

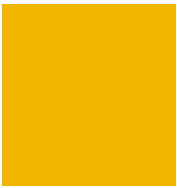


Sveriges Riksbank
Economic Review



2012:3





Sveriges Riksbank
Economic Review

2012:3

SVERIGES RIKSBANK ECONOMIC REVIEW

is issued by Sveriges Riksbank.

PUBLISHER: CLAES BERG

EDITORS: CLAES BERG, JOANNA GERWIN
AND THE COMMUNICATIONS SECRETARIAT

Sveriges Riksbank, SE-103 37 Stockholm, Sweden

Telephone +46 8 787 00 00

Advisory editorial committee: Martin W Johansson, Göran Robertsson
and Kasper Roszbach

The views expressed in signed articles are the responsibility of
the authors and are not to be regarded as representing the views of
the Riksbank on the matters concerned.

The Review is published on the Riksbank's website
www.riksbank.se/en/Press-and-published/Reports/Economic-Review

Order a link to download new issues of Sveriges Riksbank Economic Review
by sending an e-mail to: pov@riksbank.se

Dear reader,

We open this issue of the Riksbank's journal with three articles focusing on issues that are highly significant to the development of the monetary-policy work of central banks.

- Elmér, Guibourg, Kjellberg and Nessén analyse the monetary policy measures taken by the Riksbank during the financial crisis of 2008-2009, when the Riksbank implemented not only conventional measures, such as substantially cutting the repo rate to counteract the fall in demand for goods and services, but also a number of extraordinary measures. For example, the Riksbank offered SEK and USD loans to the Swedish banks at longer maturities. The aims of these loans were to ensure that there was sufficient liquidity in the financial system, to counteract the development of high risk premiums and to improve the monetary-policy transmission mechanism. These measures helped to lower both short- and long-term interest rates during the crisis, which meant, for example, that the banks' costs for funding new household mortgages could be kept down. Experience from the financial crisis shows that it is important for the Riksbank to be constantly prepared to quickly take appropriate extraordinary measures when the need arises.
- Carlsson analyses the relationship between monetary policy, inflation and unemployment. He shows that this relationship is affected by the type of theoretical model used. In the long run, unemployment is affected by the regulatory frameworks and institutions that govern the labour market. According to the older academic literature, it is unemployment's deviation from a constant long-run rate that affects inflation. In a modern new-Keynesian model with rigid wages and prices this does not apply. Here inflation is affected by the gap between the observed rate of unemployment and the rate that would prevail if all prices and wages adjusted without frictions. This difference means that it is important to clarify the underlying conceptual framework that is being used when discussing the impact of monetary policy on inflation and unemployment.
- Hokkanen, Melin and Nilson describe the methods used for, and the results of, the business survey that the Riksbank carries out three times a year. The aim of this survey is to provide information about the economic climate before the official statistical outcomes are available. The responses to the questions put to the companies thus complement the economic statistics and can therefore be used as an early basis for monetary policy assessments. They also help the Riksbank to perform a qualitative analysis that aims to discover and understand new aspects of economic development from the companies' perspective. The interviewed companies thus constitute an important network that facilitates the Riksbank's analysis of the business sector. The Riksbank also processes the companies' responses to create a quick quantitative indicator of economic activity. It was this indicator, together with the qualitative analysis,

that enabled the business survey to capture the economic downturn in 2008 and the recovery in 2009 at an early stage.

We also present two articles that highlight some economic aspects of risk and efficiency in certain parts of the financial system.

- Juks analyses how banks in Sweden and other countries encumber assets in order to reduce their funding costs. Part of this pledging may be justified in periods of financial unease, when investors seek safety at the same times as there is a shortage of low-risk securities. The possibility to encumber assets or pledge collateral thus can make funding less expensive. However, if the encumbering of assets becomes very common it may also contribute to the creation of risks in the financial system. This is because the encumbering of assets has the potential to shift risks to non-secured creditors. If unsecured creditors count on being able to transfer some of this risk to the government, and thus to the taxpayers, there is a possibility that the proportion of encumbered assets in the banking sector will be larger than is socially justified. To counteract this possibility it may be appropriate for the authorities to strengthen market discipline. This can be done, for example, by making the deposit guarantee dependent on the level of encumbrance and by increasing the transparency of the banks' asset encumbrance.
- Segendorf and Jansson investigate the efficiency of the Swedish payment system. They show that the use of cash or debit and credit cards in Sweden is relatively efficient. However, the authors' assessment is that Swedish consumers use cash more often than is socially optimal. There may therefore be good reasons to encourage consumers to use debit cards to a greater extent. One way to do this is introduce transparent and cost-based fees for different types of payment. It may be necessary to compensate the consumers in order to make it easier for them to accept such fees, for example by offering better interest rates on current accounts or lower prices in the shops. It may also be necessary to complement such fees with simple and targeted information.

Read and enjoy!

Claes Berg and Joanna Gerwin

Contents

- The Riksbank's monetary policy measures during the financial crisis – evaluation and lessons learnt **8**

Heidi Elmér, Gabriela Guibourg, David Kjellberg and Marianne Nessén

What monetary policy measures – conventional and extraordinary – did the Riksbank take during the financial crisis 2008-2009 and what effects did they have? In this article we describe the sequence of events during which the repo rate was cut by 4.5 percentage points over the course of six months, and when the Riksbank provided loans to Swedish banks that amounted at most to just over 9 per cent of GDP. We also explain why the Riksbank took these measures and analyse their effects. Our analysis shows that these measures had the desired effect. Both money market and bond yields as well as variable mortgage rates declined and monetary policy expectations fell. An important experience of the financial crisis is how important it is for the Riksbank to be prepared to quickly take appropriate extraordinary measures when the need arises.

- Monetary policy and unemployment: A conceptual review **32**

Mikael Carlsson

The relationship between monetary policy and unemployment is a frequently-discussed topic. This debate uses a number of unemployment concepts that are considered relevant for monetary policy. However, it is not always easy to understand exactly what is meant by these concepts, or what relationship they have to monetary policy. This article therefore discusses which concepts are relevant on the basis of modern theory and how they differ from older academic literature. Based on the main theories on unemployment the article first discusses concepts dealing with long-run unemployment. After that we focus on short-run concepts and here particularly the relationship between short-run variations in unemployment and inflation. An important cause of confusion in the debate, apart from unclear concepts, is that the new and older reasoning regarding the relationship between unemployment and inflation in the short run are not compatible with one another. For example, it is shown in this article that what drives inflation in one reasoning does not do so in the other.

■ The Riksbank's business survey – a quick indicator of economic activity **53**

Jyry Hokkanen, Tore Melin and Alexander Nilson

Like several other central banks, Sveriges Riksbank conducts its own business survey. The aim of the survey is to gather up-to-date information as a basis for an assessment of the current state and development of the economy ahead of the publication of official statistics. The structure of the business sector in Sweden, with only a few predominant companies in each sector, makes it possible for the Riksbank to cover the development of a large part of the economy by primarily turning to these large companies. The companies' responses are also used to forecast the development of GDP.

■ Asset encumbrance and its relevance for financial stability **67**

Reimo Juks

Increased asset encumbrance in a number of countries has raised questions about the underlying driving forces and consequences of such trends for financial stability. This article argues that some increase in asset encumbrance is natural in distressed market conditions where investors demand more safety and there is an overall lack of safe assets. Yet asset encumbrance may be also driven by banks' strong private incentives to tilt their financing towards secured funding, with adverse effects on financial stability. The banks' failure to internalise liquidity effects, rating arbitrage, risk-insensitive deposit pricing, upcoming regulatory reforms and implicit government guarantees are leading banks to ignore some of the social costs associated with secured funding. Regulatory interventions should aim to encourage the issuance of unsecured funding and eliminate or mitigate regulatory imperfections. Among the recommended policy actions are increasing the transparency of asset encumbrance and changing the deposit insurance system to reflect asset encumbrance, as these measures do not require the assessment of a socially-optimal level of asset encumbrance that varies over time and across countries.

■ **Cards or cash. How should we pay? 90***Björn Segendorf and Thomas Jansson*

Inexpensive and easy-to-use payment services promote trade and prosperity. The efficiency of the payment system is thus an important social issue. In this article we give a brief account of the social costs of cash and card payments and discuss how the payment system could be used more efficiently. We find that cash and cards are used relatively efficiently in the Swedish payment system. However, there is potential for further reducing payment costs by using debit cards to a greater extent than today. The payment system could also be used more efficiently if transparent and cost-based charges were introduced. Consumers may find it easier to accept such charges if they are compensated for them and provided with simple and targeted information.

The Riksbank's monetary policy measures during the financial crisis – evaluation and lessons learnt

HEIDI ELMÉR, GABRIELA GUIBOURG, DAVID KJELLBERG AND MARIANNE NESSÉN

Heidi Elmér worked at the Riksbank's Monetary Policy Department until September 2012 and is now employed at Skandia. The other authors work at the Monetary Policy Department of the Riksbank.

What monetary policy measures – conventional and extraordinary – did the Riksbank take during the financial crisis 2008-2009 and what effects did they have? In this article we describe the sequence of events during which the repo rate was cut by 4.5 percentage points over the course of six months, and when the Riksbank provided loans to Swedish banks that amounted at most to just over 9 per cent of GDP. We also explain why the Riksbank took these measures and analyse their effects. Our analysis shows that these measures had the desired effect. Both money market and bond yields as well as variable mortgage rates declined and monetary policy expectations fell. An important experience of the financial crisis is how important it is for the Riksbank to be prepared to quickly take appropriate extraordinary measures when the need arises.

When the US investment bank Lehman Brothers went bankrupt in September 2008, what had started out as a crisis in the US mortgage market became a global financial crisis – the most serious in the western world since the Great Depression of the 1930s. The uncertainty and suspicions regarding creditworthiness meant that financial market participants were unwilling to trade with one another and lend to one another. The crisis meant that asset prices around the world fell dramatically and the financial markets functioned poorly – some submarkets completely ceased functioning. Moreover, risk premiums rose substantially – it became both more expensive and more difficult to obtain loans.¹

This situation presented monetary policy with unexpected challenges. Monetary policy affects the economy through a number of different channels, such as bank and market interest rates, exchange rates and credit granting. Under normal circumstances, changes in the policy rate affect first of all interest rate-setting and the financial conditions in the economy and then after that demand, production and inflation. The Riksbank also publishes a repo-rate path, that is a forecast of the future level of the repo rate that is compatible with the inflation target. In this way, the Riksbank can also attempt to influence

¹ A risk premium is the additional return an investor requires as compensation for taking a higher risk. Risk premiums can be divided up into different parts. The maturity premium compensates for the risk that interest rates will develop unfavourably during the period that an investor owns the instrument. The liquidity premium compensates for the risk of not being able to sell an investment for liquid funds at all or without making a substantial loss. The credit risk premium compensates for the risk that the counterparty will not be able to meet its obligations.

expectations of future repo rates and help to steer interest rate-setting in the economy at longer maturities. It is this chain of effects of repo-rate changes on the economy as a whole that is usually referred to as the monetary policy transmission mechanism.²

During the financial crisis 2008-2009, however, the shocks to the financial markets were so great that the impact of monetary policy was adversely affected. The increased risk premiums meant that the changes in the repo rate had less effect than normal on market rates – the transmission mechanism for monetary policy quite simply was not functioning well enough. The large cuts in the repo rate also meant that the level of the repo rate approached zero, which limited the Riksbank's capacity to use further repo rate cuts if necessary to facilitate credit granting and stimulate the economy. However, further monetary policy stimulation was necessary, as GDP continued to fall heavily at the beginning of 2009.

The Riksbank, like many other central banks, therefore needed to take what are known as extraordinary measures, which were also intended to promote financial stability. Among other things, the Riksbank offered its counterparties loans in SEK at longer maturities and also loans in US dollars, it approved a larger number of securities as eligible assets and extended the circle of counterparties.³ The first loans were granted in October 2008 and after that loans were offered regularly until the end of October 2010.

These measures had several purposes. They were intended to safeguard the liquidity of the Swedish financial system, to make the financial markets function better and to facilitate credit granting and reduce various risk premiums that limited the impact of monetary policy. It is of course difficult in a crisis of the magnitude of the one we experienced in 2008-2009 to draw a clear line between the measures that were taken to safeguard financial stability and those taken for monetary policy purposes. Measures that mainly aimed to promote the functioning of the financial markets have also affected and improved the monetary policy transmission mechanism from the policy rate to the interest rates charged to households and companies. However, one of the extraordinary measures – lending at a low, fixed interest rate at a long maturity – had a clearly-stated monetary policy purpose.

In this article we focus on analysing the purpose and effects of the Riksbank's crisis measures from a monetary policy perspective. We begin by describing the monetary policy measures, both conventional and extraordinary, taken by the Riksbank during 2008-2010. After that we discuss the effects that the crisis measures implemented for monetary policy purposes were expected to have. In the third part of the article, we evaluate the actual effects of the three large loans issued by the Riksbank in July, September and October

2 See Apel and Claussen (2012) for a more detailed description of transmission channels.

3 The monetary policy counterparties are the institutions that are allowed to use the Riksbank's credit facilities. A monetary policy counterparty must be a credit institution with a registered office or branch in Sweden that is a participant in RIX, the Riksbank's system for large-value payments. See Sellin and Åsberg Sommar (2012) for a more detailed description of the various counterparty categories that had access to the Riksbank's operations and extraordinary measures. A complete list of the Riksbank's counterparties is available on the Riksbank's website, www.riksbank.se.

2009 on different market rates, the rates charged to households and companies and on the exchange rate. We then move on to describe the process of phasing out the extraordinary loans and the effects it had. The final part of the article discusses in brief the lessons the Riksbank has learnt from its experiences of extraordinary measures.

THE FINANCIAL CRISIS HIT THE MARKETS HARD

The first phase of the financial crisis began in summer 2007 when housing prices in the United States began to fall. This had a negative impact on the financial institutions that had funded a large volume of mortgages. It was difficult to assess the size of the loan losses that might arise from the US mortgages when house prices fell and this created uncertainty among investors and banks.

In the middle of September 2008 the financial crisis reached full-scale impact on both Sweden and the rest of the world. When the US investment bank Lehman Brothers filed for bankruptcy, the market participants' lack of confidence in the creditworthiness of their counterparties became acute. Financial agents in many countries were affected, not just in the United States. This meant that access to credit on the financial markets declined around the world and some markets more or less ceased functioning. Many financial institutions that had earlier funded themselves cheaply through short-term loans now had problems renewing these loans, and if they obtained new loans, these were much more expensive than before. The basis spread – the difference between a three-month interbank rate and the expected overnight rate in three months' time – rocketed sky high (see Figure 1).⁴ The international economic downturn that had already begun worsened. The downswing in world trade was particularly dramatic; it was initially larger than the decline in the Great Depression of the 1930s.

Extensive crisis measures

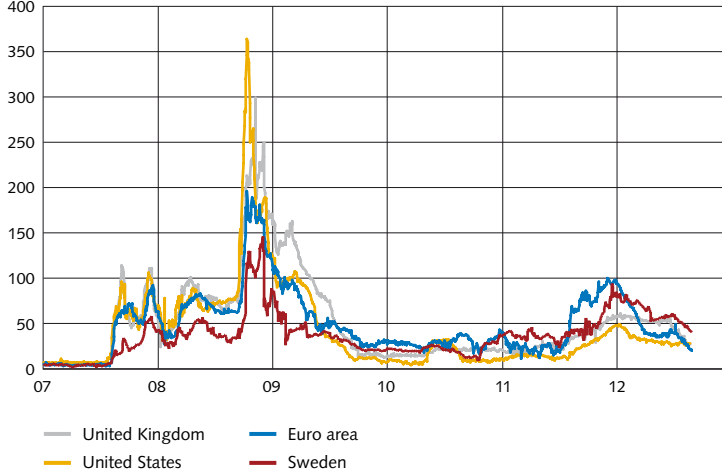
Public authorities around the world began to implement strong measures to stop the negative trend. In Sweden, the Riksbank, the Government and other Swedish authorities took a number of measures to alleviate the effects of the international financial crisis and to improve the functioning of the financial markets. Participants in the Swedish markets were also affected by the general shortage of funding in US dollars that arose after Lehman Brothers' collapse. At the end of September the Riksbank therefore began to offer loans in US dollars to Swedish banks, partly with the aid of funds from the foreign currency reserve and partly via a temporary lending facility offered by the Federal Reserve to the Riksbank and other central banks.⁵ Soon after that the Riksbank also began to lend Swedish kronor

4 The interest difference given by the basis spread measures the investors' preferences to hold liquid funds in relation to fixing them for a period of three months. The rising interest difference means that the risk premium that reflects liquidity and credit risks has increased.

5 As the loan market in dollars is global, the Federal Reserve offered other countries' central banks the opportunity to mediate dollar loans to their counterparties. This led to liquidity assistance in US dollars reaching more parts of the global financial system and helped counteract the on-going liquidity crisis. See page 23 of the Riksbank's Annual Report 2009 for more information on the Riksbank's lending in foreign currency.

to facilitate the banks' slightly longer-term funding. For example, the banks were offered loans in SEK with maturities of three and six months. Ultimately, the Riksbank's measures were aimed at facilitating the supply of credit in Sweden and reducing the various risk premiums that were limiting the impact of monetary policy on the interest rates the banks charged companies and households.

Figure 1. Difference between interbank rates and expected policy rates (basis spread), basis points



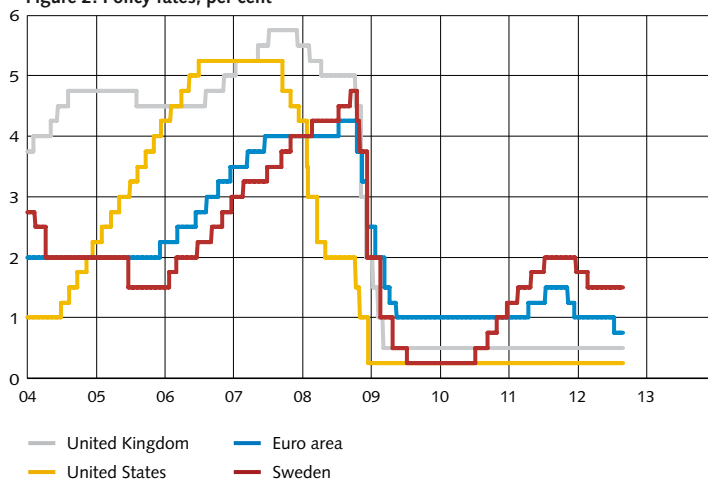
Note. The difference is calculated by subtracting the three month overnight index swap from the three month interbank rate.

Sources: Reuters EcoWin and the Riksbank.

Policy rate cut and fixed-rate loans introduced

At the same time, there was also a radical change in the traditional monetary policy conducted in many countries around the world (see Figure 2). On 8 October 2008, the central banks in the United States, the United Kingdom, the euro area, Canada, Switzerland and Sweden cut their policy rates by 0.5 percentage points in a joint action. Further cuts followed as international economic activity continued to deteriorate and had a severe impact on the Swedish economy. Sweden was hit hard by the global recession and the ensuing collapse in world trade, as the Swedish economy is strongly export-dependent. After having fallen by almost 4 per cent during the final quarter of 2008, GDP in Sweden fell by a further 5 per cent in 2009.

Figure 2. Policy rates, per cent



Source: Reuters EcoWin.

Between October 2008 and July 2009 the repo rate was cut by a total of 4.5 percentage points, to the all-time low of 0.25 per cent. In July 2009, the Riksbank announced its intention to allow the repo rate to remain at this low level for more than one year, until autumn 2010. At the same time, the Executive Board of the Riksbank decided on the first loan motivated by monetary policy: the banks were offered SEK 100 billion in loans at a fixed, low interest rate with a maturity of around 12 months (we will refer to these loans hereafter as the fixed-rate loans). Behind this decision was an assessment that the monetary policy toolbox needed to be supplemented with extraordinary measures in a situation when the repo rate was approaching its lower bound and there was still a need for monetary policy stimulation. The Riksbank’s intention was that these measures would help reduce the large difference between the repo rate and the market rates charged to households and companies during this period, and in this way contributed to monetary policy having the intended effect.

In total, the Riksbank offered three fixed-rate loans with a maturity of around one year, totalling SEK 100 billion each during the period July to November 2009. The loans were allocated through interest bidding, that is, the Riksbank’s counterparties could bid for the loan by stating both the requested volume and interest rate. Technically, the Riksbank thus did not determine the fixed interest rate. However, the Riksbank did stipulate a minimum level for the interest rate that was equal to the repo rate with an addition of 0.15 percentage points, that is, 0.40 per cent. Although the counterparties showed a great interest in the loans offered – two of the three auctions were oversubscribed – the allocation interest rates nevertheless only were a few basis points above the lowest possible bid rate.⁶ The Riksbank lent a total of SEK 296.5 billion in the three auctions of fixed-rate

⁶ The allocation interest rates were 0.45, 0.43 and 0.40 per cent.

loans. The total outstanding loan volume in Swedish kronor at both fixed and variable interest rates amounted at most to 9 per cent of GDP.

THE FIXED-RATE LOANS INTENDED TO SUPPORT MONETARY POLICY ...

The financial crisis meant that the transmission channels – primarily the interest rate channel – did not function normally. The fixed-rate loans were intended to counteract this and to contribute to monetary policy having the intended effect. To simplify, the expected effect of the fixed-rate loans on different market rates can be divided into two components: *liquidity effects* and *signalling effects*.⁷

... through liquidity effects ...

The Riksbank's fixed-rate loans in practice had two purposes: to increase the supply of funding for the bank system and at the same time reduce the banks' funding costs by making it cheaper for them to borrow from the Riksbank than via the market.

With the fixed-rate loan the Riksbank thus created a large surplus supply of kronor in the banking system, which was intended to have a number of different effects on market rates. For example, the increased access to relatively long-term and extensive funding was intended to reduce the banks' refinancing risk, that is, the risk that a bank that needs funding will not be able to obtain it through the capital markets. This could push down liquidity premiums and thus contribute to lower market rates, including lower lending rates to households and companies.

Another effect that was expected to ensue from the fixed-rate loans was that the banks would benefit from the arbitrage-like possibilities that these loans offered. By borrowing cheaply from the Riksbank and investing the money in, for instance, debt securities with a corresponding maturity but a higher interest rate (such as bank and corporate certificates, or mortgage bonds with maturities of up to one year), a bank could obtain a positive return. This type of investment could contribute to pushing up the price of these debt securities and consequently lead to the interest rate falling. Prior to the first fixed-rate loan, this type of potential arbitrage effect was expected to arise mainly in short-term government and mortgage bonds.

... and signalling effects

The fixed-rate loans were implemented as a complement to the traditional monetary policy that entailed cutting the repo rate and at the same time signalling that the repo rate would be kept at a low level over the coming year. They were thus aimed at increasing confidence in the Riksbank's intention to really conduct an expansionary monetary policy until autumn 2010. One could say that the Riksbank "invested money" in its own forecast for the repo rate by lending money at a low and fixed interest rate with a maturity of one year. To

⁷ A similar type of division of the effects can be found in other studies of various types of quantitative easing. See, for instance, Bauer and Rudebusch (2011).

make it easier to understand this reasoning, we first need to explain how the Riksbank's operational framework functions.⁸

The amount of kronor in the economy can only be regulated by the Riksbank, as the payment system is a closed system. If the Riksbank supplies liquidity in the form of loans to the banks, a surplus arises in the payment system (given that the system is in balance to start with). The Riksbank intends to balance the payment system every day, which is achieved through various market operations. The aim is to create good conditions to attain a stable overnight rate close to the Riksbank's repo rate. A surplus is balanced through weekly issues of Riksbank certificates or through daily fine-tuning transactions. The banks can thus choose to deposit their surplus with the Riksbank either in Riksbank certificates at the repo rate or in the daily fine-tuning operations at the repo rate minus 0.1 percentage points.

For the Riksbank the fixed-rate loans could have meant a loss if the repo rate was raised, as the surplus the loans created needed to be regularly drained off by means of the Riksbank's ordinary market operations, where the Riksbank pays an interest rate linked to the repo rate. However, the income from the fixed-rate loans was tied to the allocation interest rates in the auctions during the duration of the loan. As long as the Riksbank held to its forecast for the interest rate and held the repo rate unchanged at 0.25 per cent, the Riksbank was assessed to make a small profit, as the interest rate on the fixed-rate loans was around 0.4 per cent (that is, the Riksbank would have a positive net interest income). However, if the Riksbank were to abandon its forecast and raise the repo rate earlier than signalled, this could lead to losses for the Riksbank. In this way, the fixed-rate loans could strengthen the Riksbank's signalling that it was serious about holding the repo rate at a low level for a long period of time.

During spring 2009 monetary policy expectations, measured both as forward rates and according to surveys, were higher than the prevailing repo-rate path.⁹ The Riksbank expected that the decision to offer fixed-rate loans would lead to confidence in the repo-rate path strengthening and to monetary policy expectations falling. This would mean that the fixed-rate loans contributed to forward rates and other market rates remaining low or falling further and ultimately easing the financial conditions for companies and households.

THE FIXED-RATE LOANS HAD EFFECTS ON INTEREST RATES

We have used a method usually called "event study" to measure the quantitative effects of the three fixed-rate loans.¹⁰ The same type of evaluation method has also been used

8 For a detailed description of the Riksbank's operational framework, see Nessén, Sellin and Åsberg Sommar (2011).

9 There are different ways of measuring monetary policy expectations. One can use surveys to obtain answers on market agents' expectations directly. Another way is to use forward rates which are the market pricing of the monetary policy expectations.

10 See the appendix for a more detailed description of the evaluation method.

to study the effects of other central banks' extraordinary measures.¹¹ A first stage in the study is to identify occasions when new information on the fixed-rate loans became available. Such information has primarily come from the Riksbank's own announcements of its decisions, auction terms and auction results. But other news that might affect market agents' expectations of the Riksbank's measures may also have been significant. When a list of these occasions has been drawn up, a second stage involves analysing the list to see how various market rates and the exchange rate have developed at these points in time and in this way it is possible to quantify and summarize the liquidity and signalling effects of the fixed-rate loans on the financial markets.

The study focuses on measuring the impact of the fixed-rate loans on a number of different market rates, including forward rates, interbank rates and bond rates. Changes in interest rate differences (also called spreads), such as the difference between a two-year mortgage bond and a government bond, have also been studied.

It is not possible to entirely isolate the effects of the fixed-rate loans from other factors. The loans were part of the total monetary policy assessment, which also included the repo-rate decision at that time and the Riksbank's own repo-rate forecast (repo-rate path). Despite the measurement problems this entails, we have nevertheless assessed that the Riksbank's study provides a reasonable estimate of the interest rate effects the three fixed-rate loans caused.¹²

All in all, our analysis shows that the fixed-rate loans pushed down the short-term market rates by at most around 20 basis points. For example, the interbank rates and interest rates on bank certificates and commercial paper declined, which made short-term funding cheaper for both financial and non-financial companies. A study that quantifies the effects of the European Central Bank's (ECB's) fixed-rate loans in 2009 shows similar effects (ECB Monthly Bulletin November 2009).

The analysis also shows that interest rates on more long-term interest-bearing instruments, such as government and mortgage bonds, fell when the Riksbank published a repo-rate path with a forecast for a low repo rate over a long period of time and at the same time announced a fixed-rate loan. The largest effects observed were on bonds with maturities of up to two years, where rates fell in total by at most 40 basis points. Lower policy rate expectations and lower liquidity risk premiums appear to have been some important factors behind the lower bond yields, but it is difficult to completely isolate the individual effect from the fixed-rate loans as other simultaneous events may have affected the Swedish money and bond markets. On the other hand, the announcement of the fixed-rate loans appears to have had fairly little effect on the exchange rate.

11 See, for example, Gagnon et al. (2010) and Krishnamurthy and Vissing-Jorgensen (2010) for estimates of the effects on US asset prices of the Federal Reserve's various quantitative easing measures. See the article "Central Bank Communication in Periods of Heightened Uncertainty" in the ECB Monthly Bulletin November 2009 for estimates of the effects on financial markets of the European Central Bank's fixed-rate loans.

12 For further discussion of the measurement problems, see the appendix.

Money market rates fell

The money market is the market for short-term loans with maturities of up to one year. We have analysed the effects on interbank rates by studying the so-called STIBOR rates at different maturities.¹³ These interbank rates are used to a great extent as reference rates for many types of financial contract, for instance, for many bank loans and pricing of loans through commercial papers. This gives the potential effects of the fixed-rate loans on interbank rates an even broader impact that can contribute to lower interest rates for both companies and households (see below).

The interest rate changes for interbank rates with different maturities are shown in Figure 3. We report two different estimates of the total effect, one where we include all nine events that are linked to the fixed-rate loans (red bar) and one where we only include the three occasions when the respective fixed-rate loans were announced for the first time (blue bar). The general result for the study as a whole is that most of the effects arose at the time of the three announcements of the respective fixed-rate loans.

The interbank rate that has a maturity from tomorrow until the next day is called tomorrow-to-next (T/N) and has a very high correlation with the Riksbank's repo rate. In total, the T/N rate declined by 0.33 percentage points between June and December 2009, that is, during the period when the Riksbank offered its three fixed-rate loans. 0.25 percentage points of the total decline in the T/N rate can be attributed to the repo-rate cut of 0.25 percentage points in July 2009. This means that the other 0.08 percentage points were probably linked to the strong increase in the banking system's liquidity as a result of the fixed-rate loans.

The more long-term interbank rates with three and six month maturities fell by a total of 0.30 and 0.37 percentage points respectively in connection with the relevant announcements. To adjust for the effect of the repo-rate cut on 2 July and to focus on the liquidity effects of the fixed-rate loans, we need to analyse the various spread measures, the basis spread and the FRA-RIBA spread, which are shown in Figure 3.^{14, 15} These measures give the difference between the interbank rate and the average expected

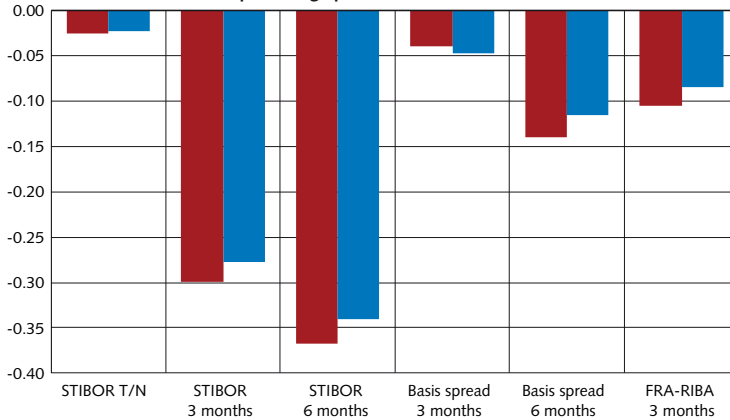
13 It is difficult to find a reliable, precise and frequent data material for interest rate listings on short-term credit markets other than for STIBOR. The STIBOR rates are five banks' (during this period they were six banks) daily assessment of the current interest rate situation for non-secured loans at various maturities on the interbank market. Although the STIBOR rates are not common market rates, and have their limits with regard to exactly reflecting actual interest rates for non-secured loans, the changes in STIBOR rates nevertheless follow fairly closely the changes in most interest rates on related markets.

14 The basis spread is the difference between the STIBOR interbank rate and the STINA rate with the same maturity. We analyse maturities of three months and six months. STINA is an interest rate derivative contract where two parties exchange a fixed interest rate flow and a variable interest rate flow respectively with one another. The variable interest-rate flows are based on the interest rate for the short term tomorrow-to-next, which is closely-related to the Riksbank's repo rate. The market-listed fixed interest rate in the STINA contracts reflects the average expected overnight rate during the term of the contract.

15 FRA and RIBA are two different types of forward contract. In a forward contract two parties agree on an interest rate at a point in the future. The time interval and the price (interest rate) in the contract are determined when the contract is entered into and will depend to a great extent on the counterparties' interest rate expectations. An FRA (Forward Rate Agreement) contract refers to a forward contract for a future three-month STIBOR, while a RIBA contract refers to future repo rates. See the note to Figure 3 for a definition of the basis spread.

overnight rate during the same time to maturity. This difference represents the lender’s interest compensation for lending surplus liquidity (compensation for liquidity risk) and for lending to a borrower that might default on payment (compensation for credit risk). The spread measures declined by around 0.04-0.14 percentage points at the times when the fixed-rate loans were announced. From June to December 2009 these spreads declined by 0.19-0.36 percentage points, but during such a long time period other factors than the fixed-rate loans may also have played a role. For example, the corresponding spreads in other countries also declined during this period, which indicates that the global financial risk declined then. This probably contributed to risk in the Swedish banking system also declining after being at a higher level throughout the global financial crisis. Nevertheless, the announcement of the fixed-rate loans appears to have contributed to bringing down short-term interest costs for banks and companies that use market funding.

Figure 3. Aggregate effect on short-term interest rates and spreads after announcements about fixed-rate loans, percentage points



- Effects from all announcements related to the three fixed-rate loans
- Effects from announcements of the decisions to offer the three fixed-rate loans

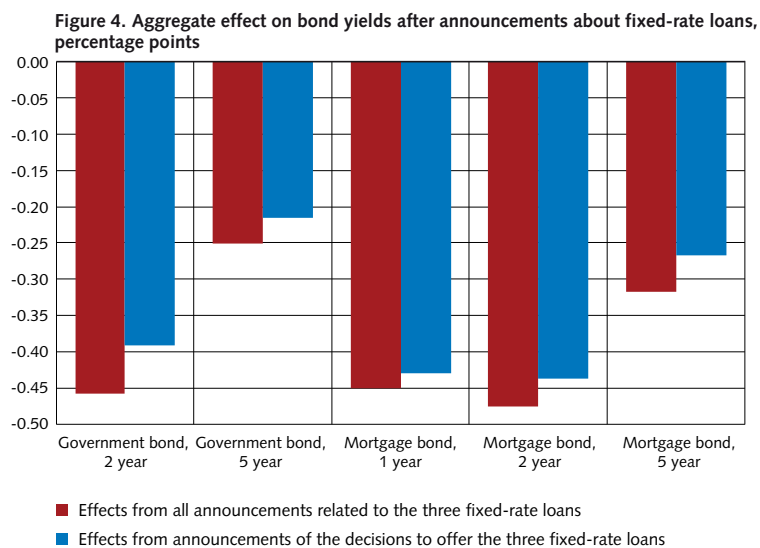
Note. Basis spread is the difference between STIBOR interbank rate and STINA rate. FRA-RIBA spread is the difference between the shortest forward contract for STIBOR 3 months (an FRA contract) and the corresponding forward contract for the repo rate (RIBA). The spread reflects the expected interbank spread for a horizon of up to three months. News related to the fixed-rate loans refers to announcements of new loans, terms for loans or auction results. See Appendix Table A1.

Source: The Riksbank.

Bond yields also fell

The fixed-rate loans were also expected to have some effect on some of the long-term interest-bearing securities, such as government bonds and mortgage bonds. As the fixed-rate loans had a maturity of one year, it is reasonable to expect that their possible effects were greatest on bonds with maturities of around two years or less. We have therefore analysed changes in both of the bonds' yield levels and different measures of the interest differences (spreads) that could be related to liquidity and credit risks.

Figure 4 shows the total yield changes for government and mortgage bonds when the fixed-rate loans were announced. Government bond yields with a maturity of two years and five years fell by a total of 0.46 and 0.25 percentage points respectively. The repo-rate cut in July 2009 explains, as far as we can judge, at most 0.10-0.20 percentage points of the total change in these yields. The changes in bond yields can probably be explained by both liquidity and signalling effects. However, it is not possible to entirely refine these effects and state how much the respective effects contributed to the total fall in bond yields.



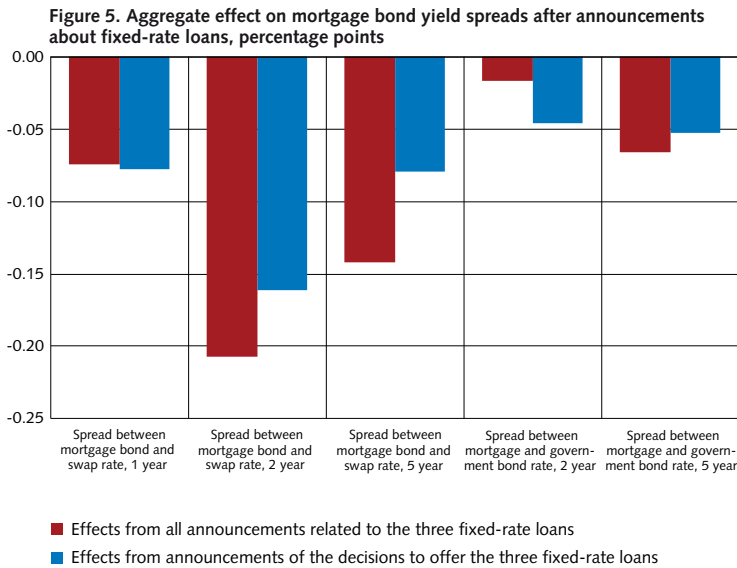
Note. The bond rates are zero coupon rates with a constant maturity, calculated from interpolated yield curves. Mortgage bond rates are based on interest rate data for Stadshypotek covered bonds. News related to the fixed-rate loans refers to announcements of new loans, terms for loans or auction results. See Appendix Table A1.

Source: The Riksbank.

Mortgage bond yields fell slightly more than government bond yields. For maturities of one, two and five years the decline was 0.45, 0.47 and 0.32 percentage points respectively. However, the effects of the fixed-rate loans on mortgage bond yields can have received additional impetus from the ECB's purchases of covered mortgage bonds that was begun on 6 July 2009 and continued over the whole of the following year. Although the ECB's purchases only concerned the markets for covered bonds issued in euro, they may have

made the total funding situation for Swedish banks and mortgage institutions who also obtain funding through issues in euro marginally easier.¹⁶

Figure 5 shows how the interest difference (spread) between mortgage bonds and corresponding interest swaps changed in connection with the fixed-rate loans.¹⁷ The spreads for maturities of one, two and five years declined by 0.07, 0.21 and 0.14 percentage points respectively. Most of the total decline in interest rate differences between mortgage bond yields and swap rates for maturities of one and two years that were noted between June and December 2009 took place in connection with the announcement of the fixed-rate loans (see the blue bar). This decline should also have affected mortgage rates charged to households. The interest rate difference is namely an important measure of the additional cost that banks or mortgage institutions have to pay to convert long-term funding costs to, for instance, a three-month rate and are thus an important factor that affects the mortgage rates charged to households.¹⁸



Note. The bond rates are zero coupon rates with a constant maturity, calculated from interpolated yield curves. Mortgage bond rates are based on interest rate data for Stadshypotek covered bonds. News related to the fixed-rate loans refers to announcements of new loans, terms for loans or auction results. See Appendix Table A1.

Source: The Riksbank.

16 The holder of a covered bond has special preferential rights in the event of a bankruptcy.

17 An interest rate swap is a bilateral agreement to exchange a specific interest rate in return for another interest rate for a predetermined period according to specific conditions. A large share of the banks' and mortgage institutions' mortgages to households have a short-term fixed interest period of three months. This means that the banks have a large share of their income from mortgages based on a three-month interest rate. At the same time, the banks' own borrowing at long maturities is through issues of mortgage bonds. A bank can use an interest rate swap to convert the fixed interest cost to a three-month rate to match its funding costs so that they correspond to the interest income. The bank's total funding costs will then be the three-month interest rate plus the interest difference between, for instance, a five-year mortgage bond yield and a swap rate with the same time to maturity.

18 For a more detailed explanation, see the article "The relationship between the repo rate and interest rates for households and companies" in the Monetary Policy Report, February 2012.

Household mortgage rates fell when funding costs decreased

One aim of the Riksbank's fixed-rate loans was to try to lower the general level of interest rates in the Swedish economy by affecting the market rates that determine the banks' funding costs. Variable mortgage rates fell by 0.42 percentage points, from 2.26 per cent in June before the first fixed-rate loan to 1.84 per cent in December after the last fixed-rate loan, which is more than the 0.25 percentage points by which the repo rate was lowered during the period.^{19, 20} As our analysis shows, the fixed-rate loans helped to hold down, for example, interbank and mortgage bond rates, which led to lower costs for the banks and mortgage institutions for the funding of new household mortgages. Above all, the fixed-rate loans helped to delay the increase in the banks' funding costs. The higher funding cost that became established after 2010 is partly due to the fact that the price of risk is now higher than it was before the financial crisis. For a detailed discussion of this we refer to the article "The latest developments on the short-term interbank market" in the Financial Stability Report 2011:1.²¹

Monetary policy expectations fell

To estimate the signalling effect of the fixed-rate loans, we study the development of the market's pricing of monetary policy expectations.²² Pricing is defined here as market-listed forward rates – or more specifically FRA and RIBA contracts. Figure 6 shows the sum total of changes in interest rates for a range of forward contracts for the three days on which the new fixed-rate loans were announced (2 July, 3 September and 22 October 2009). The figure shows five different forward contracts and specifies the changes in both RIBA and FRA contracts that measure the expected repo rate one to two years ahead.

FRA and RIBA rates can also be affected by factors (so-called risk premiums) other than monetary policy expectations, but for the daily changes we measure we assume that these factors are constant.²³ If we thus assume that the risk premiums in the FRA and RIBA rates

19 See the financial market statistics that Statistics Sweden gathers from financial institutions on behalf of the Riksbank.

20 Mortgage rates are not market rates that react immediately to new information but can be adjusted with a time lag of a few days. It is therefore not meaningful to study the immediate effects of the announcement of the fixed-rate loans on mortgage rates.

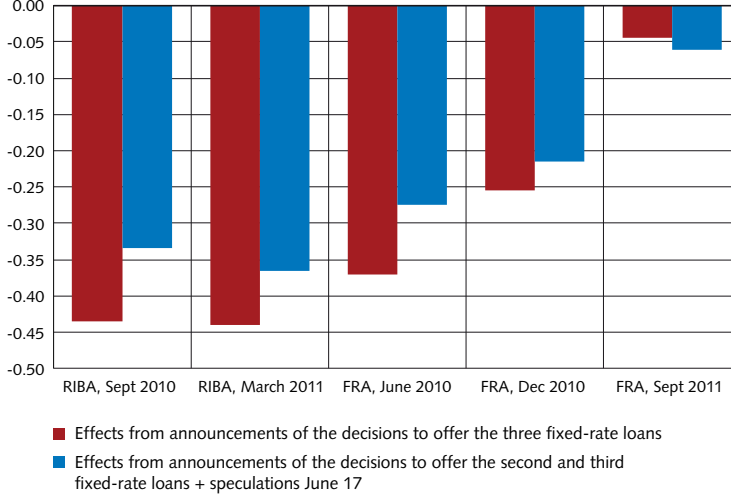
21 For a more detailed discussion of the pricing of mortgages we refer to the article "The relationship between the repo rate and interest rates for households and companies" in the Monetary Policy Report published in February 2012.

22 One way of continuously measuring monetary policy expectations is to use market-listed forward rates. Forward rates are interest rates determined today for investments or loans that will not be executed until a date in the future (settlement date) and which will then run for a predetermined duration. The forward rate can be regarded as the sum of the expected average repo rate from the settlement date to the due date, a maturity premium that compensates for interest rate risk and other risk premiums depending upon liquidity and credit risks. The market's monetary policy expectations are calculated as the forward rate adjusted for these premiums. A less continuous way of measuring monetary policy expectations is to conduct surveys.

23 In the case of FRA rates this assumption could be misleading as they are forward contracts for future STIBOR with a maturity of three months and even future STIBOR can be affected by the fixed-rate loans via liquidity effects. The signalling effect may therefore be overestimated when it is measured in terms of FRA rates. However, the results show the opposite as the changes in the FRA rates are smaller than the changes in the RIBA rates.

were unchanged on the days in question, the result shown in Figure 6 can be interpreted such that the monetary policy expectations one to one and a half years ahead fell by a total of 0.25-0.45 percentage points (see the red bars). However, the total effect on forward rates over a horizon of approximately two years (FRA September 2011) was only a decrease of 0.05 percentage points.

Figure 6. Aggregate effect on forward rates after announcements about fixed-rate loans, percentage points



Note. FRA is a forward rate for STIBOR three-month rate and RIBA is a forward rate for a three-month average of the repo rate.

Source: The Riksbank.

In connection with the first announcement of a fixed-rate loan on 2 July 2009, the Riksbank also cut the repo rate by 0.25 percentage points. This cut was almost 0.20 percentage points more than the market participants had expected on average. Part of the change in short-term forward rates that followed directly afterwards was therefore due to this surprising decision and thus had nothing to do with the fixed-rate loan. In total, forward rates fell by 0.15-0.18 percentage points on 2 July and it is probable that a large part of this change can be related to the unexpected repo-rate cut. It is therefore probable that, as an isolated event, the announcement of the first fixed-rate loan had a rather limited impact on monetary policy expectations. Another factor that contributed to this is that already in mid-June the market participants had begun to speculate about the Riksbank's possibilities to use fixed-rate loans to support a low repo-rate path after the ECB announced fixed rate loans with a maturity of one year in early May 2009. The signalling effect of the fixed-rate loan may thus have been taken into account in the forward rates

already before the announcement on 2 July. For example, forward rates fell by 0.07-0.10 percentage points already on 17 June when evident speculation began.²⁴

If we instead put together the interest rate changes in 17 June with the effects in connection with the announcements of the last two fixed-rate loans we get a potentially better measure of the total signalling effect. This is because we exclude the effects of the repo-rate cut on 2 July and take into account the fact that the first fixed-rate loan may have been partly or wholly expected since 17 June. The total effect is then 0.22-0.37 percentage points, as shown in the blue bars in Figure 5. Irrespective of which of these measurement methods we use, the results show that the fixed-rate loans had noticeable effects on monetary policy expectations.

Small effects on the exchange rate

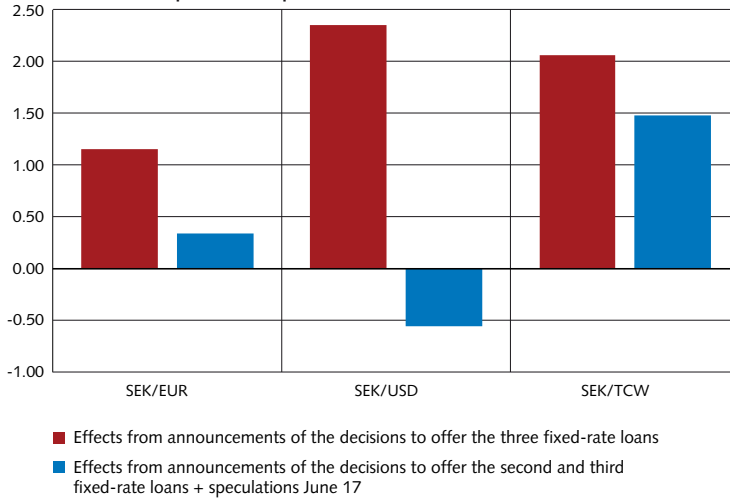
In addition to their effects on market rates, the fixed-rate loans may have had signalling effects on the exchange rate. However, as the foreign exchange market reacts to an extensive flow of information during the day it is difficult to prove that the change in the exchange rate on a day on which the Riksbank announced a fixed-rate loan is related to this announcement.

Our analysis of the changes in the TCW index and in the exchange rate for the krona in relation to the euro and the US dollar shows that that the effect of the fixed-rate loans on the exchange rate was in general not particularly significant.²⁵ However, we can note that the exchange rate for the krona was affected, above all in connection with the announcement of the first fixed-rate loan in July 2009. On this day, the krona weakened by 1.1 per cent against the euro and by 3 per cent against the US dollar. The effect on the exchange rate was not as great when the other two fixed-rate loans were announced, which means that the overall measure shown in Figure 7 is entirely dominated by the effect in July. The weakening of the krona on 2 July 2009 may also have been an effect of the fact that the repo rate was surprisingly cut to 0.25 per cent on the same day. We cannot therefore fully isolate and identify the effects of the fixed-rate loan on the krona in this case. The available data thus indicates that the fixed-rate loans did not have any tangible effect on the krona.

²⁴ We have identified this date on the basis of speculation in market newsletters and a statement by the National Institute of Economic Research that the Riksbank would probably implement quantitative measures of kind in 2009.

²⁵ The TCW index is an index that measures the value of the krona against a basket of other currencies.

Figure 7. Aggregate effect on exchange rates after announcements about fixed-rate loans, krona depreciation in per cent



Note. Positive (negative) effect means that the Swedish krona depreciated (appreciated). SEK/TCW is a weighted currency index for the Swedish krona which weights the exchange rate versus Sweden's most important trading counterparties.

Source: The Riksbank.

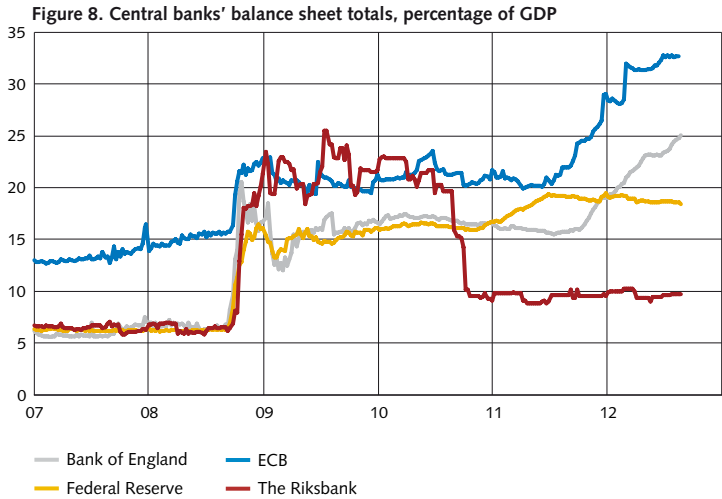
THE RIKSBANK PHASED OUT THE EXTRAORDINARY MEASURES IN 2010

The functioning of the financial markets improved during the course of 2010. For example, access to funding on the private markets improved once again for the Swedish banks and other market participants. The Swedish economy also developed much better than expected. GDP increased by 6.1 per cent in 2010, more than cancelling out the decrease in the preceding year. The upturn was broadly based and could be noted in all parts of the components of GDP. The labour market also developed better than expected and unemployment began to fall.

The Riksbank therefore began to signal its exit strategy, that is its plans for a gradual phase-out of its extraordinary lending. Already in the winter of 2009, the Riksbank began to gradually raise the interest rates on the variable-rate loans that it continued to offer the banks for stability reasons. The maturities of these loans were also gradually shortened in 2010. In this way, the Riksbank was able to gradually prepare the banks for a return to a normal situation without extraordinary lending when the time came to repay the major fixed-rate loans in the summer and autumn of 2010. As the banks' access to market funding improved and the Riksbank's lending became less favourable, the banks chose not to extend their borrowing from the Riksbank. The final outstanding fixed-rate loan was repaid in October 2010, thus bringing to an end the Riksbank's extraordinary lending.²⁶

²⁶ Only a few small variable-rate loans remained outstanding after this. The final outstanding variable-rate loan was repaid in January 2011.

Figure 8 clearly shows that the Riksbank’s extraordinary lending first led to a substantial increase in the Riksbank’s balance sheet and then to a rapid decrease in the autumn of 2010.



In parallel with the phase-out of the extraordinary lending, monetary policy as a whole also began to return to a more normal situation. The Riksbank had left the repo rate unchanged at 0.25 per cent since the beginning of July 2009. In July 2010, the repo rate was raised for the first time to 0.5 per cent and this was followed by another three increases, each of 0.25 percentage points. At the end of 2010, the repo rate was 1.25 per cent.

The Riksbank communicated its exit strategy well in advance. In practice, the extraordinary measures were phased out in a relatively simple way. As the Riksbank’s fixed-rate loans fell due in 2010 and the loans were not renewed, the Riksbank had largely wound up its extraordinary measures. Technically, the Riksbank’s exit strategy was thus uncomplicated, even though the loans that fell due comprised relatively large sums in relation to GDP. Unlike many other central banks, the Riksbank did not use the purchase of securities as a crisis measure, which probably also made it easier to plan and implement the phase out of the extraordinary measures.

UNCERTAINTY CREATED FRICTIONS DURING THE PHASE-OUT OF THE EXTRAORDINARY MEASURES²⁷

When the Riksbank ended its extraordinary lending it was reasonable to expect that interbank rates and yields on the bond market would rise once again. The aim of offering

²⁷ See also the articles “Temporary market turbulence following the phase-out of the Riksbank’s loan facilities” in Financial Stability Report 2010:2 and “The latest developments on the short-term interbank market” in Financial Stability Report 2011:1.

the fixed-rate loans had been to push down these rates by reducing the liquidity-risk premiums. When the final fixed-rate loan fell due in October 2010, interbank rates and mortgage bond rates did indeed increase, but temporary frictions also arose that led to increased volatility on, above all, the short-term money markets. The increased turbulence on the interbank markets also affected other markets for a short while, which resulted in higher market rates. Why did these frictions arise despite the fact that the Riksbank had clearly communicated its exit strategy in advance at the same time as the banks had good access to alternative funding?

Liquidity in the banking systems decreased rapidly when the banks repaid the fixed-rate loans, which primarily affected the most short-term interbank market, that is the overnight market. It was expected that there could be an effect on the interbank market, but the extent of this effect was difficult to assess in advance.

During the period in which the Riksbank provided large amounts of liquidity, trade on the overnight market fell drastically as the banks no longer needed to borrow money from each other overnight. When the loans subsequently matured and liquidity decreased, the banks were once again forced to manage their short-term funding by balancing out their daily surpluses and deficits with each other. During this transition, uncertainty arose about what sums, and at what interest rates, the banks were prepared to lend to each other from one day to the next without collateral. After having been very low and stable in relation to the repo rate since the autumn of 2008, the overnight rate therefore increased. Uncertainty on the overnight market affected other markets and also pushed up the more long-term money market rates.

The increased uncertainty lasted for a few weeks after the final major fixed-rate loan matured in October 2010. After this, the banks agreed on the new conditions that would apply on the overnight market. This reduced the uncertainty about the price of, and access to, overnight loans. This in turn led to a gradual stabilisation of the situation on the other money and bond markets.

The measures hid a change in the risk behaviour of financial agents

Alongside the increased uncertainty, there were other factors that can explain the frictions that arose. The substantial liquidity surpluses that prevailed when the fixed-rate loans were in place had hidden a new phenomenon. The financial crisis had prompted a change in the risk behaviour of the Swedish banks. It is clear that in connection with the outbreak of the crisis the banks became more aware of the risks in the banking system at the same time as their need to maintain substantial reserves of liquid funds as part of their liquidity management process had increased. The effects of this new behaviour did not become clear until the surplus liquidity decreased and the banks once again needed to borrow and lend money from and to each other on the interbank and overnight markets to a greater extent. The consequences of increased risk awareness and larger liquidity buffers were a higher price for liquidity in the form of higher liquidity-risk premiums. Prior to the financial crisis these premiums were very low. It was not surprising that average risk premiums on

the Swedish money and bond markets increased after the phase-out of the Riksbank's extraordinary measures. However, the adjustment to the new situation was probably more volatile than the market participants had expected. Changes in liquidity management and increased risk awareness are not unique to Swedish banks. Market participants abroad have also demonstrated this behaviour and now wish to hold more liquid funds than before the start of the financial crisis.

However, it was not only the change in the banks' risk behaviour with regard to liquidity management that caused the increased volatility on the market. The banks themselves were well prepared to repay their loans to the Riksbank and had access to long-term market funding. But there were a number of other market participants that had benefited from the low and stable interbank rates that the surplus liquidity contributed to. Arbitrage-like possibilities had arisen that enabled these participants to use inexpensive funding at short maturities to invest in more long-term assets, for example mortgage bonds. When the short-term interbank rates subsequently increased and the inexpensive funding was no longer available, this type of investment was no longer favourable and the participants chose to sell their holdings, which led to falling prices and higher interest rates.

It is of course natural for a certain degree of friction to arise as the market adjusts from a situation with great liquidity surpluses in the banking system to a more normal situation. However, the types of friction that arose in connection with the phase-out of the Riksbank's measures clearly demonstrate that all the market participants were not fully prepared for the consequences of the phase-out. This is despite the fact that the Riksbank communicated the phase-out in advance and conducted a dialogue with the banks ahead of the phase-out. One of the lessons learned from the Riksbank's exit strategy is thus that one can expect it to be difficult to foresee the chain reaction that can occur when, after a long period of time, one draws in a substantial liquidity surplus and returns to a more normal way of managing liquidity.

Effective measures require preparation and quick action

The Riksbank's operational framework for the implementation of monetary policy has demonstrated that it works well in both normal conditions and in situations like the recent financial crisis. In normal cases, the Riksbank uses only a few instruments to implement monetary policy. However, during the financial crisis the Riksbank made a number of changes to the framework and took a number of extraordinary measures – primarily lending in Swedish kronor and US dollars to the banks. Other central banks took similar extraordinary measures. The major difference compared to other central banks is that the Riksbank did not purchase securities as a crisis measure. Although the alternative of purchasing debt securities such as government bonds or mortgage bonds was also investigated and discussed at the Riksbank. However, the assessment of the Executive

Board was that the fixed-rate loans in combination with other monetary-policy measures would be adequate to manage the situation at that time and that further measures were not necessary.²⁸

Our evaluation of the effects of the fixed-rate loans shows that during 2009 the loans led to more favourable financial conditions in Sweden. Despite the fact that the repo rate had reached what was perceived as its lowest bound, the Riksbank was able to counteract a weak development of the economy by making monetary policy more expansionary. Money market rates and bond yields, as well as variable mortgage rates, were lower as a result of the fixed-rate loans. Our conclusion is that this type of fixed-rate loan can have an impact on market rates, at least in a situation in which the financial markets are not functioning smoothly.

However, experience from the financial crisis shows how important it is that the Riksbank is prepared to quickly take appropriate extraordinary measures when the time arises. Nor can one rule out the possibility that other measures, for instance, the purchase of securities, might be necessary in the future. A further important lesson is that changes and new measures require some preparation before they can be implemented in practice. For example, this can mean it is necessary to work out new routines, agreements and systems. It is thus an advantage if the Riksbank has an infrastructure in place in normal times so that the need for a long period of preparation is reduced in a situation in which there is a need to carry out extraordinary measures quickly. The Riksbank therefore decided in spring 2012 to supplement the monetary policy toolbox by once again acquiring a limited bond portfolio in Swedish kronor. This measure meant that the Riksbank ensured it had the systems, routines, and knowledge needed to be equipped to implement extraordinary measures at short notice in the future.

The winding up of the crisis measures also provided important experience. The Riksbank communicated its exit strategy in good time, in the hope of preparing the banks for a situation without liquidity assistance. As the banks gained better access to market funding once again and the loans from the Riksbank became more expensive and were offered at shorter maturities, the banks were encouraged to let the loans with the Riksbank mature. The banks' transition from extensive liquidity assistance to full market funding thus took place gradually and was relatively undramatic.

The volatility on the money and bond markets that arose after the Riksbank wound up its extraordinary measures in autumn 2010 lasted for a few weeks before the situation stabilised. However, the short-term money market is still, in autumn 2012, showing higher volatility and higher interest rates than prior to the financial crisis. This is probably partly due to the banks becoming more aware of the risks in the banking system, at the same time as their need to hold larger reserves of liquid funds as part of their liquidity management have increased. Moreover, the current debt crisis in Europe has contributed to higher risk premiums and increased market unease, which also affects the Swedish money and bond

28 See the minutes of the Executive Board's monetary policy meeting on 20 April 2009.

and foreign exchange markets. There are thus several reasons why risk premiums are still higher than they were prior to the outbreak of the financial crisis. Nor does it appear likely that they will return to the low levels we saw before the financial crisis broke out.²⁹

It should be emphasised that the Riksbank chose to phase out its extraordinary measures during a period of relatively favourable conditions. The situation on the financial markets had improved, the Swedish economy was recovering at a good pace and other central banks were still providing the markets with large liquidity surpluses. An important experience of the Riksbank's work with its exit strategy is thus that despite the preparations it may be difficult to foresee the full consequences of phasing out extraordinary measures.

²⁹ New regulations also play an important role for future risk premiums. See, for instance, the article "Basel III – Effects on the Swedish banks and Sweden" in the Financial Stability Report 2010:2.

References

Annual Report, 2009, Sveriges Riksbank.

Apel, Mikael and Claussen, Carl Andreas, 2012, Monetary policy, interest rates and risk taking, *Sveriges Riksbank Economic Review*, 2012:2, Sveriges Riksbank.

Bauer, Michael and Rudebusch, Glenn, 2011, The Signaling Channel for Federal Reserve Bond Purchases, *Working Paper* 2011-21, Federal Reserve Bank of San Francisco.

Beechey, Meredith, and Elmér, Heidi, 2009, "The lower limit of the Riksbank's repo rate", *Economic Commentary* no. 11, 2009, Sveriges Riksbank.

ECB Monthly Bulletin, November 2009, Central Bank Communication in Periods of Heightened Uncertainty, pp. 71-79, ECB.

ECB Monthly Bulletin, October 2010, The ECB's Response to the Financial Crisis, pp. 59-74, ECB.

Eklund, Johanna and Åsberg Sommar, Per, 2011, "The Swedish market for balancing liquidity between the banks overnight 2007-2010", *Sveriges Riksbank Economic Review*, 2011:1, Sveriges Riksbank.

Elmér, Heidi, Sellin, Peter and Åsberg Sommar, Per, 2012, "The framework for the implementation of monetary policy and experience of the financial crisis", *Economic Commentary*, no. 2, 2012, Sveriges Riksbank.

Financial Stability Report, 2010:2, Temporary market turbulence following the phase-out of the Riksbank's loan facilities, pp. 23-25, Sveriges Riksbank.

Financial Stability Report, 2010:2, Basel III – Effects on the Swedish banks and Sweden, pp. 61-67, Sveriges Riksbank.

Financial Stability Report, 2011:1, The latest developments on the short-term interbank market, pp. 26-30, Sveriges Riksbank.

Gagnon, Joseph, Raskin, Matthew, Remache, Julie and Sack, Brian, 2010, Large-Scale Asset Purchases by the Federal Reserve: Did They Work? *Staff Report* no. 441, Federal Reserve Bank of New York.

Krishnamurthy, Arvind and Vissing-Jorgensen Annette, 2011, The Effects of Quantitative Easing on Interest Rates, *Brookings Papers on Economic Activity*, Fall 2011, Brookings Institution.

Lenza, Michelle, Pill, Huw and Reichlin, Lucrezia, 2010, Monetary policy in exceptional times, *Working Paper* no. 1253, ECB.

Minutes of monetary policy meeting 20 April 2009, Sveriges Riksbank.

Monetary Policy Report, February 2012, The relationship between the repo rate and interest rates for households and companies, pp. 52-56, Sveriges Riksbank.

Nessén, Marianne, Sellin, Peter and Åsberg Sommar, Per, 2011, The framework for the implementation of monetary policy, the Riksbank's balance sheet and the financial crisis, *Economic Commentary*, no. 1, 2011, Sveriges Riksbank.

Sellin, Peter and Åsberg Sommar, Per, 2011, Review of the Riksbank's operational framework for the implementation of monetary policy, *Sveriges Riksbank Economic Review* 2012:2, Sveriges Riksbank.

Svensson, Lars E. O., 2009, Monetary policy with a zero interest rate, speech at SNS.

Söderström, Ulf and Westermark, Andreas, 2009, Monetary Policy with a Zero Interest Rate, *Sveriges Riksbank Economic Review* 2009:2, Sveriges Riksbank.

Appendix: Description of the method to estimate the effects of the fixed-rate loans

A method of the “event study” type studies the immediate price effects of the publication of new information. The method is used in this study to summarise interest rate changes at specific measured points in time that can be linked to days when new information related to the fixed-rate loans was published. The total sum of interest or exchange rate changes during the days is assumed to reflect the total effect the fixed-rate loans have had on financial prices. The same type of evaluation method has also been used to study the effects of other central banks’ extraordinary measures (see, for instance, Gagnon et al., 2010 and Krishnamurthy and Vissing-Jorgensen, 2010). Data for market rates and exchange rates has been gathered from Reuters EcoWin.

We have compiled a list of news that can be linked to the Riksbank’s fixed-rate loans. Such new information has primarily arisen from the Riksbank’s own announcements and statements on the fixed-rate loans, such as the announcement of decisions, auction terms and auction results. The list has also been supplemented with some news about the ECB’s measures, which could also have affected Swedish market rates. However, it appears that these news items do not coincide with any of the measurement points linked to the fixed-rate loans. Table A1 lists all of the dates and events. In total, nine days of news are identified as being directly linked to the Riksbank’s fixed-rate loans and consist of announcements from the Riksbank itself.³⁰

There are some properties and potential problems with the chosen method that need to be taken into account. As decisions on the three fixed-rate loans were announced in connection with the monetary policy meetings, we cannot completely isolate the effects of the fixed-rate loans, as the loans were presented as part of the monetary policy assessment, which also includes the current repo-rate decision and repo-rate path. Furthermore, the overall effects are calculated by measuring the change in market rates during a whole day.³¹ This is a sufficiently long period for the market agents to have time to act on the new information and for the market rates to have time to adapt. But at the same time, it leads to a risk that non-relevant events during the measurement day may affect market rates and thus the results. Studies using a much greater number of observations than in our case assume that disturbances from non-relevant events will be positive and negative by turns, which means they offset one another and do not have a decisive effect on the total interest rate effect. However, we have checked the few observation dates for such exogenous events and our assessment is that these events have not affected the results in any decisive way, despite the small number of observations.

30 This does not include the payment dates for the fixed-rate loans, that is the days when the money was actually paid out by the Riksbank. In terms of signalling, these days are not of interest as all events around these transactions are finally determined and published two days before.

31 One alternative would have been to use intraday data, but there was no such reliable and adequate data available at the time of this study.

The interpretation of the estimated effects should also take into account the fact that some of the effects of the fixed-rate loan could have gradually had an impact over a longer period of time and not just on the actual observation dates. The estimated effect in this study could therefore be less than the total effect of the fixed-rate loan.

Table A1. News and events related to the fixed-rate loans

	Related news and events	News on the fixed-rate loans
07-05-2009	ECB announces 'fixed rate loan' 12 m & purchase of covered bonds	
17-06-2009	Speculation that the Riksbank may offer fixed-rate loans	
24-06-2009	ECB announces results of auction 12 m	
02-07-2009	Riksbank announces fixed-rate loans 12 m + cuts repo rate to 0.25%	*
06-07-2009	ECB initiates programme for purchase of covered bonds	
09-07-2009	Riksbank announces terms for auction	*
13-07-2009	Riksbank announces results of auction	*
15-07-2009	Payment of first loan	
03-09-2009	Riksbank announces new fixed-rate loan 12 m	*
10-09-2009	Riksbank announces terms for auction	*
14-09-2009	Riksbank announces results of auction	*
16-09-2009	Payment of second loan	
22-10-2009	Riksbank announces new fixed-rate loan 12 m	*
29-10-2009	Riksbank announces terms for auction	*
02-11-2009	Riksbank announces results of auction	*
04-11-2009	Payment of third loan	

Note. Dates marked with an asterisk are used to calculate the accumulated effects on interest rates and spreads in Figures 3, 4 and 5.

Monetary policy and unemployment: A conceptual review

MIKAEL CARLSSON*

The author works at the Monetary Policy Department of the Riksbank.

The relationship between monetary policy and unemployment is a frequently-discussed topic. This debate uses a number of unemployment concepts that are considered relevant for monetary policy. However, it is not always easy to understand exactly what is meant by these concepts, or what relationship they have to monetary policy. This article therefore discusses which concepts are relevant on the basis of modern theory and how they differ from older academic literature. Based on the main theories on unemployment the article first discusses concepts dealing with long-run unemployment. After that we focus on short-run concepts and here particularly the relationship between short-run variations in unemployment and inflation. An important cause of confusion in the debate, apart from unclear concepts, is that the new and older reasoning regarding the relationship between unemployment and inflation in the short run are not compatible with one another. For example, it is shown in this article that what drives inflation in one reasoning does not do so in the other.

Many different concepts relating to unemployment are used in the debate on monetary policy. Concepts such as “equilibrium unemployment”, “non-accelerating inflation rate of unemployment” (abbreviated NAIRU) and “natural unemployment” are but a few examples. Furthermore, a distinction is often made between the concepts at different time horizons. These concepts are often considered relevant to monetary policy, but it is rare that more precise information is given with regard to what they mean.¹ As a result of this, there is widespread confusion regarding the concepts in the general debate.

A number of models have been constructed recently that combine ideas from labour market research with the canonical New Keynesian model for monetary policy analysis. One such model is the Riksbank’s own model RAMSES II.² However, the introduction of these partial labour market models into modern macro-economic theory has led to further confusion. This confusion has arisen because the unemployment concepts used in the

* The author would like to thank Claes Berg, Karin Carlsson, Joanna Gerwin, Lars E.O. Svensson, Ulf Söderström, Andreas Westermark and Karl Walentin as well as participants at several presentations for their views and suggestions. The views expressed in the article are the author’s own views and do not necessarily represent the Riksbank’s views on these issues.

1 In the cases where an attempt is made to be more precise, this often merely demonstrates how vague the concept is. A good example of this is given in Rogerson (1997), which points to 12 different varieties of “natural unemployment” in academic research.

2 See Christiano, Trabandt and Walentin (2011).

debate often arise from older theories. For example, it is not clear what role a concept such as NAIKU plays in the modern reasoning on which the New Keynesian models are based.

This article discusses which unemployment concepts are relevant on the basis of modern theory and how they differ from the concepts used in the older academic literature.

We use the main theories on unemployment as a starting point and focus first on unemployment in what is known as *steady state*, that is, the unemployment that prevails when the effects of all of the macroeconomic shocks that can affect unemployment have faded. This discussion also makes it easier to understand the older reasoning on the relationship between unemployment and inflation as presented in Layard, Nickell and Jackman (2005) (LNJ).

In the second part of the article we focus on how the short-run variation in unemployment (around the steady-state level) is related to inflation both on the basis of the older LNJ model from 1991 and on the basis of a New Keynesian model with a labour market block formulated by Blanchard and Galí (2010). We also discuss what are – and are not – relevant concepts of unemployment for monetary policy in the New Keynesian model.

Long-run equilibrium: Unemployment in steady state

An often-used concept in the theoretical analysis of the labour market and of long-run unemployment is “natural unemployment”. This was defined by Milton Friedman (1968, p. 8) in his speech as the President of the American Economic Association as follows:

The ‘natural rate of unemployment,’ in other words, is the level which would be ground out by the Walrasian system of general equilibrium equations, provided there is imbedded in them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on.

The general interpretation of the natural rate of unemployment is that it is a long-run equilibrium that the economy strives to attain over time. This idea has had considerable impact in academic research. However, Friedman’s original definition is so comprehensive that it becomes unclear, and there are thus many different examples of how the concept has been defined and used in academic research (see for instance Rogerson, 1997, for a discussion). This has in turn led to confusion over what is actually meant when talking about natural or long-run unemployment.

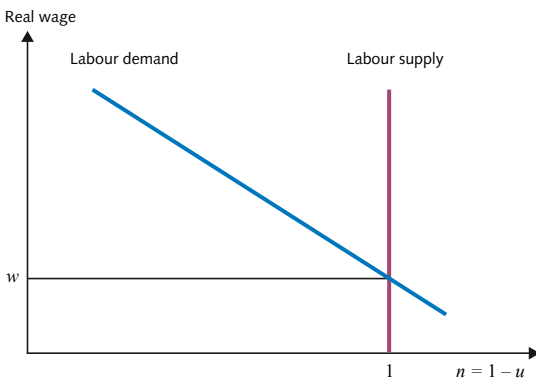
As the concepts in modern theory were not available to Friedman, we cannot know exactly what he thought in relation to them. But this is not really so important. Modern dynamic theory shows that if we let the macro economic shocks in the model fade out, then unemployment moves towards its steady state level. This rate of unemployment is thus the relevant long-run labour market equilibrium in the reasoning that forms the base for modern dynamic analysis.³

³ However, some details are unclear, such as whether this concerns a stochastic or non-stochastic equilibrium or whether or not distortions should be included in the long-run equilibrium concept.

Theories on steady-state unemployment

To gain an idea of the determinants of steady-state unemployment and to obtain some background to the older reasoning on which the LNJ model is based, we can study the theories available to explain unemployment. As a frame of reference we can first use the neo-classical (or Walrasian) labour market with perfect competition. Since labour supply will not play a major role in the latter discussion, we assume that the individuals' labour supply is insensitive in relation to real wages. Moreover, we normalise the size of the labour force to unity. Under competition, the firm will employ workers until the marginal product of labour equals the real wages, which gives an (implicit) demand function for labour. With decreasing returns to scale, the marginal product of labour declines with the number of employees and we obtain a demand function where the demanded quantity (labour) falls if the price (real wages) increases. In this case the equilibrium on a Walrasian labour market is illustrated in Figure 1.

Figure 1. Equilibrium on a Walrasian labour market



In this model the equilibrium wage is thus given at the point where the demand for labour intersects with the labour supply curve. As the labour market clears at this point, that is, all supply finds its demand at the market price, full employment will prevail. This means that if we define an unemployed person as someone who can and wants to accept employment at the current market wage but nevertheless cannot find a job, then the neoclassical theory of clearing markets cannot explain unemployment. Instead, a theory on unemployment must be based on frictions and imperfections.

The theories we have on steady-state unemployment can be divided into two main lines.

The first line is that different factors lead to wage-setting that differs from that on a Walrasian labour market (as in efficiency wage models and bargaining/trade-union models). As wage-setting in these models gives a higher wage than the one that clears the labour market with perfect competition, unemployment will arise in steady state.

The second line is based on the observation that the labour market is characterised by a never-ceasing flow of individuals moving between employment and unemployment. As

the parties on both sides of the labour market (jobseekers and employers with vacancies) differ both with regard to the qualities they have themselves and what they seek from their counterpart, it takes time and resources to fill a job vacancy (search models). If an individual loses his job, it is not likely that he will immediately find a new one. Similarly, an employer cannot expect to immediately fill a job vacancy. This leads to there always being unemployment, even when all of the macroeconomic disruptions have faded and long-run equilibrium prevails.

All of these labour market models (or combinations of them) result in a supply curve, which is known as the wage-setting curve, with a negative relationship between real wages and unemployment.⁴ We will return to the question of what lies behind this negative relationship, but if we take this for granted for now we obtain a wage-setting relationship as

$$w^n - p = \gamma_0 - \gamma_1 u, \quad \gamma_1 > 0, \tag{1}$$

where w^n is (the logarithm of) the nominal wages and p is (the logarithm of) the price level that gives (the logarithm of) real wages as $w^n - p = w$. Further, u is unemployment and γ_1 is a parameter determined by the sensitivity of wage-setting in relation to changes in unemployment. Finally, all factors determining the horizontal position of the wage-setting relationship are summarised by γ_0 . This means that factors that shift the wage-setting relationship horizontally, and give a higher or lower wage for all levels of unemployment, affect γ_0 . As we are interested in steady-state unemployment (with price and wage stability), we assume there that nominal wages (w^n) are set with knowledge of the price (p).

The demand relationship, which is known as the *price-setting relationship*, is usually derived from the firms' optimal pricing decisions in a market with imperfect competition, given the nominal wage.⁵ The result is a generally positive relationship between real wages and unemployment. The price-setting relationship can be expressed mathematically as

$$p - w^n = \beta_0 - \beta_1 u, \quad \beta_1 \geq 0, \tag{2}$$

where we assume in the same way as for the wage-setting relationship that the price (p) is set with knowledge of nominal wages (w^n). Here the parameter β_1 is determined by the sensitivity of the price markup that firms want to make on nominal wages is in relation to unemployment. It is also worth noting that unemployment ($u = 1 - n$) affects the marginal product of another employee and thereby the firms' marginal cost and optimal price.⁶

4 Real wages here refers to the real product wage, that is to say the ratio of the nominal wage to the producer price.

5 Note that a price decision on a market with imperfect competition also determines the firm's demand and production and thus its demand for labour.

6 Technically, β_1 is determined by the degree of return to scale (to factors that can be adjusted). The existence of search frictions also affects β_1 (see appendix B).

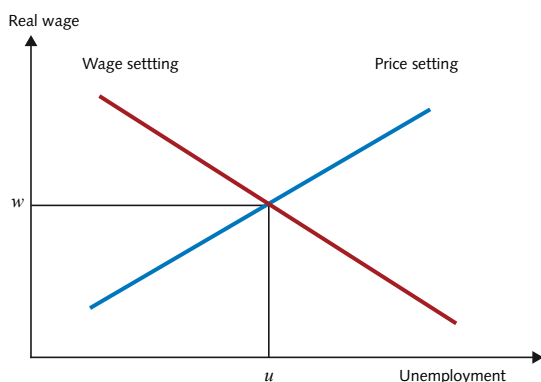
Finally, all factors determining the horizontal position of the price-setting relationship are summarised by β_0 . More specifically, β_0 declines (which reduces the price set for all levels of unemployment) with increased productivity or product market competition, and increases with increased capital costs.⁷

We can also write the price-setting curve in terms of real wages as

$$w^n - p = -\beta_0 + \beta_1 u. \tag{3}$$

Equilibrium unemployment will thus be obtained at the wage at which price-setting and wage-setting decisions are consistent with one another, which is illustrated in Figure 2 (where we have changed the terms on the horizontal axis to focus on unemployment).

Figure 2. Equilibrium unemployment in a general labour market model



What, then, determines the wage-setting relationship, or in other words, why doesn't the labour market clear in steady state? To answer this question, we must go on to analyse the three main models one by one.

EFFICIENCY-WAGE MODELS

The central idea in efficiency wage models is that there are also advantages for a firm in paying a higher wage (than the one that clears the labour market). In, for example, Shapiro and Stiglitz (1984) the advantage is that a higher wage means that the employees exert themselves more at work, even when it is difficult for the employer to monitor them. On a labour market with perfect competition, there is no reason for an employee to be afraid of losing his or her job if they are discovered not to be working, as the employee can immediately find another equivalent job. But if the firm pays more than the wage that

7 As we, for instance, take the capital cost for granted, what is discussed here is a partial analysis of the labour market. It may be good for understanding which mechanisms give rise to steady-state unemployment – as long as we bear in mind that important determinants, which are here taken for granted, are determined in an interplay with other parts of the economy.

clears the labour market, the job will be valuable to the employee. If the job has a value and there is at the same time a risk of being detected not working and of being fired, the employee will want to make an effort at work. This gives an incentive for the firm to set wages at a higher level than that which clears the labour market, and as the value of a job (given the wage) rises if it becomes more difficult to find a new one (increased unemployment), the wage-setting relationship will slope downwards.⁸

TRADE-UNION AND BARGAINING MODELS

In trade-union and bargaining models it is no longer the firm that sets the wage one-sidedly; the wage is instead set in a bargaining process, or in an extreme case with a monopoly trade union where the trade union sets the wage one-sidedly.⁹

If we think of a general bargaining model, one normally lets the parties bargain over the total surplus (see Appendix A for a more detailed review of this type of model). The total surplus will in turn stem from the difference between the value to the firm and to the employee (trade union) of an agreement and the value of the respective party's alternative to an agreement (outside option). In these models wages are usually determined by the total surplus being divided according to the respective party's bargaining power.

If the employee's alternative to an agreement is weakened by an increased unemployment, this will lead to a wage-setting relationship that declines in relation to unemployment. One example is a model where higher unemployment makes it more difficult for the employee to find a new firm to bargain with, which leads to a longer expected period of unemployment benefit if one is not finding an agreement in the bargain. Such a relationship can be found, for instance, in the general equilibrium model used by the Riksbank in its forecasting work and for alternative scenarios: RAMSES II.

Some things to note are that a higher bargaining power for the employee pushes up the wage-setting curve and leads to higher wages and higher equilibrium unemployment (compare Figure 2 with an intended wage-setting curve above the one in the figure). Moreover, if the alternative to an agreement for the employee improves by, for instance, better conditions in the remuneration system for the unemployed, this will also push up the wage-setting curve.

Whether taxes and charges affect the negotiated wages and unemployment depends on how they affect the surplus the parties are bargaining over. Taxes and charges that have a proportional effect on the surplus will not affect gross wages and thus not will they affect unemployment. Appendix A contains one such example where an equally large proportion of both wages and unemployment benefit are deducted as tax. Such taxes and charges will only affect net wages and net gains. Similarly, factors that affect the product wage

8 Another idea on this theme is that a higher wage can attract applicants who are better in respects that the firm is unable to observe (Weiss, 1980), can reduce the staff turnover (Eriksson and Gottfries, 2005), or create a loyalty between employer and employee (Akerlof and Yellen, 1990). See, for instance, Romer (2006) for a longer discussion of efficiency wage models.

9 In the latter case the trade union will strive to attain a high wage, but must also take into account the fact that a higher wage will result in higher unemployment among its members.

to consumer real wage ratio will not affect unemployment (for example, changes in the relative price of imported goods) if the surplus is affected proportionally. Note, however, that a tax change that increases the relative value of reaching an agreement and working for the employee (at a given gross wage), for example through the tax deductions for those in work, will push down the wage-setting curve. This in turn reduces unemployment.

The degree of coordination in the bargaining will also affect the labour market outcome; see Calmfors (1993). For one thing, more centralised bargaining will mean that external effects of wage increases on, for instance, unemployment and inflation, can be internalised in the bargaining. For another thing, the degree of centralisation will affect the possibility to shift the wage costs onto the product price as, for instance, the competition is stiffer between firms within the same sector than it is between sectors.

Finally, it is also worth noting that the way in which one formulates the employee's alternative to an agreement in these models will to a large degree affect the sensitivity of the total surplus, and thereby that of the wages, in relation to unemployment. It makes a considerable difference to this sensitivity, for instance, if the alternative to an agreement is a strike (but the employees stay within the firm), or if the alternative is to seek a new bargaining counterpart in the prevailing economic climate; see Hall and Milgrom (2008).

SEARCH AND MATCHING MODELS

The starting point for search and matching models is that resources are required before employers and job applicants can find one another and the firm can begin to produce goods and services.^{10, 11} These search frictions mean that the firms' incentives to create job vacancies are affected by the number of unemployed who are looking for work. This is because it is costly to keep a post vacant and to look for someone to employ (for instance, the costs of advertising and recruitment services) and because unemployment affects the time it is expected to take to fill a vacancy. This leads to a form of price-setting relationship that, like the one above, is an upward sloping function of unemployment (see appendix B for a more detailed description of search and matching models).

As search frictions mean that an employee, or job, cannot be exchanged for an equivalent employee, or job, without a period of costly searching, a value is created when an employee and a firm meet and bargain. Similar to the bargaining model we have described, the wage is set by the total surplus being divided according to the respective party's bargaining power. The wage-setting curve will once again be a downward-sloping function of unemployment, as higher unemployment will worsen the employee's alternative to an agreement.

¹⁰ See Pissarides (2000) for a detailed description of the search and matching models.

¹¹ There are also search frictions in RAMSES II, but when estimating the model these frictions have proved to have little significance; see Christiano, Trabandt and Walentin (2011). This result is consistent with the findings of Carlsson, Eriksson and Gottfries (2012) in Swedish microdata and with what Yashiv (2000) finds in Israeli macrodata.

As the wage-setting curve in the search and matching models is based on bargaining, many of the conclusions from the bargaining model carry over. For example, the wage-setting curve is pushed up with improvements in the remunerations system for the unemployed, which leads to both a higher wage and higher unemployment.

Note, however, that as the firm's revenue from having an employee is treated as an exogenous variable (or, in other words, as a variable determined outside of the model), the basic search model should also be regarded as a partial equilibrium analysis.

Short-run equilibrium: Unemployment and inflation

In the previous section we discussed the main theories on steady-state unemployment. The next stage is to move on from a theory on long-run unemployment to a theory on short-run variation in unemployment and inflation. For this we need to introduce the partial labour market analysis into a general equilibrium model, which is done in RAMSES II, for instance. This means that, for instance, the firm's revenue from having an employee in the model is no longer exogenous, but endogenously determined on the basis of the general equilibrium system. This means that RAMSES II is part of a new research programme where one has included the partial labour market models in the New Keynesian model, to be able to study outcomes for the labour market and inflation in general equilibrium in a way that agrees with modern dynamic macro theory.¹²

Theories on short-run variation in unemployment and inflation

THE LAYARD, NICKELL AND JACKMAN MODEL

There is also an aim in the older literature to link together the partial labour market models with general equilibrium, and in this way build a theory on short-run variation in unemployment and inflation. An important example of this is Layard, Nickell and Jackman (2005) (LNJ). Their reasoning on the labour market and macro economy is described in the book *Unemployment: Macroeconomic Performance and the Labor Market*, which was first published in 1991. The book has had a very large impact on how economists think about unemployment in general equilibrium and on the relationship between unemployment and inflation both in labour-market research and in the policy discussions. The LNJ model is based on a price- and wage-setting relationship (shown in Figure 2) but does not specify which of the above-mentioned mechanisms has given rise to the downward-sloping wage-setting curve.

Let us write down the price-setting relationship where firms set the price given the expected wage as

12 Other examples are Danthine and Kurmann (2004) and de la Croix, de Walque and Wouters (2009) who introduce efficiency wages, Erceg, Henderson and Levin (2000) and Carlsson and Westermark (2011) who include a monopoly trade union/bargaining model for wage setting, and Gertler, Sala and Trigari (2008) who introduce labour-market search frictions in a New Keynesian framework.

$$p - E_{-1}(w^n) = \beta_0 - \beta_1 u, \beta_1 \geq 0, \quad (4)$$

and the wage-setting relationship where the wage is set given the expected price as

$$w^n - E_{-1}(p) = \gamma_0 - \gamma_1 u, \gamma_1 > 0, \quad (5)$$

where E is an expectations operator and the index -1 indicates that expectations were formed earlier. Thus, $E_{-1}(p)$ is yesterday's expectation of today's price. We can then once again describe equilibrium unemployment in a price and wage-setting diagram as shown in Figure 2. Figure 3a shows these relationships once again, with the difference that we are clear that the relationships are based on prices and wages being realised according to expectations. Steady-state unemployment is thus given as the unemployment that makes price and wage-setting decisions compatible, at the same time as prices and wages are realised according to expectations (that is, price and wage stability).

If an expectation is not realised for some reason, a situation may arise in the short run where unemployment deviates from steady-state unemployment. To derive a short-run supply curve, LNJ further assumes that the "surprises" in prices and wages are equal, which gives the following relationship:

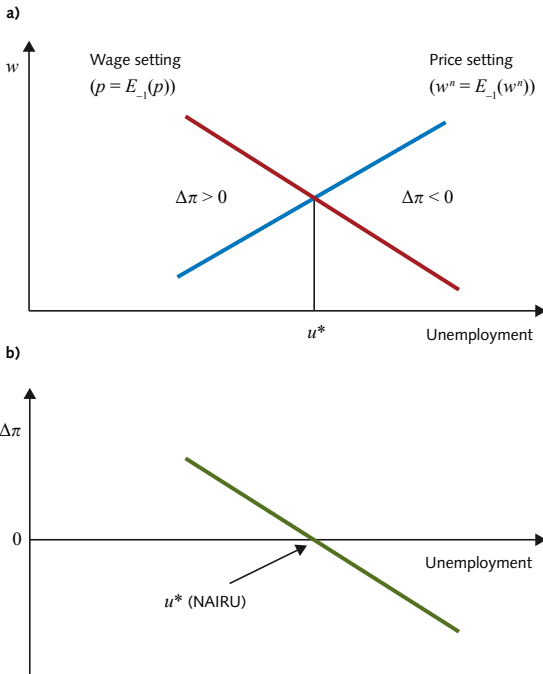
$$u - u^* = -\frac{1}{\vartheta} (p - E_{-1}(p)), \quad (6)$$

where $u^* = (\beta_0 + \gamma_0) / (\beta_1 + \gamma_1)$ is steady-state unemployment and $\vartheta = (\beta_1 + \gamma_1) / 2$ is a measure of wage and price flexibility, that is, how much prices and wages will adjust to changes in the labour market situation (see appendix C for a derivation of equation (6)). Thus, low unemployment is related to positive price and wage surprises. We can then rewrite (6) by adding and deducting the previous period's price level, p_{-1} , as

$$\pi = E_{-1}(\pi) - \vartheta(u - u^*), \quad (7)$$

where $\pi (= p - p_{-1})$ is the rate of inflation. We then see that the LNJ model leads to a traditional Phillips curve with an expectations term. However, the difference is that the LNJ model can give an idea of what determines steady-state unemployment.

Figure 3. Equilibrium in the Layard, Nickell and Jackman model



To move on, a theory is required as to how expectations of inflation are formed. Instead of trying to derive expectations consistent with the model, that is, the expectations that rational agents in the model would form if they knew the properties of the model, LNJ assumes that the agents in the model incorrectly believe that the rate of inflation is what is known as a random walk.¹³ Given that this assumption, the rational, and also the adaptive, forecast is $E_{-1}(\pi) = \pi_{-1}$. If we move the inflation forecast back to the left-hand side, we then get the following relationship:

$$\Delta\pi = -\vartheta(u - u^*), \tag{8}$$

13 That is, a process where a shock today will affect today's inflation and today's expectations of inflation in all future periods in the same way.

which is illustrated in Figure 3b.^{14, 15} Thus, the inflation rate rises when unemployment is lower than steady-state unemployment in the LNJ model. Further, the rate of inflation is constant when unemployment coincides with steady-state unemployment; u^* is thus what is usually known as “the non-accelerating inflation rate of unemployment” (NAIRU).¹⁶ What it is important to note prior to the continued discussion is that equation (8) entails there being one, and only one, level of unemployment that is compatible with a constant rate of inflation, that is steady-state unemployment u^* .

From a policy perspective, it is interesting to note that the Phillips curve in equation (8) means that only surprises in the inflation outcome will lead to variation in unemployment. Also note that the LNJ model assumes that systematic, or predictable, monetary policy does not have any effect on today’s real economic outcome via its effect on the formation of expectations. This assumption contrasts strongly with the New Keynesian model we will discuss in the next section, where the effects of systematic monetary policy on the formation of expectations is an important part of the monetary policy transmission mechanism.¹⁷

To then resolve the general equilibrium model, LNJ assumes a demand relation where aggregate demand, measured as unemployment, is given by

$$u = -\frac{1}{\lambda}(m - p), \tag{9}$$

where m is nominal GDP (adjusted for real trend growth). Exogenous shocks to m thus lead to short-run deviations between the observed and the long-run unemployment and to movements in price and wage inflation. In the long run, however, the economy moves back towards steady-state unemployment.¹⁸

A NEW KEYNESIAN MODEL OF A LABOUR MARKET

The key difference between the LNJ model and New Keynesian models with a labour market block is the degree of formalisation and the further insights this provides. New

14 Note that the assumption that inflation is perceived as a random walk is rather unusual seen from an inflation-targeting policy perspective and that it entails the agents also believing that deviations between the observed rate of unemployment and steady-state unemployment are white noise (that is, a process where today’s outcome does not help to forecast tomorrow’s outcome). One reason why the assumption was common in the academic research was that one then did not need to put one’s foot down in the 1970s debate on adaptive or rational expectations.

15 Equation (8), extended by an error term, has often been used to estimate equilibrium unemployment (as a function of estimated parameters in the regression); see for example Blanchard and Katz (1997). With an error term one can regard equilibrium unemployment as the level of unemployment that systematically gives a constant rate of inflation. However, the obvious econometric problem is the endogeneity bias. This is because both inflation and unemployment are determined in general equilibrium, and one thus cannot say that unemployment causes inflation but not vice versa.

16 A more correct term is “the non-increasing inflation rate of unemployment”, but this is how the term is used in the academic research.

17 For empirical evidence that supports expected monetary policy affecting real economic outcomes, see Mishkin (1982a, 1982b).

18 By taking the time difference of (9) and combining it with (8) we get $u = \frac{1}{\vartheta + \lambda}(\vartheta u^* + \lambda u_{-1} - (\Delta m - \pi_{-1}))$. As discussed by LNJ, u will return to u^* and π_{-1} will tend towards Δm if Δm is constant for long enough.

Keynesian models are based on assumptions of agents' preferences and restrictions (micro foundations) and solutions are based on model consistent expectations instead of ad hoc assumed aggregate relationships and expectations. Further, one can also discuss optimal monetary policy in this model class, that is, the monetary policy that maximises welfare for agents in the model, as the models are derived from the agents' preferences and restrictions. However, this difference in approach can lead to confusion over concepts. A concept that is relevant in modern dynamic theory need not be relevant in the older model without clear micro assumptions (and vice versa).

In a New Keynesian model there are several concepts that are of interest to monetary policy. These include both dynamic concepts and long-run concepts that apply when all of the macro economic shocks have faded:

- i. First is the *observed equilibrium* as shown by data or the short-run equilibrium given all frictions and imperfections in the model.
- ii. The second concept is the equilibrium that would apply if prices and wages could be adjusted without friction, which is known as the *flexible-price equilibrium*. The difference between the observed equilibrium and the flexible price equilibrium, which is usually called the *flexible-price gap*, is a measure of the imbalance that has arisen as a result of the adjustment of prices and nominal wages to new conditions not being immediate (so-called nominal frictions). The flexible price gap is thus the gap that is relevant for determining the underlying inflationary pressure and for forecasting inflation.
- iii. The third concept concerns efficiency – that is, the allocation (or the production and consumption decisions) that a planner would choose if he or she had the task of maximising the welfare of the individual given the resources restrictions and the production technology available. This is called the *efficient allocation*. The deviation between the observed equilibrium and the efficient allocation is the *welfare-relevant gap*, that is, the gap that is relevant when formulating optimal monetary policy and which is included in the central bank's target function in this model.
- iv. When all of the macro economic shocks have faded, the economy will return to the *steady state*, the fourth concept, which is the long-run equilibrium.¹⁹

An article by Blanchard and Galí (2007) shows that in the basic New Keynesian model the flexible price equilibrium and the efficient allocation move together in a “one-to-one” relationship when there are shocks to the economy. In other words, there is no classical

¹⁹ If we assume, which is common in monetary policy analysis, that fiscal policy neutralises all remaining inefficiencies in steady state, the long-run equilibrium will be the same regardless of whether the economy is governed by a planner or whether we have a decentralised solution where individual agencies make decisions in the economy. These remaining inefficiencies can, for instance, concern a firm having market power in pricing decisions. One can then use subsidies to get the companies to make efficient production decisions.

conflict between stabilising inflation and stabilising the welfare-relevant gap, a property that Blanchard and Galí (2007) call “the divine coincidence”.²⁰

However, this coincidence no longer applies if real imperfections, such as real wage rigidities, or inefficient macroeconomic shocks, such as price markup shocks, drive a wedge between the flexible-price equilibrium and the efficient allocation in the model. An example of a model where this coincidence does not occur is Blanchard and Galí (2010) (BG). This model is a New Keynesian model with a simple implementation of a labour market with search frictions. BG shows that the efficient allocation in this model has a constant unemployment rate in all periods. When one introduces wage-setting where the wages are not fully adapted to the variation in productivity (real wage rigidity) we obtain a wedge between the flexible-price gap and the welfare-relevant gap for unemployment, which annuls the “divine coincidence”. A key result from BG is that the stabilising of inflation, which in turn means that unemployment is stabilised at the flexible-price equilibrium, does not give rise to a constant unemployment.

What does this mean for a concept such as NAIRU? Here it is important to remember that the LNJ relationship between changes in inflation and the difference between the observed unemployment and steady-state unemployment is given by

$$\Delta\pi = -\vartheta(u - u^*). \quad (10)$$

This means that LNJ thinks of a world where the stabilisation of unemployment at the steady-state level gives a constant rate of inflation. In the steady state of a New Keynesian model unemployment will be equal to steady-state unemployment and inflation will be constant, but outside of the steady state – when the economy is exposed to shocks – this will no longer apply. BG shows that a monetary policy that completely stabilises unemployment at u^* creates a variation in inflation, and a monetary policy that completely stabilises inflation creates variation in unemployment. The NAIRU idea, where the deviation in unemployment from a constant long-run rate of unemployment drives inflation changes, is thus not relevant in the New Keynesian model world, as it is the flexible-price gap that is related to inflation here. Stabilising unemployment at the flexible-price equilibrium does not mean a constant rate of unemployment. On the contrary, it means that unemployment must vary in response to macroeconomic shocks.²¹

It is worth noting that one could calculate the rate of unemployment that is consistent with constant inflation on the basis of a solution of the New Keynesian general equilibrium model. But once again, this level will not be constant when the economy is subjected to shocks.

What does this mean for the Riksbank’s RAMSES II model? As RAMSES II is a New Keynesian model, the overall conclusions also apply to this model, even if the exact

20 Note that as the New Keynesian basic model includes nominal price rigidity, the deviation in inflation from steady-state inflation (the target) will be part of the central bank’s target function.

21 The gap in equation (10), that is $(u - u^*)$, is instead related to the deviation from the efficient or welfare-maximising allocation in the BG model.

expressions for the various gaps differ according to the modelling choices for, say the labour market. As in BG, there is no divine coincidence applying in RAMSES II, as there are inefficient macroeconomic shocks that drive a wedge between the flexible-price equilibrium and the efficient allocation in the model. Another property that is model-dependent is whether the efficient allocation is defined by constant unemployment. In BG this result follows on from specific assumptions of preferences, lack of capital formation in the model and a special formulation of the employment adjustment cost function, and it is unlikely that RAMSES II would also have this property. However, determining how much unemployment varies in the efficient allocation in RAMSES II is complicated, and at present unknown.

Concluding comment

It has not always been easy for the various participants in the debate on monetary policy and the labour market to understand one another. The confusion regarding concepts is due to the fact that the debaters rarely describe explicitly the models they base their reasoning on. We have therefore discussed in this article two different lines of reasoning (the LNJ model and the New Keynesian model) on the relationship between inflation and unemployment and shown that these are not compatible with one another. The factors behind inflation in one line of reasoning are not a relevant measure of inflationary pressure in the other. Such differences lead to confusion in the discussion. The debate on economic policy gains new life from the different parties having different views on how the economy functions, but if the message is to get across and if opinions are to be discussed constructively, it is important that the debaters are clear as to which line of reasoning they rely to support their arguments.

References

- Akerlof, George A. and Yellen, Janet L. (1990), The Fair Wage-Effort Hypothesis and Unemployment. *Quarterly Journal of Economics* 105, pp. 255-283.
- Basu, Susanto and Carlsson, Mikael (2012), The Aggregate Significance of Labor Reallocation. Mimeo, Sveriges Riksbank.
- Blanchard, Olivier and Galí, Jordi (2007), Real Wage Rigidities and the New Keynesian Model. *Journal of Money, Credit, and Banking* 39 pp. 35-65.
- Blanchard, Olivier and Galí, Jordi (2010), Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics* 2, pp. 1-30.
- Blanchard, Olivier and Katz, Lawrence F. (1997), What We Know and Do Not Know About the Natural Rate of Unemployment. *Journal of Monetary Perspectives* 1, pp. 51-72.
- Calmfors, Lars (1993), Centralisation of Wage Bargaining and Macroeconomic Performance – A Survey. *OECD Economic Studies*, No. 21.
- Carlsson, Mikael, Eriksson, Stefan and Gottfries Nils (2012), Product Market Imperfections and Employment Dynamics, forthcoming *Oxford Economic Papers*.
- Carlsson, Mikael and Westermark, Andreas (2011), The New Keynesian Phillips Curve and Staggered Price and Wage Determination in a Model with Firm-Specific Labor. *Journal of Economic Dynamics and Control* 35, pp. 579-603.
- Christiano, Lawrence J., Trabandt, Mathias and Walentin, Karl (2011), Introducing Financial Frictions and Unemployment into a Small Open Economy Model. *Journal of Economic Dynamics and Control*, 35, pp. 1999-2041.
- Danthine, Jean-Pierre and Kurmann, André (2004), Fair Wages in a New Keynesian Model of the Business Cycle. *Review of Economic Dynamics* 7, pp. 107-142.
- de la Croix, David, de Walque, Gregory and Wouters, Rafael (2009), A Note on Inflation Persistence in a Fair Wage Model of the Business cycle. *Macroeconomic Dynamics* 13, pp. 673-684.
- Erceg, Christopher J., Henderson, Dale W. and Levin, Andrew T. (2000), Optimal Monetary Policy with Staggered Wage and Price Contracts. *Journal of Monetary Economics* 46, pp. 281-313.
- Eriksson, Stefan and Gottfries, Nils (2005), Ranking of Job Applicants, On-The-Job Search, and Persistent Unemployment. *Labour Economics* 12, pp. 407-428.
- Friedman, Milton (1968), The Role of Monetary Policy. *American Economic Review* 58, pp. 1-21.
- Gertler, Mark, Sala, Luca and Trigari, Antonella (2008), An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining. *Journal of Money, Credit, and Banking* 40, pp. 1713-1764.
- Hall, Robert E. and Milgrom, Paul R. (2008), The Limited Influence of Unemployment on the Wage Bargain. *American Economic Review* 98, pp. 1653-1674.
- Layard, Richard, Nickell, Stephen and Jackman, Richard (2005), *Unemployment: Macroeconomic Performance and the Labor Market*. Oxford University Press.
- Mishkin, Frederic S. (1982a), Does Anticipated Monetary Policy Matter? An Econometric Investigation. *Journal of Political Economy* 90, pp. 22-51.
- Mishkin, Frederic S. (1982b), Does Anticipated Aggregate Demand Policy Matter? Further Econometric Results. *American Economic Review* 72, pp. 788-802.
- Nash, John (1950), The Bargaining Problem. *Econometrica* 18, pp. 155-162.
- Pissarides, Christopher A. (2000), *Equilibrium Unemployment Theory*. 2nd ed. MIT Press.

Rogerson, Richard (1997), Theory Ahead of Language in the Economics of Unemployment. *Journal of Economic Perspectives* 11, pp. 73-92.

Romer, David (2006), *Advanced Macroeconomics*. 3rd ed. McGraw-Hill/Irwin, NY.

Rubinstein, Ariel (1982), Perfect Equilibrium in a Bargaining Model. *Econometrica* 50, pp. 97-109.

Shapiro, Carl and Stiglitz, Joseph E. (1984), Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review* 74, pp. 433-444.

Weiss, Andrew (1980), Job Queues and Layoffs in Labor Markets with Flexible Wages. *Journal of Political Economy* 88, pp. 526-538.

Yashiv, Eran (2000), The Determinants of Equilibrium Unemployment. *American Economic Review* 90, pp. 1297-1322.

Appendix A: Trade union and bargaining models

This appendix discusses trade union and bargaining models in greater detail. We begin with a monopoly trade union that sets wages one-sidedly. In this case the trade union will strive for a high wage but must also take into account the fact that a higher wage will result in higher unemployment among its members (that is, members who can and want to work for the given wage but where there will not be sufficient demand at the given wage to employ all of them). To illustrate the points in this model more clearly, we imagine that we have a large trade union that covers the whole economy. Moreover the firm in the economy has the right to determine employment given the wage, what is known as the right-to-manage, which is the reasonable empirical case.²² As the price-setting curve shows optimal combinations between wages and unemployment for the firm, the solution will thus always lie on this curve. The monopoly trade union's problem then is to choose an optimal point on the price-setting curve, given its preferences.

Graphically, we can illustrate the monopoly trade union's wage-setting problems as in Figure 4, where the trade union chooses the combination of real wage (w^{MF}) and unemployment (u^{MF}) where its (highest attainable) indifference curve (between the members' wages and the risk of unemployment) meets the price-setting curve.

If we instead imagine that wages are negotiated between the firm and the trade union (under right-to-manage) the standard modelling is for wages to be set by Nash bargaining^{23, 24}

$$\max_w [\mathcal{A}(w) - A_0]^\rho [\Pi(w) - \Pi_0]^{1-\rho}, \quad (A1)$$

where $\rho \in [0, 1]$ is the trade union's relative bargaining power, $\mathcal{A}(w)$ is the trade union's pay-off function, which depends on the bargained wage and $\Pi(w)$ is the firm's profit. Furthermore, A_0 (Π_0) is the trade union's (firm's) alternative to an agreement. The trade union's surplus in the bargain is given by $\mathcal{A}(w) - A_0$ and in the same way the firm's surplus is given by $\Pi(w) - \Pi_0$.

If we let ρ move towards 1, we reach the monopoly trade union solution discussed above as a special case. If we instead allow ρ to move towards 0, the firm will set wages in line with the reservation wage w^R which is given by the constraint that the employees' surplus $\mathcal{A}(w) - A_0 \geq 0$.²⁵ That is, the firm must pay at least a wage that the employees are willing to work for. Note that the latter condition describes a jump in the labour supply. This

22 See Layard, Nickell and Jackman (2005) for a discussion. Further, Basu and Carlsson (2012) show evidence that the right-to-manage assumption is reasonable in Swedish micro data.

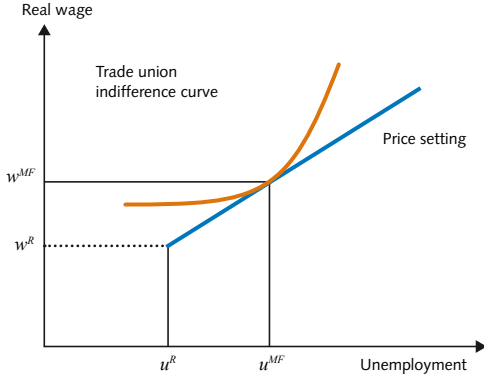
23 For theoretical support for this formula within axiomatic theory, see Nash (1950) and on the basis of non-cooperative bargaining theory, see Rubinstein (1982).

24 The main alternative to a right-to-manage formula is that the trade union and the firm negotiate on both wages and employment (what is known as efficient bargaining). In practice, one can see such a bargaining as first choosing the rate of employment to maximise the surplus in (A1) and that the wages are then determined on the basis of the relative bargaining power between the parties. As wages in this case do not determine employment, the wage is no longer allocative under this formulation of the bargaining.

25 Note that in the second special case with a monopoly trade union ($\rho = 1$) the constraint $\Pi(w) - \Pi_0 \geq 0$ will always be fulfilled as long as the trade union chooses a point on the firm's pricing curve.

is because no one is willing to work for a lower wage than the reservation wage, w^R (which in turn is determined by the unemployment benefit or the value of leisure time), while everyone is willing to work for a wage that is higher than w^R . Thus, unemployment will jump down to 0 for wage levels below w^R .

Figure 4. Equilibrium unemployment in a monopoly trade union model



In a bargaining process, the wage will end up in the interval of $w \in [w^R, w^{MF}]$. The exact level for the wages (and unemployment) will, in turn, be determined by the relative bargaining power, ρ .

The point is also made in the main text that taxes and charges that have a proportional effect on the surplus will not have any effect on wages, and thus not on unemployment; they will only affect net wages and net gains. To give an example of this, let us express $A(w) - A_0$ in equation (A1) as

$$p(w) \left(\frac{w^{1-\sigma}}{1-\sigma} - \frac{B^{1-\sigma}}{1-\sigma} \right), \tag{A2}$$

where $p(w) = N(w) / M$ is the probability of a trade union member being in work (where membership, M , is an exogenous given) and we assume that the trade union's alternative to an agreement A_0 is unemployment benefit in real terms (B). If we now introduce a payroll tax we get

$$p(w) \left(\frac{((1-\tau)w)^{1-\sigma}}{1-\sigma} - \frac{((1-\tau)B)^{1-\sigma}}{1-\sigma} \right) = (1-\tau)^{1-\sigma} (A(w) - A_0). \tag{A3}$$

That is, payroll tax has a proportional effect on the surplus and does not affect the primary conditions for (gross wages) w .

Appendix B: Search and matching models

This appendix gives a more detailed review of the basic search and matching model. To make it simple, we here look at a model with constant and exogenous search intensity. Thus, we assume that an unemployed person always looks for work. Further, we specify a matching function as $H=x(u\bar{L},v\bar{L})$, where H is the number of new hires (during a period of time), \bar{L} is the exogenous given labour force and u and v are the rate of unemployment and vacancy rate respectively.²⁶

In the search models the relevant labour market variable is not unemployment, but the demand pressure for labour (relative to supply), often known as “tightness”, which is defined as the ratio of vacancies to unemployed persons seeking work, $\theta(=v/u)$. Further, the probability of a firm filling a vacancy within a period of time $q\theta=\frac{H}{v\bar{L}}$ where we assume that $q'(\theta)<0$. The probability of an unemployed person finding a job during a period of time is given as $p(\theta)=\frac{H}{u\bar{L}}=\theta q(\theta)$ where we assume that $p'(\theta)>0$. Firms with vacancies thus find workers more easily when there are more job-seekers relative to the number of job vacancies and job-seekers find a job more easily when there are more job vacancies relative to to the number of job-seekers.

The labour market is in flow equilibrium when the inflow to unemployment is as large as the outflow. As we assume that the probability of a worker leaving his or her job during a period of time, s , is constant, we have in the flow equilibrium that $s(1-u)=p(\theta)u$ or that the inflow into unemployment is as great as the outflow from unemployment. We can write this equivalently as

$$u=\frac{s}{s+p(\theta)}. \tag{B1}$$

The equation (B1) gives us an expression of equilibrium unemployment for a given labour market tightness (θ). Given what we have assumed about the matching function's properties, we can represent equation (B1) in the vacancy/unemployment space as a downward sloping curve that is convex towards origo (see Figure 5 bottom left-hand corner). This type of curve is usually called a Beveridge curve, which is also observed as an empirical relationship in data. Note that changes in the matching process affect the $p(\theta)$ expression, and shift the Beveridge curve closer to or further from origo. That is to say, when the matching efficiency changes, more or fewer vacancies are needed to uphold the same unemployment rate.

To determine labour market tightness (θ) we then turn to the labour-demand side/ wage formation in the model. The first relationship is the wage-setting curve. As the search frictions mean that one cannot find a new bargaining partner cost-free, a value is created when the employer and employee meet. In the model, the wage is set by dividing the total surplus in a bargain in accordance with equation (A1) above. The wage-setting

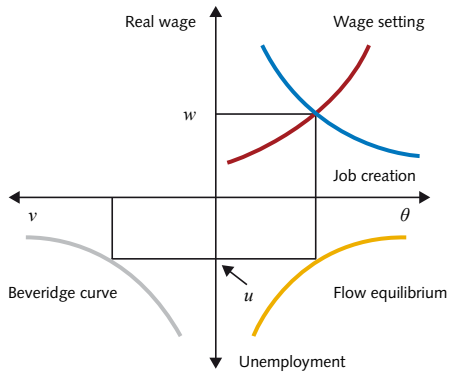
²⁶ In accordance with empirical results, we assume that H is homogenous of degree 1 and concave ($H_{11}<0, H_{21}<0$).

curve will be an upward-sloping function of θ , as more vacancies per job-seeker mean that the competition for workers increases, which strengthens the workers' alternative to an agreement.

The other relationship used to determine labour market tightness (θ), the job-creation curve, comes from the firms' incentives to create vacancies. A vacancy has a value as one can expect to recruit a worker and generate revenues in the future. If the value of the vacancies (net from the cost of holding a vacancy) is positive, the firm can create gains by posting more vacancies, but if the net value is negative, the firm will close the vacancy. Thus, the number of vacancies in equilibrium can be determined by a zero profit condition for them. This zero profit condition gives a negative relationship between wages and θ , as higher labour-demand pressure means that the firm expects it will take longer to fill a vacancy ($q'(\theta) < 0$), which reduces the value of the vacancy. To restore the equilibrium value of a vacancy to zero, the wages must thus fall. One of the key implications from the search model is thus that not only wage-setting, but also incentives for job-seeking are affected directly by the labour market situation (θ) via the expected cost of filling a vacancy.

If we combine these relationships, the equilibrium in a search model can be illustrated as in Figure 5. Alternatively, one can represent the equilibrium as a price-setting and wage-setting equilibrium as in Figure 2 in the main text, by combining both the job-creation curve and the wage-setting curve with the flow-equilibrium condition.

Figure 5. Equilibrium unemployment in a search model



Appendix C: Derivation of the short-run supply curve in LNJ

By writing equations (4) and (5) as deviations from the expectation, we write the price-setting and wage-setting relationship as

$$p - w^n = \beta_0 - \beta_1 u - (w^n - E_{-1}(w^n)), \quad (C1)$$

and

$$w^n - p = \gamma_0 - \gamma_1 u - (p - E_{-1}(p)). \quad (C2)$$

That is, a situation with a surprisingly high price and wage inflation shifts these curves and gives a short-run equilibrium with lower unemployment. This can also be shown by combining (C1) and (C2), and solving for unemployment which gives

$$u = u^* - \frac{(p - E_{-1}(p)) + (w^n - E_{-1}(w^n))}{\beta_1 + \gamma_1}, \quad (C3)$$

where u^* is defined as $(\beta_0 + \gamma_0) / (\beta_1 + \gamma_1)$, that is, equilibrium unemployment with price and wage stability or, in other words, steady-state unemployment. This level can be calculated by setting price and wage surprises to zero in equations (C1) and (C2) and then substituting for real wages in (C2) and solving for unemployment.

To derive a short-run supply curve we then assume (as in LNJ) that $(p - E_{-1}(p)) = (w^n - E_{-1}(w^n))$ which gives

$$u - u^* = -\frac{2}{(\beta_1 + \gamma_1)} (p - E_{-1}(p)), \quad (C4)$$

which corresponds to equation (6) in the main text.

The Riksbank's business survey – a quick indicator of economic activity

JYRY HOKKANEN, TORE MELIN AND ALEXANDER NILSON

The authors work at the Monetary Policy Department of the Riksbank.

Like several other central banks, Sveriges Riksbank conducts its own business survey. The aim of the survey is to gather up-to-date information as a basis for an assessment of the current state and development of the economy ahead of the publication of official statistics. The structure of the business sector in Sweden, with only a few predominant companies in each sector, makes it possible for the Riksbank to cover the development of a large part of the economy by primarily turning to these large companies. The companies' responses are also used to forecast the development of GDP.

Various types of economic statistics form the basis for the Riksbank's assessments of the state of the Swedish economy, as well as for the forecasts on which the monetary-policy decisions are partly based. However, the time lag associated with these statistics means that they may sometimes be seen to be out of date at the time of publication. This applies, for example, to the data from the quarterly National Accounts, which is usually not available until two months after the end of the quarter concerned. Consequently, the Riksbank and other decision-makers need complementary data that provides a more up-to-date picture of the situation. The Riksbank's forecasting models therefore use flash statistics that reveal economic tendencies but are often not as detailed or complete as the official statistics. The Business Tendency Survey of the National Institute of Economic Research is one example of such flash statistics that provide an indication of the economic climate.¹ Various studies have also shown that the data in the Business Tendency Surveys is highly consistent with the overall economic statistics and that the companies' assessments of the outlook for the future in the surveys thus provide a valuable basis for assessments of the development of the economy.

Since 2007, the Riksbank has conducted its own business survey in order to complement the general economic statistics. The aim of the survey is to gather reliable information on the state of the economy, but also to provide an opportunity to discuss current economic issues that are of interest to our analysis with the companies.

¹ The Business Tendency Survey of the National Institute of Economic Research is based on questions put to over 6 000 companies every month.

MANY CENTRAL BANKS CARRY OUT THEIR OWN SURVEYS OF THE ECONOMY

Many central banks gather information on the current state and development of the economy through their own contacts with the business sector. Perhaps the most well-known business survey, the so-called Beige Book, is conducted by the regional offices within the Federal Reserve System in the United States.² This type of data collection is valuable because independent central banks can often get good information from the companies in the country concerned as they are usually not bound by the political decisions of governments, nor have any vested interest in the business development of individual companies. The questions about the development of prices and economic activity that are usually asked in the central-bank surveys also relate to factors that are important to the potential development and prospects of the companies. This makes it interesting for the companies to take part in the surveys and to study the results.

The methods the central banks use to collect the data vary. In some cases the central banks carry out extensive data-collection programmes that can complement other national statistics, either because these statistics are inadequate or unavailable, while in other cases the surveys are less formal in nature. The latter may, for example, apply on special occasions when the central bank believes that close contacts with the business sector are necessary to quickly collect and also communicate information on the economic situation.

HOW THE RIKSBANK'S BUSINESS SURVEY IS CONDUCTED

In the Riksbank's business survey, the data is collected by carrying out in-depth interviews. This is one of the many qualitative methods for data collection used within a range of subject matter areas. Both qualitative and quantitative surveys measure the properties or characteristics of the subject of the survey in some way. Qualitative surveys often focus on the relationship between different conditions and on trying to find new aspects of them by adopting a freer approach to the issues that are to be investigated. In contrast to quantitative methods, which are used to collect data on the prevalence of certain properties,³ quantitative methods are used to collect data on particular aspects of different properties. This may relate to causal relationships or to conceivable reactions to certain external circumstances, but also to discovering different points of view that one did not know existed or that were assumed to be irrelevant before the data-collection process began.⁴

Other qualitative methods apart from in-depth interviews include focus groups, text analysis and participatory observation. A common feature of these methods is that they search for relationships or properties that it has not been possible to describe or delimit in advance so that they can then be measured quantitatively. For example, it is often difficult

2 Other central banks that conduct business surveys include the Bank of England, the Bank of Canada, Norges Bank, the Swiss National Bank and the Reserve Bank of Australia.

3 The results of quantitative surveys, for example surveys of the level of support for political parties, are presented in the form of frequencies and distributions that can be calculated with a certain degree of accuracy for the entire population that the sample has been taken from.

4 See Starrin et al. (1991)

to quantify points of view about the future when aiming to clarify the contexts of which they form part. Focus groups are a method commonly used in market surveys where the aim is to understand how consumers view a new product or a new brand in relation to already existing products and brands.⁵ In text analysis, one searches for statements but also for latent or underlying messages that are “hidden” in the text. Such text analyses are, for example, regularly carried out following general elections in Sweden. Participatory observation has often been used in studies of residential environments and so on.

QUALITATIVE RESULTS IN THE RIKSBANK'S SURVEY CAN APPLY TO ALL COMPANIES

Unlike the results of quantitative surveys, the results of qualitative surveys cannot be transferred from those studied to apply to an entire population using calculations in the form of statistical inference. This is not to say, however, that the results are invalid for the entire population. If this was the case, then it would not be possible, for example, to use a few focus groups, each made up of ten or so participants, to understand target groups consisting of thousands of intended purchasers.

The strength of the qualitative survey lies in the fact that it can help us to discover and understand various new aspects of a situation or to understand the assessments made by those interviewed. This distinguishes the qualitative survey from the quantitative survey in which the questions are set from the start and limited to aspects determined in advance. By using qualitative methods, one can collect a wider range of information. For example, a company executive may reason in a wide variety of ways about the appropriate action to take in certain circumstances. In a qualitative survey, these different lines of reasoning can be discussed on site with the executive concerned.

In its business survey, the Riksbank asks about aspects such as the development of economic activity and prices and tries to capture how the companies view these aspects and intend to act in relation to them. Although there may be many lines of reasoning in theory, in practice the number of options is limited. After a number of interviews, the addition of new aspects of the variables investigated therefore tends to decrease with each additional observation. The set of assessments and viewpoints associated with a certain point in time or situation in the economy thus usually becomes fairly stable already after a relatively small number of observations, such as is the case in the Riksbank's business survey.⁶

THE BUSINESS SURVEY'S SAMPLE FOCUSES ON THE LARGER COMPANIES

The Riksbank's business survey is conducted among companies that can be assumed to give multifaceted assessments of, in the first instance, the development of prices and economic activity. This means that it is particularly interesting to follow sectors in which

5 See Seymour (1992).

6 See Glaser & Strauss (1967). In methodological terms one usually talks about “theoretical saturation” when additional observations confirm rather than alter the picture that has emerged so far.

fluctuations in economic activity are clear. Companies in the retail and manufacturing sectors are therefore overrepresented in the sample in relation to their share of the business sector as a whole, while the sample is limited among companies in the other services sector.

Due to the high degree of concentration in the Swedish business sector, a limited number of interviews cover a large part of the sector in terms of turnover and the number of employees. The ten largest groups account for approximately 35 per cent of the employees in the manufacturing industry, and in the retail sector the degree of concentration is even higher: the three largest companies account for approximately 45 per cent of the number of employees. Several other sectors, such as construction and road and rail transportation, are also dominated by a few major companies.

Since September 2007, the Riksbank has interviewed 25 to 60 companies of importance to the Swedish economy three times a year. The interviews have been conducted ahead of the monetary-policy decisions that are accompanied by the publication of Monetary Policy Reports. The sample has partly been the same every time, but as different issues have arisen from time to time it has been appropriate to interview companies that have been particularly affected by various events in the economy on these different occasions. However, the core of the sample has always consisted of the largest companies, which in this case are defined as the companies with most employees in Sweden in the respective sectors. In total, about 300 companies have been interviewed over the course of five years.

At present, 45 companies are normally interviewed in May and September and 25 companies in January. The number of employees that the interviewed companies have in Sweden varies between 200 000 and 300 000 at the time of each survey, which constitutes 6 to 8 per cent of the total number of employees in the Swedish business sector.

THE QUESTIONS CAPTURE THE DIRECTION IN WHICH THE ECONOMY IS MOVING

The interviews are conducted with the help of an interview guide containing questions on the companies' views on economic activity, production, employment, the development of prices and other issues that the particular survey wishes to gather information on.⁷ In the case of international companies, information is primarily sought on their Swedish operations. The questions in the guide are phrased as direct questions, but they do not have to be put verbatim. The important thing is that the interviewers get answers to their questions and that the issues are fully covered in the interview, as the main purpose of the survey is to obtain in-depth information on how the companies perceive the current situation and how this affects their planning. The interviewers code the responses in accordance with the scale attached to each question, usually in terms of shifts such as "better" and "worse", or "unchanged". These "quantitative" questions are accompanied by straight-forward, more qualitatively-oriented questions such as "why?". The analysis of

⁷ The interview guide is the tool that the interviewers use during the interviews and contains the questions to be covered during the interview and advice and instructions for the interviewer. An example of the interview guide is presented in the Appendix.

the interviews that the interviewers compile in a summary of each interview is therefore as important as the coding of the questions.

THE RIKSBANK'S OWN STAFF CARRY OUT THE INTERVIEWS

It is natural for the Riksbank's own staff to conduct the interviews as the issues to be addressed can change quickly from one survey to another. The responses also need to be analysed and compared with other internal assessments made at the Riksbank. The issues investigated during the surveys are also such that the interviewers need to have a good grasp of both macroeconomics and the way that companies work so that they can, for example, ask relevant follow-up questions.

However, when the survey began in 2007, the Riksbank lacked methodological experience of how to carry out surveys of this kind; there was no corps of interviewers or any other similar group that had special training in the skills required to conduct surveys. To establish whether the Riksbank would be able to conduct its own survey, a pilot survey was therefore carried out which demonstrated that the Riksbank was in fact well able to perform the survey using its own staff. The experience gained during the pilot survey is now used in the introduction that new interviewers are given ahead of the first survey they take part in, and any new experience gained during the course of the surveys is also used in the introduction. More than 50 economists at the Riksbank have now acted as interviewers over the five years in which the survey has been conducted.

THE RESULTS ARE REPORTED QUICKLY

The interview guide lists the questions to be asked and describes how the interviews themselves should be conducted. The interviews are recorded unless the interviewee objects to this, which is very unusual. The recording is an important aid to the interviewer who should write a summary of the interview within two days. The summaries are read within two days by the editorial group that writes the survey report and any queries that arise are resolved between the group and the interviewer and, in some cases, by contacting the company again.

The editorial group meets four to five times during the period of the survey and the overall result of the survey emerges during the course of their work and is summarised in the report that is then published. Each survey begins with a meeting at which new questions and particularly important aspects are discussed with the interviewers, and when the interviews have been completed the interviewers are gathered together for a debriefing at which the experience from the interviews is discussed.

In this process, the individual responses from each of the companies are weighed together to form an overall picture of how the companies view the economic situation and how they intend to act on the basis of this view.

GREAT INTEREST IN PARTICIPATING IN THE BUSINESS SURVEY

The Riksbank's business survey has provided quick and useful information despite the fact that it is based on a relatively limited amount of data. This may partly be because Swedish business, as mentioned earlier, is very concentrated, with several sectors that are dominated by large companies.

In general, the Swedish companies are also interested in participating in the survey. They are accommodating and willing to provide information, and very few decline to take part. The quality of the information they provide is also consistently high and the interviews are usually carried out directly with members of the company's management. The value of the information we get from the interviews is therefore high already under normal conditions, but the information is particularly useful in the event of rapid changes in the economy.

THE RIKSBANK'S BUSINESS SURVEY IS RELIABLE

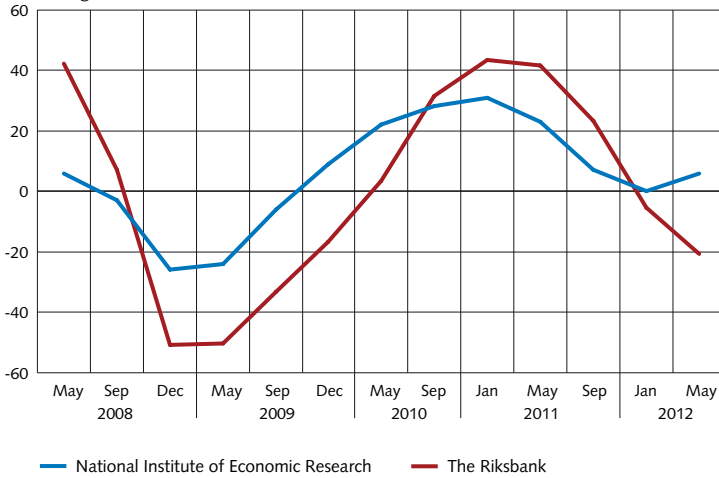
The results of the Riksbank's business survey can be compared with those of the Business Tendency Survey of the National Institute of Economic Research. This is because it is partly the same companies that are interviewed. The response distribution for similar questions in these two surveys can also be calculated in the same way thanks to the quantitative coding that is carried out alongside the qualitative analysis.⁸ A comparison with the Business Tendency Survey can thus indicate how reliable the results of the Riksbank's business survey are.⁹

Figure 1 shows how the companies perceive the economic situation according to the Riksbank's business survey and the Business Tendency Survey's confidence indicator. There is a close correspondence between the two series and both reach their trough in December 2008 and their peak in January 2011. One difference is that the Riksbank's time series varies more than the time series in the Business Tendency Survey. A possible reason for this is that the Riksbank's business survey focuses on companies that are sensitive to fluctuations in economic activity and is based on fewer companies, which means that each individual response may have an impact on the results.

8 As the size of the participating companies varies significantly in each survey, the companies' responses are weighted in terms of the number of employees they have in Sweden.

9 The questions in the two surveys are not identical. The Business Tendency Survey's confidence indicator is based on responses to three different questions, but in both cases the surveys aim to measure the "situation in" or the "state of" the business sector. The National Institute of Economic Research interprets the confidence indicator as a "measure of the situation" while the Riksbank's business survey interprets its responses as a "measure of economic activity" which should thus make the responses comparable.

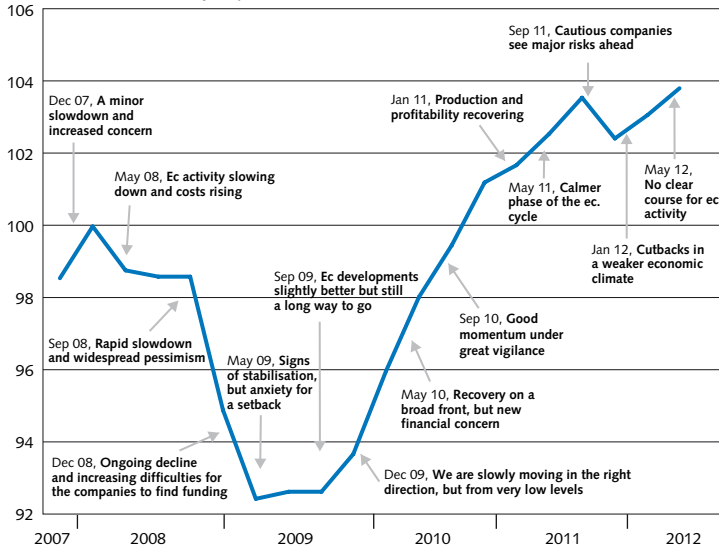
Figure 1. The economic situation according to the Riksbank's business survey compared to the confidence indicator of the National Institute of Economic Research
Net figures



The Riksbank's analysis of the companies responses and their assessments of the economic situation is summarised after survey in a published report. The headings in each report express the most important results in condensed form. Another way of assessing the reliability of the information that comes from the Riksbank's business survey is therefore to compare the headings in the reports with the development of Sweden's GDP. Figure 2 shows the headings and the development of GDP at the time each report was published. The comparison between the assessments expressed in the reports' headings and the development of GDP show that the survey has provided information that successfully captures the future development of the economy at an early stage.

Figure 2. Report headings in the Riksbank's business survey and the development of Sweden's GDP

GDP level, seasonally-adjusted data, index 2007 Q4 = 100



It is particularly interesting that the analysis in the business survey captured the downturn in the autumn of 2008 at an early stage, and thereafter also the turnaround in 2009. The headings in the surveys in September and December 2008 were “Rapid slowdown and widespread pessimism” and “Ongoing decline in economic activity and increasing difficulties for the companies to find funding” respectively. These were the messages communicated in early October 2008 and mid-January 2009, although it was not until the end of February 2009 before the dramatic fall in GDP during the fourth quarter of 2008 was fully confirmed by the official statistics.

However, it can also be noted that the report headings include events in the economy that are not automatically captured by the development of GDP. For example, financial unease and access to funding were highly important factors which affected the operations of the companies and characterised their responses during the period studied. Profitability, prices and the supply of labour are other factors that are perhaps not directly captured by the development of GDP.

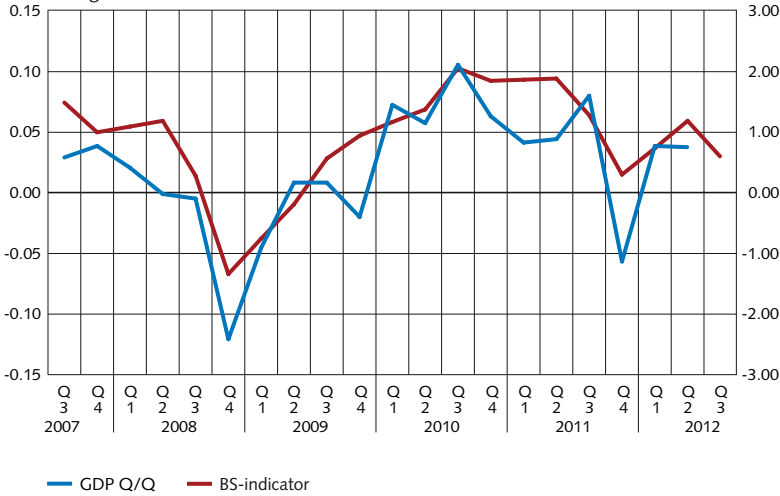
AN INDICATOR OF ECONOMIC ACTIVITY THAT PREDICTS THE GDP OUTCOME

The questions in the business survey that are coded quantitatively are used partly to support the qualitative analysis and partly to calculate a quantitative indicator of economic activity. Most of the questions in the survey relate in one way or another to the development of economic activity. By weighing together the eight questions in the survey that have been asked from the start, we produce an indicator of economic activity that reflects the expected development better than a single question does. The questions used

to produce this indicator relate to production volumes and employment (preceding three months and next three months), investment plans six months ahead, current profitability, the development of productivity up until the time of the survey and the expected development of prices in the coming year. Their weight in the indicator is determined by the ability of the respective questions to predict changes in GDP from quarter to quarter. The indicator's covariation with the quarterly change in GDP is presented in Figure 3.

Figure 3. Indicator of economic activity based on the Riksbank's business survey (BS-indicator) in comparison with quarterly changes in Sweden's GDP

The indicator's index on the left Y axis and GDP's quarterly percentage changes on the right Y axis



Note: R2 = 0.84.

There is a strong covariation between the indicator and the development of GDP, as expressed by the correlation coefficient of 0.84 and as illustrated by the curves in the graph, which follow each other relatively closely. What the figure does not show is the time difference between the calculation of the indicator and the publication of the GDP outcome. During the quarters in which the survey is conducted, the indicator is calculated up to two months before the outcome is published. This means that it is possible to paint an early picture of the development of economic activity on the basis of the Riksbank's business survey.

THE INDICATOR IS USED TO CALCULATE QUARTERLY FORECASTS FOR GDP

A third way of assessing the reliability of the business survey is to compare the forecasts for GDP produced with the help of the indicator with the results provided by the Riksbank's forecasting models. Forecasts of the change in GDP from quarter to quarter have been produced on the basis of the indicator since the third quarter of 2008.

Forecasting models can be assessed by ranking the forecasting errors that the models have made in the form of mean absolute errors (MAE) and the root mean square error

(RMSE). A comparison of the quarterly forecasts from the business survey with a weighted mean forecast from the various forecasting models shows that the survey's forecasts hold up well. Both the MAE and the RMSE are smaller in the business survey's forecasts than in the mean forecast, and they are also smaller than in the individual models from and including the third quarter of 2008 up to and including the second quarter of 2012. The difference between the MAE and the RMSE is also small in the business survey's forecasts, which means that these contain few significant forecasting errors. The reliability of the forecasts based on the indicator of economic activity in the business survey is thus relatively high.

THE BUSINESS SURVEY IS AN IMPORTANT ELEMENT OF THE RIKSBANK'S POLICY WORK

The business survey is now an established part of the monetary policy analysis at the Riksbank. Its results are highly reliable and are regularly reported in the media. The results are also noted by analysts that monitor the work of the Riksbank.

An important result of the work with the business survey is that the Riksbank has gained a greater understanding of the actions taken by the companies in different phases of the economic cycle. Information on the complexity of different markets and changed company strategies has also helped to increase our analytical skills. This may relate, for example, to understanding how exchange-rate hedging takes place or how funding has changed as a result of the financial crisis. Our experience underlines the fact that it is important to have direct contacts with the business sector in order to better understand changes in the workings of the economy and the behaviour of the players concerned. The flexibility of the survey means that it can also provide a basis for the Riksbank's other assignments apart from monetary policy. This is illustrated, for example, by the in-depth questions about the companies' funding situation that were asked during the financial crisis and that aimed to give the Riksbank a better understanding of the workings of the financial markets in a crisis situation.

The Riksbank's business survey is under constant development and several lessons have been learned during the five years in which it has been conducted: an effective internal organisation for the survey has been set up in which the interview and analytical skills of those involved are constantly improving. The number of companies interviewed is growing and the addition of new companies helps to build up the information bank. The emphasis of the recurring questions in the interviews shifts in line with the current economic situation, and specific questions are asked when this is called for. Extensive and detailed material from the survey has been compiled over the years and many different types of analysis can be performed on the basis of this material in addition to the reporting that is carried out today. The interviewed companies can also be seen as a network that can be used for the mutual exchange of information between the Riksbank and the business sector in more or less formal ways.

References

Glaser, Barney G. & Strauss, Anselm L., (1967), *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Publishing Company

Seymour, Daniel T., (1992), *Marknadsundersökningar med kvalitativa metoder*, Studentlitteratur

Starrin, Bengt, Larsson, Gerry, Dahlgren, Lars and Styrborn, Sven (1991), *Från upptäckt till presentation, Om kvalitativ metod och teorigenerering på empirisk grund*, *Teori Forskning Praktik*, Studentlitteratur

Wärneryd, Bo (1986), *Att fråga, Om frågekonstruktion vid intervjuundersökningar och postenkäter*, Statistics Sweden (SCB)

Yin, Robert K. (1989), *Case study research, Design and Methods*, *Applied Social Research Methods Series Volume 5*, Sage Publications

Ödman, Per-Johan, (1979), *Tolkning förståelse vetande*, *Hermeneutik i teori och praktik*, AWE/GEBERS

Appendix

THE INTERVIEW GUIDES – AN EXAMPLE

The interview guides contain the questions that the interviews aim to get answers to. The answers emerge from the discussion that an in-depth interview entails, which means that the answers often come up without putting the questions directly. Some central questions are, however, formulated as direct questions in the guides so that they can be simply put in the specified way if the answers that emerge in the ongoing discussion are unclear. The responses to these core questions are coded by the interviewer on what is usually a three-point scale of the type increased – unchanged – decreased.

The guides also contain instructions for the interviewers. These may relate to appropriate follow-up questions, aspects that the interviewer should pay special attention to and so on.

The guides differ slightly depending on which of the four sectors, that is manufacturing, construction, retailing and other services, that the Riksbank's business survey is directed at. The general questions about economic activity, production/sales, sales prices and so on are the same for all sectors while, for example, the question about stocks is not put to service companies.

The interview guide presented in this Appendix was used for interviews in the manufacturing sector in May 2012. The instructions to the interviewers have been removed for reasons of space.

1 a) What is your assessment of the current economic situation for your company?¹⁰

Good... Satisfactory... Poor...

b) What is your assessment of the economic situation for your company in six months' time?

Better... Unchanged... Worse...

c) What risks might the economic situation entail for your company?

d) How great are the risks from the economic situation now?

Greater than normal Normal Lower than normal

2 a) How has the production volume developed over the past three months compared with the previous three months (seasonally adjusted)?

Increased... Unchanged... Decreased...

b) How do you expect the production volume to develop over the coming three months compared with the previous three months (seasonally adjusted)?

Increase... Unchanged... Decrease...

¹⁰ Questions about the current economic situation have been included in the business survey since the round of interviews in May 2008.

c) What is your assessment of the company's stock of finished goods at present?

Too large Just right Too small

3 a) How have orders developed over the past three months compared with the previous three months (seasonally adjusted)?

Increased... Unchanged... Decreased...

b) How do you expect orders to develop over the coming three months compared with the previous three months (seasonally adjusted)?

Increase... Unchanged... Decrease...

4 a) How has employment developed over the past three months compared with the previous three months (seasonally adjusted)?

Increased... Unchanged... Decreased...

b) How do you expect employment to develop over the coming three months compared with the previous three months (seasonally adjusted)?

Increase... Unchanged... Decrease...

c) What is the size of your staff in relation to demand/production?

Too narrow Just right Too broad

5) How does the company currently view the capacity to manage an (unexpected/expected) increase in demand?

No problems... Some problems... Major problems...

6 a) How much do you expect wage costs (per hour and including employer's contributions) to increase during the next 12 months?¹¹

Increase more than before... As much as before... Increase less...

b) What is your assessment of the wage drift (i.e. wage increases over and above central wage agreements) for the next calendar year?

Increase... Unchanged... Decrease...

7) How do the company's investment plans for the coming six months look (compared with the previous six months)?

Increased... Unchanged... Decreased...

8) How has the company's access to external funding changed over the past quarter?¹²

Improved... Unchanged... Worsened...

9) What is your assessment of profitability at present?

Good... Satisfactory... Poor...

¹¹ The questions regarding the development of wage costs and wage drift have been included since the round of interviews in May 2008.

¹² The question on access to external funding has been included since the round of interviews in September 2008.

10 a) How do you intend to change your sale prices?

Current quarter (in relation to previous);	Lower	Keep unchanged	Increase
Coming quarter (in relation to current);	Lower	Keep unchanged	Increase
Coming 12 months (in relation to now);	Lower	Keep unchanged	Increase

b) Which of the following factors will drive prices upwards or downwards in the coming 12 months? (Demand, spare capacity¹³, material costs, labour costs, development of productivity, changes in margins,¹⁴ other)

11) How has productivity (level) developed in your company over the past six months compared with the previous six months?

Fallen...	Unchanged...	Risen...
-----------	--------------	----------

12) How quickly does a change in the exchange rate have an impact on your sales prices?¹⁵

0-3 months...	3-6 months...	6-12 months or longer...
---------------	---------------	--------------------------

¹³ Spare capacity is not relevant to all companies.

¹⁴ Trade margins or equivalent.

¹⁵ The question regarding the impact of exchange rate changes was introduced in the survey in December 2008.

Asset encumbrance and its relevance for financial stability

REIMO JUKS*

The author holds a PhD from the Stockholm School of Economics and works at the Financial Stability Department of the Riksbank.

Increased asset encumbrance in a number of countries has raised questions about the underlying driving forces and consequences of such trends for financial stability. This article argues that some increase in asset encumbrance is natural in distressed market conditions where investors demand more safety and there is an overall lack of safe assets. Yet asset encumbrance may be also driven by banks' strong private incentives to tilt their financing towards secured funding, with adverse effects on financial stability. The banks' failure to internalise liquidity effects, rating arbitrage, risk-insensitive deposit pricing, upcoming regulatory reforms and implicit government guarantees are leading banks to ignore some of the social costs associated with secured funding. Regulatory interventions should aim to encourage the issuance of unsecured funding and eliminate or mitigate regulatory imperfections. Among the recommended policy actions are increasing the transparency of asset encumbrance and changing the deposit insurance system to reflect asset encumbrance, as these measures do not require the assessment of a socially-optimal level of asset encumbrance that varies over time and across countries.

Introduction

The existing evidence shows that asset encumbrance, or using assets to secure claims, has increased markedly in a number of countries since 2005. This raises a host of questions that need to be answered. Why do banks prefer secured as opposed to unsecured funding? What are the consequences of asset encumbrance for investors, taxpayers and financial stability? Is there a case for regulatory interventions? And, if so, what regulatory tools should be used?

To answer these questions, the article first explains asset encumbrance and its most common sources. It then reviews the existing evidence on asset encumbrance, both in Sweden and abroad. The effects of asset encumbrance on the riskiness of different types of debts are then considered, together with potential reasons for why banks have increased asset encumbrance. Finally, the article discusses whether regulatory intervention is needed and what tools should be used.

* Contact email: reimo.juks@riksbank.se. The author would like to thank David Forsman, Anders Rydén, Olof Sandstedt, Staffan Viotti, Göran Robertsson, Claes Berg and Joanna Gerwin for useful comments and discussions. The views expressed here are those of the author and do not necessarily reflect those of the Executive Board of Sveriges Riksbank.

What is asset encumbrance?

To understand asset encumbrance, it is useful to start with a brief description of bank funding. Bank funding sources, such as deposits, market funding and capital, can be characterised by maturity, seniority and collateralisation. Maturity refers to the time period after which the security must be redeemed by the issuer. Seniority refers to the ranking of claims to the residual assets in the event of an issuer’s insolvency. Collateralisation refers to the existence or non-existence of assets that secure the claims in the event that an issuer fails to meet the claims.

It is the collateralisation feature of bank funding sources that gives rise to asset encumbrance. To secure or collateralise a claim, a bank must specify assets that creditors can possess in situations where the bank fails to honour its commitments. This process is typically called asset encumbrance, but can be also referred to as pledging or earmarking assets.

Asset encumbrance can be illustrated with a simple example of two banks (see Figure 1). In Bank A, all investors are uncollateralised or unsecured. Bank A is therefore said to have no asset encumbrance. In Bank B, the claims of investor 1 are safeguarded by the assets marked in green. The claims of investor 2 are left unsecured. Bank B is therefore said to have some asset encumbrance.

Figure 1. Unencumbered and encumbered bank

BANK A		BANK B	
ASSETS	LIABILITIES	ASSETS	LIABILITIES
All assets	Investor 1	Assets pledged to investor 1	Investor 1
	Investor 2		Investor 2
		Other assets	

Sources of asset encumbrance

Asset encumbrance may originate from market funding, derivatives and insurance claims. It cannot originate from deposits or capital as these claims are of an uncollateralised nature (see Figure 2).

Two types of market funding can lead to asset encumbrance: short term secured funding, such as repos¹, and long-term secured funding, such as covered bonds. Types of market funding that do not lead to asset encumbrance include, for instance, commercial papers and senior unsecured bonds.

¹ Security borrowing is very similar to repos and can also lead to asset encumbrance. For simplicity, we consider security borrowing to be one class of repo. We also ignore the off-balance sheet encumbrance that arises with re-hypothecations associated with reverse repos.

Derivatives and insurance claims can also lead to asset encumbrance. Both derivatives and insurance claims can be thought of as debt instruments similar to those in market funding, but with some contingent features.

When derivatives are initiated, contracts are typically designed so that counterparties have no claims against each other at origination. But, as time passes, one of the counterparties typically accumulates claims against the other. To safeguard these claims, the debtor typically must post collateral, leading to asset encumbrance.

Figure 2. Sources of bank funding and asset encumbrance

DEPOSITS	Insured	Retail deposits
	Non-insured	Interbank funds
MARKET FUNDING	Short-term	Repos
		Commercial papers
	Long-term	Covered bonds
		Senior unsecured
CAPITAL	Tier 2	Subordinated
	Non-core Tier 1	Junior subordinated
	Core Tier 1	Common equity
DERIVATIVES		Interest rate swap
INSURANCE		Life insurance

Note. Sources of bank funding that can lead to asset encumbrance are marked in green.

Insurance contracts can be also collateralised and can thus lead to asset encumbrance. In a typical insurance contract, an insurance holder pays an insurance premium to an insurer for the period of the insurance coverage. These funds are then invested in assets which are usually earmarked to the benefit of insurance holders. Depending on the type of insurance, the insurer may increase asset encumbrance, either because the market value of insurance assets decreases while the corresponding insurance obligations are pre-determined or because the insurance obligations increase while the market value of assets does not.

Deposits and capital cannot lead to asset encumbrance. For deposits, this is despite the fact that some deposits have state guarantees. These guarantees mean that the owners of deposits are compensated in the event that a bank fails to meet its obligations. However, following this compensation, the state has a claim against the bank in the form of unsecured debt.

Even though bank capital may include some debt instruments in addition to equity, this does not lead to asset encumbrance. This is because the debt under capital is always

unsecured. The difference between unsecured market debt and debt classified as capital stems from their different levels of seniority.²

Needless to say, asset encumbrance can change over time. A bank's deliberate funding choices can provide one reason for this. A completely unencumbered bank can decide to issue secured funding, which means that its asset encumbrance increases suddenly. Asset encumbrance also changes over time due to factors beyond the bank's immediate control. For instance, a margin call from the derivatives' counterparty due to an unfavourable change in the underlying asset can lead to a sudden increase in encumbrance.

Data on asset encumbrance

Data on asset encumbrance is not readily available at either the bank or the country level. The evidence provided below is thus restricted to a limited number of countries and banks in Europe and excludes some sources of asset encumbrance such as insurance and derivatives.³

Both the level and growth of asset encumbrance have varied considerably in the European countries (see Figure 3). In 2005, the level of asset encumbrance was highest in countries with strong covered bond traditions, such as Spain and Germany.

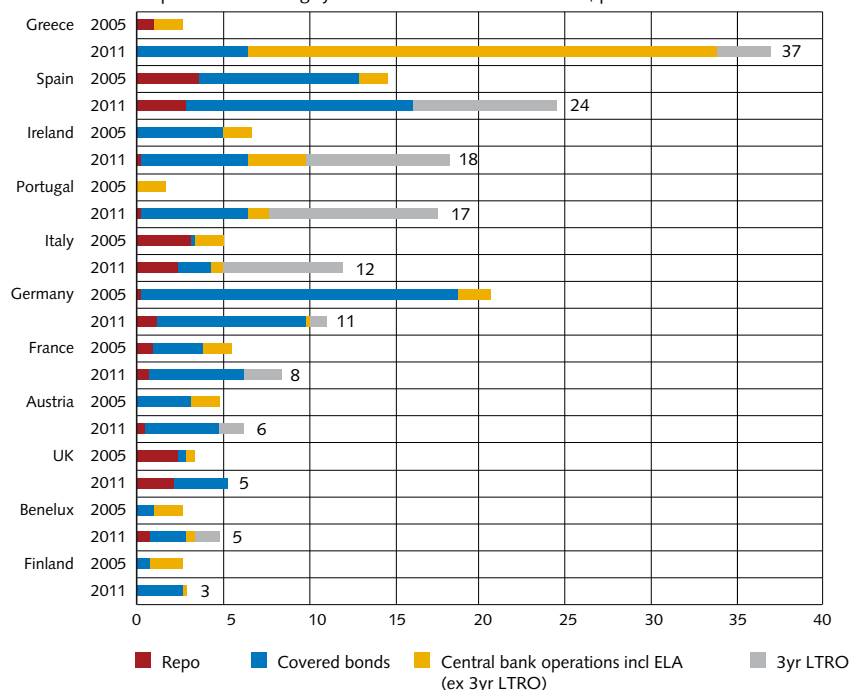
Between 2005 and 2011, asset encumbrance increased significantly in all sample countries except Germany.⁴ Even though the importance of covered bonds increased in Europe during this period, by 2011 the level of asset encumbrance was no longer highest in the traditional covered bond countries. Instead, asset encumbrance levels reached new heights in countries with problematic public finances, such as Greece, Spain, Ireland, Portugal and Italy. For these countries, the main source of asset encumbrance was repo funding associated with the liquidity assistance provided by the European Central Bank (ECB).

2 Seniority of a debt instrument determines the order of repayment in case of bankruptcy. In this example, all unsecured market debt gets paid before the debt that is classified as capital.

3 This data limitation should be kept in mind when comparisons across countries and banks are made. As illustrated by Dexia, derivatives represent an important source of encumbrance (see also Alloway 2011).

4 Asset encumbrance decreased in Germany between 2005 and 2011 due to the removal of state guarantees for German public banks in 2005.

Figure 3. Encumbrance trends in Europe
 Proportion of banking system balance sheet encumbered, per cent

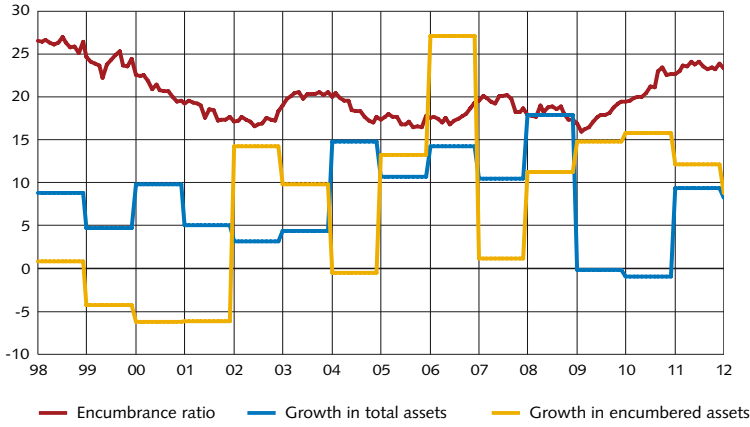


Note. LTRO and ELA stand for long-term refinancing operations and emergence liquidity assistance, respectively.

Source: Barclays Capital 2012.

In Sweden, system-wide encumbrance from covered bonds increased from 16 per cent in 2005 to 23 per cent in 2011 (see Figure 4). A clear upward trend started in 2009. Between 2009 and 2010, banks and mortgage institutions experienced zero growth in total assets. Nevertheless, the annual growth rate of covered bonds was around 15 per cent.

Figure 4. Encumbrance trend in Sweden
Per cent

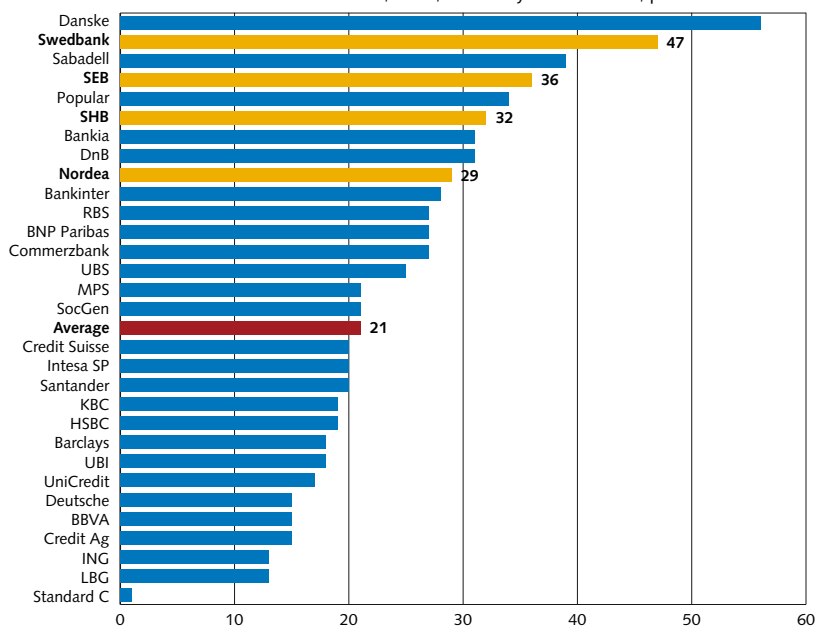


Note. Encumbrance ratio is a ratio of outstanding covered bonds to total assets adjusted for 30 per cent of over-collateralisation. This data includes the Swedish banks and mortgage institutions.

Sources: Statistics Sweden and the Riksbank.

Asset encumbrance tends to vary widely, not only across European countries, but also across European banks (see Figure 5). Consistent with the country-level evidence, banks headquartered in countries either with strong covered bond traditions (for example Danske Bank from Denmark) or problematic public finances (for example Sabadell and Bankia from Spain and Popular from Cyprus) tend to be at the top of the encumbrance list.

Figure 5. Asset encumbrance in European banks
 Encumbered assets to total assets, 2010, ranked by encumbrance, per cent



Note. This data refers to the end of 2010. Total assets exclude derivatives and encumbered assets exclude LTROs.

Source: Barclays Capital 2012.

It is also noteworthy that asset encumbrance also varies considerably across banks within a country. For instance, in Sweden, asset encumbrance varies between 30 and 50 per cent across the four largest banks.

Understanding asset encumbrance

The evidence in the previous section shows that asset encumbrance tends to vary systematically across banks and countries, and over time. In order to understand where such a systematic variation comes from and whether it poses a problem to financial stability, it is necessary to understand the effects of asset encumbrance.

EFFECTS OF ASSET ENCUMBRANCE: STRUCTURAL SUBORDINATION

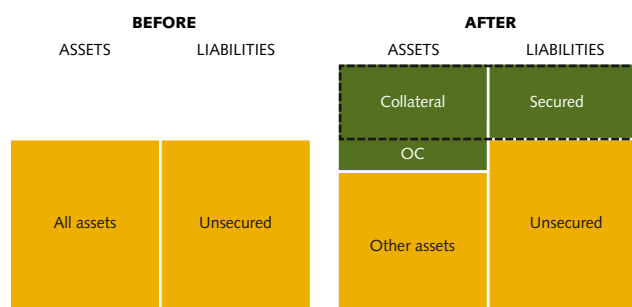
One of the effects of asset encumbrance is that it tends to shift risks to unsecured creditors, a process called structural subordination. Rational investors, however, do not need to be adversely affected since they can price in such increases in risk. Both of these statements are illustrated in the following section.

Figure 6 presents a graphical illustration of risk-shifting. Two balance sheets are shown, before and after the issuance of secured funding. The issuance of secured debt is somewhat

simplified here. It is assumed that the funds raised from secured funding are used to invest in assets which, in turn, are used as collateral for secured debt.

In this example, structural subordination of unsecured claimants takes place via two channels. Firstly, the issuance of secured funding is usually over-collateralised.⁵ This means that unsecured claimants finance some part of pledged assets or collateral. Secondly, even if the issuance of secured funding has no over-collateralisation, structural subordination usually takes place. As the value of collateral is usually risky, it may not be enough to satisfy the claims of secured investors. It is thus possible that some of the other assets could be used to satisfy the claims of secured creditors in the event of default. For both of these reasons, unsecured claimholders tend to be worse off after asset encumbrance increases. As will be illustrated below, structural subordination depends both on the degree of collateralisation as well as the relative share of secured and unsecured debt.

Figure 6. An issuance of secured funding



To understand the exact magnitude of risk-shifting, it is also useful to carry out a more rigorous analysis. In the following, we illustrate the effect of structural subordination for a hypothetical bank. We assume that secured and senior unsecured debts are the only funding sources available, and there is no equity. Asset encumbrance is captured by varying the ratio of secured debt to total assets and the degree of over-collateralisation. The bank is assumed to face a stress scenario in which its assets fall 20 per cent and it is forced to liquidate these assets.⁶

Recovery rates for unsecured and secured debt at different levels of asset encumbrance are presented in Figure 7. The average recovery rate is assumed to be 80 per cent, corresponding to a 20 per cent fall in assets.⁷ With no secured debt funding, the recovery rate for unsecured debt is equal to the average recovery rate.

As the share of secured debt in funding increases, the recovery value for unsecured debt falls below the average recovery rate, indicating that an additional risk is shifted to

5 Over-collateralisation is just another way to refer to haircut.
 6 This fall in assets does not necessarily have to come from credit risk. It can also stem from the fact that the fire-sale values of these assets might deviate from the assets' fundamental values.
 7 In professional terms, 20 per cent loss is called loss given default (LGD) and recovery rate is defined as one minus loss given default.

unsecured debt-holders. The recovery rate for secured debt also decreases, but it stays above the average recovery rate until secured debt is the only source of funding. Note also that the relation between recovery values for unsecured debt and the share of secured debt is non-linear. This means that the risk for structural subordination is larger for a bank with a high initial share of secured debt, as compared to a bank with a low initial share of secured debt to total funding.

Figure 7. Recovery rates and the share of secured debt to total funding in banks with no over-collateralisation
Per cent

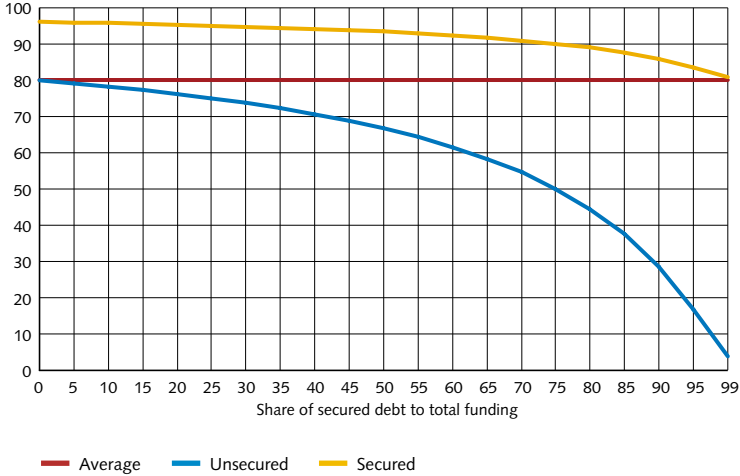
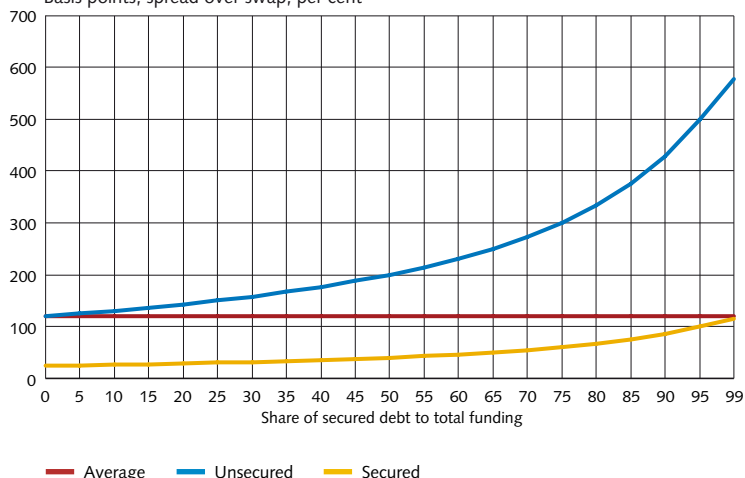


Figure 8 shows the effects of asset encumbrance on spreads for unsecured and secured debt.⁸ The calculation uses recovery rates from Figure 7. We can see that falling recovery rates from increased asset encumbrance lead to increased spreads for unsecured debt. Put another way, to compensate for the increases in risk stemming from asset encumbrance, a higher spread must be offered to investors holding unsecured debt.

⁸ This is done simply by multiplying loss given default by the risk-neutral probability of default, which we have assumed to be 6 per cent. Note also that an implicit assumption is that investors have all the information that is required to assess the riskiness of their claims.

Figure 8. Fair spreads and the share of secured debt to total funding in banks with no over-collateralisation

Basis points, spread over swap, per cent



To summarise, we have seen that asset encumbrance leads to structural subordination of unsecured claimants: their claims become riskier as asset encumbrance increases. We have also seen that rational investors can price in increased riskiness by requiring higher spreads. The analysis above can be recalculated for different levels of over-collateralisation, but the main implications remain unchanged (see Appendix).

As our analysis in this section is based on somewhat arbitrary parameter values, it is also worth mentioning how the analysis changes as parameter values change. Two parameters, the average recovery value for assets and probability of default, have direct implications for our results.

As assets are made riskier in terms of decreased average recovery value, structural subordination clearly increases: the gap between the average recovery value and the recovery value for unsecured creditors increases for any given level of asset encumbrance.

Increases in the probability of default do not affect the magnitude of structural subordination as defined above, but make the effects of asset encumbrance more relevant. In a situation where no bankruptcy ever takes place, recovery values clearly play no role. But as the probability of default increases, the effects of asset encumbrance become relevant for investors and are thus increasingly factored into spreads.

In conclusion, the effects of asset encumbrance on unsecured claimants are stronger at high levels of asset encumbrance (non-linearity) and at times of financial instability, when asset quality is low and the probability of default is high.

EFFECTS OF ASSET ENCUMBRANCE: LIQUIDITY RISKS

Secured funding and the resulting asset encumbrance can change also banks' liquidity risk, thereby affecting the banks' probability of default (see Figure 9). We focus here on two

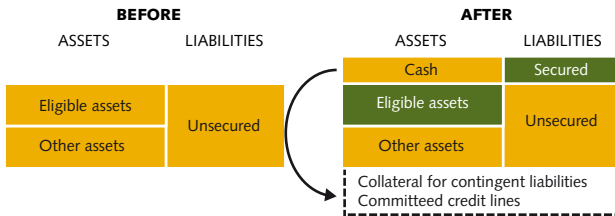
specific aspects of liquidity risks that relate to asset encumbrance and secured funding: the ability to obtain extra liquidity and the potential for increased outflows.

In the following, we ignore liquidity risks that relate to the stability of different funding sources (for example roll-over risks). The reason for this is that there is no *a priori* reason for why some funding sources are more stable than others. Indeed, certain imperfections, such as market segmentation, are necessary, and these imperfections are studied in the subsequent sections.

Unencumbered balance sheets allow banks to use eligible assets to obtain extra liquidity in the future, if needed (see Figure 9).⁹ In other words, low encumbrance of eligible assets acts as an unused liquidity buffer for banks which can be used for unexpected liquidity needs such as those from committed credit lines or margin calls related to derivatives.

It is important to note that the act of encumbering assets itself does not increase the probability of default. Indeed, the encumbrance of eligible assets tends to increase banks' cash holdings and thus helps decrease banks' immediate probability of default. But after a bank has made use of this extra cash holding, it will have a smaller liquidity buffer left to cover the remaining liquidity risks.

Figure 9. Liquidity risks and asset encumbrance



In addition to the banks' decreased ability to obtain extra liquidity, reliance on secured funding makes the banks' liquidity position more sensitive to changes in the underlying collateral. Whenever the value of collateral is decreased, the banks usually have to find additional collateral to offset this initial fall in the value of collateral. This contributes to the banks' liquidity risks during times of distress.

We conclude this section by highlighting the interaction between structural subordination and liquidity risks stemming from secured funding and asset encumbrance. Different funding sources usually have different maturity profiles. Consequently, encumbering assets and raising cash clearly benefits those investors in unsecured debt whose claims are maturing. Encumbering assets and raising cash may also benefit the remaining unsecured investors if the ultimate default is avoided by such encumbrance. In the event that the ultimate default cannot be avoided, the remaining unsecured debt holders are worse off as a result of such transactions, as the recovery value for their claims has decreased.

⁹ Traditionally, secured funding was meant to be used when funding could not be obtained in any other way (see Euromoney 2011).

Potential driving forces of asset encumbrance

The previous sections show that asset encumbrance leads to the structural subordination of unsecured creditors and can increase liquidity risks. This section takes these two effects as given and tries to understand the systemic variation of asset encumbrance across banks, countries and over time.

Dynamic aspects of asset encumbrance

As explained previously, low encumbrance of eligible assets can be viewed as an unused liquidity buffer. Given that banks act prudently in normal times, this extra liquidity buffer can and should be used in times of distress leading to counter-cyclical asset encumbrance.

Three main mechanisms lead to increased asset encumbrance in times of distress. Firstly, as the value or quality of existing collateral decreases, banks must post additional collateral to offset such a tendency. Secondly, confidence in banks usually falls in times of distress. This means that banks' counterparties may require a higher level of collateralisation in times of distress than in normal times. Thirdly, for some banks, unsecured markets may be closed entirely. In such circumstances, the only option for a bank may be recourse to secured markets.

Another potential driver of asset encumbrance over time may be changes in the demand for and supply of safe assets. Since the outbreak of the global financial crisis in 2008, investors' low risk appetite, together with a number of regulatory changes for banks, insurance companies and central counterparty clearing houses, has clearly contributed to increased demand for safe assets (see IMF 2012). At the same time, the collapse of high quality securitisation instruments, the sovereign debt crisis and the reduced re-hypothecation of collateral have decreased the supply of safe assets (see IMF 2012). In the light of these changes, the banks' issuance of secured assets represents the banks' response to the excessive demand for safe assets.

The recent changes in regulatory environment may also have contributed to asset encumbrance via their treatment of covered bonds. Many investors of bank debt enjoy extra benefits if they choose to invest in covered bonds instead of unsecured bank debt. The Basel III Accord partly allows the inclusion of covered bonds into the liquidity buffer in LCR. The Solvency II Directive requires insurance companies to hold less capital if they hold covered bonds as compared to unsecured debt. Central banks have typically lower haircuts for covered bonds than for unsecured debt (see, for example, ECB). Covered bonds are usually excluded from write-downs in many resolution frameworks, while unsecured debt is not (see, for example, the Danish system and the EU Commission's proposal on bail-in).

The dynamic aspects of asset encumbrance described above can account for some of the systematic variation in asset encumbrance observed in the data. For instance, the global financial crisis that started in 2008, together with the sovereign debt crisis, clearly seems to have contributed to the overall increase in encumbrance. Furthermore, the asymmetric allocation of the sovereign debt crisis in European countries also fits the observed increases in asset encumbrance across countries (for example Greece vs. Finland).

Structural aspects of asset encumbrance

This section focuses on asset encumbrance deriving from banks' structural preferences for secured funding. Before starting, it may be useful to recall that any risk-shifting that results from secured funding cannot, *per se*, be a driver of bank funding choice. Whenever a bank issues low-cost secured funding, this cost-saving should be exactly offset by higher spreads on unsecured debt, unless some imperfections exist that hinder such a mechanism.

Due to their safe nature, secured funding allows banks to cater to the needs of investors seeking safety. In the absence of secured funding, these investors would be forced to invest in riskier instruments than they would prefer. Thus secured funding lowers banks' funding costs, since banks escape offering the premium that would otherwise be required to attract these investors. In addition, when risks are allocated to the investors that are best suited to bear them, the roll-over risk for banks tends to decrease.

Another market imperfection that influences the banks' optimal balance of secured and unsecured funding is the possible uncertainty regarding the future encumbrance of the banks. Given that banks cannot commit to a certain level of asset encumbrance, unsecured creditors would price in such an uncertainty. As the riskiness of unsecured debt is more sensitive to changes in asset encumbrance than secured debt (see the section on structural subordination), banks can lower their funding costs by using secured as opposed to unsecured funding.

The possibility of catering to safe investors, the potentially lower roll-over risk and the relatively low sensitivity to future encumbrance all imply that the issuance of *some* secured funding can be optimal for banks. To identify their optimal funding structure, banks should balance these benefits of secured funding against the cost of secured funding, which mainly comes from increased liquidity risks from asset encumbrance (see the section on increased liquidity risks).

In reality, the optimal balance between secured and unsecured funding described above does not usually prevail due to implicit liquidity guarantees from central banks, the existence of risk-insensitive deposit insurance and potential rating arbitrage.

Implicit liquidity guarantees from central banks tilt banks' funding structures towards secured funding. In a systemic liquidity crisis, central banks usually provide extra liquidity to the system. This leads banks to ignore some of the liquidity risks from asset encumbrance, tilting bank funding toward secured funding.

Risk-insensitive deposit insurance is a clear driver of secured funding in banks with a large deposit base. The holders of senior debt and depositors face structural subordination in equal magnitudes in jurisdictions where deposits and senior unsecured debt have the same priority. But while the holders of senior unsecured debt can get compensated for this extra risk, the pricing of deposit insurance usually does not incorporate the effects of structural subordination. This means that banks with a large deposit base find it privately optimal to issue secured funding and shift risks to depositors.

The way rating agencies assign credit ratings can also tilt banks' preference toward secured funding such as covered bonds. To rate senior unsecured debt, rating agencies

focus primarily on the probability of default rather than recovery values (see Deutsche Bank 2011). This leads to rating arbitrage for banks since the increased share of secured funding does not lead to the downgrading of senior unsecured debt, even if it reduces the recovery value for unsecured creditors.

The drivers of asset encumbrance explained above are structural, that is they can explain the level of asset encumbrance but not necessarily trends. For instance, deposit insurance schemes already existed in 2005, so the presence of deposit insurance schemes to explain trends of asset encumbrance between 2005 and 2011 is limited.

Yet, when these structural drivers are combined with changes in the probability of default, they can also explain the trends. In stable times, the difference between the cost of secured and unsecured funding is minimal since default is distance. It can thus be argued that, before the outbreak of the global financial crisis in 2008, banks had relatively little to gain when they substituted unsecured funding with secured funding. But, as the probability of default increased in the post-crisis period, the gains became material and the banks had a stronger impetus for undertaking these transactions.

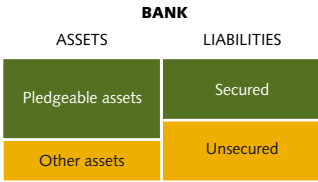
Additional aspects of asset encumbrance across countries and banks

Despite a number of factors mentioned previously, some aspects of asset encumbrance still remain unclear. For instance, it is unclear why there was such a large difference in asset encumbrance across countries in 2005 before the outbreak of the global financial crisis. Moreover, it is unclear why there is a systematic variation in asset encumbrance across banks within a given country. This section reviews two factors that can explain these puzzles: differences in countries' financial structures and banks' business models.

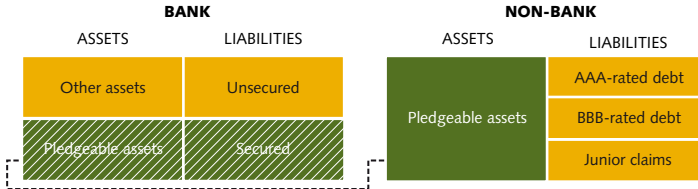
Differences in countries' financial systems form important driving forces of asset encumbrance across countries. For instance, one can make a distinction between countries with and without securitisation markets. In countries without securitisation markets (for example Sweden), banks follow an 'originate and hold to maturity' model, in which almost all loans are kept on the banks' balance sheets (see Figure 10). In such countries, asset encumbrance tends to be high, since credit intermediation takes place mainly within the banking sector and some of this credit intermediation (for example mortgages) is naturally funded with secured funding.

Figure 10. Asset encumbrance in bank-oriented and securitisation-oriented financial systems

Panel 1. Bank-oriented system



Panel 2. Securitisation-oriented system



In contrast, in countries with securitisation markets (for example the United Kingdom), banks usually engage in limited credit intermediation. A large part of credit intermediation takes place via securitisation. In this model, banks usually originate loans, but then sell some of these loans, such as mortgages, to other non-bank institutions which, in turn, fund these transactions with secured debt instruments. Clearly, in such a system, the importance of secured funding within the banking sector is reduced and the observed asset encumbrance in the banking sector tends to be low.

Another important explanatory factor behind the high variation in asset encumbrance across banks is the banks’ business model. Due to regulatory restrictions, only certain lending can be financed with secured funding. For instance, a common type of lending that is allowed in most covered-bond legislations is mortgage lending. A mortgage-focused bank thus tends to have more asset encumbrance than a bank with a non-mortgage focus.

Are regulatory actions motivated?

In the previous sections, we discussed various driving forces behind asset encumbrance. Factors such as the level of financial stress, existing and forthcoming regulatory frameworks, the balance between the demand for and supply of safe assets, banks’ business models, differences in countries’ financial markets and frictions in financial markets all tend to influence asset encumbrance. Importantly, some of these factors determine the socially-optimal level of asset encumbrance, while others tend to push asset encumbrance above the level that is socially optimal.

The socially-optimal level of asset encumbrance is an abstract concept, but has a clear definition: the point at which social benefits and costs are maximised (see Figure 11). Factors that determine the social optimum include excessive demand for safe assets, differences in countries’ financial markets, frictions in financial markets and the liquidity risk that stems from asset encumbrance.

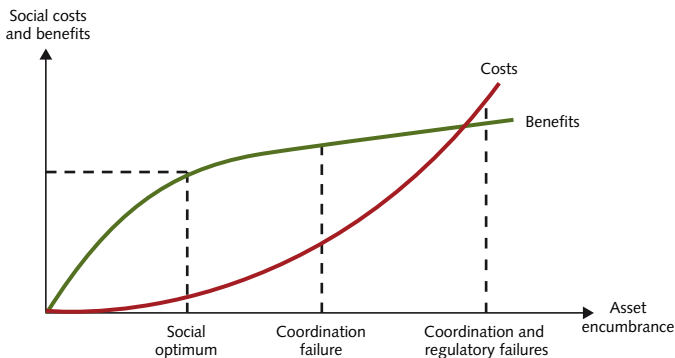
The social optimum has a clear country and time dimension, as factors mentioned above can vary across countries and over time. For instance, the excess demand for safe assets clearly varies over time, depending on the underlying demand and supply dynamics in each country. In addition, countries with more traditional banking sectors clearly have a higher level of socially-optimal asset encumbrance than countries with less traditional banking sectors, simply because more credit intermediation takes place within such a banking sector.

Regulatory actions are necessary when there is more asset encumbrance in the banking sector than is socially optimal. Excessive asset encumbrance can result from the following imperfections and market failures:

- the failure of individual banks to internalise the effects on the system, also known as coordination failure;
- regulatory imperfections;
- self-fulfilling expectations, especially in times of distress.

Coordination failure relies on the idea that a decentralised banking sector may lead to a suboptimal outcome (see Korinek 2012). In the context of asset encumbrance, coordination failure can be applied to liquidity risks. The argument is that, in the absence of counterbalancing measures, individual banks choose a level of asset encumbrance that is consistent with their own liquidity profile, but which is suboptimal at the system level. The resulting decentralised outcome has too much liquidity risk at the system level, as banks fail to incorporate possible fire-sale externalities when they choose their optimal liquidity profile (see Figure 11).

Figure 11. The optimal level of asset encumbrance



In addition to coordination failure, there are various regulatory imperfections that distort banks' funding choices, with an end result of excessive asset encumbrance (see Figure 11). Banks can derive pure private gains from the issuance of secured funding, leading to excessive encumbrance due to the following reasons:

- Secured funding tends to shift risks to tax-payers via risk-insensitive deposit pricing. Unlike other unsecured investors, depositors do not price in the increased risk at default (that is, decreasing recovery values) that may result from the issuance of secured funding.
- Secured funding shifts risks to tax-payers via implicit liquidity guarantees. Due to implicit liquidity guarantees from central banks, banks rationally tend to ignore some of their liquidity risks. To the extent that the liquidity provision is costly, these costs are shifted to tax-payers.
- Secured funding shifts risk to unsecured debt-holders, which, considering the presence of implicit state guarantees, may fail to factor in such increases in risks.
- The upcoming regulatory reforms such as the Basel III Accord, the Solvency II Directive and bail-in regulations favour secured, as opposed to unsecured, funding.

Another market failure that may lead to excessive asset encumbrance is the possibility of self-fulfilling expectations (see also Haldane 2012). Whenever investors of unsecured debt expect a high level of asset encumbrance, the rational response from banks is indeed to rely on secured funding, leading encumbrance to increase in a self-fulfilling manner. Such expectations may ultimately lead to a financial system with a maximum degree of secured funding. In such a system, the benefits of having secured funding are minimal (since almost everybody would be secured), but all the negative side-effects of secured funding, such as increased liquidity risks, would be present.

What separates the possibility of self-fulfilling expectations from coordination failure and regulatory imperfections is the underlying time dimension. Both coordination and regulatory imperfections push banks to take excessive levels of asset encumbrance in normal unstressed situations. However, self-fulfilling expectations tend to become important in times of distress, when the optimal level of asset encumbrance increases temporarily to accommodate the stress. It is after the realisation of the stress that self-fulfilling expectations may start hindering a return to the lower socially-optimal level of asset encumbrance.

Potential policy remedies

As a result of market failures and regulatory imperfections, asset encumbrance can become excessive. Various regulatory measures can be undertaken to constrain these excesses. Below, we review a gross list of regulatory measures that can be undertaken and highlight the pros and cons of each such measure.

Perhaps the most important difference between these various regulatory measures is their reliance on the measurement of the socially-optimal level of asset encumbrance. We therefore make a distinction between two groups of measures: those that increase market discipline and those that rely explicitly or implicitly on knowing the socially-optimal level of asset encumbrance.

Regulatory tools that increase market discipline

A straightforward way to increase market discipline is to increase the transparency of asset encumbrance. The current disclosure of asset encumbrance is, in general, poor and varies considerably across banks and different sources of asset encumbrance. Asset encumbrance from covered bonds is probably the best-documented form of encumbrance. Yet, even for covered bonds, there are no widely-accepted disclosure standards. Asset encumbrance from repos and derivatives has limited or no disclosure at all. For instance, repos that banks have undertaken with the ECB represent a good example in this category, as neither banks nor the ECB routinely disclose such information.

The existing low disclosure of asset encumbrance is puzzling. After all, any uncertainty around the existing level of asset encumbrance should be factored in by unsecured investors which, in turn, would give banks clear incentives to disclose this information.¹⁰ Yet, as explained previously, the presence of government guarantees can make investors somewhat reluctant to demand such information. In addition, banks may choose not to disclose such information, even if investors would strongly require it. Arguably, the market for senior unsecured debt is currently closed for a number of European banks, partly because of limited disclosure on encumbrance.

In this light, a mandatory disclosure of encumbrance would increase investors' awareness of asset encumbrance risks and contribute to a better pricing of these risks. As the effects of encumbrance are factored into the spreads, banks would have better incentives to keep a proper balance between secured and unsecured debt, lowering potentially excessive encumbrance.

Another way to increase market discipline is to reduce banks' incentives to shift risks to tax-payers via deposits. Recall that deposits are a form of unsecured funding for banks and thus are affected by asset encumbrance and structural subordination. There are three ways to limit the deposit arbitrage.

The first option is to make the deposit insurance premium sensitive to asset encumbrance. Since depositors are a form of senior unsecured creditors, the fair risk-adjusted deposit insurance premium is observable from the market.

The second option is to increase seniority of deposits above that of senior unsecured debt. This clearly reduces banks' deposit arbitrage. It also discourages other senior unsecured debt holders somewhat, at least temporarily. Yet this should not be seen as problematic as these investors can and should request compensation for it.

The third option is to separate mortgage taking from deposit taking. In countries like Sweden, this would imply that mortgage institutions would act as separate entities with no ties with their mother institutions. A downside to this solution is a changed financial system,

¹⁰ Whenever a bank does not disclose, investors would price in the expected or average encumbrance. Given this, banks with below average encumbrance would disclose since the cost of debt would decrease in this case. Investors would update their expectations of encumbrance for the remaining banks, leading to a new wave of disclosure.

which, in turn, could have unexpected consequences, such as the creation of securitisation markets.

Some of these solutions have been implemented internationally. The seniority of deposits over other senior unsecured debt has been proposed in some jurisdictions such as the United Kingdom, while, for instance, in the United States such seniority already exists (see Barclays Capital 2012). The separation of deposit taking from mortgage lending is essentially taking place in Denmark, where mortgage institutions are specialised lenders with no deposit funding or implicit funding ties from their mother companies (see ECBC 2011).

Regulatory tools with a concrete knowledge of the social optimum

A clear way to limit excessive encumbrance is to set explicit limits on asset encumbrance. Such limits would control the level of encumbrance mechanically, preventing it from increasing above the socially-optimal level. Yet such limits might also discourage unsecured creditors. In normal times, such limits control the level of encumbrance. However, in stressed situations, when unsecured markets would be closed, such limits would not allow banks to tap secured funding markets, increasing the probability of default. In essence, these limits would incentivise banks to be prudent in normal times but would also undesirably restrict banks' actions in times of distress.¹¹

In addition, such limits have a number of operational drawbacks. Firstly, it is unclear how the socially-optimal level of asset encumbrance could be determined. Clearly, financial systems in different countries tend to call for different limits. But, even within a country, such limits give rise to discriminatory effects across banks with different business models. Secondly, it is unclear which measures of asset encumbrance should be used. A large part of encumbrance, such as that from derivatives, is hard to measure. This may lead to a measure of encumbrance that focuses on the visible part of encumbrance, such as that from covered bonds.

Many of the limitations of setting explicit limits on asset encumbrance could be avoided via the usage of soft limits instead. One way of doing this could be by linking capital requirements under Pillar II¹² to the level of encumbrance. Whenever encumbrance starts to increase above the social optimum, banks would be required to hold more capital. The increased capital requirements would motivate banks to reduce encumbrance. It would also reduce the structural subordination of unsecured investors either via decreasing the probability of default or lowering the loss that must be borne by debt holders.

Importantly, such soft limits could vary over time and be based on different measures of encumbrance. The flexibility coming of the soft limits may prove to be useful as we learn how the financial system responds to such limits. These soft limits may also be used to constrain the self-fulfilling expectations that are likely to follow after the period of high

¹¹ In essence, the optimal regulation would call for the counter-cyclical limits: in good times, low limits will be desirable, while in bad times, high limits will be preferred.

¹² Pillar II refers to the supervisory review process under the Basel II Accord.

encumbrance and distress. The Pillar II treatment also has drawbacks, mainly in the form of low transparency and the banks' incentives to resist and manipulate these soft limits.

Some of the policy options outlined previously have been already considered internationally. Countries like Australia, Canada, Greece, Italy, New Zealand, Spain and the United States have placed explicit limits on the issuance of covered bonds (see ECBC 2011). In the United Kingdom and the Netherlands, the regulation of asset encumbrance follows soft limits in the sense that the limits are determined on a case-by-case basis (see Bank of America Merrill Lynch 2011).

Concluding remarks

Asset encumbrance varies over time and across banks and countries. Some of this variation can clearly be justified by the level of financial stress, the demand and supply dynamics of safe assets, and differences in countries' financial systems and banks' business models. Yet there are also good reasons to believe that asset encumbrance can become excessive when left unregulated. The failure of individual banks to internalise their liquidity effects on other banks, together with existing and upcoming regulatory imperfections, could lead to situations in which asset encumbrance is higher than is socially optimal. Regulators have a wide range of tools available to counterbalance potential excesses in asset encumbrance. Some of these tools rely on knowledge of the socially-optimal level of asset encumbrance, which may be hard to measure. Consequently, tools that rely less on this knowledge have a clear comparative advantage.

References

- Alloway, Tracy (2011), Bank Collateral Drying up in Rush for Security, *Financial Times*, 19 October 2011.
- Bank of America Merrill Lynch (2011), The Cost of Asset Encumbrance, *Bank of America Merrill Lynch Credit Analysis*.
- Barclays Capital (2012), Over-promising? Encumbrance at European banks, *Barclays Capital Equity Research*.
- Deutsche Bank (2011), Cédulas Account for 17% of YTD Covered Bond Issuance, *EUR Liquid Credit Weekly*.
- ECBC (2011), *European Covered Bond Fact Book*, 2011.
- Euromoney (2011), Covered bonds: Funds-starved Banks Risk Collateral Damage, *Euromoney*.
- Haldane, G Andrew (2012), Financial Arms Races, *Bank of England Speech*.
- IMF (2012), *Global Financial Stability Report*, April 2012.
- Korinek, Anton (2012), Systemic Risk-Taking: Amplification Effects, Externalities, and Regulatory Responses, *Working Paper*.

Appendix: The effects of over-collateralisation on recovery rates and spreads

The analysis in the section on structural subordination can also be carried out for different levels of over-collateralisation (see Figure A1). Increasing asset encumbrance via increased over-collateralisation leads to a higher degree of structural subordination for a given share of secured debt to total funding. An interesting aspect to note, however, is that an increased asset encumbrance via over-collateralisation leads to increased recovery rates for secured debt. This can be contrasted with asset encumbrance via an increased share of secured debt in total funding, which leads to decreasing recovery rates both for secured and unsecured bondholders.

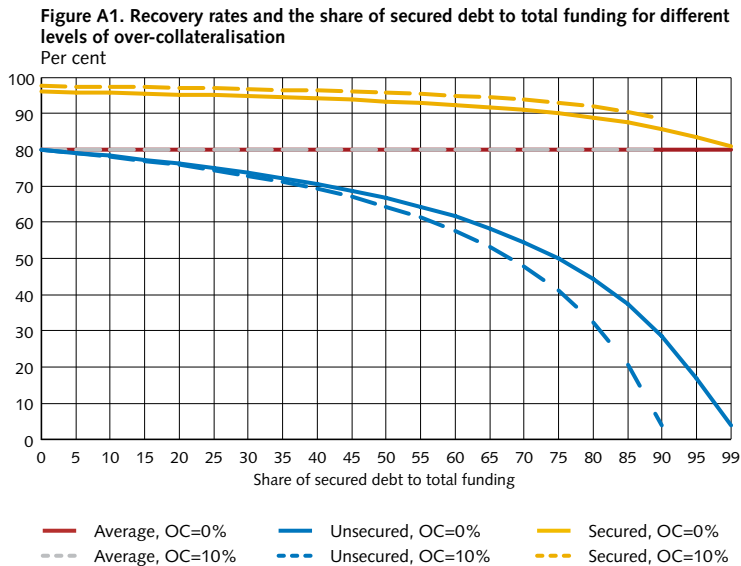
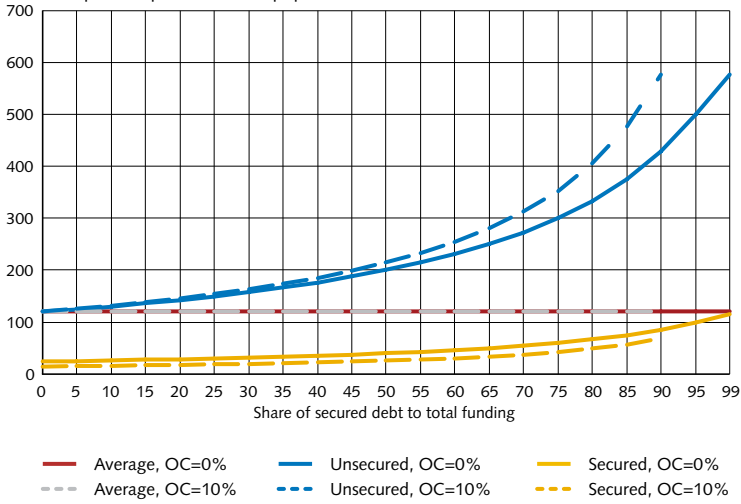


Figure A2 shows the effects of asset encumbrance on spreads for unsecured and secured debt. The calculation uses recovery values from Figure A1. We can see that falling recovery values from increased asset encumbrance lead to increased spreads for unsecured claimants. The effect of asset encumbrance on spreads for secured debt depends on the type of asset encumbrance. A new issuance of secured debt with unchanged over-collateralisation leads to increasing spreads for secured debt, while increased over-collateralisation for a given fraction of secured debt in total funding leads to decreased spreads.

Figure A2. Fair spreads and the share of secured debt to total funding for different levels of over-collateralisation

Basis points, spread over swap, per cent



Cards or cash. How should we pay?

BJÖRN SEGENDORF AND THOMAS JANSSON

Björn Segendorf works at the Financial Stability Department and Thomas Jansson at the Monetary Policy Department of the Riksbank.

Inexpensive and easy-to-use payment services promote trade and prosperity. The efficiency of the payment system is thus an important social issue. In this article we give a brief account of the social costs of cash and card payments and discuss how the payment system could be used more efficiently. We find that cash and cards are used relatively efficiently in the Swedish payment system. However, there is potential for further reducing payment costs by using debit cards to a greater extent than today. The payment system could also be used more efficiently if transparent and cost-based charges were introduced. Consumers may find it easier to accept such charges if they are compensated for them and provided with simple and targeted information.

A payment is a transfer of a monetary value from one party to another, often as compensation for a good or a service.¹ When the buyer attaches a higher value to the good or service than the seller a transaction can take place, and both parties will be better off than if no transaction had occurred. However, when the good or service is exchanged for a payment, costs arise for making the payment. These costs are always borne in one way or another by the seller or the buyer.² In practice, this means that it will only be possible to carry out the transaction if the buyer's valuation of the good exceeds the seller's by at least as much as it costs to make the payment. The greater this cost, the fewer the number of potentially mutually favourable transactions that will be carried out. Similarly, lower payment costs contribute to increased trade and, by extension, greater prosperity. It is therefore in the interests of society that expedient and inexpensive payment methods are available.

The aim of this article is to discuss how the use of cards and cash affects efficiency in the Swedish payment system. We begin by discussing different types of payment, their basic characteristics and the difference between private costs and social costs. We then examine the costs for payments using cash and cards in Sweden in more detail and also make international comparisons. Finally, we discuss whether cash and cards are used efficiently and possible measures to increase efficiency.

1 Payments have three different functions in the economy: they make it possible to trade in goods and services, to convert savings to investments and to redistribute financial risks. The term payment as used in this article refers to compensation for goods or services, but in principle the reasoning also applies to payments in other contexts.

2 There is also often a cost for delivering the goods, either directly in monetary terms or, more indirectly, in terms of time and effort. Together with the cost of making the payment, these costs constitute what is usually referred to as a transaction cost. This article focuses on that part of the transaction cost that stems from the payment.

THE DESIGN AND COST STRUCTURE OF THE PAYMENT SYSTEM

At an overall level, there are two types of payment in Sweden: *cash* payments and *account-based* payments. Cash payments are made at the point-of-sale by one of the parties to the transaction handing over the monetary value directly to the other party. In an account-based payment, the transaction is carried out by transferring the monetary value electronically between accounts in special custodial institutions, which are usually commercial banks. Here the banks act as intermediaries for the payment by receiving instructions to transfer a certain sum from one account to another and carrying out the payment without the further involvement of the payer or payee.

Cash payments require considerable manual handling of the cash on the part of both the payer and the payee, but also on the part of banks, cash-in-transit companies and others. Account-based payments, on the other hand, require a financial infrastructure within which payment information is processed. This infrastructure mainly consists of different types of IT system such as online banking systems, card terminals, switches that direct information and clearing platforms.³ In contrast to cash payments, account-based payments usually require little manual work.

A common feature of both types of payment is that they are both associated with considerable *economies of scale*. Setting up a logistical infrastructure to manage the flow of cash in the economy is costly, and it is also costly to develop the financial infrastructure required for account-based payments. However, once the infrastructure is in place, the marginal costs for handling marginally more cash or an additional account-based payment are low. This means that the average cost of a payment of a certain type tends to fall as the number of payments increases. In the case of account-based payments there are also possible synergy effects that further increase the economies of scale in that different types of account-based payment can be wholly or partly processed within the same infrastructure. This reasoning is general and applies to all types of monetary payments in Sweden. However, hereafter we will focus exclusively on payments using cash and cards. These two payment methods can be substituted for one another and are mainly used at the point-of-sale.

Both cash payments and card payments have their advantages and disadvantages. Cash payments transfer the monetary value directly when the cash is handed over and can be a quick and easy-to-use alternative when, for example, making smaller payments at the point-of-sale or for payments between private individuals. In addition, no special infrastructure is required to receive cash payments. A disadvantage from the user's perspective is that cash does not earn interest. The interest that is not earned by the person holding the cash gives rise to the corresponding interest income at the Riksbank. This revenue from the issue of banknotes and coins is called *seignorage*. Other potential disadvantages are that the purchasing power of the cash holder is limited to the amount

³ Clearing is a collective term for the processing of payment information that forms the basis for the exchange of payments between intermediaries.

of cash he or she holds and that there is a risk of counterfeiting. On the other hand, payments made using debit or credit cards give the buyer access to more purchasing power and can be less time consuming when paying somewhat larger amounts.⁴ The potential disadvantages of cards are that they are not accepted everywhere (the infrastructure is lacking), payments cannot be made if there are technical problems and that there is a risk of card fraud, for example skimming.

Different payments are also more or less appropriate in different situations, and the preferences of the user are also to some extent a matter of taste. However, the choice of payment method also governs the cost of the payment and, ultimately, the total cost for payments in Sweden. The lower the transaction cost is, the more efficient trade will be and the greater prosperity will be. Below, we therefore outline the costs for payments using cash, debit cards and credit cards and discuss to what extent the Swedish payment system is used efficiently.

Private and social costs

An important distinction in this context is that between *private costs* and *social costs*. The former are the costs incurred by an agent or a sector, for example the banking sector. These are the costs for the resources consumed by the agent or sector in connection with the payment and the fees they pay to other agents or sectors. For example, the banks pay the cash-in-transit companies for transporting cash and consume their own resources, such as labour and premises, in connection with cash management. A bank's revenues from the sale of payment services can be deducted from these private costs to arrive at the private net cost that represents the bank's profit or loss. The bank can also earn income from the retail trade by receiving daily takings and providing change.

The social costs, on the other hand, reflect the value of the overall resources consumed in connection with a payment, or in total for a certain type of payment, and thus constitute a measure of the value of other goods and services that could otherwise have been produced with these resources. However, the social costs cannot be calculated by adding together the private costs of all the agents involved as this would entail a certain degree of double counting. Chart 1 gives an example of this in the case of cash management.

Chart 1. An example of the difference between private and social costs

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th colspan="2" style="text-align: left; padding: 5px;">BANK</th></tr> <tr><td style="padding: 5px;">Consumed resources</td><td style="text-align: right; padding: 5px;">100</td></tr> <tr><td style="padding: 5px;">Fee paid</td><td style="text-align: right; padding: 5px;">100</td></tr> <tr><td style="padding: 5px;">Private costs</td><td style="text-align: right; padding: 5px;">200</td></tr> <tr><td style="padding: 5px;">- Fee received</td><td style="text-align: right; padding: 5px;">0</td></tr> <tr><td style="padding: 5px;">Net cost</td><td style="text-align: right; padding: 5px;">200</td></tr> </table>	BANK		Consumed resources	100	Fee paid	100	Private costs	200	- Fee received	0	Net cost	200	<p>Fee</p> <p>→</p> <p>100</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><th colspan="2" style="text-align: left; padding: 5px;">CASH-IN-TRANSIT COMPANY</th></tr> <tr><td style="padding: 5px;">Consumed resources</td><td style="text-align: right; padding: 5px;">100</td></tr> <tr><td style="padding: 5px;">Fee paid</td><td style="text-align: right; padding: 5px;">0</td></tr> <tr><td style="padding: 5px;">Private costs</td><td style="text-align: right; padding: 5px;">100</td></tr> <tr><td style="padding: 5px;">- Fee received</td><td style="text-align: right; padding: 5px;">100</td></tr> <tr><td style="padding: 5px;">Net cost</td><td style="text-align: right; padding: 5px;">0</td></tr> </table>	CASH-IN-TRANSIT COMPANY		Consumed resources	100	Fee paid	0	Private costs	100	- Fee received	100	Net cost	0
BANK																										
Consumed resources	100																									
Fee paid	100																									
Private costs	200																									
- Fee received	0																									
Net cost	200																									
CASH-IN-TRANSIT COMPANY																										
Consumed resources	100																									
Fee paid	0																									
Private costs	100																									
- Fee received	100																									
Net cost	0																									

4 Debit cards are cards where the payment is debited directly from the buyer's bank account. Charge cards are cards where the bank periodically invoices the card holder the total value of the payments. Credit cards give the card holder the opportunity to wait before paying all or parts of the invoiced sum. In this article, we use the term credit card collectively for both credit and charge cards.

In the example in Chart 1, a bank consumes SEK 100 in personnel costs (consumed resources) to sort cash that will be paid back to the Riksbank. The bank then gets a cash-in-transit company to transport the cash to the Riksbank at a cost of SEK 100 (fee paid). The cash-in-transit company's cost for producing the transport service is SEK 100 for personnel, fuel and the depreciation of the vehicle used (consumed resources). At the same time, the company earns income of SEK 100 (fee received). If we ask the bank and the cash-in-transit company to report their costs the bank will report SEK 200 and the cash-in-transit company SEK 100. The private costs will thus total SEK 300 while the total for the resources consumed is SEK 200, and it is the latter sum that is the social cost in this example.

The difference between the private cost and the social cost is thus SEK 100, which is equivalent to the fee paid by the bank to the cash-in-transit company. In order to convert private costs to social costs, one must thus discount the fees that different agents pay to each other. This can be done in two ways: either at the payer (the bank) or at the payee (the cash-in-transit company). If we exclude the fee from the payer we will get the value of the resources consumed by the payer. If we choose instead to include the fee as a negative cost for the payee we will get a net cost that corresponds to the payee's corporate loss or, alternatively, a negative net cost that corresponds to a corporate profit. Note that both total consumed resources and total net costs add up to SEK 200. Both ways thus give the social cost, but how this cost is shared between the parties differs depending on how one chooses to calculate it.⁵ In terms of consumed resources both parties bear a cost of SEK 100, but in terms of net cost the bank bears the entire social cost of SEK 200.

There are of course more than two parties involved in cash and card payments in society as a whole, but the principle is the same as in Chart 1. One has to take into account all the payments between the different agents that relate to the purchase of services between them to arrive at the social costs. The Riksbank's seignorage should also be discounted. For all sectors except the consumers, the costs that we present in this article include costs for various security measures such as credit assessments and losses resulting from fraud in the case of card payments and security features, inspection and losses from counterfeit banknotes and coins.

From an economic perspective, the profit that an agent makes from a fraud and the resulting loss suffered by another agent are regarded as a transfer from one party to another. On the other hand, the cost of the real resources consumed by countermeasures is a social cost. Our calculations do not take into account integrity aspects and the possible costs that, for example, consumers may perceive from knowing that their card payments are registered by the bank, and sometimes by the retailer concerned.⁶

5 Externalities are consequences that affect someone other than those that cause the event concerned. A common example is environmental pollution. For the total in the table to really reflect the social cost one needs to assume that there are no externalities or that the prices paid by the agents on the market fully take into account any externalities.

6 Our analysis differs from a traditional social cost benefit analysis in that we focus on the cost side and do not take into account any social benefits. We thus do not include, for example, the sense of satisfaction or security that consumers feel when paying in cash or by card.

THE COSTS OF CASH AND CARD PAYMENTS

In 2010 and 2011, we collected information on behalf of the Riksbank on the costs of payments made using cash, debit cards and credit cards in 2009. The information was collected from the Riksbank, cash-in-transit companies, banks and consumers. Only payments from consumers to businesses were included.⁷ Table 1 presents the calculated number of payments and the total values for different payment methods.^{8,9}

Table 1. Number and value of cash and card payments from consumers to companies in Sweden in 2009

	CASH	CARD		TOTAL
		DEBIT	CREDIT	
Total value (SEK, billions)	261	550	123	673
Number of transactions (millions)	1 034	1 337	240	1 577
Average transaction value (SEK)	252	411	513	427

Source: Segendorf and Jansson (2012).

PAYMENTS AT THE POINT-OF-SALE

Cash, debit cards and credit cards are currently the payment instruments that are used to make payments at the point-of-sale. The agents that have costs in connection with cash payments are the Riksbank, cash-in-transit companies, banks, businesses and consumers. In the case of card payments, only the latter three agents incur costs.

Table 2 presents the private costs, fees and social costs for cash payments in the form of consumed resources and net costs, as in the example in Chart 1. The private costs are presented in the central column in the table. To the left of the central column we deduct the income an agent has earned in the form of fees and seignorage and thus arrive at the net cost that reflects the profitability of cash operations. To the right of the central column we deduct the fees paid from each agent's private costs and thus arrive at the resources that each agent has consumed. Both the total for net costs and the total for consumed resources show the social costs of cash and we can also see that the totals in the left and right columns are the same. The only difference between the columns is that the respective costs are distributed in different ways between the parties. The right-hand column shows where the resources have been consumed while the left-hand column shows the net cost of a certain sector. A negative net cost for, for example, the Riksbank, entails a corporate profit.

⁷ This study used a method developed and applied in collaboration between 13 European central banks under the coordination of the European Central Bank (ECB). It means that cost calculations from different countries can be compared for the first time.

⁸ For a description of the method, cost data and results, see Segendorf and Jansson (2012).

⁹ For a definition of debit cards and credit cards, see footnote 4.

Table 2. Cost of cash payments in Sweden in 2009, SEK million

	NET COST	FEES RECEIVED	SEIGNORAGE RECEIVED	PRIVATE COSTS	SEIGNORAGE PAID	FEES PAID	CONSUMED RESOURCES
The Riksbank	-5 566	6	5 800	240		30	210
CITs	-20	1 650		1 630		6	1 624
Banking sector	2 090	1 270		3 360	540	350	2 470
Retailers	6 720			6 720	620	1 820	4 280
Consumers	5 380	50		5 430	4 640	770	20
Sum	8 604	2 976	5 800	17 380	5 800	2 976	8 604

Note. Both the column *net cost* and the column *consumed resources* add up to the social costs. The two columns therefore represent different ways of illustrating who bears the costs.

Source: Segendorf and Jansson (2012).

In the case of the Riksbank, the private costs amounted to SEK 240 million and the largest cost items were for the purchase of new banknotes and coins and storage. Seignorage amounted to SEK 5.8 billion, SEK 4.6 billion of which came from the consumers while the remaining SEK 1.2 billion was more or less evenly divided between banks and companies. Most of the cash-in-transit companies' private costs of SEK 1.6 billion arose in connection with transport and with cash management at the depots. Their income amounted to just over SEK 1.6 billion, of which the banks paid approximately SEK 350 million and the corporate sector around SEK 1.3 billion. The banks' largest private cost of almost SEK 3.4 billion stemmed from withdrawals and deposits. Part of this sum consisted of fees to the cash-in-transit companies and seignorage. The fees received by the banks consisted of fees from the corporate sector of almost SEK 500 million and of fixed card fees from consumers of SEK 770 million. Here we have divided the consumers' fixed fees for cards between cash withdrawals and card payments in proportion to the number of transactions. For their part, the businesses had substantial costs in the form of fees and personnel costs for receiving payments over the counter and checking the takings at the end of the working day. Personnel costs comprise wage costs including social security contributions of SEK 1.8 billion, calculated on the basis of over one billion cash payments which in 2009 took an average of approximately 26 seconds each. The costs for checking takings and other back-office activities amounted to approximately SEK 2.1 billion. The major costs for consumers related to seignorage, fees to the banks and the time required to make the cash payments.¹⁰ The total social cost of cash payments thus amounted to SEK 8.6 billion, which gives an average cost per cash payment of SEK 8.30. Real resources are mainly consumed in the corporate sector, but also banks and, to a lesser extent, cash-in-transit companies, consume resources when managing cash payments. However, the Riksbank and the consumers consume very little real resources. On the other hand, the distribution of net costs is different. Here the Riksbank makes a considerable profit (negative net cost) due to

¹⁰ The consumers' time cost has been calculated using a model in which the time cost is a function of the interest rate and the number of payments and withdrawals. The time cost was low in 2009, in part due to the low interest rate in this year. See Segendorf and Jansson (2012) for a discussion of this.

the income from seignorage, while the corporate sector and the consumers bear the major part of the costs. The banks also make a loss on their cash operations.¹¹

Tables 3 and 4 present the costs for payments using debit cards and credit cards. In the case of payments using debit cards, the management of payments vis-à-vis companies, including costs for IT and communication, represented the banks' largest cost item. Other large cost items were customer services, the authorisation of payments and other checks. In the case of credit cards, there were also costs for marketing, credit assessments and more extensive customer services including bonus programmes. Income consisted mainly of fees from businesses and consumers and transaction charges from companies. The average fee amounted to almost SEK 0.80 for payments using debit cards and SEK 7.30 for payments using credit cards. Note the substantial difference in the profitability of the banks (negative net cost) for debit and credit cards in the tables.¹²

The businesses' major costs comprised fees to the banks and the time taken to receive payments over the counter. It was estimated that a card payment takes 25 seconds. In the case of debit cards, the time cost was the highest at approximately SEK 2.3 billion, while fees amounted to almost SEK 1.4 billion. The situation was the reverse for credit cards where the fees were much higher (just over SEK 2 billion) than the time cost (just over SEK 400 million). The major cost item for the consumers was the fixed fees paid to the banks, while their time cost, as in the case of cash, was negligible.

The social costs for payments using debit and credit cards amounted to almost SEK 6 billion and SEK 2.8 billion respectively. The average social cost for a debit-card payment was thus just under SEK 4.50 and for a credit-card payment SEK 11.70. The considerable difference is partly explained by the advantages of scale that the larger number of debit-card payments entails compared with the relatively few credit-card payments.

Table 3. Cost of debit-card payments in Sweden in 2009, SEK million

	NET COST	FEES RECEIVED	PRIVATE COSTS	FEES PAID	CONSUMED RESOURCES
The Riksbank					
CITs					
Banking sector	230	3 240	3 470	40	3 430
Retailers	3 890		3 890	1 380	2 510
Consumers	1 840	70	1 910	1 890	20
Sum	5 960	3 310	9 270	3 310	5 960

Source: Segendorf and Jansson (2012).

11 We wish to underline the fact that whether the banks make a loss or not is highly dependent on whether one chooses to include the net interest income from the current accounts that the public have with the banks. If this is included the banks make a substantial profit. For further discussion, see Segendorf and Jansson (2012).

12 As in the case of cash, we have not included net interest income for current accounts. If net interest income is included, the profitability of debit cards increases significantly, see Segendorf and Jansson (2012).

Table 4. Cost of credit-card payments in Sweden in 2009, SEK million

	NET COST	FEES RECEIVED	PRIVATE COSTS	FEES PAID	CONSUMED RESOURCES
The Riksbank CITs					
Banking sector	-650	3 490	2 840	490	2 350
Retailers	2 470		2 470	2 010	460
Consumers	990	490	1 480	1 480	0
Sum	2 810	3 980	6 790	3 980	2 810

Source: Segendorf and Jansson (2012).

The social cost of cash payments is thus largely the same as the joint social cost of debit-card and credit-card payments. However, the average cost is significantly higher for a cash payment (SEK 8.30) than for the weighted average cost of a debit or credit-card payment (SEK 5.50). This difference is due to the fact that the economies of scale are greater for card payments than for cash payments.

Below we discuss whether the Swedish payment system is efficient. We begin by investigating whether the costs in Sweden are reasonable in an international perspective.

THE SWEDISH PAYMENT SYSTEM IS EFFICIENT IN AN INTERNATIONAL PERSPECTIVE

The costs of payments and the efficiency of the payment system have been studied in several countries. In this article we focus on comparisons with countries that are similar to Sweden and with countries that have applied the same calculation method as that used to calculate the Swedish costs. The Swedish study is part of a coordinated programme involving 13 national central banks in the EU and the ECB. So far, only Denmark and Hungary have published national reports following this programme. It is therefore natural to compare Sweden with these two countries as their results are directly comparable.

There are also other studies that have used a somewhat different method to calculate the costs of payments. The difference between these studies and those coordinated by the ECB mainly relates to what costs the studies have chosen to include. The results of the different groups of studies are not therefore directly comparable, but can nevertheless be used to give an indication of the differences between countries. Two studies that are appropriate for comparison are those conducted by Gresvik and Haare (2009) in Norway for 2007 and by Bergman et al. (2007) in Sweden for 2002. Table 5 summarises the different results for payments using cash and cards where we have chosen to add together the costs for payments using debit and credit cards.

Table 5. Costs for card and cash payments, comparison between four countries, SEK at 2009 value and exchange rate

		2002	2009			2007
		SWEDEN	SWEDEN	DENMARK	HUNGARY	NORWAY
Cash	Social cost (SEK, billions)	7 206	8 604	8 260	7 911	4 248
	Share of GDP (%)	0.27	0.28	0.38	0.8	0.15
	Social unit cost (SEK)	5.16	8.32	10.5	2.79	8.59
Cards, total*	Social cost (SEK, billions)	2 098	8 770	5 166	1 886	6 513
	Share of GDP (%)	0.09	0.28	0.24	0.19	0.23
	Social unit cost (SEK)	3.52	5.55	5.89	10.78	7.21

*Includes card payments at the point-of-sale and remote payments using debit and credit cards.

Sources: Danmarks Nationalbank (2012), Turján et al. (2011), Grevik and Haare (2009), Bergman et al. (2007), Segendorf and Jansson (2012), Statistics Norway (national accounts), Norges Bank and ECB Statistical Warehouse.

The comparison shows that it is completely reasonable for Sweden to have social cost for cash of between SEK 8 and 9 billion, which corresponds to approximately 0.3 per cent of GDP. This conclusion is also in line with other studies.¹³ A social cost of SEK 8 to 10 per payment (unit cost) for cash payments also appears to be normal.

Hungary deviates in this respect in that it has higher costs measured as a percentage of GDP and a lower average cost per payment. The reason for this is that the use of cash is much more prevalent in Hungary than in the other countries. Norway, on the other hand, has lower costs in terms of both absolute costs and as a percentage of GDP. The probable reason for this is that cash is used to a lesser extent in Norway than in the other countries, which reduces these costs. However, the average cost of a cash payment in Norway is in line with the corresponding costs in Sweden and Denmark in 2009. It should also be added that Norway has worked methodically for a long time to improve the efficiency of its payment system.

The average cost in Sweden increased between 2002 and 2009, which was due to a fall in the number of cash payments in this period. This reflects the fact that the use of cards is increasing in Swedish society, but is also due to differences in the methodologies used in the studies. The 2002 study also includes payments between private individuals, while the calculations for 2009 are based only on the number of payments from consumers to companies and authorities.

It also appears that a cost for card payments in relation to GDP of between 0.25 and 0.30 per cent is reasonable. At first glance, the level seems to be somewhat high in Sweden compared to the other countries, but not when examined more closely. This is due to the fact that the corresponding cost in Hungary is low as the use of cards is limited there. In Denmark, on the other hand, a national debit card (Dankort) is widely used, which leads to lower costs. Dankort cannot be used abroad, however, unlike most of the Swedish cards which are linked to global card systems such as VISA or Master Card. Norway also has a national card system (Bank Asept) and, furthermore, has a higher GDP per capita than

¹³ In general it is regarded as normal to be in the interval 0.3-0.4 per cent of GDP, but the level may sometimes be higher. In the Netherlands, the cost for cash was estimated to be approximately 0.48 per cent of GDP in 2002, see Brits and Winder (2005).

Sweden. It is thus not remarkable that the Swedish costs correspond to 0.28 per cent of GDP. The social unit cost for a card payment is, on the other hand, relatively low in Sweden. It is in line with the corresponding cost in Denmark and is slightly lower than in Norway.

The social costs for cards were higher in 2009 than in 2002 but, as in the case of cash, the difference is largely due to differences in methodology in that fewer types of cost were included in the study from 2002 than in the study from 2009.

All in all, there is nothing to indicate that Sweden's payment system is less efficient than the systems in the other countries in the comparison above. When more of the countries that have participated in the coordinated study publish national reports and the ECB presents a comparative study, the possibility to compare Sweden with other countries will increase.

CASH IS USED TOO MUCH

At the start of this article, we discussed the various types of infrastructure needed for cash and card payments and the existing economies of scale. The infrastructure for cash payments is mainly logistical, while that for card payments consists of IT systems. Not unreasonably, the cost of "producing" an additional payment differs between these two systems.

One example is the time cost. Examining our data from 2009, we find that an average card payment takes 25 seconds, irrespective of the amount of the payment. The average cash payment takes 26 seconds, but for smaller amounts, for example SEK 10 or SEK 20, it is probably faster, particularly if the exact amount is handed over. On the other hand, for payments of larger amounts, for example SEK 1 367, where a number of different denominations must be used, either handed over to the payee or as change returned to the payer, it can take longer to pay by cash than by card. Cash also requires more work for checking takings and may entail time spent queuing at ATMs.

The average time for a cash payment thus increases in line with the amount of the payment, while the time for a card payment remains the same regardless of the amount. As time has a value for all of the parties involved in both cases, the economic cost of carrying out a card payment thereby remains the same, in principle, regardless of the transaction amount, while the economic cost of a cash payment increases with the transaction amount. Consequently, there should be a threshold value for the transaction amount at which the cost for a cash payment is the same as for a card payment. For transaction amounts below this threshold, it would be cheaper, from a social perspective, to pay in cash, while, for those above the threshold it would be cheaper to pay by card. Below, we calculate these cut-off points between card and cash payments and contrast these with how consumers actually pay.

As the costs of a debit or credit-card payment do not depend on the amount of the transaction, we have set these at the same level as the average social cost, which is to say SEK 4.50 for payments with a debit card and SEK 11.70 for payments with a credit card. For cash payments, we have divided the social costs (SEK 8.6 billion) into fixed costs and

variable costs, where the variable costs are those that change when the total value of cash payments in the economy changes, for example the time cost. The fixed costs, which are thus not dependent on cash usage include, for example, costs for designing banknotes, certain costs for security and storage etc.¹⁴ We have divided up the fixed costs per the number of cash payments, and the variable costs on the basis of the total value of all cash payments. This gives us a simple equation that expresses the cost of a cash payment as a linear function of the value of the cash payment – see equation (1), where v is the value of the payment.

$$f(v) = \frac{\text{fixed costs}}{\text{number of payments}} + v \times \frac{\text{variable costs}}{\text{total value}} \quad (1)$$

The fixed costs of cash payments in 2009 amounted to SEK 4.25 billion, and the variable costs to SEK 4.35 billion. The number of cash payments and their value is given in Table 1.

The calculated economic cost of card payments and cash payments is illustrated in Figure 1. As we described earlier, the cost of a card payment remains constant, irrespective of the transaction amount, and thus the cost functions of debit and credit-card payments run parallel to the figure's x axis, which expresses the payment's value. In contrast, the cost of cash payments increases with the value of the payment. The slope of the cost function of cash payments is thus given by the second fractional expression on the right-hand side of equation (1). The intercept on the y axis is given by the first fractional expression. The threshold value for cash and debit-card payments is given by the intersection between the cost function for cash payments and the cost function for payments using debit cards. We can see that cash payments are less expensive from society's point of view for payments below SEK 20. Similarly, the threshold value for cash and credit-card payments is SEK 450, that is below this value cash payments are socially less expensive than credit-card payments.¹⁵

14 See Segendorf and Jansson (2012) for a more detailed description.

15 In Denmark in 2009, Jacobsen and Pedersen (2012) found that the economic cut-off point between cash payments and payments made using the domestic debit card (Dankort) was DKK 29, which is equivalent to SEK 41. The cut-off point between cash payments and debit-card payments in Sweden in 2002 was SEK 72, see Bergman et al. (2007). One reason why this cut-off point has shifted downwards over time is that the economies of scale in the card system have been exploited to a greater extent. The use of cards in Sweden increased from 66 card payments per capita in 2002 to 182 in 2009. There are many indications that the use of cash declined in the same period.

Figure 1. The economic cost of different payments as a function of the transaction amount

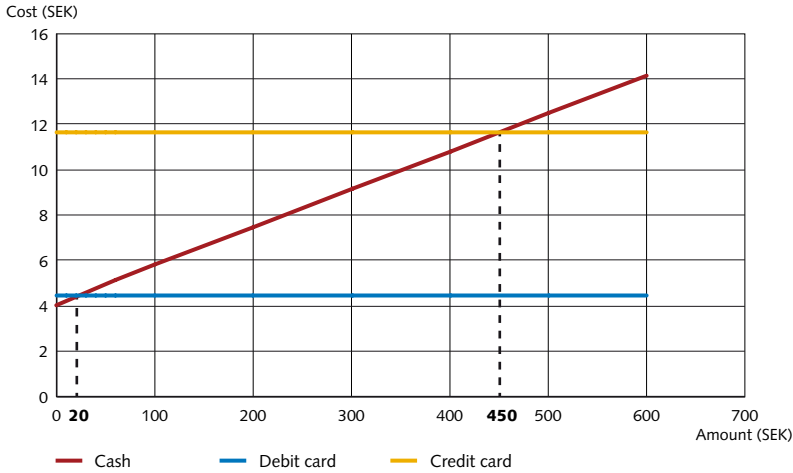


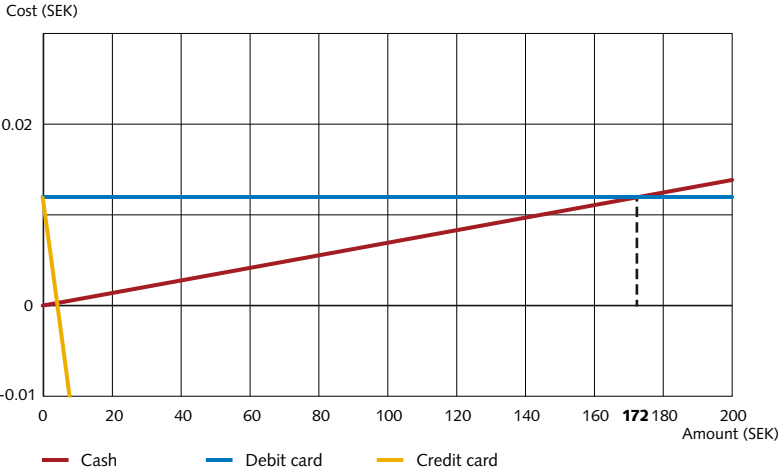
Table 1 shows that the average value of a cash payment is SEK 252. Considering the threshold values in Figure 1 and the fact that credit cards are used to a limited extent by consumers, this indicates that Swedish consumers use cash too much as the average value of cash payments is so much higher than a weighted average of the threshold values.¹⁶ Put another way, we can say that the consumers use debit cards to little. We can also note that the average cost of a credit-card payment is always higher than the cost of a debit-card payment irrespective of the transaction amount, which means that from a social point of view credit-card payments should be avoided altogether.

Can we explain the consumers' choice of payment method by studying what economic incentives they have to choose one method rather than another? The consumers almost never have to pay any transaction charges or charges for withdrawals from ATMs. We therefore assume that neither the use of cash nor the use of cards is associated with any charges to the consumer at the point-of-sale. The cost to the consumer is thus determined by the time it takes to make the payment. In the case of cash payments, this time is worth SEK 20 million per year (see the item Consumed resources in Table 2). Divided by the total value for all cash payments this gives the slope of the cost function for cash payments. In the case of card payments, we divide the corresponding value by the number of card payments as the time taken does not depend on the size of the payment. For credit cards there is also an income in terms of an interest-free loan. A purchase is paid for after an average of 45 days and the consumer pays no interest during this period – in 2009 the average short-term lending rate was 2.35 per cent. This means that the cost function for credit cards has a negative slope; that is it becomes increasingly economically favourable to pay with a credit card the higher the sum to be paid is. Figure 2 shows that the private

¹⁶ Table 1 shows that 15.4 per cent of card payments are made using credit cards. If we weigh together the threshold values for debit cards and credit cards with their respective percentages of the volume of card payments we get SEK 86.

threshold value between debit-card and cash payments is SEK 172, while between credit-card and cash payments it is only SEK 4.

Figure 2. The private cost of different payments as a function of the transaction amount



The social cut-off points in Figure 1 show how the consumers should behave from the point of view of the economy as a whole, while the private cut-off points in Figure 2 show how the consumers should behave given the economic incentives they have. The figures show that there is a considerable gap between the social cut-off points and the private cut-off points. The consumers thus have no economic incentives that promote the efficient use of the payment system. The cut-off points between cash payments and debit-card payments indicate that we should expect cash to be used to a significantly greater extent than is socially desirable.

COULD A MORE EFFICIENT PAYMENT SYSTEM BE ACHIEVED WITH TRANSACTION CHARGES?

In this article, we have demonstrated that the social costs of payments made by card and cash at points-of-sale in Sweden in 2009 amounted to SEK 17.3 billion (see Table 5). This was equivalent to 0.56 per cent of Sweden’s GDP in the same year, which is in line with the situation in other comparable countries. To minimise the social costs, cards should have been used instead of cash for payments exceeding SEK 20. However, the average cash payment amounted to SEK 252. In other words, Swedish consumers are using cash too much in comparison with the level that would minimise costs to society for these types of payment.

From these results, we can conclude that the use of cards and cash in the Swedish payment system is relatively efficient, but that there is potential to save even more of society’s resources by encouraging consumers to use their debit cards more often. How to do this is a more open question, however. To a large degree, it is a matter of increasing consumers’ private economic incentives to act in a more efficient manner in terms of the

economy as a whole. One method of achieving this would be to introduce cost-based transaction charges. The idea is to allow prices to reflect the cost of producing services or goods, as already happens on most markets.

However, on the Swedish market for card and cash payments, there are few or no such price signals at present. Even so, experience from Sweden and Norway indicates that consumers are sensitive to such price signals. In Sweden, payments by cheque were priced at the start of the 1990s, after which cheque payments decreased rapidly.¹⁷ Over-the-counter payments at bank branches have also long been subject to a charge, and few bank customers choose to pay in this way today. Instead, they pay via online banking or by postal credit transfers. A further indication that Swedish customers are sensitive to economic incentives is provided by a study by Bergman et al. (2007), which finds that the consumers' use of cards and cash coincides strongly with their private economic incentives.

Experience is similar in Norway, where a system using cost-based transaction charges to a greater extent was introduced a little over 20 years ago. Norwegian statistics show that direct transaction prices covered about 20 per cent of the variable costs in 1988. By the end of the 1990s, this figure had increased to around 60 per cent. A study by Humphrey et al. (2001) uses Norwegian data from this period and shows that the choice of payment method is largely governed by price. So there is good reason to believe that cost-based transaction charges among banks and businesses could increase the efficiency of the payment system.

However, there are a number of potential problems with the introduction of transaction charges. First, it is difficult to design these charges in a way that makes it rational for the separate banks and businesses to introduce them individually. Being first to introduce transaction charges would probably entail a significant cost in terms of dissatisfaction among customers and impaired competitiveness. Banks and businesses would probably find themselves caught in a 'prisoner's dilemma' in which nobody wants to be first to introduce charges, even though this would be best for all.^{18, 19} One possible solution may be to make hidden charges visible, which is feasible because today's consumers largely contribute towards funding payment services in an indirect way, for instance through low interest rates

17 In 1990, every seventh transaction was paid for by cheque. Cheque payments accounted for just over one-tenth of the value of non-cash payments. One bank introduced a charge of SEK 15 and other banks followed suit. In just a few years, the greater part of cheque usage disappeared. Today, cheque payments account for only about one tenth of a per cent of both the value of the payments as well as the number of transactions. See Nyberg (2008).

18 The prisoner's dilemma is the name of a classic decision-making problem in game theory. The problem can be described by supposing a situation in which two people have committed a crime and are sitting in custody, each in their own cell with no possibility of communication. Each prisoner has the possibility to testify against his companion or to remain silent. A betrayal would shorten the prisoner's own sentence, but if both choose not to testify they will receive a significantly shorter sentence or even be released due to lack of evidence. Collectively, it is more beneficial for the prisoners not to betray each other, while individually it is more rational for each prisoner to testify against his companion. If both parties are uncertain of what the other will do, they will each look to their own interests and take the course of action that is individually rational, which will result in them receiving longer sentences than if they had chosen not to testify against each other.

19 The banks in Norway introduced cost-based transaction charges individually with the support of Norges Bank. The fact that the large banks were first to introduce this type of charge probably made it easier for the smaller banks to follow suit.

on current accounts. Consequently, the banks would not have to increase their revenues but only redistribute them from hidden to visible charges. In such a situation, it is possible that consumer distrust could be overcome if consumers can simultaneously be convinced that they will not be paying more in total, but will be compensated in other ways, such as through better interest rates on current accounts and lower prices in the shops. However, it is not easy to see how this could credibly be achieved. This solution would be further obstructed by the fact that there is a cost associated with levying a transaction charge and that introducing a transaction charge would thus not always be economically motivated in the cases where the service is cheap to produce and the charge consequently very low. However, this problem can be solved by *packaging* the payment services (often called *bundling*), so that a customer gets access to one or more payment services for a fixed price over a certain period.²⁰ The net interest income the banks earn on the current accounts can also be considered to be an indirect charge. However, we are fairly certain that this indirect charge is neither sufficiently transparent nor clearly linked to the use of payment services to give consumers the incentive to use payment services in an economically-optimal way.²¹ This means that indirect funding via net interest income should be replaced by direct funding through charges, either by way of period fees or through transaction charges, depending on the circumstances.

The Riksbank's investigation also reveals that the consumers' choice of payment method is habitual to a certain extent (see Segendorf and Jansson (2012)). The use of cash increases with age and is more common among men, while card usage increases with income, education and the size of both the place of residence and the household, and is more common among women. Transaction charges would thus probably have to be complemented by other measures aimed at different groups of consumers. For example, this could be a matter of discretely influencing behaviour (known as *nudging*), which has proved to be effective in other contexts in attempts to influence attitudes and behaviour by presenting different alternatives in various ways. In the case of cards and cash, for example, the consumers' choice of payment method could be influenced with the aid of appropriately-designed information in shops about the charges shops pay to the banks for various payment services. Another possibility would be to introduce a charge on one payment method as a discount on another payment method.²² Experience from the United States, the United Kingdom and Denmark indicates that nudging can have significant

20 This is a common pricing method for the private customers of Swedish banks. Online banking services and debit cards are often combined.

21 The banks usually levy most of the charges on companies, for example in the form of transaction charges. In general, card payments generate a surplus for the banks, while cash gives rise to a deficit. See Guibourg and Segendorf (2007) for a more detailed analysis of the Swedish banks' pricing.

22 One possible example of Swedish nudging can be found in Chapter 5, Article 1 of the Payment Services Act (2010:751), which forbids Swedish businesses from demanding charges for non-cash payments, for example for credit cards, but permits discounts. Discounts and charges are equivalent in this context, as the same price structure can be attained via both discounts and charges. It is possible that the word 'discount' is used rather than 'charge' due to the negative associations of the word 'charge'.

effects in certain situations such as energy conservation, littering and pension saving.²³ However, this is a relatively new approach to the problem of optimal choice of payment method and thus needs to be studied more closely. In the meantime, nudging should be seen as a complement to cost-based transaction charges and a way of making it possible to introduce these charges.

23 The Economist (2012) provides a lucid introduction to the concept of nudging and presents a number of interesting examples of it.

References

- Bergman, M., G. Guibourg, and B. Segendorf, 2007, The Cost of Paying – Private and Social Costs of Cash and Card Payments, *Sveriges Riksbank Working Paper Series*, 212.
- Brits, H. and C. Winder, 2005, Payments are no free lunch, *Occasional Studies* 3(2), De Nederlandsche Bank.
- Danmarks Nationalbank, 2012, Costs of payments in Denmark.
- Gresvik, O., and H. Haare, 2009, Costs in the Norwegian payment system, *Staff Memo* 4, Norges Bank.
- Guibourg, G., and B. Segendorf, 2007, A note on the price- and cost structure of retail payment services in the Swedish banking sector, *Journal of Banking and Finance*, 31, pp. 2817-2827.
- Humphrey, D., M. Kim, and B. Vale, 2001, Realizing the Gains from Electronic Payments: Costs, Pricing, and Payment Choice, *Journal of Money, Credit, and Banking*, 33(2), pp. 216-234.
- Jacobsen, G. K., and A. M. Pedersen, 2012, Costs of card and cash payments in Denmark, *Monetary Review*, 2(1), pp. 109-122, Danmarks Nationalbank.
- Nyberg, L., 2008, Cash and payments – what lies ahead?, speech, www.riksbank.se.
- Segendorf, B. and T. Jansson, 2012, The cost of consumer payments in Sweden, *Sveriges Riksbank Working Paper Series*, 262.
- Swedish Code of Statutes*, Payment Services Act (2010:751).
- The Economist*, 2012, Nudge nudge, think think, 24 March.
- Turján, A., E. Divéki, E. Keszy-Harmath, G. Kóczán, and K. Takács, 2011, Nothing is free: A survey of the social cost of the main payment instruments in Hungary, *MNB Occasional Papers*, 93, Magyar Nemzeti Bank.



Sveriges Riksbank
SE-103 37 Stockholm

www.riksbank.se
Tel +46 8 787 00 00
Fax +46 8 21 05 31