CESifo
Economic Studies

Volume 53 Number 1 March 2007

www.cesifo.oxfordjournals.org

Contents

Editorial Announcement

At the end of 2006, David Wildasin retired as editor of the journal. Starting in 2007, John Whalley has joined our team of editors.

Gerhard Illing

Managing editor

CESifo Economic Studies, Vol. 53, 1/2007, 2–41, doi:10.1093/cesifo/ifm001
Advance Access publication 27 February 2007

A Framework for Independent Monetary Policy in China

Marvin Goodfriend* and Eswar Prasad†

Abstract

As the Chinese economy becomes more market based and continues its rapid integration into the global economy, having an independent and effective monetary policy regime oriented to domestic objectives will become increasingly important. Employing modern principles of monetary policy in light of the current state of China's financial institutions, we motivate and present a package of proposals to guide the operation of a new monetary policy regime. Specifically, we recommend an explicit low long-run inflation objective, operational independence for the People's Bank of China (PBC) with formal strategic guidance from the government, and a minimal set of financial sector reforms (to make the Chinese banking system robust against interest rate fluctuations). We argue that anchoring monetary policy with an explicit inflation objective would be the most reliable way for the PBC to tie down inflation expectations, and thereby enable monetary policy to make the best contribution to macroeconomic and financial stability, as well as economic growth. The management and monitoring of money (and credit) growth by the PBC would continue to play a useful role in the stabilization of inflation, but a money target would not constitute a good stand-alone nominal anchor. (JEL codes: E5 and P2)

1 Introduction and overview

As China's economy develops and becomes more market oriented, and as its integration with the world economy continues, monetary policy will need to shoulder an increasingly large burden in ensuring stable, non-inflationary growth. Rising integration, for instance, implies greater vulnerability to external shocks, and monetary policy is typically the first line of defense against such shocks. Although deeper structural reforms may be the key determinants of long-term growth, monetary policy has an

* Tepper School of Business, Carnegie Mellon University.
† Research Department, IMF, e-mail: eprasad@imf.org, and Department of Applied Economics and Management, Cornell University, e-mail: eswar.prasad@cornell.edu.
We thank Ray Brooks, Steven Dunaway, John Fernald, Jiming Ha, Robert Hetzel, Lamin Leigh, Allan Meltzer, Ronald McKinnon, Mark Stone, Sun Tao, Tao Zhang, Yanchun Zhang, and seminar participants at the Federal Reserve Banks of Cleveland and San Francisco for useful comments. We are also grateful to numerous other colleagues at the IMF, including participants in various internal seminars, for helpful discussions and comments. The comments of the editor, Gerhard Illing, and two anonymous referees helped sharpen the presentation in the article. Wellian Wiranto provided excellent research assistance. The views expressed in this article are solely those of the authors and do not necessarily reflect those of the IMF or IMF policy.

important role to play in creating a stable macroeconomic environment that is essential for those reforms to take root.

Monetary policy in China has in recent years operated under difficult constraints, including a fixed exchange rate regime, an underdeveloped financial system and numerous institutional weaknesses. Having an independent monetary policy is an important policy priority. Maintenance of an exchange rate regime with limited *de facto* flexibility exposes the economy to significant risks of macroeconomic instability. While capital controls provide some room for maneuver for monetary policy even under such a regime, this room tends to be limited in practice and could result in inadequate control of investment growth and inflationary (or deflationary) pressures. Moreover, the effectiveness of capital controls inevitably erodes over time as domestic and international investors find channels, including rising trade, to evade them.

These considerations have led the authorities to initiate a move towards a more flexible exchange rate regime. On 21 July 2005, the renminbi was revalued by 2.1 percent relative to the US dollar and it was announced that the value of the renminbi would henceforth be set with reference to a basket of currencies rather than having it pegged to the dollar. Since then, however, the renminbi has been maintained at a stable level relative to the dollar, indicating limited *de facto* flexibility. Nevertheless, the authorities have clearly stated their intention to allow for greater flexibility over time; recent fluctuations in the exchange value of the renminbi confirm this intention.

An important consequence of the move towards a flexible exchange rate is the need to adopt a new nominal anchor and an associated strategy for monetary policy. In this article, we make the case that China should adopt a low inflation objective as the new nominal anchor. Moreover, we conclude that, given the relative merits of an inflation objective and the potential problems associated with maintaining a fixed exchange rate, there are good reasons for China to adopt this new anchor expeditiously.

Theory and experience from around the world—from both advanced industrial economies as well as emerging market economies—suggest that making low inflation the main objective of monetary policy is the most reliable way to enable the People's Bank of China (PBC) to stabilize domestic inflation and employment against macroeconomic shocks. An inflation objective can accommodate fluctuations in productivity growth and changing relationships between monetary or credit aggregates and inflation, all of which are relevant considerations for a developing economy. It also has the virtue of easy communicability.

We are not advocating a full-fledged inflation targeting regime, although this could serve as a useful long-term goal. Our approach is

more practical for the foreseeable future, and it should deliver most of the benefits of formal inflation targeting. In light of the changing structure of the economy and weaknesses in the monetary transmission mechanism, our framework could accommodate a continued role for the monitoring and management of monetary (and credit) aggregates by the PBC. But money would not constitute a good stand-alone nominal anchor since the changes in China's economic structure and financial markets imply that the rate of money growth consistent with a stable rate of inflation is likely to be highly variable.

Can this framework, which accords primacy to a low inflation objective, be reconciled with the broader mandate of the PBC? The PBC Law states that "Under the guidance of the State Council, the PBC formulates and implements monetary policy, prevents and resolves financial risks, and safeguards financial stability". And would a low inflation objective be consistent with promoting sustained high employment growth, a key consideration for Chinese policymakers? Our response is that it is precisely by providing a firm and credible nominal anchor through a low inflation objective that the PBC can best contribute to overall macroeconomic stability, and best provide for sustained employment growth and financial stability.

Recent academic research and policy experiences lend strong support to this view. US Federal Reserve Board Chairman Ben Bernanke has articulated a similar position. In his confirmation hearings before the US Senate (Bernanke 2005, p. 2), he said:

> I view the explicit statement of a long-run inflation objective as fully consistent with the Federal Reserve's current policy approach, including its appropriate emphasis on the role of judgment and flexibility in policymaking. Most important, this step would in no way reduce the importance of maximum employment as a policy goal. Indeed, a key justification for this action is its potential to contribute to stronger and more stable employment growth by further stabilizing inflation and inflation expectations.

Although we are not advocating formal inflation targeting for China, some of the requirements of that regime are important for a low inflation objective as well. Principal among these is instrument (operational) independence for the central bank. The PBC should have the authority and the capability to use its monetary policy instruments, e.g., bank reserves or an interest rate, to credibly anchor inflation and stabilize the macroeconomy in general. We do not believe that broader independence for the PBC is essential, although the PBC must be empowered to build up institutional capacity necessary to support its monetary policy mission, and given the financial resources to do so. It is essential, however, that the

Chinese government explicitly acknowledge its support for a low inflation objective as the nominal anchor for monetary policy.

What would it take to put in place a low inflation objective as an effective nominal anchor? Exchange rate flexibility is of course a prerequisite for an independent monetary policy oriented to domestic objectives. But a move towards greater exchange rate flexibility is hardly the solution by itself. Indeed, enhancing the effectiveness of the monetary transmission mechanism poses difficult challenges independent of the constraints related to the exchange rate regime. Principal among these is the reform of the financial system, since it is through the banking system that monetary policy must influence economic activity.

The Chinese state-owned banking system has long labored under lending directives from the government. Progress has been made since the late 1990s in improving the commercial orientation of the banking sector, and significant strides have been made in improving banking supervision and regulation. But Chinese banks are still far from being robust commercially driven financial entities. Given the dominance of the banking sector in China's financial landscape, this has important implications for monetary policy transmission.

The essence of the challenge is to transform the banking system from an off-budget arm of fiscal policy—which uses captive savings of households to support state enterprises, whether commercially viable or not—into a banking system that can direct credit prudently to its most valued uses given correct interest rate signals. Even in the best of circumstances, it will take years for China to put in place all of the components of a modern, efficient banking system. This is especially so when one recognizes that the transition process must be supervised and regulated with great care to preserve the public's confidence in the banks and guard against moral hazard problems associated with explicit or implicit deposit insurance provided by the government. The consequences of the legacy of directed lending will also inevitably complicate the transition.

Nevertheless, we believe that it is both feasible and desirable for China to put in place a minimal set of financial sector reforms and regulations that would enable it to adopt an independent monetary policy with low inflation as the nominal anchor. These reforms would be aimed at giving the PBC full control of its balance sheet so that the central bank could manage bank reserves solely for monetary policy purposes. The reforms and regulations would also need to ensure that banks could withstand the financial stress that results from fluctuations in interest rates necessary to stabilize the macroeconomy and maintain price stability. We believe that reforms could be put in place in the next few years to achieve these ends and serve as an adequate foundation for independent monetary policy.

Our proposal has three additional attributes. First, it would allow for continuity in the operational approach to monetary policy. The PBC could continue its current operations and gradually adapt its procedures to the pursuit of independent monetary policy as supporting reforms are put in place. Our proposal would mainly entail a shift in strategic focus to a well-defined inflation anchor. Second, under present circumstances, the shift to an inflation anchor would be seamless since it would involve merely locking in the current low rate of inflation. Third, the adoption of effective independent monetary policy would facilitate various reforms that have intrinsic benefits of their own. For instance, the resulting macroeconomic stability would facilitate the modernization of the financial system. In addition, the new policy regime would necessitate improvements in the statistical base that would enhance public sector transparency and encourage better communication about policy intentions.

Our main goal in this article is to make the conceptual case for a low inflation objective as the nominal anchor for independent monetary policy in China. There are of course a number of important practical details that would need to be worked out, including the appropriate level and width of the target range for the inflation objective, the best method for communicating this objective, the appropriate measure of inflation to be used, etc. Other than discussing some of these issues from an analytical perspective, we leave the specifics as open questions to be addressed in future work.

In the next section, we address a basic issue: why an inflation target is preferable to a fixed exchange rate or a stand-alone money growth target as the nominal anchor for monetary policy. In section 3, we use the modern theory of monetary policy to review the macroeconomic principles underlying the case for adopting an inflation objective as the nominal anchor. We explain why and how stabilizing inflation also stabilizes employment and allows an economy to grow at its potential. And we describe the mechanics by which monetary policy actions work to stabilize employment and inflation. In section 4, we describe the institutional support that a central bank needs in order to implement independent monetary policy effectively.

In section 5, we describe the main features of current monetary and banking institutions in China in order to identify constraints in the Chinese financial sector that would impede the effective adoption of independent monetary policy. In section 6, we lay out our proposal for China to adopt an independent monetary policy with low inflation as its nominal anchor. In particular, we recommend a minimal set of financial sector and other reforms that China could put in place in a few years to credibly sustain low inflation and enable monetary policy to make its maximum contribution to macroeconomic stability and economic growth.

2 Low inflation objective as nominal anchor

An inflation objective—an explicit or implicit long-run range for the inflation rate and an acknowledgment that low inflation is a priority for monetary policy—has emerged as the leading nominal anchor in practice around the world. We begin by reviewing the advantages of an inflation objective compared to its two leading competitors for nominal anchor—a fixed nominal exchange rate and a money growth target. Our reasoning applies equally well to emerging market economies such as China as it does to advanced industrial economies. Note that, for the purposes of this discussion, we have adopted an expansive definition of a low-inflation objective, encompassing full-fledged formal inflation targeting regimes (e.g., Canada and New Zealand) as well as implicit inflation targeting regimes (e.g., the United States).[1]

2.1 Disadvantages of alternative nominal anchors

A fixed nominal exchange rate regime borrows the nominal anchor from abroad. If the partner country anchors *its* price level or inflation rate, domestic inflation is well anchored, too. Otherwise, the fixed exchange rate system imports the variability in foreign inflation or deflation. Even if foreign inflation remains low on average, the fixed exchange rate forces domestic policymakers to accept one of two unpleasant alternatives. Either domestic interest rates follow foreign rates closely irrespective of domestic circumstances, or the home country must impose controls on private international capital flows to try and insulate itself. Moreover, once the viability of a fixed rate regime is called into question, the expected returns to attacking the fixed rate make the regime susceptible to destabilizing speculative flows. For all of the above reasons, fixed exchange rate regimes have tended to be fragile in practice, even when supported by capital controls; and so a fixed nominal exchange rate has proven to be an unreliable nominal anchor for monetary policy (see Obstfeld and Rogoff 1995).

What of a money growth target as the nominal anchor? Monetary targets were employed by the US Federal Reserve, among other central banks, to stabilize inflation in the 1970s and 1980s (see Goodfriend and King 2005). However, central banks in most countries with well-developed banking and financial systems now employ interest rates to target low inflation, largely ignoring money growth. Because money growth and inflation tend to be highly correlated in the long run, targets for deposits, bank reserves and the monetary base can still play a useful operational

[1] See Bernanke et al. (1999) for a discussion of inflation targeting around the world, and Goodfriend (2005) on implicit inflation targeting in the United States.

role in stabilizing inflation in countries with underdeveloped financial systems and less reliable interest rate channels of monetary transmission. Nevertheless, developing economies, in particular, are subject to financial innovations that shift the short-run relationship between monetary aggregates, on one hand, and economic activity and inflation on the other. Moreover, money growth targets must be adjusted from time to time in order to accommodate changes in potential output growth and trend velocity in order to sustain low and stable inflation. Thus, a money growth target would not by itself constitute a stable anchor for inflation expectations.

2.2 Country experiences with different anchors

Numerous major central banks around the world—e.g., the Eurosystem and the Bank of England—now employ an explicit target range for inflation as the nominal anchor for monetary policy. Others, such as the US Federal Reserve, anchor monetary policy with an implicit low inflation objective. Where central banks have been given the necessary institutional support to stabilize inflation, they have been able to acquire an impressive degree of credibility for low inflation that has anchored inflation expectations firmly and contributed importantly to stabilizing actual inflation (see, e.g., Bernanke et al. 1999; Goodfriend 2005). Under such circumstances, an inflation objective appears to generate a virtuous circle of credibility. This is evident in the US, where the Federal Reserve's implicit targeting of low inflation has made an important contribution to macroeconomic stability.

Emerging market and developing economies that adopted various forms of inflation targeting, have in general had similar positive experiences. As noted by Mishkin (2000), some advantages of inflation targeting are particularly relevant for emerging market economies: (i) a stable relationship between money and inflation is not critical to its success; (ii) inflation targeting is easily understood by the public and is transparent; and (iii) inflation targeting focuses the political debate on what monetary policy can (and can not) achieve in the long-run and away from the temptation to use monetary policy to stimulate employment growth in the short-run.

Mishkin lists three major potential problems for inflation targeting in emerging markets; (i) inflation may be hard to control because of underdeveloped financial systems; (ii) inflation targeting requires fiscal policy to support the inflation target; and (iii) the exchange rate flexibility required for inflation targeting might be difficult for policymakers to allow. Based on the experiences of a few inflation targeting emerging market economies, Mishkin concludes that these problems can be overcome in practice. Moreover, a large and growing empirical literature

investigating the macroeconomic performance of emerging market economies under different monetary frameworks finds support for inflation targeting in practice.[2]

3 Principles of monetary policy geared toward targeting inflation

We now summarize briefly the principles of monetary policy geared toward sustaining an objective for low inflation.[3] We explain why stabilizing inflation also stabilizes employment over the business cycle and accommodates economic growth. And we emphasize the complications and virtues of the fact that inflation targeting depends heavily for its effectiveness on the credibility of the central bank's commitment to low inflation. We use the principles of the modern theory of monetary policy to guide our discussion; these concepts are quite general and apply, at a basic level, even to a socialist market economy such as China.

The modern theory of monetary policy has at its core monopolistically competitive firms that set product prices at a markup over marginal production costs (wages and materials costs) that is expected to maximize profits over time. Firms consider changing product prices only if demand and cost conditions threaten to compress or elevate markups significantly and persistently relative to their profit-maximizing levels. Such reasoning yields the fundamental principle of price stability: Monetary policy geared toward sustaining low inflation must manage aggregate demand so that production costs rise at the targeted rate of inflation—then firms will raise product prices at the targeted rate of inflation because they are confident that doing so will keep actual markups at profit-maximizing markups.

The principle of inflation targeting given earlier has the important implication that monetary policy geared to targeting inflation yields the best cyclical stabilization of employment (Goodfriend and King 1997; Woodford 2003). Thus, even those who care mainly about the stabilization of employment can support a low inflation objective for monetary policy. But these points also have the important implication that the best that monetary policy can do is to stabilize inflation; it should not

[2] See, e.g., Jonas and Mishkin (2005), Mishkin and Schmidt-Hebbel (2005) and IMF (2006). For a skeptical view, see Blanchard (2005).

[3] This exposition draws on Goodfriend (2002) and Broaddus and Goodfriend (2004).

be used to try to counteract fluctuations in output and employment that are due to shocks to productivity and other factors affecting aggregate supply.

Such reasoning assumes that all goods are produced by monopolistically competitive firms whose prices are sticky. In practice, products such as oil and food are produced and traded in highly competitive markets where shocks to supply and demand impact overall inflation directly. In order to stabilize overall inflation against a positive shock to the flexible-price sector, monetary policy would have to depress aggregate demand sufficiently, which would then raise the markup and create an output gap in the sticky-price sector.

Thus, monetary policy would appear to face a trade-off between inflation and employment variability in the short run with respect to shocks to the flexible-price sector. However, the problem can be circumvented if the central bank targets an objective for low *core* inflation (excluding oil and food prices, in the example above). The economy can then adjust to changes in the relative prices of oil and food while core inflation and employment are both stabilized. This would be a more stable nominal anchor than an overall inflation objective, especially at short to medium horizons, and serve as a better anchor for inflation expectations. Moreover, stabilizing core inflation and letting other prices adjust would make the economy operate most like a flexible-price economy.

Note that any rate of productivity growth is compatible with a particular inflation objective. The logic is as follows. Production costs rise over time at the rate of wage inflation minus the rate of labor productivity growth. When actual markups are stabilized at profit-maximizing markups, firms are content to raise product prices at the rate of wage inflation minus the rate of labor productivity growth. Competition for labor tends to raise wages on average over time at the rate of labor productivity growth plus an adjustment for inflation. Hence, any rate of productivity growth is consistent with the maintenance of the profit-maximizing markup and the targeted rate of inflation on average over time. The capability of a nominal anchor to accommodate highly variable productivity growth is important for a rapidly developing country like China. The discussion earlier indicates that Chinese monetary policy can and should maintain a low objective for inflation even if productivity growth fluctuates over a wide range.

In addition to the direct benefits of an inflation objective described earlier, it can improve macroeconomic performance indirectly by tying down inflation expectations. Imperfect credibility for low and stable inflation makes an economy vulnerable to fluctuating beliefs about inflation or deflation, which, in extreme cases, can take the form of

inflation or deflation scares.[4] Inflation scares present a central bank with a dilemma. Ignoring them encourages even more doubt about a central bank's commitment to low inflation. However, to restore credibility for low inflation, a central bank may have to tighten monetary policy aggressively—at the risk of precipitating a recession—in order to make sure that inflation doesn't spiral out of control.

Deflation scares also involve a credibility problem—the possibility that interest rate policy might be immobilized at the zero bound and that the central bank might be unwilling or unable to act against deflation with unconventional monetary policy at the zero bound. Moreover, a policy vacuum at the zero bound could encourage ill-advised fiscal policy actions, as well as anti-competitive policies, that could reduce potential GDP growth. The threat to potential output exacerbates a deflation scare by lowering future income prospects and causing households and firms to cut current spending, which reduces labor demand, lowers wages, and elevates markups further. Thus, a deflation scare is problematic because, like an inflation scare, it has the potential to lead to a protracted recession.

In short, inflation targeting depends heavily for its effectiveness on the credibility of the central bank's perceived power, will, and competence to maintain low and stable inflation. The reliance on credibility is a complication because a monetary policy regime geared to targeting low inflation needs substantial institutional support to guarantee the credibility of the central bank's commitment to low inflation. In addition, a country must have in place a fiscal policy regime which is not expected to need or resort to inflationary finance in the future.

However, once credible institutional support for inflation targeting has been created, the reliance on credibility is a virtue. If the public believes that the central bank has the power and the scope to use monetary policy to maintain stable inflation, then firms will be inclined to keep price adjustments on target because they will regard any deviations of actual markups from profit-maximizing markups as temporary. Thus, credibility for low inflation tends to be self-enforcing to a considerable extent. With credible institutional support in place, markets tend to be relatively

[4] Goodfriend (1993) and Goodfriend and King (2005) document how inflation scare shocks have destabilized the US economy in the past. Although inflation in China has been reasonably stable of late, China's earlier experiences suggest that it is potentially susceptible to both inflation and deflation scares. For instance, the Chinese economy endured a burst of high inflation in 1994–5, when inflation exceeded 20 percent (Figure A2). The economy experienced deflationary episodes in 1998–99 and 2002.

forgiving of temporary tactical policy mistakes that may be committed by a central bank as it acts to stabilize inflation.

4 Institutional support for independent monetary policy

A variety of institutional preconditions are needed to support operating procedures to enable a central bank to pursue independent monetary policy with a low inflation anchor. A central bank must have instrument independence—the authority and will to use its policy instruments to act quickly and decisively in response to incoming data—to maximize the potential for monetary policy to stabilize inflation, inflation expectations and employment, and to ensure financial market stability. In particular, a central bank must be prepared to move short-term interest rates quickly and aggressively over a large range if necessary. Even in the US, which has maintained low inflation consistently for two decades now, the Federal Reserve has had to allow short-term interest rates to fluctuate in a wide range to maintain low inflation.

In order for monetary policy to consistently preempt fluctuations in inflation around an inflation objective, a central bank must utilize modern statistical techniques together with comprehensive, timely, and reasonably accurate statistical indicators of macroeconomic conditions to guide its policy actions. A central bank needs reliable measures of inflation and indicators of the direction of pressures on future inflation. Such indicators could include aggregate price markups, estimates of the gap between actual and potential output, estimates of capacity utilization in the manufacturing sector, measures of employment growth relative to estimated trend labor force growth, and indicators of inflation expectations. In general, it would help to guide monetary policy to keep track of the growth of various financial aggregates such as the monetary base, bank reserves, bank deposits, and loans against estimated growth rates believed to be consistent with low inflation. Moreover, a central bank must develop techniques to produce efficient conditional forecasts of inflation and output to inform policy decisions.

Finally, the government should grant the central bank instrument independence with strategic guidance directing monetary policy to be used flexibly to stabilize employment and financial markets, subject to inflation remaining in or near an explicit low inflation objective. A public understanding of the commitment to low inflation is necessary to assure its credibility. The credibility of that commitment requires central bank instrument independence to achieve it, the government's agreement to support it, and a role for oversight by government and markets to hold the central bank accountable for carrying it out.

As a mechanical matter, monetary policy is implemented by managing the aggregate supply of bank reserves, which are deposits of commercial banks at the central bank. These include required reserves plus any excess reserves beyond those that satisfy reserve requirements. A central bank must have *full* control of aggregate bank reserves to stabilize inflation expectations credibly. Control of bank reserves is *necessary* because a central bank must manage aggregate demand over the business cycle by manipulating the supply of bank reserves.

A central bank's control of bank reserves is compromised when it is obliged to acquire or sell assets for reasons other than managing aggregate bank reserves to stabilize inflation. In general, there are three reasons why a central bank might have to do so: (i) it might be directed to buy government debt, i.e., to finance a government deficit in whole or in part with newly created bank reserves; (ii) it might be directed to lend to banks, nonfinancial firms, or state enterprises; or (iii) it might be obliged to buy foreign assets to support a managed or fixed exchange rate. For instance, when a nation such as China chooses to manage its foreign exchange rate within a tight range, the central bank must accommodate the market's excess demand or supply of foreign exchange at the stabilized exchange rate by creating or draining bank reserves.[5]

It may be possible for the central bank to offset, or sterilize, the effect of the required asset action on aggregate bank reserves by taking an opposite asset action with another asset or liability on its balance sheet. Sterilization of foreign exchange flows, however, leaves the exchange rate and domestic interest rates unaffected, and does little to mitigate the factors that gave rise to those flows in the first place. Even when supplemented with capital controls, sterilization of inflows leads to rising quasi-fiscal costs or other implicit costs associated with financial repression. And the buildup of foreign exchange reserves exposes the central bank balance sheet to risks of capital losses associated with exchange rate and interest rate fluctuations.

To reiterate, the central bank needs to be free of any significant obligations that compromise its ability to manage aggregate reserves to stabilize inflation. In particular, monetary policy credibly geared toward targeting low inflation must be accompanied by a willingness on the part of the government and the public to allow a substantial degree of flexibility in the foreign exchange rate, so that exchange rate adjustments, and not central bank purchases and sales of foreign assets, can allow the foreign exchange market to clear.

---

[5] We follow the usual presumption—that the demand for currency is unaffected by capital flows—so that unsterilized purchases or sales of foreign exchange by a central bank show up as changes in bank reserves.

Even if a central bank has full control of aggregate bank reserves, it must be willing to use its control of bank reserves to move short-term interest rates in a relatively wide range, aggressively at times, to sustain credibility for low inflation in order to implement monetary stabilization policy effectively. An inclination to smooth short-term interest rates to cushion the banking system against financial stress would compromise the central bank's ability to manage aggregate demand to sustain low inflation, severely undermine the credibility of its inflation target, and destabilize employment and inflation over the business cycle.[6]

The banking system must be financially robust to fluctuations in short-term interest rates so that the central bank is willing to move short-term interest rates as needed to manage monetary policy effectively. Even if a country has in place credible deposit insurance that protects the viability of the banking system, a central bank might be reluctant to make full use of its instrument independence against inflation if raising interest rates triggers cash flow problems for banks, with the potential to precipitate a public sector bailout. Thus, it is essential for the credibility of monetary policy geared to targeting low inflation that financial vulnerabilities of the banking system to high interest rates and to large fluctuations in interest rates be corrected.

Finally, assuming that banks are well capitalized, managed, and regulated so that they lend funds prudently, rates on bank deposits and loans as well as nonbank money market instruments should be deregulated and free to reflect the cost of loanable funds in the interbank market. This would broaden the channels by which monetary policy is transmitted to the economy and minimize disruptions in banking and credit flows that would otherwise occur because rigid interest rates caused disintermediation in certain credit markets.

5 Monetary and banking institutions in China

In this section, we present a broad overview of current monetary and banking institutions in China in order to assess the adequacy of the Chinese financial system to support independent monetary policy. We highlight certain aspects of the banking and broader financial systems that will command our attention in section 6, where we propose a package

[6] See Goodfriend and King (2005). Poole (1978, pp. 105–10), describes how interest rate smoothing contributed to the destabilization of inflation and output in the United States in the 1960s and 1970s.

of reforms that would enable China to adopt an independent monetary policy.

5.1 The banking and financial systems

The financial landscape in China is dominated by the state-owned banking system.[7] The stock and bond markets have rather limited roles. Total deposits in the banking system amount to about 160 percent of GDP, compared to an outstanding stock of government debt of about 25 percent of GDP (Figures 1 and 2). The corporate bond market is small and the stock market is relatively thin as well. With only a small number of enterprises permitted to list and about two-thirds of shares in listed enterprises held by the state and not traded, the stock market does not play a major role in intermediating household saving into corporate investment.[8]

The banking system has been dominated by four large state commercial banks (SCBs) that together account for more than half of the total assets of the banking system (56 percent as of end-2005; this share has been declining in recent years). The joint stock commercial banks (JSCBs) have expanded the size of their balance sheets quite rapidly in recent years and now account for about one-fifth of total banking system assets. A few policy banks such as the China Development bank have explicit directed lending mandates. The banking system is rounded out by a number of smaller banks, including rural and urban credit cooperatives (see Barnett 2004, for a fuller description of the structure of the Chinese banking system).

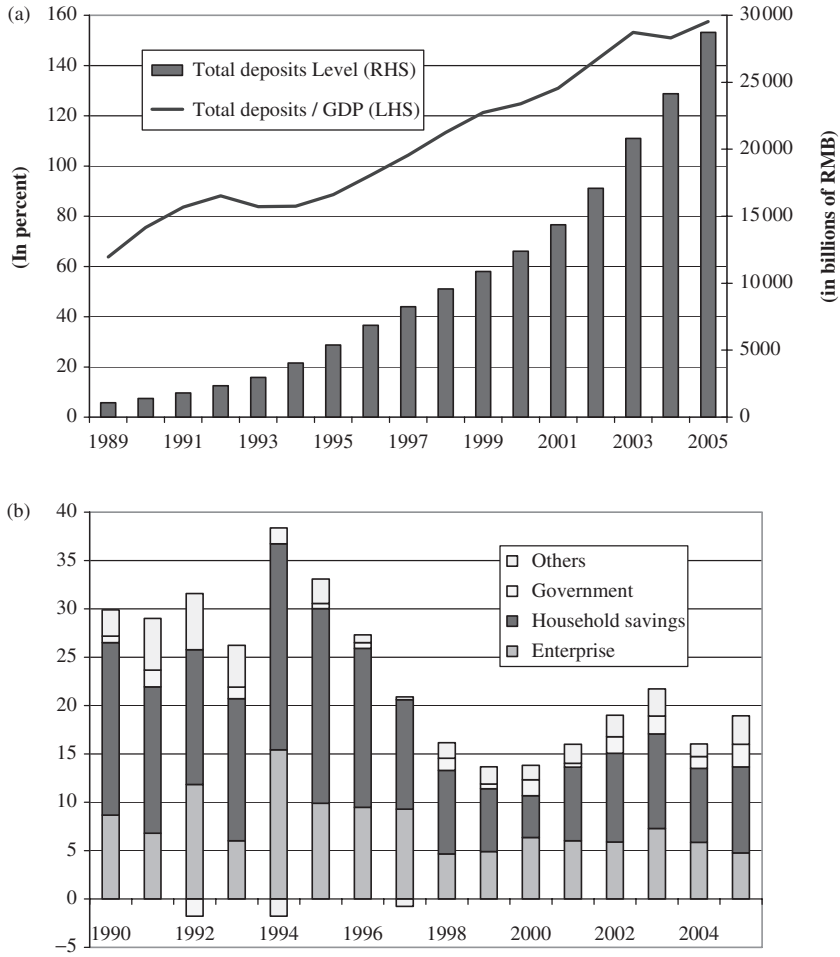
During the 1980s and through the late-1990s, Chinese banks, including the SCBs, were provided explicit official guidance on their lending operations. This approach of ''directed lending'' favored large state-owned enterprises (SOEs). Such SOEs were seen as important not only for employment generation but also because—by providing benefits such as housing, education, health and pensions to their workers—they served an important role in the delivery of social services. The availability of such directed lending meant that commercially unviable enterprises could continue getting funding for working capital and even for new investments.

The policy of directed lending was terminated around 1998. However, the legacy of directed lending stayed with the banks in the form of

---

[7] Allen, Qian and Qian (2005) argue that there is also a large unregulated informal financial sector that plays an important role in financial intermediation in China.

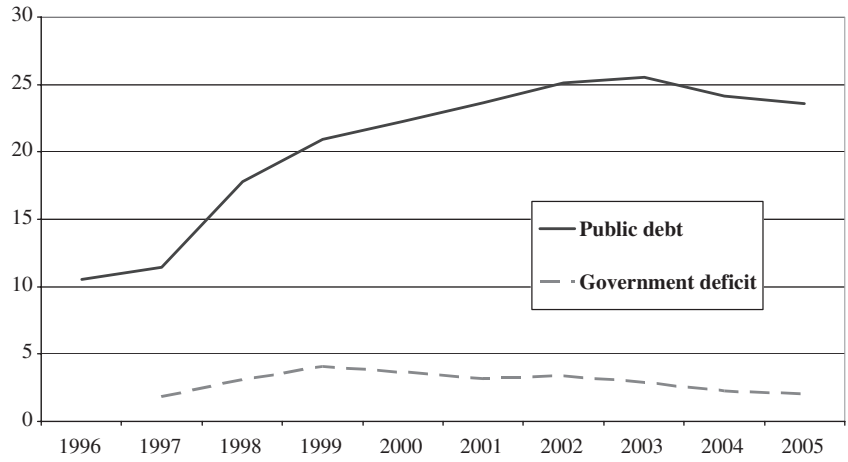
[8] The PBC's Monetary Policy Report for the fourth quarter of 2005 notes that, of the funds raised in the domestic financial market in that quarter, bank loans account for 78 percent, corporate bonds and stocks account for about 6 percent each, and the rest are government securities. In the first quarter of 2005, bank loans accounted for as much as 89 percent and corporate bonds and stocks for only about 1 percent.



Source: CEIC and authors' calculations. Government share of deposits include fiscal deposits as well as deposits by government agencies and organizations. Others refer to rural and trust deposits, and remaining components. Deposits figures are based on end-of-period data, and do not include foreign currency deposits.

Figure 1 (a) Total deposits in financial institutions (b) Growth of total deposits and contribution of components (in percent)

nonperforming loans (NPLs). Estimates of the stock of NPLs in the banking system vary across a wide range and are subject to numerous measurement problems (see Barnett 2004). But the recognition that the stock was large and a major hindrance to banking reforms led to a carve-out of some NPLs from the SCBs to asset management companies



Source: International monetary fund

Figure 2 Government deficit and public debt as ratio to GDP (in percent)

(AMCs) earlier this decade. These AMCs have had a relatively low cash recovery rate on these NPLs (on average about 20 percent of the nominal loan amounts), indicating the poor quality of the underlying assets.

More importantly, the blunt tools available for controlling credit growth—including ceilings on total loan growth at each bank—and the strong signal that banks should reduce NPL accumulation may in fact have had perverse effects on the quality of lending. These factors may together have provided an incentive for banks to continue rolling over loans to SOEs, even unviable ones, in order to prevent those loans from appearing as NPLs. At the same time, profitable SOEs are not required to pay dividends to the state and, therefore, have had an incentive to plow their retained earnings back into new investments. Thus, price increases and rising profitability in some sectors, in addition to the availability of cheap bank financing, appear to have fueled the recent investment boom. The high level and sectoral concentration of investment may both presage the build up of excess capacity in some sectors, which in turn could lead to a resurgence of NPLs over the next 2–5 years (Goldstein and Lardy 2004).

Recognizing these potential problems and the need for a robust financial sector, the Chinese authorities have redoubled their efforts on banking sector reforms. Most of these reform efforts to date have focused on three of the SCBs—Bank of China, China Construction Bank, and Industrial and Commercial Bank of China. Foreign strategic investors have been invited to take equity stakes in these banks, in the hope that this will expedite improvements in corporate governance and lead to transfers of

managerial and risk-management expertize (see Hope and Hu 2005). Each foreign strategic investor is permitted to hold up to a 20 percent equity stake in a bank, with a cap of 25 percent on the total equity stakes of all such investors.

To make progress on bank restructuring and entice foreign strategic investors, large amounts of NPLs have been transferred from these banks to asset management companies. The PBC, through a holding company called Central Huijin Investment Company, has infused capital (using portions of the stock of foreign exchange reserves) into these banks. As a consequence, these banks already meet or are close to meeting the threshold capital adequacy ratio of 8 percent, with provisioning for loan losses. These banks have also been permitted to do IPOs abroad in order to strengthen their capital bases.

In sum, these three SCBs are in stronger shape than they were a couple of years ago. Whether these banks now constitute efficient financial intermediaries is a different matter, of course. On that score, progress has been limited. A similar picture is true of the banking system at large.

5.2 Monetary policy implementation

The primary instruments of monetary policy used by the PBC are open market operations, changes in the discount rate, and reserve requirements, aided and abetted by "window guidance" to banks on their lending operations.[9] Another instrument, which is used less as it is constrained by the fixed exchange rate regime, is interest rate policy (interest rates on PBC lending and reserves held at the PBC, as well as other rates including banks' base deposit and lending rates). The PBC has recently been using growth rates of money and bank lending as explicit intermediate targets. The relationship of these aggregates to real activity has not been stable over time. Furthermore, there has been a trend decline in velocity, with the growth rate of M2 consistently being a few percentage points higher than nominal GDP growth over the last few years, complicating things further. Yet, targets for money and credit growth have become an important device for the PBC to signal its monetary policy intentions.

Reserve requirements have recently been used quite extensively as a monetary policy instrument. The required reserve ratio (ratio to a bank's deposits) was reduced from 13 percent during the early 1990s to 8 percent in 1998 and to 6 percent in 2000, in part to allow banks to better manage

[9] See Yi (2001) and Xie (2004). A telling (and probably slightly unfair!) indication of the importance accorded to nonmarket based and nonprudential measures to control credit growth is a sub-heading in the chapter on Monetary Policy Conduct in the PBC's 2004Q2 Monetary Policy Report (p. 44). It reads: "Moral suasion intensified to guide credit structure optimization."

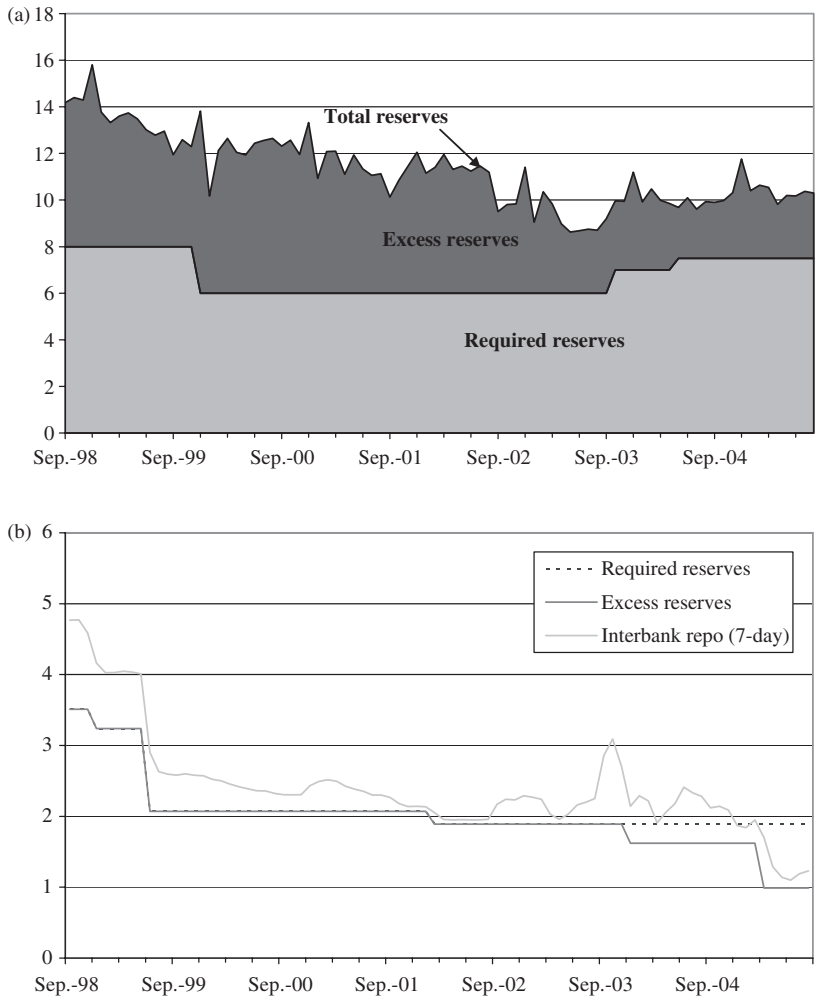
their funds (Figure 3a). This ratio was raised to 7 percent in September 2003 and further to 7.5 percent in April 2004, as part of a series of measures intended to control lending growth amidst concerns about rapid credit growth and potential overheating in the economy.

In addition to changes in reserve requirements, differentiated reserve requirements were introduced in April 2004. This affected second-tier banks, including the JSCBs that had accounted for a significant part of the surge in lending growth in 2003. Those banks in this category that did not meet certain standards in terms of the quality of their loan portfolios and capital adequacy were subject to a reserve requirement of 8 percent, half a percentage point higher than the standard required reserve ratio. Rural and urban credit cooperatives were exempt from this higher reserve requirement.

One complication in using the reserve requirement ratio as an active instrument of monetary policy is that the state banks, especially the SCBs, have tended to maintain substantial excess reserves at the PBC. A portion of these excess reserves is believed to be used for interbank settlement and liquidity management purposes, but it is difficult to discern how large the banks' perceived need for excess reserves for this purpose is. The PBC clearly has a concern that a significant amount of excess reserves makes the banks less sensitive to changes in its policy interest rates in the interbank market. This led the PBC to reduce the rate of remuneration on excess reserves to 1.62 percent in December 2003 and to 0.99 percent in April 2005 (compared to the unchanged rate of 1.89 percent on required reserves; Figure 3b).

In recent years, the amount of excess reserves maintained by banks (in percent of total bank deposits) has declined from 7.3 percent at end-2000 to 4 percent in March 2005 (Figure 3a). At that time, excess reserves were on the order of 4 percent for the SCBs, 5.3 percent for the JSCBs and 5.6 percent for the RCCs. This still suggests a high level of liquidity in the banking system. Notwithstanding the abundance of liquidity and the low interest rate on excess reserve holdings at the PBC, the lending behavior of the banks, especially the SCBs, has been held in check partly due to their objective of meeting mandated capital adequacy requirements (8 percent) by 2007.

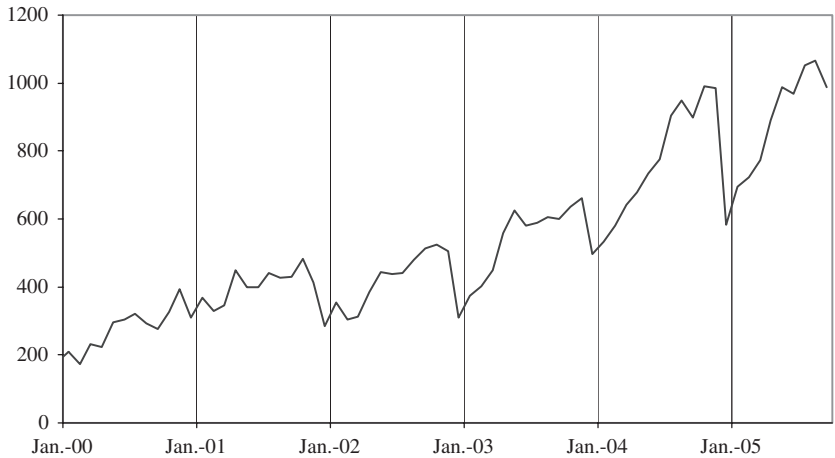
The PBC has been trying to build up the interbank market and improve its effectiveness as a channel for the transmission of monetary policy. However, the existence of substantial excess reserves undermines the predictability of monetary policy actions on reserve pressures. For instance, in October 2003, a modest change in reserve requirements caused the SCBs to build up excess reserves in anticipation of further increases in reserve requirements as the PBC sought to aggressively tighten credit. Consequently, there was a sharp spike in interbank rates as the



Source: CEIC and authors' calculations.
Notes: Bank reserves are expressed as percentage ratios of total deposits in the banking system. The rate of remuneration on required and excess reserves was the same until December 2003.

Figure 3 (a) Bank reserves (b) Rate of remuneration on reserves

smaller banks, especially the JSCBs, sought funding from the interbank market. In any event, the interbank market has expanded rapidly in recent years, with the transaction volume rising from about 1 trillion yuan in 1999 to 23 trillion yuan in 2005. In recent years, repurchases have accounted for around two-thirds of these transactions and interbank



Source: CEIC

Figure 4 Government deposits at the central bank (billions of RMB)

borrowing for about a quarter of the transactions (PBC's Monetary Policy Reports).

A further challenge faced by the PBC is that liquidity management is complicated by unpredictable seasonal fluctuations in government deposits maintained at the PBC (Figure 4). Tax and other revenue collections during the year typically lead to a buildup of deposits over the course of the year. These deposits are generally withdrawn towards the end of the year to finance various expenditure obligations (withdrawals for public investment, in particular, tend to be concentrated towards year-end). This introduces strong seasonal components—whose magnitudes can be unpredictable—in government deposits at the PBC.

In addition to its policy rate (rediscount rate) that affects the interbank market, base deposit and lending rates of the state banks have traditionally been set by the PBC with prior approval of the State Council. More recently, the PBC has been afforded some independence to change the floating bands around the base rates that provide some degree of flexibility to banks in setting deposit and lending rates. Interest rate liberalization has proceeded in steps over the last couple of years. On 1 January, 2004, the PBC increased the flexibility in the rate for loans to the private sector to 0.9–1.7 times the base rate for commercial banks and urban credit cooperatives and 0.9–2.0 times the base rate for rural credit cooperatives. Financial institutions were also given the freedom to determine lending rates for individual borrowers based on their risk

profiles and other characteristics, rather than being constrained by guidelines on pricing loans related solely to size and ownership structure of borrowers.

Interestingly, banks have taken only limited advantage of this added flexibility in lending to the private sector. A survey reported in Table 2 of the PBC's 2004Q3 Monetary Policy Report reveals that, in the first three quarters of 2004, about half of all loans were made at or below the base lending rate. The SCB, in particular, priced two-thirds of their loans at or below the base rate and did < 5 percent of their new lending at more than 1.3 times the base rate. A possible explanation is that the degree of flexibility in lending rates may simply not have been sufficient to compensate banks for lending to private sector firms, whose loans are inherently riskier than those made to SOEs.[10] The regional commercial banks and rural credit cooperatives, on the other hand, made better use of this flexibility, pricing 66 percent and 93 percent of their loans, respectively, above the base rate.

On 29 October, 2004, the ceiling on lending rates was scrapped altogether (except for urban and rural credit cooperatives).[11] The subsequent widening of the gap between the base lending rate and the actual lending rate (a weighted average based on loan volumes) indicates that banks are beginning to use this margin (Figure 5b). However, there is still little evidence that the SCBs, in particular, are using this flexibility to substantially redirect lending to the private sector at higher interest rates. A survey in the 2005Q4 Monetary Policy Report, like the one noted earlier for 2004, reveals a similar picture in terms of the pricing of loans by different types of banks. This could reflect concerns banks have about their risk-assessment capabilities, especially in an environment where there are still strong pressures to avoid accumulation of new NPLs. A less benign explanation is that banks are responding to an informal incentive structure that remains unchanged—loans made to state enterprises are still regarded as less risky in terms of reputational costs to bank managers and loan officers, while loans made to private sector enterprises that become nonperforming could entail charges of incompetence or corruption.[12]

---

[10] Bad loans to SOEs are more readily forgiven than bad loans to the private sector.

[11] See Dunaway and Prasad (2004) for an assessment of the potential benefits of this policy shift.

[12] Podpiera (2006) examines lending growth, credit pricing, and regional patterns in lending to look for evidence of changes in the behavior of SCBs following recent reforms and strengthening of their balance sheets. He concludes that the pricing of credit risk remains rather undifferentiated and that SCBs do not appear to take enterprise profitability into account when making lending decisions. For more on Chinese banking reforms, see Dobson and Kashyap (2006) and Garcia-Herrero, Gavilá and San tabarbára (2006).

Deficiencies in the legal framework may also play a role. Collateral provisions are generally difficult to enforce, so lending to the private sector carries additional risks.

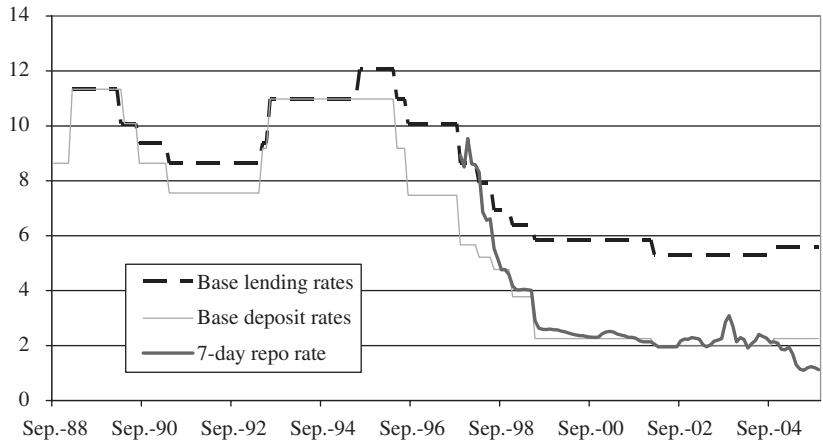
Along with the liberalization of lending rates, banks were given more freedom to make downward adjustments to deposit rates. The maintenance of a floor on lending rates and a ceiling on deposit rates appears intended to ensure that competition among banks does not drive down margins, which are seen as essential to maintain bank profitability and enable them to fortify their balance sheets by using profits to write off loan losses (Figure 5a).

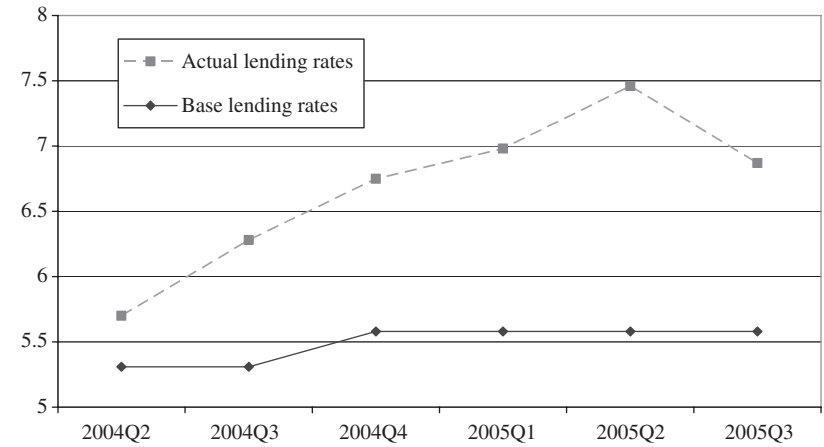
5.3 The exchange rate regime

From 1995 to 2005, the renminbi was maintained at a fixed parity relative to the USD. This regime was classified by the authorities as a managed float since the rate could in principle move by 0.3 percent around this parity. In practice, the rate remained essentially fixed at the central parity. The number of participants in the foreign exchange market, the China Foreign Exchange Trading System (CFETS), was limited to a handful including some of the SCBs. In fact, these banks acted as clearing agents for many of the trades that they settled directly without the transactions ever reaching the CFETS. This kept the foreign exchange market relatively thin and underdeveloped. But it also made the mechanics of managing the exchange rate easier for the PBC since it could quickly correct any deviations from the central parity.

On 21 July, 2005, the renminbi was revalued by 2.1 percent relative to the USD and the government announced that the external value of the renminbi would henceforth be set with reference to a basket of currencies, although neither the currency composition of the basket nor the basket weights have been publicly disclosed. The new regime also allows for fluctuations of up to 0.3 percent around the reference rate against the USD. In principle, this could mean that the exchange rate is allowed to drift up (or down) by 0.3 percent each day, which could amount to a significant appreciation (or depreciation) over a period of time. In practice, however, the renminbi has moved only by modest amounts against the dollar.

This rigidity of the exchange rate has constrained monetary policy independence by making it difficult for the PBC to use interest rates as an instrument to meet domestic policy objectives (see Eichengreen 2004 and Prasad, Rumbaugh and Wang 2005). The existence of capital controls, even though they may not be fully effective, implies some room for monetary policy independence. In practice, however, interest rate changes are tightly restricted by the financial repression and capital controls needed to keep banks solvent.

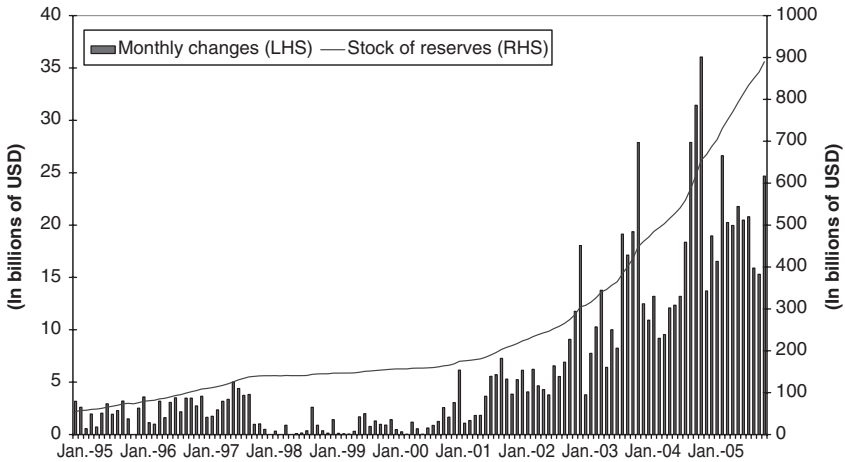




Source: CEIC, PBC monetary reports and IMF calculations

Figure 5 (a) Base lending and deposit rates (1-year) (in percent) (b) Base and actual lending rates (1-year) (in percent)

The complications created by a fixed exchange rate have been most evident in the rapid build-up of international reserves since 2001, when the renminbi began to come under appreciation pressures (Figure 6). The spike in the pace of reserve accumulation during 2001–04 is largely attributable to a surge in speculative capital inflows (through both official and unofficial channels; see Prasad and Wei 2005), although a rapid



Notes: The flow and stock numbers for foreign exchange reserves in this figure include the amounts used for bank recapitalizations: $45 billion in December 2003, $15 billion in April 2005, and $5 billion in September 2005; as well as a $6 billion fx swap that PBoC conducted with domestic banks in November 2005.
Source: CEIC and authors' calculations.

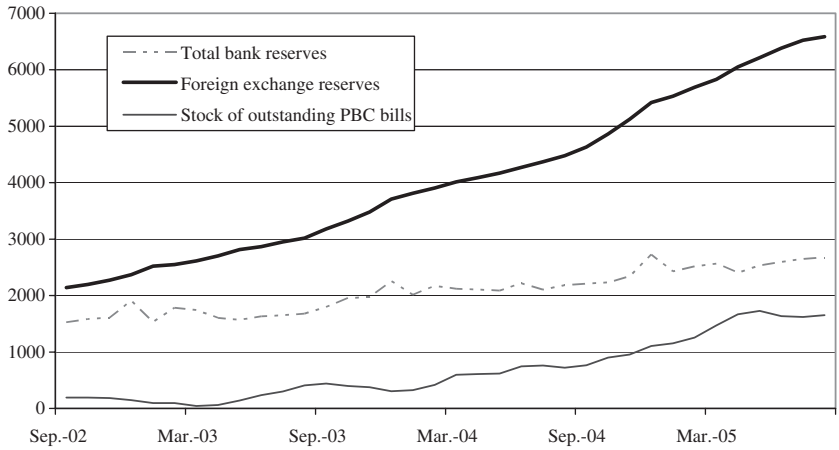
Figure 6 Foreign exchange reserves: flows and stocks

expansion in the trade surplus seems to have become more important during 2005.

Until 2002, government bonds had been used as the primary instrument for sterilization of foreign inflows. Some conversion to central bank bills (PBC bills) took place in late 2002, when the stock of government bonds available for repo transactions shrank to very low levels. The first full-fledged auction of new PBC bills took place in May 2003. PBC bills have now become the primary instrument for sterilization of capital inflows and, with the surge in inflows, the stock of outstanding PBC bills has increased rapidly (Figure 7).

The fraction of reserves sterilized by the central bank has varied over the last few years and it is not straightforward to assess exactly how much sterilization has taken place (see Green 2005a,b). By and large, the PBC seems to have had little trouble soaking up liquidity using PBC bills. While a few analysts have taken the low levels of sterilization as signaling, at least in some periods, potential problems in sterilization operations, this is far from obvious. The rate of credit growth has, after all, come down significantly relative to the very high growth rates observed in 2003–04. Furthermore, the interest rate on PBC bills remains quite low.

A confluence of forces has facilitated sterilization operations. Saving rates are very high; corporate saving, in particular, has increased sharply

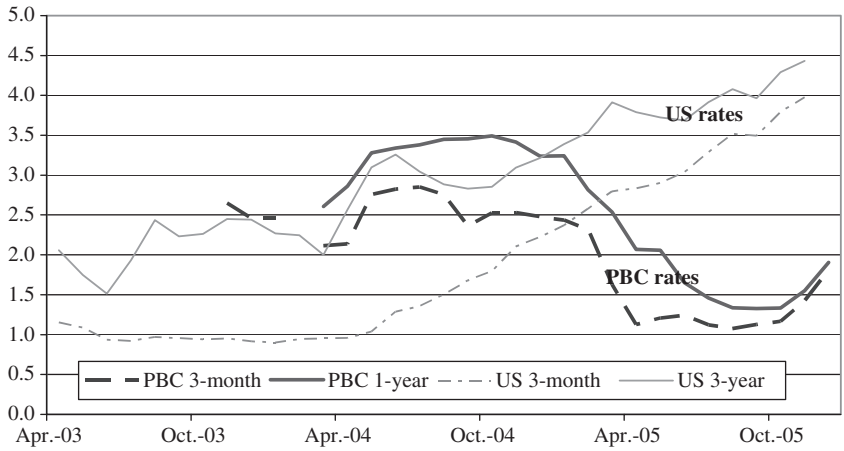


Source: PBC reports and CEIC

Figure 7 Stocks of reserves and central bank bills (in billions of RMB)

over the last year. Most of these savings flow into the banking system since there are few alternatives. This has made the banks flush with liquidity, particularly at a time when they are under pressure to hold down growth in credit and improve their balance sheets. In this context, banks have an incentive to hold PBC bills rather than increase their lending since corporate lending, for instance, carries a capital requirement of 100 percent while no capital needs to be put aside for lending to the government. Thus, there is a great deal of demand for PBC bills even at relatively low interest rates, well below the rates of return on comparable-maturity industrial country treasury bonds (Figure 8). This means that, at the margin, sterilization is essentially a money-making operation for the PBC.

But such a cost-benefit calculation can be deceptive. The lack of exchange rate flexibility not only reduces monetary policy independence, it also affects banking sector reforms. The inability of the PBC to use interest rates as a primary tool of monetary policy implies that credit growth is often controlled by much blunter and nonmarket-oriented tools, including nonprudential administrative measures. As argued by Prasad and Rajan (2005), this vitiates the process of banking reform by keeping banks' lending growth under the administrative guidance of the PBC rather than letting it be guided by market signals. This constraint has also perpetuated large efficiency costs via provision of cheap subsidized credit to inefficient state enterprises. The incidence of these and other costs of banking system inefficiency are not obvious, but they are probably



Source: CEIC and U.S. Treasury

Figure 8 PBC bill rates vs. US treasury yields

ultimately borne by depositors in the form of low real returns on their saving.

6 Independent monetary policy for China

China's declared intention to adopt a flexible exchange rate necessitates the choice of a new nominal anchor and a new strategy for monetary policy. Employing the principles of monetary policy discussed earlier in light of China's current financial institutions, we present a package of proposals to guide China's new independent monetary policy regime. We recommend a low inflation nominal anchor, operational independence for the PBC with formal strategic guidance from the government, reforms to make the Chinese banking system robust against interest rate fluctuations, and specific advice regarding the improvement of statistics, communications, and the institutional capacity of the PBC.

6.1 A low inflation nominal anchor

We believe that an explicit fixed low inflation objective would be an appropriate nominal anchor for China—it would help to firmly anchor inflation expectations and has many advantages over the alternatives, including the current *de facto* fixed exchange rate regime. This new nominal anchor could be supplemented, for the foreseeable future, with an operational role for money growth targets to help achieve the announced low inflation objective. Money growth targets would be of great help in

China, which is just beginning to modernize its banking system and to utilize indirect monetary policy instruments in lieu of direct credit controls to implement monetary policy. For the reasons discussed in section 2, however, we believe that a money growth target would not be a good *stand-alone* nominal anchor for Chinese monetary policy.

To set in motion the shift to an independent monetary policy framework, China should announce in the near future its intention to adopt an explicit low long-run inflation objective in order to lock in the current low inflation rate indefinitely. A qualitative commitment to low inflation might suffice until the details of the explicit inflation objective are worked out. The announcement should explain that monetary policy anchored by a long-run inflation objective would direct the PBC to stabilize employment and financial markets in the short run, subject to a commitment to keep inflation at or near the fixed long-run inflation objective on average over the medium term. Such a statement would be consistent with the broader mandate of the PBC and would enable it to carry out independent monetary policy flexibly.[13] In our view, it is premature and probably unnecessary for China to adopt formal, elaborate, inflation targeting procedures advocated by some economists and pursued by some central banks; although more formal inflation targeting procedures should not be ruled out for the future.

While the shift to independent policy disciplined by a low inflation anchor is in principle a major undertaking, the PBC would not need to make sharp changes in its operating procedures while supporting reforms for the new framework are put in place. Nevertheless, it is important for China to adopt a low inflation objective soon so that it is not without a nominal anchor during the transition to a flexible exchange rate regime, which is a stated medium-term objective of the authorities.

It will be important to specify the long-run inflation objective in more precise quantitative terms. As argued in section 2, there are sound analytical reasons for defining the inflation objective on the basis of a measure of core inflation.[14] Eventually, other choices and decisions about the nature of the inflation objective will need to be addressed—e.g., point target versus range; level of the target etc. In this article, our focus is more on the strategic aspects rather than technical details, so we leave these as

[13] The flexible strategy for monetary policy anchored by a long-run inflation objective that we suggest for China is closely related to that recommended for the United States by Bernanke (2004) and Goodfriend (2005). It is also related to the "inflation targeting lite" approach (Stone, 2003), although we believe that the subordination of the inflation target to other macroeconomic objectives considered there would hamper the effectiveness of monetary policy in anchoring inflation expectations.

[14] Food prices are quite volatile in China. Food accounts for about 40 percent of the consumption basket of Chinese households, giving it a large weight in the CPI. The retail price of energy is administered.

open questions for now. What we wish to emphasize here is the principle of transparency in monetary policy making, which would be embodied in an explicit numerical inflation objective. To quote Bernanke (2005, p. 2):

> Providing quantitative guidance about the meaning of "long-term price stability" could have several advantages, including further reducing public uncertainty about monetary policy and anchoring long-term expectations even more effectively.

6.2 Instrument independence for the PBC with strategic guidance from the government

China has already done much to modernize its banking and financial system. However, it must undertake additional reforms to support an independent monetary policy. The crucial requirement is that the PBC be granted instrument (operational) independence. Operational independence is necessary because the PBC must have the authority to move its policy instruments aggressively on short notice without permission from other government agencies.

In turn, there are two key prerequisites for effective instrument independence: the PBC must be given full control of aggregate bank reserves, and the Chinese banking system must be made financially robust against interest rate fluctuations. We recommend a minimal set of banking reforms below that could provide the requisite financial robustness in a few years. The modernization of the banking system will take much longer, but a fully modern banking system is not essential for monetary policy purposes.[15] In addition, we emphasize that, to make instrument independence fully effective, the PBC will need the discipline and accountability provided by formal strategic guidance from the government.

Full PBC control of bank reserves

China has already put in place some of the institutional arrangements necessary for the PBC to effectively manage aggregate bank reserves in the short-run. It has created a deep, liquid market in central bank bills through which the PBC can manage aggregate bank reserves effectively with open market operations. The Chinese have also created an active, liquid repo market that the PBC uses to manage the supply of reserves on

[15] The optimal speed of financial sector reforms in a second-best world with multiple distortions, and how it is tied in with other reforms, is a complicated issue (see Prasad and Rajan 2006). Indeed, instruments such as "window guidance" may continue to play a limited role during the transition to a more efficient banking system. However, such instruments must be utilized with care since they may not work as expected and may have perverse side-effects.

a day-to-day basis. The infrastructure for borrowing or lending reserves among banks in the interbank market on the basis of repos or on an uncollateralized basis at the China Interbank Offer Rate is well established. The level of interbank rates is determined flexibly to clear the market for borrowing and lending reserves, and the spread between the rates varies with such things as the nature of collateral backing the loan. The reserve market allows a given pressure on reserve positions to be distributed evenly across banks because each has the opportunity to obtain or release reserves into the market at common interbank interest rates.

Our positive assessment of certain aspects of the interbank market must, however, be balanced against a number of its unsatisfactory features: its relative thinness and illiquidity, the fact that major players may have excessive market power, and the fact that nonbank participants have the potential to destabilize the market. We believe that the Chinese financial authorities can and will remedy such defects before too long.

One defect deserves a little more attention. The behavior of excess reserve demand is said to complicate the transmission of monetary policy. Excess reserve demand appears to be volatile, and it has a tendency to offset the effect of changes in aggregate reserve supply on short-term interest rates, thus short-circuiting somewhat the PBC's power to influence bank lending. We have two recommendations for mitigating this problem. First, the PBC should refrain from discretionary reserve requirement adjustments because these induce volatility in excess reserve demand as banks try to anticipate and prepare for changes in reserve requirements by building up excess reserves in advance.

Second, the PBC ought to discontinue the payment of interest on excess reserves. Experience suggests that discontinuing interest on excess reserves in order to raise the opportunity cost would lower the elasticity of the excess reserve demand and greatly reduce its volatility. To the extent that banks economize on excess reserves as a result, the PBC could drain reserves with PBC bills. Those banks unable to economize on excess reserve demand would lose interest income. But the PBC could ease the loss of interest income by initially returning lost interest earnings and withdrawing the interest rate rebate from banks over time.

At present, the primary threat to the PBC's independent control of bank reserves arises from its responsibility to buy or sell foreign exchange in support of the tightly managed exchange rate.[16] It is not enough that

[16] The notions that the current exchange rate regime is sustainable indefinitely and that capital controls will continue to provide room for independent monetary policy are, in our view, either fallacious or ignore many of the attendant costs and risks. This issue is discussed at length in Appendix A of the working paper version of this article (Goodfriend and Prasad 2006).

China intends to introduce more flexibility into the exchange rate over time to facilitate adjustment in its external accounts. The government, as part of its program to grant the PBC operational independence for monetary policy, should relieve the PBC of the responsibility to support particular exchange rate objectives through its foreign exchange operations.[17] Only then can the PBC manage its balance sheet with full credibility to maximize the power of monetary policy to stabilize the Chinese macroeconomy.

Another consideration is that, over the long-run, the PBC must continually accumulate assets to provide for the trend growth of bank reserves and currency. In recent years, foreign exchange acquisitions in connection with China's exchange rate policy have provided more than enough longer-run growth of the PBC balance sheet; earlier, PBC lending to banks did so. Because the stock of foreign assets has grown so large, and because a large portion has been sterilized with PBC bills, even if foreign exchange inflows slow or reverse, the PBC should be able to provide for trend growth of its balance sheet for a while by allowing PBC bills to run off. However, once PBC bills have run off, the PBC must be prepared to acquire assets on a regular basis to provide for trend growth of currency.

One possibility would be for China to create a liquid government securities market to enable the PBC to expand its balance sheet by acquiring government securities rather than by accumulating foreign exchange or by lending to banks. Under this arrangement, the government would issue enough new securities every year for the PBC to purchase. The PBC would need to have the discretion to purchase just enough debt to allow the stock of currency to grow at a rate consistent with its inflation objective. This would facilitate the conduct of monetary policy by making PBC asset acquisition independent of foreign exchange policy and of bank supervision and regulation. The PBC would return to the government the interest paid on the securities it buys, net of its operating needs, and the government would obtain the revenue generated from the growing demand for currency at stable prices. For this arrangement to work well, the inflation objective would of course have to be explicitly backed by the government (including the fiscal authority).

Robust banking against interest rate fluctuations

China has taken a number of steps to modernize its banking system, and has created much of the institutional flexibility for the PBC to transmit

[17] This would not preclude the PBC from limited foreign exchange intervention at the margin. See Broaddus and Goodfriend (1996) for an analysis of the Federal Reserve's foreign exchange operations.

monetary policy actions effectively to aggregate demand—through a liquid bank reserves market, with flexible, competitively determined interbank interest rates managed by open market operations. To prepare China for independent monetary policy, it is now essential that Chinese banks be made financially robust to fluctuations in short-term interest rates. This is necessary both for banks to manage lending prudently and for the PBC to allow interbank interest rates to fluctuate as needed to manage independent monetary policy effectively.

The fundamental source of the financial robustness problem in China's banking system is two-fold: (i) China's banks have long been a primary means of financing SOEs, and (ii) many of China's banks are themselves run by local managers politically motivated (or under pressure of provincial government officials) to direct credit to SOEs. The high rate of saving and the lack of alternative investment opportunities in China provide the banking system with ample loanable funds to finance questionable loans to SOEs. With the government's tacit approval, moreover, banks have an incentive to carry loss-making SOE loans on their books indefinitely.[18]

The problem is that banks whose interest earnings are significantly impaired due to NPLs have cash flow sufficient only to pay relatively low interest on loanable funds acquired in the interbank and deposit markets. Higher interbank rates associated with more restrictive monetary policy would put weak banks under stress. Since banks are tightly connected through the payments system and the network of interbank balances, the financial distress would threaten the entire banking system. The distorted incentives for bank managers that lead to the accumulation of NPLs must be overcome if the Chinese banking system is to be made financially robust to flexible interest rates. In short, a financially fragile banking system has the potential to undermine central bank independence by, for instance, making the PBC reluctant to raise interest rates to head off inflationary pressures.

The financial vulnerability of the Chinese banking system to interest rate fluctuations is a difficult problem. Clearly, the authorities must complete the removal of NPLs from banks in order to fortify the banking system against flexible interest rate policy. China must also reform its banking system so that bank managers are free from political pressures to lend to underperforming SOEs and are instead motivated to make prudent loans to viable enterprises. Otherwise, the banking system is likely to be weakened again by a resurgence of NPLs.

[18] The ceiling on deposit rates and the floor on loan rates together have kept cash flows in the banking system positive, in spite of the large share of NPLs on bank balance sheets.

Chinese banks are essentially involved in a *fiscal policy* function—the financing of SOEs to support employment in the state sector until it can be absorbed by the growing private sector. Although that financing is provided in the form of "loans", it is not *banking policy*. The problem is that bank managers cannot be asked to lend prudently, with an expectation that loans be repaid and bank capital preserved, when managers are rewarded by the political system for directing fiscal transfers to state firms, and then largely excused for loan losses in the state sector.

We conclude that, in order to make the banking system robust against interest rate fluctuations going forward, the Chinese government must disentangle bank lending from the financing of nonviable SOEs. In Goodfriend and Prasad (2006, Appendix B), we propose that this could be accomplished by channeling financial support for nonviable SOEs through a separate government agency, an essential step that we believe could be completed in a few years.

An important related point is that, in order to protect against moral hazard in connection with the implicit insurance of Chinese bank deposits, supervisors and regulators must be empowered to ensure that bank capital remains above required minimums, and to intervene promptly to restrict the disposition of bank assets in the event that a bank's capital falls below required minimums. Furthermore, to make enforcement fully credible, Chinese bank regulators must have access to sufficient funding, e.g., a deposit insurance fund, to pay off depositors promptly if a bank is unable to abide by its regulations.

To sum up, the financing of SOEs through the banking system in China impedes the development of banking, fiscal, and monetary policies. Chinese banks cannot be governed according to good banking practice and regulated with the help of good banking policy, unless they are relieved of their responsibility for financing SOEs. The separation of fiscal policy support for SOEs from banking is the key to making Chinese banks financially robust against interest rate fluctuations. The robustness of Chinese banks, in turn, is necessary to provide a sufficient degree of separation of monetary policy from both banking and fiscal policies so that the PBC can conduct monetary operations independently and effectively.[19]

We believe that China could complete the reforms outlined earlier in a few years, in large part because Chinese financial authorities have been working hard to strengthen the banking system. The authorities clearly recognize the need for bank reform, China has the resources to deal with the NPL problem, and the authorities appreciate the urgency for doing what is necessary to support independent monetary policy for China.

[19] As long as inflation remains low, open market operations sufficient to implement monetary policy have relatively minor fiscal implications.

It will take much longer for Chinese banks to modernize fully, in particular to adopt methods for efficiently pricing loans according to risk. Nevertheless, Chinese monetary policy can be transmitted effectively through a banking system that may be far from the efficient banking frontier, as long as the banking system is financially robust against interest rate fluctuations and the exchange rate regime does not inhibit the PBC from employing the full range of interest rate variability to stabilize inflation and economic activity.

It would be very helpful at some point for China to deregulate bank deposit and lending rates further, although full deregulation of interest rates is not critical for monetary policy to achieve its primary macroeconomic objectives—the PBC can manage bank reserves to achieve whatever restraint on the growth of money and credit is needed.[20] In any case, full relaxation of interest rate restraints must be undertaken in conjunction with regulatory improvements to minimize moral hazard problems connected with deposit insurance.

Strategic guidance from the government

We emphasize that instrument independence must be granted in tandem with strategic guidance from the government. The operationally independent PBC should be instructed by the government to pursue the objectives for monetary policy enumerated in the PBC law, subject to a commitment to keep inflation at or near the fixed long-run inflation objective on average over the business cycle. Government support for operational independence is necessary to encourage the PBC to take potentially difficult monetary policy actions that may be needed on occasion. Explicit government direction must also serve as the basis upon which the PBC can be held accountable in some way for achieving its mandated objectives—perhaps through regular monetary policy oversight hearings. Without such strategic guidance from the government, the PBC would be deprived of the credibility essential to make independent monetary policy work well.

6.3 Statistics, communications, and institutional capacity

The transitional nature of the Chinese economy creates unique problems in the production and interpretation of statistical indicators, including those relevant for monetary policy decisions. Even with accurate data

[20] For instance, interest rate ceilings in the US did not prevent the Federal Reserve from achieving its macroeconomic objectives with monetary policy in the period before interest rates were deregulated fully by the Depository Institutions Deregulation and Monetary Control Act of 1980. Nevertheless, interest rate regulations inhibited the flexibility of the economy to respond to monetary policy actions.

in hand, there would be relatively little history on which to base forecasts of inflation. Moreover, productivity growth, which is difficult to forecast even in the US, is very difficult to predict in an emerging market economy like China where growth rates could vary over a much wider range. Given such complications, it would be useful for the PBC, in guiding monetary policy, to keep track of the growth of various money and credit aggregates against estimated growth rates believed to be consistent with low inflation.

At a minimum, the PBC must have timely access to accurate and comprehensive data on Chinese macroeconomic and financial conditions. It will be important for the government to make the gathering of accurate data (including at the provincial levels) a high priority for enabling effective macroeconomic management. We recognize that improving the consistency, reliability, and timeliness of macroeconomic data is no small challenge, but our view is that the PBC, with the help of other government statistical agencies, should be able to satisfy adequately its basic statistical needs within a few years. In addition, the PBC should use its regional branches to construct a nation-wide network to capture the latest anecdotal information on current economic and financial conditions in China.

The PBC must also acquire the analytical capacity to decide how to move its instruments flexibly in response to developments in the economy. That capability involves the acquisition of the relevant hardware as well as a staff of well-trained economists and statisticians. The PBC must be empowered to build up the institutional capacity necessary to support its monetary policy mission, and given the financial resources to do so.

In preparation for the day when it is called upon to manage monetary policy independently, the PBC should continue to improve and broaden its published assessments of economic and financial conditions in China, its monetary policy communications, and its judgments about future economic conditions.[21] Communications should motivate the steps that China is taking to modernize its banking and financial systems. The PBC and the China Banking Regulatory Commission (CRBC), which was spun off from the PBC in 2003, should continue advertising and explaining institutional reforms that have been undertaken in this regard.

We think it would be useful, as well, for the PBC and the CRBC to explain the reforms, in part, as necessary to adopt a framework for

[21] The PBC issues quarterly monetary policy reports. In addition, the PBC recently released its first Financial Stability Report (PBC 2005); this is expected to be an annual report.

independent monetary policy. The PBC, in particular, should explain, along the lines we've outlined in this article, the need to prepare China for independent monetary policy and its advantages for China over the long run. Talking about the need for reforms to facilitate the effectiveness of monetary policy would both motivate the reforms and help build credibility for the government's commitment to low inflation.

7 Concluding remarks

A flexible independent monetary policy oriented to domestic objectives is fast becoming indispensable for the effective management of the Chinese economy. We have attempted to provide both motivation and direction for China's transition to an independent monetary policy. Given the current underdeveloped state of the Chinese banking and financial systems, some may think our focus on this issue is premature. We think otherwise. Given China's intention to move to a flexible exchange rate, there are good reasons for China to begin to build the institutional foundations for the transition now. In particular, China must choose a new nominal anchor for monetary policy as it introduces flexibility into its nominal exchange rate.

There is a clear case for making a low long-run inflation objective that new nominal anchor, and little reason to delay its adoption. It will take many years to modernize China's financial system fully, but we have argued that China could put in place in the next few years a modest package of reforms that would serve as an effective foundation for independent monetary policy anchored by a low inflation objective.

The key is to grant the PBC operational monetary policy independence, which requires that the PBC be given full control of bank reserves and that the Chinese banking system be made robust against interest rate fluctuations. To satisfy the former requirement, the government must allow a substantial degree of flexibility in the foreign exchange rate, so that exchange rate adjustments, and not PBC purchases and sales of foreign assets, can clear the foreign exchange market. Strengthening of bank balance sheets, including by removal of NPLs, is necessary to satisfy the latter requirement. Bank lending must also be disentangled from the financing of nonviable SOEs, which we proposed could be accomplished by channeling financial support for nonviable SOEs through a separate government agency.

In addition, we emphasized that PBC operational independence must be granted with formal strategic guidance from the government. Without such strategic guidance, monetary policy would lack credibility and the PBC would be deprived of the support needed to take potentially difficult

monetary policy actions. Finally, we underscored the need for the PBC to be given financial resources and encouragement by the government to build up the institutional capacity necessary to support its monetary policy mission.

References

Allen, F., J. Qian and M. Qian (2005), "Law, Finance and Economic Growth in China", *Journal of Financial Economics* **77(1)**, 57–116.

Barnett, S. (2004), "Banking Sector Developments," in E. Prasad, ed., *China's Growth and Integration into the World Economy: Prospects and Challenges*, IMF Occasional Paper No. 232, International Monetary Fund, Washington.

Bernanke, B. (2004), "Inflation Targeting", *Federal Reserve Bank of St. Louis Review* **July/August**, 165–68.

Bernanke, B. (2005), "Testimony of Ben S. Bernanke", Nomination Hearing Before the Committee on Banking, Housing and Urban Affairs, US Senate, November 15, http://www.federalreserve.gov/boarddocs/testimony/2005/.

Bernanke, B., T. Laubach and F. Mishkin et al. (1999), *Inflation Targeting: Lessons from International Experience*, Princeton University Press, Princeton, NJ.

Bernanke, B. and M. Woodford, eds. (2005), *The Inflation-Targeting Debate,* University of Chicago Press, Chicago, IL.

Blanchard, O. (2005), "Discussion of Inflation Targeting in Transition Countries: Experience and Prospects", in B. Bernanke and M. Woodford, eds., *The Inflation-Targeting Debate*, University of Chicago Press, Chicago, IL.

Broaddus, J.A. and M. Goodfriend (1996), "Foreign Exchange Operations and the Federal Reserve", *Economic Quarterly*, Federal Reserve Bank of Richmond, Winter, 1–20.

Broaddus, J.A. and M. Goodfriend (2001), "What Assets Should the Federal Reserve Buy?", *Economic Quarterly*, Federal Reserve Bank of Richmond, Winter, 7–22.

Broaddus, J.A. and M. Goodfriend (2004), "Sustaining Price Stability", *Economic Quarterly*, Federal Reserve Bank of Richmond, Summer, 3–20.

Cole, H. and L. Ohanian (2001), "New Deal Policies and the Persistence of the Great Depression: A General Equilibrium Analysis," Federal

Reserve Bank of Minneapolis, Research Department Working Paper, No. 597, May.

Dobson, W. and A. Kashyap (2007), "The Contradiction in China's Gradualist Banking Reform," *Brookings Panel on Economic Activity*, forthcoming.

Dunaway, S. and E. Prasad (2004), "Interest Rate Liberalization in China," Op-ed article in *International Herald Tribune*, December 3.

Eichengreen, B., (2004), "Chinese Currency Controversies." CEPR Discussion Paper No. 4375, Center for Economic Policy Research, London, UK.

García-Herrero, A., S. Gavilá and D. Santabárbara (2006), "China's banking Reform: An Assessment of its Evolution and Possible Impact", *CESifo Economic Studies* **52**, 304–63.

Goldstein, M. and N.R. Lardy (2004), "What Kind of Landing for the Chinese Economy?" Policy Briefs in International Economics, No. PB04-7, Institute for International Economics, Washington.

Goodfriend, M. (2002), "Monetary Policy in the New Neoclassical Synthesis: A Primer", *International Finance*, pp. 165–92, *Reprinted in Economic Quarterly*, Federal Reserve Bank of Richmond, Summer 2004, pp. 21–45.

Goodfriend, M. (1993), "Interest Rate Policy and the Inflation Scare Problem: 1979–1992," *Economic Quarterly*, Federal Reserve Bank of Richmond, Winter, 1–23.

Goodfriend, M. (2004), "Understanding the Transmission of Monetary Policy," Manuscript, Federal Reserve Bank of Richmond. Presented at Joint China-IMF High Level Seminar on China's Monetary Policy Transmission Mechanism, Beijing, May 2004.

Goodfriend, M. (2005), "Inflation Targeting in the United States?", in B. Bernanke and M. Woodford, eds., *The Inflation-Targeting Debate*, University of Chicago Press, Chicago, IL.

Goodfriend, M. and R. King (1997), "The New Neoclassical Synthesis and the Role of Monetary Policy", in B. Bernanke and J. Rotemberg, eds., *NBER Macroeconomics Annual*, MIT Press, Cambridge, MA, 231–95.

Goodfriend, M. and R. King (2005), "The Incredible Volcker Disinflation", *Journal of Monetary Economics* **52, July**, 981–1016.

Goodfriend, M. and E. Prasad (2006), "A Framework for Independent Monetary Policy in China," IMF Working Paper, WP/06/111, International Monetary Fund, Washington.

Green, S. (2005a), "Making Monetary Policy Work in China: A Report from the Money Market Front Line," Manuscript, Standard Chartered Bank, Shanghai, China.

Green, S. (2005b), "On the Ground—China" various research reports of Standard Chartered Bank Global Research, Shanghai, China.

Ho, C. and R. McCauley (2003), "Living with Flexible Exchange Rates: Issues and Recent Experiences in Inflation Targeting Emerging Market Economies," BIS Working Paper No. 130.

Hope, N. and F. Hu (2005), "Reforming Chinese Banking: How Much Can Foreign Entry Help?" Manuscript, Stanford University, Stanford, CA.

International Monetary Fund (2004), (2005), *People's Republic of China: Article IV Consultation—Staff Report*, IMF Staff Country Reports 04/351 and 05/411. Available on the web at www.imf.org.

Jonas, J. and F.S. Mishkin (2005), "Inflation Targeting in Transition Countries: Experience and Prospects", in B. Bernanke and M. Woodford, eds., *The Inflation-Targeting Debate*, University of Chicago Press, Chicago, IL.

McKinnon, R.I. (2006), "China's Exchange Rate Trap: Japan Redux?" *American Economic Review*, **96**(2), 427–431.

Mishkin, F.S. (2000), "Inflation Targeting in Emerging-Market Countries", *The American Economic Review, Papers and Proceedings* **90(2)**, 105–09.

Mishkin, F.S. and K. Schmidt-Hebbel (2005), "Does Inflation Targeting Make a Difference," Manuscript, Columbia University.

Obstfeld, M. and K. Rogoff (1995), "The Mirage of Fixed Exchange Rates", *Journal of Economic Perspectives* **9, Fall**, 73–96.

People's Bank of China, *China Monetary Policy Report*, various issues Beijing, China.

People's Bank of China (2005), *China Financial Stability Report 2005*, Beijing, China.

Podpiera, R. (2006), "Progress in China's Banking Sector Reform: Has Bank Behavior Changed?" IMF Working Paper, 06/71.

Poole, W. (1978), *Money and the Economy: A Monetarist View*, Addison-Wesley Publishing Co., Reading.

Prasad, E., T. Rumbaugh and Q. Wang (2005), "Putting the Cart before the Horse? Capital Account Liberalization and Exchange Rate

Flexibility in China,'' IMF Policy Discussion Paper 05/1, International Monetary Fund, Washington.

Prasad, E. and S.-J. Wei, (2005), ''China's Approach to Capital Inflows: Patterns and Possible Explanations,'' IMF Working Paper 05/79. Forthcoming in an NBER Volume on Capital Flows edited by Sebastian Edwards.

Prasad, E. and R. Rajan (2005), ''China's Financial Sector Challenge,'' Op-ed article in *Financial Times*, May 3.

Prasad, E. and R. Rajan (2006), ''Modernizing China's Growth Paradigm,'' *American Economic Review*, **96**(2), 331–336.

Setser, B. (2006), ''The Chinese Conundrum: External Financial Strength, Domestic Financial Weakness'', *CESifo Economic Studies* **52**, 364–95.

Stone, M. (2003), ''Inflation Targeting Lite,'' IMF Working Paper 03/12, International Monetary Fund, Washington.

Svensson, L.E.O. (2003), ''Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others'', *Journal of Economic Perspectives* **17**, **Fall**, 145–66.

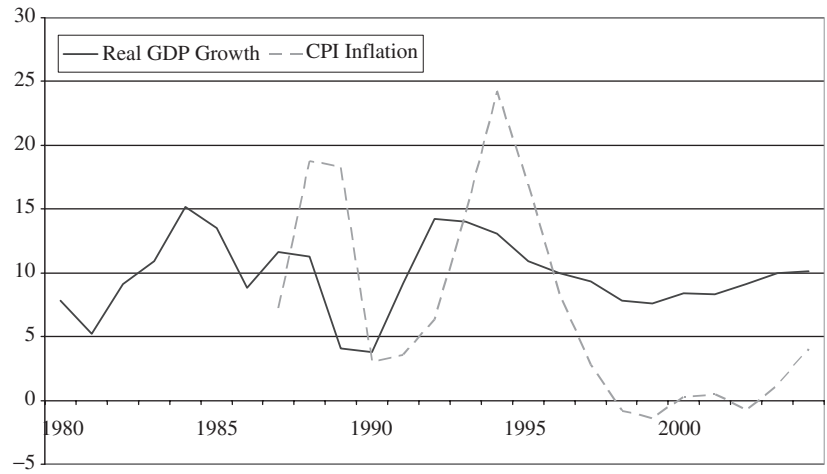
Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton, NJ.

Xie, P. (2004), ''China's Monetary Policy: 1998–2002,'' Stanford Center for International Development Working Paper No. 217, Stanford University, Stanford, CA.

Xie, P. and X. Luo (2001), ''Taylor Rule and China's Monetary Policy,'' Manuscript, People's Bank of China, Beijing, China.

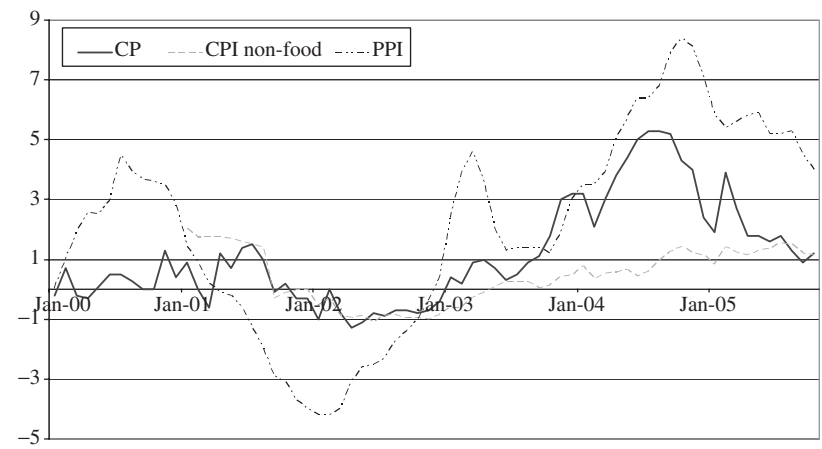
Yi, G. (2001), ''The Framework of China's Monetary Policy,'' Manuscript, People's Bank of China Beijing, China. Presented at the PBC-IMF International Seminar on Monetary Policy Operations, May 2001, Suzhou, China.

Appendix



Source: CEIC and IMF's international financial statistics.

Figure A1 GDP growth and inflation rates



Source: CEIC and IMF's international financial statistics

Figure A2 Alternative measures of inflation (monthly data, year on year changes in prices)

CESifo Economic Studies, Vol. 53, 1/2007, 42–68, doi:10.1093/cesifo/ifm002
Advance Access publication 17 February 2007

Perspectives from the Happiness Literature and the Role of New Instruments for Policy Analysis

Bernard M. S. van Praag*

Abstract

After having been ignored for a long time by economists, *happiness* is becoming an object of serious research in 21st century economics. In Section 2 we sketch the present status of happiness economics. In Section 3 we consider the practical applicability of happiness economics, retaining the assumption of ordinal individual utilities. In Section 4 we introduce a cardinal utility concept, which seems to us the natural consequence of the happiness economics methodology. In Section 5 we sketch how this approach can lead to a normative approach to policy problems that is admissible from a positivist point of view. Section 6 concludes. (JEL codes: B21, B41, D63, I31, I38)

Keywords: Happiness economics, subjective well-being, equivalence scales, economic policy.

1 Introduction

During the last years we see a small revolution in economic science. I have in mind the rise of ''happiness economics'' as a serious branch of economics. Until quite recently the concept of ''happiness'' as a research subject for economists was anathema. The aim of this article is a consideration of the relevance of this new development in economic science for economic science as such and for economics as a tool for policy making.

The fundamental aversion among twentieth-century economists towards the concepts of happiness, well-being,[1] etc. may be traced back to the influence of Behaviourism in general and the influential tract of Robbins (1932) in particular.[2]

Although Robbins recognized that psychology plays a role in the explanation of economic behaviour, he excludes the idea that feelings, and especially feelings of happiness, could be measured in an operational way. As we assume that economic behaviour, that is making choices between alternatives, is motivated by maximization of utility, satisfaction, well-being or happiness,[1] this stand leaves the economist in an

* Bernard M.S. van Praag is affiliated to University of Amsterdam, e-mail: B.M.S.vanPraag@uva.nl

[1] We will make no difference between those notions. The reason therefore is that all those notions are metaphysical concepts. Without an operational measurement method they remain empirically indistinguishable.

[2] I will refer to the second printing.

awkward position. He observes only the results of a choice process, but he does not observe the underlying components of the choice process. In terms of the traditional consumer model, he observes only the preferred commodity bundle, but not the indifference curves behind it. Happiness economics makes an attempt to observe and estimate the indifference curves on a space of alternatives. Such a space may involve various combinations of income, family size, health, etc. The same method could theoretically be applied to assess the shape of indifference curves on the traditional commodity space. As far as we are aware of, this has not been done yet.

Since at least from Pareto (1909) economists differentiate between ordinal utility and cardinal utility. Mainstream economists mostly do not talk of happiness but of utility. As we said before, the choice of the word is just a matter of taste without consequences. We may also talk of ordinal and cardinal happiness. Ordinal utility/happiness is just a labelling system for indifference curves, where a higher label corresponds to a higher utility level. If three indifference curves in a space of alternatives are labelled 1,2,3, respectively, it implies that the individual is indifferent between alternatives with the same label, that is on the same indifference curve, while 1-alternatives are considered as worse than 2-alternatives and 3-alternatives as better than 2-alternatives. The curves may be relabelled as 2, $2\frac{1}{2}$ and 5 without any consequences. In the case of a cardinal utility/happiness interpretation the re-labelling has consequences. In the first labelling system the utility differences between the three curves were equal, while in the second labelling system the utility difference between the first two is slight, while the difference with the third is considerable. Let us assume a population of three persons, one on each curve. In the first system average utility/happiness would be 2 and in the second labelling system $3\frac{1}{6}$. It is obvious that for a meaningful evaluation of average or national happiness or for the inequality in the happiness distribution the additional assumption of cardinality is essential.

In the early economic literature there was a definite cardinal flavour (e.g. Marshall, Pigou). Economics as part of the social sciences was assumed to play a role in the solution of social problems, e.g. the alleviation of poverty and inequality, the promotion of employment and economic growth. Just taxation implied a comparison between individuals of losses in happiness caused by the imposition of taxes. Even if classical economists were unable to assess happiness and differences in cardinal happiness empirically, there was some perspective that in the future this might become possible. Since Robbins' (1932) influential tract, the assumption of cardinal utility fell gradually into discredit. It got the seemingly final blow by Samuelson (1947) and Houthakker (1950,1961) with their ''revealed preferences'' approach and the ensuing erosion of the utility concept.

Actually, this development of economic science over the last 60 years is a mixed blessing. By the erosion of the utility concept and the denial of a cardinal character to the utility concept economic science lost much of its relevance for practical socio-economic policy, which deals for a large part with distribution problems requiring comparisons of utility or happiness between individuals. Prudent economists do not dare to call one distribution not worse than an alternative distribution, except if that distribution is not worse than the alternative for all individuals involved. For example, no one will have qualms when stating that an income distribution between two persons $y^{(1)} = (3000, 2000)$ is better than $y^{(2)} = (2000, 1000)$. The first situation is better in the sense of Pareto-optimality. However, at the same time this is a rather trivial statement. In most cases where distributions or social situations have to be compared we do not meet this simplistic situation. It is only possible to say that some people are better off in one situation than in the other and that the inverse statement holds for other people involved. Hence, we cannot make a statement about the ranking of the two situations in terms of social desirability.

If economists are declaring themselves unable to make normative evaluations and comparisons, at the same time they declare themselves unable to offer help for most political decision situations, which as a matter of fact are and can only be based on such comparisons.

It seems that theoretical mainstream economics has chosen for the more prudent, let us say, *scientific* position, sticking to Pareto-optimality as the ultimate tool, in this way reducing (perhaps rightly) the ambition of economics to be a policy instrument.

However, it is not exaggerated to notice here a certain schizophrenia, hidden in many economists. On the one hand in theoretical work they detest to make normative statements, while in their applied work they do it as a matter of routine. But even in many "pure" fields of economics we find hidden normative judgements. For instance, the literature on the construction of income inequality measures on an axiomatic basis is full of hidden normative statements. For instance, the Dalton axiom that says that income inequality is reduced by transferring money from a rich person to a poor one is based on a normative vision (however, sympathetic) on interpersonal utility comparison and the validity of Gossen's First Law, never scientifically shown to be true.

The question we would like to tackle here is whether the new "happiness economics approach" offers new perspectives for economics as a science and for its application in economic and social policy. Our reaction will be in the affirmative.

The structure of this article is as follows. In Section 2 we sketch the present status of happiness economics and we link it with the earlier

approach in the seventies of the last century known as the Leyden School (LS). In Section 3 we consider the practical applicability of happiness economics, retaining the assumption of ordinal individual utilities. In Section 4 we introduce a cardinal utility concept, which seems to us the natural consequence of the happiness economics methodology. In Section 5 we sketch how this approach can lead to a normative approach to policy problems that is admissible from a positivist point of view. In Section 6 we draw some conclusions.

2 What is happiness economics?

Let us start to consider what happiness economics is. The underlying principle is simple and intuitively plausible. If you like to know how happy an individual is with "something", a straightforward method to get an idea about it is to ask an individual about his happiness with that "something". As happiness is a rather emotionally laden word, in practice we ask how "satisfied" an individual is.

The point of departure are so-called "satisfaction questions" like

*How satisfied are you with **your income**?*
*How satisfied are you with **your job**?*
*How satisfied are you with **your health**?*
*How satisfied are you with **your life as a whole**?*

Those questions have been posed in psychological surveys for quite some time.

The discrete response is either in terms of verbal categories like "horrible", "bad",...,"not bad", "delighted" or in numerical categories like 1,2,...,5 or 0,1,2,...,10.

It is now generally accepted that such questions make sense. More precisely, most respondents are willing and are able to answer such questions, and respondents in similar circumstances give similar responses. As shown before, such questions may refer to many different things, like in this example "income", "job", "health" or "life- as a whole". It may be aspects of life, or as psychologists say "life domains".

Up to this moment the response is nearly always discrete, although conceptually a continuous response scale is possible as well. In the latter case we may think of a continuous line interval from A (worst) to B(best), where respondents put a cross at the point that corresponds with their feeling on a (worst,best) scale.

Modern satisfaction researchers or "happiness economists" are not only interested in observing the degree of satisfaction with different "domains of life" and in being able to state that in the EU average satisfaction with

income is a 7 on a 10-scale and in Russia a 5. They are interested in the question *why* one person is satisfied to such and such a degree with a specific life aspect or life "domain". Their primordial research question may be summarized as: What are the determinants of satisfaction?

The basic methodology to assess the influence of different factors x is to assume a latent utility variable u for which holds $u = f(x; \beta, \varepsilon)$, where β is a parameter vector and ε a random disturbance term. It is assumed that all respondents that give the same response enjoy the same utility/happiness value u. That is, they are situated on the same indifference curve in an ordinal context. The most simple model is then to use a linear description and to assume for respondent n that $u_n = \beta_1 x_{1,n} + \beta_2 x_{2,n} + \cdots + \beta_0 + \varepsilon_n$. Such a model with discrete response is traditionally estimated by Ordered Probit or Ordered Logit. The equation $\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_0 = constant$ describes an indifference surface and the ratios β_i/β_j are the trade-off coefficients between the factors x_i and x_j. It is mostly possible to assess the shadow price of a variable x. Let us assume x_2 stands for income and x_1 is changed into $x_1 + \Delta x_1$. If that increase is happiness increasing and if more income is also preferred to less income, then the change into $x_1 + \Delta x_1$ has to be compensated by an income decrease Δx_2 such that the individual stays on the same indifference curve. More exactly, we require $\beta_1(\Delta x_1) + \beta_2(\Delta x_1) = 0$, yielding a shadow value $(\beta_1/\beta_2)(\Delta x_1)$.

The beginning of happiness economics is frequently set in the beginning of the nineties (see Frey and Stutzer 2002). That is not entirely true. The work of the so-called LS in the seventies of the previous century, named after the Dutch university to which the research group was affiliated, can be seen as part of happiness economics as well, although in the beginning of the seventies the word "happiness" was avoided by the LS in favour of the less presumptuous term "individual welfare of income". It is interesting to line out what are the differences and similarities between the LS-approach and the contemporary happiness approach (CHS). This is not only interesting from a scientific historical perspective. It might be that LS has things to offer that fill in a lacuna in contemporary happiness economics. Moreover, there seems to be some confusion among modern happiness economists about the question in how far the old Leyden School was on the same track. Therefore, we shall firstly devote some lines on the LS-approach.

In 1968 I published my book on Individual Welfare functions and Consumer Behavior. One of its main theses was that utility or welfare might be an operational and measurable concept. In Van Praag (1971), challenged by some empirical opportunities, I formulated the so-called Income Evaluation Question (IEQ). It was a question module by which something should be measured, that I called *welfare derived from income*. The idea was that only by *questioning* the individual himself we might be

able to get information about his feelings on welfare (happiness, etc.). This idea of asking people about their *feelings* and accepting this as valuable information is a basic and common point of departure for LS and CHS. In 1971 most economists did not accept this simple truth, denying themselves by this refusal very valuable information.

The IEQ, has been posed in the years since then to many thousands of respondents. It runs as follows:

The Income Evaluation Question (IEQ) (mid-interval version)

Whether you feel an income is good or not so good depends on your personal life circumstances and expectations.

In your case you would call your net household income:
a very low income if it would equal € ——
a low income if it would equal € ——
a still insufficient income if it would equal € ——
a just sufficient income if it would equal € ——
a good income if it would equal € ——
a very good income if it would equal € ——

The differences between this IEQ and modern satisfaction questions can be summarized as follows:

(i) The IEQ deals with income, that is with financial satisfaction only. Satisfaction questions may deal with financial satisfaction, job satisfaction, life as a whole, etc. Hence, we have to compare the IEQ with the *financial* satisfaction (FINSAT)-question.

(ii) The FINSAT-question asks for an evaluation of current income y_c. This is so to say the *stimulus* to the respondent. The respondent's reaction is an evaluation on a numerical or verbal evaluation scale. The IEQ provides *six* stimuli $U_1, \ldots, U_6$ in terms of evaluations on a verbal evaluation scale. The six reactions of the respondent are denoted by $c_1, \ldots, c_6$. There are two possibilities to utilize the results provided by the IEQ: an *ordinal* and a *cardinal* one.

(iii) In the ordinal representation we compare answers $c_{i,n}$ by different respondents n with different individual characteristics x_n. As the respondents with different characteristics are observed to need different income levels $c_{i,n}$ to reach the same FINSAT-level i, we may look for a relationship $c_i n = \beta_1 x_{1,n} + \beta_2 x_{2,n} + \cdots + \beta_i + \varepsilon_n$. For fixed i it describes an indifference curve for the level i. (see, e.g. Van Praag and Van der Sar 1988).

(iv) In the cardinal interpretation the LS translated the verbal evaluations $U_1, \ldots, U_6$ into numbers $\frac{1}{12}, \frac{3}{12}, \ldots, \frac{11}{12}$. That this is a reasonable way of translating words into numbers in this context, is empirically confirmed in Van Praag (1991). Then a log-normal distribution function $U_n(c) = \Lambda(c; \mu_n, \sigma_n)$ was fitted through the points $\{(c_{in}, (2i - 1/12))\}_{i=1}^{6}$. The individual parameters μ_n, σ_n are estimated[3] by the log-mean and the log-standard-deviation of the responses $c_{1n}, \ldots, c_{6n}$. It follows that by means of the IEQ an individual utility function (or in the LS-terminology an individual welfare function) $U_n(c)$ can be estimated for each separate individual. As μ_n, σ_n differ over individuals those utility functions are different.

(v) The individual parameter μ_n (like the separate answers $c_{1n}, \ldots, c_{6n}$) can be well explained by individual characteristics x_n, including current income y_c. As the coefficient of income is significantly positive, it implies that someone's utility function shifts with increasing income. More precisely, financial needs increase with a rise of current income. This shift of norms with respect to income with the individual situation was called in van Praag (1971) *preference drift*. It is the same adaptation phenomenon that the psychologists Brickman and Campbell (1971) discovered in the same year in another context and empirical setting. They called it the *hedonic treadmill*. The own situation serves as an anchor point; if the own situation changes norms about satisfaction will shift accordingly.

(vi) As it is empirically found that $U_n(c) = \Lambda(c; \mu_n(y_c), \sigma_n)$, it is easy and attractive to estimate how individual n evaluates his current income. There holds $U_n(y_c) = \Lambda(y_c; \mu_n(y_c, x_n), \sigma_n) \stackrel{def}{=} \tilde{U}(y_c; x_n)$. The last term is the true evaluation by an individual of his own income. Hence, already in the LS (see, Van Praag 1971) there was made a distinction between an *ex ante* welfare function, which describes how a specific individual evaluates all income levels and an *ex post* welfare function, which describes how all members of the population evaluate their own income. Here there is some analogy with the distinction between *decision utility* and experienced *utility*, made by Kahneman, Wakker and Sarin (1997).

(vii) The link between LS-results and CHS-results lies in the fact that the Leyden *ex post* welfare function is identical to the financial satisfaction function, that is, it has the same indifference curves. It implies that for financial satisfaction the same trade-offs are found as when using the financial satisfaction question.

[3] In the early LS-publications a slightly different method was used to estimate μ and σ; later on the log-mean and log-S.D. appeared to give almost identical results.

Then it is no wonder that many of the LS-results can be found back in the MHS-literature, e.g. for the effect of reference groups, external effects, family size, etc. We refer to Van Praag and Ferrer-i-Carbonell (2004) for an extensive review. Earlier reviews may be found in Van Praag (1976, 1993, 1999).

(viii) Although in most of the Leyden literature the parameter μ is derived in a cardinal context, the parameter itself makes also sense in an ordinal context. Most Leyden results do not need that cardinal interpretation and keep their validity in an ordinal environment (cf. Van Praag, Van der Sar, 1988). The cardinality assumption is only essential for the analysis of inequality and redistribution problems.

(ix) In sum, the LS-literature is based on a somewhat more demanding question module. It deals only with financial satisfaction, while CHS deals with various domains and global welfare/happiness. In the domain of financial satisfaction it gives the same results as CHS, while the very useful distinction between *ex ante* and *ex post* evaluations is added.

Although the LS provided remarkable results and got some sympathizers all over the world, the truth is that it was never accepted by mainstream economics, because mainstream economics in the seventies maintained that such measurements were impossible in the first place. Sociologists and psychologists mostly did not know of the results, because those results were not submitted to their journals and in those days there reigned a terrible provincialism between the behavioural sciences, where sociologists and psychologists strictly rejected the idea that economists could have to offer something meaningful on the subject of ''feelings''.

Somewhat later, also independently, Easterlin (1974) as a rare exception in those times studied happiness and discovered the famous Easterlin paradox, saying that differences in individual happiness between citizens within a country may be explained for a good deal by differences in material circumstances, while he found average reported happiness to be about the same in some strongly different countries, poor and rich all over the world.

3 The practical applicability of happiness economics, retaining the assumption of ordinal individual utility

In the previous pages we noticed already that satisfaction responses may be (partly) explained by objective factors x. Now we shall pursue this line.

The resulting outcomes of the satisfaction questions in surveys may be summarized by satisfaction distributions, as shown in Table 1, which we derived from the German Socio-Economic Panel (GSOEP). For instance, 10.93% evaluate their job by "5", 18% by "7" and about 28% by "8". For other domains of life we find other evaluations.

If such job evaluations are meaningful responses, we have to assume that the 18% respondents who evaluate their job situation by a 7 have about the same feeling of satisfaction with respect to their job. It cannot be that a "7" means for one respondent "very good" and for another "very bad". This is a not unreasonable assumption for respondents who have about the same cultural and linguistic background. However, let us realize that this assumption cannot be *proven* to be *true*, for, therefore, we would need some kind of psychometer, which tells us by psycho-physical means how satisfied one is. Hence, the "common background"-assumption is a *primitive* assumption. However, it is an assumption, which is tacitly accepted in nearly all empirical surveys. This assumption is also called the assumption of interpersonal comparability of satisfaction. Clearly, this assumption is not strange at all except for the most extreme economic curmudgeons.[4] Most humans will be in no doubt when they see somebody laughing or smiling; they will interpret it as a person in a happy mood. Alternatively, when we see a weeping or frowning person we will interpret it as that the person is unhappy. We interpret such signs of moods similarly and we are also able to identify those feelings of the other in terms of own feelings. The same holds for the interpretation of wordings in a questionnaire.

If we assume interpersonal ordinal comparability to be a valid assumption, our data set can be split up into categories $0, 1, \ldots, 10$ of respondents who belong to the same satisfaction class. For the time being we do not assign yet a cardinal significance to those satisfaction levels, that is, we do *not* assume that a respondent belonging to class 8 feels *twice* as happy as one belonging to class 4. However, we observe that two respondents, who are in objectively different circumstances, may be equally happy, for they evaluate their situation by the same response. They are situated on the same indifference curve.

Now we shall consider how such a data set may be used for policy making. We focus as an example on the effect of family size on financial satisfaction. Notice, however, that this is just intended as an example of a general method. Subsequently, we will point to some other applications. Our example will deal with the satisfaction with income, or financial

[4] The definition is according to the dictionary: An ill-tempered person full of resentment and stubborn notions.

Table 1 Satisfaction distributions, Workers West Germany, 1996 (in percentages)

	0	1	2	3	4	5	6	7	8	9	10
Job Sat.	0.80	0.55	1.73	3.04	4.32	10.93	9.96	17.97	27.89	13.81	9.0
Financ.	0.31	0.33	0.77	1.89	3.46	9.35	10.62	22.19	30.10	13.46	7.53
Health	0.68	0.54	2.03	3.59	4.51	12.13	10.11	17.49	26.06	13.25	9.62
Housing	0.93	0.50	1.51	2.49	3.50	7.89	7.68	14.85	25.72	17.87	17.06
Leisure	1.02	1.33	3.57	5.49	6.32	13.68	11.51	16.86	21.40	10.01	8.81
Envir.	0.81	0.71	1.78	4.73	6.64	16.95	14.56	22.15	20.29	7.68	3.69
General	0.23	0.25	0.71	1.35	2.76	9.72	11.07	23.98	33.51	11.63	4.77

satisfaction FINSAT.This point is illustrated by Figure 1. Let us assume that we find from our (financial) satisfaction question that family[5] A with a monthly income $y = €.2000$ and no children ($fs = 0$) is equally satisfied with its financial situation as family B with an income $y = €.3000$ and two children ($fs = 2$). Or in formula $(2000,0) \sim (3000,2)$, where the sign "$\sim$" has to be read as "equally satisfied". This observation leads straight ahead to a *political recipe* for the construction of a family equivalence scale.

Assume that the government looks for an answer to the question how much family allowance we should give to a family C with income $y = €.2000$ and two children ($fs = 2$), in order that it will be at the same satisfaction level as family B. It stands to reason that the financial situation of family C is evaluated less than that of the family A with €.2000, since family C has two additional children to support. In order that A and C feel *equally* satisfied the family C should get an income increase of €1000 in order to arrive at the same curve as A. In fact, the first two families A and B are situated on the same "satisfaction indifference"- curve, while family C is situated on an indifference curve corresponding to a lower satisfaction level. A few of those indifference curves are sketched in Figure 1.

We see that an increase in family size fs has to be compensated by an increase in income y in order to keep the family at the same level of financial satisfaction as before the family size increase took place. This is what economists call the compensating income. If this compensation is not given, the family shifts to a lower satisfaction level, which corresponds in Figure 1 to a *higher* curve.

The family compensation to generate equality is €1000, If we compensate by less, say €500, then the family will be *under*-compensated. Would we like to compensate by more, say €1500, the family would be over-compensated.

The amount of €1000 is what Kapteyn and Van Praag (1976) called a *welfare-neutral* family compensation. What we do not know, if we stick to the ordinal interpretation, is: how serious is the welfare difference caused by the €500 under-compensation? Is the emotional difference slight or even negligible, or is the under-compensation really painful?

Obviously, the government does not have to follow the rule of welfare-neutrality in practice. It may be that policy makers willingly prefer to over-compensate low-income families and to under-compensate high-income families for political reasons. Then the political instrument of the family assistance schedule is an instrument to change the distribution of income

[5] We do not dwell here on the difference between the satisfaction of an *individual* and that of a *household*/*family*.

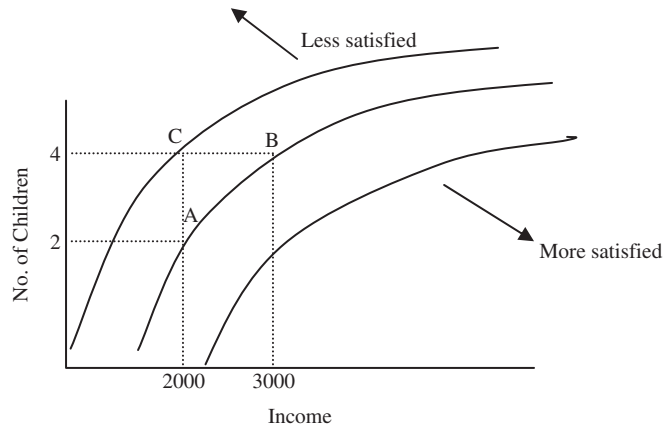


Figure 1 FINSAT-Indifference curves in (*fs*, income)-space

and indirectly that of happiness. But, the welfare neutral scheme is at least a *benchmark for policy* makers to find out in how far the enacted schedule satisfies welfare neutrality and to see which income classes are over-compensated or under-compensated.

As we suggested already, the crux of the matter is whether those indifference curves are estimable. The answer is yes. We may derive such curves from the answers on satisfaction questions, where the basic assumption is that individuals, who state that they belong to the same satisfaction response category, *by definition* are situated at the same indifference curve. The technical methodology is at this place beyond the scope. Let it suffice to say that the methodology has now been facilitated so much, also by means of standard software now available, that any economist or other social scientist with a slight knowledge of statistics and computer programming can perform such estimations. Actually, there are different variants of the method, e.g. Probit or Logit regression. As Ferrer-i-Carbonell and Frijters (2004) have shown, the many experiences thus far yield remarkably stable and robust results, which only marginally depend on the specific method applied. In Van Praag and Ferrer-i-Carbonell (2004, 2006) it has been shown that simple OLS-variants may be constructed that are as good as Probit or Logit, but computationally much easier and faster to implement than the traditional methods. A moment of thought explains much of this robustness. All these specifications amount to different specifications of the labelling system of the underlying indifference curves, but the indifference curves themselves are unchanged and it is these indifference curves which are estimated, either by Ordered Probit, Logit or what else.

The family equivalence scale derived from data from the GSOEP is tabulated in Table 2. We see that a couple with one child would need

Table 2 Family Equivalence Scales, derived from financial satisfaction questions

Household size	Correction factor
1	81%
2	100%
3	113%
4	123%
5	132%
6	140%
7	146%
8	152%

(from VanPraag and Ferrer-i-Carbonell 2004, Table 2.3)

13% more of net income in order to be at the same financial satisfaction level as the couple without child.

The real significance of this method becomes clear when we compare it with the traditional methods to define a family equivalence scale (see Van Praag and Warnaar 1997). The most traditional method in economic literature is to assess the cost of children by means of budget analysis or by means of normative budgets, defined by nutritional experts. The latter approach is clearly paternalistic, as it depends on the opinion of a commission of experts. The former approach derived from demand analysis is based on assumptions, which are more or less arbitrary. For instance, that it would be possible to distinguish specific commodities and services as "adult goods" (Deaton and Muellbauer 1986). Then two families of different size are assumed to be at the same welfare level if they have the same consumption of adult goods. There are two basic problems with this method. First, what are adult goods: alcoholics, cigarettes, a haircut? The second point is that a couple with a child may have a completely different consumption pattern as a couple without adults without being less or more satisfied. It is probable that most parents will drink less than couples without child, but would that imply a lower level of satisfaction? The basic problem, however, is that equivalence can only be defined if we know or can observe when individuals are equally satisfied. The only way in which we can find this out is by asking people how satisfied they are and by assuming that individuals who use the same expression in order to express their level of satisfaction are equally satisfied. The adult good approach or any other approach that is not based on this simple observation must be seen as arbitrary (see for a more detailed analysis, Van Praag and Warnaar 1997).

It is obvious that the satisfaction approach is also less exact than it looks like and has many limitations we need to be aware of.

First, satisfaction may vary on a continuous scale. If we observe it discretely in terms of a few response categories, a loss of accuracy is incurred. Moreover, satisfaction methods assume a model as well, which may be more or less realistic. For instance, the equivalence scale tabulated in Table 2 has been based on a model equation

$$u_i = \alpha \ln(fs) + \beta \ln(y) + \gamma$$

where i stands for a specific response category, that is, level of financial satisfaction. This is actually the equation describing the satisfaction indifference curve of level i.

In the log-case an increase by $\Delta \ln(fs) = \ln(fs) - \ln(fs_0)$ has to be compensated by an income increase

$$[\ln(y) - \ln(y_0)] = \frac{\alpha}{\beta}[\ln(fs) - \ln(fs_0)]$$

This specification is the basis for Table 2. We see that the compensation depends on the initial family size fs_0, where we have a "cheaper by the dozen"-effect (due to the lgogartihmic specification of family size), and on the initial income level y_0, which implies that richer people get more compensation in absolute amounts than poorer people.

Who says that this is the correct equation?

One of the problems with this equation is that the resulting income-compensation schedule depends on the initial income level. It is a fixed percentage of income. Let us now consider the alternative equation where logarithms have been replaced by absolute values:

$$u_i = \alpha \cdot fs + \beta \cdot y + \gamma$$

For the linear case the income compensation is given by

$$y - y_0 = \frac{\alpha}{\beta}[fs - fs_0]$$

In this case there is no "cheaper by the dozen"-effect; each additional child is compensated for by the same money amount.

The model equation

$$u_i = \alpha \cdot \ln(fs) + \beta \cdot y + \gamma$$

would yield a compensation

$$y - y_0 = \frac{\alpha}{\beta}[\ln(fs) - \ln(fs_0)]$$

Here poor and rich people would get the same compensation, but the "cheaper by the dozen"-effect is maintained.

It follows that the choice of the model equation is rather crucial with respect to the resulting compensation schedule.

A second point is whether the indifference curves are parallel or that they are specific for the level i of satisfaction reached. Or in formula,

$$u_i = \alpha_i \ln(fs) + \beta_i \ln(y) + \gamma_i$$

In that case the compensation schedule changes with the level of satisfaction. This might entail that rich people with a high level of satisfaction might have a smaller trade-off ratio α_i/β_i than poor people. Indeed, there is evidence found for this case (see Van Praag and Ferrer-i-Carbonell 2004).

A third complication is caused by the question whether financial satisfaction is *only* depending on the variables family size *fs* and net income *y* or that there are other variables relevant as well, e.g. region, age, price level.

Certainly there are omitted but relevant variables. On top of that there is a random disturbance term, standing for all unobservable variables like individual psychological traits. We can use panel data models to control for individual psychological traits that typically do not change across time, for example optimism and intelligence. Nevertheless, we cannot control for those traits that change across time. For example, individual's capacity to adapt by changing expectations.

It follows that this satisfaction approach is also not without difficulties. However, these specification and choice of variables difficulties are shared by both the traditional and the subjective satisfaction approach. Eventually, it depends on subjective choices, made by the researcher and sometimes by politicians, who like to get politically feasible results.

But there is an essential difference between the traditional methods and this subjective satisfaction approach. The traditional approach begins with a naïve artificial criterion to answer whether families are at the same satisfaction level or not. In the subjective approach this is not necessary, as satisfaction equality can be empirically observed. If two respondents are equally satisfied, they will evaluate their satisfaction with the same value, either numerically or verbally. This opens the way to find *empirically* which equation and combination of variables fits the data, i.e. the responses on satisfaction questions, best.

It is evident that the same methodology may be used to assess the effects of other variables on FINSAT.

Up to now we have exclusively considered the relation between financial satisfaction FINSAT and the size of the family. However, it is possible to apply the same methodology to the satisfaction with other domains of life, like job satisfaction, health satisfaction or satisfaction with life as a whole.

At first researchers were interested in determinants such as income, unemployment (Clark and Oswald 1994), etc. Later on, the focus expanded to the effects of reference groups (Ferrer-i-Carbonell 2005; Luttmer 2005), inflation (Boes, Lipp and Winkelmann 2007), and to policy oriented issues such as health economics (Ferrer-i-Carbonell and Van Praag 2002), environmental issues like airport noise (Van Praag and Baarsma 2005).Then the range of topics widens more and more to include crime and terrorism (Frey, Luechinger and Stutzer 2004).

Clark, Oswald and Warr (1994) looked at job satisfaction and how it varies with age. They found that job satisfaction is U-shaped in age. Many other studies using other data sets seem to confirm this. Job satisfaction, as many domain satisfactions and life satisfaction, seems to reach a minimum around an individual middle age after which point satisfaction increases together with age. The only exception is health satisfaction, which shows a negative relationship with age for any age (see Ferrer-i-Carbonell and Van Praag 2004).

Di Tella et al. (2001) and Blanchflower and Oswald (2004) looked at the effect of unemployment on life satisfaction.

Frey and Stutzer (2000) looked at the impact of more or less democracy on life satisfaction, based on differences in regional government in the Swiss cantons. They found that individuals living in more democratic cantons were happier under ceteris paribus conditions.

Ferrer-i-Carbonell and Van Praag (2002) give an example where the monetary counter-values of losses in health, caused by various chronic diseases, are assessed.

Van Praag and Baarsma (2005) evaluate the damage by airplanes, caused by various noise levels on inhabitants in the neighbourhood of Amsterdam Airport. They found that airplane noise damage could be compensated by giving the inhabitant about 2% of income as a compensation subsidy.

Senik (2004), Ferrer-i-Carbonell (2005) and Luttmer (2005) investigate the effect of the neighbours' income on individual satisfaction. The reference effect was investigated by the LS as well with similar empirical outcomes. It was then called *reference drift* (see, Kapteyn 1977; Van Praag, Kapteyn and Van Herwaarden 1979; Kapteyn and Van Herwaarden 1980).

To these examples we might add many others from the now flourishing happiness literature. This method has been applied in many papers.

One striking paper is that by Clark and Oswald (2002). They calculate the impact of different life events upon human well-being. According to that paper "getting married, for instance, is calculated to bring each year the same amount of happiness, on average, as having an extra £70 000 of income per annum. The psychological costs of losing a job greatly exceed

those from the pure drop in income. Health is hugely important to happiness. Widowhood brings a degree of unhappiness that would take, on average, an extra £170 000 per annum to offset.''

Actually, the production of new papers and new results is so large during the last years that it is simply impossible to do justice to all authors and their valuable contributions. One of the recent surveys (Clark, Frijters and Shields 2006) states

> Studying the causes and correlates of human happiness has become one of the hot topics in economics over the last decade, with both the size and depth of the literature increasing at an exponential rate (Kahneman and Krueger, 2006). To illustrate, a search of ECONLIT for journal articles with either ''Happiness'', ''Life Satisfaction'' or ''Well-being'' in the title, identifies 465 published articles between 1960 and 2006. Of these 363 (78%) have been published since 1995, 285 (61%) have been published since 2000 and one-third of the literature (37%, or 173 articles) has appeared in print in just the last three years. Focusing only on the period 2000–2006, this measure of economists' interest in the ''economics'' or ''science'' of happiness places that interest at roughly half the level of interest in ''Wages'' and just below that in ''Discrimination''. However, happiness is quickly catching up: it is in its ''industrial revolution'' stage . . ..

Consequently, we refer to the surveys by Di Tella and MacCulloch, (2006), Clark, Frijters and Shields (2006), Frey and Stutzer (2002), Senik (2005), Layard (2005), Dolan (2006), Van Praag and Ferrer-i-Carbonell (2004) for more information.

Our conclusion is that this methodology offers a viable way to assess the monetary costs and the compensation amounts needed to neutralize specific effects of changes in satisfaction determinants plus other interesting insights such as the extent of poverty, and the effect of relative income and economic growth. We notice in passing that most of the studies cited do not take into account the mechanisms of preference and reference drift/Hedonic treadmill and comparison income. As both mechanisms combined nearly neutralize income changes, the long-term effects of monetary compensations become questionable. However, we observe that for most problems there do not exist any other viable methods.

This makes the new methodology extremely attractive for the study of policy problems.

Finally, there are domains for which income is not a significant determinant of satisfaction. An example is health satisfaction, where the income coefficient β was found not to differ from zero. In such a case the trade-off ratios in terms of additional income cannot be calculated.

More or less money has no influence on satisfaction. In such situations the model remains politically relevant, if compensation can be given in other determinants, i.e. less working hours, or more safety in the streets.

4 The step to cardinality

Up to this point we have stuck to the ordinality assumption. That is, we assume that an evaluation of an "8" compared to a "7" indicates that the individual evaluates the first situation as better than the second. However, we do not know *how much* better. As already said, economists have always been reluctant to make the step from ordinality to cardinality. This was caused by the strictly behaviouralist approach, where one looks only at revealed preference behaviour. That point of view admits only the following statement. If alternative A is chosen above alternative B, it implies that satisfaction, anticipated from A is more than that of B, but it does not yield information about the value of the difference. Is alternative A "slightly preferred to" or "enormously better" than B (see also, Van Praag and Frijters 1999)?

For policy making this is a really nasty situation. First, if we like to compare the average happiness of populations (see, e.g. Easterlin 1974); Blanchflower and Oswald 2004) this is strictly impossible if we adhere to the ordinal point of view. Let us demonstrate this with the following simple example. Let us assume we have two individuals. One is stating his happiness is 6 and the other evaluates his happiness by 7. The average is 6.5. If we give only ordinal significance to these answers, we may, according to the definition of ordinality, apply any monotonic transformation to the happiness values. Let us take the simple transformation of adding 1 to all values, so that the average becomes 7.5. It shows that taking average happiness as an index of the happiness of a population (in this case just two persons) becomes meaningless. Or put otherwise, all papers and authors, to begin with the famous and important paper by Easterlin (1974), who present such averages, are implicitly embracing *a* cardinal interpretation, even if they do not mention this explicitly. I do not blame them for doing that. It just demonstrates how natural this interpretation is and how unnatural it is to stick to an ordinal interpretation only. It may strike some as strange that happiness would be bounded from below and/or above, as response scales like "$1, \ldots, 10$" or "very bad, $\ldots$, excellent" seem to suggest. However, in reality we have never met a respondent who, when faced with the question to evaluate his happiness on a (0,10)-scale, would refuse to answer because his happiness level, being a "12" was not included in the scale presented. Each respondent accepts and understands a finite scale,

where the lower bound stands for "Complete Misery" and the upper bound for "Perfect Bliss".

Similarly, it is interesting to consider the inequality of the happiness distribution. Let us continue this simple two-person example. We define inequality simply by the standard deviation of happiness. In this case the standard deviation is 0.5. It is easily seen that it does not change if we add one unit to both evaluations. Therefore, let us apply another simple transformation to the happiness responses. Let us now square the happiness indicator yielding 36 and 49 as happiness values. The average is 42.5 and the standard deviation becomes 6.5. It is obvious: aggregates, averages and inequality measures make no sense if we do not assume cardinality to begin with.

Also happiness data do not give much support for the analysis of redistribution problems, if we stick to the ordinal interpretation. Let us assume an ordinal utility function $U(y,fs)$, which increases in income y. Two individuals A and B have utilities $U_A = U(y_A,fs_A)$ and $U_B = U(y_B,fs_B)$, respectively. Now one thing is sure, the ordinal interpretation ensures that we may redistribute the sum income such that both individuals get the same utility by solving the equation $U_A = U(y_A - \Delta y,fs_A) = U(y_B + \Delta y,fs_B) = U_B$ under the provision that redistributed incomes are positive. The individuals A and B are then on the same indifference curve. However, apart from this utility equalization we cannot make any statement, if we do not admit for the cardinal interpretation. If we would redistribute by Δy but fall short of equalization, we cannot say whether the utility loss of A $U(y_A,fs_A) - U(y_A - \Delta y,fs_A)$ is greater or smaller than the utility gain $U(y_B + \Delta y,fs_B) - U(y_B,fs_B)$ of B. The net aggregate effect of any redistribution cannot be assessed.

The obvious solution is to accept the cardinal significance of subjective self-evaluations of happiness or satisfaction as meaningful. Or more directly, if a respondent evaluates his happiness by a "6" or an "8", we have to accept that answer as his or her degree of happiness. It implies that we accept a transition from 6 to 7 or from 7 to 8 to stand for equal improvements of happiness.

Is this such a strange assumption? In my eyes not. In the first place it is a question of semantics, and common culture. Language is a common good, created in an evolutionary way in order to have a means of communication between people. That is, words represent (roughly) the same meaning to all in the language community. We all believe that happiness is bounded from below by a state of absolute despair (not necessarily to be equalized with death) and from above by a state of perfect happiness or "bliss". It follows that it is natural that evaluations will be on a bounded scale like school grades, e.g. from 0 to 10 or their

verbal equivalents "very bad" to "excellent", used in many schools as well. And in the same way the most efficient way to use the gradations in between, is to assume that each subsequent grade stands for an equal jump on the ladder between 0 and 10.

What is the difference between this cardinal satisfaction concept and the shunned concept of cardinal utility in consumer theory? The difference is that from the observation of purchasing or more generally choice behaviour we cannot derive cardinal information (revealed preferences), while satisfaction questions provide that information, however inaccurately (stated preferences). There is no mystery about how it is done either. It is just a repeatable question module yielding a straightforward answer.

This is just the same method as is used in physics for some centuries, when measurement units and methods are developed to measure electrical currents, light intensity, temperature and so on. It is evident that there is some arbitrariness involved in how to translate the phenomenon in terms of a specific scale, but after that choice has been made the measurement unit gets a life of its own. By its usage it gains significance and after some time the results get a common language meaning, just as the strength of an electric current is described in Volts for the technician, or the dioptre for an individual wearing glasses. For instance Wikipedia defines:

> A **dioptre** is a unit of measurement of the optical power of a lens or curved mirror, which is equal to the reciprocal of the focal length measured in metres (i.e. 1/metres). For example, a 3 dioptre lens brings parallel rays of light to focus at 1/3 metre.
> Quantifying a lens in terms of its optical power rather than its focal length is useful because when relatively thin lenses are placed close together their powers approximately add. Thus a thin 2 dioptre lens placed close to a thin 0.5 dioptre lens yields almost the same focal length as a 2.5 dioptre lens would have. . . .

We see here exactly the same procedure: the unit of measurement is defined by some observation. The usefulness of this particular choice is then explained because the observation, so measured, is handsome to use in important calculations where an (approximately) additive law holds.

Transferring the argument to the measurement of happiness we see that there is nothing arbitrary in the measurement of happiness by survey questions. It is a well-defined empirical operation. If we accept it, the way is open to defining averages, inequalities and evaluating redistributions, in short to the use of happiness or satisfaction calculus for normative social policy. The only thing we have to accept is that our measurements are less exact and more volatile over time than physical measurements. Hence, it is easier and more reliable to make statements

on the average happiness of a population of individuals, where the measurement errors will compensate each other, than to make statements about a single individual.

5 A glance at the political usage of cardinal satisfaction

As shown before, if we accept the assumption of interpersonal ordinal comparability, there are a lot of relevant socio-economic applications in the field of equivalence scales. However, if we also accept the assumption of cardinality, there opens up a new world. Happiness of individuals is defined by a measurable and repeatable operation, where the measurement does not depend in any way on the observer. It is an objective measurement. That is, if the response categories are $1,2,3,\ldots,10$ it is assumed that all respondents scale their happiness between 0 and 10 and that all respondents evaluate the adjacent jumps between categories as equal. Clearly, this assumption cannot be proven to be true in an empirical way. This could only be validated in an objective way, if we had an alternative way of measuring individual "happiness". However, "happiness" as such is a metaphysical concept. It becomes a physical concept only by defining it by means of an empirical measurement method. The outcome of such measurements should not depend on the person who measures and it should yield (roughly) the same outcomes in repeated measurements. Last but not least the measurement outcomes should fit our predictions more or less. The first two requirements are obviously met. Whether the third requirement is met is a matter of taste. If most people feel it is met, it does.

There is one point of confusion that is frequently raised. How is it possible that one individual evaluates a specific situation as "extremely satisfactory", while another individual evaluates the *same* situation as "extremely unsatisfactory"? Is this not tantamount to saying that the happiness-meter is extremely unreliable and unstable? The answer to this apparent paradox is that each individual evaluates situations in his own way, given his personal situation, his past experiences, his social environment, etc. We would only have a problem if two very *similar* persons would evaluate the same situations differently. Fortunately, this is not the case. It is also therefore, that a systematic explanation of satisfaction answers is possible, yielding statistically significant effects.

If we assume that a cardinal interpretation is meaningful, we may think of new applications.

The first result is that we can produce a *social happiness atlas* of a society. See the society as a landscape of persons. The co-ordinates are social characteristics like age, education, region of the country, civil status,

income, political party, employment status, etc. For each point on the map we get an altitude co-ordinate, standing for "happiness" or satisfaction with life. We may also think of several co-ordinates standing for several domain satisfactions. We refer to Van Praag, Frijters and Ferrer-i-Carbonell (2003), where it is shown that happiness with life as a whole may be seen as an aggregate of domain satisfactions. It follows that the policy maker gets an idea of how happiness is distributed over society, and how domain satisfactions are distributed as well. As he may estimate the impact of various policies on individual satisfactions, this gives also the key to formulating policies that may enhance the happiness of groups of individuals or of the society as a whole. As measures generally will favour some citizens and worsen the position of other citizens, it becomes also possible to get some idea of a Kaldor compensation in terms of increments and decrements of happiness. We refer especially to Kapteyn and Van Herwaarden (1980), who studied the effect of changing tax schedules on the distribution of financial satisfaction, taking into account reference and adaptation effects. In short, this is a fascinating piece of information for any policy maker.

Aggregation of individual happiness would also lead to an index for Gross National Happiness (see also, Kahneman 2004). In essence this is the Benthamite Social Welfare Function (SWF).

Then the government's objective might be to maximize this social welfare function. However, there is a problem here: the weighting problem. Should everybody be weighted equally or should for example the poor be over-weighted in the SWF? It is well-known that the SWF may be defined more generally by the use of weights as $\mathrm{SWF} = \sum_n w_n U_n$ or still more generally as $\mathrm{SWF} = W(U_1, \ldots, U_N)$. Each individual or each social group gets its own weight. We see that knowledge of the individual happiness levels does not give automatically a happiness index for *policy* makers(See also, Frey and Stutzer (2005)). Here is a *political value* judgement required about the functional specification of the SWF. Needless to say that political parties as a rule will embrace different SWF-specifications.

However, there are several less ambitious objectives where the new apparatus can play a significant role.

The first field is poverty analysis. If we define everybody as poor if his $U < \alpha$, where α is a cutoff-point, either chosen by policy makers or by means of a survey by the population concerned, then we can estimate how many individuals or households are poor in society and, what is very important, we can identify the poor according to their social co-ordinates. The value of α may be taken to equal 4 ("bad") on a 10-scale or less severely 5. This is the so-called subjective poverty approach. We refer to the first contribution in this line by Goedhart et al. (1977) and for a multi-dimensional domain satisfaction approach to Van Praag and

Ferrer-i-Carbonell (2007). Hence, we get a poverty map of society and this is helpful for policy, because policy measures can become better targeted. First with respect to social coordinates, that is, which specific groups need help and can be effectively supported? Second, because we can differentiate help with respect to life-*domains*. Some are "health-poor", because their health satisfaction is low, and need more health care or health cost subsidy. Others are "job-poor" and need better working conditions. Again, others are feeling financially poor and need more income support.

The second field is the impact assessment of political measures on various social segments. If we know the determinants of satisfaction and happiness and the government changes the values of those determinants, then we are also able to assess the impact of such changes in terms of individualized happiness gains and losses and changes in (subjective) poverty gaps.

Up to now the reader may have got the idea that the measurement of happiness and satisfaction is only relevant for policy makers. However, this is by no means true. The third field deals with more commercial applications. We think on the cardinal measurement of self-rated satisfaction with health. Changes in health can be assessed in terms of "health gains" and "health gains per dollar" (Van Praag and Ferrer-i-Carbonell 2004; Dolan and Kahneman 2006). This is clearly not only relevant for the shaping of health policy, but also for pharmaceutical firms, health insurance companies, and all other kinds of health agents and authorities. In modern countries the value of therapies and medicines has to be assessed in order to get them in the insured package. Therefore a cost-benefit analysis is needed, that takes account of the costs and labour productivity effects. However, an important dimension of the benefits is the so-called intangible benefits. These are frequently neglected or estimated in a somewhat arbitrary way. For instance, by assigning a value of $100 000 to each healthy life year. How much better feels the patient and perhaps his family really? Here the subjective measurement is coming in the picture. This cost-benefit assessment (including the intangibles) may also be applied to subgroups of the population, differentiated according to age, health risk profiles, or specific diseases.

Similarly, job satisfaction evaluations may be used as a tool for evaluating personnel policy. Another field of application might be establishing the money value of health injuries and the corresponding damage amounts to be paid.

It lies at hand to apply the methodology in marketing analysis. We may measure in the same way the satisfaction with a car, a house or peanut butter. This may give indications to marketeers how to direct their selling efforts and/or to model the product materially or in image.

Then it is only one step to define descriptions of real products and fictitious ones by means of vignettes to be evaluated by respondents in surveys or in a laboratory situation. Notice that the latter evaluations are based on *ex ante* utility conceptions.

6 Conclusion

At the moment we see in the circles of some economists still a certain reluctance to recognize the significance of the happiness economics methodology, and especially to admit for a cardinal significance of satisfaction questions. This is no doubt the result of half a century of cursing cardinalism. It is our prediction that in the next decade the measuring of cardinal utility or satisfaction by means of satisfaction questions will become a matter of routine. Obviously, the extension to normative policy via the admission of cardinalist interpretations will not devalue the ordinal usage for compensation questions, etc. The admission of cardinalism is an extension of the methodology for new objectives, not a substitution of one method for another.

This implies that the methodology of what is now called "happiness economics" probably will become one of the major instruments of socio-economic policy. At the moment we stand just at the beginning.

References

Blanchflower, D.G. and A.J. Oswald (2004), "Well-Being over Time in Britain and the USA", *Journal of Public Economics* **88**, 1359–1386.

Boes S., M. Lipp and R. Winkelmann (2007), "Money Illusion under Test", forthcoming in *Economics Letters*.

Brickman, P. and D.T. Campbell (1971), "Hedonic Relativism and Planning the Good Society", in M.H. Apley, eds., *Adaptation-Level Theory: A Symposium*, Academic Press, NY, pp. 287–302.

Clark, A.E. and A.J. Oswald (1994), "Unhappiness and Unemployment", *Economic Journal* **104**, 648–659.

Clark, A.E., A.J. Oswald and P.B. Warr (1994), "Is Job Satisfaction U-shaped in Age?", CEPREMAP Working Paper No. 9407.

Clark, A.E. and A.J. Oswald (2002), "A Simple Statistical Model for Measuring How Life Events Affect Happiness", *International Journal of Epidemiology* **31**, 1139–1144.

Clark, A.E., P. Frijters and M.A. Shields (2006), "Income and Happiness: Evidence, Explanations and Economic Implications", Working Paper No. 2006-24, Paris-Jourdan Sciences Economiques.

Deaton, A.S. and J. Muellbauer (1986), ''On measuring Child Costs: With Applications to Poor Countries'', *Journal of Political Economy* **94**, 720–744.

Di Tella, R., R.J. MacCulloch and A.J. Oswald (2001), ''Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness'', *American Economic Review* **91(1)**, 335–341.

Di Tella, R. and R. MacCulloch (2006), ''Some Uses of Happiness Data in Economics'', *Journal of Economic Perspectives* **20(1)**, 25–46.

Dolan P. and D. Kahneman (2006), ''Interpretations of Utility and their Implications for the Valuation of Health'', Working Paper 2006, Tanaka Business School, Imperial College, London.

Easterlin, R. (1974), ''Does Economic Growth Improve the Human Lot?'', in P.A. David and M.W. Reder, eds., *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Academic Press, Inc., New York.

Ferrer-i-Carbonell, A. and B.M.S. Van Praag (2002), ''The Subjective Costs of Health Losses due to Chronic Diseases: An Alternative Model for Monetary Appraisal'', *Health Economics* **11**, 709–722.

Ferrer-i-Carbonell, A. and P. Frijters (2004), ''How Important Is Methodology for the Estimates of the Determinants of Happiness?'', *Economic Journal* **114(497)**, 641–659.

Ferrer-i-Carbonell, A. (2005), ''Income and Well-Being: An Empirical Analysis of the Comparison Income Effect'', *Journal of Public Economics* **89(5–6)**, 997–1019.

Frey, B.S. and A. Stutzer (2000), ''Happiness, Economy and Institutions'', *Economic Journal* **110**, 918–938.

Frey, B.S. and A. Stutzer (2002), ''What can Economists Learn from Happiness Research?'', *Journal of Economic Literature* **40(2)**, 402–35.

Frey, B., S. Luechinger and A. Stutzer (2004), ''Calculating Tragedy: Assessing the Costs of Terrorism'', Working Paper, Zurich.

Frey, B. and A. Stutzer (2005), ''National Indicators of Well-Being: A View from Political Economy'', presented at Annual Meeting of the European Public Choice Society, December 2005.

Goedhart, Th., V. Halberstadt and A. Kapteyn (1977), ''The Poverty Line: Concept and Measurement'', *The Journal of Human Resources* **12**, 503–520.

Houthakker, H.S. (1950), ''Revealed Preference and the Utility Function'', *Economica* **17**, New Series, 159–174.

Houthakker, H.S. (1961), ''The Present State of Consumption Theory'', *Econometrica* **29**, 704–740.

Kahneman, D., P.P. Wakker and R. Sarin (1997), ''Back to Bentham? Explorations of Experienced Utility'', *Quarterly Journal of Economics* **112**, 375–405.

Kahneman, D. (2004), ''Toward National Well-Being Accounts'', *American Economic Review* **94(2)**, 429–434.

Kapteyn, A. and B.M.S. Van Praag (1976), ''A New Approach to the Construction of Family Equivalence Scales'', *European Economic Review* **7(4)**, 313–335.

Kapteyn A. (1977), *A Theory of Preference Formation*, PhD Thesis, Leyden University.

Kapteyn, A. and F.G. van Herwaarden (1980), ''Interdependent Welfare Functions and Optimal Income Distribution'', *Journal of Public Economics* **14**, 375–397.

Layard, R. (2005), *Happiness*, Allen Lane, London.

Luttmer, E.F.P. (2005), ''Neighbors as Negatives: Relative Earnings and Well-Being'', *Quarterly Journal of Economics* **120(3)**, 963–1002.

Pareto, V. (1909), *Manuel d'économic politique*, Paris: Giard & Brière.

Robbins, L. (1935), *An Essay on the Nature & Significance of Economic Science*, 2nd edition. MacMillan, London (first edition 1932).

Samuelson, P.A. (1947), *Foundations of Economic Analysis*, Harvard University Press, Cambridge, USA.

Senik, C. (2004), ''When Information Dominates Comparison: Learning from Russian Subjective Panel Data'', *Journal of Public Economics* **889–10**, 2099–2123.

Senik, C. (2005), ''Income Distribution and Well-Being: What can we Learn from Subjective Data?'', *Journal of Economic Surveys* **19**, 43–63.

Van Praag B.M.S. (1968), *Individual Welfare Functions and Consumer Behavior - A Theory of Rational Irrationality*, North-Holland Publishing Company Amsterdam, Contributions to Economic Analysis, nr. 57.

Van Praag, B.M.S. (1971), ''The Welfare Function of Income in Belgium: an Empirical Investigation'', *European Economic Review* **2**, 337–369.

Van Praag, B.M.S. (1976), ''The Individual Welfare Function and its Offspring'', in J.S. Cramer et al. eds., *Relevance and Precision*, North-Holland Publishing Company, Amsterdam, pp. 263–293.

Van Praag, B.M.S., A. Kapteyn and F.G. van Herwaarden (1979), ''The Definition and Measurement of Social Reference Spaces'', *The Netherlands' Journal of Sociology* **15**, 13–25.

Van Praag, B., Th. Goedhart and A. Kapteyn (1980), ''The Poverty Line - a Pilot Survey in Europe'', *The Review of Economics and Statistics* **62(3)**, 461–465.

Van Praag, B.M.S. and N.L. van der Sar (1988), ''Household Cost Functions and Equivalence Scales'', *Journal of Human Resources* **23**, 193–210.

Van Praag B.M.S. (1991,1994), ''Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept'' *Journal of Econometrics* **50**, 69–89. Also published in R. Blundell, I. Preston, I. Walker eds., *The Measurement of Household Welfare*, Cambridge University Press, pp. 86–110.

Van Praag, B.M.S. (1993), ''The Relativity of the Welfare Concept'', in A. Sen and M. Nussbaum, eds., *Quality of Life*, Clarence Press, Oxford.

Van Praag, B.M.S. and M.F. Warnaar (1997), ''The Cost of Children and the Use of Demographic Variables in Consumer Demand'', in O. Stark and M.R. Rosenzweig eds., *Handbook of Population and Family Economics*, vol. 1A, Elsevier Science, Amsterdam, pp. 241–273.

Van Praag, B.M.S. and P. Frijters (1999), ''The Measurement of Welfare and Well-Being; the Leyden Approach'', in D. Kahneman, E. Diener and N. Schwarz, eds., *Well-Being: The Foundations of Hedonic Psychology*, Russell Sage Foundation, New York, pp. 413–433.

Van Praag, B.M.S. and P. Frijters (1999), ''Different Data Sources for Studying Behaviour'', in M.G. Fase and W. Kanning, eds., *Economics, Welfare Policy and the History of Economic Thought: Essays in Honour of Arnold Heertje*, Cheltenham, UK, pp. 290–309.

Van Praag, B.M.S., P. Frijters and A. Ferrer-i-Carbonell (2003), ''The Anatomy of Subjective Well-Being'', *Journal of Economic Behavior and Organization*, **51**, 29–49.

Van Praag, B.M.S. and A. Ferrer-i-Carbonell (2004), *Happiness Quantified: A Satisfaction Calculus Approach*, Oxford University Press, Oxford and New York.

Van Pragg, B.M.S. and B. Baarsma (2005), ''Using Happiness Surveys to Value Intangibles: The Case of Airport Noise'', *Economic Journal* **115** 224–246.

Van Praag, B.M.S. and A. Ferrer-i-Carbonell (2007), ''The Subjective Approach to Multidimensional Poverty Measurement'', in N. Kakwani and J. Silber, eds., *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave Macmillan.

CESifo Economic Studies, Vol. 53, 1/2007, 69–96, doi:10.1093/cesifo/ifm004
Advance Access publication 27 February 2007

Efficiency in Family Bargaining: Living Arrangements and Caregiving Decisions of Adult Children and Disabled Elderly Parents*

Liliana E. Pezzin†, Robert A. Pollak‡ and Barbara S. Schone§

Abstract

In this article, we use a two-stage bargaining model to analyze the living arrangement of a disabled elderly parent and the assistance provided to the parent by her adult children. The first stage determines the living arrangement: the parent can live in a nursing home, live alone in the community, or live with any child who has invited coresidence. The second stage determines the assistance provided by each child in the family. Working by backward induction, we first calculate the level of assistance that each child would provide to the parent in each possible living arrangement. Using these calculations, we then analyze the living arrangement that would emerge from the first stage game. A key assumption of our model is that family members cannot or will not make binding agreements at the first stage regarding transfers at the second stage. Because coresidence is likely to reduce the bargaining power of the coresident child relative to her siblings, coresidence may fail to emerge as the equilibrium living arrangement even when it is Pareto efficient. That is, the outcome of the two-stage game need not be Pareto efficient. (JEL classification: D1, J1, J2)

† Medical College of Wisconsin, Milwaukee, WI, USA. e-mail: lpezzin@mcw.edu

‡ Department of Economics and Olin School, Washington University; St. Louis, MO, USA and NBER, IZA and CESifo, e-mail: pollak@wustl.edu

§ Agency for Healthcare Research and Quality; Rockville, MD, USA and Georgetown University, e-mail: bschone@ahrq.gov

* An earlier version of this article appeared as NBER Working Paper No. 12358, July 2006. An even earlier version of much of the material in this article was included in 'Long-Term Care and Family Decision Making: Coresidence and Efficiency'. We are grateful to Shelly Lundberg, Jon Skinner, and two anonymous referees for helpful comments, to Joanne Spitz for editorial assistance, and to NIH grant R01 AG24049-01 for financial support. Pollak is grateful to the John D. and Catherine T. MacArthur Foundation for their support. Earlier versions of this material were presented at Washington University in St. Louis, the University of Pennsylvania, the NBER Cohort Studies group, the University of Kansas, the 2003 ASSA meetings in Washington, DC, the 2003 ESPE meetings in New York, the University of Michigan, the University of Victoria, the University of British Columbia, Carleton University, the University of Colorado, the University of California at Riverside, Williams College, the 2006 CESifo Area Conference on Employment and Social Protection, and the University of Illinois. The views in this paper are the authors'. No official endorsement by the Agency for Healthcare Research and Quality of the Department of Health and Human Services is intended or should be inferred.

1 Introduction

Many countries are currently facing the aging of their populations and aging populations imply an increasing number of disabled elderly persons requiring assistance with their everyday activities. The United States, many European countries, Australia, and Japan have all been actively investigating policies to reform their long-term care systems for the disabled elderly to control costs, improve access, and ensure quality.[1] Countries differ in the role of the government in financing long-term care, the extent to which they provide cash rather than in-kind assistance, and in the degree to which long-term care is means-tested rather than universal. Most countries, however, are committed to providing those disabled elderly who wish to remain in the community the opportunity to do so.[2] For example, in the US nearly two-thirds of the 5.5 million elderly with chronic disabilities rely, often exclusively, on family members for help with basic activities of daily living (Spillman and Pezzin 2000). Much of the assistance provided by family members takes place in the context of coresidence of the disabled elderly with adult children. Although evidence suggests a secular increase in living alone among the elderly (Kotlikoff and Morris 1990), coresidence of disabled elderly parents with their adult children remains fairly common in the US (Crimmins and Ingegneri 1990; Davis et al. 1997) in 2003 nearly one-fifth of all family households with a member 65 or older contained at least one of the householder's children aged 18 or older (US Census Bureau). Coresidence of elderly parents with adult children is even more prevalent in other parts of the world. In Europe, over one-quarter (26 percent) of all individuals 60 or older live with an adult child, with figures ranging from 43 percent in Spain to 4.1 percent in Denmark (United Nations 2005). Coresidence rates in Africa, Asia, and Latin America range from two-thirds to three-quarters (United Nations 2005).

Coresidence has been linked to both the well-being and health of the elderly. For example, disabled elderly persons who coreside with adult children are less likely than their noncoresiding counterparts to make a transition into a nursing home (Garber and MaCurdy 1990; Kemper and Pezzin 1996; Dostie and Leger 2005). Moon (1983) provides some evidence

[1] Of the 30 oldest countries, as measured by the percentage of the population over age 65, 25 are European countries (Gibson, Gregory and Pandya 2003).

[2] There are important social and cultural differences across countries that affect the role of families in providing care. For example, in France and Italy the expectation is that adult children will provide care, and public assistance is contingent on them providing care. In Germany family caregivers may be eligible for public pensions. See Gibson, Gregory and Pandya (2003) for a general discussion of differences in long-term care across Europe.

that coresidence promotes the health of the disabled elderly: she finds that those cared for in the home of a family member have better health outcomes than those in similar health who were cared for in an institution.

Coresidence may also benefit the child. Household public goods can be jointly consumed or produced; additionally, economies of scale and other efficiency gains in consumption and production make coresidence less costly than providing equivalent services without coresidence. For example, by sharing a household, adult children may lower the time cost of providing care since travel costs can be avoided. Coresidence, however, reduces privacy for the adult child and, depending on the alternative living arrangement, may also reduce privacy for the elderly parent.

The assistance provided to disabled family members is often the product of numerous individual and joint decisions by family members with different preferences facing different constraints. Family members not only make caregiving decisions on behalf of disabled family members but often provide hands-on care themselves and share the financial consequences of caregiving decisions. Moreover, the preferences of the disabled elderly may differ from those of their spouses and their adult children, and the preferences of one child may differ from those of another. Differences may arise about the type of care desired for the disabled elderly and the setting in which they receive it. For example, children may want a parent to enter a nursing home, while the parent prefers to live independently; or, a brother may want his sister to care for their mother while the sister wants her brother or his wife to be the primary caregiver. The possibility of conflict regarding caregiving and the roles of different family members in providing care suggest that family members may have incentives to behave strategically.

In this article we develop a model of family caregiving for the unpartnered disabled elderly that captures important features of the complex interactions between elderly parents and adult children and among the children. The partnered (e.g., married) disabled elderly are likely to receive care from their partners (e.g., spouses) rather than from children (Dwyer and Coward 1991; Pezzin, Pollak and Schone 2005b) and are less likely than the unpartnered to enter an institution (Freedman 1996). Thus, because of the costs they impose on government budgets, the unpartnered disabled elderly are a group of particular policy interest and we restrict our attention to them. To avoid analytical complications, we focus on the case of a widowed parent with unmarried children. This focus avoids the complications associated with another parent who may also

require care, and with the adult children's spouses who may have caregiving responsibilities for their own parents.[3]

We propose and analyze a two-stage game with three players: a disabled parent and two adult children. The first stage determines the living arrangements and the second stage determines intrafamily transfers. The stages are related because first-stage decisions affect second-stage bargaining power; the stages are distinct, however, because we assume that family members cannot or will not make binding commitments regarding their future behavior. We show that even if the second stage is conditionally efficient (i.e., efficient given the living arrangements determined in the first stage), the equilibrium of the two-stage game may be inefficient.

We argue that government long-term care policies have efficiency as well as distributional implications. The distributional implications of government long-term care policies are widely recognized and are analogous to the distributional implications of old-age benefits and health insurance for the elderly. That is, government policies redistribute the burden of caring for the elderly between generations, within generations, between families, and within families. Discussions of the efficiency implications of government long-term care policies often treat the disabled elderly as individuals rather than as members of families. When families are recognized, the long-term care literature had generally assumed that families respond efficiently to the incentives created by government policy. We relax that assumption.

Long-term care for the disabled elderly has sparked a growing literature. The economics literature, surveyed in Norton (2000), focuses on the supply and demand for nursing home care and on long-term care insurance, but pays little attention to the family. Early studies that discuss the role of the family concentrate primarily on support from children to parents in the form of shared housing, analyzing the determinants of living arrangements (Börsch-Supan 1989; Börsch-Supan, Kotlikoff and Morris 1991; Ellwood and Kane 1990; Kotlikoff and Morris 1990; Börsch-Supan et al., 1992; Börsch-Supan, McFadden and Schnabel, 1996).

The first generation of research on families' care arrangements relied on Becker's model of the family (Wolf and Soldo 1994; Ettner 1995 and 1996;

[3] Laferrère and Wolff (2006) survey the literature on family transfers. We ignore the possibility that transfers from children to their disabled elderly parents are motivated by the expectation of bequests, as suggested by the strategic bequest model of Bernheim, Shleifer and Summers (1985). Pollak (1988) argues that, if the parents would prefer to divide their bequethable wealth equally among their children, then the threat to disinherit a child who fails to provide sufficient attention is not credible. Perozek (1998) discusses empirical problems in Bernheim, Shleifer, and Summers (1985). McGarry (2006) surveys the literature on bequests and inheritances.

Kemper and Pezzin 1996). More recent work has used game-theoretic bargaining models to examine family care arrangements. Two examples will suffice. Pezzin and Schone (1999a, 2002a) specify and analyze a model of living arrangements, informal caregiving, labor force participation, and cash transfers. Their game involves two players, an elderly parent and an adult daughter, each with a utility function defined over a vector of private goods, leisure (for the daughter), and a public good, representing the parent's physical health. Conditional on the parent's disability, the production of the parent's physical health requires that the parent receive some form of care, whether formal (i.e., paid) or informal (i.e., unpaid). Parent and daughter make decisions that determine the levels of private consumption, leisure, cash transfers from the daughter to the parent, the combination of formal and informal care used to produce the parent's physical health or well-being, and the living arrangement (either separate or coresidence). As in the separate spheres model of Lundberg and Pollak (1993), Pezzin and Schone assume that intrahousehold allocation is determined as the solution to a cooperative Nash bargaining game in which the threat point is the Cournot–Nash equilibrium of a noncooperative game.

Heidemann and Stern (1999) and Engers and Stern (2002) develop a game theoretic model of family bargaining designed to motivate a structural empirical model of family long-term care decisions. In particular, they focus on determining whether the parent enters a nursing home, lives independently with no care provided by her children, or, if the parent does receive care from her children, which child becomes the primary caregiver. In their formulation, each adult child decides independently whether to attend a meeting in which living and care arrangements for the disabled parent are determined. Both voluntary and compulsory participation versions are analyzed and estimated, with results favoring the voluntary model. The children who participate reach a binding agreement while the non-participating children are excluded from family decision making and bear no responsibility for caring for the parent. For each child, the decision of whether to attend the meeting depends on the value she places on participating in the decision, the side payments that she anticipates, and the effect that she anticipates her presence at the meeting would have on the family's decision.

Much of the research examining family caregiving has focused on the parent–child dyad. Although most studies have included variables summarizing the characteristics of the remaining family network (Kotlikoff and Morris 1990; Pezzin, Pollak and Schone 2006a; and 2006b, Pezzin and Schone 1999b and 2002b; Stern 1994 and 1995), little work has analyzed interactions among the adult children. Papers that analyze caregiving with interactions of two or more children include

Engers and Stern (2002), Checkovich and Stern (2002), Pezzin and Schone (2006), and Pezzin, Pollak and Schone (2005a).

To analyze interactions among adult children, we build on research that has modeled intrahousehold allocation within a game theoretic framework (Manser and Brown 1980; McElroy and Horney 1981; Lundberg and Pollak 1993, 1994 and 2003). Game theoretic models are especially suitable for analyzing intergenerational living and transfer arrangements because they recognize the divergent and often conflicting interests of family members. They also specify a process for translating these divergent interests into outcomes. In the next three sections we analyze various three person games involving a disabled elderly parent and two adult children. We examine living arrangements, interhousehold transfers, and intrahousehold transfers. In section 2 we describe the two-stage caregiving game. Like all dynamic games, our two-stage sequential game is solved by backward induction. In section 3 we analyze the second-stage game. We argue that, for plausible specifications, the equilibrium of the second stage game is likely to be Pareto efficient conditional on the living arrangement when the parent lives on her own in the community, or in a nursing home, and may be Pareto efficient when she lives with one of her children. In section 4 we analyze the first stage game which determines the living arrangement. We show that, even if the equilibrium of the second stage game is conditionally efficient, the equilibrium of the full game need not be Pareto efficient. Section 5 is a brief conclusion.

2 A caregiving game with two children

We consider four possible living arrangements for the parent: living in a nursing home (A^n), living on her own in the community (A^o), living with child 1 (A^1), and living with child 2 (A^2). The parent's utility in each of these living arrangements depends on the living arrangement itself and on her consumption of a private consumption good (C). We use the superscripts (n, o, 1, 2) to indicate the living arrangement and the subscripts (p, 1, 2) to indicate family members. Thus, C_2^o denotes private consumption by child 2 when the parent lives alone and C_1^2 denotes consumption by child 1 when the parent lives with child 2. We assume that children care about their own private consumption and about the parent's private consumption. We also assume that the children care about the parent's living arrangement, which affects the parent's well being and the child's privacy.

Economists' usual assumption about preferences—what Becker (1981) calls "altruistic" preferences—implies that the children defer to the disabled parent's preferences over her own consumption and, in some

cases, over her living arrangements as well. For example, Becker's children would defer to the parent's preference between living in a nursing home and living on her own in the community. We regard Becker's altruistic preferences as an implausible special case even when the parent is mentally competent and especially implausible when the parent is cognitively impaired (e.g., with Alzheimer's). We also dislike Becker's terminology, which forces us to say that a child who rejects a disabled parent's preference for living alone (e.g., because the child believes that the parent cannot safely live alone) is not altruistic.[4] In this article we treat consumption as one dimensional, avoiding the issue of merit goods except in regard to living arrangements. Pezzin and Schone (1999a, 2002a) analyze a model with two goods, one of which is a merit good.

We model family interactions as a two-stage game in which both stages may contain substages. The first stage is noncooperative and determines the living arrangement. The second stage determines consumption. We assume that family members cannot or will not make binding agreements at the first stage regarding care giving at the second stage. Hence, the assistance that a child provides at the second stage, although it may be predictable at the first stage, is determined at the second stage.[5]

Although we cannot directly observe the impact of not having binding agreements on decisions made at the second stage, the lack of binding agreements may partially explain some empirical regularities associated with caregiving. For example, the concentration of caregiving by coresident caregivers and the relatively small amount of assistance provided by the noncoresident children may reflect the fact that, once the parent begins to live with one child, the other children reduce the care they provide confident that the coresident child will make up the difference.

Suppose, for example, that the first stage begins with the children deciding, separately and simultaneously, whether or not to invite the parent to coreside and ends with the parent choosing among the feasible living arrangements: she can move into a nursing home, live on her own,

---

[4] In the introduction to the 1991 edition of his *Treatise on the Family*, Becker concedes that "The most unsatisfactory aspect of my discussion . . . [in the 1981 edition is] . . . the failure to combine the discussion of 'merit goods' and altruism" (p. 10). Just as merit goods can motivate paternalistic governments to provide tied transfers (e.g., food stamps), merit goods can motivate paternalistic family donors, when they have the ability, to provide tied rather than untied transfers. Pollak (1988) proposed a model in which family members have paternalistic preferences. In place of Becker's term, "altruistic" preferences, Pollak (2003) argues that "deferential" preferences is more descriptive.

[5] As a referee points out, living arrangements are not determined in a one-shot game. Most obviously, a decision that the parent live on her own in the community can be revisited if her health deteriorates. Even the decision to move into a nursing home is not irrevocable.

or accept the invitation of any child who has invited her to coreside. At the second stage, taking as given the living arrangement determined at the first stage, the children and the parent make decisions that determine resource allocation under that living arrangement. We can model the second stage as a noncooperative game or as a cooperative game; alternatively, we can finesse some but not all of the difficulties of modeling the second-stage game by postulating an "allocation rule". We define an allocation rule as a specification of each family member's second-stage behavior as a function of the living arrangement, which is determined in the first stage, of the economic and demographic characteristics of all players, and of any relevant policy parameters (e.g., voucher programs, tax subsidies for caregivers, allowances for dependent care).

We have adopted and adapted the idea of an allocation rule from Chiappori (1988, 1992) who proposed a "sharing rule". In the context of allocation between spouses within marriage, the sharing rule specifies the utility of the two spouses as a function of variables assumed to determine bargaining power. Chiappori assumes that the sharing rule is single-valued and Pareto efficient. Chiappori postulates the sharing rule directly, and does not attempt to derive it from an underlying model of bargaining. Unlike Chiappori's sharing rule, our allocation rule describes the allocation of goods rather than utility and, again unlike Chiappori's sharing rule, our allocation rule need not be single-valued or Pareto efficient. By beginning with the allocation rule, we avoid not only the need to analyze the second-stage game but also the need to specify it, or even to specify whether it is cooperative or noncooperative.

A priori, it is unclear whether family decision making should be modeled as a cooperative or a noncooperative game. Shubik (1989, p. 103) asserts that noncooperative game theory "is generally not so useful to describe complex, loosely structured social interaction". Family bargaining—whether between spouses within marriage, between an adult child and a disabled parent, or among adult children—exemplifies such interactions. But the usefulness of cooperative game theory is also problematic. First, although cooperative game theory allows us to proceed without specifying the "rules of the game"—the strategies available to each player, or in extensive form, the sequence of moves and the information available to each player at each move—noncooperative game theory teaches that the rules of the game are often crucial determinants of the outcome. Thus, by modeling interactions as a cooperative game we necessarily disregard the strategic factors that may determine the outcome. Second, the efficiency of social arrangements and practices is a central concern of economics and of public policy, yet cooperative bargaining models assume that outcomes are Pareto efficient. Hence, cooperative bargaining models are incapable of investigating the conditions that make it possible to achieve and sustain

efficient outcomes because cooperative models refuse to recognize the possibility of inefficiency. Because complex, loosely structured social interactions are very difficult to model, we regard the choice between modeling family interactions as a noncooperative or a cooperative game not a matter of principle but of research strategy. In the following section we consider alternative specifications of the second stage game.

3 The second stage game: child to parent transfers

For each of the four possible living arrangements, we consider several formulations of the second stage game which determines transfers from children to the parent. We assume that each child's utility is increasing in the parent's consumption and in her own consumption, but is independent of her sibling's consumption. Hence, a child's utility increases if the sibling's contribution to the parent increases.[6]

3.1 A^o. Suppose that the parent lives on her own in the community

When the parent lives independently in the community or in a nursing home, we assume that the parent accepts whatever transfers the children offer her. Hence, with these living arrangements, the parent is not a strategic player in the second stage game.[7] The children may play a one-shot noncooperative game, a repeated cooperative game, or a cooperative game. We consider each of these in turn.

A^o.i. A noncooperative, one-shot, voluntary contribution game

The one-shot Cournot-Nash game is familiar but not very plausible. Using "reaction functions" which show each child's best response to the transfers made by the other, we can calculate the equilibrium. (Figure 1). When public good provision is a simultaneous move one-shot game, the public good will be underprovided. More specifically, if both children make positive contributions to the family public good, then the equilibrium is Pareto inefficient: there exist strictly greater contribution levels that both children would prefer to the Cournot–Nash equilibrium.

This underprovision result does not hold if only one child makes positive contributions in the Cournot–Nash equilibrium. In the one

[6] Consider two polar cases: (i) the child responds to the increase in the siblings's contribution by reducing her own contribution by the same amount, thus increasing her own consumption and leaving the parent's consumption unchanged. (ii) the child responds by leaving her own contribution to the parent unchanged, thus increasing the parent's consumption and leaving her own consumption unchanged.

[7] When the parent coresides with one of the children, we discuss cases in which the parent is a strategic player.

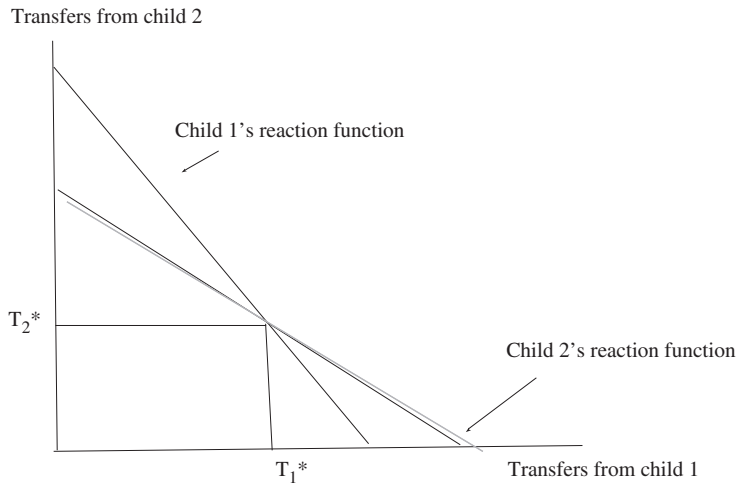


Figure 1 Children's reaction functions for transfers

contributing child case, just as in the one-child case, provision is Pareto efficient.

A^o.ii. A noncooperative, repeated, voluntary contribution game

Now, suppose the children play a repeated voluntary contribution game. More specifically, suppose that at the beginning of every period each child has the opportunity to make transfers to the parent. To simplify the analysis, suppose that neither the parent nor the children can carry over resources from one period to another, so that the stage games played in successive periods are identical. Hence, the stage games are related only because the children can punish each other for misbehavior by reducing their own contributions to the public good.

If the children are sufficiently patient, then the folk theorem asserts that any feasible, individually rational allocation is a subgame perfect equilibrium of the repeated game. Hence, the repeated game has a very large set of subgame perfect equilibria some of which are Pareto efficient, but many of which are not. Unless we introduce additional assumptions about how family members play the game, we cannot narrow down the set of subgame perfect equilibria or assign probabilities to them. A cogent objection to modeling children's transfers to a disabled elderly parent as an infinitely repeated game is the assumption that the game will continue indefinitely. Because the game will eventually stop, end game considerations may affect behavior from the outset. The theoretical argument that games with known end points unravel is strong, but empirical evidence

suggests that unraveling need not occur.[8] Stochastic termination is more plausible, perhaps because of the death of the parent or one of the children or a dramatic change in the parent's health status (e.g., a stroke requiring hospitalization, making continued independence infeasible). But stochastic termination complicat the analysis because the stage games played in successive periods are not identical.

A^o.iii. A cooperative game

Now suppose, as Shubik would have us assume in analyzing a "complex, loosely structured social interaction", that the second-stage game is cooperative. We can conclude immediately that, conditional on the living arrangement, the second stage equilibrium is Pareto efficient. Unlike noncooperative games, cooperative games make no mention of strategies or moves, but instead require us to specify both the payoffs attainable by each coalition and a solution concept. We discuss two solution concepts: the Nash bargaining solution and the core. The Nash bargaining solution, the leading solution concept in bargaining models of marriage, selects a particular Pareto-efficient outcome as *the* solution (Figure 2). Which Pareto-efficient allocation it selects depends on the specification of the threat point (T^1, T^2). Bargaining models of marriage have emphasized Nash bargaining and neglected other cooperative bargaining models and solution concepts. For example, although Manser and Brown considered both the Nash and the Kalai–Smorodinsky (1975) bargaining solutions, subsequent work on bargaining in families has virtually ignored Kalai–Smorodinsky. Gugl (2004) provides an interesting exception, considering both the Nash and Kalai–Smorodinsky bargaining solutions. Gugl's work suggests that the difficulty of doing comparative statics with Kalai-Smorodinsky may account for its eclipse by the Nash bargaining solution.

The core is the set of feasible, undominated allocations—allocations that cannot be improved upon by any coalition. With two players, the core is the set of utility allocations corresponding to points on the frontier in Figure 2 (i.e., the set of Pareto-efficient allocations). With three players, core allocations must not only be Pareto efficient (i.e., the coalition of all three players cannot improve upon the allocation), but must satisfy the additional condition that no coalition of two players can improve their own well-being at the expense of the third player. A drawback of the core as a solution concept is that it may fail to predict a unique equilibrium and that the set of equilibria may be large.[9] Without additional assumptions,

[8] The "centipede game" is the standard example of an experimental game in which unraveling does not occur. See Kreps (1990) and McKelvey and Palfrey (1992).

[9] If the players are sufficiently patient, the folk theorem implies that the set of equilibria in the repeated game is much larger than the core.

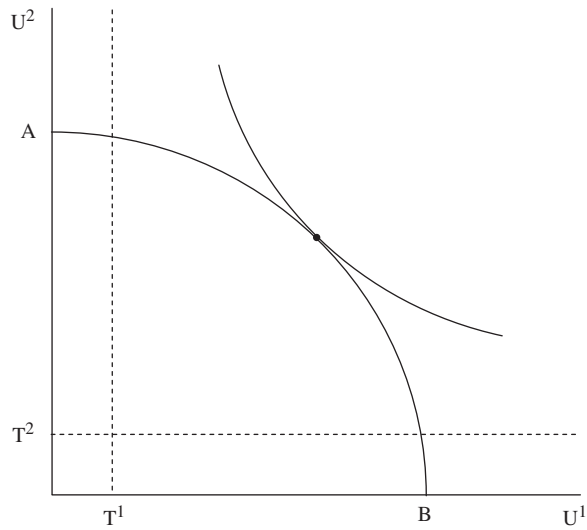


Figure 2 The Nash bargaining solution between the children

we cannot reduce the set of equilibria or assign probabilities to the elements of the set. The core, despite its prominence in game theory, has received almost no attention as a solution concept in the economics of the family, perhaps because it does not yield a unique solution in two-person games, and perhaps because, in games with more than two players, the core may be empty.

A^o.iv. Pareto efficiency when the parent lives on her own in the community

When the parent lives on her own in the community Pareto efficiency is plausible but not necessary. The Coasian tradition holds that efficient outcomes emerge from bargaining, at least when the number of agents is small, information is symmetric, and transaction costs are negligible. But the Coasian tradition rests on assumption, not on argument. Indeed, a critic might argue that the Coasian tradition amounts to little more than following Shubik's suggestion and assuming that bargaining is a cooperative game.

3.2 A^n. The parent lives in a nursing home

The logical structure of the second-stage game when the parent lives in a nursing home is essentially identical to its structure when she lives on her own in the community. As when the parent lives on her own in the

community, in the second stage game the children take the parent's living arrangement as given. Not surprisingly, some specifications of the second stage game imply efficient equilibria, other specifications imply inefficient equilibria, and still others imply multiple equilibria, some efficient and some inefficient. As when the parent lives on her own in the community, Pareto efficiency is plausible but not necesssary. Because at this level of abstraction the efficiency of transfers when the parent lives in a nursing home presents no new conceptual issues, we discuss it no further.

3.3 A^1, A^2. The parent resides with a child

Coresidence increases the strategic asymmetry between the children and weakens the bargaining power of the coresident child. Of course strategic asymmetry is always present. Even when the parent lives independently in the community or in a nursing home, the children may differ in gender, family responsibilities, labor force attachment, and attachment or proximity to the parent.[10] Like coresidence, many of these differences are endogenous. For definiteness, and without loss of generality, for the remainder of this section in which the parent coresides with a child, we suppose that the parent coresides with child 1.

We emphasize the strategic asymmetry between the coresident and the noncoresident child because it has implications for the choice of living arrangements in the first stage game. We begin by assuming that the parent is a passive spectator rather than a strategic player. Using the model proposed by Weiss and Willis (1985) in the context of child support following divorce, we examine the implications of coresidence for the noncoresident child's ability to monitor the way transfers are used by the coresident child. We then consider the way in which coresidence and the frequency of contact it implies is likely to affect the coresident child's awareness of the parent's needs or her attachment to the parent. Finally, we allow the parent to be a strategic player, assuming that allocation within the coresident household is the outcome of a cooperative game between the coresident child and the parent.

Weiss and Willis provide a framework for analyzing the effect of coresidence on bargaining power. Their concern is child support following divorce, but the strategic position of the noncoresident child contemplating contributing to the coresident household is similar to that of the noncustodial parent contemplating child support. In Weiss and Willis the child's well-being is a parental public good valued by both parents.

[10] Konrad et al. (2002) argue that older children exploit their first mover advantage by moving away from their parents, leaving younger children to bear a disproportionate share of long-term care responsibilities.

Each parent, however, is also concerned with his or her private consumption and unconcerned with the private consumption of the ex-spouse. The noncustodial parent, for definiteness, the divorced father, because he does not coreside with the child, is poorly positioned to monitor his ex-wife's allocation of child support between herself and the child. Weiss and Willis view the inability of the father to monitor the mother's allocation of resources between herself and the child as the crucial feature of the strategic situation. The inability to monitor precludes binding, enforceable agreements between the parents: the father is rationally concerned that if he increases his contribution, his ex-wife will respond by reducing hers.

Weiss and Willis model child support as a one-shot Stackelberg game: the first mover, the father, contributes resources to the mother; the mother then allocates resources between herself and the child. As Weiss and Willis show, the equilibrium allocation is Pareto inefficient: both parents would prefer an allocation in which they both reduced their private consumption and increased their transfers to the child. The Stackelberg game captures the intuition that the child will receive less than the Pareto-efficient level of resources. The Weiss and Willis conclusion is appealing, although asymmetric information, inability to monitor, and the consequent inability to make binding agreements, play no role in their formal model.

Neither child support nor long-term care is a one-shot Stackelberg game, but the Weiss and Willis insight about the strategic importance of the inability to monitor applies to both. The analogy between long-term care and the Weiss and Willis Stackleberg model of child support is closest when the parent has a cognitive disability such as severe Alzheimer's that precludes her active participation in the allocation process. Under these circumstances, the coresident child allocates resources between herself and the parent, just as in Weiss and Willis the mother allocates resources between herself and a child. Replacing the one-shot Stackelberg game by a sequence of Stackelberg games would not alter the analysis. If the noncoresident child is unable to monitor the allocation of resources in the coresident household, then a repeated game would not remove the problems because the noncoresident child lacks the information required to know when to "punish" the coresident child. The analysis of both long-term care and child support becomes more complicated if the individual receiving care is an active player rather than a passive recipient of care.

Once coresidence is established, the coresident child may have incentives to maintain it because termination would impose high psychic costs or adversely affect instrumental or affective relationships with other family members. That is, once coresidence becomes the status quo, the coresident child may find termination difficult and costly. We can interpret the coresident child's incentives to continue coresidence in terms of rewards

offered for continuing or, equivalently, in terms of punishments threatened for terminating. The noncoresident child, knowing that her sister cannot easily terminate coresidence, realizes that if she reduces her contribution, the coresident child will respond by increasing hers. England and Folbre (2003, p.73), describing the predicament of paid care workers, write: "these emotional bonds [to those receiving care] put care workers in a vulnerable position. We might call the workers 'prisoners of love'; a kind of emotional 'hostage effect' comes into play". The logic of their argument applies with even greater force to care provided by family members.[11]

We now drop our assumption that the parent is a passive spectator and assume instead that she is an active player. We begin with the one-child case and consider the implications of coresidence for bargaining between the coresident child and the parent. We then return to the two-child case, and consider the implications of coresidence for bargaining between the noncoresident child, the coresident child, and the parent.

Following Pezzin and Schone (1999a, 2002a), we assume that when the parent and a child coreside their interactions are cooperative but when they do not coreside their interactions are noncooperative. We assume that, within the coresident household, control over resources affects allocation. That is, government and family transfers to the coresident child have a different effect than transfers to the parent. More formally, resources controlled by the coresident child and resources controlled by the parent are separate arguments of the coresident household's allocation rule. Empirical work by Hayashi (1995) and by Pezzin and Schone (1997) on allocation within two-generation households finds that resources controlled by the elderly parents have a different effect on household expenditure patterns than resources controlled by their coresident adult children.[12]

Government programs that provide direct payments, in-kind services, or tax incentives to households with a disabled elderly member are uniform across households rather than tailored to the allocation rules of particular households. Government transfers, like private transfers, affect allocations in the coresident household through its allocation rule. Hence, government policies that provide resources to disabled elderly parents will have systematically and predictably different effects than policies that provide the same resources to coresident children.

---

[11] In a dynamic model, contributions by the noncoresident child might decrease over time while contributions by the coresident child might increase; our static models, however, cannot accommodate this behavior.

[12] In a similar vein, empirical work by Lundberg, Pollak and Wales (1997) on allocation within married couple households finds that resources controlled by husbands have a different effect on household expenditure patterns than resources controlled by wives.

We now return to the two child case. To model the asymmetry between the positions of the noncoresident child and the coresident child, we assume that the second stage game contains two substages: in the first substage, the noncoresident child makes transfers to the coresident child and to the parent; in the second substage, the coresident child and the parent play a cooperative game. For definiteness, think of the coresident household's behavior as determined by a cooperative Nash bargaining game in which the threat point depends on the allocation of transfers between the coresident child and the parent.[13] Thus, the noncoresident child must decide not only how much to transfer to the coresident household, but also on the allocation of transfers between the coresident child and the parent.[14] We represent the allocation that emerges from the full second stage game by an allocation rule.

Pareto efficiency is less plausible when the parent coresides with one of the children. Within the coresident household, Pareto efficiency depends on assuming either that the parent is not an active player or, alternatively, that the parent and the coresident child bargain to a Pareto-efficient allocation. Of course Pareto efficiency in the coresident household is guaranteed if the parent and the coresident child play a cooperative game. The difficulty, however, is not within the coresident household but in interactions between the coresident household and the noncoresident child. The situation is similar to that modeled by Weiss and Willis in their analysis of child support by noncustodial (and noncoresident) parents. Analogous to the noncustodial parent in Weiss and Willis, the noncoresident child might be willing to contribute or contribute more to the disabled elderly parent, but refrains from doing so because she is concerned that the coresident child will "tax" the increased contribution and spend some portion of it on herself. Thus, when the parent coresides with one of her adult children, Pareto efficiency in the second stage game is possible but not as plausible as when the parent lives on her own in the community or in a nursing home.

---

[13] We do not assume that the parent and the coresident child play as a "team". The team assumption, like the unitary model, would imply that the parent and the coresident child act as a single player. That is, the coresident household would act as if it had a single preference ordering and a single budget constraint (i.e., as if the parent and the coresident child pooled their resources). If the coresident child and the parent are a team, then the behavior of the coresident household would be independent of how the noncoresident child or the government allocated transfers between the coresident child and the parent.

[14] The government may allocate a portion of its transfers to the coresident child rather than the disabled elderly parent, and by doing so may induce a child to invite coresidence. A child cannot use the prospect of such transfers to induce a sibling to offer coresidence unless at least of one the children is willing and able to make a binding commitment.

4 The first stage game: living arrangements

In this section we analyze the first stage game. In section 4.1 we show that the equilibrium of our two-stage game can be Pareto inefficient even when the second stage game is Pareto efficient conditional on the living arrangement. In section 4.2 we show that the equilibrium may depend on the precise specification of the first-stage game (e.g., when the game is sequential, the equilibrium may depend on which child moves first), and we consider games in which the parent is an active player. Finally, in section 4.3 we show that the game may have multiple equilibria; more specifically, we consider a game in which all family members move simultaneously and show that five of its 12 strategy profiles are equilibria.

We begin by describing the first stage game. Because the parent cares about living arrangements as well as private consumption, she might prefer to live independently with fewer consumption goods than live with child i with more.Suppose that the second-stage game associates with each living arrangement unique consumption values for the parent and the children. We assume that the parent can rank the four possible alternatives $[(C_p^n, A^n), (C_p^o, A^o), (C_p^1, A^1), (C_p^2, A^2)]$ or, if the parent is also concerned with the children's consumption, $[(C_p^n, C_1^n, C_2^n, A^n), (C_p^o, C_1^o, C_2^o, A^o), (C_p^1, C_1^1, C_2^1, A^1), (C_p^2, C_1^2, C_2^2 A^2)]$. In either case, these preferences over consumption and living arrangements induce a preference ranking over living arrangements by the parent and by each of the children. These induced preferences play a crucial role in our analysis of the first stage game. Each child's preferences reflect his or her concern for the parent's private consumption as well as the implications of each living arrangement for the child's private consumption.[15]

As an initial example, we model the first stage as consisting of simultaneous moves by the siblings, followed by a decision by the parent, who chooses among the living arrangements available to her. Each child has two moves: inviting coresidence (v) or not inviting coresidence (v′). For each profile of moves (e.g., both invite coresidence; child 1 invites coresidence and child 2 does not, etc.), we assume that the resulting levels of utility are known to each child (e.g., if both invite coresidence, they know that the parent will choose to live with child 1), or, more precisely, they can assign probabilities to each living arrangement.

[15] Problems can arise if the second-stage game associates more than one consumption value with some living arrangement. More specifically, the analysis proceeds with no difficulty provided probabilities are associated with each of the possible consumption outcomes. Without such probabilities, however, the analysis is dead in the water.

In the first stage example described above, the parent will choose her preferred living arrangement from the available options determined by the children's invitations. This choice is based on the parent's calculation of the utility levels attainable in each living arrangement. The parent faces at most four alternatives, depending on whether both children invite coresidence, neither child invites coresidence, or one child invites coresidence and the other does not. Two loose ends remain: nonuniqueness in the solution to the second stage game, and nonuniqueness in the parent's choice.

If the allocation rule associates a unique consumption pattern with each living arrangement, then the parent will prefer one to the other or be indifferent between them. We assume that the parent's ranking of living arrangements is an ordering. If the ordering is strict (i.e., no ties), then we can proceed directly to the analysis of the first stage game. If the ordering is not strict, then the parent will sometimes face situations in which the "best" living arrangement is not unique. Although such ties pose no problem for the parent, they do for the children who must decide at the first stage whether to invite coresidence. Henceforth, we assume that the ordering is strict to avoid the difficulties associated with ties.

A multiplicity of equilibria in the second stage game poses a troubling problem.[16] The problem arises because the parent cannot choose among living arrangements unless she can assign probabilities to each allocation in the set. If the allocation rule associates probabilities with each allocation in the set, then choosing among living arrangements is like choosing among lottery tickets. In the absence of such probabilities, however, we encounter difficulties modeling parental choice.

Before turning to our examples, we dispose of an expositional complication. We want to treat family members as if they have direct preferences over living arrangements. Provided the parent and the children can predict the transfers that would take place in each possible living arrangement, we can legitimately focus on induced preferences over living arrangements, relying on the fact that each living arrangement is associated with unique levels of private consumption, of care for the parent, and of privacy for the children and for the parent. Hence, instead of carrying forward notation that explicitly recognizes the role of private consumption for the children and for the parent, we work with the induced preferences over living arrangements.

[16] Multiple equilibria can arise in cooperative second stage games (e.g., when the core is the solution concept) and in noncooperative second stage games (e.g., in repeated games as a consequence of the folk theorem).

4.1 Pareto inefficient equilibria

To construct an example of a game with an inefficient equilibrium, we begin by specifying the preferences of each family member. Suppose preferences for the parent's living arrangement (taking account of the transfers that would be made and the implied consumption pattern) are represented by:

	Parent's ranking	Child 1's ranking	Child 2's ranking
	Parent lives:		
First choice	with child 1	with child 2	with child 1
Second choice	with child 2	independently	independently
Third choice	independently	with child 1	with child 2
Fourth choice	in nursing home	in nursing home	in nursing home

That is, the parent prefers to live with child 1, but would rather live with child 2 than live independently. Each child prefers that the parent coreside with the other child, and each child would prefer that the parent live independently rather than coreside with the parent. The unique equilibrium, indeed, the dominant strategy equilibrium, of any game in which the children have these preferences has the parent living independently. With these preferences, the nursing home living arrangement is an option that the parent would never choose, and one that the children know that she would never choose.

Suppose, however, that each child would invite coresidence if she knew her sister would contribute "enough", and that the sister would rather contribute enough than have the parent live independently.[17] An omniscient and omnipotent social planner could impose a solution on the family that would make everyone—the parent and both children—better off by requiring that the parent live with child 1 and requiring child 2 to contribute "enough". But the family cannot achieve this or any other Pareto-efficient solution and is misguided, as if by an invisible hand, to a Pareto inefficient equilibrium. The demonstration of inefficiency depends on comparing a living arrangement and transfer pattern that a social planner might impose with the living arrangement and transfer pattern

[17] Anna Rubinchik-Pessach suggests formalizing this by supposing that there are two possible contribution levels, 0 and T**, and writing down preference profiles corresponding to each of them. Thus, everyone prefers the outcome in which child 2 contributes T** and mom lives with child 1, but that outcome is not an equilibrium of the two-stage game.

that would emerge as the equilibrium of a two-stage game. Our argument does not establish nor do we claim that for all configurations of preferences the equilibrium of the two-stage game is inefficient. We have established that for some configurations of preferences the equilibrium is inefficient.

Lundberg and Pollak (2003) describe and analyze a related two-stage game in the context of bargaining within marriages: the "two-earner couple location problem". In Lundberg and Pollak, spouses play a two-stage game in which the first stage determines the location (e.g., whether the couple moves to the husband's preferred location or the wife's preferred location), and the second stage determines allocation within marriage. When the spouses prefer different locations, inefficient outcomes (e.g., inefficient divorces) are possible even when the second stage game is efficient conditional on the location determined in the first stage. An analogous result holds in our long-term care game—the equilibrium of the two-stage long-term care game may be an inefficient living arrangement, even when second stage transfers are efficient conditional on the living arrangement. The crucial features of both the two-earner couple location game and our long-term care game are that first-stage decisions affect future bargaining power and that family members cannot or will not make binding, enforceable agreements.

4.2 Outcomes depend on the structure of the first-stage game

To construct an example in which the equilibrium living arrangement depends on the precise specification of the first stage game, we again begin by specifying each family member's preferences.

	Parent's ranking	Child 1's ranking	Child 2's ranking
	Parent lives:		
First choice	with child 1	with child 2	with child 1
Second choice	with child 2	with child 1	with child 2
Third choice	independently	independently	independently
Fourth choice	in nursing home	in nursing home	in nursing home

That is, the parent prefers to live with child 1, but would rather live with child 2 than live independently. Each child prefers that the parent coreside with the other child, but each child prefers coresidence with the parent to having the parent live independently. Both children and the parent prefer having the parent live independently rather than in a nursing home.

Using these preferences, we consider alternative specifications of the first stage game. We first consider three specifications of the noncooperative first-stage game in which the children move before the parent, then three specifications in which the parent moves before the children. Finally, we consider a specification in which the parent and the children move simultaneously.

Children move before the parent

We consider two sequential games that differ in which child moves first and then consider a simultaneous move game. When child 1 moves first, she does not invite (v') the parent to coreside; the best response of child 2 is to invite (v) the parent to coreside, and the parent accepts the invitation. When Child 2 moves first, she does not invite the parent to coreside; the best response of child 1 is to invite the parent to coreside, and the parent accepts the invitation. The simultaneous move game has two pure strategy equilibria: (i) Child 1 invites coresidence, and child 2 does not; the parent accepts the invitation of child 1. (ii) Child 2 invites coresidence, and child 1 does not; the parent accepts the invitation of child 2.[18]

These examples demonstrate that the equilibrium of our two-stage game can depend on the precise specification of the game (e.g., which child moves first in the sequential game) and that some specifications of the game (e.g., the game in which the children move simultaneously) can have multiple equilibria. In section 4.3 we offer an example with a richer set of equilibria.

Parent moves before the children

When the parent moves first, she can be a strategic player. Suppose that the parent can commit herself to reject particular invitations, if she should receive them.[19,20] More specifically, the game begins with the parent choosing among three moves.

(i) preemptively reject an invitation from Child 1 (r1)
(ii) preemptively reject an invitation from Child 2 (r2)
(iii) make no preemptive rejection of any invitation (r′)

---

[18] The simultaneous move game also has a mixed strategy equilibrium which we shall not discuss.

[19] Shelly Lundberg suggests that the parent might achieve the same result by insulting a child's spouse.

[20] We assume that the parent cannot commit herself to reject living independently or in a nursing home. If she could commit herself to rejecting one or both of these living arrangements (e.g., by starving herself to death), she might be able to force coresidence.

The children then move, either sequentially or simultaneously, as described above. Finally, the parent chooses a living arrangement: she can live independently or in a nursing home or accept any invitation she has received except those she has preemptively rejected.

The analysis of these games is straightforward. The parent begins by committing herself to reject an invitation from child 2. The equilibrium of all three games—the two sequential games and the simultaneous game—is an invitation from child 1, which the parent accepts. This example shows that the ability of a family member (in this case, the parent) to commit can alter the equilibrium outcome and, in this case, commitment enables the parent to achieve the outcome she prefers.

4.3 Multiple equilibria

The parent and the children move simultaneously

The simultaneous game has 12 strategy profiles and 5 of these are equilibria. Recall that the disabled parent has 3 possible strategies (r1, r2, r′), and each child has two possible strategies (v, v′). The reader can verify that the 5 equilibria are:

(r1, v′, v) parent lives with child 2
(r2, v, v′) parent lives with child 1
(r2, v, v) parent lives with child 1
(r', v′, v) parent lives with child 2
(r', v, v′) parent lives with child 1.

As this example shows, some specifications of the game have a large number of Nash equilibria.[21]

4.4 Agency for the parent

As we have already seen, the parent becomes a strategic player in the first-stage game if she has the ability to reject certain invitations. An alternative specification of the first-stage game can increase the parent's bargaining power. Instead of assuming that the children issue invitations which the parent must accept or reject, suppose that the parent issues invitations which the children must accept or reject.

For the configuration of preferences in our first example, the parent derives no strategic advantage from issuing invitations: regardless of what invitations she issues, the only equilibrium is living independently

[21] If the parent moves first and the children observe her move and then play a simultaneous move game, then all of these Nash equilibria are subgame perfect.

in the community.[22] For the configuration of preferences in our second example, the parent can achieve her preferred outcome (living with child 1) if she can issue invitations. For example, if the rules of the game require her to issue all invitation simultaneously, then she can accomplish her desired outcome by issuing an invitation only to child 1. If the rules of the game allow sequential invitations, she can accomplish her desired outcome by first inviting child 2 to coreside and, when child 2 rejects her invitation, inviting child 1 to coreside.

5 Conclusion

We have used a two-stage bargaining model to analyze the living arrangement of a disabled elderly parent and transfers to the parent from her adult children. The first stage determines the living arrangement, the second child-to-parent transfers. Working by backward induction, we first calculate an allocation rule that specifies the level of transfers that each child would provide to the parent in each living arrangement. We then analyze the living arrangement(s) that emerge as equilibria of the first stage game. Because the living arrangement affects bargaining power in the second stage game, and because family members at the first stage are unwilling or unable to make binding agreements regarding transfers at the second stage, the equilibria of the two-stage game may be Pareto inefficient even if the equilibrium of the second stage subgame is conditionally efficient.

A better understanding of the process by which families come to assume the responsibility and share the burden of caring for the disabled elderly is essential for designing and evaluating long-term care policies. Governments have long recognized that their long-term care policies have distributional implications. That is, long-term care policies affect how the burden of long-term care is shared between generations, within generations, between families, and within families. As governments increasingly explore policies to address the needs of their growing disabled elderly populations, the possibility that families' long-term care decisions may result in inefficient outcomes suggests an additional role for public policy: promoting efficiency. Initiatives may be tailored to family living arrangements, for example, by taxing non-coresiding adult children or subsidizing co-residing adult children for both their informal care services and relative loss of bargaining power, in order to promote more

[22] The outcome might be different if the parent could commit herself to moving into a nursing home, the worst outcome for her and for the children, unless child 1 would coreside with her.

efficient outcomes.[23] While government transfer schedules that are independent of family living arrangements are analytically simpler than schedules that are not, contingent transfer schedules may limit the deadweight loss associated with government efforts to ensure the availability of long-term care. Public long-term care policy will be enhanced by recognizing that the caregiving behavior of family members responds to the incentives created by public programs.

References

Becker, G.S. (1981, 1991), *Treatise on the Family*, Harvard University Press, Cambridge, MA.

Bermheim, B.D., A. Shleifer and L. Summers (1985), "The Strategic Bequest motive", *Journal of Political Economy* **93**, 1045–1076.

Börsch-Supan, A. (1989), "Household Dissolution and the Choice of Alternative Living Arrangements Among Elderly Americans", in D. Wise, ed., *The Economics of Aging*, University of Chicago Press, Chicago, pp. 119–150.

Börsch-Supan, A., L. Kotlikoff and J. Morris (1991), "The Dynamics of Living Arrangements of the Elderly: Health and Family Support", in J. Pacolet and C. Wilderom, eds., *The Economics of Care of the Elderly*, Avebury, London, pp. 114–135.

Börsch-Supan, A., V. Hajivassiliou and L. Kotlikoff, et al. (1992), "Health, Children and Elderly Living Arrangements", in D. Wise, ed., *Topics in the Economics of Aging*, University of Chicago Press, Chicago, pp. 79–103.

Börsch-Supan, A., D. McFadden and R. Schnabel (1996), "Living Arrangements: Health and Wealth Effects", in D. Wise, ed., *Advances in the Economics of Aging*, University of Chicago Press, Chicago, pp. 193–212.

Centers for Medicare and Medicaid Services (2004), http://www.cms.hhs.gov/medicaid/1915c/default.asp

Checkovich, T.J. and S. Stern (2002), "Shared Caregiving Responsibilities of Adult Children with Elderly Parents", *Journal of Human Resources* **37**, 441–478.

[23] In the US, examples of such an initiative are state programs under the Home and Community-Based Waivers that either condition eligibility for services on the presence of an "active" family caregiver or provide more, rather than less, hours of paid assistance to elderly persons who coreside with a family caregiver (Centers for Medicare and Medicaid Services 2004).

Chiappori, P.A. (1988), ''Rational Household Labor Supply'', *Econometrica* **56**, 63–70.

Chiappori, P.A. (1992), ''Collective Labor Supply and Welfare'', *Journal of Political Economy* **100**, 437–467.

Crimmins, E.M. and D. Ingegneri (1990), ''Interactions and Living Arrangements of Older Parents and their Children'', *Research on Aging* **12**, 3–35.

Davis, M.A., D.J. Moritz and J. Neuhaus, et al. (1997), ''Living Arrangements, Changes in Living Arrangements and Survival Among Community Dwelling Older Adults'', *American Journal of Public Health* **87**, 371–377.

Dostie, B. and P.T. Leger (2005), ''The Living Arrangement Dynamics of Sick, Elderly Individuals'', *The Journal of Human Resources* **40**, 989–1014.

Dwyer, J.W. and R.T. Coward (1991), ''A Multivariate Comparison of the Involvement of Adult Sons Versus Adult Daughters in the Care of Impaired Adults'', *Journal of Gerontology: Social Sciences* **46**, S259–269.

Ellwood, D. and T. Kane (1990), ''The American Way of Aging: An Event History Analysis'', in D. Wise, ed., *Issues in the Economics of Aging*, University of Chicago Press, Chicago, pp. 121–147.

Engers, M. and S. Stern (2002), ''Long-term Care and Family Bargaining'', *International Economic Review* **43**, 1–44.

England, P. and N. Folbre (2003), ''Contracting for Care'', in M. Ferber and J. Nelson, eds., *Ten Years Beyond Economic Man*, University of Chicago Press, Chicago, pp. 61–79.

Ettner, S.L. (1995), ''The Impact of Parent Care on Female Labor Supply Decisions'', *Demography* **32**, 63–80.

Ettner, S.L. (1996), ''The Opportunity Costs of Elder Care'', *Journal of Human Resources* **31**, 189–205.

Freedman, V.A. (1996), ''Family Structure and the Risk of Nursing Home Admission'', *Journal of Gerontology: Social Sciences* **51B**, S61–69.

Garber, A. and T. MaCurdy (1990), ''Predicting Nursing Home Utilization Among High Risk Elderly'', in D. Wise, ed., *Issues in the Economics of Aging*, University of Chicago Press, Chicago, pp. 173–204.

Gibson, M.J., S. Gregory and S. Pandya (2003), ''Long-Term Care in Developed Nations: A Brief Overview'' AARP Public Policy Institute Research Report 2003–13.

Gugl, E. (2004) ''Joint vs. Individual Taxation and Intrafamily Distribution'', University of Victoria Working Paper.

Hayashi, F. (1995), ''Is the Japanese Extended Family Altruistically Linked? A Test Based on Engel Curves'', *Journal of Political Economy* **103**, 661–674.

Heidemann, B. and S. Stern (1999), ''Strategic Play Among Family Members When Making Long-Term Care Decisions'', *Journal of Economic Behavior and Organization* **40**, 29–57.

Kalai, E. and M. Smorodinsky (1975), ''Other Solutions to Nash's Bargaining Problem'', *Econometrica* **43**, 513–518.

Kemper, P. and L.E. Pezzin (1996), ''The Effect of Public Provision of Home Care on Living and Care Arrangements: Evidence from the Channeling Experiment'', in R. Eisen and F. Sloan, eds., *Alternatives for Ensuring Long-Term Care*, Kluwer Press, Amsterdam, pp. 125–146.

Konrad, K., H. Kunemund, K.E. Lommerud and J.R. Robledo (2002), ''Geography of the Family'', *American Economic Review* **92**, 981–998.

Kotlikoff, L. and J. Morris (1990), ''Why Don't the Elderly Live with Their Children? A New Look'', in D. Wise, ed., *Issues in the Economics of Aging*, University of Chicago Press, Chicago, pp. 149–169.

Kreps, D.M. (1990), *A Course in Microeconomic Theory*, Princeton University Press, Princeton.

Laferrere, A. and F.-C. Wolff (2006), ''Microeconomic Models of Family Transfers'', in S. Kolm and J. Mercier-Ytier, eds., *Handbook on the Economics of Giving, Reciprocity and Alturism*, North Holland, Amsterdam.

Lundberg, S. and R.A. Pollak (1993), ''Separate Spheres Bargaining and the Marriage Market'', *Journal of Political Economy* **101**, 988–1010.

Lundberg, S. and R.A. Pollak (1994), ''Noncooperative Bargaining Models of Marriage'', *American Economic Review* **84**, 132–137.

Lundberg, S. and R.A. Pollak (2003), ''Efficiency in Marriage'', *Review of Economics of the Household* **1**, 153–168.

Lundberg, S. J., R.A. Pollak and T.J. Wales (1997), ''Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit'', *Journal of Human Resources* **32**, 463–480.

Manser, M. and M. Brown (1980), ''Marriage and Household Decision Making'', *International Economic Review* **21**, 31–44.

McElroy, M. and M.J. Horney (1981), ''Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand'', *International Economic Review* **22**, 333–349.

McGarry, K. (2006), "Inheritances and Bequests", forthcoming in *The Palgrave Dictionary of Economics*.

McKelvey, R. and T. Palfrey (1992), "An Experimental Study of the Centipede Game", *Econometrica* **60**, 803–836.

Moon, M. (1983), "The Role of the Family in the Economic Well-Bring of the Elderly", *The Gerontologist* **23**, 45–50.

Norton, E. (2000), "Long-Term Care", in A. Culyer and J. Newhouse, eds., *Handbook of Health Economics,* Vol. 1B, North-Holland, Amsterdam, pp. 955–994.

Perozek, M. (1998), "Comment: A Reexamination of the Strategic Bequest Motive", *Journal of Political Economy* **106**, 423–445.

Pezzin, L.E. and B.S. Schone (1997), "The Allocation of Resources in Intergenerational Households: Adult Children and Their Elderly Parents", *American Economic Review* **87**, 460–464.

Pezzin, L.E. and B.S. Schone (1999a), "Intergenerational Household Formation, Female Labor Supply, and Informal Caregiving: A Bargaining Approach", *Journal of Human Resources* **34**, 475–503.

Pezzin, L.E. and B.S. Schone (1999b), "Parental Marital Disruption and Intergenerational Transfers: An Analysis of Lone Elderly Parents and their Adult Children", *Demography* **36**, 287–298.

Pezzin, L.E. and B.S. Schone (2002a), "Intergenerational Transfers of Time and Elderly Living Arrangements: A Bargaining Model of Family Resource Allocation Decisions", Working paper, AHRQ, Rockville, MD.

Pezzin, L.E. and B.S. Schone (2002b), "Exploring the Theoretical Foundations of Altruism: An Analysis of the Effects of Divorce and Remarriage on Intergenerational Living and Care Arrangements", in P.K. Chopra and B.N. Ghosh, eds., *Gender and Development*, Wisdom House Press, UK, pp. 93–118.

Pezzin, L.E., R.A. Pollak and B.S. Schone (2005a), "Bargaining Power and Intergenerational Coresidence: Adult Children and Their Disabled Elderly Parents", Working paper, AHRQ, Rockville, MD.

Pezzin, L.E., R.A. Pollak and B.S. Schone (2005b), "Long-Term Care of the Disabled Elderly: Do Children Increase Caregiving by Spouses?", Working paper AHRQ, Rockville, MD.

Pezzin, L.E. and B.S. Schone (2006), "Examining Motives for Intergenerational Transfers: A New Test Based on Siblings' Behavior", Working paper, AHRQ, Rockville, MD.

Pezzin, L.E., R.A. Pollak and B. Schone (2006a), ''Marital Disruption, Step Children, and Transfers to the Elderly'', *Swiss Journal of Economics and Statistics* **S-16**, 103–106.

Pezzin, L.E., R.A. Pollak, and B. Schone (2006b), ''Marital History, Family Structure, and Transfers to the Elderly'', Washington University Working Paper.

Pollak, R.A. (1988), ''Tied Transfers and Paternalistic Preferences'', *American Economic Review* **78**, 240–244.

Pollak, R.A. (2003), ''Gary Becker's Contributions to Family and Household Economics'', *Review of Economics of the Household* **1**, 111–141.

Shubik, M. (1989), ''Cooperative Games'', in J. Eatwell, M. Milgate and P. Newman, eds., *The New Palgrave, Game Theory*, W.W. Norton, New York, pp. 103–107.

Spillman, B. and L.E. Pezzin (2000), ''Potential and Active Family Caregivers: Changing Networks and the 'Sandwich' Generation'', *Milbank Quarterly* **78**, 347–374.

Stern, S. (1994), ''Two Dynamic Discrete Choice Estimation Problems and Simulation Method Solutions'', *Review of Economics and Statistics* **96**, 695–102.

Stern, S. (1995), ''Estimating Family Long-Term Care Decisions in the Presence of Endogenous Child Characteristics'', *Journal of Human Resources* **30**, 551–580.

US Census Bureau (2005), *Statistical Abstract of the United States*, US Department of Commerce, Washington, DC.

United Nations (2005), *Living Arrangements of Older Persons Around the World*, Department of Economic and Social Affairs, Population Division, United Nations, New York, NY, USA.

Weiss, Y. and R.J. Willis (1985), ''Children as Collective Goods in Divorce Settlements'', *Journal of Labor Economics* **3**, 268–292.

Wolf, D. and B. Soldo (1994), ''Married Women's Allocation of Time to Employment and Care of Elderly Parents'', *Journal of Human Resources* **29**, 1259–1276.

CESifo Economic Studies, Vol. 53, 1/2007, 97–114, doi:10.1093/cesifo/ifm005
Advance Access publication 21 February 2007

Being on Sick Leave: Possible Explanations for Differences of Sick-leave Days Across Countries

Rigmar Osterkamp and Oliver Röhn*

Abstract

Sick-leave days differ widely among industrialised countries. For the US it is 5, for Sweden 20 and for Poland 26 days per year and per employee. The possible causes for these differences have apparently not been systematically analysed. Two groups of contributing factors are considered: (i) objective causes, like the general health situation, employment of women and older persons, and (ii) behavioural reactions (a) to macroeconomic conditions, like unemployment or the possibility to work outside the official labour market, and (b) to the design of institutions, like the generosity of granting sick leave or the strictness of employment protection. On the basis of a panel for 20 countries and for the years 1996–2002, it is econometrically shown that the main explanatory factors are the generosity of granting sick leave, the strictness of employment protection and the employment of older persons. The unemployment rate and the employment of women—contrary to the result of some single-country studies—do not contribute to the explanation of sick-leave differences between countries. (JEL classification: I12, I18)

1 Introduction

There are large differences in sickness absence across industrialised countries (Figure 1).[1] The difference between the country with the lowest level of days absent due to illness (USA) and the one with the highest level (Poland) is more than fivefold. It is also striking that the Eastern European transition countries (primarily Poland, Slovakia and Czech Republic, but also Hungary) are all among the countries with the highest number of absence days. This article seeks to explain these differences.

Section 2 describes what the literature has to say to that question and what it does not say. The annex contains an annotated list of the literature. The next section discusses possible explanatory variables in the light of the literature. Section 4 contains the econometric analysis.

* Ifo Institute for Economic Research, Munich, Germany. e-mail: osterkamp@ifo.de; roehn@ifo.de

We are grateful for the critical comments of two anonymous referees.

[1] Figure 1 shows average values for the period 1996–2002. Within this time span there are fluctuations but a clear trend within a country and/or diverging or converging trends across countries cannot be observed. However, in a longer time horizon some countries exhibit rising or declining trends. Yet, due to data restrictions on the explanatory side we do not focus on longer-term analyses.

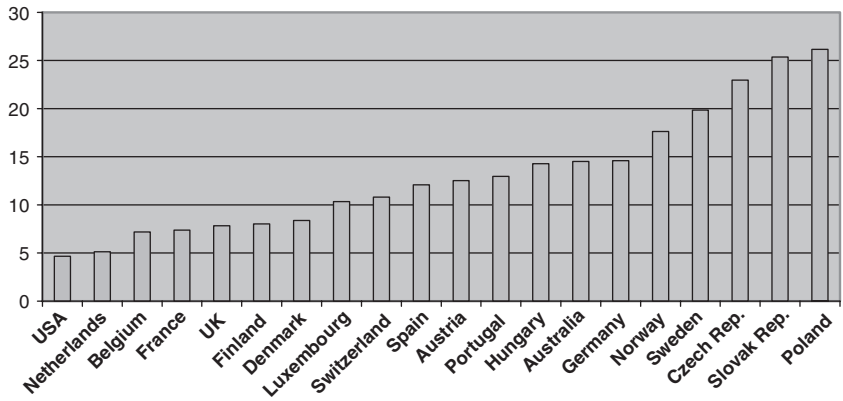


Figure 1 Sick-leave days per year and per employee, 1996–2002
Source: OECD Health Data, 2005

Section 5 summarizes the results. Section 6 relates the results to the political task of reducing overly high rates of absenteeism and provides some ideas for further research.

2 The literature and open questions

There is a growing literature on sickness absence. It can be classified into country studies, international comparisons and single topic studies. The Annex Table gives an annotated overview. Most studies (at least most of the studies listed) are by economists, but also sociologists and psychologists are engaged in the debate. All studies, with one exception, are of an empirical nature, albeit using methods of quite different degree of sophistication. The exception is the study of Holmlund (2004), which presents a theoretical model of employee behaviour under specific benefit schemes applied to different labour force states (at work, on sick-leave, unemployed). The majority of studies focus on single countries, sometimes with a look to other countries.

Compared to single country studies, there is a much lower number of international comparisons. An even smaller number covers many countries, *and* tries to assess the determinants of differences in sickness absence. One recent example for the latter type of studies is the investigation by Frick and Malo (2005).

The existing literature does mention a number of possible determinants of sick-leave differences across countries. These determinants will be incorporated into this analysis. At first one might consider the

objectively given health situation of the population as a determining effect on sickness absence. But there seems to be no study in which this is suggested to be the case.[2] In many studies the determining effect of the *unemployment rate* (specifically: of its change) has been established. The employment of *women and older persons* also seems to contribute to a higher level of sickness absence. Many studies stress the importance of *institutional regulations* that determine the *degree of generosity* to which sick-leave absence is granted or made possible. But there is no study, to our knowledge, which establishes systematically a relation between a measure of generosity to the amount of sickness absence in a country-comparative context.[3]

One possible (co-)determining factor for sickness absence, which especially relates to the high number of absence days in the transition countries, is, to our knowledge, never mentioned in the literature. The authors of this article have some anecdotal and personal evidence that workers of Eastern European countries earn income in the unofficial labour market of neighbouring (and higher income) countries while reporting at home to be sick. This hypothesis, too, is tested.

3 Explanatory variables

The aim is to explain the differences of sickness absence across countries. The potential explaining variables are taken from the aforementioned literature and from further considerations of our own. The variables can be grouped in *objective causes* and *behavioural reactions*.

3.1 Objective causes

Differences in the *general health condition of the population* might, at a first glance, be able to explain different sickness absence rates. However, it is not very plausible that the general health conditions in industrialised countries vary so much that the large differences in sickness absence rates could be explained. Moreover, the ''objectively given general health of a population'' is difficult if not impossible to measure. Single measures like mortality rates, longevity or quality-adjusted life-years (QALYs) do not seem to cover adequately what is meant by ''general health of a population''. Alternatively, there do exist individual,

[2] This could be different if countries of largely diverging per-capita income and health-care provision levels were analysed. But all studies mentioned in the Annex Table relate to industrial countries.

[3] With the exception of a preliminary study by Osterkamp (2002).

self-reported judgements of own health in cross-country studies (e.g. Banks et al. 2004). But the results are difficult to assess. It has been found that more than half of the differences in the self-reports are based on response scale differences (Banks et al. 2004). Thus, this factor has not been included in the list of potential explaining variables.

Research at the enterprise level (e.g. Moreau et al. 2004, Barmby et al. 2000) shows that in most cases women and older persons are more often absent due to sickness than men or younger persons (by about 10 percent in both cases). *Female and old-age participation rates* have, therefore, been incorporated in the study.

3.2 Behavioural reactions

The *unemployment rate* and its change is the most intensively studied single factor for explaining different (and changing) absence rates. In most countries there is a clear pro-cyclical behaviour of sickness absence. This relation is specifically pronounced e.g. in Germany, Norway and Sweden. For Norway, Askildsen, Bratbera and Nilsan (2002) have shown that the pro-cyclical behaviour of sickness absence cannot be explained by a composition effect (changing age and health structure of an enterprise's employees during the cycle) but must be due to a disciplining effect of the cycle. The level of unemployment, thus, has been included in the research.

Sickness absence (but also unemployment) can be regarded as the result of a worker's conscious decision. Such a view relates to the efficiency wage hypothesis (for an early analysis, see e.g. Yellen 1984). Monitoring the effort and output of workers is costly and, thus, incomplete. This enables workers to shirk. The usual assumption is that the level of shirking depends on the opportunity costs faced by the worker. The opportunity costs for reporting to be sick are summed up here by seven institutional factors that characterise the regulatory environment in which the worker has to make his sickness absence decision:

- Is there a waiting period (and of how many days), after which sick-leave pay starts?
- Is there the possibility of self-certification for being sick (and for how many days)?
- The official sickness certificate is issued by whom—by the patient's own doctor or by an independent examining doctor who works on behalf of the employer or the sickness fund?
- In case of sickness absence, how long does the employer continue to pay the salary, and is there any reduction? (two variables)
- In case of sickness absence, how long does the sickness fund continue to pay the salary, and is there any reduction? (two variables)

These factors are condensed into one measure of opportunity costs, which is called *measure of generosity of granting sick leave*. For the first three factors it is straightforward to integrate them into the measure of generosity. A longer waiting period reduces generosity, more self-certification days increase it. The issuance of the official sickness certificate by a doctor of the *patient's confidence* is of high generosity,[4] while the issuance by a doctor of the *employer's or the sickness fund's confidence* is of lower generosity. The latter two cases mean that there is an external proof. Also straightforward is the amount (percentage) of continuation of payment, either by the employer or by the sickness fund. The higher the amount, the higher the generosity. With respect to the question of who pays—the employer or the sickness fund—we assume that the longer the employer pays and the later the sickness fund steps in, the lower the generosity is. The reason is that the employer–employee relation is less distant and more personal than the relation between sickness fund and employee. Moreover, the sickness fund draws on anonymous funds, while the employer uses his own money for paying sick leave. Thus, the latter has better instruments and more incentives to monitor an employee going or being on sick-leave.

If a sickness is not too serious—all the more if there is no sickness at all—an employee has an advantage from being on sick leave. He (she) avoids the disutility of work, can perform simple tasks at home or can even offer working hours on the black labour market, while, at the same time, receiving sick-leave pay. The major enabling factor for this behaviour is the practice and generosity of granting sick-leave. The driving factor of such behaviour is the size of the opportunity income (of various forms) that can be earned this way. In this study only one form of opportunity income is taken into consideration, namely the income that can be earned from engaging in the labour market of a neighbour country with higher per capita income.

4 Econometric analysis

4.1 The data

To test our hypothesis that the level of generosity of granting sick leave exhibits an impact on the number of sick leave days empirically,

[4] Privately practicing doctors compete for patients. Issuance of sickness certificates is one of their important instruments to meet the competition. To our knowledge, there is only one systematic survey that assesses the behaviour of doctors in issuing sickness certificates (Hussey et al., 2004, for general practitioners in Scotland). The study reports that most doctors are neither willing nor do they feel able to differentiate between really sick patients and malingerers. The study concludes that there "appeared to be important deliberate misuse of the system by general practitioners"—and, we could add, by patients.

we construct a panel of 20 OECD countries for the years 1996–2002. Our dependent variable is the number of sick leave days per employee and year. An alternative would have been the (probability of) sickness spells. It would have been interesting to also look into differences of short- and long-term absences. However, sufficient data for a panel analysis in both cases are not available. Data on sick leave absence are provided by the OECD (Health Database) and the WHO (Health For All Database). As a first source we use the OECD data. In some cases data points were missing for single years. In these cases we added information from the WHO database.

Using partly two sources of data for the dependent variable is by no means unproblematic. However, information on overlapping observations shows that the two sources differ in detail, but show similar patterns. The meaning of sickness absence has been unambiguously defined by the OECD and the source of the data is carefully reported. Nevertheless, there remain some doubts whether all reporting countries have stuck to this definition. Thus, the following calculations must be looked at under the proviso that the comparability of the data for our central variable to be explained is at least sufficient. Even more precaution is appropriate with respect to our main explanatory variable, the generosity index, which is explained subsequently. The quantitative assessment of its seven elements is unavoidably tainted by personal judgements.

To investigate the effects of a country's generosity of granting sick leave we aggregate the seven different measures of generosity discussed in the previous section to form a single composite index. To construct the index of generosity we normalised each measure to lie within a range of 0 and 1.[5] Each single index is coded in such a way that the higher the value of the sub-index the higher the level of generosity, according to our reasoning in the previous section. The aggregate index of generosity is then simply the sum of the seven sub-indices. Thus, the aggregate index of generosity can take values between 0 and 7, where a value of 7 indicates the highest level of generosity.

Information on the institutional factors has been gathered from several sources for the time period 1996–2002. The main source was the database *Social Security throughout the World* (2004) and the database *Reformmonitor* (2004) of the German Bertelsmann Foundation. We do not possess information on the institutional factors on an annual basis. We assume that the institutional factors

[5] We calculate the indices according to the following formula: Observed value/Max (all values).

Table 1 Generosity index

Country	Generosity Index (0/7 Scale)
Australia	4.10
Austria	5.46
Belgium	4.38
Canada	3.52
Czech Republic	5.15
Denmark	5.40
Finland	2.60
France	5.24
Germany	6.11
Hungary	4.75
Netherlands	3.40
Norway	6.43
Poland	4.58
Portugal	4.75
Slovak Republic	5.00
Spain	4.75
Sweden	6.73
Switzerland	5.09
UK	3.87
USA	2.70

Notes: Generosity Index is the unweighted sum of seven variables: Waiting period, Self-certification, Total maximum duration of payment, Employer maximum duration of payment, Employer amount of payment, Sickness fund amount of payment, External proof.

are valid for the whole time period under study and, hence, our generosity index is time invariant. Since institutional arrangements are usually rigid, this assumption seems reasonable. Table 1 lists each country's index score.

Our set of control variables consists of the following. First, to control for the general health condition of the work-force we include female and old age participation rates. Old age participation is measured as the percentage of 55–64 year olds employed to the total number employed. Second, to control for behavioural reactions other than the level of generosity we include the standardised unemployment rate and a variable measuring the strictness of employment protection from the OECD. The strictness of employment protection is measured on a scale from 0 to 6, where a score of 6 indicates the highest level of employment protection. Additionally, to proxy for opportunity costs we include a country's GDP per capita difference to the

Table 2 Summary statistics

	No. of Observations	Mean	Standard deviation	Minimum	Maximum
Sick leave days	140	12.55	6.70	4.4	29.1
Generosity index	140	4.70	1.09	2.60	6.73
Unemployment rate	140	7.67	3.90	2.20	19.90
Employment of the elderly	140	42.33	14.03	16.56	68.44
Employment of women	140	60.30	9.12	34.07	75.14
Strictness of employment protection	140	2.21	0.99	0.20	4.30
Agricultural employment share	140	5.62	4.13	1.39	22.10
Industrial employment share	140	28.26	5.65	20.34	41.99
Neighbour GDP p.c. above own GDP p.c.	140	0.37	0.53	0	1.76

neighbouring country.[6] Finally, we also control for structural differences in employment and production by including the share of employment in goods-producing industries and in the agricultural sector. The intuition is that employees in the agricultural or manufacturing sector are more likely to be accident-prone and subject to bad working conditions than employees in the service sectors.[7] The data source for all control variables is the OECD. Table 2 provides summary statistics for all variables.

4.2 Estimation approach and results

In order to empirically test the hypothesis that the level of generosity exhibits an impact on the days of absenteeism due to sickness, we estimate the following equation:

$$s_{i,t} = \alpha + \beta I_i + \gamma X_{i,t} + \mu_i + \varepsilon_{i,t} \quad (1)$$

where $s_{i,t}$ is the number of sick leave days per employee in country i and year t, I_i is our index of generosity and $X_{i,t}$ is a matrix of control variables. Our standard set of control variables includes the standardised

[6] If a country has several neighbouring countries, only the difference to the neighbour with the highest GDP per capita is considered. If a country has no neighbouring country or all neighbouring countries have lower GDP per capita, this variable is set to zero.

[7] We thank an anonymous referee for suggesting these control variables.

unemployment rate, the fraction of the labour force aged 55–64 years, the female employment rate, the strictness of employment protection, the industry employment share and the agricultural employment share. Additionally, we include in this set of control variables a measure of income differences between neighbouring countries. μ_i is a country specific effect.[8]

Since our variable of main interest, the generosity index I_i, is time invariant, we estimate equation (1) using the random effects (RE) estimator. To produce consistent coefficient estimates, the RE estimator assumes that the country-specific effect μ_i is uncorrelated with any of the explanatory variables. This is a very strict assumption and we test this assumption by performing a Hausman (1978) test in all specifications. The Hausman test tests the null hypothesis that the coefficients of the RE estimator and the fixed effects (FE) estimator are *not* systematically different. The FE estimator produces consistent coefficient estimates even in the case of correlation between the country effects and the explanatory variables. Rejection of the null hypothesis would imply that we have to reject the assumption of no correlation between the country specific effects and the explanatory variables.

Table 3 displays our regression results. Column 1 shows the results of our baseline specification. We find a strongly positive and significant effect of the generosity index on the level of sick leave days. The coefficient indicates that an increase of the index of generosity by one point, i.e. for example from the level of generosity of Switzerland to the level of generosity of Germany, would increase the number of sick leave days per employee and year by about two days. We also find a significantly positive effect of the strictness of employment protection and GDP per capita differences on sick leave days. The latter result would lend support to our hypothesis that income differentials can be interpreted as opportunity costs and thus provide an incentive to call in sick and supply labour in the informal sector of the (higher income) neighbour state. However, the null hypothesis of the Hausman test is rejected, and thereby raises doubt about the RE assumption that country-specific effects are uncorrelated with the explanatory variables.

Inspection of Figure 1 reveals that the Eastern European countries rank at the top in the level of sick leave days in the sample. To control for the special effects of the transition countries, we include a dummy variable for Eastern European countries in our second specification (column 2).

[8] We performed the Breusch–Pagan Lagrange multiplier (LM) test to test for the presence of country specific effects. In all specifications the test rejected the null hypothesis of the variance of μ_i being zero. Hence, we include a country specific effect in our estimation equation.

Table 3 Regression results

	Dependent variable: sick-leave days per employee and year		
	(1)	(2)	(3)
Generosity index, (0,7) Scale	2.07**	2.61***	2.62***
	(0.81)	(0.64)	(0.65)
Unemployment rate	0.04	0.09	0.09
	(0.13)	(0.11)	(0.11)
Employment of elderly	0.12	0.13*	0.12*
	(0.08)	(0.07)	(0.07)
Employment of women	−0.04	−0.02	−0.02
	(0.13)	(0.11)	(0.11)
Strictness of employment protection, (0,6) Scale	1.93**	1.67**	1.60**
	(0.89)	(0.71)	(0.77)
Neighbour GDP p.c. above own GDP p.c.	7.87***	−3.05	−2.91
	(1.84)	(2.63)	(2.71)
Agricultural employment share	−0.25	0.03	0.04
	(0.19)	(0.16)	(0.18)
Industrial employment share	−0.09	−0.22*	−0.19
	(0.14)	(0.12)	(0.14)
Eastern European dummy		17.82***	17.33***
		(3.40)	(3.79)
Constant	−3.37	−4.77	−4.91
	(8.82)	(7.27)	(7.45)
R^{2a}	75.03%	82.77%	83.20%
Time dummies	No	No	Yes
Hausman test: H0 rejected[b]	Yes[c]	No[d]	No[d]
Number of countries	20	20	20
Number of observations	140	140	140

***significant at 1% level, **significant at 5% level, *significant at 10% level.

[a] *R*-square relates to overall *R*-square.

[b] Null hypothesis of Hausman test: coefficients of random effects estimator and fixed effects estimator are *not* systematically different.

[c] Null hypothesis of Hausman test can be rejected at the 5% level of significance.

[d] Null hypothesis of Hausman test cannot be rejected at the 5% level of significance.

The coefficient of the generosity variable increases to 2.6 and is now significant even at the 1 percent level. The coefficient of employment protection also stays positive and significant. However, we find that the effect of per capita income differences turns negative, albeit insignificant, and the Eastern European dummy is positive and strongly significant.

This result suggests that the income differential variable simply picks up the special role of the transition countries in our sample without adding any additional explanatory power. This might be due to the fact the GDP per capita differences are too crude a measure of opportunity costs. Additionally, we find a weakly significant positive effect of the old age employment share and a counterintuitive negative effect of the industry employment share. In this specification, the Hausman test cannot be rejected which supports the validity of the RE estimator.

Finally, we control for time-dependent shocks that are common to all countries by including time dummies (column 3). The results are virtually unchanged compared to our second specification, except that the coefficient of the industry employment share is no longer significant. Again the null hypothesis of the Hausman test cannot be rejected.

5 Summary

We find that institutional variables are an important factor in explaining differences in days of absenteeism due to sickness across countries. Both the level of generosity of granting sick leave as well as the strictness of employment protection are positively and significantly related to the number of sick leave days in our regressions. Further, we find some support of a positive relationship between the labour force participation rate of the 55–64 year olds. One of our hypotheses, namely that the opportunity to work in neighbouring high income countries while reporting sick at home, could not be proven convincingly. Contrary to some results of single country studies, we do not find a significant effect of the unemployment rate or the female participation rates on the level of sick leave days.

6 Policy conclusions and possible avenues for further research

The national discussion in many countries on how to reduce sickness absence seems to be determined by the idea that *working makes you sick*. Accordingly, the linchpin of redressing excessive absenteeism (as far as this is the case) is often seen in requiring changes of the working conditions, above all: less repetitive work and more moral and social encouragement of the employees. While such factors may be at work, our analysis suggests that the generosity of granting sick leave plays an important role in explaining sick-leave days.

For further research it would be desirable, first, to scrutinise the sick-leave data as well as the factors of generosity on the basis of a detailed country-by-country survey. A review of individual country sources might also lead to an enlargement of the number of countries incorporated in the analysis.

Second, we have assumed that the institutional factors that were summed in the index of generosity for granting sick leave are time invariant. An in-depth assessment, however, may reveal some change over time.

References

Aronsson, G., K. Gustafsson and M. Dallner (2000), ''Sick but yet at work. An empirical study of sickness presenteeism'', *Journal of Epidemiology and Community Health* **54**, 502–509.

Askildsen, J.E., E. Bratberg and Ö.A. Nilsen (2002), ''Unemployment, Labour Force Composition and Sickness Absence: A Panel Data Study'', IZA Discussion Paper 466.

Banks, J., A. Kapteyn, J.P. Smith and A.V. Soest (2004), ''International Comparisons of Work Disability'', IZA Discussion Paper 1118.

Barmby, T.A., M.G. Ercolani and J.G. Treble (2002), ''Sickness Absence: An International Comparison'', *The Economic Journal*, **112(480)**, F315–F331.

Beatty, C., S. Fothergill and R. MacMillan (2000), ''A Theory of Employment, Unemployment and Sickness'', *Regional Studies*, **34**, 617–630.

Bengtsson, T. and K. Scott (2002), ''Immigrant Consumption of Sickness Benefits in Sweden, 1981–1991'', Working Paper, Department of Economic History, Lund University.

Bertelsmann Foundation (2004), *Reformmonitor*.

Biffl, G. (2002), ''Der Krankenstand als wichtiger Arbeitsmarktindikator'' [Sick leave as an important labour market indicator], *WIFO-Monatsberichte* **1**, 39–52.

Boss, A. (1999), ''Lohnfortzahlung und Krankenstand'' [Sick-leave payment and sickness absence], *Zeitschrift für Wirtschaftspolitik* **49(1)**, 27–50.

Campioletti, M. and J.N. Lavis (2000), ''Disability Expenditures in Canada, 1970–1996: Trends, Reform Efforts and a Path for the Future'', *Canadian Public Policy – Analyse de Politiques*, XXVI, 2.

Ercolani, M. (2000), ''A Simple Empirical Model of Sickness Absence Appkied to UK Survey Data'', Working Paper, Institute for Labour Research, University of Essex.

European Foundation for the Improvement of Living and Working Conditions (1997), *Preventing Absenteeism at the Workplace*, Luxembourg.

Eurostat (2004), *Work and Health in the EU – A Statistical Portrait, 1994–2002*, 2003 edition.

Frick, B. and M.A. Malo (2005), "Labour Market Institutions and Individual Absenteeism in the European Union", Mimeo, Faculty of Management and Economics, Witten/Herdecke University.

Danish Ministry of Employment (2003), "*Analyse af det danske sygefravaer*" [Analysis of Danish sickness absence], Copenhagen.

Gimeno, D., F.G. Benavides, J. Benach and B.C. Amick III (2004), "Distribution of sickness absence in the European Union", *Occupational and Environmental Medicine* **61**, 867–869.

Hausman, J. (1978), "Specification Tests in Econometrics", *Econometrica* **46**, 1251–1271.

Holmlund, B. (2004), "Sickness Absence and Search Unemployment", CESifo Working Paper 1227.

Hussey, S., P. Hoddinott and Ph. Wilson et al. (2004), "Sickness certification system in the United Kingdom: qualitative study of views of general practitioners in Scotland", *British Medical Journal* **328**, 88–92.

Ichino, A. and R.T. Riphahn (2001), "The Effect of Employment Protection on Worker Effort: A Comparison of Absenteeism During and After Probation", IZA Discussion Paper 385 and CESifo Working Paper 596.

Jensen, Ch., A. Aronsson, O. Björnstad and H.K. Gunnarsdottir (2003), *Sygefravaer i Norden* [Sick leave in the Nordic countries], Copenhagen.

Kelly, K. and W. Nichol (1988), *Sickness Beneficiaries – Trends and Characteristics*, Department of Social Security, Canberra.

OECD (2004), *Health Database*.

Osterkamp, R. (2002), "Work Lost Due to Illness – An International Comparison", *CESifo Forum* December, 36–40.

Moreau, M., F. Valente and R. Mark, et al. (2004), "Occupational Stress and Incidence of Sick leave in the Belgian Workforce: the Belstress Study", *Journal of Epidemiology and Community Health* **58**, 507–516.

Thalmeier, A. (1999), "Bestimmungsgründe von Fehlzeiten: Welche Rolle spielt die Arbeitslosigkeit?" [Determinants of sickness absence: which role for unemployment?], IZA Discussion Paper 62.

Thornton, P. (2003), *Disability Management – Statement and Comments*, Disability Management in the Netherlands, November 24–25.

WHO (2004), *Health for all Database*.

Yellen, J. L. (1984), "Efficiency Wage Models of Unemployment", *American Economic Review, Papers and Proceedings* **74**, 200–205.

Annex Table Annotated overview on the literature on sickness absence

Author, year, author's profession	Title	Countries	Content
Single country studies			
Kelly and Nichol, 1988, (economists)	Sickness Beneficiaries—Trends and Characteristics	Australia	Strong increase of sick leave days (SLD) was accompanied by growing unemployment in 1980s (i.e. anti-cyclical development). The increase of SLD is explained by higher use of sick-leave payment instead of (less attractive) unemployment compensation.
Boss, 1999, (economist)	Lohnfortzahlung und Krankenstand (Sick-leave payment and sickness absence)	Germany	The generosity of granting sick leave plays a dominant role in explaining the level and development of SLD. Moreover, SLD behave pro-cyclically. There is a short look also at the sick- leave regulations of other countries (NL, USA, UK and Sweden).
Thalmeier, 1999, (economist)	Bestimmungsgründe von Fehlzeiten: Welche Rolle spielt die Arbeitslosigkeit? (Determinants of sickness absence: which role for unemployment?)	Germany	Main determinant of SLD is unemployment (and, thus, the cycle). SLD develop pro-cyclically. Changes of generosity have had minor effects on SLD.
Aronsson, Gustafsson and Dallner, 2000, (sociologists)	Sick but yet at work. An empirical study of sickness presenteeism	Sweden	"Presenteeism" means *not* being on sick leave although sick. The study is based on self-reported data. It finds that presenteeism is combined with working in care, welfare or teaching occupations, with low wages as well as with high sickness absence.

Campioletti and Lavis, 2000, (economists)	Disability Expenditures in Canada, 1970–1996: Trends, Reform Efforts and a Path for the Future	Canada	Description of the various relevant social support systems for disability (sickness included), of expendituretrends and of ongoing reforms. Further necessary reforms are seen in a better program coordination and benefit integration.
Ercolani, 2000, (sociologist)	A Simple Empirical Model of Sickness Absence Applied to UK Survey Data	UK	An ''Underlying Propensity for Sickness Absence'' is measured by individual socio-economic characteristics.
Askildsen, Bratberg and Nilsen, 2002, (economists)	Unemployment, Labour Force Composition and Sickness Absence: A Panel Data Study	Norway	There is a pro-cyclical development of sickness absence that can be explained by effects of the cycle on workers' discipline, but *not* by the composition effect of the cycle
Bengtsson and Scott, 2002, (economists)	Immigrant Consumption of Sickness Benefits in Sweden, 1981–1991	Sweden	Part of the high Swedish SLD figures is explained by the relatively intensive use of the sick-leave possibility by immigrants.
Biffl, 2002 (economist)	Der Krankenstand als wichtiger Arbeitsmarktindikator (Sick leave as an important indicator for the labour market)	Austria	Main determinants of SLD are seen to be the development of unemployment (i.e. SLD behave pro-cyclically) and of labour force participation rates (gender and age). The latter factor is influenced by societal developments to facilitate early retirement and disability pensions.
Danish Ministry of Employment, 2003	Analyse af det danske sygefravaer (Analysis of Danish sickness absence)	Denmark	SLD are mainly attributed to institutional conditions (generosity of granting sick leave). Short comparisons of those conditions with Sweden, Norway and UK
Thornton, 2003, (economist)	Disability Management—Statement and Comments	NL and UK	The contribution reviews critically the ongoing reform developments in the Netherlands to increase the responsibility of employers for reducing sickness absence (disability management practices) and the possible transfer of this method to the UK.

(continued)

Annex Table Continued

Author, year, author's profession	Title	Countries	Content
Hussey et al., 2004, (physicians)	Sickness certification system in the United Kingdom: qualitative study of views of general practitioners in Scotland	UK	The study concludes that sick-leave certificates are issued to an important degree in deliberate misuse of the system.
Moreau et al., 2004, (physicians and public health economists)	Occupational stress and incidence of sick leave in the Belgian workforce: the Belstress study	Belgium	Study on the enterprise level. Sickness absence is mainly attributed to strained jobs with low social support.
Single topic studies			
Beatty, Fothergill and MacMillan, 2000, (economists, geographers)	A Theory of Employment, Unemployment and Sickness	UK data as an example	The study concludes that job dismissals may result in higher recorded sickness (absence) instead of higher recorded unemployment.
Ichino and Riphahn, 2001, (economists)	The Effect of Employment Protection on Worker Effort: A Comparison of Absenteeism During and After Probation	An Italian firm as an example	The study shows that after the end of the probation period (beginning of employment protection) sickness absence more than doubles.
Holmlund, 2004, (economist)	Sickness Absence and Search Unemployment	Theoretical paper	Different labour force states are considered (employed, on sick-leave, unemployed with or without searching due to health conditions). It is assumed that the benefit structure applied for the different states influences the choice of the state. It is shown that there might be a socially optimal benefit structure with differentiated benefits across labour force states.

International comparisons European Foundation for the Improvement of Living and Working Conditions, 1997	Preventing Absenteeism at the Workplace	EU-15 + Norway	Differences in institutional regulations (generosity of granting sick leave) across countries are described but are not seen as explanatory factors for differences in SLD. The main part of the study relates to the enterprise level, presents "models of good practice" (from eight countries) and formulates nine recommendations for enterprises to reduce sickness absence.
Barmby, Ercolani and Treble, 2002, (sociologists)	Sickness Absence: An International Comparison	8 EU countries + Canada	Sickness absence is related to individual socio-economic factors (age, gender, marital status, income, sector and tenure). Moral hazard as an explanatory factor is mentioned but not analysed.
Osterkamp, 2002, (economist)	Work Lost Due to Illness – An International Comparison	18 countries	Sickness absence is related to a rough indicator of generosity of granting sick leave. To the usual measure of total health-care costs of the economy the non-production due to sickness absence is added.
Jensen et al., 2003, (sociologists, economists)	Sygefravaer I Norden (Sick leave in the Nordic countries)	DK, N, Swed., Icel.	Sickness absence is mainly seen as a result of individual health and of the type of work and conditions of the work place.
Eurostat, 2004	Work and Health in the EU – A Statistical Portrait	EU total	The publication provides commented tables and graphs. Sickness absence is only reported by sector, not by country.

(continued)

Annex Table Continued

Author, year, author's profession	Title	Countries	Content
Banks et al., 2004 (sociologists, psychologists)	International Comparisons of Work Disability	US, UK, NL	The amount of self-reported work disability differs considerably across countries. The study shows that more than half of that difference can be explained by response scale differences instead of by differences of the objectively given health status.
Gimeno et al., 2004, (public health economists)	Distribution of sickness absence in the European Union	15 EU countries	Self-reported data for sickness absence of at least one day are used. Institutional regulations as possible explaining factors for differences across countries are mentioned.
Frick and Malo, 2005 (economists)	Labour market institutions and individual absenteeism in the EU	12 EU countries	Data from the "European Survey on Working Conditions"; explaining variables as employment protection and sickness benefits are far less relevant than individual worker characteristics.

CESifo Economic Studies, Vol. 53, 1/2007, 115–152, doi:10.1093/cesifo/ifm003
Advance Access publication 26 February 2007

From Deficits to Debt and Back: Political Incentives under Numerical Fiscal Rules

Marco Buti*, João Nogueira Martins* and Alessandro Turrini†

Abstract

Under numerical fiscal rules, such as those underpinning EMU, governments have strong temptations to use accounting tricks to meet the fiscal constraints. Given these political incentives, fiscal variables that in the past were regarded as a mere residual acquire a strategic role. This is the case of the so-called stock-flow adjustment (SFA) which reconciles deficit and debt developments. We develop a simple theoretical model where deficits and two distinct SFA components (one that could be used to reduce the deficit figures and the other to impact debt figures instead) are determined as a result of a constrained optimisation by fiscal authorities. Econometric evidence provides results consistent with the model findings. The SFA component related to the purpose to hide deficits rises with the recorded deficit, while the sales of financial assets designed to keep the debt under control rise with both debt and deficit. When deficits are in excess of the 3 percent limit, accounting gimmicks become more sensitive to the size of deficits. The SGP *per se* does not appear to increase the extent to which higher deficits trigger more accounting gimmicks. However, the SGP seems associated with a more intense use of accounting gimmicks irrespective of the level of deficit. Such accounting practices have greatly contributed to the loss of credibility of Economic and Monetary Union's fiscal rules. If properly implemented, the reformed Pact, which stresses durable adjustment and long-run sustainability, should help curb such perverse incentives. (JEL codes: E61, H62, H87)

Keywords: Stability and Growth pact, government accounting, stock-flow adjustment, fiscal gimmicks

1 Introduction

Europe's Economic and Monetary Union (EMU) is built on strong fiscal discipline foundations. The budgetary autonomy of the members of the euro area is subject to the numerical constraints of the Maastricht Treaty and the Stability and Growth Pact (SGP). The Treaty prescribes that budget deficits should not exceed 3 percent of GDP, unless exceptional

* European Comission DG ECFIN, B-1049 Brussels, Belgium.
† European Comission and CEPR, European Commission, DGECFIN, B-1049 Brussels, Belgium. e-mail: alessandro@turrini.ec.europa.eu
The views expressed in this article are those of the authors and do not necessarily reflect those of the European Commission. We would like to thank an anonymous referee, Fabrizio Balassone, Mark De Broeck, Xavier Debrun, Elena Flores, Luca Onorante, Karsten Wendorff, Guntram Wolff and several participants at the 8th Banca d' Italia Workshop on Public Finance, Perugia, 30 March–1 April. We also thank V. Ernesto Reitano and Vittorio Gargaro for excellent research assistance.

circumstances occur and, even in this case, the excess should remain limited and temporary. Public debt should not exceed 60 percent of GDP or, if this is the case, it should be maintained on a downward trend. While the numerical parameters of the Treaty were seen as a screening device to select the members of the euro area, the goal of the SGP—which set medium-term objectives for EU Member States—was to make fiscal discipline a permanent feature of EMU.

Such rules triggered a strong fiscal adjustment in the run up to EMU: the average budget deficit of the euro area was reduced from a peak of 6 percent of GDP in 1993 to less than 1 percent at the turn of the century. Public debt, which registered a quasi continuous increase from about 30 percent of GDP in the mid 1970s, reached about 70 percent of GDP in mid-1990s and started to decrease, albeit very slowly, thereafter. While both variables went into reverse in the last few years under the effect of poor economic growth and "adjustment fatigue", the threat of going back to the reckless behaviour of the pre-Maastricht era has not materialised.

A dark side of EMU's fiscal rules, however, quickly emerged. Accounting tricks, one-off operations, exotic transactions and legally dubious data manipulations to circumvent the constraints on deficits and debt became frequent. The political incentives in evading real adjustment was recognised in the early days of the new Treaty: "Maastricht encourages financial engineering to avoid underlying fiscal adjustment. Even when privatisation is desirable for efficiency reasons, it is bad economic policy to do the right (structural) thing for the wrong (financing) reasons" (Buiter et al. 1993).[1]

In spite of indications that window-dressing activities associated with fiscal rules could be sizable and anecdotal evidence piled up, a major difficulty in carrying out empirical research has been the lack of systematic information. One way to overcome this difficulty is a "bottom-up" estimation obtained summing up the value of the operations, which have been identified as falling under the category of creative accounting. This is the route that followed Dafflon and Rossi (1999) and, more recently, Koen and van den Noord (2006). The latter construct measures of "fiscal gimmickry" taking into account both one-off measures improving budget balances and creative accounting operations and find that the probability

[1] This is obviously not a wholly new or EMU-specific phenomenon. Actually, most of the empirical evidence comes from US states' fiscal behaviour. Von Hagen (1991) and Bunch (1991) show that borrowing constraints imposed on US states have led to the substitution of non-constrained for constrained debt instruments. Strauch (1998) shows that expenditure ceilings at US state level have induced a shift from the constrained current budget to investment expenditure which is unconstrained. See also Bunch (1991) and Sbragia (1996). Besides fiscal policy rules, the temptation of artificially reducing the deficit and debt figures is also connected with fooling financial markets and electorates. On how financial markets react to creative accounting, see Bernoth and Wolff (2006).

of fiscal gimmickry increases with the level of deficits in EU countries. Of course, the limitation of this approach is that it cannot be exhaustive: many operations aimed at strategically manipulating the statistical measurement of deficits will not be captured.

A second option is the one followed by Easterly (1999) and Milesi-Ferretti and Moriyama (2006) who take a "balance sheet approach" to analyse fiscal adjustment.[2] If a government has not embraced fiscal adjustment in earnest (via a lowering of its discount rate), it will respond to a budgetary constraint by reducing its asset accumulation or increase its hidden liabilities by an equal amount, leaving net worth unchanged. Hence, the improvement in the fiscal balance or the reduction in the debt ratio do not reduce the need for future higher taxes. Looking at EU countries in the 1990s, Easterly (1999) notices that privatisation in the original 11 countries of the euro area after Maastricht (i.e. after 1991) more than quadrupled while it fell in the UK, Sweden and Denmark. He concludes that the combination of basically unchanged public expenditure growth, one-off measures, the proliferation of privatisation and public investment reduction suggests that at least part of the fiscal retrenchment in response to the Maastricht constraint was illusory. Milesi-Ferretti and Moriyama (2006) find that during the run-up to EMU, the change in the stock of debt is positively correlated with changes in government assets during the same period while it is weakly correlated with changes in net worth. Despite a fall in the stock of public assets, net worth deteriorated between 1992 and 1997 in almost all EU countries. By contrast, in the period 1998–2002, net worth improved notwithstanding declining government assets. The authors interpret this as due to a lower pressure to use non-structural measures arising from the lower penalties for missing the fiscal targets once in the euro area.

In this article, we choose a largely complementary approach by studying in detail the so-called stock-flow adjustment (SFA), namely the discrepancy between the accounting value of deficit and the change in debt. This option is all the more relevant in the EU context as the relative attention to the two criteria of the Treaty has changed over time and, since the introduction of the SGP, the focus has increasingly shifted on the deficit criterion which has been the only one to receive serious attention in the public debate.

In the literature on public debt accumulation, the SFA is usually disregarded or treated as a mere residual. Such a shortcut is acceptable only to the extent that the SFA is small and cancels out over time. This is what

[2] For an early suggestion of a balance sheet approach, see Buiter (1985).

one would expect from a composite residual variable made up of several items moving in opposite directions. However, if one tried to reconstruct the debt series for years 1991–2005, disregarding the SFA, the cumulated error would exceed 4 percent of GDP for the euro area as whole. More importantly, for a number of Member States, the difference is much larger and may reach 10 percent of GDP in single specific years, or approach 40 percent of GDP cumulated over the above-indicated period. The non-weighted average of yearly SFA-to-GDP ratios in the euro-area countries from 1991 to 2005 is 1.1 percent of GDP and the absolute average is 1.9 percent of GDP.[3]

Economic analysis has started only recently to pay attention to the behaviour of the SFA. In a seminal article, von Hagen and Wolff (2006), building on the theoretical model by Milesi-Ferretti (2003), show that under the SGP where greater attention is paid to the deficit, governments tended to shift expenditure to below the line thus increasing the SFA. The authors find a systematic relationship between SFA and deficit after the introduction of the SGP. Under such a fiscal rule, where the deficit criterion receives considerable more attention than the debt (or than the yearly debt developments), governments do have an incentive in under-reporting their deficits or in packaging or describing their transactions in such a way that the gap between deficit and debt widens.

However, von Hagen and Wolff (2006) use an aggregate measure of the SFA. This does not allow capturing the different political incentives in an environment in which the degree of stringency of the deficit and debt criteria may vary across time and countries, and the costs associated to SFA also vary among its components. In particular, first, there may be legitimate reasons for persistently positive and even large value of SFA: as shown by the authors themselves (see their Table 1), amongst the largest values of the SFA over the last two decades are found in Finland, Denmark and Luxembourg, all countries where the fiscal constraints did not bite (apart from a short period at the beginning of the 1990s in the two Nordic countries. Second, low total SFA—as in Italy or Portugal—may be the result of hidden expenditure offset by sales of financial assets (privatisations).

In this article, we identify distinct SFA components that are associable with accounting gimmicks aimed at embellishing the deficit and at reducing the debt. On the basis of a simple theoretical model where deficits and two distinct SFA components (one that could be used to hide part of the deficit and the other to reduce the debt figures instead)

[3] See also European Commission (2005) for an illustration of recent developments in the SFA in EU countries.

are determined as a result of a constrained optimisation by fiscal authorities; we provide econometric evidence on the strategic use of the SFA components.

The remainder of the article is structured as follows. Section 2 presents a descriptive analysis of the SFA by focussing on size, composition and the ability of fiscal authorities to strategically use them. Section 3 illustrates the basic structure and the predictions of a simple model of a government with short-term growth objectives, but with its room for manoeuvre thwarted by fiscal constraints akin to those of EMU. Section 4 provides empirical evidence on the determinants of government operations that affect the SFA in an environment in which the budget criteria of Maastricht and the SGP constrain fiscal behaviour. The final section summarises our finding, provides some considerations on how the reform of the SGP will affect such perverse political incentives and suggests further work. The Appendix presents the theoretical model in detail.

2 The stock-flow adjustment: why does it exist?

2.1 A composite entity: the main components of the SFA

This section describes the components of the SFA. It elaborates on each of the components and reports on the available data. It then discusses on their manipulability by government with the aim of painting their deficit and debt figures—especially the former—in rosier tones. We have found useful to break down the SFA into three components. Two components reflect basic differences in the accounting basis for the deficit and the debt, while the third category gathers residual adjustments.

Differences between the accrual and cash bases of recording transactions

The government expenditure and revenue are recorded on an accrual basis, that is, at the time of the underlying transaction irrespective of effective cash flows. In contrast, the debt is a cash concept; it increases or decreases with effective debt issuance or redemptions. These ultimately depend on effective cash payments and receipts, not on the underlying revenue and expenditure.

The transactions that have been recorded as expenditure or revenue—and therefore have contributed to increase or reduce the deficit but for which the effective cash payment or receipt has not yet taken place are accounts receivable or payable, i.e. differences between cash and accruals. The issuance of zero-coupon bonds, the reimbursement of bonds that do not regularly pay coupons, the accumulation of revenue arrears, the settlement of payment arrears and the payment of expenditure in advance, the reimbursement of taxes, etc. result in positive SFAs (debt increase by

more than the deficit in a specific period). Symmetrically, interest accrued by zero-coupon bonds, or by other bonds that do not regularly pay coupons, the accumulation of payment arrears, the collection of revenue in arrears, and the collection of excessive taxes that will need to be reimbursed at a later stage, etc. lead to negative SFAs.[4]

It needs stressing that the differences between cash and accruals accounting should in principle cancel out over the years. In the medium-term (let us say 5 years), the cumulated flows of accounts receivable and payable should converge to zero, or simply reflect nominal growth.

Figure 1 shows data on the cash and accrual discrepancies for each of the EU Member States. Cash and accrual discrepancies for interest and other revenue and expenditure items are shown separately. The difference between cash and accruals appear relatively small for most countries. However, the data for Greece, Italy and, to an extent, Portugal, Cyprus and Poland stand out: even over a 6-year period during which plus and minuses should cancel out, the cumulated discrepancy is positive and large. Taking into account that these countries, in particular Greece, Italy and Portugal, were under pressure to avoid excessive deficits and, the first two have a large stock of debt, the political incentives to hide budget deficits may have been at work.[5]

Differences between the net and gross recording in relation to financial transactions

A second major difference between deficit and debt accounting—in fact, the one that has the largest impact on data—concerns the net accumulation of financial assets.

The government deficit is a net concept. It is defined as government *net* borrowing, which corresponds to the balance between revenue and expenditure excluding financial transactions. In contrast with the deficit, the debt is measured in gross terms; government assets are not netted from the government liabilities when compiling the debt. When the government accumulates financial assets and, therefore, needs to finance their

[4] We have merged statistical discrepancies in the Member States' financial accounts into accounts receivable and payable. This is because experience suggests that most statistical discrepancies (that is, differences between deficit and debt figures that statisticians are unable to attribute to any specific SFA component) are the result of the intricacies of accrual accounting. Moreover, in a number of countries, statistical discrepancies between financial and non-financial accounts are not specifically identified in their accounts but merged under accounts receivable and payable.

[5] It should be noted also that, since the extraction of the data for this article in spring 2005, there have been significant revisions in the accounts of Greece, Italy and Portugal. The reclassification of items from below to above the line implied upward revisions in their deficits and led to smaller differences between cash and accrual data. This means that data initially reported by these Member States underestimated the accrual-based deficit.

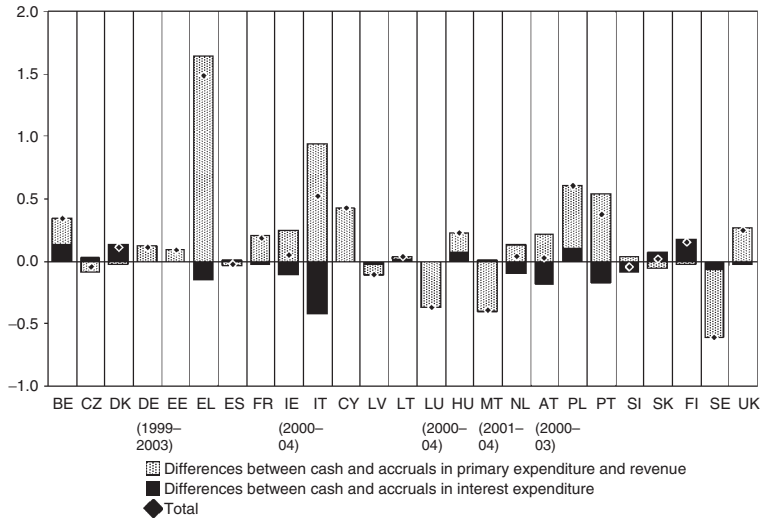


Figure 1 Time of recording: cash and accruals (% of GDP; 1999–2004, except new Member States (2000–04) or as otherwise indicated). The bars in Figure 1 correspond to the difference between cash and accrual accounting. Positive figures mean that the cash deficit is higher than the accrual-based deficit. The figure distinguishes differences between cash and accruals in the accounting of (i) primary expenditure and revenue and (ii). interest, since the latter is less under the control of government. Statistical discrepancies in government accounts appear as differences between cash and accruals in primary expenditure and revenue. Although cash and accruals figures can differ quite significantly in specific years, one would expect these differences to offset each other over a series of years. Persistent positive differences between cash and accrual accounting may suggest an underestimation of the official accrual-based deficits that are relevant for budgetary surveillance

acquisition, the gross debt increases even if the deficit and net worth would remain unchanged. Therefore, the accumulation of financial assets leads to positive SFA, and the decumulation of financial assets (e.g. privatisation) implies negative SFA.

Data available allow distinguishing the net accumulation of financial assets in four categories: liquidities (i.e. deposits with banks), loans, securities other than shares (that is, bonds issued by the private sector and traded in the stock exchanges) and shares and other equity. Data on these components are shown in Figure 2.

The Member States that have registered the largest accumulation of financial assets are those that have been in surplus and have relatively small debts, such as Denmark, Estonia, Luxembourg, Finland and Sweden. Given that the government gross debt in these countries is

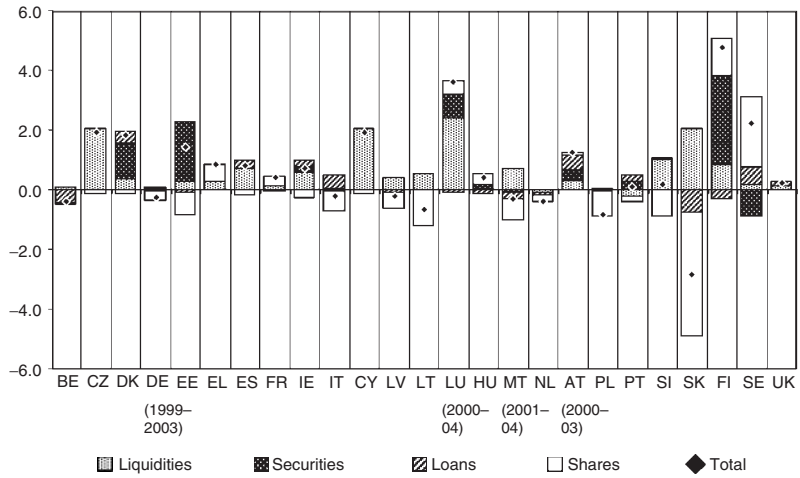


Figure 2 Accumulation of financial assets (% of GDP; 1999–2004, except new Member States (2000–04) or as otherwise indicated). The bars in Figure 2 correspond to transactions with financial assets. Positive figures mean that the government increased its holdings of the respective assets: liquidities, securities (other than shares), loans and shares. Shares cover equity in public and private enterprises, either quoted or unquoted. Large negative values (in particular of shares) often correspond to privatisation of public enterprises

low or very low, the accumulation of assets is preferable to redemption of debt.

Valuation effects and other statistical adjustments

The third component of the SFA concern valuation changes, reclassifications and other technicalities. Figure 3 shows data on this SFA category. The adjustments because of exchange rate movements are now very small in almost all Member States. They used to be much larger in several countries before the creation of the euro area.[6] Revaluations of debt can also take place in the context of early debt redemptions: two interesting cases concerned Italy (the replacement of a low-interest bond with a high interest bond and lower face value at the end of 2002) and Netherlands and Sweden (the reverse transaction in 2000). Concerning the residual ''other'', it is worth referring to large reclassifications as government debt of liabilities in connection with banking restructuring in the Czech Republic and Slovakia, which appear under ''other''.

[6] For example, the revaluation of foreign currency-denominated liabilities increased the Greek debt by 2.9 percent of GDP per year in both 1999 and 2000.

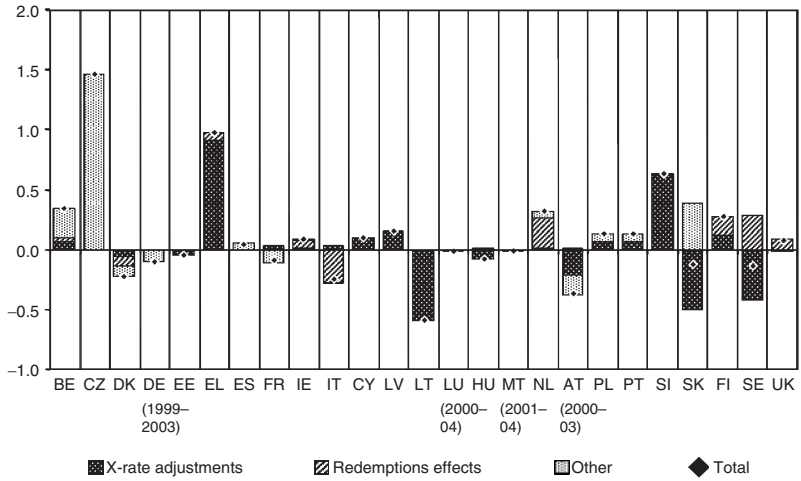


Figure 3 Valuation effects and residual adjustments (% of GDP; 1999–2004, except new Member States (2000–04) or as otherwise indicated). The bars in Figure 3 correspond to increases or reductions in government debt because of valuation effects and other adjustments. A depreciation (appreciation) of national currency leads to an increase (decrease) in the value of foreign-denominated debt expressed in national currency. The redemption of debt before maturity above (below) par leads to an increase (reduction) in the debt at nominal value and appears in the figure with a positive (negative) sign. ''Other'' includes residual adjustments, in particular reclassification of units (and respective debts and consolidating assets) within or outside government

2.2 Not so much a residual: the strategic use of the SFA

The basic question this article addresses is whether and how the SFA can be used strategically by a government that bears political costs when the deficits violate some numerical constraint, while, at the same time, it wishes to increase public spending and reduce taxes to stimulate economic activity or please the electorate.

As the previous section has shown, the SFA is the result of many different developments. A large SFA that depends predominantly on the accumulation of assets quoted in the stock exchange by a government in surplus has a considerably different nature from a large positive SFA because of the increase in the share capital of distressed public enterprises, a depreciation of national currency, because the government had to settle a large stock of spending arrears or simply because cash and accrual statistics do not match.

We propose here two alternative indicators based on SFA components that signal a strategic use by fiscal authorities to reduce deficits in an

artificial manner. To simplify language, in the remainder of the article, we will refer to the set of these SFA items as ''hidden'' deficits, ''disguised deficits'', or ''accounting gimmicks''.

A conservative measure of ''hidden'' deficits: differences between cash and accrual deficits

In an economy with liquidity restrictions, one may expect that cash receipts and payments could be of a more direct use to appease the electorate and accelerate economic activity than government revenues and expenditure. However, since the latter are those that are relevant in the EU budgetary surveillance, governments have an interest in reducing the deficit in accruals and increasing the cash deficits, by increasing the effective revenue collection lags and reducing the cash payment lags. In principle, one could expect that this strategy would only be used in specific critical moment, such as immediately before an election or in the vicinity of a commitment related with fiscal discipline frameworks like the SGP, as the difference between cash and accrual accounting is just a matter of timing. Yet, the message conveyed by Figure 1 is at variance with these considerations. The differences between accruals and cash in the accounting of revenue and expenditure for a few countries—with specific budgetary concerns—seem to be persistent. This suggests that government might also try to minimise their deficits via inconsistent cash and accrual statistics. This means that some governments have an interest in keeping low-quality statistical systems if this results in an underestimation of their deficits.

A comprehensive measure of ''hidden'' deficits: taking into account disguised government subsidies

The aforementioned measure of ''hidden'' deficits can be made more comprehensive. Indeed, there are a number of transactions in financial assets that may also take place first and foremost with the aim of hiding deficits.

Not all assets are equal. The purchase of bonds or blue-chip shares by social security investing its surpluses is not of the same nature of an injection in the share capital of a loss-making public enterprise by central government. Therefore, it would be useful to distinguish between high-quality assets and low-quality assets. The former are the financial investments which take place at market conditions and which would be accepted by a profit-maximising private investor. The latter are those which the government enters into for public policy purpose, in particular those that replace deficit-increasing subsidisation and may be determined by the wish of hiding expenditure.

Liquidities and securities can be safely classified in the first group. Loans are a less clear case. For a government which attempts to minimise its deficit, it may be preferable to grant a loan than to directly provide a subsidy. In some cases, given national budgetary rules, it may also be easier to grant loans rather than to provide direct subsidisation. Ultimately, one would have to distinguish loans granted by government according to beneficiaries' rating, and the specific conditions of each loan. However, as this information is not available, we will assume that loans granted by government do contain an element of hidden expenditure.

In relation to shares, the distinction between good and bad assets could be attempted by separating the shares which are quoted in the stock exchange and the unquoted shares, in particular in enterprises which are controlled by government. However, there are severe data-availability problem that do not allow such a distinction.[7] On the basis of the available data, we found useful to consider three categories, namely, transactions in shares by social security (which we presume to correspond to a profit-maximising behaviour of investing surpluses in high-quality assets), privatisation[8] (that is, large sales of government-owned shares, which ultimately lead to shift the control of a public firm to private hands) and other transactions in shares (where we assume on the basis of anecdotal evidence that the accumulation of low-quality shares, hiding subsidisation, predominates). Therefore, a comprehensive measure of "hidden" deficits corresponds to the conservative measure plus transactions in shares other than by social security and not related to privatisation.

Figure 4 shows data on this comprehensive measure of "hidden" deficit. As for the previous figures, data are shown as 6-year (1999–2004) averages. According to this indicator, the countries that are more likely to have used the SFA to reduce their deficits are Greece, Italy, Austria, Hungary and Portugal.

In both approaches to identify "hidden" deficits, we have disregarded valuation effects and other statistical adjustments, given their heterogeneous and erratic nature. Note, however, that in some circumstances, Member States may also put in place a number of deficit-minimising strategies, which would be reflected in this category. A case in point is the issuance of foreign debt. Assuming covered interest parity, it should be invariant to issue debt in low-yield currencies for which there is an expectation of appreciation or in high-yield currencies for which there is

[7] The ESA95 rules and nomenclature do distinguish between quoted and unquoted shares, and according to who controls the enterprise. However, most countries do not publish these data.

[8] We have considered privatisation separately to avoid that data on privatisation and on capital injections in public enterprises offset each other.

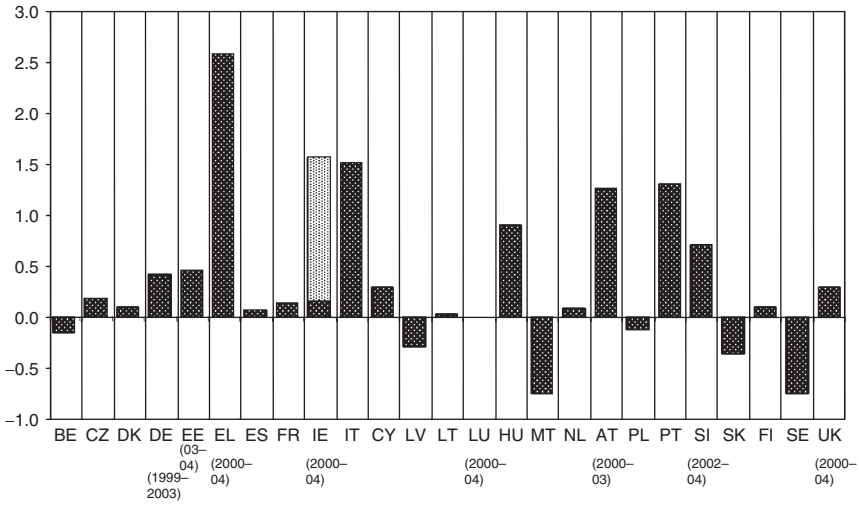


Figure 4 A comprehensive indicator of "hidden" deficits (% of GDP; 1999–2004, except new Member States (2000–04) or as otherwise indicated). Figure 4 shows data on a comprehensive indicator of "hidden" deficits. Such an indicator has been calculated as the difference between cash and accrual accounting for primary expenditure and revenue (the light coloured bar in Figure 1) plus the accumulation of assets which are more susceptible of hiding expenditure (i.e. loans and net accumulation of shares by government units other than social security and not related to privatisation). The indicator of "hidden" deficits should not be understood as an exhaustive or precise measure of creative accounting. Given the volatility of SFA components, the indicator has been calculated as a medium-term average rather than a yearly figure. For Ireland, the light-coloured part of the bar corresponds to accumulation of shares by the National Pensions Reserved Fund; which is formally classified in central government and not in social security

an expectation of depreciation. The expected revaluation of foreign debt is a cost similar to interest expenditure. However, costs stemming from the revaluation of foreign debt are booked as SFA and not in the deficit. Governments that wish to minimise their deficits could have an interest in issuing debt in low-yield currencies, even if this would increase their exchange-rate exposure. Reclassification of units (e.g. indebted public enterprises that are reclassified from the corporate sector to government) may also be the result of the hiding-deficit strategies: shifting costs to public enterprises classified in the corporate sector, though a reclassification of the enterprises and of their debts into government may take place in future.

3 Political incentives under numerical rules: a simple model

As pointed out earlier, numerical rules for deficit and debt provide incentives for creative accounting. To analyse formally how such incentives affect budgetary behaviour, we develop a simple static model of a government subject to numerical deficit and debt constraints. In spite of strong simplifications (notably, the absence of the inter-temporal implications of the SFA), the model helps capturing relevant aspects of the trade-off faced by policy makers in devising budgetary policies and schemes to embellish the budgetary position. The model set up, solution, and comparative statics are described in detail in the Appendix. Here we illustrate the basic idea and the main implications for empirical analysis.

The idea underlying the model is simple. Fiscal authorities use fiscal policy to achieve short-term growth objectives. However, because of the operation of the EU fiscal rules, budget deficits, as well as debt developments diverging from those compatible with the EU Treaty, are perceived as costly. Accounting gimmicks, as reflected in the SFA, may alleviate the perceived cost of deficits and debt, by permitting to improve the figures which are used in EU budgetary surveillance. Such operations, however, may also entail a cost, associated mainly with reduced transparency (accounting gimmicks are badly perceived by the public opinion and EU institutions) and distortions in the composition of government balance sheets. Governments will therefore trade off the benefits of higher deficits (short-run boost to growth) with their cost (the formal and informal sanctions of the EU fiscal framework). Analogously, the marginal gains from undertaking SFA operations will be equated to their marginal costs.

The model permits to derive some testable predictions concerning the relation of the SFA with deficits, debt, rules for fiscal discipline, and factors affecting government preferences, like elections. The SFA components associated with accounting gimmicks to hide deficits and those associated with transactions with financial assets that have an impact on the debt are expected to be driven by alternative explanatory factors. The model predicts that the ''hidden'' deficit component of the SFA is positively related with deficits, while the SFA component associated with purchases of financial assets is negatively related. The intuition is as follows: when the level of the deficit is high, there is a higher benefit from hiding it via accounting gimmicks to comply with the EU rules but also a bigger cost related to rising debts stemming from purchases of financial assets. Moreover, while ''hidden'' deficits are expected not be related with the stock of debt, the accumulation of financial assets by the government are expected to be negatively related

with the gross debt. Finally, a stronger perceived political cost of deficits (related to more stringent EU rules or tighter surveillance on "Maastricht" deficits) would raise the sensitivity of accounting gimmicks to the level of the deficit.

The model permits to solve for each SFA component and the level of deficits simultaneously. The deficit is therefore an endogenous variable in the model, which turns out to depend on the level of the debt, on a growth objective by the government and on parameters reflecting other aspects of the institutional setting affecting government preferences. In terms of the implementation of the empirical analysis, the fact that the deficit is endogenous calls for appropriate estimation methods to address this issue.

4 Empirical evidence

4.1 The sample

The sample used in our empirical analysis consists of observations on all EU 25 countries over the period 1994–2004. For the ten new Member States, the sample is shorter.[9] The source of the data is Eurostat, and, for what concerns SFA data prior to 2001, the European Central Bank.[10]

Table 1 reports some summary statistics for the different SFA variables used in the empirical analysis subsequently, as dependent variables. The first is total SFA excluding the valuation effects and other residual adjustments, which are most often outside the control of fiscal authorities. This measure is akin to that used in the analysis by Von Hagen and Wolff (2006). The second measure isolates the SFA components which are more susceptible of representing the "hidden" deficit.[11] Data on two proxies for this variable are reported. The first proxy is the difference in the time of recording of transactions according to the cash and accrual principles. As illustrated in section 2.2 earlier, this subset of the SFA can be seen as a conservative measure of the "hidden" deficit. The second proxy is more comprehensive and also includes the accumulation of low-quality assets

[9] The sample for the ten new Member States starts in 2000; there are severe data quality problems for earlier years. We assume that these countries started behaving strategically in a similar way to Member States in the years immediately before accession. Though they were formally subject to the SGP from 2004 only, they were already reporting their data to the European Commission—under the same format of EU Member States — under the pre-accession fiscal reporting. Moreover, the European Commission assessed their fiscal policies in the framework of the pre-accession economic programmes.

[10] The Eurostat data on fiscal and macroeconomic variables used in the analysis were those available in the AMECO database of European Commission's DG ECFIN after the release of the Commission services' Economic Forecast of spring 2005. All data are according ESA95 accounting rules including for the years prior to 2000, when the accounting standard was the old ESA79 system.

[11] This corresponds to variable x in our model (see Appendix).

Table 1 Deficit and SFA data used in regression analysis. Sample averages (% GDP)

	Whole sample (EU-25, 1994–2004)	EU-12, 1994–2004	EU-25, 1999–2004	EU-12, 1999–2004	Large countries*, 1994–2004
No. of observations	194	112	135	68	53
"Maastricht" deficit (1)	2.21	2	1.61	0.81	2.73
Total SFA (2)	0.41	0.74	0.72	1.2	0.27
"Hidden" deficit SFA (difference between cash and accrual measures of deficit) (3)	0.23	0.31	0.18	0.33	0.23
SFA related with accumulation of assets (4)	0.17	0.42	0.53	0.89	0.03
True deficit (5) = (1) + (4)	2.44	2.31	1.8	1.13	2.95
No. of observations	167	98	127	66	44
"Maastricht" deficit (6)	1.96	1.63	1.66	0.82	2.56
"Hidden" deficit SFA [comprehensive proxy: sum of (i): differences in the recording of revenue and primary expenditure (accounts receivable and payable) and statistical discrepancies; (ii) accumulation of government loans; (iii) accumulation of shares and other equities not held by social security and non-related to privatisations] (7)	0.318	0.46	0.37	0.69	0.43
True deficit (8) = (6) + (7)	2.28	2.09	2.04	1.5	2.99

*Large countries are: France, Germany, Italy, Spain and the UK.

that could correspond to disguised subsidies as discussed in section 2.2. Our conservative proxy for the "hidden" deficit will be the variable used in the baseline regressions that follow, although regressions using the comprehensive proxy will also be performed as a countercheck. The third measure of the SFA is the accumulation of financial assets by the government.[12]

The SFA is on average 0.4 percent of GDP per year over the whole sample, but the average is higher if the sample is restricted to euro-area countries or years after the introduction of the SGP. Large countries instead do not exhibit a higher SFA.[13] Over the whole sample, the SFA is roughly equally split between the component related to difference between cash and accrual deficits and the one associated with the accumulation of financial assets. The difference between cash and accrual deficits does not change significantly on average when restricting the sample to euro-area countries or to years after the introduction of the SGP. When looking at the comprehensive indicator for the "hidden" deficits, it is observed that the indicator is higher on average for euro-area countries and after the introduction of the SGP; large countries also tend to exhibit higher "hidden" deficits. On average, the indicator of "hidden" deficits is generally below half a point of GDP per year, irrespective of the conservative of comprehensive indicator. However, there is considerable cross-country dispersion in the data: some countries exhibit on average much larger "hidden" deficits than others (see Figures 1 and 4).

Table 2 reports correlation coefficients among the various SFA components and the "Maastricht" deficit across the sample. The two alternative measures for the "hidden" deficit are strongly positively correlated. The "Maastricht" deficit appears to be positively (yet quite weakly) correlated with both indicators of "hidden" deficit, while exhibiting a relatively strong negative correlation with the accumulation of financial assets. The "Maastricht" deficit and the total SFA are negatively correlated. A similar correlation pattern is observed between the "Maastricht" debt and the various SFA components: while the indicators of "hidden" deficit are weakly and positively correlated, a relatively strong correlation exists with the accumulation of financial assets.

Overall, *prima-facie* evidence indicates that the relation between deficits and the SFA is quite different depending on the specific SFA component considered.

---

[12] This corresponds to variable z in our model taken with minus sign (see Appendix).

[13] On how large and small countries react to the SGP constraints, see Buti and Pench (2004).

Table 2 Correlation among SFA components and "Maastricht" deficits (EU-25, 1994–2004)

	Total SFA (%GDP)	"Hidden" deficit SFA (difference between cash and accrual measures of deficit) (%GDP)	"Hidden" deficit SFA (comprehensive proxy) (%GDP)	SFA related with accumulation of assets (%GDP)	"Maastricht" deficit (%GDP)	"Maastricht" debt (%GDP)
Total SFA (%GDP)	1					
"Hidden" deficit SFA (difference between cash and accrual measures of deficit) (%GDP)	0.23	1				
"Hidden" deficit SFA (comprehensive proxy) (%GDP)	0.29	0.56	1			
SFA related with accumulation of assets (%GDP)	0.92	−0.14	0.0.8	1		
"Maastricht" deficit (%GDP)	−0.44	0.13	0.07	−0.5	1	
"Maastricht" debt (%GDP)	−0.10	0.21	0.15	−0.19	0.25	1

4.2 Baseline specifications

We perform fixed effect panel regressions over this sample to analyse the determinants of the SFA and its different components. The aim is to explain the determinants of three alternative measures of the SFA: the total SFA, the "hidden" deficit and the SFA associated with financial assets. The estimation method is two-stage least squares. The standard errors of the regression coefficients are robust with respect to heteroschedasticity and the possible correlation of disturbances within countries.[14]

Table 3 presents the results from the baseline specifications. The explanatory variables considered are the lagged debt/GDP ratio and the ratio of the "Maastricht" deficit over GDP. This second variable is instrumented with its own lag, the lagged debt/GDP ratio and the lagged real GDP growth rate.[15] Finally, we introduce a dummy taking value 1 for euro-area countries starting from 1999, the year in which the SGP entered into force (SGP dummy). The purpose of the dummy is to capture whether the modification of the fiscal framework associated with the creation of the EMU and the establishment of the SGP led to a different behaviour of government as far as the operations affecting the SFA are concerned.

Two specifications are considered. Specification (1) in Table 3, assumes that the SGP dummy affects only the constant term. The idea is to understand whether the introduction of the SGP increased or reduced the SFA taking the other explanatory factors constant. Specification (2) additionally allows the coefficient for the deficit to vary as a result of the introduction of the SGP. The table therefore presents two entries for the deficit coefficient. The first is the value of the coefficient in absence of the SGP. The second is the deficit interacted with the SGP dummy, which captures by how much the coefficient of the deficit changes as a result of the SGP.

---

[14] Given that the panel is unbalanced, panel Durbin–Watson tests (e.g. Bhargava et al. 1983) cannot be performed. However, the residuals from the regressions in Table 1 do not show a clear trend for most countries, which indicates that residual autocorrelation may not be a major issue. Conversely, Cook–Weisberg tests for heteroschedasticity computed on the basis of usual residual standard errors indicate that the null hypothesis of equal residual variance across panels can be rejected at almost 90 percent. Hence, in view of a possible heteroschedasticity problem, inference on regression coefficients is based on robust Huber–White residual standard errors which also take into account possible non-independence of residuals within panels.

[15] The theoretical model presented in the Appendix is solved with respect to the "true" deficit d rather than the "Maastricht" deficit d^M. Using the "Maastricht" deficit as a regressor permits not to lose observations when instrumenting, since data on "Maastricht" deficits are available over longer series compared with SFA data. Our model suggests that this choice should not pose problems in terms of the interpretation of the sign of the regression coefficient since $d^* = ((\theta_1 + \theta_3)/\theta_3)d^{M*}$, with both θ_1 and θ_3 positive (see Appendix).

Table 3 The determinants of the stock-flow adjustment: evidence from regression analysis. Baseline specifications (EU-25, 1994–2004)

Dependent variables Explanatory variables	Total SFA	"Hidden" deficit SFA (Differences between cash and accrual measures of deficit)	SFA related with accumulation of financial assets	Total SFA	"Hidden" deficit SFA Differences between cash and accrual measures of deficit	SFA related with accumulation of financial assets
Specification	(1)			(2)		
Lagged debt (% GDP)	−0.072***	0.018	−0.0918***	−0.063**	0.023	−0.086**
	(0.02)	(0.01)	(0.11)	(0.03)	(0.01)	(0.03)
"Maastricht" deficit (% GDP)	−0.223*	0.093**	−0.318***			
	(0.11)	(0.04)	(0.11)			
"Maastricht" deficit (% GDP), if no SGP				−0.149	0.127***	−0.279**
				(0.10)	(0.04)	(0.10)
"Maastricht" deficit (% GDP), Δ if SGP				−0.413**	−0.195***	−0.218
				(0.19)	(0.05)	(0.21)
Dummy SGP	−0.123	0.505**	−0.628	0.665	0.879***	−0.209
	(0.45)	(0.18)	(0.44)	(0.64)	(0.18)	(0.69)
No. of observations	191	191	191	191	191	191
R^2	0.50	0.21	0.53	0.50	0.28	0.52

Estimation method: two-stage least squares, fixed-effects panel. Hausman tests accept the hypothesis of endogeneity for the "hidden" deficit SFA component at the 90% level. Standard errors are robust with respect to heteroschedasticity and within-panels error correlation. The "Maastricht" deficit is instrumented with its own lag, the lagged debt, and the lagged real GDP growth rate.

Coefficient standard errors are reported in parentheses. *, **, ***denote, respectively, statistics significant at the 10%, 5% and 1% level.

The coefficients for the fixed effects and the constant term are not reported. The SGP dummy takes value 1 for euro-area countries after 1998. The change in the "Maastricht" deficit coefficient due to the SGP is computed as the coefficient of the "Maastricht" deficit times the SGP dummy.

Consider first specification (1).[16] Looking at the determinants of overall SFA, it turns out that operations that increase (reduce) debt ratios via the SFA are less (more) likely the higher the starting level of debt and the higher the "Maastricht" deficit. Moreover, looking at the insignificant coefficient of the SGP dummy, it appears that the introduction of the SGP did not have a significant impact on the level of aggregate SFA. Results change when looking separately at the alternative components of the SFA. In the case of the SFA component that proxies accounting gimmicks, the impact of debt is not significant, while the level of the deficit has a positive and significant effect. Each percentage point of GDP of additional deficit increases the "hidden" deficit by about 0.1 percent of GDP. These results are in line with the predictions of the model. Furthermore, the SGP dummy is positive and statistically significant: the introduction of the SGP produces an increase in "hidden" deficits by about 0.5 percent of GDP. Finally, regarding sales of financial assets, these are negatively and significantly affected both by the debt and by the deficit, a result consistent with the predictions of our model. No significant impact for the SGP dummy is found instead.

Looking at specification (2), the coefficient for the debt variable is substantially unchanged, for all measures of the SFA, compared to specification (1). For the total SFA, the coefficient of the deficit becomes more negative as a result of the SGP as in Von Hagen and Wolff (2006). By considering separately the two SFA components, it turns out that the reduction in the coefficient of the deficit is mostly associated with "hidden" deficits. This means that in absence of the SGP, accounting gimmicks would normally lead to an increase in the SFA slightly above 0.1 percent of GDP as the deficit rises by 1 percent of GDP, while the impact of additional deficits on SFA is almost negligible with the SGP. This evidence can be explained by a lower number of cases of countries with deficits above the 3 percent limit before entering into force of the SGP and outside the euro area. The evidence reported in Tables 5a and 5b supports this interpretation (see section 4.4). Again, it is confirmed the positive and significant SGP dummy in the case of the "hidden" deficit variable. The impact of the SGP on the reaction of financial operations aimed at reducing the debt to the level of deficits turns out instead being non-significant.

Several robustness checks have been performed starting from the baseline equations illustrated earlier which are not reported. First,

[16] The instruments chosen explain a large share of the variance of the deficit variable in all specifications estimated in the paper (R^2 above 0.75 in all cases). In the specifications in Table 1, Hausman tests accept the hypothesis of endogeneity for the "hidden" deficit at the 90% level.

alternative specifications including also the lagged dependent variable as a regressor to check for a richer dynamic structure have been tried. Results show that the lagged dependent variable is never statistically significant.[17] Second, regressions have also been run including only the countries with a debt/GDP ratio below 60 percent. For these countries, the debt-reduction rule does not apply, and the results from the model presented in the previous sections could not be extended. The estimates indeed perform less well. Third, results have been checked with respect to the exclusions from the sample of Luxemburg, Finland and Sweden: countries in which the respective governments have been constantly engaged in substantial accumulation of financial assets because of the large surpluses recorded by social security. Qualitative results are unchanged. Fourth, the same specifications as in Table 3 were tried using the primary cyclically adjusted (''Maastricht'') deficit instead as a measure for the (''Maastricht'') deficit. The most relevant difference compared with the results shown in Table 3 concerns the determinants of the proxy for ''hidden'' deficit: the estimated impact of the primary cyclically adjusted deficit is smaller (in absolute value) and not statistically significant. This result confirms the findings in Von Hagen and Wolff (2006): it seems that the component of deficits that is mostly offset by accounting gimmicks operations is the cyclical component (as suggested in theory by Milesi-Ferretti 2003).

4.3 Focus on euro-area countries and the impact of elections

Political incentives to circumvent EU fiscal constraints are likely to be particularly strong in electoral periods. As shown in Buti and van den Noord (2004), fiscal policies had an expansionary bias in correspondence to political elections in the early years of EMU. This seems to indicate that the objective of boosting growth via fiscal policy dominated over other objectives, in line with the literature on electoral budget cycles (see references therein).

Table 4 reports the results from two alternative empirical specifications compared with the baseline case. First, the same equations as in specification (2) in Table 3 (allowing for a coefficient for the debt that may change as a result of the SGP) is estimated restricting the sample to euro-area countries only [specification (1) in Table 4]. This permits to disentangle to what extent the impact of the SGP is associated with a different behaviour of fiscal authorities to countries outside the euro-area (and therefore, not subject to the SGP sanctions) or whether it is rather the

[17] Dickey–Fuller tests indicate that the SFA/GDP series used as dependent variables are stationary in a number of countries (about 1/3 of the cases as far as the overall SFA is concerned).

Table 4 The determinants of stock flow adjustment: evidence from regression analysis. Euro-area countries (EU-12, 1994–2004)

Dependent variables Explanatory variables	Total SFA	"Hidden" deficit SFA (Differences between cash and accruals measures of deficit)	SFA related with accumulation of financial assets	Total SFA	"Hidden" deficit SFA Differences between cash and accruals measures of deficit	SFA related with accumulation of financial assets
Specification	(1)			(2)		
Lagged debt (% GDP)	−0.0005	0.024	−0.024	0.009	0.026	−0.016
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
"Maastricht" deficit	0.100	0.152**	−0.048	0.064	0.146**	−0.078
(% GDP), if no SGP	(0.21)	(0.05)	(0.21)	(0.20)	(0.04)	(0.21)
"Maastricht" deficit	−0.597***	−0.208**	−0.393*	−0.628**	−0.217**	−0.415**
(% GDP), Δ if SGP	(0.19)	(0.07)	(0.20)	(0.16)	(0.07)	(0.17)
Dummy SGP	2.043**	0.966**	1.102	2.103**	0.986**	1.143
	(0.80)	(0.32)	(0.81)	(0.69)	(0.32)	(0.72)
Election dummy				0.701*	0.136	0.572*
				(0.33)	(0.09)	(0.28)
No. of observations	110	110	110	110	110	110
R^2	0.52	0.32	0.51	0.54	0.32	0.53

Estimation method: two-stage least squares, fixed-effects panel. Standard errors are robust with respect to heteroschedasticity and within-panels error correlation. The reported adjusted R square pertains to second-stage regressions. The "Maastricht" deficit is instrumented with its own lag, the lagged debt, and the lagged real GDP growth rate.

Coefficient standard errors are reported in parentheses. *, **, ***denote, respectively, statistics significant at the 10%, 5% and 1% level.

The coefficients for the fixed effects and the constant term are not reported. The SGP dummy takes value 1 after 1998. The change in the "Maastricht" deficit coefficient due to the SGP is computed as the coefficient of the "Maastricht" deficit times the SGP dummy. The election dummy takes value 1 in years where general elections take place.

Table 5a The determinants of stock flow adjustment: evidence from regression analysis. The impact of excessive deficits (EU-25, 1994–2004)

Dependent variables Explanatory variables	Total SFA	"Hidden" deficit SFA (Differences between cash and accruals measures of deficit)	SFA related with accumulation of financial assets	Total SFA	"Hidden" deficit SFA (Differences between cash and accruals measures of deficit)	SFA related with accumulation of financial assets
Specification	(1)			(2)		
Lagged debt (% GDP)	−0.073**	0.019	−0.092**	−0.060***	0.021	−0.081**
	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)
"Maastricht" deficit	−0.322	0.117	−0.442**			
(% GDP)	(0.16)	(0.05)	(0.13)			
"Maastricht" deficit				−0.703***	−0.038	−0.744***
(% GDP), if deficit < 3%				(0.16	(0.07)	(0.15)
"Maastricht" deficit				0.809***	0.168*	0.641***
(% GDP), Δ if deficit ≥3%				(0.21	(0.08)	(0.23)
Dummy SGP	−0.007	0.477	−0.483	0.193	0.519***	−0.324
	(0.49)	(0.19)	(0.49)	(0.41)	(0.17)	(0.44)
Dummy deficit ≥3%	0.75*	−0.184	0.942**	−1.219*	−0.594***	−0.621
	(0.38)	(0.23)	(0.31)	(0.7)	(0.15)	(0.67)
No. of observations	191	191	191	191	191	191
R^2	0.48	0.24	0.52	0.54	0.26	0.57

Estimation method: two-stage least squares, fixed-effects panel. Standard errors are robust with respect to heteroschedasticity and within-panels error correlation. The reported adjusted R square pertains to second-stage regressions. The "Maastricht" deficit is instrumented with its own lag, the lagged debt, and the lagged real GDP growth rate.

Coefficient standard errors are reported in parentheses. *, **, ***denote, respectively, statistics significant at the 10%, 5% and 1% level.

The coefficients for the fixed effects and the constant term are not reported. The SGP dummy takes value 1 for euro-area countries after 1998. The 3% dummy takes value 1 if the lagged value of the "Maastricht" deficit is above 3% of GDP. The change in the "Maastricht" deficit coefficient due to deficits being above 3% is computed as the coefficient of the "Maastricht" deficit times the 3% dummy.

Table 5b The determinants of the stock flow adjustment: evidence from regression analysis. The impact of excessive deficits (EU25, 1994–2004)

Dependent variable: "Hidden" deficit SFA (Differences between cash and accruals measures of deficit) Explanatory variables	Only euro-area countries (1)	If deficit < 3% and no SGP (2)	If deficit ≥3% and no SGP (3)	If deficit <3% and SGP (4)	If deficit ≥3% and SGP (5)
Lagged debt (% GDP)	−0.014	0.042***	0.012	0.05	0.129
	(0.01)	(0.01)	(0.6)	(0.08)	(0.08)
"Maastricht" deficit	−0.728***	0.014	0.332***	−0.05	0.21***
(% GDP), if deficit < 3%	(0.22)	(0.18)	(0.12)	(0.1)	(0.05)
"Maastricht" deficit	1.040***				
(% GDP), Δ if deficit ≥3%	(0.25)				
"Maastricht" deficit					
(% GDP), if no SGP					
"Maastricht" deficit					
(% GDP), Δ if SGP					
Dummy SGP	0.838				
	(0.60)				
Dummy deficit ≥3%	−1.731*				
	(0.886)				
No. of observations	110	53	70	53	15
R^2	0.57	0.33	0.31	0.22	0.77

Estimation method: two-stage least squares, fixed-effects panel. Standard errors are robust with respect to heteroschedasticity and within-panels error correlation. The reported adjusted R square pertains to second-stage regressions. The "Maastricht" deficit is instrumented with its own lag, the lagged debt, and the lagged real GDP growth rate.

Coefficient standard errors are reported in parentheses. *, **, *** denote, respectively, statistics significant at the 10%, 5% and 1% level.

The coefficients for the fixed effects and the constant term are not reported. The SGP dummy takes value 1 for euro-area countries after 1998. The 3% dummy takes value 1 if the lagged value of the "Maastricht" deficit is above 3% of GDP. The change in the "Maastricht" deficit coefficient due to deficits being above 3% is computed as the coefficient of the "Maastricht" deficit times the 3% dummy.

result of a changed behaviour in euro-area countries after the introduction of the SGP. A second specification [specification (2) in Table 4] extends the empirical model by introducing an election variable.[18] It is not *a-priori* obvious whether the SFA should be larger or smaller in the presence of elections. A first interpretation, in line with electoral cycle theories, is that under elections incumbent governments set more ambitious growth objectives.[19] According to this hypothesis, as illustrated in Figure 4, one would expect a positive coefficient for the election variable in the case of "hidden" deficits, while the coefficient is expected to be negative in the case of financial operations (a slower accumulation of financial assets under elections). However, there are also reasons why one may expect instead that the pace of accumulation of financial assets by the government increases during elections. Under elections, governments may want to keep a high degree of control on economic activities and have, therefore, lower incentives to carry out privatisation programmes, or may be more prone to bail out private or public corporations via the acquisition of share capital.

Repeating the baseline regressions to a sample of euro-area countries only [specification (1) in Table 4] broadly confirm those obtained considering EU 25 countries. However, there are some noteworthy differences. The sign for the debt variable is as expected, but the coefficient is smaller and not significantly different from zero. The deficit variable instead becomes significantly more negative after the introduction of the SGP. Again, in the case of the SFA component that we assumed to indicate "hidden" deficits, the coefficient turns from positive to negative. In the case of financial operations the coefficient of the deficit becomes much more negative after the SGP. The SGP dummy affects the constant term much more significantly compared with the baseline case. The overall SFA increases by more than 2 percent of GDP after the introduction of the SGP, with an almost equal contribution of its two components. Overall, restricting the analysis to euro-area countries, the upward jump in the SFA after the SGP appears more evident.

Specification (2) includes the election dummy among the explanatory factors. The dummy is positive and significant when the dependent variable is the overall SFA. Elections increase the SFA by more than 0.7 percentage points of GDP. The election dummy is positive for both SFA components, with the coefficient only barely significant in the case of the SFA component associated with disguised deficits. This evidence

[18] A dummy variable taking value 1 if general elections took place in that particular country, in that particular year.

[19] Parameter $\hat{y}$ in terms of the model presented in the Appendix.

seems to indicate that, consistently with electoral cycle arguments, in the presence of elections the pressure to embellish deficit figures increases, while the accumulation of financial assets by government increases for the possible reasons listed earlier.

4.4 The impact of excessive deficits

Which impact had the occurrence of deficits in excess of the 3 percent Maastricht reference value on the SFA? We investigate this issue in the specifications presented in Tables 5a and 5b. Specification (1) in Table 5a adds a constant dummy that takes value 1 when the ''Maastricht'' deficit is in excess of the reference value. Specification (2) in the same table also admits a different coefficient for the deficit depending on whether the recorded ''Maastricht'' deficit is above or below 3 percent of GDP. In order to overcome endogeneity issues (i.e. the fact that the deficit is above or below 3 percent may depend on the SFA dependent variable) the 3 percent dummy is constructed using the lagged value of the deficit. The variable captures therefore those cases for which, in the preceding year, the deficit is above the Maastricht reference value.

Results from specification (1) in Table 5a indicate that the fact that the deficit is above the reference value does not affect *per se* the overall level of the SFA. The dummy is never significant, for any SFA component. It is confirmed instead that the SGP raises significantly the SFA component associated with ''hidden'' deficits.

When the coefficient of the deficit variable is allowed to vary when deficits are ''excessive'' [specification (2) in Table 5a], results change quite considerably. In the case of the regression concerning the overall SFA, the deficit coefficient is significantly negative when deficits are below 3 percent and rises significantly when the coefficient becomes ''excessive''. Looking at the regression for the component capturing accounting gimmicks, it turns out that the deficit coefficient is roughly nil if the deficit is below the Maastricht reference value, while it is significantly positive when the deficit is above. This suggests that the whether or not countries are in excessive deficit may matter for the impact of the SGP on the behaviour of accounting gimmicks in response to the deficit and may help interpreting the results in Table 3 indicating that the reaction of ''hidden'' deficits to ''Maastricht'' deficits seem to have weakened with the SGP.

Table 5b displays further regressions aimed to shed light on the behaviour of the ''hidden'' deficit component of the SFA. Specification (1) repeats the same analysis as in Table 5a but limiting the sample to euro-area countries. The aim is disentangling whether it is mainly the behaviour of euro-area countries during the run up to EMU that affects the result in

Table 5a or rather that of non-euro-area countries, and notably that of the countries that acceded to the EU in 2004. Specifications (2)–(5) restrict the sample, respectively, to country/year combinations such that: (i) (lagged) deficits are below 3 percent of GDP and the SGP is not into force; (ii) deficits are above 3 percent and the SGP is not operating; (iii) deficit is below 3 percent and the SGP is into force; (iv) deficits are above 3 percent and the SGP is in action. The objective in this case is to check the effect played by the interaction of excessive deficits and the SGP in driving the reaction of accounting gimmicks to deficits.

The estimates in Table 5b indicate that when limiting the analysis to euro-area countries the results in Table 5a turn out being broadly confirmed. The evidence for euro-area countries does not appear to be radically different compared with that referring the sample of all EU25 countries. Results in specifications (2)–(5) reveal that when deficits are in excess of the 3 percent ceiling there is more intensive use of accounting gimmicks the large is the deficit, irrespective of whether the SGP is or is not into force. Conversely, the operation of the SGP *per se* does not seem to play any relevant role on the reaction of accounting gimmicks to deficits.

Overall, an interpretation of the interplay between the evolution of the EU fiscal framework and the incentives to carry out accounting gimmicks could be as follows. The excess of the deficit over 3 percent of GDP trigger a use of accounting gimmicks which is more intense the bigger the size of the deficit. This holds irrespective of the operation of the SGP. During the run-up to EMU governments had mainly an incentive to disguise their deficits when their deficits exceeded the 3 percent of GDP threshold, as their principal endeavour was to qualify to the euro. A similar reasoning applies currently to the countries that acceded in 2004. With the SGP, countries in excessive deficits need instead to ensure a rapid budgetary adjustment to avoid sanctions. The SGP *per se* does not seem to modify the reaction of accounting gimmicks to the size of deficits. However, our estimates show that the SGP is associated with a higher amount of SFA associable with accounting gimmicks, irrespective of the size of the deficit (as indicated by the significant constant SGP dummy). This result could be explained by the medium-term commitment to reach a budgetary position of close to balance introduced with the SGP, which applies both to cases where the deficit is above or below the 3 percent deficit threshold.

4.5 An alternative breakdown for the SFA

The results illustrated so far use a breakdown of the SFA which provides a comprehensive measure of financial operations but a rather partial measure of accounting gimmicks aimed at reducing the deficit. We therefore repeat the regressions presented in the baseline specifications

using the comprehensive measure for "hidden" deficits discussed in section 2.2. In this case, the financial operations that could be carried out by the government as an alternative to the provision of subsidies (i.e. via loans and shares or other equities non-held by social security and unrelated to privatisation) are moved from financial operations which contribute to reduce the debt into the "hidden" deficit component. This finer decomposition has a large cost in terms of reduced data availability but permits to countercheck our main results.

Table 6 displays the results for regressions relating to the same baseline specifications as presented in Table 3. Although the loss of observations translates into a reduction in the degree of significance of the estimates the sign of the coefficients is the expected one and the results are qualitatively the same as those obtained with the breakdown adopted in the previous regressions. In particular, it is confirmed that the SFA components associable with "hidden" deficits increases significantly after the introduction of the SGP, while this is not the case for the remaining SFA. The comprehensive measure of "hidden" deficits used in the regressions presented in Table 6 is the sum of differences between cash and accrual and of accumulation of assets that could represent disguised government subsidies. By repeating the analysis using only this second component as dependent variable, regression coefficients appear largely statistically insignificant. This suggests that the results obtained with the comprehensive measure of "hidden" deficits are mainly due to the difference between cash and accruals.

4.6 Summarising the empirical results

The main messages from the empirical results presented above can be summarised as follows.

(i) The overall SFA appears to be negatively related (yet not significantly) to deficits. However, the aggregate SFA masks relevant differences for different SFA components. While the relation is positive for the "hidden" deficit component, it is negative for financial operations.
(ii) The level of the debt has also a different impact on different SFA components: not significant for "hidden" deficits, negative for the accumulation of financial assets.
(iii) The level of the accounting gimmicks components of the SFA increases significantly with the SGP.
(iv) Accounting gimmicks are unaffected by deficits below the 3 percent ceiling, but increase strongly with deficits when these are above the Maastricht threshold.

Table 6 The determinants of the stock-flow adjustment: evidence from regression analysis. An alternative break-down for the SFA, baseline specifications (EU-25, 1994–2004)

Dependent variables Explanatory variables	Total SFA	"Hidden" deficit SFA, comprehensive proxy (A)	"Hidden" deficit SFA unrelated with differences between cash and accrual (B)	SFA related with accumulation of other financial assets (C)	Total SFA	Hidden" deficit SFA, compre-hensive proxy (A)	"Hidden" deficit SFA unrelated with differences between cash and accrual (B)	SFA related with accumulation of other financial assets (C)
Specification	(1)				(2)			
Lagged debt (% GDP)	−0.079**	0.002	−0.005	−0.082*	−0.081**	0.002	−0.005	−0.083*
	(0.02)	(0.01)	(0.018)	(0.04)	(0.03)	(0.19)	(0.017)	(0.04)
"Maastricht" deficit (% GDP)	−0.182	0.058	−0.045	−0.241				
	(0.17)	(0.08)	(0.10)	(0.15)				
"Maastricht" deficit (% GDP), without SGP					−0.010	0.117	−0.062	−0.127
					(0.13)	(0.09)	(0.09)	(0.14)
"Maastricht" deficit (% GDP), change due to SGP					0.447*	−0.152	0.023	−0.295
					(0.25)	(0.11)	(0.28)	(0.20)
Dummy SGP	0.234	0.618**	0.102	−0.383	1.031	0.889**	0.023	0.142
	(0.56)	(0.28)	(0.91)	(0.56)	(0.79)	(0.28)	(0.28)	(0.75)
No. of observations	166	166	166	166	166	166	166	166
R^2	0.50	0.33	0.24	0.52	0.52	0.34	0.25	0.53

Estimation method: two-stage least squares, fixed-effects panel. Standard errors are robust with respect to heteroschedasticity and within-panels error correlation. The "Maastricht" deficit is instrumented with its own lag, the lagged debt and the lagged real GDP growth rate.

Coefficient standard errors are reported in parentheses. *, **, ***denote, respectively, statistics significant at the 10%, 5% and 1% level. The coefficients for the fixed effects and the constant term are not reported. The SGP dummy takes value 1 for euro-area countries after 1997. The change in the "Maastricht" deficit coefficient due to the SGP is computed as the coefficient of the "Maastricht" deficit times the SGP dummy.

(A) = sum of (i): differences in the recording of revenue and primary expenditure (accounts receivable and payable) and statistical discrepancies; (ii) accumulation of government loans; (iii) accumulation of shares and other equities not held by social security and non related to privatisations.

(B) = (A) − (i)

(C) = accumulation of following assets by government: (i) liquidities; (ii) securities other than shares; (iii) shares held by social security; (iv) shares related to privatisation.

(v) Elections affect positively all SFA components.

Result (i) permits to better qualify findings already reported in Von Hagen and Wolff (2006). The authors attribute the negative relation of the SFA to deficits mainly to creative accounting related to a strategic use of financial operations (e.g. disguised subsidies accounted for as stock acquisitions by government). Our analysis permits to disentangle different SFA components and indicates that the negative relation between the SFA and deficits can be mainly associated with financial operations, while the relation between deficits and the SFA component more strictly associable with ''hidden'' deficits is on average positive. The model presented in this article can provide an explanation to these findings. The ''hidden'' deficits are positively related to deficits because the higher the deficit, the stronger the incentive to engage in creative accounting. Transactions with financial assets are instead negatively related because their purpose is to contain the growth of the debt. The same qualitative result is obtained by using a finer breakdown of the SFA which recognises that a number of financial transactions may also be used to hide expenditure.

The presence of the SGP is associated with significantly more ''hidden'' deficits irrespective of the deficit level (as reflected in the significant value of the SGP dummy both in the regressions for EU 25 and EU 12 countries), while the presence of excessive deficits increases the sensitivity of accounting gimmicks to deficits [results (iii) and (iv)]. These results are consistent with existing findings pointing to a positive effect of deficits above the Maastricht threshold on the probability of carrying out accounting tricks and one-off operations (Koen and van den Noord 2006).

Finally, we learn from result (v) that elections matter for the SFA, controlling for other determinants. The fact that in the current year general elections take place affects positively all SFA components. The positive coefficient for ''hidden'' deficits fits with the interpretation that creative accounting and a lower degree of fiscal transparency enhance the capacity of governments to put in place electoral cycles, as for instance recently highlighted empirically in Alt and Lassen (2005). On the other hand, the positive and significant coefficient in the case of financial operations is of a less obvious reading. Among the interpretation there is that under elections governments have less incentives to prepare large privatisations.

5 Conclusions

The aim of this article was to study how the budgetary rules of EMU give rise to political incentives for manipulating fiscal variables with the purpose of hiding deficits and reduce gross debt. We show both

theoretically and empirically that such powerful incentives were at work during the run-up and in the early years of EMU. Governments used a number of operations to conceal the true size of their deficits and put in place financial operations to stem the increase in the public debt. The former increased in importance after the advent of the SGP, which shifted the focus of policy surveillance on deficits with initially scant attention to the means used to ensure the respect of the deficit rule and little weight to the debt rule. Consistently with our model, increased weight to the deficit criterion in EU surveillance resulted into lower "Maastricht" deficits but also into a higher incidence of stock-flow adjustments potentially connected with accounting tricks to keep "Maastricht" deficits low. We show that such incentives were reinforced in electoral periods.

In the recent reform of the SGP, more emphasis is put on the debt rule and, more generally, on long-term sustainability, and on the need to ensure a durable correction in the excessive deficit via structural adjustment. In our analysis, a stronger emphasis on the debt would reduce deficits under most likely conditions.[20] However, such shift in focus may induce governments to carry out more sales of financial assets. To prevent this, as suggested, for instance, by Easterly (1999) and Coeuré and Pisany-Ferry (2005), fiscal surveillance oriented on a comprehensive notion of government assets and liabilities (net debt) would contribute to reduce the incentives to decumulate assets to reduce the gross debt and, indirectly to keep deficit and debt low. A higher attention to structural adjustment implies an increase in the political cost of deficit-reducing one-off operations and would reduce the incidence of accounting tricks. In the same direction goes the call for increased statistical transparency—as required by the reformed SGP—fostered by a public opinion becoming more adverse to fiscal gimmicks.[21] The shift to a more comprehensive fiscal surveillance based on multiple indicators may help to reduce the incentives to data manipulations (see, e.g. Balassone, Franco and Zotteri 2006).

While the empirical analysis carried out in this article appears relatively robust and the results broadly consistent with those in the literature, several avenues for further research can be pursued. First, SFA

[20] On the reform of the SGP, see Buti, Eijffinger and Franco (2005), Buti (2006), European Commission (2005 and 2006), and Beetsma and Debrun (2007).

[21] There is anecdotal evidence pointing to a higher perceived cost of creative accounting activities. Not only EU authorities are putting greater focus on statistical transparency, but also financial markets and credit agencies are increasingly aware of the long-term consequences for public finances of window dressing activities (see, e.g. "EU securitisation may have passed peak", *Financial Times*, 7 December 2005).

components do not capture all the means throughout which governments can manipulate fiscal variables. For instance, operations that allow cashing an immediate receipt in exchange for higher pension liabilities, sales of real estate, as well as other one-off revenues do not affect any SFA component. A comprehensive analysis should integrate the three research approaches recalled in the introduction (bottom-up identification of tricks, balance sheet and analysis of SFA components). Second, in the implementation of the EU fiscal rules, the first outcomes are the figures that really matter, though revisions in deficit, debt and SFA components are frequent. Though large revisions may be detrimental for the credibility of the concerned Member States (notably Portugal in 2002, Italy in 2005 and especially Greece in 2004) and of the SGP itself, data revisions are relatively irrelevant for the implementation of the Pact. Therefore, governments have a specific interest in portraying healthy public finances in the first deficit notification, even if the data are subsequently revised upwards. In order to capture more effectively the political incentives to manipulate fiscal accounting, one should use real-time SFA.[22] We conjecture that the repetition of our empirical analysis in real-time data would confirm and reinforce, our conclusions.

References

Alt, J. and D. Lassen (2005), ''The Political Budget Cycle Is Where You Can't See It: Transparency and Fiscal Manipulation'', EPRU Working Paper Series, 05-03, Economic Policy Reasearch Unit, University of Copenhagen.

Balassone, F. and D. Monacelli (2000), ''EMU Fiscal Rules: Is There a Gap?'', Temi di Discussione (Economic working papers) 375, Bank of Italy, Economic Research Department.

Balassone, F., D. Franco and S. Zotteri (2006), ''EMU Fiscal Indicators: A Misleading Compass?'', *Empirica* **33(2)**, 63–87.

Beetsma, R. and X. Debrun (2007), ''The New Stability and Growth Pact: A First Assessment'', *European Economic Review*, forthcoming.

Bernoth, K. and G. Wolff (2006), ''Fool the Markets? Creative Accounting, Fiscal Transparency and Sovereign Risk Premia'', Deutsche Bundesbank Discussion Papers, 19/2006.

[22] See, e.g. Forni and Momigliano (2004) for an analysis of fiscal behaviour using real-time output gaps.

Bhargava, A., L. Franzini and W. Narendranathan (1983), ''Serial Correlation and the Fixed Effects Model'', *Review of Economic Studies* **49**, 533–549.

Buiter, W. (1985), ''A Guide to Public Sector Debt and Deficits'', *Economic Policy* **1**, 13–79.

Buiter, W., G. Corsetti, N. Rubini et al. (1993), ''Excessive Deficits: Sense and Nonsense in the Treaty of Maastricht'', *Economic Policy* **8**, 57–100.

Bunch, B.S. (1991), ''The Effect of Constitutional Debt Limits on State Governments' Use of Public Authorities'', *Public Choice* **68**, 57–69.

Buti, M. (2006), ''Will the New Stability and Growth Pact Succeed? An Economic and Political Perspective'', in F. Breuss, ed., *The Stability and Growth Pact - Experiences and Future Aspects*, New York: Springer Wien, pp. 156–181.

Buti, M. and L.R. Pench (2004), ''Why Do Large Countries Flout the Stability Pact? And What Can Be Done About It?'', *Journal of Common Market Studies* **42(5)**, 1025–32.

Buti, M. and P. van den Noord (2004), ''Fiscal Discretion and Elections in the Early Years of EMU'', *Journal of Common Market Studies* **39(4)**, 737–56.

Buti, M., S. Eijffinger and D. Franco (2005), ''The Stability Pact Pains: a Forward-Looking Assessment of the Reform Debate'', CEPR Discussion Paper, 5216.

Buti, M., D. Franco and H. Ongena (1998), ''Fiscal Discipline and Flexibility in EMU – The Implementation of the Stability and Growth Pact'', *Oxford Review of Economic Policy* **14(3)**, 81–97.

Coeuré, B. and J. Pisani-Ferry (2005), ''Fiscal Policy in EMU: Towards a Sustainability and Growth Pact?'', *Oxford Review of Economic Policy* **21(4)**, 598–617.

Dafflon, B. and S. Rossi (1999), ''Public Accounting Fudges towards EMU: A First Empirical Survey and Some Public Choice Considerations'', *Public Choice* **101**, 59–84.

Easterly, W. (1999), ''When is Fiscal Adjustment an Illusion?'', *Economic Policy* **28**, 55–86.

European Commission (2005), *Public Finances in EMU*, European Economy No. 3.

European Commission (2006), *Public Finances in EMU*, European Economy No. 3.

Forni, L. and S. Momigliano "Cyclical Sensitivity of Fiscal Policies based on Real-time Data", Temi di Discussione (Economic working papers) 540, Bank of Italy, Economic Research Department.

Gros, D. (2003), "A Stability Pact for Public Debt?", *CEPS Policy Brief*, No. 30.

von Hagen, J. (1991), "A Note on the Empirical Effectiveness of Formal Fiscal Restraints", *Journal of Public Economics* **44**, 199–210.

von Hagen J. (2005), "Fiscal Rules and Fiscal Performance in the EU and Japan", IMES Discussion paper, 2005-E-5.

von Hagen, J. and G. Wolff (2006), "What Do Deficits Tell Us about Debts? Empirical Evidence on Creative Accounting with Fiscal Rules", *Journal of Banking and Finance* **30**, 3259–79.

Koen, V. and P. van den Noord (2006), "Fiscal Gimmickry in Europe: One-off Measures and Creative Accounting", in P. Wierts, S. Deerose, E. Flores and A. Turrini, eds., *Fiscal Policy Surveillance in Europe*, Bansingstoke: Palgrave McMillan, forthcoming.

Milesi-Ferretti, G. (2003), "Good, Bad, or Ugly? On the Effects of Fiscal Rules with Creative Acconting", *Journal of Public Economics* **88**, 377–94.

Milesi-Ferretti, G. and K. Moriyama (2006), "Fiscal Adjustment in EU Countries: A Balance Sheet Approach", *Journal of Banking and Finance* **30**, 3281–98.

Sbragia, A. (1996), *Debt Wish*, Pittsburgh: University of Pittsburgh Press.

Strauch, R. (1998), "Budget Processes and Fiscal Discipline: Evidence from the US States", ZEI Working Paper.

Appendix: A simple model

The first relation in the model links deficits and short-run growth. In the short-run, prices are sticky, so that output is demand-determined. It follows that fiscal policies that increase deficits by cutting government receipts or by raising payments contribute positively to short-run growth. Assuming for ease of notation an initial deficit equal to zero, we write

$$y = \phi d \tag{1}$$

where y is real output growth, d is the deficit as a share of output, and ϕ is the fiscal multiplier. A key assumption is that economic activity

is influenced by this notion of "true" deficit in cash terms.[23] As in the EU fiscal rules, however, the numerical constraint applies not to the true deficit in cash terms, d, but to the national accounts, accrual-based definition, of budget deficit, d^M (where the superscript M stands for "Maastricht"):

$$d^M = d - x \tag{2}$$

where x corresponds to the "hidden" deficit, namely the SFA component that may be "manipulated" by fiscal authorities with the specific aim of affecting the "Maastricht" deficit.

The debt accumulation identity, disregarding inflation, can be approximated as follows:

$$\dot{b} = d - yb - z \tag{3}$$

where b is the debt/output ratio, $\dot{b}$ is the time change in this ratio, and z denotes financial operations carried out by governments that do not affect the deficit but reduce the debt (i.e. z is the measure of the SFA associated with purchases of financial assets taken with minus sign, so that the total SFA is $x - z$).

The numerical rule on public debt states that the debt to GDP ratio, b, as long as it is above a given value, $\overline{b}$, should preferably be reduced at a speed α, implying that the distance of the debt from its target value should be reduced by α points a year:

$$\dot{b} = \alpha(\overline{b} - b) \tag{4}$$

Equation (4) formalises the Treaty requirement that the debt, until it is above the reference value of 60 percent of GDP, should be reduced at a satisfactory pace.[24]

We postulate that the government aims at attaining a given level of output growth, call it $\hat{y}$. The government has three instruments at its disposal: its effective cash receipts and payments which lead to the "true" deficit in cash terms (d); the "hidden" deficit (x), and transactions in financial assets which contribute to reduce the government debt (z). Fiscal authorities need to respect as far as possible constraints on deficit and debt

[23] This measure of deficit overlaps only partly with that advocated by Balassone and Monacelli (2000) and Balassone, Franco and Zotteri (2006) as the most correct indicator of annual budgetary behaviour. In those definitions most SFA components enters the deficit. Our definition of deficit also overlaps only partly with the "bare" deficit as defined by Koen and van den Noord (2006) who strip out of the "Maastricht" deficit also one-off operations.

[24] Such an expression has been used to represent the debt Maastricht rule by e.g. and Buti, Franco and Ongena (1998). Gros (2003) shows arithmetically that with constant 5% nominal growth of GDP, a constant budget deficit of 3% of GDP ensures a speed of reduction of the debt in excess of 60% of GDP of 5% a year.

similar to those in the EU. Accounting gimmicks (x) can be used to limit the deviation from the deficit objective, but they are assumed to carry a political cost as such operations trigger enhanced surveillance due to the suspicion that the government may be engaged in unlawful accounting practices. Similarly, financial operations (z) permit the government to come closer to the objective for the debt path, but also these measures carry a cost, related in this case to the possibility of a sub-optimal composition of government balance sheets.[25]

The following government loss function attempts to capture in a simplified fashion the objectives of and constraints on government budgetary behaviour:

$$L = (y - \hat{y})^2 + \theta_1(d - x)^2 + \theta_2\left[\dot{b} - \alpha(\overline{b} - b)\right]^2 + \theta_3 x^2 + \theta_4 z^2 \qquad (5)$$

The government minimises the L with respect to d, x and z. For convenience, and without an impact on qualitative results, it is assumed that fiscal authorities have a target corresponding to a zero "Maastricht" deficit and that the "political cost" of a discrepancy between the actual value of d^M and the target increases more than proportionally with the size of such discrepancy (i.e. with the square). As will be clearer later, a sufficiently high value of $\hat{y}$ guarantees that the deficits are always constrained from above, i.e. that the true budget balance chosen by fiscal authorities is negative and that an increase in parameter θ_1 would reduce the true deficit. Moreover, the above loss function holds as long as $b > \overline{b}$ (in the opposite case, it is assumed $\theta_2 = 0$).

The solution of the optimisation problem for x and z can be expressed as a function of the true deficit level as follows:

$$x^* = \frac{\theta_1 d^*}{\theta_1 + \theta_3} \qquad (6)$$

$$z^* = \frac{\theta_2[(\alpha(b - \overline{b}) + d^*(1 - \phi b)]}{\theta_2 + \theta_4} \qquad (7)$$

As for the solution of the true deficit, it is determined in the following way:

$$d^* = \frac{\phi\hat{y} - (1 - \phi b)[\alpha(b - \overline{b})](\theta_2\theta_4)/(\theta_2 + \theta_4)}{\phi^2 + (1 - \phi b)^2(\theta_2\theta_4)/(\theta_2 + \theta_4) + (\theta_1\theta_3)/(\theta_1 + \theta_3)}$$

[25] For instance, a privatisation programme pursued during times of weak demand by financial markets may lead to undervaluation of the assets previously held by the government and to a reduction in the government net worth.

We suppose that $\phi b < 1$, which guarantees that that higher deficits lead to an increase in the debt/output ratio, namely, $\partial \dot{b}/\partial d > 0$. This amounts to assuming that the impact of deficits on the numerator of the debt/output ratio prevails over the denominator effect: a condition that is likely to hold in reality and consistent with the dynamic efficiency of economies. In our case, this condition also guarantees $\dot{b} > 0$ whenever $d > 0$, so that the minimum rate of debt reduction is always binding at the solution.

Accounting gimmicks, x, depend positively on the deficit. This result follows from the fact that the model permits to analyse the interaction and the determination of deficits and that of SFA components. If fiscal authorities find it optimal to run high deficits, they will also have more incentives to hide such deficit and avoid that d^M exceeds a given threshold.[26]

Financial operations z depend directly both on the debt and the deficit. Under the assumption $\phi b < 1$ and $b > \overline{b}$, there is more decumulation of assets by the government if deficits increase (e.g. privatisations, becomes more likely and accumulation of assets less likely). To meet the target rate of debt reduction, fiscal authorities need to compensate via the SFA the increase in the debt/output ratio implied by higher deficits. The impact of the debt (at given deficit) on government financial operations is *a-priori* ambiguous. On the one hand, if the debt/output ratio is high, a stronger reduction in debt is necessary to comply with the target debt-reduction rate, and this leads to more sales of financial assets. On the other hand, when the debt ratio is high, expansionary budgetary policies will lead to a stronger pace of debt reduction so that a lower z would suffice. The prevalence of the first effect requires $\alpha > \phi d^*$, i.e. that the target reduction rate of debt is relatively high compared with the equilibrium level of the deficit. Under the conditions prevailing in the EU and in light of the prescriptions of the EU fiscal framework, the above assumption is likely to be satisfied.[27]

---

[26] It is to notice that such result differs from that in Von Hagen and Wolff (2006), where the relation between deficits and the SFA is negative because the SFA is used to accommodate the difference between a targeted change in debt and the deficit threshold.

[27] Parameter α is not spelled out in the EU fiscal framework. However, a lower bound for such a parameter could be inferred from the path for convergence of the debt ratio towards 60% of GDP consistent with a deficit of 3% of GDP, nominal GDP growth of 5% and disregarding SFA. (Such a benchmark has been used by the European Commission since 2004 in the assessment of stability and convergence programmes.) In this case, α would be 5%. Such a value for α would most likely satisfy the condition $\alpha > \phi d$.

After plugging (8) into (6) and (7) it is possible to perform comparative statics of the solutions for x and z with respect to model parameters. Table 7 presents in synthetic form the results.

A higher growth target $\hat{y}$ raises the deficit and then both SFA components. A higher pace of debt reduction α (or a lower debt target $\bar{b}$) reduces the deficit and therefore x. Even if the impact of α on z is *a-priori* ambiguous (z rises directly but falls via d) the first direct positive effect prevails. The debt/output ratio has an ambiguous impact on the deficit and therefore also on x and z.

Parameter θ_1, the weight given in the government loss function to the deficit objective, has two opposing effects on accounting gimmicks. On the one hand, a higher θ_1 rises x at given deficit. On the other hand, d falls, and this entails a lower value for x at equilibrium. It can be shown that the first effect always prevails. As for the impact of θ_1 on z, it only comes through d, and is unambiguously negative.

The effect of θ_2 on d is negative when $d > 0$ (i.e. if the true budget balance records a deficit). The impact of θ_2 on x is negative because less accounting gimmicks help ensuring faster debt reduction. However the effect of θ_2 on z would be ambiguous, due to two opposite effects: the direct one, which raises z, and indirect one, via a lower d. Finally, an increase in θ_3 and θ_4, the parameters capturing the political cost of increasing the SFA, lead to both lower deficit and lower SFA.

Table 7 Political incentives under numerical rules: a simple model. Main results from comparative statics

	Impact on deficit (d)	Impact on "hidden" deficit (x)	Impact on financial operations (z)
Growth objective ($\hat{y}$)	+	+	+
Pace of debt reduction (α), debt target ($\bar{b}$)	−	−	+
Debt level (b)	?	?	?
Political cost of deficit (θ_1)	−	+	−
Political cost of debt (θ_2)	−	−	?
Political cost of "hidden" deficit (θ_3)	−	−	−
Political cost of "financial operations" (θ_4)	−	−	−