In what follows, it will be assumed that the government determines optimally υ_I, with υ_E determined through (10).

Setting $\partial\gamma/\partial\upsilon_I = 0$ in equations (17), (18) and (19), and using (10), yields the following system of equations in τ, $\varepsilon_{\tilde{k}_G/\upsilon_I}$ and $\varepsilon_{\tilde{h}/\upsilon_I}$:

$$\alpha\varepsilon_{\tilde{k}_G/\upsilon_I} + \beta\varepsilon_{\tilde{h}/\upsilon_I} = 0, \tag{27}$$

$$\theta_1\varepsilon_{\tilde{k}_G/\upsilon_I} - \theta_2\varepsilon_{\tilde{h}/\upsilon_I} = \kappa_1\upsilon_I^*/(1-\upsilon_I^*), \tag{28}$$

$$(1-\alpha)\varepsilon_{\tilde{k}_G/\upsilon_I} - \beta\varepsilon_{\tilde{h}/\upsilon_I} = 1, \tag{29}$$

where θ_1 and θ_2 are as defined before. Combining (27) and (29) yields

$$\varepsilon_{\tilde{k}_G/\upsilon_I} = 1,$$

which indicates that the elasticity of the steady-state value of the public–private capital ratio with respect to the share of spending on infrastructure must be equal to unity at the optimum. From (27), this implies that $\varepsilon_{\tilde{h}/\upsilon_I} = -\alpha/\beta$. Substituting both results in (28) yields

$$\frac{\upsilon_I^*}{1-\upsilon_I^*} = \kappa_1^{-1}\left(\theta_1 + \frac{\alpha\theta_2}{\beta}\right).$$

Further manipulations lead to the following proposition:

Proposition 4. With a constant tax rate, the growth-maximizing allocation of spending between education services and infrastructure investment is given by

$$\upsilon_I^* = \frac{\alpha\kappa_1 + (\alpha+\beta)\kappa_2}{(\alpha+\beta)(\kappa_1+\kappa_2)} \in (0, 1), \quad \upsilon_E^* = 1 - \upsilon_I^*. \tag{30}$$

Thus, in general, the growth-maximizing share of investment in infrastructure depends not only on the parameters characterizing the goods production technology, α and β, but also on those characterizing the human capital technology, κ_1 and κ_2. In particular, the following corollary can be established from Proposition 4:

Corollary 1. An increase in the elasticity κ_1 (κ_2) in the human capital technology lowers (increases) the growth-maximizing share of spending on infrastructure investment.

It can also be verified that when $\kappa_2 = 0$, $\upsilon_I^* = \alpha/(\alpha + \beta)$, as shown elsewhere in a flow specification (see Agénor 2005, 2008a). Put differently, if infrastructure capital has no effect on the human capital technology, the optimal allocation of public expenditure depends solely on the relative importance of elasticities of goods production with respect to human and public capital. And naturally enough, if government spending on education services has no effect on the production of human capital (so that $\kappa_1 = 0$), then $\upsilon_I^* = 1$. The same result is obtained if human capital has no effect on production ($\beta = 0$), regardless of the values of κ_1 and κ_2.

The following result can also be readily established from (30):

Corollary 2. The growth-maximizing share of investment on infrastructure is such that $\upsilon_I^* > \alpha$, regardless of whether $\kappa_2 \gtrless 0$.

In other words, the growth-maximizing spending share on infrastructure investment exceeds the elasticity of output with respect to public capital in infrastructure, as predicted by the Barro rule. This result holds even if $\kappa_2 = 0$. The reason of course is that the goods production technology implies that a higher stock of public capital in infrastructure raises the marginal productivity of *both* private capital and human capital. Note, however, that if $\kappa_2 = 0$, then κ_1 has no effect on the optimal spending share; put differently, the education technology matters for the optimal policy only if public capital in infrastructure is an essential input in the production of knowledge.

Welfare-maximizing policy

Consider now the first-best welfare-maximizing policy, which involve a planner choosing optimally all quantities and policy instruments so as to maximize (subject to appropriate constraints) the household's discounted lifetime utility. Again, because the government budget constraint must hold at all times, one of the spending shares must be determined residually; in what follows, it will be assumed that υ_I is determined optimally while υ_E is set residually through (10).

To begin with, note that the economy's aggregate resource constraint (12) can be rewritten, using (9) and (10) with $\upsilon_U = 0$ for simplicity, as

$$(31) \qquad \dot{K}_P + \dot{K}_G = [1 - \tau(1 - \upsilon_I)]Y - C,$$

whereas from (6), (9) and (10),

$$(32) \qquad \dot{H} = [\tau(1 - \upsilon_I)]^{\kappa_1} Y^{\kappa_1} K_G^{\kappa_2} H^{1-\kappa_1-\kappa_2}.$$

The social planner's problem is thus to maximize (2), subject to (1), (31) and (32), with respect to C, τ, υ_I, K_P, K_G, H and Y. The Hamiltonian can be written as

$$\Lambda = \ln C + \zeta_1\{[1 - \tau(1 - \upsilon_I)]Y - C\}$$
$$+ \zeta_2[\tau(1 - \upsilon_I)]^{\kappa_1} Y^{\kappa_1} K_G^{\kappa_2} H^{1-\kappa_1-\kappa_2} + \mu(K_G^{\alpha} H^{\beta} K_P^{1-\alpha-\beta} - Y),$$

where ζ_1 and ζ_2 denote the co-state variables associated with constraints (31) and (32), respectively, and μ is the Lagrange multiplier associated with the production function (1). Optimality conditions are given by

$$(33) \qquad \frac{\partial \Lambda}{\partial C} = 0 \quad \Rightarrow \quad \frac{1}{C} = \zeta_1,$$

(34) $$\frac{\partial\Lambda}{\partial\tau} = 0 \quad \Rightarrow \quad -\zeta_1(1-\upsilon_I)Y = -\zeta_2\frac{\kappa_1\dot{H}}{\tau},$$

(35) $$\frac{\partial\Lambda}{\partial\upsilon_I} = 0 \quad \Rightarrow \quad \zeta_1\tau Y = \zeta_2\frac{\kappa_1\dot{H}}{1-\upsilon_I},$$

(36) $$\frac{\partial\Lambda}{\partial Y} = 0 \quad \Rightarrow \quad \zeta_1[1-\tau(1-\upsilon_I)] + \zeta_2\frac{\kappa_1\dot{H}}{Y} = \mu,$$

(37) $$\frac{\partial\Lambda}{\partial K_P} = \mu(1-\alpha-\beta)\left(\frac{Y}{K_P}\right) = -\dot{\zeta}_1 + \rho\zeta_1,$$

(38) $$\frac{\partial\Lambda}{\partial K_G} = \zeta_2\frac{\kappa_2\dot{H}}{K_G} + \mu\alpha\left(\frac{Y}{K_G}\right) = -\dot{\zeta}_1 + \rho\zeta_1,$$

(39) $$\frac{\partial\Lambda}{\partial H} = \zeta_2\frac{(1-\kappa_1-\kappa_2)\dot{H}}{H} + \mu\beta(Y/H) = -\dot{\zeta}_2 + \rho\zeta_2,$$

together with (1), (31) and (32), and the transversality conditions

(40) $$\lim_{t\to\infty}\zeta_1 K_P\exp(-\rho t) = \lim_{t\to\infty}\zeta_1 K_G\exp(-\rho t) = \lim_{t\to\infty}\zeta_1 H\exp(-\rho t) = 0.$$

A complete characterization of the dynamics of the model can be conducted along the lines outlined in the previous sections. It is omitted here because it is not necessary to highlight some of the key features of the optimal solution and the role of κ_2. Indeed, note first that from (35) and (36), $\zeta_1 = \mu$. From (37) and (38),

$$\mu(1-\alpha-\beta)\left(\frac{Y}{K_P}\right) = \zeta_2\frac{\kappa_2\dot{H}}{K_G} + \mu\alpha\left(\frac{Y}{K_G}\right).$$

Second, from (36) and $\zeta_1 = \mu$, $\zeta_2\dot{H} = \tau(1-\upsilon_I)\mu Y/\kappa_1$. Substituting this result in the above expression yields

$$\mu Y\frac{(1-\alpha-\beta)}{K_P} = \frac{\kappa_2\tau(1-\upsilon_I)\mu Y}{\kappa_1 K_G} + \mu\alpha\left(\frac{Y}{K_G}\right),$$

which can be rearranged to give

(41) $$k_G = \frac{1}{1-\alpha-\beta}\left\{\alpha + \frac{\kappa_2\tau(1-\upsilon_I)}{\kappa_1}\right\}.$$

Along the balanced growth path, the public–private capital ratio is constant at $\tilde{k}_G$. From (41), this implies that either the tax rate and the spending share on infrastructure are both constant as well or, if they are time-varying, they move in exactly the same direction, that is, $\dot{\upsilon}_I/\upsilon_I = \dot{\tau}/\tau$. In addition, in the first case the model displays no transitional dynamics with respect to k_G, which cannot deviate from its steady-state value. Intuitively, the social planner chooses the allocation between the two types of capital so that their shadow values are equal. As a result, the optimal capital ratio remains constant at all times. This implies also that one of the two fiscal policy instruments can be set arbitrarily. Intuitively, what this means is that from a welfare perspective, it is the *total claim* on the economy's resources that matters; the optimal share of infrastructure spending in output can be achieved by combining either a high tax rate and a low share of spending, or a low tax rate and a high share of expenditure. These results can be summarized in the following proposition.

Proposition 5. Under the first-best welfare-maximizing solution, constancy of the public–private capital ratio along the balanced growth path implies that either the tax rate and the spending share on infrastructure are both constant or, if they are time-varying, they move in the same direction. In the first case the model displays no transitional dynamics with respect to the public–private capital ratio, and one of the fiscal instruments can be set arbitrarily.

A final result that is worth pointing out is that inspection of (41) reveals that with $\kappa_2 = 0$, $k_G = \alpha/(1 - \alpha - \beta)$, which is nothing but the ratio of the elasticities of output with respect to each physical capital stock (see equation (1)). In that case, regardless of whether or not the tax rate and the spending share on infrastructure are constant, the model cannot exhibit transitional dynamics with respect to the public–private capital ratio. Thus accounting for the impact of infrastructure on the human capital technology has important implications for the dynamics of the economy under the welfare-maximizing solution as well.

V. Concluding Remarks

The purpose of this paper has been to analyse trade-offs between public spending on infrastructure investment and the provision of education services in an endogenous growth model where, as in Lucas-type models, knowledge is disembodied. Public infrastructure in the model is treated as a stock. A crucial feature of the model is that the production of human capital requires not only government spending on education services but infrastructure capital as well. The key idea is that to go to school, students need access to roads; to read and study (at night), or to connect a computer, they need electricity; to get children to attend school, adequate water and sanitation facilities must be provided. A number of recent microeconomic studies have indeed provided evidence that infrastructure has a significant impact on education outcomes in developing countries.

After deriving the balanced growth path of the model, conditions for saddlepath stability were examined. It was shown that the equilibrium path may be locally indeterminate, and that the source of indeterminacy is fundamentally related to the externalities associated with public infrastructure in the production of goods and human capital. Numerical techniques were then used to examine the transitional and long-run effects of a budget-neutral switch in government spending from education to infrastructure, for different values of the parameters characterizing the human capital technology. Under a plausible calibration for a low-income country, it was shown that reallocating funds from education to infrastructure may increase the growth rate (even though the stock of human capital grows at a slower rate), even if public infrastructure capital has only a moderate effect on the production of human capital.

The growth-maximizing tax rate and shares of government spending on infrastructure investment and education services were then determined. The optimal tax rate was shown to be equal to the sum of the elasticities of output with respect to public infrastructure and human capital, in line with results obtained elsewhere with flow models. It was also shown that the growth-maximizing allocation of public spending depends not only on the elasticity of output with respect to infrastructure capital and human capital (as would be expected), but also on the parameters characterizing the technology for producing human capital. In particular, the growth-maximizing share of spending on infrastructure exceeds the elasticity of output with respect to public capital in infrastructure, as predicted by the Barro rule. Moreover, the lower the ratio of elasticities of the education technology with respect to human capital and infrastructure capital, the lower the share of spending on infrastructure investment. It was noted that the optimal allocation rule does not depend on whether the impact of infrastructure on the human capital technology is measured as a flow or a stock effect. Because models in which infrastructure spending is modelled as a flow are much more tractable analytically, this result carries some importance for future work in this area. Finally, it was shown that under the first-best welfare-maximizing solution, constancy of the public–private capital ratio along the balanced growth path implies either constancy of both the tax rate and the spending share on infrastructure or, if they are time-varying, that they move in the same way. In the first case the model displays no transitional dynamics and one of the fiscal instruments can be set arbitrarily.

The analysis can be extended in several directions. Two possibilities would be to account for a more general production structure by using a multi-level CES function (along the lines of Krusell *et al.* 2000, for instance) or account for public capital in education, which would require treating schools, libraries, and so on, as a stock. Another direction would be to consider a mixed (private–public) education system as, for instance, in Chen (2005). According to Chen (2005, Table 2), private school enrolment as a proportion of total enrolment in secondary schools varies significantly even among low-income countries. For instance, it ranges from about 9% in Mali to 48% in Burkina Faso and 84% in Haiti. This extension (which would need to account for differences in quality as well) would allow a discussion of the potential trade-offs between subsidies to private schools and direct government involvement in education—a key policy issue for developed and developing countries alike. Finally, a third extension would be to introduce health considerations. Recent evidence for developing countries suggests that infrastructure may affect not only the creation and transmission of knowledge (as in the present study), but also the efficiency of government-provided health services. In addition, the evidence suggests that health services enhance not only the productivity of workers, but also household utility—healthier individuals may enjoy consumption more—and the ability to learn and study. Agénor and Neanidis (2006) provide a general framework for analysing how these interactions affect the optimal allocation of government spending between infrastructure, education and health services.

APPENDIX

The balanced growth equilibrium is determined as follows. First, the household budget constraint (equation (3)) can be rewritten, using (1), as

$$\dot{K}_P = (1-\tau)\left(\frac{K_G}{K_P}\right)^{\alpha}\left(\frac{H}{K_P}\right)^{\beta} K_P - C,$$

that is, with $c = C/K_P$, $k_G = K_G/K_P$ and $h = H/K_P$,

$$\text{(A1)} \qquad \frac{\dot{K}_P}{K_P} = (1-\tau)k_G^{\alpha}h^{\beta} - c.$$

Similarly, equation (4) can be rewritten as

(A2) $$\frac{\dot{C}}{C} = sk_G^{\alpha}h^{\beta} - \rho.$$

Equation (7) gives

(A3) $$\frac{\dot{H}}{H} = \left(\frac{G_E}{H}\right)^{\kappa_1}\left[\left(\frac{K_G}{K_P}\right)\left(\frac{K_P}{H}\right)\right]^{\kappa_2} = \left(\frac{G_E}{H}\right)^{\kappa_1}\left(\frac{k_G}{h}\right)^{\kappa_2},$$

so that, noting that $G_E/H = \tau(G_E/\tau Y)(Y/H)$, and using (9),

$$\frac{\dot{H}}{H} = (\tau \upsilon_E)^{\kappa_1}\left[\left(\frac{Y}{K_P}\right)\left(\frac{K_P}{H}\right)\right]^{\kappa_1}\left(\frac{k_G}{h}\right)^{\kappa_2}.$$

Using (1), which implies that $Y/K_P = k_G^{\alpha}h^{\beta}$, yields

(A4) $$\frac{\dot{H}}{H} = (\tau \upsilon_E)^{\kappa_1}k_G^{\theta_1}h^{-\theta_2},$$

where θ_1 and θ_2 are defined in the text.

From (11), using (1) and (9), and noting that $Y/K_G = (Y/K_P)k_G^{-1}$,

(A5) $$\frac{\dot{K}_G}{K_G} = \frac{\upsilon_I \tau Y}{K_G} = \tau \upsilon_I k_G^{\alpha-1}h^{\beta}.$$

Combining equations (A1), (A2), (A4) and (A5) yields equations (13), (14) and (15) in the text.

Setting $\dot{c} = \dot{h} = \dot{k}_G = 0$ in equations (13), (14) and (15), and using (10) with $\upsilon_U = 0$ for simplicity, yields

(A6) $$\tilde{c} = \rho - \Omega \tilde{k}_G^{\alpha}\tilde{h}^{\beta},$$

(A7) $$[\tau(1-\upsilon_I)]^{\kappa_1}\tilde{k}_G^{\theta_1}\tilde{h}^{-\theta_2} - (1-\tau)\tilde{k}_G^{\alpha}\tilde{h}^{\beta} + \tilde{c} = 0,$$

(A8) $$[\tau \upsilon_I \tilde{k}_G^{-1} - (1-\tau)]\tilde{k}_G^{\alpha}\tilde{h}^{\beta} + \tilde{c} = 0.$$

Substituting (A6) in (A7) and (A8) yields the following system in $\tilde{h}$ and $\tilde{k}_G$:

(A9) $$Q_1(\tilde{h}, \tilde{k}_G) = [\tau(1-\upsilon_I)]^{\kappa_1}\tilde{k}_G^{\theta_1}\tilde{h}^{-\theta_2} - s\tilde{k}_G^{\alpha}\tilde{h}^{\beta} + \rho = 0,$$

(A10) $$Q_2(\tilde{h}, \tilde{k}_G) = (\tau \upsilon_I \tilde{k}_G^{-1} - s)\tilde{k}_G^{\alpha}\tilde{h}^{\beta} + \rho = 0,$$

which can be linearized in the vicinity of the balanced-growth equilibrium to solve for $\tilde{h}$ and $\tilde{k}_G$.

To determine the steady-state impact of a budget-neutral increase in υ_I (that is, with $d\upsilon_I = -d\upsilon_E$), these last two equations can be manipulated to yield

(A11) $$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} d\tilde{h} \\ d\tilde{k}_G \end{bmatrix} = \begin{bmatrix} b_{13} \\ b_{23} \end{bmatrix} d\upsilon_I,$$

where

$$b_{11} = -\theta_2(\tau\upsilon_E)^{\kappa_1}\tilde{k}_G^{\theta_1}\tilde{h}^{-\theta_2-1} - \beta s\tilde{k}_G^{\alpha}\tilde{h}^{\beta-1} < 0,$$

$$b_{12} = \theta_1(\tau\upsilon_E)^{\kappa_1}\tilde{k}_G^{\theta_1-1}\tilde{h}^{-\theta_2} - \alpha s\tilde{k}_G^{\alpha-1}\tilde{h}^{\beta},$$

$$b_{13} = \kappa_1\tau^{\kappa_1}\upsilon_E^{\kappa_1-1}\tilde{k}_G^{\theta_1}\tilde{h}^{-\theta_2} > 0,$$

$$b_{21} = \beta(\tau\upsilon_I\tilde{k}_G^{-1} - s)\tilde{k}_G^{\alpha}\tilde{h}^{\beta-1},$$

$$b_{22} = \alpha(\tau\upsilon_I\tilde{k}_G^{-1} - s)\tilde{k}_G^{\alpha-1}\tilde{h}^{\beta} - \tau\upsilon_I\tilde{k}_G^{\alpha-2}\tilde{h}^{\beta},$$

$$b_{23} = \tau\tilde{k}_G^{\alpha-1}\tilde{h}^{\beta} < 0.$$

To establish the sign of b_{12}, note that, using (A9),

$$b_{12} = (\theta_1 - \alpha)s\tilde{k}_G^{\alpha-1}\tilde{h}^{\beta} - \theta_1\rho/\tilde{k}_G,$$

so that if $\theta_1 > \alpha$, as mentioned in the text (a necessary condition for $a_{23} > 0$), b_{12} is in general ambiguous. However, for typical values of ρ, it is likely that $b_{12} > 0$.

Similarly, to establish the signs of b_{21} and b_{22}, note that, using (A10),[18]

$$b_{21} = -\beta\rho/\tilde{h} < 0,$$

$$b_{22} = -\alpha\rho/\tilde{k}_G - \tau\upsilon_I\tilde{k}_G^{\alpha-2}\tilde{h}^{\beta} < 0.$$

Solving equations (A11) and noting that $b_{11}b_{22} - b_{12}b_{21} > 0$, it can be shown that

$$\text{(A12)}\qquad \frac{d\tilde{h}}{d\upsilon_I} = \frac{b_{13}b_{22} - b_{12}b_{23}}{b_{11}b_{22} - b_{12}b_{21}} \lesseqgtr 0, \qquad \frac{d\tilde{k}_G}{d\upsilon_I} = \frac{b_{11}b_{23} - b_{21}b_{13}}{b_{11}b_{22} - b_{12}b_{21}} > 0.$$

Thus an increase in υ_I always raises $\tilde{k}_G$, but has an ambiguous effect on $\tilde{h}$. Moreover, it can readily be verified that even with $\kappa_2 = 0$, in which case $\theta_2 \equiv (1-\beta)\kappa_1$, this ambiguity remains. If $\kappa_1 = 0$, then $b_{13} = 0$, and the above expression yields $d\tilde{h}/d\upsilon_I > 0$. At the same time, a higher value of κ_2 makes b_{13} smaller, making it more likely (everything else equal) that $d\tilde{h}/d\upsilon_I > 0$.

By implication, the effect of υ_I on the steady-state growth rate and the consumption–capital ratio is also ambiguous in general. In addition, it will depend also directly on α and β, as can be inferred from (17) and (A6), in addition to the indirect effects stemming from (A12).

ACKNOWLEDGMENTS

I am grateful to Nihal Bayraktar for assistance in preparing the numerical simulations reported in this paper, and to Kyriakos Neanidis and two anonymous referees for helpful discussions and comments. The views expressed here, however, are my own.

NOTES

1. Glomm and Ravikumar (1994) assume that private and public capital depreciate fully each period. Given constant returns to scale in production, the economy is thus always on a balanced growth path. In contrast, in the present setting, transitional dynamics are explicitly analysed.
2. In a previous paper (Agénor 2005), I considered the case where the *flow* of government services in infrastructure affects the economy's ability to produce educated labour. The optimal rules derived in that paper will later be compared with those obtained here with a stock treatment.
3. As shown by Turnovsky (1996), however, if private investment is subject to adjustment costs that fall with public services, the positive impact of government spending on growth will be more pronounced. The Barro rule would then underestimate the optimal spending share and tax rate.
4. The assumption of mandatory schooling may seem at variance with the evidence cited in the Introduction, which suggests that children may not be able to go to school due, for instance, to inadequate means of transportation. Moreover, in general, education in developing countries is compulsory only for the early years of schooling. The model takes a simplified approach in the sense

that it imposes 'full employment' of a particular type—individuals are either in work or in school, but never idle. This is a reasonable assumption in a model with complete wage flexibility.

5. The time subscript t is omitted whenever there is no risk of confusion. A dot over a variable is used later to denote the derivative of that variable with respect to time.
6. See also Eicher and Turnovsky (1999) for a discussion of the importance of the Cobb–Douglas specification to ensure constant per capita growth rates in endogenous growth models.
7. Changes in the stock of infrastructure assets could be made a function also of the human capital stock, in the form $\dot{K}_G = G_I^{\phi} H^{1-\phi}$, where $\phi \in (0, 1)$. Note, however, that an indirect effect of H is already captured in (11), given that H affects output and thus G_I through (9). Subsequent results would therefore not change much as long as ϕ is not too small.
8. From (A1), there is a fourth equivalent form, $\gamma = (1 - \tau)\tilde{k}_G^{\alpha}\tilde{e}^{\beta} - \tilde{c}$. However, given (20) below, this expression is identical to (17).
9. Note that in a more 'standard' model of endogenous growth with public capital, in which $\kappa_2 = 0$, this condition cannot be satisfied, given that $0 < \alpha, \kappa_1 < 1$. Thus in that case $a_{23} < 0$. Based on the subsequent discussion, it is easy to verify that saddlepath stability cannot be established unambiguously either.
10. The modified Routh–Hurwitz conditions, as given in Murata (1977, pp. 92–4), do not yield more tractable expressions.
11. A necessary and sufficient condition for excluding the possibility of local indeterminacy is $\text{tr}\,\mathbf{J} > 0$, because in this case the three roots cannot be all negative.
12. See Harrison and Weder (2002) and Hu *et al.* (2008) for a discussion of the role of sector- and factor-specific externalities.
13. See Armstrong *et al.* (1998) for an alternative solution technique, based on a Newton 'stacked-time' algorithm. The computer programmes used for this study are available upon request.
14. Note, however, that the Arestoff–Hurlin estimates are based on the perpetual inventory method, which consists essentially in cumulating total capital expenditure flows by central governments. These flows include items that are not, strictly speaking, related to infrastructure or productive spending, as defined earlier.
15. Blankenau *et al.* (2007) found that the elasticity of human capital with respect to government spending on education is close to zero for low-income countries, but this runs counter to intuition. It also does not account for the heterogeneity in public school enrolment discussed in the Conclusion.
16. In the present context, Rioja's results correspond to the case where $\kappa_2 = 0$. In that particular case, numerical simulations also indicate that the expenditure shift considered here would lower the steady-state growth rate if κ_1 is relatively large with respect to α.
17. Note that the optimal tax rate is considerably higher than the value actually used in the numerical simulations presented in the previous section (0.6, compared to 0.2). However, it should be kept in mind that the present analysis abstracts from tax collection costs, which are sizeable in low-income countries. As shown by Agénor and Neanidis (2007), accounting for these costs would lead to much lower optimal rates, in line with those observed in poor countries. An alternative approach would be to introduce tax evasion, as in Lin and Yang (2001) and Chen (2003).
18. Alternatively, b_{22} can be written as $b_{22} = -(1 - \alpha)\tau \upsilon_I \tilde{k}_G^{\alpha-2}\tilde{h}^{\beta} - \alpha s \tilde{k}_G^{\alpha-1}\tilde{h}^{\beta} < 0$.

REFERENCES

AGÉNOR, P.-R. (2005). Infrastructure, public education and growth with congestion costs. Working Paper no. 47, Centre for Growth and Business Cycle Research, University of Manchester.

——— (2008a). Fiscal policy and endogenous growth with public infrastructure. *Oxford Economic Papers*, **60**, 57–88.

——— (2008b). Health and infrastructure in models of endogenous growth. *Journal of Macroeconomics*, **30**, 1407–22.

——— (2009). Public capital, health persistence and poverty traps. Working Paper no. 115, Centre for Growth and Business Cycle Research, University of Manchester.

——— and MONTIEL, P. J. (2008). *Development Macroeconomics*, 3rd edn. Princeton, NJ: Princeton University Press.

——— and NEANIDIS, K. (2006). The allocation of public expenditure and economic growth. Working Paper no. 69, Centre for Growth and Business Cycle Research, University of Manchester.

——— and ——— (2007). Optimal taxation and growth with public goods and costly enforcement. Working Paper no. 89, Centre for Growth and Business Cycle Research, University of Manchester.

ARESTOFF, F. and HURLIN, C. (2005a). The productivity of public capital in developing countries. Unpublished, University of Orléans.

——— and ——— (2005b). Threshold effects in the productivity of public capital in developing countries. Unpublished, University of Orleans.

ARMSTRONG, J., BLACK, R., LAXTON, D. and ROSE, D. (1998). A robust method for simulating forward-looking models. *Journal of Economic Dynamics and Control*, **22**, 489–501.

ATOLIA, M., CHATTERJEE, S. and TURNOVSKY, S. J. (2008). How misleading is linearization? Evaluating the dynamics of the neoclassical growth model. Unpublished, Florida State University.

BAIER, S. L. and GLOMM, G. (2001). Long-run growth and welfare effects of public policies with distortionary taxation. *Journal of Economic Dynamics and Control*, **25**, 1007–42.

BARRO, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy*, **98**, s103–s25.

BIRD, R. M. and ZOLT, E. M. (2005). The limited role of the personal income tax in developing countries. *Journal of Asian Economics*, **16**, 928–46.

BLANKENAU, W. F., SIMPSON, N. B. and TOMLJANOVICH, M. (2007). Public education expenditures, taxation and growth: linking data to theory. *American Economic Review*, **97**, 393–7.

BOSE, N., HAQUE, M. E. and OSBORN, D. R. (2007). Public expenditure and economic growth: a disaggregated analysis for developing countries. *Manchester School*, **75**, 533–56.

BRENNEMAN, A. and KERF, M. (2002). Infrastructure and poverty linkages: a literature review. Unpublished, World Bank.

CANNING, D. (1999). Infrastructure's contribution to aggregate output. Policy Research Working Paper no. 2246, World Bank.

CANTON, E. (2001). Fiscal policy in a stochastic model of endogenous growth. *Economic Modelling*, **18**, 19–47.

CERRA, V., TEKIN, S. and TURNOVSKY, S. J. (2008). Foreign aid and real exchange rate adjustments in a financially constrained dependent economy. Working Paper no. 08/204, International Monetary Fund.

CHEN, B.-L. (2003). Tax evasion in a model of endogenous growth. *Review of Economic Dynamics*, **6**, 381–403.

——— (2007). Factor taxation and labor supply in a dynamic one-sector growth model. *Journal of Economic Dynamics and Control*, **31**, 3941–64.

——— and LEE, S.-F. (2007). Congestible public goods and local indeterminacy: a two-sector endogenous growth model. *Journal of Economic Dynamics and Control*, **31**, 2486–518.

CHEN, H. (2005). Educational systems, growth and income distribution: a quantitative analysis. *Journal of Development Economics*, **76**, 325–53.

COLE, M. A. and NEUMAYER, E. (2006). The impact of poor health on total factor productivity. *Journal of Development Studies*, **42**, 918–38.

EASTERLY, W. and REBELO, S. (1993). Fiscal policy and economic growth: an empirical investigation. *Journal of Monetary Economics*, **32**, 417–58.

EICHER, T. and TURNOVSKY, S. J. (1999). Non-scale models of endogenous growth. *Economic Journal*, **109**, 394–415.

ESFAHANI, H. S. and RAMÍREZ, M. T. (2003). Institutions, infrastructure, and economic growth. *Journal of Development Economics*, **70**, 443–77.

FAIR, R. C. and TAYLOR, J. B. (1983). Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. *Econometrica*, **51**, 1169–96.

FISHER, W. H. and TURNOVSKY, S. J. (1998). Public investment, congestion, and private capital accumulation. *Economic Journal*, **108**, 399–413.

FUTAGAMI, K., MORITA, Y. and SHIBATA, A. (1993). Dynamic analysis of an endogenous growth model with public capital. In T. M. Andersen and K. O. Moene (eds), *Endogenous Growth*. Oxford: Basil Blackwell.

GARCÍA-PEÑALOSA, C. and TURNOVSKY, S. J. (2005). Second-best optimal taxation of capital and labor in a developing economy. *Journal of Public Economics*, **89**, 1045–74.

GHOSH, S. and ROY, U. (2004). Fiscal policy, long-run growth, and welfare in a stock-flow model of public goods. *Canadian Journal of Economics*, **37**, 742–56.

GLOMM, G. and RAVIKUMAR, B. (1994). Public investment in infrastructure in a simple growth model. *Journal of Economic Dynamics and Control*, **18**, 1173–87.

GÓMEZ, M. A. (2004). Optimal fiscal policy in a growing economy with public capital. *Macroeconomic Dynamics*, **8**, 419–35.

HARRISON, S. G. and WEDER, M. (2002). Tracing externalities as sources of indeterminacy. *Journal of Economic Dynamics and Control*, **26**, 851–67.

HU, Y., OHDOI, R. and SHIMOMURA, K. (2008). Indeterminacy in a two-sector endogenous growth model with productive government spending. *Journal of Macroeconomics*, **30**, 1104–23.

HULTEN, C. R. (1996). Infrastructure capital and economic growth: how well you use it may be more important than how much you have. Working Paper no. 5847, National Bureau of Economic Research.

KRUSELL, P., OHANIAN, L. E., RÍOS-RULL, J.-V. and VIOLANTE, G. L. (2000). Capital–skill complementarity and inequality: a macroeconomic analysis. *Econometrica*, **68**, 1029–53.

LIN, W.-Z. and YANG, C. (2001). A dynamic portfolio choice model of tax evasion: comparative statics of tax rates and its implication for economic growth. *Journal of Economic Dynamics and Control*, **25**, 1827–40.
MURATA, Y. (1977). *Mathematics for Stability and Optimization of Economic Systems*. New York: Academic Press.
ORTIGUEIRA, S. (1998). Fiscal policy in an endogenous growth model with human capital accumulation. *Journal of Monetary Economics*, **42**, 323–55.
OTTO, G. D. and VOSS, G. M. (1998). Is public capital provision efficient? *Journal of Monetary Economics*, **22**, 47–66.
PALIVOS, T., YIP, C. K. and ZHANG, J. (2003). Transitional dynamics and indeterminacy of equilibria in an endogenous growth model with a public input. *Review of Development Economics*, **7**, 86–98.
RIOJA, F. K. (2005). Roads versus schooling: growth effects of government choices. *Topics in Macroeconomics*, **5**, 1–22.
——— and GLOMM, G. (2003). Populist budgets and long run growth. Unpublished, Indiana University.
RIVAS, L. A. (2003). Income taxes, spending composition and long-run growth. *European Economic Review*, **47**, 477–503.
SAGHIR, J. (2005). Energy and poverty: myths, links, and policy issues. Working Note no. 4, Energy and Mining Sector Board, World Bank.
SONG, L. L. (2002). Public capital, congestion and private production in Australia. Working Paper no. 23/02, University of Melbourne.
TSOUKIS, C. and MILLER, N. J. (2003). Public services and endogenous growth. *Journal of Policy Modeling*, **25**, 297–307.
TURNOVSKY, S. J. (1996). Fiscal policy, adjustment costs, and endogenous growth. *Oxford Economic Papers*, **48**, 361–81.
——— (1997). Fiscal policy in a growing economy with public capital. *Macroeconomic Dynamics*, **1**, 615–39.
——— (2000). Fiscal policy, elastic labor supply, and endogenous growth. *Journal of Monetary Economics*, **45**, 185–210.
UNITED NATIONS MILLENNIUM PROJECT (2005). *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals*. New York: Report to the United Nations Secretary General.
WORLD BANK (2006). Infrastructure: lessons from the last two decades of World Bank engagement. Unpublished, Washington DC: Infrastructure Network, World Bank.
YAKITA, A. (2004). Elasticity of substitution in public capital formation and economic growth. *Journal of Macroeconomics*, **26**, 391–408.

Economica

Economica (2011) **78**, 133–158
doi:10.1111/j.1468-0335.2009.00804.x

Work-sharing During the Great Depression: Did the 'President's Reemployment Agreement' Promote Reemployment?

By Jason E. Taylor

Central Michigan University

Final version received 28 January 2009.

The President's Reemployment Agreement (PRA) of 1933 directed firms to reduce workweeks during the Great Depression so existing jobs could be spread into additional employment opportunities. Similar 'work-sharing' policies have recently been implemented across Europe in hopes of reducing unemployment. I find that, *ceteris paribus*, the work-sharing aspects of the PRA created nearly 2.5 million new employment opportunities in around four months. However, the programme also required firms to raise hourly wage rates, offsetting close to half of these gains. Furthermore, most of the remaining employment gains were wiped out after cartel-oriented industry-specific codes of fair competition supplanted the PRA.

Introduction

> The essence of the plan is a universal limitation of hours of work per week for any individual by common consent, and a universal payment of a wage above a minimum . . . I am asking the employers of the nation to sign this common covenant . . . in the name of patriotism and humanity.
>
> (President Franklin D. Roosevelt, 24 July 1933)

Following the passing of the National Industrial Recovery Act (NIRA) on 16 June 1933, executives from hundreds of industries began the long process of negotiating the contents of their cartel-oriented 'codes of fair competition'. Unhappy with the speed of these negotiations, on 20 July 1933 the Roosevelt administration outlined a new programme termed the 'President's Reemployment Agreement' (PRA). The PRA asked firms to declare 'a truce on selfishness' by quickly signing and abiding by a so-called 'blanket code' agreement to reduce the average workweek and raise hourly wage rates—two major components that, to be approved, would ultimately have to be included alongside more business-friendly, cartel-oriented provisions in all industry-specific codes of fair competition.

The logic behind the PRA's labour provisions was twofold. First, it was thought that higher wage rates would increase aggregate purchasing power, spending, production and, ultimately, employment. The second rationale behind the PRA was that by reducing the workweek from 45–50 hours to around 35 hours, work could be 'spread' or 'shared' among more people—three jobs could effectively be created where there were previously two, even holding constant any other effects of New Deal fiscal, monetary or structural policy. Of course, similar 'work-sharing' plans have been instituted in Europe during the past 30 years—most recently in France, where the Aubry Act of 2000 reduced the legal standard work week from 39 to 35 hours—in hopes of reducing unemployment. The economics literature, both theoretical and empirical, on the employment effects of work-sharing has come to conflicting findings regarding the efficacy of such measures.

The Roosevelt administration claimed that the reemployment programme was a fantastic success, stating in November 1933 that it had put 4 million Americans to work.

A contemporary study by Lyon *et al.* (1972) was sceptical of the magnitude of this claim, but still estimated that 1.75 million jobs were created by late 1933. To more precisely quantify the economic effects of the PRA as a programme separate from the cartel-oriented NIRA, particularly in light of other important changes in fiscal and monetary policy, this paper employs an industry-level monthly panel dataset. Such an empirical approach is necessary since some industries were covered by the PRA provisions for several months, while others were covered only briefly or not at all before passing their industry-specific NIRA code whereby the blanket code was superseded.

While my primary objective is to estimate the effects of the PRA upon reemployment—both through number of jobs and, more broadly, through its impact on aggregate employment hours worked—I also examine how industry payroll, output and prices, among other variables of interest, were affected during the months between the PRA's passage and the implementation of an industry-specific code. I find that the 'share the work' goal of the PRA was achieved, as an additional 2.47 million Americans were able to draw (albeit smaller) paychecks thanks to work-sharing provisions, *ceteris paribus*. I also find that the PRA's wage-increasing provisions had the economic consequence predicted by orthodox economic theory—lower firm output and higher prices, i.e. a negative supply shock—rather than the positive demand effects predicted by the Roosevelt administration. In fact, my empirical analysis suggests that around 45% of the job gains created by the PRA work-sharing aspects were wiped out by hourly wage rate increases. Still, while the PRA has received little attention by economists, the results suggest that the programme put 1.34 million persons to work in around four months.

I. The NIRA, the PRA and the Blue Eagle

As many within the Roosevelt administration believed that the Great Depression resulted from 'ruinous' or 'cut-throat' competition, it appeared logical that collusion could help to promote industrial recovery. To that end, the NIRA suspended anti-trust laws for participating firms and required industries to negotiate and abide by a code of economic conduct which could include restrictions on pricing, output, new capacity, production practices and labour conditions. In particular, the Roosevelt administration noted that collective action with respect to labour policy was necessary to prevent the cycle wage cuts.[1]

Although the administration hoped that industry-specific 'codes of fair competition' would be implemented swiftly, the process of getting an NIRA code passed was cumbersome since the contents of these codes were not left solely to the discretion of industrial executives. To better allow the codes to represent the interests of all affected parties, industry-proposed codes had to be approved by a diverse advisory board composed of a deputy administrator of the National Recovery Administration (NRA)—the NIRA's enabling body—as well as a labour representative, a consumer representative, an industry representative and two representatives from the NRA legal division. Furthermore, the text of the NIRA stated that the codes could not be used to 'promote monopoly', placing the NRA advisory board into a delicate balancing act between firms' desires to fix prices and reduce output and the somewhat nebulous concept of 'fair competition'.

The cotton textile industry code was the first to be approved, on 9 July 1933. Three other industries followed with approved codes in July—coats and suits, shipbuilding, and woollen textiles. The sluggish speed of code passage—only four codes passed in over six weeks of negotiation, while hundreds of other industries continued to haggle—provided the impetus for the PRA 'blanket code'. In contrast to industry-specific 'permanent' codes, some of which were scores of pages long, the PRA blanket code was to be a brief agreement on wage rates and workweeks between a *firm* and the President. With the exception of the four industries noted above, firms within every other industry were to be subject to the PRA 'blanket code' from 1 August 1933, continuing until their industry-specific code was passed. In some cases, such as the iron and steel industry, whose code was approved on 19 August, this meant a few weeks; in other cases, such as the laundry and hat industries, whose codes were not passed until February 1934, or the Zinc ore industry, whose code was not passed until March 1935, this meant many months.

As firms were unlikely to voluntarily sign and abide by an agreement to raise wage rates and reduce workweeks absent some exogenous incentive, the administration announced that firms which signed on to the PRA could display the patriotic Blue Eagle emblem on their products and advertisements, and in store windows.[2] What was to have given this emblem economic significance was President Roosevelt's call for households to engage in a nationwide shopping spree at Blue Eagle firms and to effectively boycott those without the emblem. In a radio address on 24 July 1933, Roosevelt said: 'In war, in gloom of night attack, soldiers wear a bright badge on their shoulders to be sure that comrades do not fire on comrades. On that principle, those who cooperate in this program must know each other at a glance . . . and I ask that all those who join me shall display [the Blue Eagle] prominently' (*New York Times*, 25 July 1933, p. 2). Roosevelt's address appears to have inspired firm owners nationwide as the White House received over 20,000 telegrams and letters over the following 48 hours from businesses large and small pledging to raise wage rates and cut hours in exchange for the right to display the compliance emblem (*New York Times*, 28 July 1933, p. 9).

The PRA 'blanket' agreement that firms were asked to sign consisted of three parts. First, a firm generally had to agree to shorten workweeks to no more than 35 hours (40 hours for clerical and sales workers). Second, the firm had to agree to raise its minimum hourly wage rates, generally to 40 cents an hour, depending on city population. Furthermore, in the spirit of the agreement, wage rates of those workers already above this level were to be increased, or, at the very least, were not to be cut. Finally, the firm had to state that it would recognize the rights of workers to bargain collectively. Mail carriers delivered generic copies of the reemployment agreement to all known firms. Firm owners were asked to sign and pledge immediate adherence to PRA guideposts. Upon receipt of the signed pledge at the post office, the code would be approved by the NRA, generally within 24 hours. Once a firm's 'blanket code' was accepted, its name would be included on the NRA 'Honor Roll' of complying firms at the local post office. In many small towns these Honor Rolls, or new additions to them, were printed in local newspapers.[3] Furthermore, the firm owner could obtain—free of charge—posters, stickers and other Blue Eagle paraphernalia at the post office. Some firms chose to advertise compliance via newspaper advertisements—an example from the 1 August 1933 *New York Times* is shown in Figure 1.

In late August 1933, the NRA began a drive to get consumers to formally pledge their support by signing a 'Statement of Cooperation' reading 'I will cooperate in the reemployment by supporting and patronizing employers and workers who are members of the NRA'. The NRA mobilized 1.5 million volunteers to go door to door canvassing with a goal of obtaining 20 million consumer signatures. The *Chicago Tribune* noted that 'such a mobilization has not been seen since the liberty bond drives in the world war' (29 August, p. 5). Signers of this pledge typically received lapel pins and posters (to be placed in windows) that showed the Blue Eagle accompanied by the phrase 'NRA Consumer'.[4]

FIGURE 1. Stern Brothers advertisement, *New York Times*, 1 August 1933.

While signing the PRA code gave the firm the initial right to display the patriotic emblem, to maintain that right, the firm had to continuously follow through on their signed agreement. A firm that was found by the NRA Compliance Director to be in violation of its agreement would be ordered to 'cease displaying the Blue Eagle and deliver all NRA insignia in his possession to the local postmaster' (US Committee of Industrial Analysis 1937, p. 70). The Blue Eagle, or more specifically, consumers' reaction to the emblem's presence, then, was critically important to the PRA having any meaningful effect on economic outcomes. If firms felt that they had nothing to gain from displaying the Blue Eagle, then they would have been unlikely to have signed on to, much less abided by, the wage and hour requirements of the reemployment agreement, leaving the programme inept.

Once an industry-specific code was in place, firms within that industry legally had to comply with the rules set out within it or face prosecution, with penalties of up to $500 and six months' incarceration, as well as loss of the Blue Eagle emblem. In truth, the Roosevelt administration rarely sought legal recourse in the courts, but instead relied only on the threat of removing the Blue Eagle to attain compliance with the industry-specific codes. Firms did not legally have to sign a PRA agreement, however, so the potential impact that the presence (or absence) of the Blue Eagle was expected to have on a firm's business was the only tool the administration had with respect to compliance with this programme. While the vast majority of firms signed a PRA code, the Ford Motor Company was a notable exception. Biles (1991, p. 92) shows that Ford maintained market share in the domestic automobile market despite not being able to display the

Blue Eagle. Consumers may not have punished Ford because the company was already exceeding the wage and hour guideposts of the PRA and, later, the NIRA automobile code, even though it never signed either.

II. Contemporary Accounts of the PRA

The extent to which firms and consumers actually signed on to PRA pledges is a critically important factor with regard to how much impact the programme could have potentially had. A survey of newspapers, both large and small, between August and November 1933, as well as the results of the NRA's periodic surveys of firms and households, strongly suggests that the vast majority of firms signed on to a PRA agreement. Furthermore, we can be reasonably confident that at least some measure of compliance with the PRA wage and hour restrictions generally followed from signing since employees themselves would presumably have a strong incentive to report cheating to the local NRA compliance board or the media.

To illustrate, during the weeks following the PRA's announcement, the *Chicago Tribune* published nearly daily updates on the number of area firms that turned in a PRA pledge form, and listed the identities of the largest firms to sign each day. On 19 August, it was reported that '3,442 additional concerns in this area sent in pledges to join President Roosevelt's reemployment program, raising the total of employees under the NRA standard to 1,192,829'. On 20 August (p. 2), the *Tribune* noted that '532 more signed on covering 9,384' workers. The paper listed Goss Printing Press with 400 employees, J. Greenbaum Tanning with 1236, and Gordon Baking Company with 450, amongst the area's largest firms which had signed on in the past 24 hours.

The 31 August *Washington Post* (p. 2) provided several examples of specific area firms increasing employment: 'Sanitation Grocery Co., it was announced has increased its pay roll $9,500 a week for Washington and Maryland. The Great Atlantic and Pacific Tea Co. . . . has raised its Nation-wide annual payroll by $8,000,000 and added 8,340 employees.' Other Washington DC area firms were simply listed alongside the number of new workers they had hired since signing on to the reemployment agreement: 'Riggs National Bank, 16; Texas Oil Co., 7 at South Washington plant and 2 men per station at all filling stations; American Oil Co., 14 at Rosslyn plant and 100 at Lord-Baltimore filling stations; Continental Bakeries, 14; Havenner Bakery, 11; Chevy Chase Dairy, 35; Thompson's Dairy, 15; Palace Laundry, 34, Sherwood Brothers filling stations, 192; Cities Service filling stations, 10.' Of course, these numbers are self-reported so there is no way to know the extent of their individual accuracy.

On 28 August 1933, the Roosevelt administration began its first door-to-door 'census' of firms and households to estimate the extent of compliance and the number of jobs that the PRA had created. In the weeks that followed, newspapers around the nation reported the specific results of this survey. For example, on 21 September the *Syracuse Herald* (p. 3) reported that the PRA had created 2700 jobs in Syracuse and 51,808 in upstate New York. Given the 1930 populations (209,326 in Syracuse and around 4.5 million in upstate New York) and an assumption of a 40% labour force participation rate, this represents around 3% of the labour force obtaining employment in just over a month in each locale. In Tulare, California, it was reported that 'through NRA compliance 97 jobless Tulareans have been given regular employment' (*Fresno Bee Republican*, 18 November 1933, p. 8). Again, under an assumption of a 40% labour force participation rate, 97 workers represents approximately 4% of Tulare's labour force given its 1930 population of 6207. Interestingly, the article also noted that an additional 69 Tulareans had gained employment through New

Deal work relief programmes. This suggests that the PRA work-sharing programme may have been more important in many communities than better-known work relief aspects of the New Deal with respect to reemployment.

In Jefferson City, Missouri, the *Post Tribune* of 15 September 1933 (p. 1) noted that 'practically all local employers are operating under the president's re-employment agreement or special codes' and that 75% of the city's households were displaying the emblem in their windows. The *Journal and Star* of 10 September (p. 16) reported that 94% of Lincoln, Nebraska's 2015 employers had signed the PRA. In Fitchburg, Massachusetts, the *Sentinel* of 2 September (p. 1) reported that 1100 firms had signed the PRA, which had lead to the creation of 688 new jobs. The *Bee* of 3 September (p. B-3) reported that 2030 Fresno firms had signed on, creating 1914 jobs. Furthermore, the Fresno County Welfare Board reported a 50% drop in requests for food aid since 1 August and attributed this to PRA job creation. The *Chronicle–Telegram* of 21 September (p. 2) reported that 99.32% of Elyria, Ohio households canvassed had signed onto the consumers' pledge of cooperation—only 68 households refused to sign. While this evidence is clearly anecdotal, it provides micro-level insight into the potential impact of the PRA.

With respect to nationwide aggregate numbers, the NRA periodically released a running score from its reemployment censuses. The first figures on the PRA's effects were announced on 30 August, when the NRA declared that the reemployment programme had created 2 million new jobs in just one month (*Washington Post*, 30 August 1933, p. 9). On 13 September, NRA Administrator Johnson noted that 85% of employers nationwide had signed on to the PRA or were covered by an NRA code (*Middletown Times Herald*, 13 September 1933, p. 11). On 14 October, it was reported in an Associated Press article that the administration had tallied 3 million jobs created by the PRA (*Lansing State Journal*, 14 October 1933, p. 1). Finally, in an Associated Press article of 7 November 1933, Johnson credited the reemployment programme as bringing the nation 'a quarter of the way out' of the depression, saying that it 'had put 4 million men back to work and raised the wages of millions of others'.[5] If these numbers are accurate, it would offer important insight into the current 'work-sharing' policy debate. In the following sections, I employ industry-level panel data to see whether the media-reported anecdotal evidence above and the government 'reemployment census' can be supported empirically.

III. Data and Empirical Issues

Many empirical studies, including Cole and Ohanian (2004), Taylor (2002), Vedder and Gallaway (1997), Bernanke (1986) and Weinstein (1980), have examined the effects of the NIRA on either the labour market specifically, or the macro economy more generally. These studies generally conclude that the legislation's labour and cartelization policies were important factors behind the weak recovery of the 1930s as the cartels reduced output and higher wage rates further exacerbated the unemployment problem. Other scholars, such as Alexander (1994, 1997), Krepps (1997) and Taylor (2007), have employed the NIRA experiment to gain insight into cartel theory.[6] While past studies have generally examined the NIRA labour and cartel policies across the legislation, none has examined the specific effects of the PRA, which likely had a far more dramatic short-run impact on labour outcomes during the late summer and autumn of 1933 via significant increases in wage rates and reductions in work hours.

In terms of priors, the employment effects of a shortening of the workweek from a standard of 45–50 hours to the PRA-suggested 35 are controversial. While there is a popular belief that unemployment can be reduced via cuts in the workweek—a belief that has driven 'work-sharing' policy initiatives in Germany, France and Belgium, among other European nations in the past three decades—theoretical and empirical studies by economists have found ambiguous results. On the theoretical side, Calmfors and Hoel (1988) and Brunello (1989), for example, generally predict negative employment effects, while FitzRoy *et al.* (2002), Marimon and Zilibotti (2000) and Rocheteau (2002) predict positive ones. The ambiguity generally stems from work-sharing restrictions making labour less attractive than capital and thus causing firms to substitute these factors. With respect to empirical studies of recent 'work-sharing' programmes, Crepon and Kramarz (2002) find a negative employment effect from France's 1982 workweek reduction. Logeay and Schreiber (2006) conclude that the Aubry Act's (2000) policy mix of shorter workweeks, greater managerial flexibility, and government subsides to social security payments had beneficial employment effects. However, when Schreiber (2006) isolates the effects of shorter workweeks, he concludes that the non-work-sharing reforms drove these employment gains—in fact, shorter workweeks had adverse employment effects. Studies of West Germany's work-sharing reforms by Franz and Konig (1986) and Hunt (1999) have come to opposite conclusions, with the former suggesting positive employment effects and the latter negative ones.[7]

While providing jobs to the unemployed was the primary goal of the PRA, I also wish to examine, more broadly, the impact that the programme had on aggregate hours worked in the economy. The priors regarding this variable differ markedly between the predictions of modern orthodox economic theory and the 'high-wage doctrine' theory prevalent in the 1920s and1930s. Orthodox theory suggests that the mandated exogenous increases in hourly wage rates under the PRA would, other factors constant, reduce total employment hours by causing a leftward movement along the labour demand curve. The Roosevelt administration, however, repeatedly stated belief that higher wage rates would ultimately increase production and employment by boosting aggregate demand (O'Brien 1989; Taylor and Selgin 1999).

In addition to quantifying the PRA's effect on the total number of jobs and aggregate hours worked, I also examine, more generally, whether the legislation provided a boost to other economic factors of interest, such as output, the aggregate wage bill to labour, average take-home pay and prices. As past research generally demonstrates that the NIRA codes of fair competition brought about the cartel outcome of lower output, it seems unlikely that the *industry codes themselves* could have been anything but counterproductive with respect to the aforementioned policy goals. Still, since industries typically took several months to pass their specific codes of fair competition, an empirical study of the PRA's general economic impact provides insight into an important, but largely neglected, aspect of New Deal history.

The priors on movements of these variables are worthy of brief discussion. With respect to the total wage bill, under modern orthodox theory, it is unclear which way this variable would be expected to move under the PRA since higher hourly wage rates would cause fewer hours of work in aggregate.[8] The 'high-wage doctrine' would, of course, predict an unambiguous rise in the wage bill. With respect to the PRA's impact on prices and output, orthodox neoclassical theory would expect the exogenous rise in firms' labour costs to cause output to fall and prices to rise, consistent with a negative supply shock brought about by higher marginal costs. Still, the high-wage doctrine theorizes that the higher hourly wage rates would boost demand so as to offset any negative supply or employment effects they may cause. Of course, output could also have fallen (and prices have risen) from collusion under the PRA—although firms could not legally implement cartel strategies until official code passage, simply having the ability to openly discuss collusive strategies could have facilitated cartel outcomes.

To separate the impact of the PRA from the effects of the NIRA as a whole, I create two dummy variables, together comprising the NIRA time period, by employing the date of passage for each industry's specific code. Because the PRA was to take effect on 1 August 1933, the PRA dummy takes on a value of 1 for each month between August 1933 and the month an industry's specific code was passed.[9] If an industry's code was passed later than the fifteenth of the month, that month was counted as a PRA month. The 'industry code' dummy takes on a value of 1 for each month between code passage and May 1935, when the NIRA was ruled unconstitutional under the *Schechter* decision.

Note that these dummy variables account for the time periods in which firms within each industry were supposed to comply with either the PRA or industry-specific codes. However, past studies (e.g. Brand 1988; Alexander 1994, 1997; Taylor 2002; Taylor and Klein 2008) have demonstrated that non-compliance was a major issue under the NIRA, particularly after the first few months. This adds to the importance of examining the empirical effects of the PRA. If the coefficients on the PRA dummy variables are small or insignificant, this would suggest that non-compliance was an issue from the start of the NIRA. On the other hand, if I find statistically different and large movements in economic variables during the months when industries were covered by the PRA, this would suggest that compliance was widely maintained, at least initially. This would be consistent with the 'compliance crisis' literature which suggests that firms complied with the NIRA rules at the outset, but later defected, causing many of the industry cartels to break down.

My sample consists of industries covered by an NIRA 'code of fair competition' whose employment, payroll, work hours, output or prices were reported by month in the *National Bureau of Economic Research Macrohistory Database* for the time period January 1927 to December 1937. The 1927 start point is largely due to a lack of monthly data prior to this time, and the end date is chosen to avoid any movements in the data caused by Second World War production. These data generally came, in roughly equal proportions, from two sources: the National Industrial Conference Board and the Bureau of Labor Statistics. The data cover mainly large industries, so my sample covers a significant percentage of the NIRA-covered workforce. I have an index of employment (number of workers on payrolls) for 28 industries. Because I employ log differences (growth rates) in my empirical analysis, the employment index behaves identically to a measure of raw employment numbers. Fifteen of these industries also report average hours of work per week, allowing me to examine percentage changes in *aggregate employment hours for labour as a whole* within each industry (growth rate in employment index multiplied by average hours per week) rather than simply growth in number of workers on the industry's payroll (jobs). Price data are reported for 38 industries, and output data are reported for 66 industries that were covered by the NIRA.

Table 1 lists the 75 industries in the sample along with the date the 'code of fair competition' that covered each industry was passed and a column describing what data I have for each industry. The median date of industry-specific code passage in the sample was 17 November 1933, meaning that the median firm was covered by the PRA blanket code for around 4 months (the average in the sample was close to 5 months) and by the NIRA specific code for 18 months, although, as Table 1 demonstrates, there was a great deal of variability in code passage dates by industry. Since employment is the primary variable of interest, Table 2 reports the annual growth rates in employment for the 28 industries for which I have this data. (Monthly rates are not reported, in the interests of saving space.) Finally, Figure 2 shows movements in indexes of manufacturing sector output and real wage rates, as well as the money supply across the sampled time period.

IV. Empirical Results

Although I report specifications with subsets of these independent variables and employ alternate dependent variables, the general model that I estimate is

$$\begin{aligned}\text{GrthEMPLOYMENT}_{it} = \beta_0 + \beta_1\text{PRA}_{it} + \beta_2\text{NIRACODE}_{it} + \beta_3\text{GrthWAGERATE}_{it} \\ + \beta_4\text{GrthMONEY}_t + \beta_5\text{GrthGOVSPD}_t + \beta_6\text{GrthGOVREV}_t \\ + \beta_7\text{NLRA}_t + \beta_8\text{TIMETREND}_t + \varepsilon_{it},\end{aligned}$$

where

- $\text{GrthEMPLOYMENT}_{it}$ is the growth rate in the number of wage earners employed in industry i at time t
- PRA_{it} is a dummy variable equal to 1 for months during which industry i was covered by the President's Reemployment Agreement blanket code of labour provisions
- NIRACODE_{it} is a dummy variable equal to 1 for months during which industry i was covered by an industry-specific NIRA cartel code of fair competition
- GrthWAGERATE_{it} is the growth rate average real hourly wage rate for industry i in month t[10]
- GrthMONEY_t is the growth rate of the money supply in month t
- GrthGOVSPD_t is the growth rate in real government spending in month t
- GrthGOVREV_t is the growth rate in real government revenues collected in month t
- NLRA_t is a dummy variable equal to 1 for months after April 1937
- TIMETREND_t is a monthly time dummy which rises by one unit per month.

I also employ 11 month dummies (January is the base month) to control for seasonality. Although other specifications and methods (such as period fixed effects, discussed later) are employed, I primarily estimate generalized least squares (GLS) panel regressions using industry fixed effects. In recognition that my control variables may affect each industry differently, I estimate the model using cross-section-specific coefficients for each independent variable that does not vary by industry (e.g. money supply, government spending, month dummies)—this creates a separate coefficient for each industry for these variables. I also employ cross-section weights so as to estimate feasible GLS specifications assuming the presence of cross-section heteroscedasticity. Standard errors are calculated using a White period method which is robust to arbitrary serial correlation and time-varying variances in the disturbances. The major results of the paper do not change if I use an ordinary coefficient covariance method. Finally, I use growth rates in all non-dummy variables both to help control for serial correlation, which is strongly present when these regressions are run in log levels, and to allow the expansion of my data set to include both indexed and raw data.

Specifications (1) and (4) of Table 3 report the results of extremely parsimonious regressions of the NIRA's impact on employment measured both as *number of workers on payroll* (i.e. jobs) in specification (1) and as *aggregate employment hours* worked in the economy in specification (4). These regressions show that the growth in the number of workers on firms' payrolls rose when covered by both the PRA and the NIRA industry-specific cartel codes; however, growth in aggregate employment hours of labour was not significantly different during these months. In terms of magnitude, specification (1) suggests that during the months when firms were covered by the PRA, the number of workers with industrial employment grew 1.65% faster than otherwise. Since, on average, a firm in the 28-industry sample was covered by the PRA for 4.43 months, this translates into a 7.5%

TABLE 1
DATE OF INDUSTRY-SPECIFIC CODE PASSAGE FOR INDUSTRIES IN DATASET, AND THE DATA SERIES FOR EACH INDUSTRY

Industry	Code passage	Data	Industry	Code passage	Data
Alcohol	21 Aug 1934	Q, P	Lumber	19 Aug 1933	Q, P, E
American cheese	2 Feb 1935	Q	Machinery	17 Mar 1934	Q, E, W, H
Asphalt	6 Nov 1933	Q, P	Meat	4 Jan 1934	Q, E, W, H, P
Auto parts	8 Nov 1933	Q, P	Men's clothes	26 Aug 1933	E
Baking	28 May 1934	E	Men's shoes	3 Oct 1933	Q
Beef and veal	4 Jan 1934	Q, P	Mercht. pig iron	19 Aug 1933	Q, P
Bituminous coal	18 Sep 1933	Q, P	Metal	2 Nov 1933	Q, P
Books	17 Feb 1934	Q	Milk, NY	4 Jan 1934	Q
Bricks	26 Mar 1934	Q, P	Newsprint	17 Nov 1933	Q, P
Butter	4 Jan 1934	Q, P	Paper and pulp	17 Nov 1933	Q
Cement	27 Nov 1933	Q, E	Paper production	17 Nov 1933	Q, E, W, H, P
Cheese	2 Feb 1935	Q, P	Passenger cars	26 Aug 1933	Q, E, W, H, P
Chemicals	10 Feb 1934	Q, E, W, H, P	Pig iron	19 Aug 1933	Q, W, H, E
Condensed milk	4 Jan 1934	Q, P	Pork	4 Jan 1934	Q, P
Construction	31 Jan 1934	Q	Raw silk	7 Oct 1933	Q, E, W, H, P
Copper	21 Apr 1934	Q, P	Rayon	26 Aug 1933	Q, E, W, H
Corn grinds	4 Jan 1934	Q, P	Rayon yarn	26 Aug 1933	Q, P
Cotton	9 Jul 1933	Q, P, W	Refined lead	24 May 1934	Q, P
Cotton goods	17 Nov 1933	Q, E	Rice	4 Jan 1934	Q
Crude petroleum, appl.	19 Aug 1933	Q, P	Rubber	15 Dec 1933	Q, E, W, H, P
Crude petroleum, RTS	19 Aug 1933	Q	Slab zinc	26 Mar 1935	Q, P
Douglas fir lumber	19 Aug 1933	Q, P	Small cigarettes	9 Feb 1935	Q, E
Electrical manufacturing	4 Aug 1933	W, H	S. pine lumber	19 Aug 1933	Q, P

Evaporated milk	4 Jan 1934	Q, P	Steel ingots	19 Aug 1933	Q
Fertilizers	31 Oct 1933	Q	Steel sheets	19 Aug 1933	Q, E, W, H, P
Fine paper	17 Nov 1933	Q	Total shoes	3 Oct 1933	Q, P, W, H, E
Furniture	7 Dec 1933	E, W, H	Trucks	26 Aug 1933	Q
Glass	3 Oct 1933	Q, E, P	Tyre pne. casings	21 Dec 1933	Q
Hardware	2 Nov 1933	E	Tyre tubes	21 Dec 1933	Q
Hats	5 Feb 1934	E	Tyres	21 Dec 1933	Q, W
Ice cream	4 Jan 1934	Q	Wheat flour	4 Jan 1934	Q, P
Knit outerwear	18 Dec 1933	E	Women's clothing	31 Oct 1933	E
Large cigars	19 Jun 1934	Q, E	Women's shoes	3 Oct 1933	Q
Laundry	16 Feb 1934	E	Woodwork machinery	14 May 1934	Q, E, W, H
Lead ore	24 May 1934	Q, P	Wool	26 Jul 1933	Q, E, W, H, P
Leather	7 Sep 1933	Q, E, W, H, P	Wrapping paper	17 Nov 1933	Q
Locomotives	16 Feb 1934	Q	Zinc ore	26 Mar 1935	Q
Lubricants	19 Aug 1933	Q, P			

Notes
Median date of code passage is 17 November 1933. The 'Data' column denotes which data series I have for each industry, where Q is output, E is employment and payroll, W is hourly wage rate, H is average workweek, P is prices.

TABLE 2
ANNUAL EMPLOYMENT (NUMBER OF WORKERS ON PAYROLL) GROWTH RATES FOR SAMPLED INDUSTRIES, 1927–1937

Industry	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937
Baking	4.26	9.60	6.76	− 8.12	− 5.39	− 2.15	14.42	5.30	6.09	6.26	− 0.14
Cement	− 12.24	− 0.47	− 11.21	− 16.97	− 23.14	− 29.90	14.71	10.76	− 1.37	43.85	− 2.08
Chemicals	3.43	7.43	− 0.84	− 12.16	− 0.75	− 4.22	27.77	3.99	− 4.33	9.75	− 1.29
Cotton goods	− 2.48	− 3.32	− 7.98	− 20.09	− 1.92	1.26	28.77	− 3.33	− 4.77	14.22	− 12.40
Furniture	− 2.84	3.11	− 8.14	− 25.20	− 13.58	− 21.42	17.06	12.21	13.68	26.00	− 12.71
Glass	− 7.16	6.07	− 2.75	− 22.08	− 11.48	− 10.34	56.41	2.84	7.74	2.75	1.34
Hats	5.52	− 5.96	− 6.44	− 13.03	1.32	− 19.53	4.26	0.28	6.33	10.45	5.23
Hardware	− 6.78	11.07	0.81	− 23.48	− 16.19	− 11.06	35.90	0.64	11.91	1.60	− 10.57
Knit outerwear	0.59	0.78	− 2.02	− 6.69	8.64	− 1.65	3.16	0.38	2.19	− 16.22	− 0.72
Large cigars	5.36	− 9.22	− 4.78	− 9.19	− 6.21	− 8.78	− 0.63	7.89	− 5.99	5.44	− 3.24
Laundry	4.33	3.67	3.07	− 3.97	− 4.14	− 10.30	1.64	1.38	5.20	12.21	− 0.36
Leather	− 1.41	− 8.37	− 0.33	− 13.30	− 8.63	− 0.85	29.02	1.20	3.49	0.84	− 19.48
Lumber	− 6.47	3.85	− 6.56	− 29.56	− 22.67	− 9.15	24.49	9.78	14.10	11.63	− 12.75
Machinery	− 13.05	18.36	− 1.15	− 31.75	− 30.09	− 17.86	43.48	6.14	22.46	16.63	7.99
Meat	− 1.85	5.97	− 1.88	− 8.76	− 6.84	− 6.52	26.49	2.35	− 6.95	12.93	− 7.96
Men's clothes	7.76	− 3.87	1.31	− 14.33	− 1.40	− 6.56	4.57	10.07	4.43	21.02	− 13.61
Paper production	− 3.75	3.10	5.14	− 15.22	− 7.62	− 7.77	26.18	4.54	1.06	6.12	− 3.00
Passenger cars	5.02	29.44	− 30.13	− 13.39	0.14	− 18.70	41.81	25.73	7.48	6.37	− 20.62
Pig iron	− 6.55	10.13	− 3.42	− 21.58	− 18.35	− 14.87	38.10	10.16	12.94	20.13	− 9.73
Raw silk	− 1.47	− 0.20	0.80	− 8.23	− 7.24	− 4.07	9.71	14.56	− 8.41	1.34	− 21.41
Rayon	− 1.47	− 0.20	0.80	− 8.23	− 7.24	− 4.07	9.71	14.56	− 8.41	1.34	− 21.41
Rubber	2.56	4.99	− 13.35	− 20.25	− 10.45	− 3.69	32.21	− 2.38	− 1.66	15.10	− 10.75
Small cigarettes	5.36	− 9.22	− 4.78	− 9.19	− 6.21	− 8.78	− 0.63	7.89	− 5.99	5.44	− 3.24
Steel sheets	− 6.55	10.13	− 3.42	− 21.58	− 18.35	− 14.87	38.10	10.16	12.94	20.13	− 9.73
Total shoes	− 4.29	− 0.43	2.36	− 14.75	1.10	0.73	5.42	8.47	0.63	8.94	− 9.14
Women's clothing	8.04	10.58	3.12	− 6.32	− 11.81	− 2.75	7.27	22.50	9.21	11.77	− 13.36
Woodwork machinery	− 13.05	18.36	− 1.15	− 31.75	− 30.09	− 17.86	43.48	6.14	22.46	16.63	7.99
Wool	− 11.58	− 1.13	− 0.89	− 21.85	− 14.64	10.79	25.57	2.83	17.50	4.58	− 27.09

Notes
Growth rates are calculated from January to January using the monthly dataset described in the text. Monthly growth rates are not reported for each industry, in the interest of saving space.

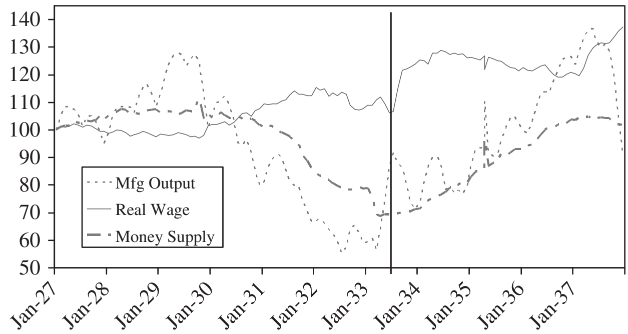

FIGURE 2. Monthly movements in manufacturing output, real wage rates and the money supply.

Notes: Solid line at August 1933 denotes passage of the NIRA. Although my empirical analysis employs industry-level data on real wage rates and output, the data above are aggregate data. Sources are 'US index of production in manufacturing' (NBER Series 01175), 'Money stock, commercial banks plus currency held by public' (NBER Series 14144), 'Average hourly earnings for twenty-five manufacturing industries' (NBER Series 0812) and 'Index of general price level' (NBER Series 04051).

increase in the number of industrial jobs brought about by the PRA. Given that the NIRA covered an estimated 22 million workers at its midpoint, the regression coefficients suggest that 1.65 million more Americans had jobs because of the PRA's labour provisions. This is relatively close to the 1.75 million estimate for job growth in the fall of 1933 provided by Lyon *et al.* in their 1935 appraisal of the NIRA, which was based on a straightforward examination of BLS industry-level data comparing average employment in April–June 1933 to that of September–December 1933. This non-*ceteris-paribus* analysis can be seen visually in Figures 3 and 4, which show a pronounced jump in jobs, but little movement (and initially a decline) in aggregate employment hours after passage of the NIRA.

To go beyond the parsimonious analysis and better estimate the effect that the legislation had on employment, I must control for other important factors that affected movements in employment during this time. For example, the money supply increased by 24.9% between the June 1933 passage of the NIRA and the May 1935 ending of the legislation. Likewise, government spending rose much faster than revenues during these months. To control for changes in fiscal and monetary policy unrelated to the NIRA, or its subprogramme the PRA, I include monthly measures of the money supply, real government spending and real government revenues.[11] I also include a time trend dummy since employment may rise over time due simply to population growth rather than economic conditions, as well as month dummies to control for seasonality, although my results are not generally sensitive to the inclusion of either of these controls. Finally, I include a dummy variable accounting for the months after the Supreme Court upheld the National Labor Relations Act in April 1937, since this legislation also impacted the way my variables interacted during the sample time period, although again, the results are not sensitive to its inclusion.

Specifications (2) and (5) (Table 3) include these six control variables as cross-section-specific coefficients. This estimation method is less restrictive than a common coefficient one as it allows these control variables to interact differently with each industry. I find that during months when firms were covered by the President's Reemployment Agreement, the number of workers with industrial employment grew 1.35% faster than

TABLE 3

INDUSTRY FIXED EFFECTS PANEL ESTIMATION OF THE EFFECTS OF THE PRA ON EMPLOYMENT GROWTH

	Dependent variables (log differences) Sample: January 1927 to December 1937					
	Industrial employment (number of workers on payroll)			Aggregate employment hours (number of workers × hours per week)		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	− 0.00178	− 0.00089	− 0.00036	− 0.00362	− 0.00100	− 0.00028
	(− 9.28)**	(− 4.62)**	(− 1.39)	(− 8.61)**	(− 2.16)*	(− 0.65)
PRA dummy (Aug 1933 to code passage)	0.01645	0.01345	0.03302	− 0.00748	− 0.01987	0.00375
	(5.15)**	(4.16)**	(9.63)**	(− 0.66)	(− 2.07)*	(0.47)
NIRA code dummy (code passage to May 1935)	0.00361	− 0.0032	− 0.00436	0.00361	− 0.01565	− 0.01187
	(4.00)*	(− 2.76)**	(− 2.86)**	(1.75)	(− 4.56)**	(− 3.74)**
Log difference industry wage rate			− 0.14994			− 0.69023
			(− 2.51)*			(− 4.79)**
Includes cross-section-specific coefficients on						
Dummies for each month	No	Yes	Yes	No	Yes	Yes
Log diff. money supply	No	Yes	Yes	No	Yes	Yes
Log diff. real government spending	No	Yes	Yes	No	Yes	Yes
Log diff. real government revenue	No	Yes	Yes	No	Yes	Yes
Time trend	No	Yes	Yes	No	Yes	Yes
NLRA dummy	No	Yes	Yes	No	Yes	Yes
Number of cross-sections	28	28	15	15	15	15
Number of observations	3668	3668	1965	1965	1965	1965
R^2	0.395	0.438	0.350	0.234	0.320	0.356

Notes

Cross-section-specific coefficients (variables in italics) and industry intercepts are not reported in the interest of saving space. I report *t*-statistics computed from White period standard errors which are robust to arbitrary serial correlation and time-varying variances in the disturbance.

*Significant at 5% level; **significant at 1% level.

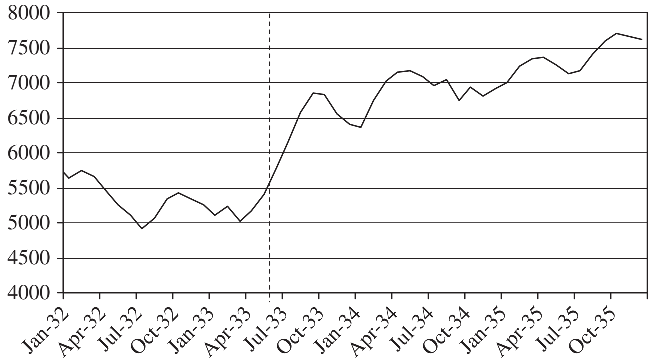

FIGURE 3. Number of workers on payrolls (thousands), manufacturing, 1932–1935.

Notes: Dashed line coincides with passage of the NIRA. Source is 'Production worker employment, manufacturing total' (NBER Series 8010b).

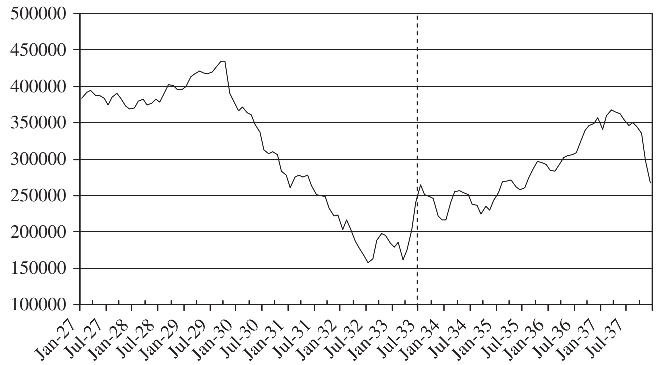

FIGURE 4. Aggregate employment hours (thousands) in manufacturing, 1927–1937 (number of workers × average workweek).

Notes: Dashed line coincides with passage of the NIRA. Series is derived by multiplying 'US average hours of work per week, manufacturing total' (NBER Series 8029a) by 'Production worker employment, manufacturing total' (NBER Series 8010b).

otherwise, *ceteris paribus*, which translates to 1.34 million additional jobs created by the PRA's labour provisions. Interestingly, it is clear that work-sharing (rather than aggregate employment creation) was the sole basis for this increase in the number of workers on payrolls. In terms of growth in aggregate hours worked (specification (5)), PRA months actually saw this measure fall nearly by 2% per month, translating to a 9.1% drop over the 4.43 months when the average industry was covered by the PRA.[12]

Furthermore, once the industry-specific NIRA cartel codes were passed, the employment picture became quite bleak. The coefficient on the NIRA code dummy in specification (2) shows that cartelized firms hired 0.32% fewer workers per month, which translates to a loss of 1.26 million jobs over the 17.5 months when a typical firm was covered by an industry-specific code. Aggregating the total effect of the NIRA from

August 1933 to May 1935, specification (2) suggests a net effect of only around 80,000 jobs created—1,340,000 gained under the PRA and 1,260,000 lost under the industry-specific cartel codes. Specification (5) shows a still bleaker picture with respect to aggregate hours worked once the industry-specific codes were passed, as aggregate employment growth fell by 1.57% per month, *ceteris paribus*. Since the average industry in my 15-industry sample was covered by its specific code for 18.73 months, this translates to a 33.6% drop in aggregate hours worked when firms were covered by NIRA industry-specific codes, *ceteris paribus*. Figure 4 reveals a drop in the raw data on aggregate hours worked, but not of such a large magnitude. My results suggest that the fiscal and monetary stimulus of the period offset much, but not all, of the negative effect that cartelization and high wage rates had on aggregate hours worked.

The finding of a strongly negative aggregate hours worked effect under the industry-specific NIRA cartel codes is consistent with the contemporary conclusions of Lyon *et al.* (1972, p. 844), who pointed out the negative impact that cartelization and higher wage rates likely have on employment: 'It is our view that the NRA has had the effect of restricting production below the levels it would otherwise have attained, hence that it has reduced the total amount of employment as measured by the number of man hours of work done. . . . Merely dividing a smaller amount of work among more workers is neither recovery nor a good substitute for it.' The finding is also broadly consistent with more recent empirical studies of the NIRA's macroeconomic effects such as Weinstein (1980), Vedder and Gallaway (1997), Taylor (2002) and Cole and Ohanian (2004). Finally, the finding of a modestly positive *net* impact of the NIRA on employment—80,000 jobs—is in line with Bernanke (1986, p. 99) who used an eight-industry sample and found that 'the NRA tended to increase employment and reduce hours', although he concludes that the legislation's overall effects on employment were small.

It would be interesting to examine the effects of the PRA's 'work-sharing' provisions *themselves* on employment. As previously mentioned, 15 of the 28 industries for which employment data are reported also report hourly wage rates. If one includes industry hourly wage rate as an independent variable along with the PRA and NIRA code dummy variables in the same regression, the policy dummies take on new meaning—the coefficients can be interpreted as *the effect of the policy independent of its mandated wage rate increases*. In essence this can help to separate out the effects of a pure work-sharing policy via workweek reduction only. In specification (3), I hold wage rates constant and find that the PRA's non-wage share-the-work provisions increased job growth, as measured by the number of workers on payrolls, by 3.3% per month. Since the 15 industries in this sample were covered by the PRA for an average of 3.27 months, this translates to 2.47 million new employment opportunities *attributable to the PRA's work-sharing provisions themselves*, *ceteris paribus*. With 11,086,000 Americans reported as unemployed at the beginning of the PRA, this estimate would have represented an astounding 22.2% drop in unemployment in less than four months.[13]

Combining the information from specification (3) with that from specification (2), which suggested 1.34 million net jobs gained, of the 2.47 million additional jobs attributable to work-sharing, over 1.1 million were effectively offset by the workweek reduction's concurrent requirement that firms pay higher hourly wage rates. Hence the total effect of the PRA (the 1.34 million jobs gained) was to reduce the ranks of unemployed by around 12.1%, *ceteris paribus*. This estimate should be viewed as an upper bound since, from a general equilibrium framework, some of the additional jobs created in NIRA-covered industries may have come at the expense of jobs in non-covered ones such as agriculture. Still, it appears that the PRA was worthy of the massive, and largely positive, media attention given it in the late summer and early autumn of 1933. In terms of raw National Industrial Conference Board (NICB) unemployment data, the quantity of unemployed workers fell from 11,086,000 in July 1933 to 9,206,000 in October 1933, before rising again beginning in November, when industry-specific NIRA codes were being enacted *en masse*. While some of the nearly 1.9 million person drop in unemployment can be attributed to fiscal and monetary expansion during these months, my results suggest that the majority of employment gains are attributable to the PRA's work-sharing provisions. Still, my regressions suggest that the employment gains could have been much larger had it not been for the concurrent increases in hourly wage rates mandated by the programme.

With regard to aggregate hours worked, all 15 industries for which I have data on employment and hours of work also report wage rate data, so the sample size is not diminished by holding wage rates constant. The results reported in specification (6) suggest that the PRA's non-wage (i.e. workweek reduction) provisions had no significant impact on aggregate hours worked, although the sign of the coefficient switches from negative to positive. This squares with expectations of such a policy—the idea of work-sharing is to spread existing work hours to more people rather than to create new employment. Interestingly, specifications (3) and (6) suggest that the drop in both measures of employment under the industry-specific NIRA codes was caused primarily by cartelization (rather than wage increases), since the employment growth declines are significant even holding wage rates constant, and the coefficient on the NIRA dummy does not decline much (around 24%) between specifications (5) and (6). In summary, the results reported in Table 3 suggest that while PRA work-sharing (i.e. a reduction in hourly workweeks) promoted reemployment by putting more workers on company payrolls, growth in aggregate hours worked actually fell under the PRA and continued to do so when firms became covered by their industry-specific cartel codes.

The finding of a relatively large positive impact on job creation under the PRA's work-sharing provisions may be viewed as evidence that workweek reductions today could help to alleviate unemployment, as has been the goal of more recent policies in Europe. A key difference, however, between the PRA and these programmes is the use of an overtime payment scheme. The PRA essentially placed a hard cap on work hours, while most work-sharing programmes today allow firms to hire workers for hours beyond the cap but require a premium, typically between 25% and 50%, to be paid for overtime hours. In fact, the assumption of overtime provisions is a key factor driving many of the theoretical predictions that work-sharing policies may harm employment. Additionally, it should be noted that my analysis of the PRA is over an extremely short time period. In the long run, the PRA's work-sharing restrictions may, indeed, have caused firms to shift from labour to capital and harmed employment, again as predicted by some models. A final difference to consider is that the PRA work-sharing was implemented to cure a large measure of cyclical unemployment, while recent European programmes primarily target structural unemployment.

Non-employment economic effects of the PRA

While reemployment was clearly the primary goal of the PRA, it may be interesting to briefly examine how the programme affected other economic variables. Table 4 reports the results of regressions examining the effects of both the the PRA and the industry-specific NIRA codes on real industry payroll (the real wage bill to labour), real take-home pay for the average worker, industry output and industry prices. The regressions

TABLE 4
INDUSTRY FIXED EFFECTS PANEL ESTIMATION OF THE EFFECTS OF THE PRA ON FOUR ECONOMIC VARIABLES

	Dependent variables (log differences) Sample: January 1927 to December 1937			
	Industry real payroll (1)	Take-home pay for average worker (2)	Industry output (3)	Industry prices (4)
Constant	– 0.00171	0.00043	– 0.03090	– 0.00212
	(– 5.98)**	(2.01)*	(– 3.54)**	(– 11.53)**
PRA dummy (Aug 1933 to code passage)	0.00908	– 0.01128	– 0.02359	0.00845
	(4.84)**	(– 2.32)*	(– 3.44)**	(3.23)**
NIRA code dummy (code passage to May 1935)	– 0.00298	– 0.00453	– 0.01392	– 0.00119
	(– 1.40)	(– 2.97)**	(– 5.96)**	(– 0.94)
Includes cross-section-specific coefficients on				
Dummies for each month	Yes	Yes	Yes#	Yes
Log diff. money supply	Yes	Yes	Yes#	Yes
Log diff. real government spending	Yes	Yes	Yes#	Yes
Log diff. real government revenue	Yes	Yes	Yes#	Yes
Time trend	Yes	Yes	Yes#	Yes
NLRA dummy	Yes	Yes	Yes#	Yes
Number of cross-sections	28	16	66	38
Number of observations	3668	2096	8372	4978
R^2	0.523	0.230	0.099	0.179

Notes
See Table 3 notes.
#With 66 cross-sections, I was unable to compute cross-section-specific coefficients for these variables, so I estimated common coefficients in specification (3).
*Significant at 5% level; **significant at 1% level.

duplicate those from specifications (2) and (5) of Table 3, employing the money supply, real government spending, real government revenue, NLRA dummy, month dummies and a time trend as independent variables.

In specification (1), the sample includes 28 industries for which aggregate payroll data are reported. Like the employment data, the payroll measure is an index (1923–1925 = 100); however, since I use monthly growth rates, it makes no difference whether I have indices or raw numbers. Again, I divide these payroll data by a macroeconomic price index to convert to real values. A stated goal of the Roosevelt administration was to increase the 'purchasing power' of labour. I find that the PRA did increase the total real wage bill to labour. Payrolls grew 0.91% per month faster under the PRA, *ceteris paribus*, translating into a 4.1% growth in real income to labour for a typical industry in the sample covered by this programme for 4.4 months, although, again, this varied from industry to industry since some were covered by the PRA for far longer than others.

Specification (2) examines the growth rate of real take-home pay of an average worker by multiplying the industry's average hourly nominal wage rate by the industry's average workweek for the 16 industries for which data are reported for these two variables—again, I divide this nominal measure by the price level to obtain real take-home pay. I find that real take-home pay fell for the typical worker under both the PRA and the industry-specific codes. Specifically, under the PRA, take-home pay growth fell by 1.1% per month—since industries in this subsample were covered by the PRA for an average of 3.3 months, this translates to a total drop in real take-home pay of 3.8%, *ceteris paribus*. Thus it appears that the increases in hourly wage rates (and prices) under the PRA were not enough to overcome the shortening of the workweeks to 'fatten pay envelopes', or even keep them steady as the Roosevelt administration had suggested was a goal. The work-sharing provisions, then, brought an income burden on those who had a job and kept it throughout, but, as shown in Table 3, created more opportunities for employment.

Specifications (3) and (4) examine the PRA's effect on industry output and prices, respectively. While the Roosevelt administration suggested that higher hourly wage rates would create more aggregate demand, which should increase output and prices, neoclassical economic theory suggests that higher labour costs will cause profit-maximizing firms to reduce output (and employment) and, generally, raise prices, as would be consistent with a negative supply shock. I find evidence supporting neoclassical theory rather than the high-wage doctrine. Under the PRA, output growth was 2.4% per month lower, *ceteris paribus*. Since the average industry in this sample was covered by the PRA for 4.98 months, this translates into a 12.3% drop in industry output across the PRA. Not surprisingly, output growth continued to fall under the industry-specific cartel codes—around 26.5% in total across the 17 months. Additionally, I find that prices rose by 0.8% per month under the PRA. In this case the 38 industries in the sample were covered by the PRA for an average of 3.92 months, so this translates to a 3.4% rise in prices under the programme. I do not, however, find any significant impact on growth of industry prices under the NIRA codes. This suggests that the majority of the price increases brought about by the NIRA, as documented by Romer (1999) and Weinstein (1980), occurred when firms were covered by the PRA blanket code rather than by the NIRA cartel codes themselves.

I also ran a two-stage least squares specification because of the potential endogeneity of prices and output. The results of specifications (3) and (4) are unchanged, except that the magnitude of the output decline under the PRA rises, and there is now a statistically significant increase in prices under the industry-specific NIRA codes. As another check, I ran specification (3) with lagged industry prices and specification (4) with lagged industry output. Here the results were largely unchanged from those reported in Table 4.

V. Robustness Checks

Potential impact analysis

Manufacturing sector level data clearly show that a pronounced increase in the average hourly wage rate and a drop in the average hourly workweek accompanied the PRA in August 1933. Specifically, between July and September 1933, the average hourly wage rate rose from 45.6 cents to 53.6 cents, while the average hourly workweek fell from 42.9 hours to 36.3 hours.[14] Of course, the correlation could be spurious if factors other than the PRA are driving these movements. If such variables are omitted from the regressions, the PRA and CODE dummy variables may be picking up effects not related to the legislation. Given the plethora of policy changes and exogenous events during the New

TABLE 5
AVERAGE HOURLY EARNINGS (CENTS PER HOUR) AND WORKWEEKS IN MAY 1933 AND MAY 1934 IN 18 (AND 16) INDUSTRIES

Industry	Hourly earnings May 1933	Hourly earnings May 1934	Workweek May 1933	Workweek May 1934
Chemicals	45.7	56.3	40.1	38.4
Cotton	29.7	44.4		
Electrical manufacturing	53.5	65.5	35.5	34.6
Furniture	37.9	53	33.2	32.9
Leather	40.6	55.2	45	37.3
Machinery	54.5	62.2	31.8	37.9
Meat	39.5	52.4	49.3	40.2
Paper production	41.5	50.4	41.2	37.7
Passenger cars	56.7	72.4	33.4	31.9
Pig iron	47.7	64.6	35.4	36.6
Raw silk	33.4	49.9	40.9	29
Rayon	33.4	49.9	40.9	29
Rubber	55.8	75.3	34.7	33.1
Steel sheets	47.7	64.6	35.4	36.6
Total shoes	40.3	56.6	40.5	37.4
Tyres	55.8	75.3		
Woodwork machinery	54.5	62.2	31.8	37.9
Wool	34.3	51.9	41.6	32.8

Deal, empirical studies of the time period are particularly susceptible to such a bias. In this section I offer a 'potential impact analysis' as is common in the policy evaluation literature, to support to the notion that the PRA affected firm behaviour with respect to wage rates and hours, and in doing so ultimately affected employment and the other dependent variables that I employ in Tables 3 and 4.

Since the PRA set specific guideposts for wage rates (40 cents) and workweeks (35 hours), I can analyse the extent that each industry differed from these guideposts *prior to* the policy change to gauge the potential impact that the legislation would have had on each industry. For example, the PRA's minimum wage guidepost would potentially have a large impact on the silk industry, whose May 1933 average hourly wage rate was 33.4 cents, but would have a much smaller potential impact on the automobile industry, since its average wage rate of 56.7 cents was well above the guidepost. Likewise, the workweek guidepost would potentially have a high impact on the meat industry, whose average workweek was nearly 50 hours, but would have a low potential impact on the electrical manufacturing industry, whose average workweek in May 1933 was 35.5 hours. Table 5 reports the average hourly earnings and workweeks for the industries in my dataset for May 1933 and May 1934. I chose May 1933 since the NIRA was passed in June—this helps to better measure the impact of the PRA and the industry codes by examining their values before any policy effects could have been experienced—and compare this to May 1934 to minimize seasonal factors. A cursory examination of this table shows that hourly earnings rose in every industry, and workweeks fell in most industries.

The key question is whether the lowest-wage industries saw the largest gain, and whether the highest-workweek industries saw the largest drops. To examine this empirically, I create a series of industry-level dummy variable representing high, medium and low wage rates and workweeks. Specifically, for the variable High Wage Industry, I assign an industry a 1 if the reported wage rate was above 50 cents per hour in May 1933, and 0 otherwise. The Medium Wage Industry dummy variable likewise denotes industries whose wage rates were between 35 and 50 cents per hour in May 1933. The base group consists of Low Wage Industries—i.e. those with hourly wage rates below 35 cents an hour in May 1933. With respect to workweeks, the dummy High Hour Industry takes on a value of 1 for those industries where workweeks were greater than 45 hours in May 1933. The Medium Hour Industry likewise represents those industries whose workweeks were between 38 and 45 hours. The workweek base group is Low Hour Industries, which are industries with lower than a 38-hour workweek in May 1933.

I duplicate the analysis in Table 3, but employ hourly wage rates and hours of workweek as the dependent variables. I add four interaction dummies which multiply the PRA and NIRA code dummies by the High and Medium impact dummies described above. (High and Medium Wage Industry dummies are employed in specification (1), while High and Medium Hour Industry dummies are employed in specification (2).) If the legislation itself drove changes in these variables, the coefficients on the interactions should be significant. Specifically, one would expect to find that High and Medium Wage industries would see *less of an increase* in wage rates caused by the PRA's 40 cents per hour guidepost than would the Low Wage industries (a negative coefficient on these terms). Furthermore, one would expect to find that the High and Medium Hour industries would have seen *a larger drop* in hourly workweek than would Low Hour industries (again a negative coefficient on the interaction terms). If, on the contrary, the rise in wage rates and cuts in workweeks documented above were caused by, say, a broad change in aggregate demand via increases in consumer confidence or some other measure that would have generally affected industries equally (or randomly), then the coefficient on the interaction dummies should be insignificant.

The results reported in Table 6 offer strong empirical support that the PRA and NIRA affected firm behaviour with respect to wage rates and workweeks. The results reported in specification (1) suggest that while wage rates grew by 18.1% per month under the PRA, *ceteris paribus*, wage rates grew 16.1% *less quickly* in High Wage industries than in Low Wage ones. Furthermore, wage rates grew 13% less quickly in Medium Wage industries than in Low Wage ones. Finally, wage rates grew 3.1% faster (16.1–13) in Medium Wage industries than in High Wage ones. With respect to movements in hours of work per week, specification (2) suggests that while workweeks fell by 1.96% per month under the PRA, High Hour industries saw workweeks fall 4.5% faster than the Low Hour base group industries. Furthermore, Medium Hour industries saw workweeks fall 3.6% faster than Low Hour ones. Finally, High Hour industries saw workweeks fall 0.9% faster (4.5–3.6) than Medium Hour ones. These results offer empirical support to the notion that the PRA itself drove the pronounced movements in wage rates and workweeks that coincided with the legislation. Those industries where the wage and hour provisions had the most potential impact were those that saw the largest increases in wage rates and the largest reduction in workweeks under the PRA and NIRA codes.

Robustness to other specifications

The results of this study are robust to a variety of alternative specifications and methods. For example, an alternative approach to controlling for macroeconomic factors would be

TABLE 6
INDUSTRY FIXED EFFECTS PANEL ESTIMATION PRA POTENTIAL IMPACT ANALYSIS FOR WAGE RATES AND HOURS OF WORK

	Dependent variables (log differences) Sample: January 1927 to December 1937	
	Hourly wage rates (1)	Hours of workweek (2)
Constant	– 0.00063	– 0.00640
	(– 3.64)**	(– 1.69)
PRA dummy (Aug. 1933 to code passage)	0.1806	– 0.01967
	(6.69)**	(– 95.70)**
NIRA code dummy (code passage to May 1935)	0.01493	– 0.00981
	(12.04)**	(– 10.28)**
Interaction PRA impact dummy (PRA dummy × high wage/hour industry)	– 0.16069	– 0.04463
	(– 5.91)**	(– 2.31)*
Interaction PRA impact dummy (PRA dummy × medium wage/hour industry)	– 0.12967	– 0.03604
	(– 4.43)**	(– 2.78)**
Interaction NIRA impact dummy (code dummy × high wage/hour industry)	– 0.00920	– 0.00001
	(– 4.45)**	(– 0.01)
Interaction NIRA impact dummy (code dummy × medium wage/hour industry)	– 0.00922	– 0.00007
	(– 6.20)**	(– 0.21)
Includes cross-section-specific coefficients on dummies for each month	Yes	Yes
Log diff. money supply	Yes	Yes
Log diff. real government spending	Yes	Yes
Log diff. real government revenue	Yes	Yes
Time trend	Yes	Yes
NLRA dummy	Yes	Yes
Cross-sections	18	16
Number of observations	2358	2096
R^2	0.302	0.271

Notes
See Table 3 notes.
*Significant at 5% level; **significant at 1% level.

to employ period fixed or random effects. A potential shortcoming arises, however, in that the PRA dummy activates for every industry at the same time, August 1933. This prevents me from employing period fixed effects along with the PRA and industry-specific code dummies. I am able to employ cross-section fixed effects and period random effects specifications, although the *F*-statistics suggest that the results should be interpreted with caution. These specifications suggest that the PRA created 1.52 million new jobs, versus 1.34 million in specification (2) of Table 3. When real wage rates are held constant, I find that the PRA's non-wage provisions created 2.46 million new jobs, versus the 2.47 million found in specification (3) of Table 3. With respect to the aggregate hours worked, an important difference between these results and those in Table 3 is that the

PRA dummy is positive and significant when wage rates are held constant. Specifically, total hours worked in the manufacturing sector rose 2.9% per month faster during PRA months, which translates to a 9.8% rise in employment hours over the course of the legislation. In the regressions using macroeconomic and monthly control variables, the coefficient was positive but insignificant.

As another robustness check, I created a pre-NIRA dummy variable for the months of June and July 1933 to control for the surge in output that followed just before and immediately after the NIRA was passed, but before the 1 August implementation of the PRA and the subsequent passage of industry-specific codes. I include this dummy in all the regressions from Tables 3, 4 and 6. None of the major results change, and the magnitude of the coefficients are basically unaltered. In fact, I do find that output and employment rose significantly during June and July 1933, which is why I thought such a control variable may be warranted. One possible explanation for this is that firms expanded output in a race for more market share in anticipation of production quotas or other production-oriented restrictions being included in their industry-specific codes. Another possibility is that a wave of consumer confidence caused the surge.

I also experimented with log levels of employment (with a lagged dependent variable) rather than log differences. While these specifications suffer from a serial correlation problem, they confirm the major results. In addition, I tried lagging the fiscal and monetary policy variables under the assumption that such a stimulus may have a delayed effect. Again, the results are largely unchanged. Finally, I tried variations upon the macroeconomic control variables, replacing or supplementing the growth rate of the money supply and government budget variables with the growth rate of an index of business activity, an index of durable output, interest rates and various price indices. Again, the major results are robust to all these variations.

VI. Conclusions

Recent popular press studies by Powell (2003) and Shlaes (2007) embody a recent trend of reexamining the New Deal in a less than flattering light. Roosevelt's call to 'bold, persistent experimentation' clearly resulted in some successful policies as well as some failures—studies that focus on only one of these two aspects can easily paint Roosevelt as either a saviour of capitalism or a leader who stood in way of economic recovery. Consistent with the New Deal as a whole, the President's Reemployment Agreement—a subprogramme of the National Industrial Recovery Act—contained both an economic policy that brought employment relief and one that worked against this goal.

The PRA's 'work-sharing' mandate, accomplished by slashing workweeks, appears by any measure to have been strongly successful with respect to the goal of short-term reemployment. I estimate that, *ceteris paribus*, this policy added 2.47 million workers to private sector payrolls, which would represent a nearly 22% decline in economy-wide unemployment in just a few months. Unfortunately, the PRA's concurrent wage rate increasing mandate offset over 1.1 million of these potential new jobs. Still, in offering an answer to the question embedded in the title of this paper, I conclude that the PRA promoted reemployment in the late summer and early autumn of 1933 when the nation sorely needed it. When I look more broadly at the NIRA as a whole, once industries passed their specific 'codes of fair competition', I find that cartelization had a strongly negative effect on both the number of jobs and the aggregate quantity of hours worked. My analysis of output and prices under the NIRA codes suggests that cartelization brought about a negative supply shock, as would be predicted by microeconomic theory. This is consistent with recent studies such as Cole and Ohanian (2004), which conclude that the legislation as a whole (PRA included) reduced output and retarded recovery. On net, my analysis suggests that 80,000 jobs were created by the PRA/NIRA between August 1933 and May 1935, but the legislation brought a large drop in aggregate hours worked sector-wide.

Although caution should be observed when applying the experiences of certain countries or time periods to others, perhaps the main lesson that the New Deal work-sharing experiment can provide to the modern policy debate is as follows. Work-sharing, through mandated shorter workweeks, can be an effective short-run tool in combating major episodes of cyclical unemployment. However, while work-sharing may enhance employment opportunities, the political compromises that may be required to enact these provisions may impose significant costs from a macroeconomic perspective.

ACKNOWLEDGMENTS

I thank the editor, three anonymous referees and Burak Dolar, as well as participants at the Economic and Business Historical Society Conference and other various workshops, for their comments and suggestions.

NOTES

1. In a radio address on 24 July 1933, Roosevelt stated: 'If all employers in each competitive group agree to pay their workers the same wages—reasonable wages—and require the same hours—reasonable hours, then higher wages and shorter hours will hurt no employer. Moreover [it] makes more buyers for his product' (*New York Times*, 25 July 1933, p. 2).
2. In the recent French 'work-sharing' Aubry Act, firms were given an incentive to comply with the 35-hour week via reductions in social security contributions for firms that effectively reduced the workweek and guaranteed a certain level of employment. See Logeay and Schreiber (2006) for more institutional details on French work-sharing.
3. For example, the *Midland Republican* (Michigan) printed the names of all 88 businesses in Midland who were on the Honor Roll on 17 August 1933. As a newspaper serving a much larger metro area, the *Lansing State Journal*, for example, generally published firms 'coming under the provisions of the NRA' from noon the previous day to noon the day of publication during August 1933. Newspapers from the largest metro areas, such as the *Chicago Tribune* and *Washington Post*, generally printed aggregate numbers in terms of the number firms signing on the previous day as well as the names of the largest concerns to do so.
4. Taylor and Klein (2008) employ a game-theoretic model of the NIRA cartel compliance mechanism and present empirical evidence that firms' beliefs in the importance of the Blue Eagle compliance emblem affected economic decisions. They offer evidence that waning enthusiasm for the emblem in the spring of 1934 help lead to cartel breakdown.
5. This was quoted in the *Helena Independent*, 7 November 1933, p. 2. The 4 million job number was also cited in NRA Release no. 1874, p. 5. It is unclear, but this estimate by the Roosevelt administration may also include employment provided by the Public Works Administration which was set up under the NIRA. The analysis here focuses on the NIRA's labour provisions and hence treats such government expenditures as exogenous from the PRA and the NIRA.
6. Alexander (1994) concludes that relaxation of anti-trust law under the NIRA acted as an important coordination device which helped industries to sustain collusion after the legislation expired. Krepps (1997), however, finds that when the sample is expanded to include industries not covered by the NIRA to be used as a comparison group, the NIRA codes did not facilitate collusive outcomes after the NIRA time period. Alexander (1997) also explores the role that cost and product heterogeneity played in facilitating collusion under the NIRA. Taylor (2007) examines how specific attributes of the NIRA codes, such as the presence of production quotas, data filing requirements and restrictions on productive capacity, affected the ability of industries to attain collusive outcomes under the legislation.
7. Work-sharing programmes have also been contentious in the political arena. In May 2007 Nicolas Sarkozy said that it was a 'stupid idea to believe that it is by working less that we will create more wealth and more jobs'. As President of France, Sarkozy drafted a May 2008 bill that would scrap the 35-hour limit (*Wall Street Journal*, 30 May 2008, p. A9).
8. Bernanke (1986) creates a model that could help to explain how hourly wage rates rose while workweeks fell during the Great Depression, apart from any exogenous effects that New Deal

legislation may have had. Of course, to the extent that firms complied with the PRA provisions, rising wage rates and falling workweeks were largely exogenous during these months.

9. Technically, the contracts signed under the PRA expired on 31 December 1933. Rather than duplicating the administrative burden of having all firms not under a 'permanent' industry code sign and get approval of another blanket agreement, Roosevelt announced by Executive Order that 'Display of the Blue Eagle on or after January 1, 1934 . . . shall be deemed an acceptance' of the terms of the firm's previous blanket agreement (*Washington Post*, 21 December 1933, p. 1).
10. All real variables are computed by dividing nominal variables by the 'Index of the general price level' (NBER series 04051).
11. Technically, Title II of the NIRA authorized \$3.3 billion of government spending for the financing of public works, so more specifically, regressions that include government spending as a control variable are estimating the NIRA's and PRA's non-Title-II impacts. The private sector employment impact of such relief spending has been analysed by Neumann *et al.* (2009). The authors employ a panel VAR method and find that an increase in work relief spending led to a decline in private employment consistent with the complaints of contemporary employers who claimed that work relief competed with the private sector.
12. The finding that aggregate employment under the PRA fell while the number of jobs rose is interesting in light of Bernanke and Powell (1987). These authors find that during the interwar period, variations in the workweek account for around half of the variation in aggregate employment hours. Downturns often caused reductions in workweeks rather than massive layoffs. However, in the postwar era the relatively steady workweek meant that changes in aggregate employment hours in the face of a demand shock were largely accomplished through layoffs or increased hires.
13. Unemployment data are from the series 'Unemployment', 1929–1944, by the National Industrial Conference Board (NBER Series 08084). Unemployment was measured as the difference between the number of people in the labour force and the number of people working.
14. Wage data are from 'US average hourly earnings, twenty-five manufacturing industries, National Industrial Conference Board' (NBER Series 8142). Workweek data are from 'US average hours of work per week, manufacturing industries, total wage earners, National Industrial Conference Board' (NBER Series 8029a).

REFERENCES

ALEXANDER, B. (1994). The impact of the National Industrial Recovery Act on cartel formation and maintenance costs. *Review of Economics and Statistics*, **76**, 245–54.

——— (1997). Failed cooperation in heterogeneous industries under National Recovery Administration. *Journal of Economic History*, **57**, 322–44.

BERNANKE, B. S. (1986). Employment, hours, and earnings in the depression: an analysis of eight manufacturing industries. *American Economic Review*, **76**, 82–109.

——— and POWELL, J. L. (1987). The cyclical behavior of industrial labor markets: a comparison of pre-war and post-war eras. In R. J. Gordon (ed.), *The American Business Cycle: Continuity and Change*. Chicago: University of Chicago Press, pp. 583–621.

BILES, R. (1991). *A New Deal for the American People*. DeKalb: Northern Illinois University Press.

BRAND, D. (1988). *Corporatism and the Rule of Law: A Study of the National Recovery Administration*. Ithaca: Cornell University Press.

BRUNELLO, G. (1989). The employment effects of shorter working hours: an application to Japanese data. *Economica*, **56**(224), 473–86.

CALMFORS, L. and HOEL, M. (1988). Work sharing and overtime. *Scandinavian Journal of Economics*, **90**, 45–62.

COLE, H. L. and OHANIAN, L. E. (2004). New Deal policies and the persistence of the Great Depression: a general equilibrium analysis. *Journal of Political Economy*, **112**, 779–816.

CREPON, B. and KRAMARZ, F. (2002). Employed 40 hours or not employed 39: lessons from 1982 mandatory reduction of the workweek. *Journal of Political Economy*, **110**, 1355–89.

FITZROY, F., FUNKE, M. and NOLAN, M. A. (2002). Working time, taxation and unemployment in general equilibrium. *European Journal of Political Economy*, **18**, 333–44.

FRANZ, W. and KONIG, H. (1986). The nature and causes of unemployment in the federal republic of Germany since the 1970s: an empirical investigation. *Economica*, **53**, S219–44.

HUNT, J. (1999). Has work-sharing worked in Germany? *Quarterly Journal of Economics*, **114**, 117–48.

KREPPS, M. B. (1997). Another look at the impact of the National Industrial Recovery Act on cartel formation and maintenance costs. *Review of Economics and Statistics*, **79**, 151–4.

LOGEAY, A. and SCHREIBER, S. (2006). Testing the effectiveness of the French work-sharing reform: a forecasting approach. *Applied Economics*, **38**, 2053–68.

LYON, L. S., HOMAN, P. T., LORWIN, L. L., TERBORGH, G., DEARING, C. L. and MARSHALL, L. C. (1972). *The National Recovery Administration: An Analysis and Appraisal*. New York: Da Capo Press (Reprint of 1935 first edition).

MARIMON, R. and ZILIBOTTI, F. (2000). Employment and distributional effects of restricting working time. *European Economic Review*, **44**, 1291–326.

NEUMANN, T. C., FISHBACK, P. V. and KANTOR, S. (2009). The dynamics of relief spending and the private urban labor market during the New Deal. *Journal of Economic History*, forthcoming.

O'BRIEN, A. P. (1989). A behavioral explanation for nominal wage rigidity during the Great Depression. *Quarterly Journal of Economics*, **104**, 719–35.

POWELL, J. (2003). *FDR's Folly: How Roosevelt and His New Deal Prolonged the Great Depression*. New York: Crown Forum.

ROCHETEAU, G. (2002). Working time regulation in a search economy with worker moral hazard. *Journal of Public Economics*, **84**, 387–425.

ROMER, C. D. (1999). Why did prices rise in the 1930s? *Journal of Economic History*, **59**, 167–99.

SCHREIBER, S. (2006). Did work-sharing work in France? Evidence from a structural co-integrated VAR model. *European Journal of Political Economy*, **24**, 478–90.

SHLAES, A. (2007). *The Forgotten Man: A New History of the Great Depression*. New York: Harper-Collins.

TAYLOR, J. E. (2002). The output effects of government sponsored cartels during the New Deal. *Journal of Industrial Economics*, **50**, 1–10.

——— (2007). Cartel code attributes and cartel performance: an industry-level analysis of the National Industrial Recovery Act. *Journal of Law and Economics*, **50**, 597–624.

——— and KLEIN, P. G. (2008). Anatomy of a cartel: the National Industrial Recovery Act of 1933 and the compliance crisis of 1934. *Research in Economic History*, **26**, 235–71.

——— and SELGIN, G. (1999). By our bootstraps: origins and effects of the high-wage doctrine and the minimum wage. *Journal of Labor Research*, **20**, 447–62.

US COMMITTEE OF INDUSTRIAL ANALYSIS (1937). *The National Recovery Administration: Report of the President's Committee of Industrial Analysis*. Washington DC: Department of Commerce.

VEDDER, R. K. and GALLAWAY, L. E. (1997). *Out of Work: Unemployment and Government in Twentieth-century America*. New York: New York University Press.

WEINSTEIN, M. M. (1980). *Recovery and Redistribution under the NIRA*. Amsterdam: North-Holland.

Economica

Economica (2011) **78**, 159–186
doi:10.1111/j.1468-0335.2009.00806.x

Labour Market Institutions and Technological Employment

By Arnaud Chéron†‡, Francois Langot†§ and Eva Moreno-Galbis†¶

†*GAINS-TEPP, University of Maine, France* ‡*EDHEC* §*PSE and IZA*
¶*CEPREMAP*

Final version received 2 February 2009.

Our paper seeks to gain insights into the effect of labour market institutions on the dynamics of the labour market during the diffusion process of new technologies. We develop an endogenous job destruction matching framework, with heterogeneous workers, where the segmentation of the labour market between workers having the required ability to do a technological job and the rest of the workers is endogenous. The dynamics of this segmentation may follow a monotonous decreasing path or a non-monotonous U-shaped path depending on the unemployment benefit system. If benefits are generous, we are in the U-shaped case.

Introduction

The relationship between labour market institutions and the effects of technological diffusion on the dynamics of the labour market remains an unexplored research field. As shown by Figure 1, a negative correlation seems to arise between technological employment (information and communication technology (ICT) related occupations, by broad definition) and the generosity of labour market institutions. More precisely, countries with generous unemployment benefit (UB) systems have a lower proportion of technological employment. As new technologies become increasingly diffused, the situation worsens (the relationship was more negative in 2007 than in 1995).

An overview of the existing literature (see Section I) allows us to distinguish between two stages in the diffusion process of information and communication technologies. First, at the beginning of the technological diffusion, the adoption of new technologies was associated with an increase in the relative demand for skilled labour (non-production workers) compared to that for unskilled (production) workers, in what became widely known as the skill biased technological change (SBTC). Second, there seems to have been a progressive replacement of workers doing routine tasks by machines. Surprisingly, this type of worker is, in general, medium-qualified. Putting both stages together yields a kind of U-shaped progression in the relative unemployment rate of medium-skilled workers, who initially had access to the expanding technological sector and then were excluded from it. Whereas there is not enough literature on the subject to talk about a stylized fact, the progression of the French medium-skilled relative unemployment rate seems to correspond well to our intuition (see Figure 2).

The objective of this paper is to develop a theoretical framework allowing us to make the link between labour market institutions and the proportion of workers employed in positions requiring the use of ICT during the diffusion process of new technologies. Comparing the United States and European experiences, we realize that the diffusion of ICT has promoted a rise in wage inequalities when UBs are low (the USA), whereas it has fostered an increase in unemployment rates in Europe, where UBs protect the living standards of people out of work and prevent an adjustment through wages. This highlights the prominent role of UBs on labour market outcomes. All in all, European unemployment and US wage inequality are two sides of the same coin.

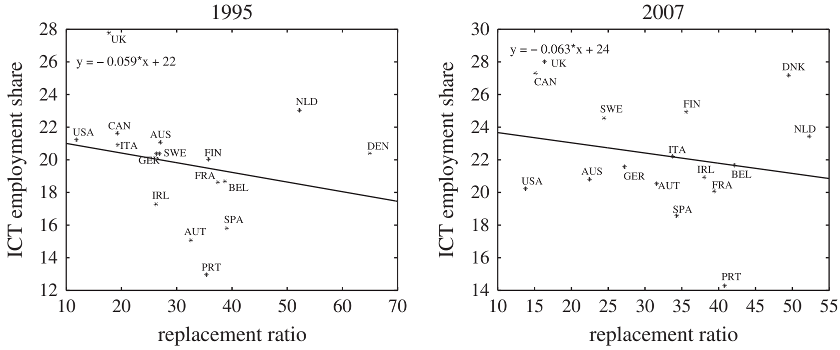

FIGURE 1. Correlation between the replacement ratio and the proportion of ICT-related occupations in various OECD countries, 1995, 2003.

Source of the average replacement ratio: OECD, Benefits and Wages Database (see Bassanini and Duval 2006). Available from the authors on request. Source concerning ICT-related occupations (broad definition): OECD Information Technology Outlook 2004, Chapter 6. The data for 2007 can be found in OECD Information Technology Outlook 2008. For Australia, Finland and Sweden, the first year is 1997 and not 1995.

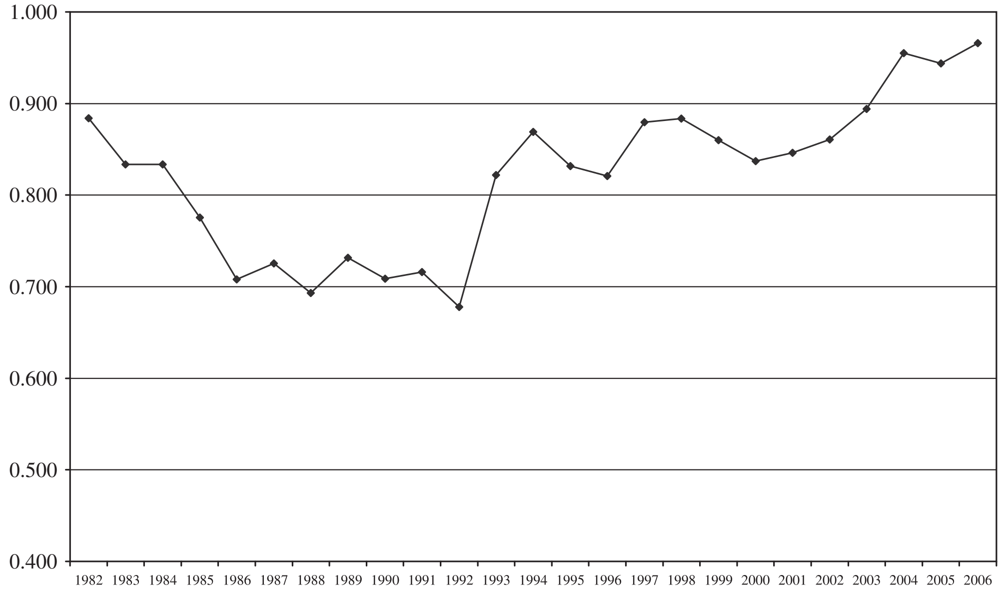

FIGURE 2. The relative unemployment rate of medium-qualified (baccalaureate or equivalent diploma) workers, France 1982–2006. (Source: French National Statistical Institute.)

We propose to analyse these features using an equilibrium unemployment model with heterogeneous workers and jobs. We assume that the economy is composed of a continuum of workers individually characterized by a given skill or ability level.[1] We also suppose that ICT and workers' abilities are complementary. We then distinguish between two types of jobs: simple jobs, where novel technologies are not used, so the worker's ability level does not influence productivity; and complex jobs, where, in contrast, new technologies are used and therefore the worker's productivity is proportional to his ability level. Compared to the related literature on task biased technological change (see Section I), complex jobs imply the performance of both routine (programmable) and non-routine (non-programmable) cognitive tasks. The least qualified workers in complex positions (which corresponds to medium-qualified workers) are supposed to perform routine tasks, and the highest-qualified workers perform non-routine tasks. This correspondence between ability level and performed task constitutes the only way that we have to distinguish between routine and non-routine tasks within complex jobs. Similarly, simple jobs in our framework correspond to what are traditionally known in the literature as manual jobs. Nevertheless, because this paper focuses on technological positions (complex jobs), the analysis of the simple segment remains fairly general and we assume that all abilities are in competition for a simple position (no possible distinction between routine and non-routine simple tasks can be supposed, as we do in the complex segment).

We extend the work of Mortensen and Pissarides (1999) by introducing an endogenously segmented labour market between workers having at least the threshold ability level giving access to complex positions, and the rest of the workers, who can occupy only simple positions. Firms offering a complex job support a setup cost, but the adoption of new technologies is supposed to improve their productivity. Furthermore, we assume a kind of learning process à la Greenwood and Jovanovic (2000), so that the setup cost of complex jobs decreases exponentially with technological diffusion (and thus the aggregate productivity associated with complex jobs increases).

The originality of our approach is to focus on the role of labour market institutions, which is introduced by indexing UBs to aggregate productivity. We abstract from the factor substitution relationship between computer capital and labour input in routine tasks proposed by Autor *et al.* (2003) or Autor and Dorn (2007) using an aggregate production function. We assume a one job, one firm setup and also consider search frictions. Our framework can be viewed as a complement to the existing literature, in which labour market frictions and institutions are omitted (see Section I). Concerning job competition, we extend the studies of Gautier (2002) and Dolado *et al.* (2000) where the segmentation of the labour market is completely exogenous, and also the framework presented in Albrecht and Vroman (2002), where the ability level of the workers does not play any role in this segmentation. As in Mortensen and Pissarides (1999), we consider an endogenous job destruction framework where the labour market is endogenously segmented between simple and complex jobs, but contrary to them, we allow different skill levels to compete for simple jobs.

By assuming an endogenous job destruction framework, we allow the minimum skill requirement to occupy a complex position (which determines the segmentation of the labour market) to follow either a non-monotonous or a monotonous path, depending on the generosity of the UB system. This constitutes a clear advantage with respect to an exogenous job destruction framework, which predicts a monotonously decreasing threshold ability level to occupy complex positions, independently of the hypothesis made on labour market institutions. Our results also reveal that UBs can rapidly exclude medium-qualified workers from complex positions, depending on whether the reduction in the setup costs and the productivity gains induced by the diffusion of ICT manage or not to compensate for wage increases (generated by the indexing of UBs to average productivity). Furthermore, because the reduction in the setup cost is not linear, a given ability level initially having access to complex positions may be excluded from this type of job during the diffusion process of ICT.

Numerical simulations suggest that reducing the replacement ratio by half allows us to increase the number of people qualifying for complex positions by 30%. This exclusion process for medium-qualified workers is also found in terms of job instability. In the presence of generous UBs, a given ability level is associated with a higher firing rate than in an economy with a restrictive unemployment insurance system. Similarly, in some periods of the technological diffusion process, unemployment rates in the presence of high UBs are twice as high as the ones found in an economy with low UBs. Aggregate unemployment rates are also increased by the presence of generous labour market institutions. More precisely, the unemployment rates in the simple segment and the complex segment are, on average, five percentage points and two percentage points higher than in an economy with low UBs.

The paper is organized as follows. The next section develops a brief survey of the related literature. Section II describes our theoretical framework and its assumptions, as well as the agents' behaviour. The steady-state equilibrium is described in Section III, and the effects of a biased technological shock in Section IV. Numerical simulations are implemented in Section V. Section VI concludes.

I. Related Literature

The impact of new technologies on the labour market has been widely analysed by economists. An overview of the empirical literature on the subject supports our introductory claim: technological diffusion can be understood in two stages. During the first stage, the adoption of new technologies, embodied or disembodied in the capital stock, favours the relative demand of qualified labour (skill biased technological change), either because of technological requirements (see Berman *et al.* 1994 for an analysis in the USA, and Machin and Van Reenen 1998 for several OECD countries) or because of induced organizational changes within firms (see Aguirregabiria and Alonso-Borrego 2001 for an example in the Spanish economy, or Caroli and Van Reenen 2001 for the United Kingdom and France).

In the second stage, the effect of new technologies is analysed in terms of tasks rather than in terms of labour qualification. More precisely, recent studies consider that the production process consists of both routine (programmable) tasks and non-routine (non-programmable) tasks. Routine tasks are performed by middle-skilled workers and can be either cognitive or manual (for example, bookkeeping, clerical work and repetitive production tasks). Non-routine tasks concern, on the one hand, abstract tasks performed by educated professionals and managers, the productivity of which depends essentially on access to abundant information and analysis. On the other hand, they also include low-skilled manual tasks, such as picking up irregularly scattered objects or walking through a crowd of moving people, that are difficult to automate or outsource since they require interpersonal and environmental adaptability as well as direct physical proximity. (See Table A1 in Appendix A for a definition of the tasks.)

Autor *et al.* (2003) note that computers are best at executing routine tasks that follow clearly defined procedures. At the same time, computers struggle with tasks that are less repetitive or require a large degree of environmental adaptability. These include abstract cognitive and interactive tasks, usually performed by high-skilled workers, but also manual tasks usually performed by low-skilled workers (see also Autor *et al.* (2006) and Autor and Dorn (2007) for other analysis on US data, Goos and Maning (2007) for the UK, Spitz-Oener (2006) for Germany, or Maurin and Thesmar (2005) for France). This task biased technological change (TBTC) fosters a progressive polarization of the labour market between 'lousy and lovely jobs' (Goos and Maning 2007), that is, between non-routine cognitive positions, generally occupied by high-skilled labour, and non-routine manual positions, traditionally reserved for non-qualified labour. As a result, the proportion of medium-skilled workers, normally associated with routine positions, decreases.

The development of all this empirical literature has been accompanied by a strand of theoretical papers providing some macro and micro foundations. From a macroeconomic point of view, Moreno-Galbis and Sneessens (2007) present a general equilibrium model with heterogeneous jobs and workers in order to analyse the relationship between the diffusion of ICT and the rise in low-skilled unemployment during the period 1975 to 2000. Ngai and Pissarides (2007) develop a multi-sector model of growth with differences in the total factor productivity (TFP) growth rates across sectors. They manage to reproduce the simultaneous growth in the relative prices and employment shares of stagnant sectors (such as community services). Their model predicts a shift of employment away from sectors with a high rate of technological progress towards sectors with low growth. At the limit, all employment converges to only two sectors, the sector producing capital goods and the sector with the lowest rate of productivity growth.

From a micro-economic point of view, Lindbeck and Snower (2000) develop a theoretical framework to show how the introduction of new information and communication systems, flexible machine tools, programmable equipment and the widening of human capital have fostered a workplace restructuring process of firms towards a holistic organization requiring multi-skilled workers. Beaudry and Green (2003) propose an endogenous technology adoption model in which geographic variation in computer adoption is driven by the relative abundance or scarcity of skilled workers, who are complemented by computer technology.

Autor *et al.* (2003) develop a stylized model where the production process includes abstract tasks implemented by high-skilled workers, routine tasks implemented by medium-skilled workers or computer capital, and manual tasks performed by low-skilled workers. Computerization is a complement to abstract rather than manual tasks. Specifically, it complements workers in abstract tasks by greatly reducing the cost of one of their primary inputs: information. The exogenous driving force of the model is the decline in the price of computer capital, which lowers the price of routine task input and increases the demand for routine tasks. Finally, Autor and Dorn (2007) complement the theoretical framework presented in Autor *et al.* (2003) by equating manual tasks with service producing occupations. Technical advances in routine tasks impact service occupations via the reallocation of labour (workers previously employed in routine tasks are reallocated to manual tasks) and changes in consumption patterns (the demand for services is increased).

In summary, during the first stages of technological diffusion, the emergence of repetitive tasks in the production process raises the demand for medium-qualified workers, who could easily develop these tasks. However, subsequent technological improvements make the human presence in repetitive tasks less useful.

II. The Model

In this section, we first present the main assumptions of the model: technologies, matching process, and firm and worker heterogeneities. Second, we derive the optimal behaviours of the labour market participants, and define labour market institutions and the wage bargaining process.

Assumptions

Worker's ability and technology of production We assume that the economy consists of a continuum of workers individually characterized by a given skill or ability level.

The ability levels $a(i)$ are drawn from the distribution $g(a(i))$ over an interval $[\underline{a}, \overline{a}] = [0, 1]$.

The firm may offer either a complex job, in which the worker implements cognitive tasks, or a simple job, in which the worker implements more manual tasks. The simple job is associated with a fixed coefficient technology requiring one worker to produce $h + \varepsilon$ units of output per period, where ε represents the random idiosyncratic productivity shock, and h represents the deterministic productivity component. In this type of task, the worker's ability level does not enhance productivity. A complex job is associated with a fixed coefficient technology requiring one worker with a skill level above a threshold value $a(i^f)$ to produce $p \cdot a(i) + \varepsilon$ units of output per period, where ε is a random idiosyncratic productivity shock, $a(i)$ stands for the worker's skill level, and p represents the unitary productivity associated with each ability level (state of technology). Because complex jobs are naturally more productive than simple ones, p must be greater than or equal to h.

The matching process Firms know perfectly the distribution of abilities among workers. Moreover, when opening a vacancy, they specify the required ability level to qualify for it (directed search[2]). The vacancy may then be filled and the worker starts producing, or remains empty and the employer continues searching.

We distinguish between two large categories of workers: those having at least an ability level $a(i^f)$ giving access to complex positions, and those having a qualification below $a(i^f)$ who can apply for only simple positions. In spite of having the possibility of occupying a complex vacancy, it may be in the interest of a worker with an ability level above $a(i^f)$ to search in the simple segment if her probability of finding a complex job is too low. Let us call $a(i^w) \geq a(i^f)$ the threshold skill level below which it is in the interest of a worker to search for a simple job. The segmentation of the labour market will then be determined by $a(\tilde{i}) = \text{Max}[a(i^f), a(i^w)]$.

The number of contacts per period in the simple segment (M_t^S) is represented by the constant returns to scale matching function $M_t^S = m\left(v^S, \sum_{j=1}^{\tilde{i}} u^j\right)$. The labour market tightness in the simple segment is then given by $\theta^S = \frac{v^S}{\sum_{j=1}^{\tilde{i}} u^j}$. In this segment, workers with divergent ability levels compete for a given type of job. The larger the number of people with different skill levels looking for a simple job, the lower the labour market tightness (θ^S) of this segment and the more intense the job competition. The probability of filling a simple vacancy equals $q(\theta^S) = M_t^S / v^S$, and the probability of finding a simple job is represented by $p(\theta^S) = M_t^S / \sum_{j=1}^{\tilde{i}} u^j = \theta^S q(\theta^S)$.

For workers searching for complex positions $(a(i) > a(\tilde{i}))$, we consider an infinitely segmented labour market in which, for each ability level, the labour market tightness equals $\theta_i^M = v_i^M / u^i$, where v_i^M stands for the number of complex vacancies directed to a skill level $a(i)$, and u_i represents the number of unemployed workers with an ability level $a(i)$. The number of contacts per period (M_{it}^M) is given by $M_{it}^M = m(v_i^M, u_i)$. The probability of filling a vacancy requiring a skill level $a(i)$ equals $q(\theta_i^M) = M_{it}^M / v_i^M$, and the probability that a worker with ability $a(i)$ will find a job is $p(\theta_i^M) = M_{it}^M / u_i = \theta_i^M q(\theta_i^M)$.

The job productivity shocks We assume an endogenous job destruction framework: if the stochastic productivity of a firm is below a given reservation level, then the optimal policy is to close the job. The values of the random idiosyncratic productivity parameter ε are drawn from the distribution Φ over the interval $[\underline{\varepsilon}, \overline{\varepsilon}]$. The process that changes this idiosyncratic term is the same for both types of job, and it follows a Poisson distribution with arrival rate $\lambda \in [0, 1]$. Therefore for every period there exists a probability λ that the firm is hit by a shock such that a new value of ε has to be drawn from Φ. Because search and hiring activities are costly, the new productivity level arising after the shock may indeed be too low to compensate either party for their efforts. The reservation productivity level will be denoted ε_i^M for the complex job and ε^S for the simple job. We assume that the first period idiosyncratic productivity in both types of job is at its maximum level, $\bar{\varepsilon}$, so that all jobs last at least one period.

The setup cost The IT revolution has fostered the emergence of complex positions where the worker is required to use computer capital. The first firms introducing these jobs had to bear high setup costs. However, as new technologies became diffused across the economy and complex jobs became increasingly abundant, these setup costs decreased thanks to positive spillovers: the follower firms did not make the same mistakes as the leaders made, so we can assume that the setup costs associated with the creation of complex positions fall as their number increases. This kind of learning process is close to the one introduced in Greenwood and Jovanovic (2000).

To formalize this idea, we simply define the process

$$K(p) = e^{-\gamma(p-p_0)}, \tag{1}$$

where γ represents the speed of adjustment, p is the current state of technology (productivity of complex positions), and p_0 stands for the final or potential technology level. Initially $p < p_0$, and at the end of the catch-up $p = p_0$, implying that $K(p) = 1$. Because the setup cost is given by an inverted exponential function, it will start decreasing faster than the linear increase in p. However, at the end of the ICT diffusion process, the subsequent decreases in $K(p)$ will be lower than the rise in p. Finally, note that since we are not considering a growth framework but rather a gradual technological shift, the value of p is upward bounded.

Agent behaviours

A vacancy can remain unfilled and the employer continues searching, or be filled and the worker starts producing. The associated asset value to each of these situations is represented by $V^M(a(i))$ (respectively V^S) when a complex (respectively simple) vacancy is empty, and by $J^M(a(i), \varepsilon)$ (respectively $J^S(\varepsilon)$) when the complex (respectively simple) vacancy is filled. In the same way, the value to the worker in a complex (respectively simple) job is denoted as $W^M(a(i), \varepsilon)$ (respectively $W^S(\varepsilon)$). Finally, the average expected return on the worker's human capital when looking for a job is represented by $U^M(a(i))$ (respectively U^S) when the worker's skill level is above (respectively below) the threshold value $a(\tilde{i})$.

The firms When the firm opens a vacancy, it bears a cost c per unit of time, whatever the skill level required to fill the vacancy.[3]

There are probabilities $1 - q(\theta_i^M)$ and $1 - q(\theta^S)$ that the complex and simple vacancies, respectively, remain empty next period. On the other hand, there are probabilities $q(\theta_i^M)$ and $q(\theta^S)$ that the complex and simple vacancies are filled. The asset values associated with a searching vacancy are then

$$\begin{aligned} V^M(a(i)) = & -c + \beta(1 - q(\theta_i^M))V^M(a(i)) \\ & + \beta q(\theta_i^M)(J^M(a(i), \bar{\varepsilon}) - K(p)), \end{aligned} \tag{2}$$

$$V^S = -c + \beta(1 - q(\theta^S))V^S + \beta q(\theta^S)J^S(\bar{\varepsilon}), \tag{3}$$

where β is the discount factor.

From the second period of the match, the asset values associated with complex and simple jobs are respectively defined as[4]

$$\begin{aligned} J^M(a(i),\varepsilon) = p \cdot a(i) + \varepsilon - w^M(\varepsilon, a(i)) + \beta(1-\lambda) \\ \text{Max}[J^M(a(i),\varepsilon), V^M(a(i))] \\ + \beta\lambda \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \text{Max}[J^M(a(i),x), V^M(a(i))]\text{d}\Phi(x), \end{aligned} \tag{4}$$

$$\begin{aligned} J^S(\varepsilon) = h + \varepsilon - w^S(\varepsilon) + \beta(1-\lambda)\text{Max}[J^S(\varepsilon), V^S] \\ + \beta\lambda \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \text{Max}[J^S(x), V^S]\text{d}\Phi(x), \end{aligned} \tag{5}$$

where $w^M(\varepsilon, a(i))$ and $w^S(\varepsilon)$ represent, respectively, the wages paid to a worker in a complex job and in a simple job. The firm opens vacancies until all rents are exhausted ($V^M(a(i)) = 0$, $V^S = 0$), that is,

$$\frac{c}{q(\theta_i^M)} = \beta(J^M(a(i),\bar{\varepsilon}) - K(p)), \tag{6}$$

$$\frac{c}{q(\theta^S)} = \beta J^S(\bar{\varepsilon}). \tag{7}$$

Similarly, it is not in the interest of the firm to continue a match for all productivity levels below $J^S(\varepsilon^S) = 0$ in simple jobs and $J^M(a(i), \varepsilon_i^M) = 0$ in complex jobs of ability $a(i)$. These job creation and job destruction rules will allow us to determine the equilibrium labour market tightness and critical productivity level in each segment.

The workers An unemployed worker receives a flow of earnings w_i^u including unemployment benefits, leisure, domestic productivity, etc. For simplicity and realism we will assume that the unemployment benefit earned by someone in the complex segment remains above the unemployment benefit obtained by someone in the simple segment, $w_M^u > w_S^u$.

A job seeker with ability $a(i) > a(\tilde{i})$ comes into contact with a complex vacant slot at rate $\theta_i^M q(\theta_i^M)$, while a job seeker looking for a simple job comes into contact with a vacancy at rate $\theta^S q(\theta^S)$. The asset value of unemployment to each type of worker is given by

$$\begin{aligned} U^M(a(i)) = w_M^u + \beta(1 - \theta_i^M q(\theta_i^M))U^M(a(i)) \\ + \beta\theta_i^M q(\theta_i^M)W^M(a(i),\bar{\varepsilon}), \end{aligned} \tag{8}$$

$$U^S = w_S^u + \beta(1 - \theta^S q(\theta^S))\, U^S + \beta\theta^S q(\theta^S)\, W^S(\bar{\varepsilon}). \tag{9}$$

The asset value of unemployment to a worker searching in the simple segment is independent of her ability because her job opportunities and unemployment benefits do not depend on it.

One of the interesting contributions of our theoretical framework is to allow workers to search for a job in the segment where their expected value of unemployment is higher. In this sense, we have the following.

- If $a(i) > a(i^f)$, then $U^M(a(i)) = \text{Max}[U^S, U^M(a(i))]$. All workers having an ability level below a critical $a(i^w)$ for which $U^M(a(i^w)) = U^S$ will search in the simple segment in spite of having the required skills to apply for complex jobs.[5]
- If $a(i) < a(i^f)$, then $U^S = \text{Max}[U^S, U^M(a(i))]$. Since these workers have a zero probability of finding a complex job, it is always in their interest to remain in the simple segment: $U^S = \text{Max}[U^S, U^M(a(i))]$.

The present values to the worker of a complex and a simple job solve the following:[6]

$$(10)\quad \begin{aligned} W^M(a(i), \varepsilon) &= w^M(a(i), \varepsilon) + \beta(1-\lambda)\text{Max}[W^M(a(i), \varepsilon), U^M(a(i))] \\ &\quad + \beta\lambda \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \text{Max}[W^M(a(i), x), U^M(a(i))]\text{d}\Phi(x), \end{aligned}$$

$$(11)\quad \begin{aligned} W^S(\varepsilon) &= w^S(\varepsilon) + \beta(1-\lambda)\text{Max}[W^S(\varepsilon), U^S] \\ &\quad + \beta\lambda \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} \text{Max}[W^S(x), U^S]\text{d}\Phi(x). \end{aligned}$$

The unemployment benefit

We introduce the role of labour market institutions by assuming that unemployment benefits are indexed to general productivity. As noted in Pissarides (1990), indexing unemployment benefit to the average wage creates complications in a model like ours where the equilibrium is characterized by a conditional wage distribution and not by a unique wage rate. In these cases a convenient shortcut is to define benefits in terms of the general productivity parameter p, which is unique and exogenous. Whether indexed to the average wage or to the productivity parameter, the intuitive idea behind this formalization of unemployment benefit is that the European unemployment system is redistributive, therefore improvements in general standards of living are reflected in unemployment benefit.

For high-productivity workers (unemployed people in the complex segment) the unemployment benefit is indexed to their productivity level: $w^u_M = \delta_M \cdot p$, with δ_M being a positive constant smaller than 1. Similarly, for low-productivity workers (simple segment) the unemployment benefit is defined as $w^u_S = \delta_S \cdot p$, with δ_S being a positive constant smaller than δ_M.

The wage bargaining process

Since search and hiring activities are costly, when a match is formed, a joint surplus is generated. At the beginning of every period, the firm and the employee renegotiate wages through a Nash bargaining process that splits the joint surplus into fixed proportions at all times. We denote the bargaining power of workers as $\eta \in (0, 1)$. We assume a two-tier wage contract where the first tier wage rate $(w^{M1}(a(i)))$ includes the setup costs borne by the firm during the first period of the match. In the subsequent bargaining problem

$(w^M(a(i)))$, the employer threat point does not include setup costs—those are sunk.[7] In the simple segment, there are no setup costs, therefore we have a unique wage. The wage contracts are

$$w^{M1}(a(i)) = (1-\eta)w^u_M + \eta(p\cdot a(i) + \varepsilon + c\theta^M_i - K(p)(1-\beta(1-\lambda))), \tag{12}$$

$$w^M(a(i)) = (1-\eta)\, w^u_M + \eta(p\cdot a(i) + \varepsilon + c\theta^M_i), \tag{13}$$

$$w^S = (1-\eta)\, w^u_S + \eta(h + \varepsilon + c\theta^S), \tag{14}$$

where we realize that wages result from a weighted average between the worker's outside option on the one hand, and the worker's productivity and labour market conditions, on the other hand.

III. The Steady State

Job creation and job destruction rules: labour market equilibria

As long as the joint surplus (the one obtained by the firm plus that of the worker) is positive, the job goes on. If the joint surplus becomes negative, the match breaks down. For each type of job there thus exists a critical productivity level—ε^M_i for complex jobs and ε^S for simple ones—below which the surplus is negative and it is not profitable to pursue the job.[8]

Integrating by parts yields the following expressions for the complex and simple job destruction rules:

$$p\delta_M + \frac{\eta}{1-\eta}c\theta^M_i = pa(i) + \varepsilon^M_i + \frac{\beta\lambda}{1-\beta(1-\lambda)}\int_{\varepsilon^M_i}^{\bar{\varepsilon}}(1-\Phi(x))\mathrm{d}x, \tag{15}$$

$$p\delta_S + \frac{\eta}{1-\eta}c\theta^S = h + \varepsilon^S + \frac{\beta\lambda}{1-\beta(1-\lambda)}\int_{\varepsilon^S}^{\bar{\varepsilon}}(1-\Phi(x))\mathrm{d}x, \tag{16}$$

where the left-hand side stands for the search value of unemployment and the right-hand side stands for the firm's minimum asset value obtained from a job.

At the steady state, the firms open vacancies until no more benefit can be obtained, that is, all rents are exhausted and the free entry condition applies: $V^M(a(i)) = 0$ and $V^S = 0$. From equations (2), (3), (4), (5), the wage equations (12), (13) and (14), and the critical productivity levels provided from expressions (15) and (16), we derive the following job creation rules:

$$\frac{c}{q(\theta^M_i)} = \beta(1-\eta)\left[\frac{(\bar{\varepsilon}-\varepsilon^M_i)}{1-\beta(1-\lambda)} - K(p)\right], \tag{17}$$

$$\frac{c}{q(\theta^S)} = \beta\frac{(1-\eta)(\bar{\varepsilon}-\varepsilon^S)}{1-\beta(1-\lambda)}, \tag{18}$$

where $K(p) = \mathrm{e}^{-\gamma(p-p_0)}$.

For simple jobs (occupied by workers with an ability level lower than $a(\tilde{i})$), we have the traditional job creation–job destruction theoretical framework introduced by Mortensen and Pissarides (1994). In this setup, the job creation curve (equation (18)) is strictly decreasing in the space $[\varepsilon, \theta^S]$, and the job destruction curve (equation (16)) is strictly increasing, guaranteeing the existence of a unique equilibrium point $[\varepsilon^{S}*, \theta^{S}*]$.

In the complex segment the equilibrium is also characterized by the job creation and destruction curves. However, when dealing with complex positions we must distinguish between three variables of interest: the reservation productivity (ε_i^M), the labour market tightness (θ_i^M), and the ability level corresponding to the considered labour market segment ($a(i)$). We first consider the job creation and job destruction curves in the $[\theta_i^M, \varepsilon_i^M]$ space. For any $a(i) > a(\tilde{i})$, the equilibrium labour market tightness and reservation productivity of the ability segment are determined by the intersection between the job creation (equation (17)) and job destruction (equation (15)) curves. While the former curve is negatively sloped, the latter has a positive slope, guaranteeing the existence of a unique equilibrium.

The endogenous segmentation of the labour market

The labour market is segmented between workers occupying complex positions and those occupying simple positions. The ability level determining this segmentation is represented by $a(\tilde{i})$. As shown below, $a(\tilde{i})$ may result from either the firm's choice or the worker's choice.

The firm's critical ability level for complex jobs Because firms offering a complex position must support a setup cost equal to $K(p)$ during the first period of the contact, they will direct their vacancies to workers having at least the ability level required to compensate this cost. Nevertheless, congestion effects imply that it is not optimal for the firms to direct all the job offers towards the most highly qualified workers. The minimum ability level required to exactly compensate the setup cost of opening a complex position is given by $a(i^f)$.

Proposition 1. The lowest ability level, $a(i^f)$, hired for a complex position satisfies

$$p\delta_M = pa(i^f) + (\bar{\varepsilon} - K(p)(1 - \beta(1 - \lambda))) + \frac{\beta\lambda}{(1 - \beta(1 - \lambda))} \int_{\bar{\varepsilon} - K(p)(1 - \beta(1 - \lambda))}^{\bar{\varepsilon}} (1 - \Phi(x))\mathrm{d}x. \tag{19}$$

Proof. The $a(i^f)$ slot stands for the least profitable ability in complex jobs: nobody creates a complex job for an ability level $a(i)$ lower than $a(i^f)$. At this point the number of vacancies converges to zero, leading to the following free entry condition:

$$\frac{\bar{\varepsilon} - \varepsilon_{i^f}^M}{1 - \beta(1 - \lambda)} = K(p) = \mathrm{e}^{-\gamma(p - p_0)}. \tag{20}$$

The firm must then determine the ability level, $a_i = a(i^f)$, supporting this reservation productivity. To this end, we set θ_{i^f} to zero in the job destruction rule (equation (15)):

$$p\delta_M = pa(i^f) + \varepsilon_{i^f}^M + \frac{\beta\lambda}{1 - \beta(1 - \lambda)} \int_{\varepsilon_{i^f}^M}^{\bar{\varepsilon}} (1 - \Phi(x))\mathrm{d}x. \tag{21}$$

Combining equations (20) and (21) yields equation (19). □

For a very low degree of ICT diffusion (low p), setup costs will be so high that it will not be in the interest of the firm to offer any complex position ($a(i^f)$ high). Indeed, high setup costs must be compensated by a low idiosyncratic productivity component (low $\varepsilon_{i^f}^M$); for given labour market institutions (δ_M), the threshold ability level $a(i^f)$ will be high. Moreover, generous unemployment benefits (high δ_M) are associated with higher levels of $a(i^f)$, reducing the proportion of technological employment.

The worker's critical ability level for complex jobs Even if someone has an ability level above $a(i^f)$, it might be in her interest to search in the simple segment if the number of complex vacancies open in her ability slot is too low. A worker with an ability level $a(i) > a(i^f)$ may decide to search in the simple side of the labour market if her asset value of unemployment when remaining in the complex segment is below her asset value of unemployment when looking for a simple job. This trade-off can be represented using the asset values of unemployment:

$$(1-\beta)(U^M(i) - U^S) = w_M^u - w_S^u - \left(\frac{1-\eta}{\eta}\right)c(\theta^S - \theta^M(a_i)).$$

A worker having the required ability level to occupy a complex job may decide to search in the simple segment ($U^M(i) < U^S$). This arises when unemployment benefits perceived while searching for a complex position are not high enough to compensate the low probability to find a complex job. Note that even if $w_M^u > w_S^u$, a worker having $a(i) > a(i^f)$ may prefer to search for a simple position if the probability to find a complex job is too low ($\theta^S > \theta^M(a_i)$). In this equilibrium, the segmentation of the labour market is determined by the worker's threshold value $a(i^w)$, defined by $U^M(i^w) = U^S$. There will be no posted vacancies within the interval $[a(i^f), a(i^w)]$.

Conversely, if unemployment benefits are high enough to compensate the low probability of finding a complex job, all workers with an ability level above $a(i^f)$ remain searching in the complex segment. In this equilibrium, the segmentation is given by the firm's threshold value $a(i^f)$.

IV. The Effects of a Biased Technological Diffusion

In this section we analyse the effects of a gradual diffusion of new technologies. We consider an initial situation characterized by a very low p and assume a progressive increase in p towards p_0. To be illustrative, we impose an initial situation where p is low and there are no complex positions in the economy, i.e. $a(\tilde{i}) = \overline{a(i)} = 1$. As the economy starts its technological development, p rises, reducing the gap between p and p_0 and fostering the appearance of complex positions, whose productivity is enhanced. The diffusion of complex jobs promotes a gradual reduction in the setup costs borne by firms, i.e. a fall in $K(p) = \mathrm{e}^{-\gamma(p-p_0)}$.

Simple jobs

The equilibrium in the simple segment is characterized by the job creation and job destruction curves, respectively given by (16) and (18).

Proposition 2. A biased technological change exclusively favouring the productivity of complex positions decreases the labour market tightness and increases the reservation productivity in the simple segment. By the Beveridge curve, unemployment in the simple sector increases. Unambiguously, the number of simple jobs in the economy decreases.

Proof. The proof is straightforward from (16) and (18)—see Appendix B. □

Even if the productivity of workers employed in simple jobs is not affected by the technological change, their reservation wage increases following the upturn in p (w_S^u is partly indexed to p). This yields a reduction in the number of simple vacancies and an increase in the reservation productivity.

Complex jobs

Determining the effects of technological diffusion on the complex segment becomes a slightly more complicated issue. Our analysis covers two stages. We first analyse the impact of a biased technological change on the labour market equilibrium of a particular ability slot. In the second stage, we study the effects of the technological change on the endogenous segmentation of the labour market. By means of this two-step analysis we provide a complete picture of the equilibrium in the complex segment.

How does ICT shift the labour market equilibrium in each ability slot? We analyse the effect of a biased technological progress on the labour market equilibrium of an ability slot that remains open in the complex segment. Both the job creation and job destruction curves are affected by this technological change. The job creation curve shifts due to the setup cost, while the job destruction curve shifts due to the improvement in the workers' productivity and the rise in the outside option (see Figure A2 in Appendix B for a graphical representation).

Proposition 3. In the complex segment, the diffusion of novel technologies increases the reservation productivity and decreases the labour market tightness of those workers having low ability levels. In contrast, for workers having intermediate and high ability levels, the reservation productivity decreases and the labour market tightness increases.

Proof. See Appendix B. □

Concerning the job creation decision, the diffusion of ICT is associated with a progressive reduction of the setup cost borne by firms offering complex positions. The opening of complex vacancies is thus stimulated.

The direct impact of the diffusion of ICT on the job destruction decision is governed by the gap between the marginal increase of the firm's output (a_i) and the marginal increase in the wage (δ_M). For given θ_i^M and $a(i)$, there exists an ability level such that $a(i) = \delta_M$, and the job destruction curve remains unaffected by the change in p. On the other hand, if $\delta_M < a(i)$, there is an improvement in the stability of these jobs. Finally, if $\delta_M > a(i)$, there is a rise in the reservation productivity required to pursue the match.

Combining both the movement of the job creation curve and the one of the job destruction curve, three possible situations can arise at the equilibrium.

- If $a(i)$ equals δ_M or is nearby, only the shift of the creation curve is significant: the new equilibrium will be characterized by a higher labour market tightness and then a higher reservation productivity. In an extreme case we might find a situation where the shift of the job destruction curve is such that $\varepsilon_i^M(\theta_i^M)$ remains unaffected and $\theta_i^M(\varepsilon_i^M)$ increases (case A in Figure A2 in Appendix B).

- For high ability levels, deterministic productivity gains ($pa(i)$) largely compensate for the rise in wages resulting from the increase in the outside option ($p\delta_M$) and in the labour market tightness (θ_i^M). Labour hoarding strategies thus become more profitable,[9] implying that job duration is enlarged (fall in the reservation productivity) and labour market tightness increases (case B in Figure A2 in Appendix B).
- Conversely, for a 'sufficiently low' ability level, labour hoarding strategies are no longer profitable (increase in the reservation productivity) and the labour market tightness falls[10] (case C in Figure A2 in Appendix B).

How does ICT affect the segmentation of the labour market? All workers included in the interval $[a(i^f), a(i^w)]$ can theoretically apply for a complex position; however, when the number of vacancies offered in their ability slot is low, it may be in their interest to search rather in the simple segment. At $a(i^w)$, the worker is indifferent between searching in the simple or the complex segment. In this section we analyse the effects of a progressive increase in p on the size of each labour market segment, that is, on $a(\tilde{i}) = \text{Max}[a(i^f), a(i^w)]$.

Case 1: The segmentation is determined by the firm's threshold value When unemployment benefits manage to compensate for the low probability of finding a complex job, all workers with an ability $a(i) > a(i^f)$ search for a complex position and no mobility is observed between segments. Even if they bear high unemployment rates, it is in the interest of these workers to remain searching in the complex segment because unemployment benefits are high. In this context $a(i^w)$ equals $a(i^f)$.

Proposition 4. For a given $a(i^f)$, there exists a $\tilde{\delta}_M(p)$ such that for $\delta_M > \tilde{\delta}_M(p)$ ($\delta_M < \tilde{\delta}_M(p)$) we have $\partial a(i^f)/\partial p > 0$ $(\partial a(i^f)/\partial p < 0)$. Then the size of the complex segment is reduced (augmented) during the diffusion process of new technologies.

Proof. To determine the impact of a biased technological change on the critical ability level $a(i^w) = a(i^f)$, we use the job creation and job destruction curves defined for $\theta_{i^f} = 0$ (equations (20) and (21)). The effect on the final critical ability level required in complex positions is given by

$$\text{(22)} \qquad \frac{\mathrm{d}a(i^f)}{\mathrm{d}p} = \frac{\delta_M - a(i^f) - K(p)\gamma(1 - \beta(1 - \lambda\Phi(\varepsilon_{i^f})))}{p}.$$

Let us define $\tilde{\delta}_M$ such that $\mathrm{d}a(i^f)/\mathrm{d}p = 0$:

$$\text{(23)} \qquad \tilde{\delta}_M = a(i^f) + K(p)\gamma(1 - \beta(1 - \lambda\Phi(\varepsilon_{i^f}))).$$

We then easily deduce that

$$\text{if } \delta_M > \tilde{\delta}_M, \quad \text{then } \frac{\mathrm{d}a(i^f)}{\mathrm{d}p} > 0,$$

$$\text{if } \delta_M \leqslant \tilde{\delta}_M, \quad \text{then } \frac{\mathrm{d}a(i^f)}{\mathrm{d}p} \leqslant 0. \qquad \square$$

If productivity gains (determined by $a(i^f)$), together with the actualized reduction[11] in the setup cost, do not manage to overcome the rise in the reservation wage (given by δ_M), then the skill requirement to occupy a complex position rises. This implies that a lower proportion of the labour force has access to complex positions.

The endogenous job destruction framework presents the main advantage of not imposing a predetermined path on the labour market segmentation. A detailed analysis of equation (22) permits us to better understand under which conditions the segmentation of the labour market may follow a non-monotonous path. Because the setup costs are represented via an inverted exponential function, the initial reductions in the setup cost will be considerable. This implies that even if the critical ability level $a(i^f)$ is such that $\delta_M > a(i^f)$, the firm may accept a lower threshold ability if the reduction in the setup cost is sufficiently great to compensate for the gap between δ_M and $a(i^f)$: in this case we have $\delta_M - a(i^f) < K(p)\gamma(1-\beta(1-\lambda\Phi(\varepsilon_{i^f})))$. In contrast, as new technologies become increasingly diffused (high p), the decrease in the setup cost falls and therefore it may no longer compensate the gap between δ_M and $a(i^f)$: we then have $\delta_M - a(i^f) > K(p)\gamma(1-\beta(1-\lambda\Phi(\varepsilon_{i^f})))$. Hence firms raise their skill requirements for complex positions. During the rising path of p we may find a U-shaped trend of the ability level required in complex vacancies.

Case 2: The segmentation is determined by the worker's threshold value When unemployment benefits are not sufficient to compensate for the low probability of finding a job, all workers within the interval $[a(i^f), a(i^w)]$ prefer to search for a job in the simple segment. The progression of $a(i^w)$ during the ICT diffusion is determined by both the labour market tightness in the simple segment and the changes in the $w_S^u - w_M^u$ relationship, since $\theta_{i^w} = \theta^S + (1-\eta)/(c\eta)(w_S^u - w_M^u)$. Therefore, to determine the effect of a biased technological change on the dynamics of $a(i^w)$, we need to analyse the evolution of both θ^S and $w_S^u - w_M^u$ along the increasing path of p.

Proposition 5. When p increases, $a(i^w)$ converges towards $a(i^f)$.

Proof. By Proposition 2, we know that $\partial\theta^S/\partial p < 0$. Moreover, the variation of unemployment benefits when technological progress accelerates is such that $\mathrm{d}(w_S^u - w_M^u)/\mathrm{d}p = (\delta_S - \delta_M) < 0$. This implies that $\partial\theta_{i^w}/\partial p < 0$. Then $a(i^w)$ converges towards $a(i^f)$ when p increases. □

Productivity improvements fostered by the diffusion of new technologies favour the convergence of $a(i^w)$ towards $a(i^f)$, since the progressive increase in unemployment benefits, together with the reduction of the labour market tightness in the simple segment, makes it decreasingly profitable for people qualifying for complex positions to search in the simple segment. Because the labour market tightness in the simple segment falls and the divergence in the unemployment benefits decreases along the rising path of p, $\theta_{i^w}^M$ will tend progressively towards zero, implying that $a(i^w)$ will converge towards $a(i^f)$.

The role of the endogenous job destruction rate The main objective of this paper is to analyse the impact of labour market institutions on the segmentation of the labour market (between complex and simple jobs) during the diffusion process of new technologies. To do so, two possibilities arise. We can consider either an exogenous job destruction framework or an endogenous job destruction framework. The former presents the main advantage of simplifying computations; however, it has a very constraining limitation: the segmentation of the labour market necessarily follows a monotonous decreasing path. This implies that the proportion of workers having access to complex positions continuously increases (see Appendix D for more details).

In an endogenous job destruction framework, the deterministic instantaneous profit of the firm $a(i) - \delta_M$ does not need to be positive for the match to continue, since the firm may expect a positive shock ε to arrive so that the negative deterministic difference is compensated. This allows a non-monotonous segmentation path to arise since we can have a critical ability level $a(i^f)$ such that $\delta_M > a(i^f)$. In this case, we may find that the initial reductions in the setup cost are sufficiently big to compensate for the gap between δ_M and $a(i^f)$, so $\partial a(i^f)/\partial p$ is negative (see equation (22)). Similarly, at a given moment in time, the reductions in the setup cost may no longer compensate for the gap between δ_M and $a(i^f)$, so that $\partial a(i^f)/\partial p$ becomes positive. In summary, by allowing the instantaneous profit of a match to be negative, the endogenous job destruction framework can foster a non-monotonous segmentation of the labour market.

V. Numerical Simulations

The quantitative implications of the model concerning the effects of a biased technological progress on the ability requirements to occupy complex jobs, on labour flows and on unemployment are clearly presented as a result of computational exercises. The results reported in this section are based on the following additional specification assumptions. A matching function of the Cobb–Douglas form is assumed with elasticity with respect to vacancies equal to ψ. The distribution of idiosyncratic shocks is assumed to be uniform on the support $[\underline{\varepsilon}, \overline{\varepsilon}]$, i.e. $F(x) = (x - \underline{\varepsilon})/(\overline{\varepsilon} - \underline{\varepsilon})$ defined on $[-3, 3]$. The baseline parameters used in computations are shown in Table 1. The elasticity with respect to vacancies (ψ), the bargaining power of workers (η), the arrival rate of a productivity shock (λ), the discount factor (β), and the recruiting cost (c) are calibrated with the values adopted by Mortensen and Pissarides (1994). All other structural parameters are chosen so that at the various steady states computed for different stages

Table 1
Baseline Parameter Values

	French-type		Counterfactual
Unemployment benefit index: complex segment	$\delta_M = 0.50$		$\delta_M = 0.35$
Unemployment benefit index: simple segment	$\delta_S = 0.30$		$\delta_S = 0.15$
Severance tax	$T = 0.2$		$T = 0.2$
Discount factor		$r = 0.02$	
Matching elasticity		$\psi = 0.5$	
Bargaining power		$\eta = 0.5$	
Matching efficiency		$m_0 = 0.3$	
Recruiting cost		$c = 0.3$	
Productivity shock frequency		$\lambda = 0.1$	
Speed of the catch-up process		$\gamma = 0.4$	
Deterministic productivity for simple jobs		$h = 1$	
Deterministic productivity for complex jobs		$p = 1$	

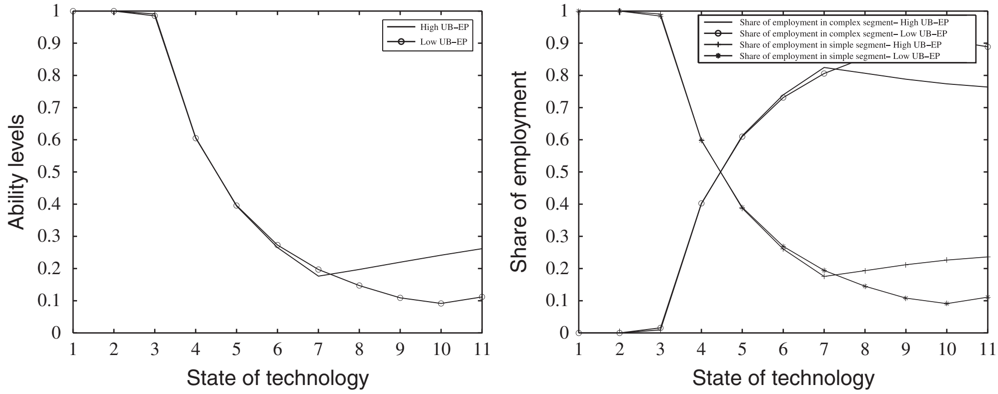

FIGURE 3. The minimum skill level found in complex positions and the proportion of employment in each segment during the ICT diffusion process.

of technological diffusion, unemployment rates, job destruction rates and average duration of unemployment spells match the average experience of France.[12] We compare this benchmark situation (French-type economy case) with a counterfactual one that would have arisen if the degree of indexing of the unemployment benefit had been lower. As in Mortensen and Pissarides (1994), the severance tax is assumed to be equal to 20% of the best productivity level that can be attained in a position.[13] We introduce this employment protection simply to keep as close as possible to the French case. When we eliminate it, our conclusions hold and there is only a scale effect.

We consider an initial situation where ICT is non-existent and all jobs are simple, so that workers of different abilities compete for the same positions and everyone receives a wage w^s. (Wage divergences arise simply from the idiosyncratic productivity component ε.) We then simulate the effects of a progressive diffusion of technologies (increase in p to attain p_0) and analyse the minimum skill level found in complex positions $a((\tilde{i}))$), the size of each labour market segment (θ_i^M and θ^S), job stability (ε_i^M and ε^S) and unemployment (u^M and U^S).

The left-hand panel in Figure 3 summarizes the minimum ability level found in complex positions along the rising path of p. During the first half of the technological diffusion, both economies display a similar decreasing path concerning the minimum ability level occupying complex positions. Furthermore, along this downward trend, the segmentation of the labour market is determined by the worker's threshold value $a(i^w)$. The situation is modified as soon as we consider the second half of the ICT diffusion process. From this point, the economy with high unemployment benefits sharply increases the skill requirements in complex positions. Furthermore, along this upward trajectory the firm's skill requirements are binding, that is, $a(i^w) = a(i^f)$. Medium-skilled workers in the neighbourhood of i^f prefer to remain searching in the complex segment because the high unemployment benefits more than compensate for their low probability of finding a job.

The U-shaped path followed by skill requirements in complex positions implies that workers previously qualified for these jobs no longer have access to them and are forced to search in the simple segment where they compete with lower qualified workers. At the end of the ICT diffusion process there will be 30% more excluded workers from the complex segment in the presence of high UB than in the presence of low UB, where the

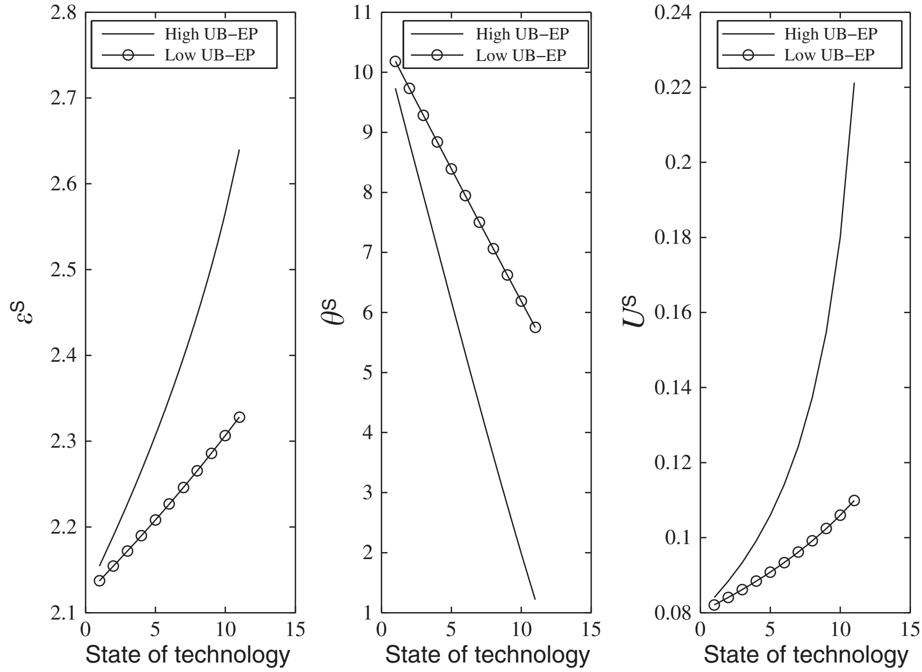

FIGURE 4. Job stability, labour market tightness and the unemployment rate in the simple segment during the ICT diffusion process.

segmentation follows a monotonously decreasing path. Conversely, there will be 28% more employed people in the simple segment (see the right-hand panel of Figure 3) when the unemployment system is generous (the two percentage point divergence is explained by the unemployment rate). Related to Figure 1, presented in the Introduction, the proportion of technological employment is lower in the presence of high UB.[14]

Result 1. The presence of high UB excludes a larger fraction of workers from the complex segment and exacerbates job competition in the simple segment. Low UB allows a larger proportion of workers with heterogeneous abilities to have access to technological employment.

The dynamics of job stability, labour market tightness and unemployment are strongly differentiated depending on the labour market segment considered. In the simple segment (Figure 4), job instability continuously increases whatever the type of economy considered. The labour market tightness of this segment is also progressively reduced and unemployment rates rise. The deterioration of the labour market conditions is deeper in a French-type economy, where the probability of being fired during the diffusion process of ICT is an average 2 percentage points higher, the probability of finding a simple job is around 17 percentage points lower (market tightness is an average 36% smaller), and unemployment rates are 3 percentage points higher (reaching a 5 percentage point differential at the end of the diffusion process) than in an economy characterized by low UB.

Result 2. The diffusion of ICT does not simply contract the simple segment, it also deteriorates the situation of workers employed in it. This degradation is more marked in the presence of high UB.

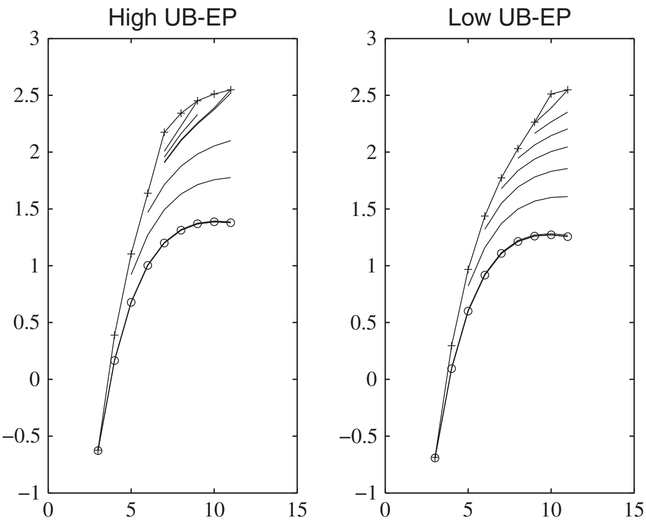

FIGURE 5. Job stability in the highest and lowest ability slots of the complex segment during the ICT diffusion process.

Key: x-axis: states of technology; y-axis: reservation productivity level ε_i^M required to pursue a match. – ○ – line: evolution of ε_i^M for $i = 1$ (highest ability slot present in the complex segment). – + – line: envelops the critical productivity levels associated with the lowest ability level present in the complex segment at each stage of technology ($a(i^w)$ or $a(i^f)$). — lines: represent the dynamics of the ε_i^M associated with the minimum ability levels ($a(i^w)$ or $a(i^f)$) that entered the complex segment at a given p and that remain for a while in it (they are no longer the lowest abilities of the segment).

Moreover, as displayed by Figure 3, the proportion of workers concerned by these worse labour market conditions is higher in the French-type economy, where a larger fraction of workers occupies simple jobs.

In the complex segment, the analysis must distinguish between the highest ($\overline{a_i} = 1$) ability level present in the segment and the lowest one ($a(\tilde{i})$), which essentially corresponds to medium-skilled workers in the economy. Job stability and labour market tightness progress very differently for the two types of worker, even if they are occupied in the same type of job.

Let us start with the dynamics of job stability. By comparing both economies (high UB vs low UB), Figure 5 allows us to draw three conclusions. First, for a given ability level, job instability is greater in an economy characterized by generous unemployment benefits. Actually, if we consider the highest ability slot ($a_i = 1$), we estimate that the probability of being fired is on average 3 percentage points higher in the presence of high UB. Second, as new technologies are diffused (increase in p), all ability slots bear a more important reservation productivity.[15] However, the largest instability is borne by the lowest skill levels in the complex segment (medium-skilled workers). Finally, in the French-type economy, the reservation productivity levels associated with each of the ability slots are less dispersed than in the presence of low UB. Actually, in an economy characterized by low UB, there is a greater heterogeneity of ability levels occupying complex positions (see Figure 3) but there is also a greater dispersion in the job stability borne by each ability. As a result, wage dispersion will also be greater in the presence of low UB than in the presence of high UB.

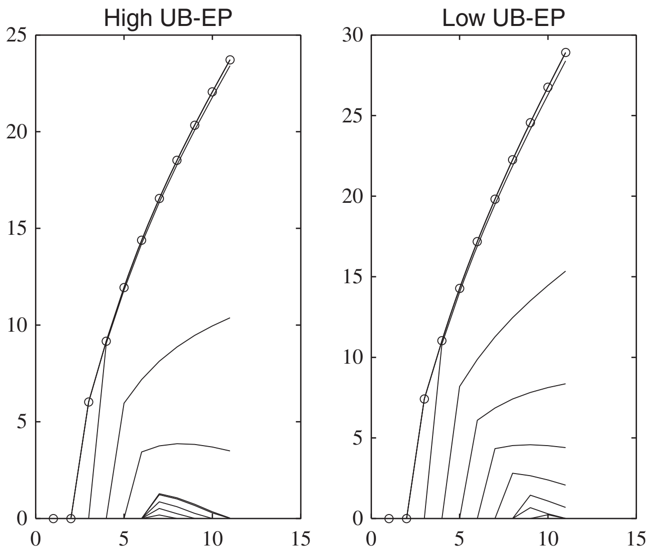

FIGURE 6. Labour market tightness of the highest and lowest ability slots of the complex segment during the ICT diffusion process.

Key: x-axis: states of technology; y-axis: market tightness, θ_i^M. – ○ – line: dynamics of the labour market tightness of the highest ability slot present in the complex segment, θ_1^M. — lines: labour market tightness associated with those ability levels ($a(i^w)$ or $a(i^f)$) that entered the complex segment at a given state of technology and that then remain in this segment (they are no longer the lowest abilities of the segment).

Figure 6 presents the progression of the labour market tightness (θ_i^M) in various ability slots as p increases. Because of the U-shaped evolution of the minimum ability level present in complex positions, some slots appear and disappear during the ICT diffusion process. This yields an inverted U-shaped path for the labour market tightness associated with some of the slots.

While θ_i^M increases sharply and continuously for the highest ability level in both economies, the attained labour market tightness is, on average, 34% greater in the presence of low UB, leading to a probability of finding a job around 25 percentage points higher. Concerning the rest of the ability levels, the more pronounced U-shaped path followed by the minimum ability level in the presence of high UB is reflected in the inverted U-shape observed for the labour market tightness. In contrast, in the economy characterized by low UB, fewer workers are excluded from the complex segment (see Figure 3), implying a smoother progression of the labour market tightness associated with the lowest ability slots (the inverted U-shape is only observed at the end of the rising path of p).

Figure 7 represents, by means of a histogram, the unemployment rates associated with all the ability slots present in the complex segment at each state of technology (see Appendix C for the computation of the unemployment rates). The larger the number of abilities included in this segment, the darker will be the bars of the histogram. The left-hand side of each of the bars corresponds to the unemployment rates of the lowest ability slots present in the complex segment. These slots stand for much lower skill levels in the presence of low UB than in the presence of high UB (see Figure 3). As a result, when comparing the histograms of both economies, we observe that those associated with the situation where UBs are low are darker on the left-hand side (they represent a larger number of ability levels).

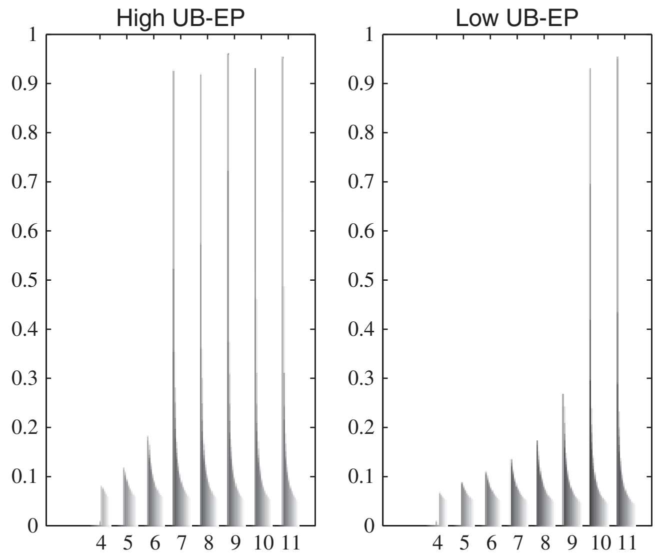

FIGURE 7. Unemployment in the highest and lowest ability slots of the complex segment during the ICT diffusion process.

Key: x-axis: states of technology; y-axis: unemployment rates, u_i^M.

The most striking result when comparing both economies is that the unemployment rates borne by the lowest ability slots of the complex segment in the French-type economy become extremely high from the early stages of the technological diffusion process. In these economies, workers prefer to remain searching in the complex segment even if their probability of finding a job converges to zero, because unemployment benefits are very high. Actually, they remain in the complex segment until they are excluded from it by firms (when firms increase the skill level required to fill a complex position, $a(i^f)$).

In contrast, in an economy characterized by low UB, the lowest ability workers prefer to search in the simple segment rather than to bear huge unemployment rates in the complex segment (the segmentation of this labour market is given by $a(i^w)$). Unemployment rates in complex positions remain thus fairly moderate until the end of the rising path of p, when UBs become sufficiently high to compensate the low probability of finding a job, and workers thus decide to remain in the complex segment.

Result 3. High UB is the main factor responsible for the high unemployment rates for medium-qualified workers traditionally occupied in complex positions. High UB reduces the incentive to search in the simple segment since it becomes more profitable for medium-skilled workers to remain unemployed in the complex segment.

All in all, estimated unemployment rates are, on average, higher in a French-type economy, 2 percentage points higher in the complex segment, and 5 percentage points higher in the simple one. Furthermore, if we define medium-qualified workers as those within the $\pm 25\%$ interval around $a(\tilde{i})$, we observe, from Figure 8, that their relative unemployment rate follows a U-shaped path such as the one observed in French data (see Figure 2 in the Introduction). Relative unemployment rates of medium-skilled workers are clearly higher in the presence of generous UB.

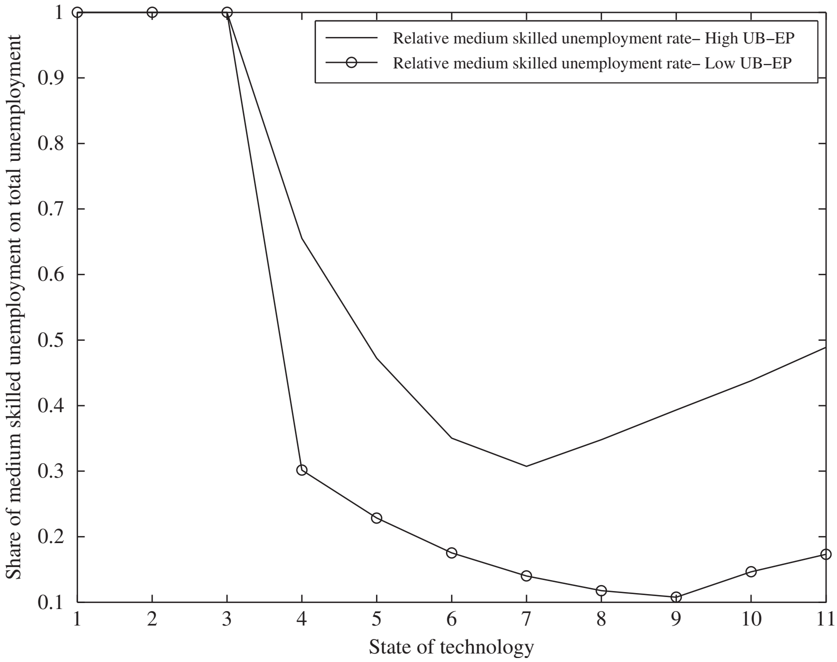

FIGURE 8. Relative medium-skilled unemployment rate during the ICT diffusion process.

VI. CONCLUSION

This paper tries to gain insights into the effect of labour market institutions on the labour market dynamics observed during the diffusion process of new technologies. More precisely, we want to determine under which conditions the generosity of the unemployment benefit system can lead to a monotonous or to a non-monotonous segmentation path of the labour market during the diffusion process of ICT. To do so, we develop an endogenous job destruction framework where new technologies asymmetrically affect productivity in simple and complex jobs, where unemployment benefits are indexed to aggregate productivity, and where there are positive spillovers linked to the expansion of complex positions (reduction in the setup costs). We find that the larger the redistributive component of the unemployment benefits, the larger the fraction of skills excluded from complex jobs, that is, the more pronounced is the U-shaped progression of the minimum skill level required in complex positions.

By working with an endogenous job destruction framework, we also provide predictions for the stability of jobs in both the complex segment and the simple segment, as well as for the evolution of unemployment rates. We find that simple positions suffer from greater job instability as new technologies are diffused. This instability increases with the generosity of the unemployment benefit system. On the other hand, people occupied in complex jobs may see their job stability increase or decrease depending on their ability level. Low-ability workers employed in complex positions experience greater job instability.

Concerning unemployment rates, high unemployment benefits induce larger aggregate unemployment rates, not only in the simple segment but also in the complex segment. Finally, our numerical simulations show that the presence of generous unemployment benefits provides an incentive to medium-qualified workers to remain unemployed in the complex segment rather than searching for a job in the simple segment. As a result, medium-qualified workers experience higher unemployment rates.

Even though the paper focuses exclusively on the role of unemployment benefits in labour market dynamics, we realize that the impact of input substitution must not be omitted, as already highlighted by Autor *et al*. (2003), Autor and Dorn (2007), Maurin and Thesmar (2005) and Goos and Maning (2007). Future research aimed at evaluating the effect of both labour market rigidities and input substitution on the segmentation of the labour market should rather consider a framework similar to Cahuc *et al*. (2007), where market frictions are introduced and firms can employ heterogeneous workers.

APPENDIX A

Table A1
Definitions of the Various Types of Tasks

Cognitive/complex tasks		Manual/simple tasks	
Routine	Non-routine	Routine	Non-routine
Workers carry out a limited and well-defined set of cognitive activities, that can be accomplished following explicit rules	Problem-solving and communication activities	Workers carry out a limited and well-defined set of manual activities, that can be accomplished following explicit rules	Workers carry out manual activities that could not be accomplished following explicit rules

APPENDIX B: THE EFFECT OF A SKILL BIASED TECHNOLOGICAL CHANGE

Simple jobs: Proof of Proposition 2 and graphical illustrations

At the equilibrium, the impact of technological progress depends on the behaviour of the job creation and job destruction curves. A variation in p shifts the job destruction curve up. This shift (see Figure A1) is given by

$$\frac{\partial p}{\partial \varepsilon^S} = \frac{\delta[1-\beta(1-\lambda\Phi(\varepsilon))]}{1-\beta(1-\lambda)} > 0.$$

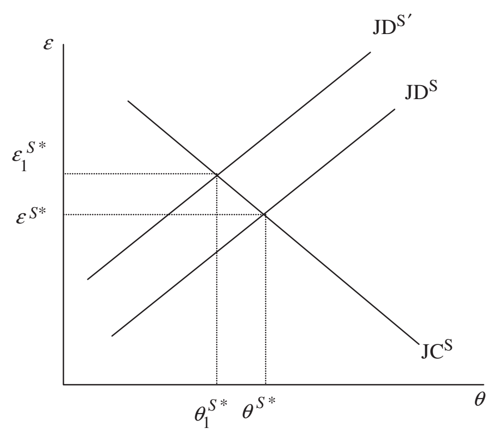

Figure A1. Effects of a biased technological in the specialized segment.

In contrast, the job creation curve does not shift. Then, the final impact of a variation in p is determined by the sign of the slope of the job creation curve, which is negative. Differentiating equation (18) with respect to ε^S yields

$$\frac{c(1-\beta(1-\lambda))}{\beta(1-\eta)}\frac{q'(\theta^S)}{q^2(\theta^S)}\frac{\partial\theta^S}{\partial\varepsilon^S} = 1.$$

Because $0<\beta<1$, $0<\eta<1$ and $c>0$, the first term on the left-hand side, $[c(1-\beta(1-\lambda))]/\beta(1-\eta)$, is positive. Therefore $\text{sign}(\partial\theta^S/\partial\varepsilon^S) = \text{sign}(q^2(\theta^S)/q'(\theta^S))$. As $q^2(\theta^S)$ is always positive and $q'(\theta^S)$ is negative, we find that $q^2(\theta^S)/q'(\theta^S)<0$. The job creation curve is negatively sloped. Then we deduce that $\partial\theta^S/\partial p<0$. □

Complex jobs: Proof of Proposition 3 and graphical illustrations

For a given ability level $a(i)\neq a(i^f)$, we differentiate both the job creation and job destruction curves of the corresponding segment, leading to

$$\text{(A1)}\qquad d\theta_i^M = Z\left[\frac{-d\varepsilon_i^M}{1-\beta(1-\lambda)} + \gamma K(p)\,dp\right],$$

$$\text{(A2)}\qquad d\varepsilon_i^M\left(1-\frac{\beta\lambda}{1-\beta(1-\lambda)}(1-\Phi(\varepsilon_i^M))\right) = \frac{\eta}{1-\eta}c\,d\theta_i^M + dp\cdot(\delta_M - a(i)),$$

where

$$Z = -\frac{\beta(1-\eta)q(\theta_i^M)}{c\dfrac{q'(\theta_i^M)}{q(\theta_i^M)}}>0$$

since $q'(\theta_i^M)<0$. Combining equations (A1) and (A2) allows us to determine how the reservation productivity will be affected at the equilibrium:

$$d\varepsilon_i^M\left[1-\frac{\beta\lambda}{1-\beta(1-\lambda)}(1-\Phi(\varepsilon_i^M)) + \frac{c\eta}{1-\eta}\frac{Z}{1-\beta(1-\lambda)}\right]$$
$$= dp\left(\frac{\eta}{1-\eta}cZ\gamma K(p) + \delta_M - a(i)\right).$$

If $[\eta/(1-\eta)]\,cZ\gamma K(p)+\delta_M>a(i)$, then $d\varepsilon_i^M/dp>0$. If $[\eta/1-\eta]\,cZ\gamma K(p)+\delta_M<a(i)$, then $d\varepsilon_i^M/dp<0$.

Concerning the impact of p on the labour market tightness, two possible cases arise.

If $d\varepsilon_i^M/dp<0$, then by equation (A1) we unambiguously obtain $d\theta_i^M/dp>0$. This corresponds to high ability levels (case B in Figure A2).

Conversely, if $d\varepsilon_i^M/dp>0$, then results are ambiguous (see equation (A1)). On the one hand, for very low ability levels, $d\varepsilon_i^M/dp>0$ is sufficiently large to overcome the impact of the setup costs on the creation curve (the term $\gamma K(p)dp$ in equation (A1)). Then $d\theta_i^M/dp<0$ (case C in Figure A2). On the other hand, for intermediate ability levels, the term $d\varepsilon_i^M/dp>0$ may not be sufficiently large to overcome the impact of the term $\gamma K(p)dp$ in equation (A1). Then $d\theta_i^M/dp>0$. Among these intermediate ability levels, there exists an $a(i)$ such that $[\eta/1-\eta]\,cZ\gamma K(p)+\delta_M = a(i)$. In this case $d\varepsilon_i^M/dp = 0$ and then $d\theta_i^M/dp>0$ (case A in Figure A2). □

APPENDIX C: THE LABOUR MARKET FLOWS AND THE EQUILIBRIUM RATE OF UNEMPLOYMENT

We assume an ability distribution function $G'(a(i)) = g(a(i))$. Total unemployment is then given by

$$\text{(A3)}\qquad U = \int_{\underline{a}}^{\overline{a}} u_i\, g(a(i))\,di.$$

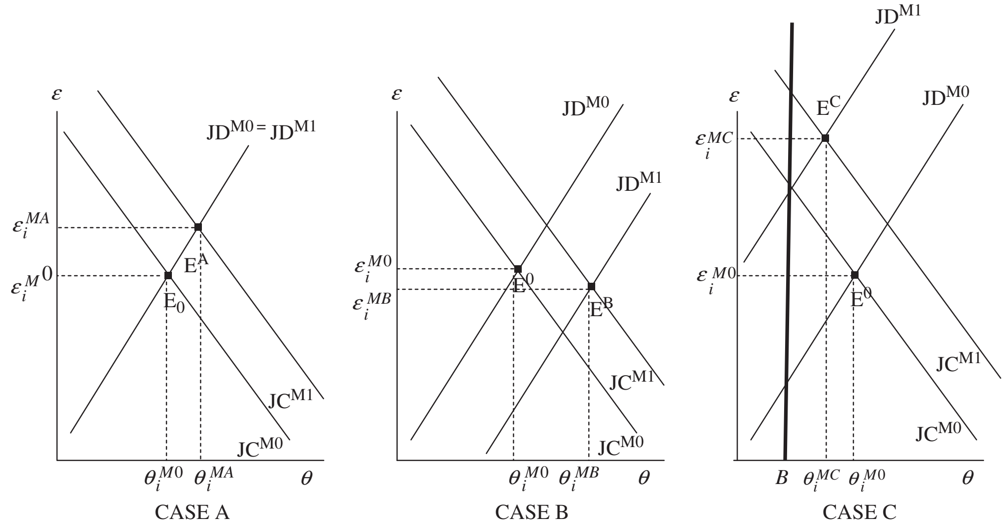

FIGURE A2. Effects of a biased technological shock on the job creation and job destruction curves for $a(i) > a(i^f)$.

If $a(i) > a(\tilde{i})$, the inflows to unemployment are hence equal to $[g(a(i)) - u_i]\lambda\Phi(\varepsilon_i^M)$, and the outflows are equal to $u_i\, p(\theta_i^M)$. In the steady state these two flows are identical, implying the equilibrium equation

(A4) $$[g(a(i)) - u_i]\lambda\Phi(\varepsilon_i^M) = u_i p(\theta_i^M) \quad \text{for all } i \geqslant \tilde{i}.$$

The number of unemployed workers with a particular skill level $a(i)$ above $a(\tilde{i})$ is then given by

(A5) $$u_i = \frac{\lambda\Phi(\varepsilon_i^M) g(a(i))}{\lambda\Phi(\varepsilon_i^M) + p(\theta_i^M)} \quad \text{for all } i \geqslant \tilde{i}.$$

If $a(i) < a(\tilde{i})$, the number of employed workers with a skill level smaller than $a(\tilde{i})$ is given by $G(a(\tilde{i})) - U_s$, where $G(a(\tilde{i}))$ stands for all workers having an ability level below $a(\tilde{i})$, and U_s represents the number of these workers being unemployed. The inflow into unemployment is equal to $(G(a(\tilde{i})) - U_s)\lambda\Phi(\varepsilon^S)$, whereas the outflow from unemployment is equal to $U_s\, p(\theta^S)$. At the steady state the inflows and outflows from unemployment must be identical:

(A6) $$(G(a(\tilde{i})) - U_s)\lambda\Phi(\varepsilon^S) = U_s p(\theta^S) \quad \text{for all } i \leqslant \tilde{i}.$$

This leads to the following equilibrium unemployment for all $i \leqslant \tilde{i}$:

(A7) $$\frac{U_s}{G(a(\tilde{i}))} = \frac{\lambda\Phi(\varepsilon^s)}{p(\theta_s) + \lambda\Phi(\varepsilon^s)}.$$

Aggregate unemployment can be obtained through the addition of low-skill and high-skill unemployment (equations (A5) and (A7)):

$$U = U^s + \int_{a(\tilde{i})}^{\bar{a}} u(a(i))g(a(i))\,\mathrm{d}a(i)$$

(A8) $$\equiv u^s \int_{\underline{a}}^{a(\tilde{i})} g(a(i))\,\mathrm{d}a(i) + \int_{a(\tilde{i})}^{\bar{a}} u(a(i))g(a(i))\,\mathrm{d}a(i)$$

$$= \int_{\underline{a}}^{\bar{a}} u(a(i))g(a(i))\,\mathrm{d}a(i),$$

where $u(a(i)) = u^s$ for all $i \leqslant \tilde{i}$. Because $a(\tilde{i})$, the labour market tightness and the critical productivity levels are known, we can directly determine the equilibrium unemployment levels.

APPENDIX D: THE EXOGENOUS JOB DESTRUCTION FRAMEWORK

If, instead of considering an endogenous job destruction framework where jobs are destroyed when the idiosyncratic productivity component is such that the surplus associated with the match becomes negative, we had considered an exogenous job destruction model, the results would have been substantially modified. In an endogenous job destruction framework, the instantaneous deterministic profit associated with a match $pa_i - p\delta_M$ may be negative, since the firm can expect a high idiosyncratic productivity shock to arrive so that the total instantaneous profit would become positive, $pa_i + \varepsilon - p\delta_M > 0$. In contrast, in an exogenous job destruction framework, $pa_i - p\delta_M$ must always be positive, since there is no uncertainty concerning any productivity component. This will have important consequences on the path followed by the segmentation of the labour market when new technologies are diffused.

The stationary equilibrium in the presence of exogenous job destruction is given by the following job creation rules:

$$\text{(A9)} \qquad \frac{c}{\beta q(\theta_i^M)} - K(p) = \frac{p \cdot a(i) - w^{M1}(a(i))}{1 - \beta(1-\chi)},$$

$$\text{(A10)} \qquad \frac{c}{\beta q(\theta^S)} = \frac{h - w^S}{1 - \beta(1-\chi)},$$

where χ stands for the exogenous job destruction rate.

Unemployment rates are obtained by equating the inflows to the outflows from unemployment. The number of unemployed workers with a particular skill level $a(i)$ above $a(\tilde{i})$ is then given by

$$\text{(A11)} \qquad u_i = \frac{\chi g(a(i))}{\chi + p(\theta_i^M)} \quad \text{for all } i \geq \tilde{i}.$$

If $a(i) < a(\tilde{i})$ (simple segment), the Beveridge curve is given by

$$\text{(A12)} \qquad \frac{U_s}{G(a(i^w))} = \frac{\chi}{p(\theta_s) + \chi}.$$

As in the endogenous job destruction framework, setting $\theta_i^M = 0$ in the job creation rule gives the equilibrium value

$$\text{(A13)} \qquad a(i^f) = \frac{1}{p}[p\delta_M + K(p)(1 - \beta(1-\chi))].$$

While being qualified to fill a complex position, all workers having a skill level within $[a(i^f), a(i^w)]$ prefer to search in the simple segment. The critical ability level $a(i^w)$ is determined by the equality $U^M(i^w) = U^S$, leading to

$$\text{(A14)} \qquad \theta_{i^w}^M = \theta^S + \frac{(1-\eta)}{c\eta}(w_S^u - w_M^u).$$

A technological diffusion exclusively favouring the productivity of complex positions decreases the labour market tightness[16] (θ^S) and then increases the unemployment rate in the simple segment. In contrast, in the complex segment, the biased technological change necessarily increases labour market tightness.[17]

Concerning the segmentation of the labour market, deriving (A13) with respect to p leads to

$$\text{(A15)} \qquad \frac{\partial a(i^f)}{\partial p} = -\frac{1}{p}\left[a(i^f) + \gamma K(p)(1 - \beta(1-\chi)) - \delta_M\right].$$

Because $\gamma K(p)(1-\beta(1-\chi))$ decreases continuously and because $a(i^f) > \delta_M$, the minimum skill level required in complex positions should continuously fall. At the limit, all ability levels should have access to positions requiring the use of new technologies.

ACKNOWLEDGMENTS

We thank the CEPREMAP and the Region Pays de la Loire for their financial support. We also thank Jean Olivier Hairault for helpful discussions, as well as all participants at the TEPP seminars and the EALE and SED conferences 2007. Any remaining errors are ours.

NOTES

1. We prefer to refer to skills or ability levels rather than to education since education can be seen as an individual choice, whereas ability levels are exogenous. With this restrictive interpretation, one can perform counterfactual simulations without considering the reaction of skill accumulation. We leave this for further research.
2. As in Mortensen and Pissarides (1999) we assume that if the skill requirement of a position is above the worker's skill level, then production is nil. Conversely, if the worker's skill level is above the skill requirement of a job, then the productivity of this position will equal the skill required by the firm. Therefore it is optimal for both agents to direct their search.
3. Note that at the end of the technological diffusion process, the recruitment cost in complex jobs is relatively small with respect to productivity. In other words, when new technologies are widely diffused, the relative cost of creating a complex position must be lower than in a situation where technologies are not diffused: this result must be interpreted as if ICT reduced the information costs induced by the search process.
4. The asset values associated with the starting period of the match differ from the continuing asset values because of the fact that the idiosyncratic productivity is assumed to be at its maximum level and wages in complex positions interiorize the existence of a setup cost.
5. For simplicity, we assume that as soon as a worker of ability level $a(i) < a(i^w)$ starts searching for a job in the simple segment, she becomes eligible for w_S^u.
6. For the first period of the match, the idiosyncratic productivity component is set to $\bar{\varepsilon}$ and wages in complex positions interiorize the existence of a setup cost.
7. The surplus sharing rule during the first period of the match equals $\eta[W^M(a(i),\varepsilon) - U^M(a(i))] = (1-\eta)[J^M(a(i),\varepsilon) - V^M(a(i)) - K(p)]$, whereas in the following periods setup costs are sunk and we thus find $\eta[W^M(a(i),\varepsilon) - U^M(a(i))] = (1-\eta)[J^M(a(i),\varepsilon) - V^M(a(i))]$.
8. In the first period of the match, the idiosyncratic productivity is set to its maximum value, such that all matches pursue at least one more period.
9. If $\delta_M < a(\tilde{i})$, all workers in the complex segment will benefit from a larger increase in their productivity than in their wages. In contrast, for $\delta_M > a(\tilde{i})$ we might have some workers for whom productivity improvements overcome the rise in wages, whereas for other workers the situation will be the opposite.
10. Note that this shift is downward bounded by $\theta_{i^w}^M = \theta^S + [\beta(1-\eta)/c\eta](w_S^u - w_M^u)$ (vertical line B in Figure A2), since all individuals in an ability slot having a labour market tightness below this bound will prefer to search in the simple segment.
11. The effective variation in the setup costs must take into account the fact that not all complex positions manage to survive after the shock on ε. On the other hand, notice that $K(p)$ is decreasing in p, and γ represents the speed of adjustment, implying that $\gamma K(p)$ stands for the reduction in the costs derived from an increased diffusion of ICT. This reduction must then be corrected by the fraction $(1-\Phi(\varepsilon_{i^f}))$ of complex positions that does not manage to survive.
12. At the final stage of the technological diffusion, the unemployment rate of high-skilled workers is around 6%, whereas that of low-skilled workers reaches 20%. Yearly average complex job destruction rates equal 14%, and those of simple jobs equal 27%. Finally, the unemployment duration of high-skilled workers is less than six months, whereas that of low qualified workers is around a year and a half.
13. The equilibrium equations in the presence of a severance tax are available upon request.
14. Note too that by giving birth to complex jobs, technological diffusion yields the appearance of more wage inequality. Initially, when only simple jobs are present in the economy, all wage differentials responded to ε. As complex jobs start appearing, two new types of inequality arise. On the one hand, there are wage differentials between wages in simple and complex jobs, arising from the fact that complex jobs respond more to p and from the fact that market tightness in the simple segment falls, whereas it increases in the complex segment. On the other hand, within the complex segment there are also wage differentials between the various ability levels.
15. For the highest ability slots we find a situation corresponding to case A of Figure A2, where the labour market increases but so does the reservation productivity level.

16. $$\frac{\partial p}{\partial \theta^S} = \frac{1}{(1-\eta)\eta}\left[-\eta c - (1-\beta(1-\chi))\left(\frac{-cq'(\theta^S)}{\beta(q(\theta^S))^2}\right)\right] < 0$$

since $q'(\theta^S) < 0$.

17. $$\frac{\partial p}{\partial \theta_i^M} = \frac{\eta c + \dfrac{-c(1-\beta(1-\chi))q'(\theta_i^M)}{\beta(q(\theta_i^M))^2}}{(1-\eta)(a_i + \gamma K(p)(1-\beta(1-\chi)) - \delta_M)}.$$

The numerator is always positive, since $q'(\theta_i^M) < 0$. Because in the exogenous job destruction framework $a_i > \delta_M$ and because $K(p)(1-\beta(1-\chi)) > 0$, we know that the denominator will always be positive, so $\partial p / \partial \theta_i^M > 0$.

REFERENCES

AGUIRREGABIRIA, V. and ALONSO-BORREGO, C. (2001). Occupational structure, technological innovation and reorganization of production. *Labour Economics*, **8**, 43–73.

ALBRECHT, J. and VROMAN, S. (2002). A matching model with endogenous skill requirements. *International Economic Review*, **43**, 283–305.

AUTOR, D. and DORN, D. (2007). Inequality and specialization: the growth of low-skill service jobs in the United States. Mimeo.

———, LEVY, F. and KEARNEY, M. (2006). The polarization of the US labor market. *American Economic Review*, **96**(2), 189–94.

———, ——— and MURNANE, R. (2003). The skill content of recent technological change: an empirical exploration. *Quarterly Journal of Economics*, **118**(4), 1279–333.

BASSANINI, A. and DUVAL, R. (2006). Employment patterns in OECD countries: reassessing the role of policies and institutions. Social, employment and migration Working Papers no. 35.

BEAUDRY, P. and GREEN, D. (2003). Wages and employment in the United States and Germany: what explains the differences? *American Economic Review*, **93**(3), 573–602.

BERMAN, E., BOUND, J. and GRILICHES, Z. (1994). Changes in the demand for skilled labor within U.S. manufacturing: evidence from the Annual Survey of Manufacturers. *Quarterly Journal of Economics*, **109**, 367–97.

CAHUC, P., MARQUE, F. and WASMER, E. (2007). A theory of wages and labor demand with intrafirm bargaining and matching frictions. CPER Discussion Paper no. 4605.

CAROLI, E. and VAN REENEN, J. (2001). Skilled biased technological change? Evidence from a panel of British and French establishments. *Quarterly Journal of Economics*, **116**(4), 1449–92.

DOLADO, J., FELGUEROSO, F. and JIMENO, J. (2000). Youth labor markets in Spain: education, training and crowding-out. *European Economic Review*, **44**, 943–56.

GAUTIER, P. A. (2002). Unemployment and search externalities in a model with heterogeneous jobs and heterogeneous workers. *Economica*, **69**, 21–40.

GOOS, M. and MANING, A. (2007). Lousy and lovely jobs: the rising polarization of work in Britain. *Review of Economics and Statistics*, **89**(1), 118–33.

GREENWOOD, J. and JOVANOVIC, B. (2000). Accounting for growth. Rochester Center for Economic Research Working Papers no. 475.

LINDBECK, A. and SNOWER, D. (2000). Multitask learning and the reorganization of work: from Tayloristic to holistic organization. *Journal of Labor Economics*, **18**(3), 353–76.

MACHIN, S. and VAN REENEN, J. (1998). Technology and changes in skill structure: evidence from seven OECD countries. *Quarterly Journal of Economics*, **113**, 1215–44.

MAURIN, E. and THESMAR, D. (2005). Changes in the functional structure and demand for skills. *Journal of Labor Economics*, **22**(3), 639–64.

MORENO-GALBIS, E. and SNEESSENS, H. (2007). Unemployment, capital-skill complementarity and embodied technological progress. *Louvain Economic Review*, **73**, 241–72.

MORTENSEN, D. and PISSARIDES, C. (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, **61**, 397–415.

——— and ——— (1999). Unemployment responses to skill-biased technology shocks: the role of labor market policy. *Economic Journal*, **109**(April), 242–65.

NGAI, R. and PISSARIDES, C. (2007). Structural change in a multi-sector model of growth. *American Economic Review*, **97**, 429–43.

PISSARIDES, C. (1990). *Equilibrium Unemployment Theory*. Cambridge, MA: MIT Press.

SPITZ-OENER, A. (2006). Technical change, job tasks and rising educational demands: looking outside the wage structure. *Journal of Labor Economics*, **24**(2), 235–70.