Sveriges Riksbank Economic Review

2021:2

S V E R I G E S R I K S B A N K

SVERIGES RIKSBANK ECONOMIC REVIEW
is issued by Sveriges Riksbank.

Editors: MARIANNE NESSÉN AND ULF SÖDERSTRÖM

Sveriges Riksbank, SE-103 37 Stockholm, Sweden
Telephone +46 8 787 00 00

The Review is published on the Riksbank's website
www.riksbank.se

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

ISSN 2001-029X

Dear readers,

This year's second edition of Sveriges Riksbank Economic Review comprises a smorgasbord of questions relevant to a central bank. The first article takes a closer look at households' expectations of where the economy is headed, while the second article analyses banks that are regarded as systemically important. These are followed by an article on the development of the Swedish krona over a longer period of time, and the final article concerns how pandemics have affected Sweden over the past 200 years.

- **Can households predict where the macroeconomy is headed?**

  Kamil Kladívko and *Pär Österholm* of Örebro University analyse Swedish households' expectations of inflation and unemployment. These expectations are interesting as households' decisions can be affected by their expectations of future events. Moreover, households' inflation expectations reflect their confidence in the ability of monetary policy to attain the inflation target. In this article, Kladívko and Österholm analyse how well households have managed to predict changes in inflation and unemployment in the coming period.

- **Systemically important banks: is there a TBTF premium?**

  Marianna Blix Grimaldi, Mats Christoffersson, Yuuki Ikeda and *Jonas Niemeyer* describe questions concerning systemically important banks, that is to say, banks that, if they were to fail, would cause large problems for the financial system and the economy as a whole. As these banks are often protected by the state if they suffer financial problems, their financing costs are lower than those of other banks. The authors calculate the size of this premium for a large sample of international banks, and discuss how the premium varies over time and between regions.

- **The development of the Swedish real exchange rate over a longer perspective**

  Carl-Johan Belfrage takes a closer look at how the Swedish krona's real exchange rate has developed in recent decades, and also in a 100-year perspective. He compares and analyses different measures of the real exchange rate and discusses various possible explanations for the long-term development of the krona. In addition, he presents an empirical estimate of the trend for the real exchange rate.

- **How lasting are the economic consequences of pandemics? 220 years of Swedish experiences**

  Stefan Laséen analyses how the pandemics that have affected Sweden since the early 19th century have affected both demographics – for instance, the number of births and the total population – and the economy – for instance, GDP and inflation. In his analysis he uses data from the Riksbank's historical monetary statistics. He also discusses what possible conclusions can be drawn with regard to the effects of the current pandemic.

Read and enjoy!

Marianne Nessén and Ulf Söderström

Contents

Can households predict where the macroeconomy is headed?

Kamil Kladívko and Pär Österholm*

Kamil Kladívko is Associate Senior Lecturer in Economics at the School of Business, Örebro University. Pär Österholm is Professor of Economics at the School of Business, Örebro University, and is an affiliated researcher at the National Institute of Economic Research.

Survey data of households' expectations of macroeconomic variables might provide useful information to those who analyse or forecast the economy. In this article, we evaluate whether households can predict in which direction inflation and the unemployment rate will move over the coming year. The analysis is conducted using monthly Swedish data from the National Institute of Economic Research's *Economic Tendency Survey* over the period from January 1996 until August 2019. Our results indicate that households can forecast in what direction the unemployment rate is headed, but they fail to predict the direction of future inflation.

1 Introduction

For an economic policymaker, such as a central bank, the expectations of the economy's agents might be of interest from several perspectives. For example, long-run inflation expectations could be informative regarding the credibility of the inflation target. Other types of expectations, such as short- or medium-term expectations of GDP growth, inflation, wage growth or the unemployment rate, can provide useful input for policymakers since the actions of the agents – and thereby macroeconomic outcomes – tend to depend on the agents' expectations. Yet another aspect is that expectations might be good forecasts that the policymaker could take into account in order to improve its own forecasts. Various properties of the expectations might also reveal how expectations are formed and evidence of deviations from rationality, for example, could affect how a policymaker both conducts policy and communicates. Accordingly, it is not surprising that quite some effort is put into measuring agents' expectations. For example, in Sweden, two surveys are conducted on a monthly basis which (among other things) address the issue of inflation expectations; households are interviewed in the National Institute of Economic Research's *Economic Tendency Survey* ('*Konjunkturbarometern*') and money-market participants are interviewed in a survey commissioned by Sveriges Riksbank, commonly referred to as the *Prospera Survey*.[1]

In this article, we analyse households' survey expectations. The reason for this focus is the simple fact that households constitute an important part of the economy; for example, household consumption's share of GDP is approximately 45 per cent in Sweden. More specifically, we evaluate households' directional forecasts of inflation and the unemployment rate in Sweden. This is done using monthly data from the *Economic Tendency Survey*.

In conducting this analysis, we follow a line of research that can be seen as being concerned with the forecasting properties of the expectations themselves; see, for example,

* We are grateful to Marianne Nessén, André Reslow and Ulf Söderström for valuable comments.

1 Businesses are also interviewed regarding their inflation expectations in the *Economic Tendency Survey*. However, this is done at a quarterly frequency. In a similar manner, employee organisations, employer organisations, manufacturing companies and trade companies are interviewed in the *Prospera Survey* at a quarterly frequency. The *Economic Tendency Survey* is conducted by Origo Group. The *Prospera Survey* is conducted by Kantar Sifo.

Batchelor and Dua (1989), Thomas (1999), Trehan (2015), and Berge (2018). That is, we are interested in whether the households can predict where the economy is headed. Our focus is accordingly different from that of the fairly voluminous literature which utilises household survey data in order to generate model-based macroeconomic forecasts.[2] Another novel aspect of our article is that we study directional forecasts. While it is not uncommon for such forecasts to be analysed in the macroeconomic literature – see, for example, Ash et al. (1998), Greer (2003), Thomas and Grant (2008), Baghestani et al. (2015), and Driver and Meade (2019) – it is nevertheless the case that numerical forecasts tend to be the focus in the overwhelming majority of empirical studies.[3] No rigorous analysis has previously been conducted on the directional forecast data that we study in this article. Our study should hence bring new information to policymakers and others who analyse and forecast the Swedish economy.

2 Data

We use monthly data from the *Economic Tendency Survey* ranging from January 1996 to August 2019.[4] In this survey, 1,500 randomly sampled Swedish households are interviewed each month.[5] The respondents are asked a number of questions related to their own economic situation and the Swedish economy at an aggregate level. This is Sweden's most important household survey and it is part of the European Commission's *Joint Harmonised EU Programme of Business and Consumer Surveys*.

As a key survey in Sweden, data from the *Economic Tendency Survey* have of course been analysed previously. For example, based on micro-level data, Jonung (1981), Jonung and Laidler (1988), Batchelor and Jonung (1989), Palmqvist and Strömberg (2004), and Hjalmarsson and Österholm (2019, 2020, 2021) have investigated various aspects of perceived inflation, inflation expectations, mortgage-rate expectations and housing-price expectations. Aggregate time series from the survey – such as confidence indicators or the mean expectation of a variable – are also commonly used for macroeconomic forecasting and analysis; see, for example, Hansson et al. (2005), Assarsson and Österholm (2015), Hjalmarsson and Österholm (2017), and Jönsson (2020).

In this article, we analyse data on the two questions in the survey that concern directional forecasts – that is, questions 6 and 7 in the survey. These questions pertain to inflation and the unemployment rate. Their formulations, and the answers available to the respondents, are given below:[6]

> Question 6: Compared to the situation today, do you think that in the next 12 months prices in general will...?
>
> i. Increase faster
> ii. Increase at the same rate
> iii. Increase at a slower rate
> iv. Stay about the same
> v. Fall slightly
> vi. Don't know

2 See, for example, Carroll et al. (1994), Easaw and Herevi (2004), Dreger and Kholodilin (2013), Assarsson and Österholm (2015), and Campelo et al. (2020) for just a few contributions.

3 Additional studies addressing directional forecasts of inflation or the unemployment rate include Sinclair et al. (2010), Ahn and Tsuchiya (2016), Chen et al. (2016), Ahn (2018), Pierdzioch et al. (2018), and Sosvilla-Rivero and Ramos-Herrera (2018).

4 The survey started out as a quarterly survey in 1973. Since 1993 it has been conducted on a monthly basis. The starting date for the time series studied here is January 1996.

5 The number of respondents in the survey has varied over time. During the sample that we are employing, it has ranged between 1,500 and 2,100. The present number of respondents is 1,500 per month.

6 It should be noted that question 6 has the phrasing stated here if the respondent's 'perceived inflation now' (which is question 5 in the survey) is positive. If the respondent's 'perceived inflation now' is non-positive, the phrasing of the question is adjusted somewhat in order to make it consistent with non-positive inflation today.

Question 7: How do you think the level of unemployment in the country will change over the next 12 months? Will it...?

i. Increase sharply
ii. Increase slightly
iii. Remain the same
iv. Fall slightly
v. Fall sharply
vi. Don't know

We evaluate the survey expectations against the outcomes for CPI inflation and the unemployment rate (seasonally adjusted, age group 16 to 64 years); the last available observation for the outcomes is from August 2020.

In order to econometrically analyse the forecasting performance of the survey data, we generate a directional forecast. This is achieved by first taking the balance, b_t, of the share of respondents (as a percentage) that at time t predicted an increase, minus the share that predicted a decrease. This balance is similar to diffusion indices that are commonly generated from survey data; see, for example, OECD (2000) and Pinto et al. (2020). We then turn the balance into a directional forecast, x_t, according to the rule $x_t = 1$ (indicating an increase) if $b_t > 0$ and $x_t = 0$ (indicating a decrease) if $b_t \leq 0$.[7]

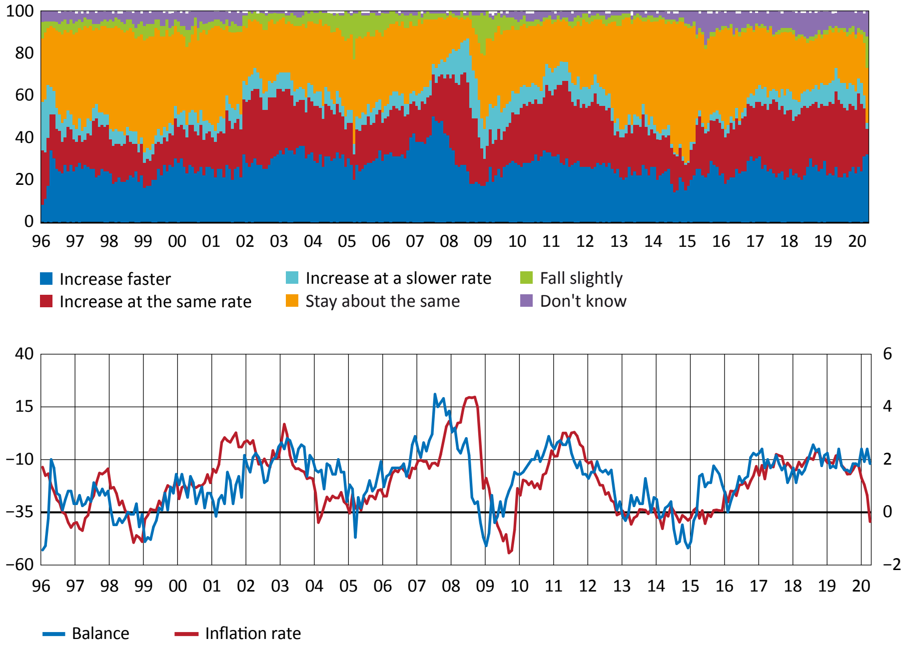
To construct the balance for the inflation question, we take the share of respondents choosing the first alternative among the possible answers, minus the total share choosing the third, fourth and fifth alternatives. Two things deserve to be pointed out concerning this issue. The first is that the question is phrased in terms of prices rather than inflation. This might add a layer of complication if the respondent is used to thinking in terms of inflation. This is not unlikely to be the case in Sweden, since formal inflation targeting was introduced in 1993 and communication typically concerns inflation (rather than the price level).[8] Second, only one of the possible answers implies that inflation will increase, whereas three alternatives imply that inflation will decrease. This feature has the possibility of skewing the respondents' answers due to the so-called *end aversion bias*, which means that respondents tend to avoid the endpoints of a response scale and prefer alternatives closer to the midpoint.[9] For the unemployment rate, the balance is generated as the share of respondents choosing the first two alternatives minus the share choosing the fourth and fifth.

As an illustration of how the balance and forecasts are constructed, consider the unemployment-rate question in January 1996. 6 per cent of respondents answered that the unemployment rate would 'increase sharply', 33 per cent answered 'increase slightly', 42 per cent answered 'remain the same', 16 per cent answered 'fall slightly' and 0 per cent answered 'fall sharply'; finally, 3 per cent answered 'don't know'. The balance is given as $b_{January\ 1996} = 6 + 33 - 16 - 0 = 23$ and the directional forecast accordingly becomes $x_{January\ 1996} = 1$, indicating that respondents predicted an increase in the unemployment rate.

The shares of responses (six for each variable) over time are shown in the top panels of Figures 1 and 2; the bottom panels of Figures 1 and 2 show the resulting balances along with the actual rates of inflation and unemployment.

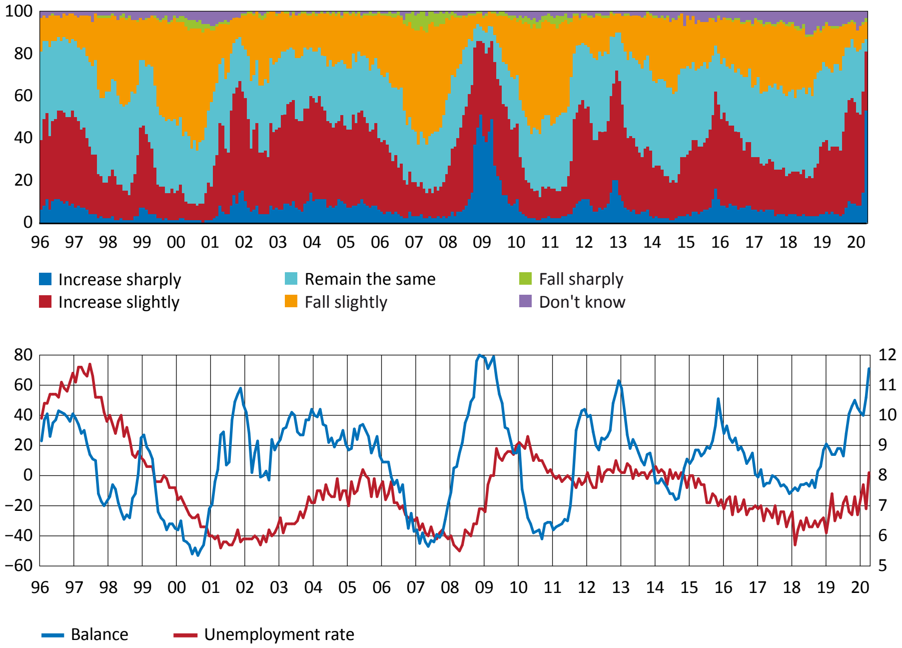
7 We have removed the possibility of having 'unchanged' as a category by merging $b_t = 0$ and $b_t < 0$. This is reasonable though as $b_t = 0$ in only three cases for inflation and two cases for the unemployment rate.
8 Formally, the Riksbank announced in January 1993 that the target for monetary policy would be 2 per cent inflation, starting in 1995.
9 This bias is related to the more general behavioural phenomenon *extremeness aversion*; see, for example, Neumann et al. (2016) for a discussion.

Figure 1. Shares of different answers concerning inflation, and related balance and inflation rate
Shares (top panel) in per cent. Balance (bottom panel, left axis) in percentage points. Inflation (bottom panel, right axis) in per cent



Sources: National Institute of Economic Research, Macrobond and authors' calculations

Figure 2. Shares of different answers concerning the unemployment rate, and related balance and unemployment rate
Shares (top panel) in per cent. Balance (bottom panel, left axis) in percentage points. Unemployment rate (bottom panel, right axis) in per cent



Sources: National Institute of Economic Research, Macrobond and authors' calculations

Looking at Figure 1, it can be seen that the share of respondents that say that prices will 'increase faster' (which indicates the opinion that inflation will increase) has been 26 per cent on average. A substantially higher share can be found in 2007, when it was in the interval 38 to 50 per cent. This was a time period when inflation in Sweden was rising and there was a discussion about increased inflationary pressure; see, for example, Sveriges Riksbank (2007). However, the share predicting an increase in inflation came down substantially in 2008; interestingly, this process started well before the financial crisis hit its peak in the autumn. In addition, the share of respondents answering that prices will 'increase at the same rate' – that is, that inflation will stay the same – has historically had a similar average, namely 23 per cent. It is noteworthy though that, on average, only seven per cent of respondents have said that prices will 'increase at a slower rate', whereas 35 per cent have chosen the alternative that prices will 'stay about the same'. If inflation is above zero – which it typically has been – the latter statement is also a statement about falling inflation but a more specific one. Finally, the share that says that prices will 'fall slightly' has, on average, been small – approximately four per cent. Concerning the balance regarding the inflation question, a striking feature is the fact that the series almost never takes on positive values. In fact, the balance is positive only in eleven cases, implying that it is very rare that a majority of the households forecast increasing inflation.

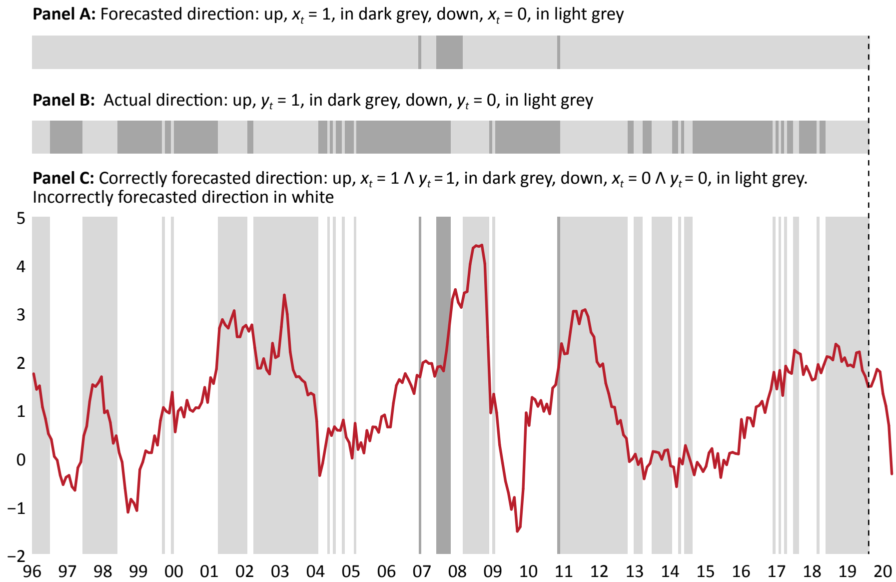
Turning to Figure 2, this shows a fair bit of variation over time in the shares for the unemployment rate. For example, for a few months in the year 2000, less than 10 per cent of the respondents said that the unemployment rate would increase; in December 2008, this figure peaked at 86 per cent. It can be noted that during the financial crisis and its more immediate aftermath many respondents also said that the unemployment rate would 'increase sharply', making this period stand out. In general, most of the variation is due to changes in the three central alternatives ('increase slightly', 'remain the same' and 'fall slightly'). Very few respondents – approximately one per cent on average over time – suggest that the unemployment rate will 'fall sharply'. It is reasonable that households seem reluctant to predict a sharply falling unemployment rate. Stylised facts regarding business cycles do not suggest that unemployment rates tend to decrease rapidly. The balance for the unemployment rate question is more centred around zero and appears to have a clearer cyclical pattern than the balance for the inflation question.

3 Empirical analysis

We now turn to an empirical analysis of our data and first give a graphical illustration. Figures 3 and 4 show the actual rates of inflation and unemployment, along with the directional forecasts and the actual directional changes. In each figure, Panel A displays the directional forecast, x_t, that was calculated from the balance, b_t. Forecasts of an increase ($x_t = 1$) are indicated with dark grey and forecasts of a decrease ($x_t = 0$) are indicated with light grey. Panel B records the actual directional change of the forecasted variable. It should be noted that this has been aligned with the forecast origin date – that is, at a given date, it indicates the directional change between that date and twelve months later. The actual directional change, y_t, is coded analogously to x_t, namely $y_t = 1$ if the variable increases over the twelve-month horizon and $y_t = 0$ otherwise. Note that the value of y_t becomes known at time $t + 12$. An increase ($y_t = 1$) is indicated with dark grey and a decrease ($y_t = 0$) is indicated with light grey. Finally, panel C captures the match of the directional forecast with the actual directional change (and also displays the actual rates of inflation and unemployment). Correctly forecasted directions are reported using dark and light grey shaded areas, while incorrectly forecasted directions are reported using white areas. The dark grey areas correspond to the case when an increase in the variable was correctly forecasted and the light grey areas indicate when a decrease in the variable was correctly forecasted. Observe that the correctly forecasted directions in panel C follow from the intersection of directions in panels A and B.

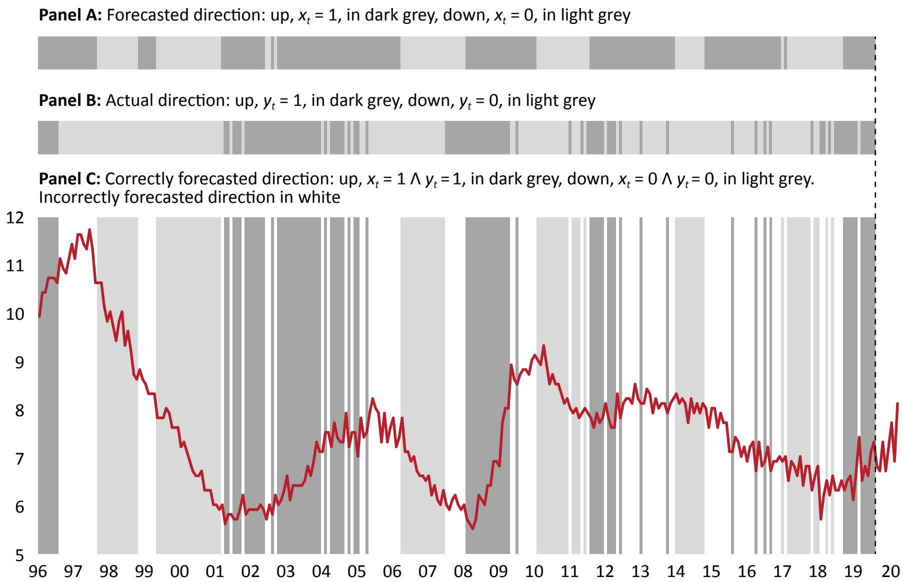
The share of correct forecasts is 46 per cent for inflation and 62 per cent for the unemployment rate (see Table 1). It is noteworthy how an increase in inflation is almost never correctly forecasted; in almost all cases where the outcome was an increase in inflation, a decrease had been predicted. This is, of course, related to the fact pointed out above, namely that households almost always predict a decrease in inflation.

Figure 3. Directional forecast of inflation
Inflation (panel C) in per cent



Note. Panel A: Forecasted directional change aligned with the forecast origin date. The vertical dashed line indicates the last forecast that can be evaluated. Panel B: Actual directional change of inflation aligned with the forecast origin date. Panel C: The red line gives CPI inflation (year-on-year). Correctly forecasted direction in dark and light grey; incorrectly forecasted direction in white.
Sources: Macrobond and authors' calculations

Figure 4. Directional forecast of the unemployment rate
Unemployment rate (panel C) in per cent



Note. Panel A: Forecasted directional change aligned with the forecast origin date. The vertical dashed line indicates the last forecast that can be evaluated. Panel B: Actual directional change of the unemployment rate aligned with the forecast origin date. Panel C: The red line gives the unemployment rate (in per cent). Correctly forecasted direction in dark and light grey; incorrectly forecasted direction in white.
Sources: Macrobond and authors' calculations

We assess the accuracy of the directional forecasts by employing the Pesaran and Timmermann (1992) test with a Newey-West correction for the presence of serial correlation. This test is effectively a test of independence between the directional forecast x_t and the actual directional change y_t; see the Appendix for a detailed description of the test.

The 2×2 tables to test the independence between the forecast and the realized directional change for our two variables, as well as the test results, are reported in Table 1.

The test statistic, t_{PT}^{NW}, clearly confirms what is suggested by the figures. The null hypothesis of independence of forecasted and actual direction cannot be rejected for inflation but is forcefully rejected for the unemployment rate. We accordingly conclude that Swedish households are unable to forecast where inflation is headed, whereas they have highly significant ability in forecasting the direction of the unemployment rate.

Table 1. Results from the directional accuracy test.

	Inflation		Unemployment rate	
	Actual up $y_t = 1$	Actual down $y_t = 0$	Actual up $y_t = 1$	Actual down $y_t = 0$
Forecast up $x_t = 1$	7	4	86	89
Forecast down $x_t = 0$	149	124	20	89
Proportion of correct forecasts, $\hat{P}$	0.46		0.62	
Estimated expected proportion of correct forecasts, $\hat{P}_{H0}$	0.45		0.47	
Test statistic, t_{PT}^{NW}	0.43		3.47	
p-value	0.664		<0.001	

Note. The top part of the table gives the 2×2 contingency tables of 284 forecasts to test the independence between the households' forecast and actual direction. $\hat{P}$ is the proportion of correct forecasts (see equation (4) in the Appendix), whereas $\hat{P}_{H0}$ is an estimate of the expected proportion of correct forecasts under the null hypothesis of independence (see equation (6)). t_{PT}^{NW} is the Pesaran-Timmermann (1992) test statistic with Newey-West correction (see equations (10)–(12)).

The fact that households have some success in predicting the direction of the unemployment rate is not completely surprising. While macroeconomic forecasting by no means is a trivial exercise, the unemployment rate appears to have fairly distinctive cyclical swings (as can be seen from Figures 2 and 4). It likely also helps that the unemployment rate is a concept to which it should be reasonably easy for households to relate.

The failure when it comes to predicting the direction of inflation is perhaps no surprise either. Given the somewhat mixed evidence in the previous literature, no unambiguous conclusions can be drawn regarding different agents' ability to forecast the direction of inflation. Our results are nevertheless in line with recent studies that point to households not being successful at this task; see, for example, Ahn and Tsuchia (2016) and Ahn (2018).[10] It should be kept in mind, however, that this international evidence is based on households that face economic environments that are quite different to that in Sweden. We believe that contributing factors to the failure are the phrasing of the question and the available answers, which were discussed above. One should also consider that a substantial part of the investigated sample comes from a period where inflation may have been quite difficult to predict. This is related to the fairly widespread claim that the connection between the real economy and inflation in many countries is weaker today than previously or, put differently, that the Phillips curve has become flatter; see, for example, Bean (2006), Gaiotti (2010), Kuttner and Robinson (2010), IMF (2013), and Occhino (2019).[11] It accordingly does not seem unreasonable to conclude that the conditions under which households have been forecasting inflation have, at least in parts, been non-trivial.

10 In contrast, some studies indicate that professional forecasters are somewhat more successful at forecasting the direction of inflation; see, for example, Chen et al. (2016) and Sosvilla-Rivero and Ramos-Herrera (2018).

11 This is not an undisputed claim though; see, for example, Fitzgerald et al. (2013) and Berger et al. (2016). For some additional recent contributions concerning the Phillips curve, see, for example, Coibion and Gorodnichenko (2015), Blanchard (2016), Leduc and Wilson (2017), and Karlsson and Österholm (2020).

4 Concluding remarks

In this article, we have shown that Swedish households have statistically significant ability in forecasting the direction of the unemployment rate but that they fail in forecasting where inflation is headed. Despite the failure regarding the directional forecasts of inflation, it can still be worth monitoring these expectations since flawed expectations can still contain useful information to a policymaker, for example. Of course, it is also of interest to know that the expectations have shortcomings.

The finding that the households fail in forecasting the direction of inflation can, to some extent, probably be explained by the fact that inflation objectively has been difficult to predict during a substantial part of the analysed sample. However, we believe that another relevant aspect is that the phrasing of the question in the survey and the answers available are somewhat problematic. The question is phrased in terms of prices, which might complicate things for a respondent who is used to thinking in terms of inflation. Concerning the answers, the respondents' choices could be affected by the fact that only one of the alternatives implies that inflation will increase. To conclude, it does not seem unlikely that the inflation question might be perceived as complicated by the respondents and it could be the case that some respondents do not have sufficient 'economic literacy' to pass this hurdle. This is something that designers of household surveys ought to keep in mind. Considering that this question is part of the *Joint Harmonised EU Programme of Business and Consumer Surveys* (European Commission, 2016), this is likely to be an issue of relevance beyond the Swedish context.

References

Ahn, Young Bin (2018), 'Directional accuracy of urban consumers' inflation forecasts in China', *Emerging Markets Finance and Trade,* vol. 54, pp. 1414–1424.

Ahn, Young Bin and Yoichi Tsuchiya (2016), 'Directional analysis of consumers' forecasts of inflation in a small open economy: Evidence from South Korea', *Applied Economics*, vol. 48, pp. 854–864.

Ash, J. Colin K., David J. Smyth and Saeed M. Heravi (1998), 'Are OECD forecasts rational and useful? A directional analysis', *International Journal of Forecasting*, vol. 14, pp. 381–391.

Assarsson, Bengt and Pär Österholm (2015), 'Do Swedish consumer confidence indicators do what they are intended to do?', *Applied Economics Quarterly*, vol. 61, pp. 391–404.

Baghestani, Hamid, Mohammad Arzaghi and Ilker Kaya (2015), 'On the accuracy of Blue Chip forecasts of interest rates and country risk premiums', *Applied Economics*, vol. 46, pp. 113–122.

Batchelor, Roy A. and Pami Dua (1989), 'Household versus economist forecasts of inflation: A reassessment', *Journal of Money, Credit and Banking*, vol. 21, pp. 252–257.

Batchelor, Roy A. and Lars Jonung (1989), 'Cross-sectional evidence on the rationality of the mean and variance of inflation expectations', In: Grunert, Klaus G. and Folke Ölander (eds.), *Understanding Economic Behavior*, Kluwer Press, Boston.

Bean, Charles (2006), 'Globalisation and inflation', *Bank of England Quarterly Bulletin*, vol. 46, pp. 468–475.

Berge, Travis J. (2018), 'Understanding survey-based inflation expectations', *International Journal of Forecasting*, vol. 34, pp. 788–801.

Berger, Tino, Gerdie Everaert and Hauke Vierke (2016), 'Testing for time variation in an unobserved components model for the U.S. economy', *Journal of Economic Dynamics and Control*, vol. 69, pp. 179–208.

Blanchard, Olivier (2016), 'The Phillips curve: Back to the '60s?', *American Economic Review*, vol. 106, pp. 31–34.

Blaskowitz, Oliver and Helmut Herwartz (2014), 'Testing the value of directional forecasts in the presence of serial correlation', *International Journal of Forecasting*, vol. 30, pp. 30–42.

Campelo, Aloisio, Viviane Seda Bittencourt and Marco Malgarini (2020), 'Consumers confidence and households consumption in Brazil: Evidence from the FGV survey', *Journal of Business Cycle Research*, vol. 16, pp. 19–34.

Carroll, Christopher D., Jeffrey C. Fuhrer and David W. Wilcox (1994), 'Does consumer sentiment forecast household spending? If so, why?', *American Economic Review*, vol. 84, pp. 1397–1408.

Chen, Qiwei, Mauro Costantini and Bruno Deschamps (2016), 'How accurate are professional forecasts in Asia? Evidence from ten countries', *International Journal of Forecasting*, vol. 32, pp. 154–167.

Coibion, Olivier and Yuriy Gorodnichenko (2015), 'Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation', *American Economic Journal: Macroeconomics*, vol. 7, pp. 197–232

Dreger, Christian and Konstantin A. Kholodilin (2013), 'Forecasting private consumption by consumer surveys', *Journal of Forecasting*, vol. 32, pp. 10–18.

Driver, Ciaran and Nigel Meade (2019), 'Enhancing survey-based investment forecasts', *Journal of Forecasting*, vol. 38, pp. 236–255.

Easaw, Joshy Z. and Saeed M. Heravi (2004), 'Evaluating consumer sentiments as predictors of UK household consumption behaviour. Are they accurate and useful?', *International Journal of Forecasting*, vol. 20, pp. 671–681.

European Commission (2016), *Joint Harmonised EU Programme of Business and Consumer Surveys. User Guide*, European Commission, Brussels.

Fitzgerald, Terry, Brian Holtemeyer and Juan Pablo Nicolini (2013), 'Is there a stable Phillips curve after all?', *Federal Reserve Bank of Minneapolis Economic Policy Paper* 13-6.

Gaiotti, Eugenio (2010), 'Has globalization changed the Phillips curve? Firm-level evidence on the effect of activity on prices', *International Journal of Central Banking*, vol. 6, pp. 51–84.

Greer, Mark (2003), 'Directional accuracy tests of long-term interest rate forecasts', *International Journal of Forecasting*, vol. 19, pp. 291–298.

Hansson, Jesper, Per Jansson and Mårten Löf (2005), 'Business survey data: Do they help in forecasting GDP growth?', *International Journal of Forecasting*, vol. 21, pp. 377–389.

Hjalmarsson, Erik and Pär Österholm (2017), 'Households' mortgage-rate expectations – More realistic than at first glance?', *Sveriges Riksbank Economic Review*, no. 2, pp. 57–64.

Hjalmarsson, Erik and Pär Österholm (2019), 'A micro-data analysis of households' expectations of mortgage rates', *Economics Letters*, vol. 185, Article 108693.

Hjalmarsson, Erik and Pär Österholm (2020), 'Heterogeneity in households' expectations of housing prices – Evidence from micro data', *Journal of Housing Economics*, vol. 50, Article 101731.

Hjalmarsson, Erik and Pär Österholm (2021), 'Anchoring in surveys of household expectations', *Economics Letters*, vol. 198, Article 109687.

IMF (2013), *World Economic Outlook*, April 2013.

Jönsson, Kristian (2020), 'Machine learning and nowcasts of Swedish GDP', *Journal of Business Cycle Research*, vol. 16, pp. 123–134.

Jonung, Lars (1981), 'Perceived and expected rates of inflation in Sweden', *American Economic Review*, vol. 71, pp. 961–968.

Jonung, Lars and David Laidler (1988), 'Are perceptions of inflation rational? Some evidence for Sweden', *American Economic Review*, vol. 78, pp. 1080–1087.

Karlsson, Sune and Pär Österholm (2020), 'A note on the stability of the Swedish Phillips curve', *Empirical Economics*, vol. 59, pp. 2573–2612.

Kuttner, Ken and Tim Robinson (2010), 'Understanding the flattening Phillips curve', *North American Journal of Economics and Finance*, vol. 21, pp. 110–125.

Leduc, Sylvain and Daniel J. Wilson (2017), 'Has the wage Phillips curve gone dormant?', *Federal Reserve Bank of San Francisco Economic Letter* 2017-30.

Neumann, Nico, Ulf Böckenholt and Ashish Sinha (2016), 'A meta-analysis of extremeness aversion', *Journal of Consumer Psychology*, vol. 26, pp. 193–212.

Newey, Whitney K. and Kenneth D. West (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica*, vol. 55, pp. 703–708.

Newey, Whitney K. and Kenneth D. West (1994), 'Automatic lag selection in covariance matrix estimation', *Review of Economic Studies*, vol. 61, pp. 631–653.

Occhino, Filippo (2019), 'The flattening of the Phillips curve: Policy implications depend on the cause', *Federal Reserve Bank of Cleveland Economic Commentary* 2019-11.

OECD (2000), *Main Economic Indicators. Sources and Definitions*, OECD, Paris.

Palmqvist, Stefan and Lena Strömberg (2004), 'Households' inflation opinions – A tale of two surveys', *Sveriges Riksbank Economic Review*, no. 4, pp. 23–42.

Pesaran, M. Hashem and Alan G. Timmermann (1992), 'A simple nonparametric test of predictive performance', *Journal of Business and Economic Statistics*, vol. 10, pp. 461–465.

Pierdzioch, Christian, Monique B. Reid and Rangan Gupta (2018), 'On the directional accuracy of inflation forecasts: Evidence from South African survey data', *Journal of Applied Statistics*, vol. 45, pp. 884–900.

Pinto, Santiago, Pierre-Daniel G. Sarte and Robert Sharp (2020), 'Learning about consumer uncertainty from qualitative surveys: As uncertain as ever', *International Journal of Central Banking*, vol. 16, pp. 47–99.

Sinclair, Tara M., Herman O. Stekler and Lindsay Kitzinger (2010), 'Directional forecasts of GDP and inflation: A joint evaluation with an application to Federal Reserve predictions', *Applied Economics*, vol. 42, pp. 2289–2297.

Sosvilla-Rivero, Simón and María del Carmen Ramos-Herrera (2018), 'Inflation, real economic growth and unemployment expectations: An empirical analysis based on the ECB survey of professional forecasters', *Applied Economics*, vol. 50, pp. 4540–4555.

Sveriges Riksbank (2007), Separate minutes of the Executive Board, Meeting on 29 October 2007.

Thomas, Lloyd B. (1999), 'Survey measures of expected U.S. inflation', *Journal of Economic Perspectives*, vol. 13, pp. 125–144.

Thomas, Lloyd B. and Alan P. Grant (2008), 'The accuracy and rationality of US and Australian household inflation forecasts: A comparative study of the Michigan and Melbourne Institute surveys', *Economic Record*, vol. 84, pp. 237–252.

Trehan, Bharat (2015), 'Survey measures of expected inflation and the inflation process', *Journal of Money, Credit and Banking*, vol. 47, pp. 207–222.

Appendix

In this appendix, we provide the details of the econometric test employed in Section 3 to assess the accuracy of the directional forecasts.

The directional forecast is denoted x_t and the actual directional change y_t. Note that any directional variable is Bernoulli distributed. We define

(1) $P_x = \mathbb{P}(x_t = 1)$ and $P_y = \mathbb{P}(y_t = 1)$,

where $\mathbb{P}$ is the probability function. We further introduce the variable z_t which takes on the value 1 if the forecast is correct, and the value 0 if the forecast is wrong. By using the conjunction operator, $\wedge$, we write $z_t = 1$ if $x_t = 1 \wedge y_t = 1$ or $x_t = 0 \wedge y_t = 0$, and $z_t = 0$ if $x_t = 1 \wedge y_t = 0$ or $x_t = 0 \wedge y_t = 1$ (see panel C in Figures 3 and 4). The probability of $z_t = 1$ is thus given by

(2) $P = \mathbb{P}(x_t = 1, y_t = 1) + \mathbb{P}(x_t = 0, y_t = 0)$.

Under the null hypothesis that x_t and y_t are independent – that is, if x_t has no power to predict y_t – then it follows from the definition of independence that the probability of $z_t = 1$ is given by

(3) $P_{H0} = P_x P_y + (1 - P_x)(1 - P_y)$.

The probability P is efficiently estimated as the proportion of correct directional forecasts in a data set with T observations, and thus the estimate is given by

(4) $\hat{P} = T^{-1} \sum_{t=1}^{T} z_t$.

Under the null hypothesis of no predictive power, $T\hat{P}$ has a binomial distribution with expected value TP_{H0} and variance $TP_{H0}(1 - P_{H0})$. In the case in which P_x and P_y are known, one can use the approximate test for the Bernoulli parameter P. The test statistic is asymptotically standard normal and given by

(5) $$t = \frac{\hat{P} - P_{H0}}{\sqrt{T^{-1} P_{H0}(1 - P_{H0})}}.$$

For example, assuming a symmetric random walk behaviour implies $P_y = 0.5$ since an up-move and a down-move of the forecasted variable are equally likely. It is then natural for any forecast to have $P_x = 0.5$. In this case the test statistic simplifies to $t = \sqrt{T}(2\hat{P} - 1)$. However, in practice, P_x and P_y are not known and need to be estimated from sample data. Their efficient estimates are given by $\hat{P}_x = T^{-1} \sum_{t=1}^{T} x_t$ and $\hat{P}_y = T^{-1} \sum_{t=1}^{T} y_t$, and consequently P_{H0} is replaced by

(6) $\hat{P}_{H0} = \hat{P}_x \hat{P}_y + (1 - \hat{P}_x)(1 - \hat{P}_y)$.

Pesaran and Timmermann (1992) derive that in this case the test of predictive performance of x_t can be based on

(7) $$t_{PT} = \frac{\hat{P} - \hat{P}_{H0}}{\sqrt{\widehat{\text{var}}(\hat{P}) - \widehat{\text{var}}(\hat{P}_{H0})}},$$

where $\widehat{\text{var}}(\hat{P}_{H0}) = T^{-1}(2\hat{P}_y - 1)^2 \hat{P}_x(1 - \hat{P}_x) + T^{-1}(2\hat{P}_x - 1)^2 \hat{P}_y(1 - \hat{P}_y) + 4T^{-2} \hat{P}_y \hat{P}_x(1 - \hat{P}_y)(1 - \hat{P}_x)$ and $\widehat{\text{var}}(\hat{P}) = T^{-1} \hat{P}_{H0}(1 - \hat{P}_{H0})$. The t_{PT} test statistic is asymptotically standard normal.

The suggested approach implicitly assumes that the forecast and actual process are serially independent. However, serial correlation if often present in economic applications. Blaskowitz and Herwartz (2014) suggest a Newey-West correction for the directional accuracy test we consider. First note that independence of Bernoulli variables x_t and y_t is equivalent to zero covariance between x_t and y_t. It then follows that

$$P - P_{H0} = 2\text{cov}(x_t, y_t), \qquad (8)$$

where cov(·,·) denotes the covariance operator.[12]

Consequently $\widehat{\text{cov}}(x_t, y_t) = 0$ if and only if $t_{PT} = 0$. We can thus alternatively test for zero covariance between the directional forecast x_t and the actual directional change y_t. We follow the exposition in Blaskowitz and Herwartz (2014) and decompose

$$x_t = P_x + w_t \text{ and } y_t = P_y + v_t, \qquad (9)$$

where w_t and v_t are binary zero mean random errors which may be serially correlated. It follows that the null hypothesis of $\text{cov}(x_t, y_t) = 0$ is equivalent to $\mathbb{E}[w_t v_t] = 0$, where $\mathbb{E}[\cdot]$ denotes the expectation operator. To bring the model to data we estimate $\hat{w}_t = x_t - \hat{P}_x$, $\hat{v}_t = y_t - \hat{P}_y$, and $\overline{wv} = \widehat{\text{cov}}(x_t, y_t) = T^{-1} \sum_{t=1}^{T} \hat{w}_t \hat{v}_t$. The test of predictive performance is then based on the test statistic

$$t_{PT}^{NW} = \frac{\widehat{\text{cov}}(x_t, y_t)}{\sqrt{T^{-1} \hat{s}_T^{NW}}}, \qquad (10)$$

which is asymptotically standard normal, and where $\hat{s}_T^{NW}$ is the heteroscedasticity and autocorrelation consistent variance estimator (Newey and West, 1987) for $\widehat{\text{cov}}(x_t, y_t)$. In particular,

$$\hat{s}_T^{NW} = \widehat{\text{cov}}(w_t v_t, w_t v_t) + 2\sum_{g=1}^{G} \left(1 - \frac{g}{G+1}\right) \widehat{\text{cov}}(w_t v_t, w_{t+g} v_{t+g}), \qquad (11)$$

$$\widehat{\text{cov}}(w_t v_t, w_{t+g} v_{t+g}) = \sum_{t=1}^{T-g} (\hat{w}_t \hat{v}_t - \overline{wv})(\hat{w}_{t+g} \hat{v}_{t+g} - \overline{wv}), \qquad (12)$$

and the truncation lag G is – as is commonly done in the literature – set equal to the integer part of $4(T/100)^{2/9}$, that is, we set $G = 5$; see Newey and West (1994) for a technical discussion.

12 Note that $\text{cov}(x_t, y_t) = \mathbb{E}[x_t y_t] - \mathbb{E}[x_t]\mathbb{E}[y_t] = \mathbb{P}(x_t = 1, y_t = 1) - \mathbb{P}(x_t = 1)\mathbb{P}(y_t = 1)$. Introduce $a_t = 1 - x_t$, $b_t = 1 - y_t$ and note that $\text{cov}(a_t, b_t) = \text{cov}(x_t, y_t)$. Since $P - P_{H0} = \mathbb{P}(x_t = 1, y_t = 1) + \mathbb{P}(x_t = 0, y_t = 0) - \mathbb{P}(x_t = 1)\mathbb{P}(y_t = 1) - \mathbb{P}(x_t = 0)\mathbb{P}(y_t = 0)$, the result follows.

Systemically important banks: is there a TBTF premium?

Marianna Blix Grimaldi, Mats Christoffersson, Yuuki Ikeda and Jonas Niemeyer*

Jonas Niemeyer works in the Financial Stablility Department of the Riksbank. Marianna Blix Grimaldi works at the Swedish National Debt Office but was visiting the Riksbank's Research Department when working on this article. Mats Christoffersson works at the Swedish National Debt Office and Yuuki Ikeda is at Bank of Japan but was at the Bank for International Settlements when working on this article.

Too big to fail (TBTF) is a catchy phrase used to describe a systemically important bank (SIB) that is so entwined with the economy that its failure would impose significant losses on other firms or seriously impede the functioning of the financial system with consequent risks to the broader economy. As a result, when such large banks have been close to default, governments have traditionally used public funds to ensure payments of the bank's debt, to avoid it defaulting. Such a bail-out of a TBTF bank limits systemic risk, but it has economic costs. In fact, treating a bank as TBTF extends unlimited protection to all of the bank's creditors, not just depositors, giving the bank a funding advantage and incentives to take on more risk than other banks. In other words, the TBTF banks benefit from a TBTF premium. After the global financial crisis and in order to limit the TBTF problem, regulators imposed new stricter rules on the largest banks in an effort to control risk-taking and introduced a new process for resolving failures of large banks in a way that subjects the creditors of such banks to losses. However, as the financial system adapts, the TBTF problem remains a public concern. This article contributes to the literature and previous policy work on TBTF issues in several aspects. We estimate TBTF premiums from the period before the global financial crisis to March 2021 for a sample of 53 SIBs. We present novel evidence on regional variation of TBTF premiums, also within Europe. We show that, while the driving factors of TBTF premiums have not changed, their relative relevance has shifted significantly. Finally, we show the significant variation of TBTF premiums that exists among large banks.

1 Introduction

Some banks may be so large and complex that market participants think that public authorities would find it difficult to let them fail when they face financial problems. Such banks are typically called systemically important banks (SIBs) as they are considered to be too-big-to-fail (TBTF). When the market perceives a bank to be TBTF, it has real consequences. If market participants think that such a bank will be saved (i.e. bailed-out) by public authorities, the risk for its shareholders and bondholders will fall. Under such

* We would like to thank Meredith Beechey Österholm, Tor Jacobson, Ola Melander, Marianne Nessén and Tamás Vasi for their valuable comments. All errors and deficiencies remain the responsibility of the authors. The views in this article are the authors' views and should not be interpreted as the views of the Riksbank or its Executive Board.

presumptions, shareholders will gain when the bank makes a profit but will not lose their investment if the bank defaults. Heads I win, tails you lose. Similarly, bondholders would be almost certain to get their investment and earned interest back. Therefore, if market participants think a bank is a SIB, it will affect the pricing of the bank's shares and bonds. If the perceived risk is lower, expected returns will be lower and thus prices higher. Thus, such banks will benefit from a TBTF premium (see FSB 2021). By using market prices, it is possible to estimate this market-implied TBTF premium. A relatively large literature, using different methodologies, has showed the existence of such a premium for SIBs in different countries.[1]

To estimate the TBTF premium in this article, we use a Merton-type model, in which a firm defaults when its asset value falls below a determined boundary. Such models estimate the probability that a firm defaults, and this probability depends on the firm's capital structure, i.e. its own funds in relation to its debt. While it is not the only approach available for estimating the TBTF premium, it is well suited to the problem at hand, see, for example, IMF (2014).

We contribute to the literature and previous policy work in several ways. First, by using data from SIBs from Europe, North America and Asia (Japan), we add novel evidence on the regional variation of TBTF premiums, especially for some European banks. Second, we investigate the determinants of TBTF premiums and how they vary by region. Third, we investigate the time dynamics of the determinants of TBTF premiums. Fourth, we analyse the differences in the TBTF premiums of global systemically important banks (G-SIBs) versus domestic systemically important banks (D-SIBs) in our sample.

The rest of the article is organized as follows. In section 2 we discuss why some banks are large and why this is a public policy concern as well as what has been done in terms of global standards to reduce the TBTF-problem after the global financial crisis (GFC) in 2008–2009. In section 3, we present the estimates of TBTF premiums. We document changes in TBTF premiums over time and across regions. We also present our analysis on the determinants of TBTF premiums and, assess the heterogeneity among SIBs. Section 4 concludes.

2 The TBTF issue

There is no universal definition of what constitutes a bank. Some countries define a bank as an entity that has a banking licence, which becomes a rather circular definition. Other countries focus their definition on banks' provision of payment services. Other countries focus their definition on banks' ability to take deposits and grant loans. Regardless of how they are defined, in most countries banks play a vital role in the economy by: a) providing payment services, b) taking deposits and granting loans and in doing so transforming liquid deposits into illiquid loans and c) being an intermediary to more advanced financial risk management for households and firms.

Similarly, there is no clear definition of which banks are SIBs and thus which banks could benefit from a TBTF premium. On a yearly basis, the Basel Committee on Banking Supervision (BCBS) establishes a list of about 30 banks that are defined as global systemically important banks (G-SIBs) for the coming year. The 75 banks with the largest leverage ratio exposure measure constitute the G-SIB sample banks.[2] For each of these banks, data are collected, 12 indicators are calculated and basic global market shares are calculated. In addition, supervisors may exercise discretion and adjust the mechanistic score, if warranted. As markets and banks develop, the yearly G-SIB list is not static but may change from one year to another. During the last few years, the BCBS has designated about 30 banks as G-SIBs.

1 A systemically important bank is defines as a bank that is deemed to be of such importance that the bank's failure may trigger a wider financial crisis, see BCBS (2013). Colloquially, such banks are often described as TBTF. FSB (2021) lists a number of references outlining TBTF premiums and how to estimate them.

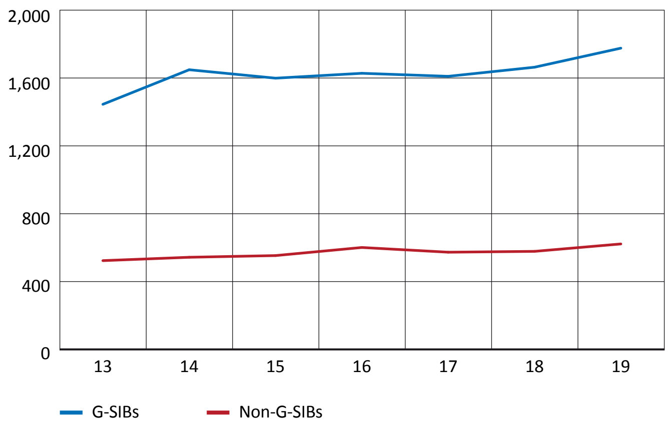
2 In reality, the G-SIB sample consists of 76 banks, see BCBS (2013). Somewhat simplified, a bank's leverage ratio exposures measure consists of its on-balance sheet items and most of its off-balance sheet items weighted with a factor, see BCBS (2017a).

Although some banking groups have entered the list and some exited, the list has been rather stable over the years. The G-SIB list is then endorsed by the Financial Stability Board (FSB) and published in November every year (see FSB 2020). The G-SIBs are the largest – and by the public authorities deemed the most systemically important – banking groups in the world. On a yearly basis, there are about 30 G-SIBs and about 75 banks in the G-SIB sample.

In addition, countries may define a bank as a domestic systemically important bank (D-SIB). This designation may also vary over time. Most – but not all – G-SIB sample banks are also D-SIBs. On the other hand, many D-SIBs are not included the G-SIB sample as they may be large in their home country but have limited global scale, so focusing on SIBs more generally would include more banks than are included in the G-SIB sample. In most of the empirical part of the article from section 3 and onwards, we use data from 53 banks, 26 that have been included in the G-SIB list at any point in time and 27 D-SIBs (see Appendix B).

On a global scale, many banks are large. For 2019, the World Bank reports an overall GDP for the world of 87,800 billion USD or approximately 78,150 billion EUR.[3] At the end of 2019, the aggregate size of the G-SIB sample banks was 81,320 billion EUR (BCBS G-SIB-database).[4] So the aggregate size of these 76 banks is larger than the total world GDP.

Figure 1. Average size of systemically important banks in the world
Billion EUR



Note. Size is measured as the leverage ratio exposure measure at year-end, averaged across the banks. Non-G-SIBs are the banks within the G-SIB sample that are not designated as G-SIBs.
Sources: www.bis.org/bcbs/gsib/, Bank reports and authors' own calculations

The SIBs have also remained large since the GFC. Using data from end 2013 to end 2019, Figure 1 plots the average size of the G-SIBs as well as that of the G-SIB sample banks that are not designated as G-SIBs (Non-G-SIBs). For G-SIBs, the average size (in terms of leverage ratio exposure measure) has been around 1 600 billion EUR. In Figure 2, two things stand out. On average, G-SIBs are substantially larger than Non-G-SIBs. Also, the G-SIBs have become slightly larger since 2013. It is also worth noting that most of the G-SIBs are highly complex. While there is no universal good measure of complexity, the average G-SIB has over a thousand subsidiaries in over 40 jurisdictions (FSB 2021). Most of these banks are also active in many different lines of banking business, further increasing complexity.

A relevant question in this context is why some banks are large and complex. There are several reasons for this.

A first reason is the existence of large non-financial corporates. When these large corporates need banking services they typically turn to large banks. To offer some services, such as corporate actions and fixed income offerings, to these large corporates, the bank will

3 The source for global GDP is the World Bank World Development Indicators https://data.worldbank.org/indicator/NY.GDP.MKTP.CD and this has been transformed using the end 2019 exchange rate.
4 In this context, size is measured as the leverage ratio exposure measure.

need a fairly large balance sheet, as it may have to absorb fairly large risks for a short period of time.

A second reason is that certain banking services may also include economies of scale and scope.[5] Payment services is an area where the marginal cost of an additional payment may be very low while there may be substantial fixed costs. This creates opportunities to profit from economies of scale and scope. A larger bank may also have greater opportunities to diversify risks. However, diversification necessitates good risk data aggregation and as banks grow in size that may become more difficult. Also, in some areas, the economies of scale and scope may be less pronounced. The empirical literature on this subject is not new but the evidence remains mixed. Berger et al. (1999) perform a literature review and find scant evidence of any substantial economies of scale or scope. More recently, Beccalli and Rossi (2017) document some economies of scope in the European banking sector. Also, Boot (2017) argues that technological developments such as fintech may increase the fixed costs and reduce some variable costs, i.e. potentially creating – or alternatively further enlarging – the scope for economies of scope and scale.

A third reason is that many large corporations, including banks, have a natural tendency to grow. Executives often prefer to manage larger firms as that increases their power and remuneration. This empire-building feature may be a driving force for the ever-increasing size of banks (Jensen and Meckling 1976 and Jensen 1986). Many banks have a disperse ownership which may accentuate this agency problem.

A fourth reason is that banks tend to become larger as a result of government actions. Given the important role that banks play in the economy, it is also very difficult to close down a bank. When a non-financial corporation faces economic challenges, it will shrink, close down or even go bankrupt. Few banks take that route. Instead, banks facing economic problems often elicit public interventions. A common public solution to a bank facing problems is to merge it with another bank. Indeed a common result of financial distress is that banks become even larger and fewer. Nordea is a case in point. It was created by merging a number of smaller banks following the various financial crises in the 1990s in the Nordic countries. Presently, consolidating some parts of the European bank sector is also being discussed as one of the possible solutions to the low profitability of some European banks (see, for example, ECB 2020).

A fifth reason is that some banks may be perceived to have an implicit government guarantee. Such guarantees would reduce the bank's funding costs and lower the risk of both shareholders and bondholders. This lower risk may reinforce empire-building features. Also, the controlling effect of shareholders on bank management may be undermined.

In the GFC 2008–2009, it became clear that some large and complex banks posed a substantial risk to the global economy. The failure of Lehman Brothers in 2008 had repercussions all over the world. It became clear that some banks needed stricter rules than other banks, simply because they posed larger risk to society. These banks were seen as TBTF and their existence became a public concern.

2.1 TBTF – a public policy concern

The existence of large and complex banks that are SIBs or TBTF banks is a public policy concern. Their potential failure or default can create large negative spill-over effects on the financial system and the real economy (see BCBS 2011 and FSB 2021). In other words, there are serious negative externalities of such a failure. Given the externalities, public authorities may be unwilling to let such a bank fail. Instead, in a crisis, public authorities may do what they can to save the bank. The perception that a SIB would be saved from default by public authorities diminishes the risk to shareholders and bondholders even in normal times.

5 Economies of scale arise when there are cost advantages of producing a single good, while economies of scope arise when there are cost advantages of producing a variety of similar but different goods.

This reduces the funding costs of these banks as market participants will demand a lower risk premium for such funding than for other banks that are not perceived as systemically important. Thus, one way to assess if market participants perceive a bank to be systemically important is to see if it has lower funding costs, i.e. if it benefits from a TBTF premium due to implicit government support.[6]

The GFC also showed that the negative externalities had wide cross-border implications. The failure of Lehmann Brothers had serious consequences for the financial sector in many countries. Many financial markets experienced severe stress and GDP fell in most parts of the world. Therefore, the need for coordinated action to limit these risks was abundantly clear. From a small-country perspective, the need to ensure sufficiently strict global minimum rules is even more important. The interdependency is larger for small countries than larger countries, but the main message from the crisis is that when there are financial problems in some of the worlds' largest and most complex banks, small and large countries are all likely to feel the economic consequences of those problems.

2.2 Responses to the TBTF issue after the GFC

In the aftermath of the GFC, G-20 leaders met in Pittsburgh in September 2009 to discuss financial markets and the world economy. There was full agreement to launch a vast series of global reforms to strengthen the financial regulatory framework with the aim of increasing the resilience of the global financial system, and to do so in a coordinated manner. Among other things, they agreed (G-20 2009):

1. to 'develop resolution tools and frameworks for the effective resolution of financial groups';
2. 'that prudential standards for systemically important institutions should be commensurate with the costs of their failure'; and
3. that there was a need to develop 'more intensive supervision and specific additional capital, liquidity, and other prudential requirements'.

The first bullet above indicates a desire to develop a more efficient way to deal with banks in distress than through a public bail-out. Notably, the second and third bullets highlight that some banks pose a greater risk to financial stability than others and that these banks need stricter prudential standards. Up until then, the Basel Committee, being the global standard-setter for banks, had developed common regulatory standards for all globally active banks, but largely without making any distinction between smaller and larger banks. In 2009 – for the first time – there was a political commitment for more stringent requirements for a subset of these banks, the SIBs.

With the agreement to impose stricter requirements on a subset of banks, the first important question was how to define these SIBs. In 2011, the Basel Committee developed a method to single out banks that it considers the most global systemically important – the G-SIBs.[7] G-SIBs have a surcharge on their capital requirement in the form of a separate G-SIB buffer. The higher the ranking among the G-SIB sample banks, the higher the surcharge. All relevant jurisdictions have implemented the G-SIB framework, and ensured that it has become a legal requirement for these banks (see BCBS 2016).

6 There may of course be other reasons why a bank benefits from lower funding costs. In our model, see section 3, we try to take these into account.

7 The framework was first agreed in November 2011, see BCBS (2011), and later updated in 2013, see BCBS (2013), and revised in 2018, see BCBS (2018). As part of the method, the Basel Committee collects data from all internationally active banks with a leverage ratio exposure measure larger than 200 billion EUR. The data consist of a number of balance sheet items, transaction items and other items – in total about 65 items – categorized into 12 indicators. The method then calculates the market share of all of these indicators and weighs them together to create a global weighted market share.

The Basel Committee also recognized that other banks may be systemically important on a domestic or regional basis, even if the failure of such a bank may not have global repercussions. It therefore developed a framework for dealing with D-SIBs (BCBS 2012). It is less prescriptive than the G-SIB framework but provides a number of principles to guide jurisdictions in dealing with D-SIBs. Jurisdictions should make an assessment whether any bank is a D-SIB but there is no requirement for defining any bank as such, nor to apply any surcharge, even if the framework recognizes such a surcharge as a useful tool.

Apart from the surcharge, SIBs are also subject to more intensive supervision. The exact way to implement and organize this is largely up to each jurisdiction to define.

Following the GFC, FSB also developed new standards for the recovery and resolution of G-SIBs (see FSB 2014). These standards have four basic components. First, banks should develop recovery plans for what they should do in a stressed scenario. Second, jurisdictions should develop a legal structure to facilitate resolution of a bank in a crisis. Third, resolution authorities should develop resolution plans for banks that may fail. Fourth, banks should have a certain level of Total Loss Absorbing Capacity (TLAC) (see FSB 2015).[8] TLAC-instruments include capital but also debt instruments that can be bailed-in, i.e. written down or converted to equity, when the bank is in resolution. The main purpose of the TLAC requirements is to ensure that a failing bank's shareholders and creditors can absorb sufficient losses in resolution, so that the public sector would not have to resort to using public funds in a bail-out. By ensuring that SIBs have sufficient private capital at risk, the purpose is to reduce the expectation that public funds would need to be used in a crisis. This would also force the market to better assess the risks as bank shareholders and holders of the bail-in-able debt would potentially suffer losses if the bank is put into resolution.

The first G-SIB buffer surcharge is only implemented for capital requirements in terms of risk-weighted assets. In 2017, the Basel Committee agreed that all G-SIBs should be subject to an additional surcharge buffer on their leverage ratio requirement (see BCBS 2017b). The new leverage ratio surcharge buffer was due for implementation on January 2022 but due the Covid-19 pandemic, implementation may be postponed by one year (see BCBS 2020).

Many jurisdictions apply the TBTF-standards to a larger set of banks, than only G-SIBs. D-SIBs are common in many countries and they typically also face higher capital requirements than other banks.[9] Also, requirements on bail-in-able debt for D-SIBs are becoming frequent in many countries. Some countries even apply such requirements on bail-in-able debt to other banks that are not formally SIBs[10]. Recovery and resolution planning as well as legal changes in the implementation of resolution frameworks have been approved in many countries. In that respect, the global standards have materially changed the system in many countries around the world.

3 Development of TBTF premiums

As discussed before, the presumption of government support gives rise to a so-called TBTF premium, which is embedded in market equity prices and credit spreads. It is worth noting that all market prices-based models measure market participants' *perceptions* of the development of TBTF premiums and are therefore imperfect measures of the premium in itself, which is unobservable. In the following, for the sake of brevity, we refer to those imperfect measures as TBTF premiums.

In this article, we focus on estimates based on a Merton-type structural credit-pricing model. It is the same model as in FSB (2021) and closely related to the model in Schweikhard

8 The European version is called Minimum Requirement for own funds and Eligible Liabilities (MREL).

9 In Sweden, Finansinspektionen has determined that three banks are D-SIBs, Handelsbanken, SEB and Swedbank and they all face a D-SIB surcharge.

10 In Sweden, nine banks are subject to requirements on bail-in-able debt, see www.riksgalden.se.

and Tsesmelidakis (2012) and Tsesmelidakis and Merton (2012). Appendix A describes the model.[11]

A commonly convenient way to define the premium is in terms of the difference between equity market-implied CDS, or fair-value CDS (*FVCDS*), and the observed CDS spread as below:

(1) $TBTF_{i,t} = FVCDS_{i,t} - CDS_{i,t}$

where *i* denotes bank$_i$ and *t* denotes time. The modelling of the *FVCDS* is based on the insight that a firm's equity and debt can be valued as options on the asset value of the firm.[12] This insight was originally introduced by Black and Scholes (1973) and subsequently developed by Merton (1974). In this framework, default occurs when the value of the firm's assets becomes lower than the value of its debt and is therefore insufficient to allow the firm to meet its contractual obligations. The firm's market asset value is unobservable, but it can be inferred from equity prices, the firm's liabilities structure and the business (asset value) risk of the firm. Over the past decades, this model has been widely applied as it provides a forward-looking measurement of the default probability and equity-implied credit spreads.

Another source of credit spreads is the CDS market. A CDS contract provides insurance against the risk of a default by a particular firm and provides a relatively direct and unbiased measure of the risk of default. In case of default, the seller of the CDS compensates the buyer for the losses that the buyer would otherwise incur on the amount insured. The insurance seller receives fixed periodic payments in return. The annual sum of those payments is usually expressed as a percentage of the face value of the CDS and quoted in basis points and is referred to as CDS spread. In this way, the CDS spread reflects the default risk of the firm.

The economic intuition behind the comparison as in (1) is that observed CDS spreads reflect *both* the probability of bank default *and* the likelihood and size of government support in case of default. In contrast, under the assumption that the government does not bail-out equity holders, the equity prices contain only information on the probability of default.[13] As a result, the equity price information allows the calculation of a hypothetical, 'equity-implied' *FVCDS* that is not affected by the probability of bailout.[14] The difference between the *FVCDS* and the observed *CDS* spread can therefore be interpreted as a measure of investors' expectations of government support.

3.1 Data used in the analysis

To compute *TBTF* premiums, we use data from Bloomberg and Markit (CDS). In the following analysis, we use macroeconomic variables that we collect from Bloomberg, with the exception of the variable Probability of Crisis, which we collect from the Stern-NYU's V-Lab initiative, which is based on Engle and Ruan (2019).[15] The source for the bank balance sheet variables such as return on equity, total capital ratio and total bank assets is the Fitch database.

11 There are different approaches to estimating TBTF premiums. The estimates of the premiums can differ significantly depending on the approach used. However, while the estimated level may differ, the pattern over time of the evolution of TBTF premiums is comparable, see for example IMF (2014) and FSB (2021).

12 A financial option is the right to buy or sell an underlying asset at a specified price on or before a specified maturity date. An option is particularly valuable when investors are more uncertain about the future value of the underlying asset.

13 The capital injection in a bailout can take several forms, such as loans, stocks, bonds or cash. While there is a possibility of the bank's shareholders being bailed out depending on the form of bailout, historically, for the large part, shareholders have suffered losses and not been bailed out in case of bank default. It is also worth noting that while both shareholders and bank creditors benefit from an implicit bank subsidy in terms of share prices and lower risk *ex-ante*, it is mainly bank creditors that benefit from a bail-out, *ex post*.

14 *FVCDS* and CDS may differ for other factors than the probability of bailout. For example, at times the *FVCDS* may be lower than the CDS. While it is not possible to exclude the influence of such factors, they have proved to be mostly transitory, see Dwyer et al. (2010).

15 See Volatility Laboratory (V-Lab), Stern Business School, New York University at http://vlab.stern.nyu.edu/welcome/risk/.

Market-based variables such as equity prices and volatility, CDS spreads, interest rates are at daily frequency whereas bank balance sheets are at quarterly frequency. Therefore, in order to mitigate unduly noise but keep as much information as possible from market-based data, we compute the TBTF premium on a monthly frequency by aggregating the market-based data and interpolating the quarterly bank balance sheet. We compute TBTF premiums for 53 SIBs. We select the banks according to the availability and quality of the data. In Appendix B, we list the banks in our sample and present some basic summary statistics of the data we use in our regression analysis.

3.2 Time variation of TBTF premiums

To evaluate TBTF premiums over time, we divide the time period into several sub-periods, corresponding to the period pre-GFC, the GFC, the post-GFC period and pre-reform period, the reform implementation period and the Covid-19 pandemic period as shown in Table 1. Figure 2 shows the TBTF premiums averaged across all countries in our sample for these sub-periods.

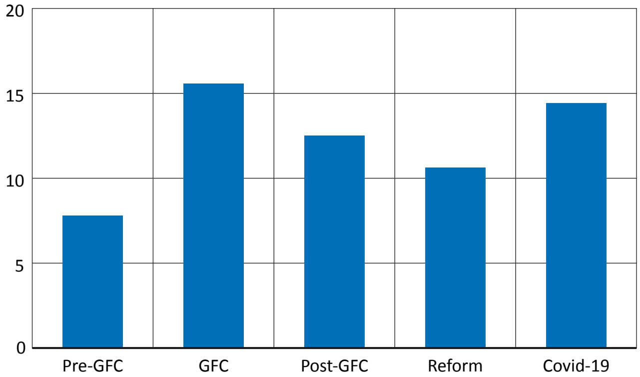
Table 1. Sub-periods analysed

Pre-GFC	GFC	Post-GFC	Reform	Covid-19
2004–2007	2008–2009	2010–2011	2012–2020:2	2020:3–2021:3

Note. The end of our sample is March 2021.

Figure 2. The TBTF premiums across sub-periods

Per cent



Note. Each bar represents the average TBTF premium in per cent in each sub-period. Pre-GFC denotes pre global financial crisis, i.e. 2004–2007; GFC denotes the global financial crisis, 2008–2009; Post-GFC denotes the period 2010–2011, i.e. post GFC and before the reform implementation period; Reform denotes the reform implementation period, 2012–(February) 2020; Covid-19 refers to the Covid-19 period, between March 2020 and the end of our period, i.e. March 2021. The TBTF premiums are computed at monthly frequency.
Sources: Bloomberg, Markit and authors' own calculations

Figure 2 shows three key results. First, TBTF premiums have declined from the peak of the GFC not only in the post-GFC period, but also continued to decline during the reform implementation period. The average premium has declined by about 6 percentage points corresponding to a decline in TBTF premiums of 60 per cent from the GFC to the reform implementation period. Second, TBTF premiums have not come down to pre-crisis levels but have remained 2–3 percentage points above throughout the reform implementation period. Thirdly, TBTF premiums have increased during the Covid-19 period, highlighting that theses premiums are highly time-dependent and indicating a positive correlation with the amount of market stress.

In order to corroborate our analysis above, we assess whether TBTF premiums are also statistically different across sub-periods by performing a panel regression analysis as follows:

(2) $$TBTF_{i,t} = \beta_0 + \beta_1 SubPeriod_1 + \beta_2 SubPeriod_2 + \beta_3 SubPeriod_3 + \beta_4 SubPeriod_4 + \varepsilon_{i,t}$$

where the dependent variable $TBTF_{i,t}$ is the TBTF premium for each $bank_i$ at time t. $SubPeriod_j$ with j equal to 1, 2, 3, and 4 denotes a dummy variable that takes value 1 for each corresponding period and zero otherwise. β_0 captures the average TBTF premium in the baseline period. The difference in TBTF premiums between the baseline and j period is measured by β_j.

We have five sub-periods. To measure the difference between the baseline and the other four sub-periods, we then run the analysis as in equation (2) five times, changing the baseline period to each of the five sub-periods as in Table 1. Table 2 shows the β_j values.

Table 2. TBTF premium differences among sub-periods, regression results

Sub-periods \ Baseline	Pre-GFC	GFC	Post-GFC	Reform
GFC	6.61*** (0.00)	–	–	–
Post-GFC	4.28*** (0.00)	−2.34*** (0.00)	–	–
Reform	2.70*** (0.00)	−3.91*** (0.00)	−1.57*** (0.00)	–
Covid-19	7.22*** (0.00)	−0.60 (0.50)	2.94*** (0.00)	4.52*** (0.00)

Note: Pre-GFC denotes pre global financial crisis, i.e. 2004–2007; GFC is the global financial crisis, 2008–2009; Post-GFC denotes the period during 2010–2011; Reform denotes the reform implementation period, 2012–2020:2; Covid-19 denotes the time period from 2020:3–2021:3. Robust p-values are presented in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The results relate to equation (2) where we include fixed-effects to control for the unobserved time-invariant heterogeneities across individual banks and countries. The baseline period is in columns. Each row reports the values of the estimated β_j from equation (2), i.e. the difference of the period on the row compared to the baseline (in column). Results are robust to different choices of the variance-covariance matrix estimator. Table 2 results are based on a variance-covariance matrix with cluster at bank level. The number of observations in each regression is 9,772.
Sources: Bloomberg, Markit and authors' own calculations

The differences in TBTF premiums across the sub-periods are all statistically significant, with the exception of the difference between the GFC and the Covid-19-period. For a further discussion of the time difference of TBTF premiums, see FSB (2021). The reasons why TBTF premiums in the reform implementation period have not fallen to pre-GFC levels are not fully clear. Later in the article, we therefore analyse how the *determinants* of TBTF premiums vary across time.

3.3 Regional variation of TBTF premiums

In this section, we analyse regional variation in the evolution of TBTF premiums. Similarly to Sarkar (2020), we analyse the dynamics of TBTF premiums for banks in Asia, Europe and North America, but we divide Europe in two sub-regions, the Nordic countries and the other European countries. Diversity within a certain region has been much less investigated in the literature and previous policy work on TBTF issues, and therefore our analysis contributes to that work with novel evidence on Europe.[16]

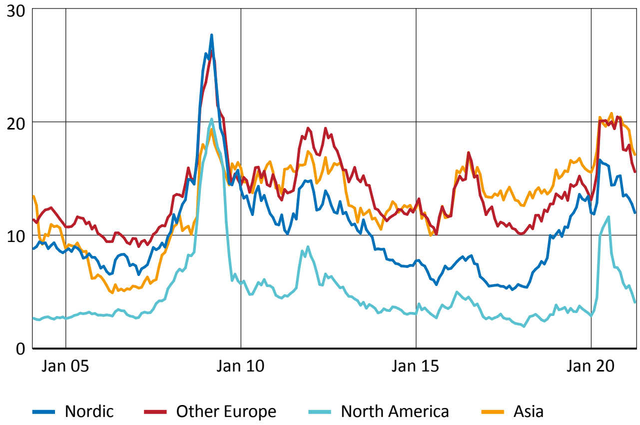
In this context, we define the *Nordic* region as Denmark, Norway, Sweden and Finland (from 2018), which share a similar structure of the economy, a well-integrated bank system within the region as well as cultural ties and history that – in many aspects – set them as a

16 In our analysis, we include banks from FSB countries such as Canada, France, Germany, Italy, the Netherlands, Japan, Spain, Switzerland, the UK and the US. We also include banks from Austria, Belgium, Denmark, Norway, Portugal, Sweden and Finland (from 2018). The period covered is January 2004–March 2021.

group apart from the other European countries.[17] Furthermore, none of them is a member of the Financial Stability Board. We label the other European countries as Other Europe (OE).[18]

Figure 3 illustrates the TBTF premium dynamics for the four regions – *Nordic*, *Other Europe*, *North America* and *Asia*.[19]

Figure 3. The TBTF premiums across regions
Per cent



Sources: Bloomberg, Markit and authors' own calculations

Figure 3 clearly displays regional variation in the evolution of TBTF premiums. TBTF premiums are highest for Asian banks and lowest for banks in North America. It also shows that there is a significant regional variation within Europe. The banks in the Nordic region appear to have lower TBTF premiums than in the other European countries in the sample, possibly reflecting structural differences among the regions and among banks. After the GFC, while banks in the Nordic region benefitted from an economic rebound and relative financial stability, banks in Other Europe have had relatively low profitability and comparably high levels of non-performing loans. The comparatively large increase in premiums for the Nordic banks observed during 2019 is possibly related to the fallout of money laundering investigations that involved several banks both in Denmark and Sweden which decreased equity shares prices and increased volatility, affecting their probability of default and TBTF premiums.

Figure 3 also shows that changes occurred in premiums between periods. While premiums significantly increased for all regions from the pre-GFC to the GFC-period, they have declined more for the North American and the Nordic banks in the following period. During the Covid-19 pandemic crisis, premiums rose for banks in all regions but increased significantly more for banks in the North American region, although from a much lower level. Compared to Europe, for example, premiums for the North American region almost tripled from the beginning of the pandemic to the peak during the summer 2020, whereas they increased about 30 per cent in Europe during the same period. By the end of the period in this analysis – end of March 2021 – premiums had declined significantly for both regions but not yet returned to pre Covid-19 crisis levels.

To investigate whether the differences we observe in Figure 3 are meaningful also in statistical terms, we perform a similar exercise as in the previous section, i.e. we run a panel regression analysis as in equation (3):

17 In October 2018 Nordea moved its headquarters from Sweden to Finland, becoming the largest D-SIB bank in Finland. Therefore *Nordic* includes banks in Finland but only from 2018.

18 Other Europe (OE) includes banks in Austria, Belgium, France, Germany, Italy, the Netherlands, Spain, Switzerland, Portugal and the UK.

19 North America includes banks in the United States and Canada; Asia denotes banks in Japan.

(3) $TBTF_{i,t} = \beta_0 + \beta_1 Region_1 + \beta_2 Region_2 + \beta_3 Region_3 + y_t + \varepsilon_{i,t}$

where y_t denotes time (year) fixed-effects and $Region_j$ with $j = 1, 2$ and 3 is a dummy variable that takes value 1 for the corresponding region and zero otherwise. The intercept β_0 measures the average TBTF premium in the baseline region. Table 3 shows the results. The heading of the columns shows the baseline region. The numbers in each row correspond to the estimated value of the coefficients β_j in equation (3) and measure the difference in TBTF premiums between regions. For example, in the first column, the coefficient for Other Europe (OE) is about 4, i.e. the TBTF premium for the banks in OE is 4 percentage points higher than for banks in the Nordic region. The difference between North America (NA) and the Nordic region is about −3.6, which means that the premiums are about 3.6 percentage points lower for banks in NA. Notably, differences *within* Europe are compelling. In fact, the difference between NA and OE is close to −7.6 percentage points and therefore more than double compared to the Nordic region. In addition, the difference between OE and Asia is relatively small and not statistically significant. In contrast, the premiums for banks in Asia compared to banks in the Nordic region are about 3 percentage points higher and statistically significant. Finally, the TBTF premiums of North American banks are significantly lower than those of Asian banks.

Table 3. Differences among regions, regression results

Regions \ Baseline	Nordic	OE	Asia
OE	4.00*** (0.00)	–	–
Asia	3.23*** (0.03)	−0.77 (0.60)	–
NA	−3.59*** (0.00)	−7.59*** (0.00)	−6.82*** (0.00)

Note. Nordic denotes banks in the Nordic region, OE denotes banks in Other Europe, NA denotes North American banks. Robust p-values are presented in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The results relate to equation (3) where we include time-fixed effects. Each row reports the difference of that region on the row compared to the baseline (in column). Results are robust to different choices of the variance-covariance matrix estimator. In Table 3, we report the results based on a variance-covariance matrix with a cluster at bank level. The number of observations in each regression is 9,772.
Sources: Bloomberg, Markit and authors' own calculations

3.4 The determinants of TBTF premiums

3.4.1 Methodology and variables

In the previous section, we showed a significant regional variation. As discussed before, it is possible that this variation is related to structural differences at country level and to individual bank characteristics. In this section, we investigate the determinants of TBTF premiums that can be related to the observed variation by running the following panel regression model with fixed effects:

(4) $TBTF_{i,c,t} = \alpha + \gamma_i + \beta_m Macro_{c,t} + \beta_b Bank_{i,t} + \delta_c * year_t + Covid\text{-}19_t + \varepsilon_{i,t}$

where the explanatory variables $Macro_{c,t}$ and $Bank_{i,t}$ account for macro-financial variables for a given country c and bank-specific variables for a given bank i, respectively.[20] We briefly describe them below. The variable γ_i denotes bank fixed-effects, which controls for differences among banks.[21] $\delta_c * year$ is a country-year fixed-effect, which controls for country-

20 In an alternative specification, we have included the one-period-lagged TBTF among the controls. We find that the regression results are by and large unaffected by this change.

21 For convenience, γ_i can be thought of as shorthand for a set of dummy bank variables each multiplied by their respective regression coefficients, i.e. a dummy variable for each bank multiplied by its regression coefficient.

specific economic and financial environment that is not already captured in the $Macro_{c,t}$, including the effects of the Basel III regulation and resolution reform framework.[22] $Covid\text{-}19_t$ is a variable that takes value 1 during the period of the Covid-19 pandemic – defined as the period from March 2020 to the end of the sample March 2021, and zero otherwise. Finally, α – the intercept – gives us the average TBTF premium, given zero values for all of the explanatory variables and $\varepsilon_{i,t}$ is the error term.

The vector $Macro_{c,t}$ includes the following five variables:

a) the VIX index as a measure of market risk aversion,
b) a probability of financial crisis developed by the Stern-NYU's V-Lab initiative, which is based on Engle and Ruan (2019),
c) monetary policy influence as measured by the two-year country-specific government bond yield,
d) a measure of longer-term, structural development of the economy as measured by the natural rate of interest r*[23], and
e) the sovereign debt to GDP ratio.

The rationale for including the risk aversion and the probability of financial crisis measures is that default probabilities jump up during downturns and periods of heightened risk aversion and financial stress. Therefore, the implicit guarantees underlying TBTF premiums become particularly valuable in times of crisis. On these grounds, we expect changes in TBTF premiums to be positively correlated with both the VIX and the probability of financial crisis measure. We expect positive long-term, structural developments to be negatively associated with TBTF premiums if they result in economic growth improvements and higher competition in the bank sector (see Boyd and De Nicolo 2005, Boyd et al. 2007 and Schaeck et al. 2009).

The influence of monetary policy (as measured by the short-term government bond yield) is *a priori* ambiguous. On one hand, monetary policy tightening may increase the probability of bank default and thus potentially increase TBTF premiums. On the other hand, higher interest rates may signal buoyant economic conditions and increases in asset values and therefore lower probability of default and potentially lower TBTF premiums.

Similarly, the impact of sovereign debt is *a priori* unclear. Countries that have a lower debt-to-GDP ratio have more fiscal capacity to support banks that may eventually fail or run into problems and therefore may be more inclined to intervene in the banking sector when needed, implying higher TBTF premiums the lower the debt-to-GDP ratio. On the other hand, countries that run relatively large debts may be more dependent on banks absorbing a significant share of issued government debt. In the latter case, the sovereign sector and the bank sector are more tightly linked and dependent on each other than otherwise, resulting in an increase of the bailout probability and therefore of TBTF premiums.

We include in $Bank_{i,t}$ the following three variables:

a) the total capital ratio as a measure of the solvency of a bank,
b) the return on equity for measuring bank profitability, and
c) bank size as measured by total assets (in log terms)

We expect higher capital buffers and better profitability to lower the probability of bank default and therefore to be negatively associated with TBTF premiums. The effect of bank

22 A more direct measure of progress on the implementation of resolution reforms is the Resolution Reform Index (RRI) presented in FSB (2021). The scope of the RRI is limited to the FSB members and therefore is not available for all countries in our sample. Given this data limitation, we control for the influence of regulatory changes indirectly through the time-varying country fixed-effects.

23 r* denotes the natural interest rate, i.e. the real interest rate expected to prevail when an economy is at full strength and inflation is stable. Changes in r* relate to shifts in demographics, slowdown in trend productivity growth and global factors affecting real interest rates, see Holston et al. (2017).

size on TBTF premiums is ambiguous *a priori*. Larger banks may tend to be more systemic and therefore more likely to receive some support from public authorities in case of distress. On the other hand, larger banks are more likely to benefit from economies of scale and scope. They also tend to have more advanced risk management and investment diversification strategies (see Laeven et al. (2014) and our discussion in section 2). We summarize the expected effects of the different variables in Table 4.

Table 4. Variables in the analysis and their expected impact on TBTF premiums

Variables	Description	Direction
Size	Log of total assets	+/–
Solvency	Total Capital Ratio (TCR)	–
Profitability	Return-on-equity (ROE)	––
ProbCr	Probability of systemic crisis	+
IntRate	2 year government bond yield	+/–
r*	Structural economic developments as measured by the natural interest rate	–
SovDebt	Ratio of sovereign debt to GDP	+/–
VIX	Market risk aversion	+

Note. Total Capital Ratio is defined as total capital held by a bank divided by its risk-weighted assets; VIX is the Chicago Board Options Exchange Volatility Index.

3.4.2 Regression results

Table 5 reports the results of our panel regression analysis.[24] The column *World* shows the regression results for all banks included in our sample. We find that higher bank capital and better profitability tend to lower the TBTF premium. The coefficient of bank size is positive with statistical significance. This suggests markets perceive that larger banks tend to be associated with larger TBTF premiums.

Next, turning to the estimates of macro variables in column *World*, as expected, higher probability of financial crisis and higher market risk aversion are associated with higher TBTF premiums. Higher interest rates are associated with higher TBTF premiums, giving some support to the hypothesis that tightening of monetary policy increases the probability of default of banks and therefore their TBTF premiums. Better structural economic developments (as measured by r^*) are associated with lower TBTF premiums whereas sovereign debt turns out to be statistically insignificant. The Covid-19 dummy is also insignificant.

24 The results presented in Table 5 are robust to changes in variables and period length. In particular, for our robustness checks we have used the Tier 1 capital ratio, ROA, the difference between the 10- and 2 year government bond yield and sovereign CDS.

Table 5. Panel regression results

TBTF / *Variables*	World	Nordic	OE	Asia	NA
Size	0.75*** (0.00)	6.22*** (0.00)	1.19*** (0.00)	1.48*** (0.00)	−0.01 (0.97)
TCR	−0.10*** (0.00)	−0.08*** (0.00)	0.01 (0.81)	−0.19*** (0.00)	−0.07 (0.10)
ROE	−0.11*** (0.00)	−0.16*** (0.00)	−0.14*** (0.00)	−0.01 (0.46)	−0.03*** (0.00)
ProbCr	0.06*** (0.00)	0.09*** (0.00)	0.06*** (0.00)	0.04*** (0.00)	0.07*** (0.00)
IntRate	0.53*** (0.00)	0.50** (0.03)	0.61*** (0.00)	−3.56*** (0.00)	−0.11 (0.39)
r*	−1.82*** (0.00)	−1.89*** (0.00)	−0.62 (0.12)	−1.03** (0.05)	−3.16*** (0.00)
SovDebt	−0.02** (0.02)	0.04 (0.24)	−0.00 (0.81)	0.01 (0.52)	−0.16*** (0.00)
VIX	0.13*** (0.00)	0.15*** (0.00)	0.13*** (0.00)	0.10*** (0.00)	0.13*** (0.00)
D_Covid-19	0.34 (0.17)	−2.45*** (0.00)	0.83** (0.03)	−0.29 (0.44)	1.21*** (0.00)
No. Obs.	9,067	1,221	4,415	1,082	2,349
Adj. R-squared	0.89	0.91	0.86	0.90	0.79
Fixed-effects	Yes	Yes	Yes	Yes	Yes

Note. Size is the bank's total assets (log), TCR is the bank's total capital ratio, ROE denotes the bank's return-on-equity, ProbCr is the probability of systemic crisis, IntRate is the 2-year government bond yield, r* is the natural interest rate, SovDebt denotes the sovereign debt-to-GDP ratio, VIX is the Chicago Board Options Exchange Volatility Index and D_Covid-19 is a dummy variable with value 1 for the Covid-19 period covered by our analysis, i.e. from March 2020 to the end of the sample, March 2021. We include bank and country by year fixed-effects. Robust p-values are presented in parentheses *** p<0.01, ** p<0.05, * p<0.1.

When examining the results by regions, several differences emerge. Among macroeconomic variables, positive structural, long-term economic developments (measured by r^*) appear to be associated with lower TBTF premiums, as expected. Higher probability of crisis and VIX are associated with higher TBTF premiums for all regions whereas the results for sovereign debt are mixed depending on the region. An increase in the government bond yield increases the premiums for banks in Europe and it lowers them for banks in Asia, possibly reflecting structural differences in the banking sector and the economy. One interpretation is that in countries characterized by both weak economic growth and a dominant banking sector, an increase in interest rates is more likely to lead to an increase in bank probability of default and therefore higher TBTF premiums. This may be the case for many countries in Europe.

Turning to the effect of bank characteristics, there are large differences in the impact of the bank *size* variable. Specifically, the coefficient of bank size in the Nordic region is higher than in other regions, especially compared to Other Europe. This may reflect the fact that the banking sector in the Nordic region is particularly large in relation to the size of the economy. Thus, all else equal, even a small change in bank size has a comparatively larger impact on TBTF premiums in the Nordic region compared to in Other Europe. Bank profitability is negatively associated with TBTF premiums for every region, although the impact is larger for Europe than Asia and North America. The coefficients for banks' capital ratios have the expected sign for every region except Other Europe, and they are not statistically significant in Other Europe and North America. More specifically for the Nordic region, a one percentage point increase in TCR is associated with a reduction of 8 basis points in TBTF premiums. It is not obvious how to assess the economic significance of these estimates. Assuming – just for illustrative purposes – that a third of the outstanding debt could be bailed-out in a crisis, the analysis indicates that a one percentage point increase of TCR corresponds to a reduction in the implicit subsidy of about 2 per cent in terms of average GDP of the region (at current prices, 2019).[25] These estimates are highly uncertain but indicate that the economic impact of these TBTF premiums is not insignificant.

Finally, we examine the effect of the Covid-19 pandemic through the variable: *D_Covid-19*. TBTF premiums increased during the pandemic crisis and reached a peak during the summer of 2020. After that, they started to gradually decline for most countries and, in fact, for several countries by the end of our sample, i.e. March 2021, they had returned to levels close to those prevailing before the Covid-19 pandemic. The analytical results in Table 5 show that the Covid-19 shock increased TBTF premiums for the *World*, but this is not statistically significant. However, when analysed by region, there are significant differences. In fact, the Covid-19 period is associated with an increase in TBTF premiums for banks in Other Europe and North America whereas it has a negative sign for banks in the Nordic region (significant) and Asia (non-significant). These results are broadly in line with the evidence of the larger negative economic implications of the pandemic crisis for Other Europe and North America compared to the other regions we analyse; see, for example, IMF (2021). However, the results must be taken *cum grano salis* given the relatively short length of the Covid-19 period analysed. In addition, the interventions by public authorities eased market conditions and significantly influenced the ability of firms to withstand the economic consequences of the pandemic with potential confounding effects on TBTF premiums.

All in all, the analysis shows that the main driving factors of TBTF premiums and their significance differ across regions. Even within Europe, the determinants of TBTF premiums for banks in the Nordic region appear to be different from those of other European banks. In particular, bank size and capital appear to matter more for TBTF premiums in Nordic region than in Other Europe.

3.4.3 The relative influence of TBTF determinants has changed

In the previous section, we showed that bank size matters for the dynamics of TBTF premiums. Laeven et al. (2014) lists three main reasons why bank size matters for systemic risk and, ultimately, for TBTF premiums: a) large banks may benefit from better diversification, which reduces risks. They may also carry out market-based activities more competitively; b) the cost of debt for large banks is lower if they are perceived to be TBTF and c) large bank have specific corporate governance challenges, which increase their leverage and risk-taking. Thus, a larger bank size would be associated with lower TBTF premiums in the first case and higher in the following two.

Yet, the relative significance of the determinants of TBTF premiums may be changing over time. Antill and Sarkar (2018) document that bank size has become a less significant driver of systemic risk compared to other factors such as financial system interconnectedness and complexity. The reason for this change is still not fully clear.

In order to test whether this finding is also true for the TBTF premiums of banks in our sample, we deploy a modified specification of equation (2) that includes interaction terms between variables of interest and year dummy variables. This dynamic set-up enables us to assess whether a coefficient of interest evolves in a statistically significant manner over time.[26] Figure 4 shows the average coefficients of bank size by sub-periods. The larger the coefficient, the larger the impact that a change in assets has on premiums. The coefficient increased during GFC. Since then, it has declined. This suggests that the importance of bank size, while still being a key determinant of TBTF premiums, has been declining.

25 To translate the premiums in monetary terms we have multiplied it by the amount of debt that could be bailed out. However, the bailout-able debt is unknown *ex ante* and it is one reason why translating the premiums into monetary terms is not straightforward. In the text, we have assumed that one third of the outstanding debt would be bailed out, but these estimates are only for illustrative purposes.

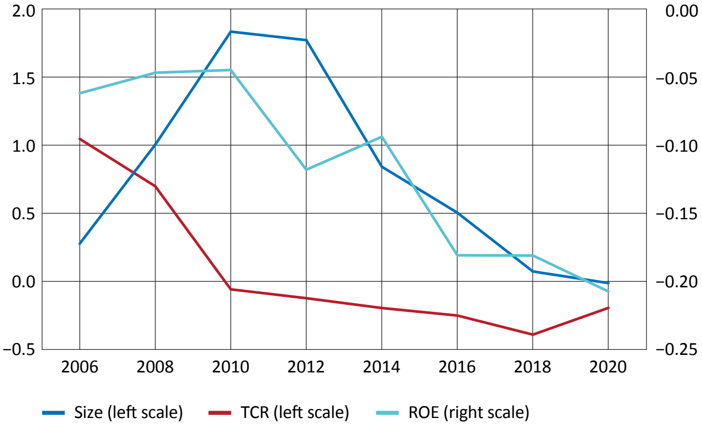
26 We provide detailed results in Appendix C.

Figure 4 also shows that while the importance of bank size has fallen, the significance of bank profitability as an explanatory variable has steadily increased over the same period, as the coefficient has become increasingly negative. Goel et al. (2021) find that profitability has been a key determinant but often overlooked when analysing the response of SIBs to regulatory reforms. They argue that profitability has been a more important determinant of a bank's response to the TBTF regulation than factors such as the business model or domicile. They find that only banks that are less profitable have reduced their systemic footprint. Our results show that profitability has become more of a key driver of TBTF premiums after the GFC, suggesting that as banks' response to regulation was determined by the level of profitability, investors' perceptions of TBTF premiums correspondingly pivoted towards bank profitability.

In Figure 4, we also investigate the impact of bank capital on TBTF premiums. It has considerably changed over time. In the pre-GFC period, an increase in bank capital is associated with higher TBTF premiums, but this relationship has changed significantly over time. The impact of capital on TBTF premiums turned negative in the reform implementation period, implying that an increase in capital is associated with a reduction in TBTF premiums. This positive sign of the impact of capital on TBTF premiums before the GFC is in line with previous evidence on the inability of the pre-GFC capital regulation to prevent a crisis and the need for governments around the world to step in with emergency support and bail out the financial sector. In fact, capital adequacy ratios of banks that benefitted from bailout during the GFC tended to be even higher than those of other banks, see IMF (2009). In addition, many banks that were rescued during the GFC appeared to be in compliance with minimum capital requirements shortly before and even during the crisis. During the post-GFC period, Basel III rules made capital requirements more stringent and the requirements were gradually implemented. Our results suggest that market participants have come to view the new set of capital standards (and total bank capital) as increasingly more informative about bank performance and ability to withstand losses. Starting from 2010, total capital requirements are negatively associated with TBTF premiums. However, our analysis cannot exclude that these results may also be driven by the normalization of financial and economic conditions after the GFC.

Figure 4. The impact of bank size, profitability and capital on TBTF premiums over time

Coefficient



Note. Size is measured by bank total assets (log), TCR denotes total capital ratio and ROE is return on equity.

Sources: Bloomberg, Markit and authors' own calculations

From the perspective of the TBTF issue, the results on bank size, profitability and bank capital suggest that market participants' perception of what determines a bank to be TBTF has moved away from sheer size. It is reasonable that after the GFC and implementation of Basel III regulatory changes and bank resolution reforms, market perceptions have been recalibrated and that bank profitability and larger capital buffers have increasingly become more relevant for market participants' assessment of TBTF premiums.

Notably, the changing perceptions may have contributed to the higher level of TBTF premiums during the reform implementation period compared to in the pre-GFC period. As bank profitability declined dramatically after the GFC for a majority of banks in our sample and remained subdued for most of the period we examine, the heightened attention on bank profitability as a key driver of TBTF premiums may have contributed to keeping them at higher levels than during the pre-GFC period. In addition, our results point to a relatively larger influence of bank profitability compared to, for example, total capital. This suggests that the influence of increased capital ratios on market participants' assessment of TBTF may have been counterbalanced by the relatively poor development of bank profitability during the reform implementation period in many countries.

3.5 Heterogeneity among types of SIBs

In the literature and previous policy work, SIBs are commonly treated as a relatively uniform sample of banks.[27] But is this the case in reality? In this section, we analyse whether and to what degree G-SIBs and D-SIBs differ with respect to the evolution of TBTF premiums and whether and to what degree determinants of TBTF premiums vary between the two groups.[28]

In the following analysis, we classify our sample into G-SIBs and D-SIBs. We include a bank in the group of G-SIBs if it appears at least once on the G-SIB list, as published by the FSB, and apply its status through all the periods. We include banks in the D-SIB group that have been designated as such by national authorities. We limit this part of the analysis to European banks. There are two reasons for this: a) we avoid mixing potentially larger regional differences between European countries and other regions; b) we have a more balanced group of G-SIBs and D-SIBs from Europe.[29] The banks included can be found in Appendix B. In the analysis of this section, we use the subset of banks, i.e. 14 G-SIBs and 18 D-SIBs from the Nordic region and Other Europe.

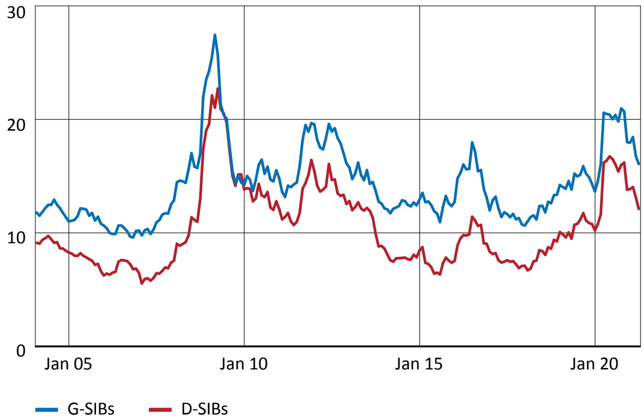
Figure 5 shows TBTF premiums for both European G-SIBs and D-SIBs over time.

27 BCBS (2019) contains a more general analysis of the differences between G-SIBs and other banks in the G-SIB sample. However, it does not deal with the TBTF premiums of these banks.

28 With regard to the impact of resolution reforms on TBTF premiums, FSB (2021) uses the RRI index in regression analysis and points out that its coefficient is larger for D-SIBs than G-SIBs and statistically significant only for D-SIBs. See FSB (2021) – Addendum to the Technical Appendix for details.

29 We use the list of D-SIBs regularly published by the European Banking Authority (EBA). We select the banks that have been continuously present over the period of our analysis and for which market data are available. Our D-SIB group includes Erste Bank, Raiffeisen Bank, KBC, Commerzbank, Danske Bank, BBVA, BFA (Bankia), Caixia, Banco Popular Español, Banco de Sabadell, Lloyds, Banca Monte dei Paschi di Siena, Intesa Sanpaolo, DNB, Banco BPI, SEB, Handelsbanken, and Swedbank. The G-SIB group includes Credit Suisse, UBS, Deutsche Bank, Santander, Crédit Agricole, BNP Paribas, Société Générale, Barclays, HSBC, RBS, Standard Chartered, Unicredit, ING, and Nordea. Nordea changed from being classified as a G-SIB to a D-SIB in 2019, but for sake of simplicity and because the period is relatively short, we keep it in the G-SIB group. Note that due to the availability of data for several countries, we have only G-SIBs (for example, Switzerland) or only D-SIBs (for example, Denmark).

Figure 5. TBTF premiums: G-SIBs vs. D-SIBs
Per cent



Sources: Bloomberg, Markit and authors' own calculations

Figure 5 shows that D-SIBs have lower TBTF premiums. After the GFC, the difference in premiums for the two groups has become larger as D-SIBs premiums have declined more than those of G-SIBs. In addition, the difference has reached a level close to that prevailing before the GFC.

What drives these results? We rerun the empirical analysis as in equation (4) but we add interaction terms between a G-SIB dummy and the variables in the vector *Bank*, i.e. size (total assets), capital (total capital ratio) and profitability (ROE), and in the vector *Macro* as follows:

$$(5) \quad TBTF_{i,t} = \alpha + \gamma_i + \beta_{m1}Macro_{c,t} + \beta_{m2}Macro_{c,t} * G\text{-}SIB_{i,t} + \beta_{b1}Bank_{i,t} + \beta_{b2}Bank_{i,t} * G\text{-}SIB_{i,t} + \delta_c * year_t + Covid_t + \varepsilon_{i,t}$$

The G-SIB dummy takes value one for banks that are in the G-SIB group and zero otherwise.[30] The interaction terms measure the difference in the effect of being a G-SIB compared to other bank types, i.e., in our sample, D-SIBs.

30 The interaction term can be regarded as an adjustment to the slope coefficients of the variables in the vector *Bank*.

Table 6. Panel regression: G-SIBs vs. D-SIBs

Variables	TBTF
Size	1.66*** (0.00)
Size*G-SIB	−1.33*** (0.00)
TCR	−0.14*** (0.00)
TCR*G-SIB	0.14*** (0.00)
ROE	−0.13*** (0.00)
ROE*G-SIB	−0.11*** (0.00)
ProbCr	0.07*** (0.00)
ProbCr*G-SIB	−0.01*** (0.00)
IntRate	0.54*** (0.00)
IntRate*G-SIB	0.41*** (0.00)
r*	−0.62* (0.07)
r* * G-SIB	−0.87*** (0.00)
SovDeb	0.03** (0.03)
SovDebt*G-SIB	−0.05*** (0.00)
VIX	0.12*** (0.00)
VIX*G-SIB	0.02** (0.05)
D_Covid-19	−0.17 (0.25)
No. Obs.	5,636
Adj. R-squared	0.88
Fixed Effects	*Yes*
Region	Europe

Note. Size is the bank's total assets (log), TCR is total capital ratio, ROE denotes return-on-equity, ProbCr is the probability of a systemic crisis, IntRate is the 2-year government bond yield, r* is the natural interest rate, SovDebt denotes the sovereign debt-to-GDP ratio, VIX is the Chicago Board Options Exchange Volatility Index and D_Covid-19 is a dummy variable with value 1 for the Covid-19 period covered by our analysis, i.e. from March 2020 to the end of the sample, March 2021. We only include banks from the Nordic region and Other Europe, see Appendix B for a list of the banks in the analysis. We include bank and country by year fixed effects. We present robust p-values in parenthesis, *** p<0.01, ** p<0.05, * p<0.1.

The results in Table 6 confirm our previous finding that, in general, an increase in bank size is associated with an increase in TBTF premiums. Larger banks have higher TBTF premiums. However, when the interaction term with the dummy for G-SIBs is included, the coefficient turns negative, indicating that the positive correlation between size and TBTF premium is driven by the D-SIBs and that the effect of size on TBTF premiums for G-SIBs is muted. Thus, size seems to be less of a TBTF premium driving factor for G-SIBs, potentially because they are already seen as TBTF, and that any additional size does not affect that assessment.

Similarly, in line with the overall results from Table 3, as a bank's capital increases, the TBTF premium falls significantly. However, that does not seem to be the case for G-SIBs, as we get a significantly positive coefficient for the interaction term between capital and G-SIB status. Thus if a G-SIB increases its capital, the bank's TBTF premium does not fall. Market participants still seem to assess it as TBTF.

Interestingly, the coefficient for the probability of a crisis also has a different sign for the multiplicative term. The coefficient for the probability of a crisis is similar to our results in Table 3, indicating that when the probability of a crisis increases, TBTF premiums increase significantly. However when looking at the G-SIBs only, the coefficient turns statistically significant and negative. A potential explanation is that G-SIBs already are considered to be

TBTF. It is also possible that the uncertainty about how to coordinate a resolution of a G-SIB is larger than for a D-SIB. Thus, in line with the result from Figure 5, as G-SIBs have higher TBTF premiums the additional increase may be lower as a crisis looms.

The effect of the sovereign debt-to-GDP is interesting. In the regression for the *World* (Table 3), the coefficient is negative and significant, indicating that the higher the sovereign debt is, the lower the TBTF premiums are. Looking at the Nordic region and Other Europe, the coefficients are not statistically significant. When we look at the same coefficient in Table 4, with data only from European banks, the coefficient becomes statistically significant and positive, indicating that as sovereign debt increases, TBTF premiums also increase. However, when multiplied by the G-SIB dummy, the coefficient is negative. In general terms, our results suggest that country-specific variables, such as public debt, matter less for G-SIBs than D-SIBs, potentially reflecting G-SIBs' less dependence on country-specific characteristics.

These results need to be interpreted with care because the number of European banks included in the analysis is relatively small. Nonetheless, they suggest that when factors such as bank size and bank capital are considered, there are clear differences in TBTF premiums between G-SIBs and D-SIBs, and these differences refer both to the levels of these premiums but also to the driving forces. This heterogeneity warrants further analysis and a continued monitoring by policy makers.

4 Conclusion

In this article, we provide an overview of the systemically important banks and the reforms to address too-big-to-fail (TBTF) problem. The overarching aim of the reforms was to increase the resilience of the banking system and mitigate the TBTF problem. The key elements of the framework comprise additional loss-absorbing capacity and resolution requirements for SIBs.

One quantitative approach to assess the effects of the reforms is to measure TBTF premiums. They increase because SIBs may benefit from an implicit funding subsidy that reflects the probability of bailout. Incentives to take on risk in financial institutions can also push TBTF premiums higher. This, in turn, increases both the likelihood and cost of bank failures.

In this article, we calculate the premiums using a Merton-type structural credit pricing model. We find that TBTF premiums diminished after 2012. This development coincides with the time when post-crisis reforms were announced and implemented for the countries in our sample.

However, our results show that there is a significant regional variation in the evolution of TBTF premiums, with reductions in premiums for Europe and Unites States but less for Asia. In addition, we show that while TBTF premiums have decreased for large European banks, there is considerable variation also within Europe. We provide novel evidence on banks in the Nordic region having lower TBTF premiums than other European banks.

We also analyse the determinants of TBTF premiums and find that the factors that affect the premiums have changed over time, in particular after the GFC. Bank size appears to matter less now than earlier, whereas the significance of bank capital and bank profitability as drivers of TBTF premiums has increased considerably. Notably, before 2010, higher bank capital signalled higher TBTF premiums, while after the GFC, higher bank capital signals lower TBTF premiums.

Furthermore, we investigate whether there are differences in the determinants of TBTF premiums of G-SIBs versus D-SIBs. Notably, we find that size is positively associated with TBTF premiums but primarily for D-SIBs. Compared to D-SIBs, TBTF premiums for G-SIBs are negatively associated with size. We also find some evidence that bank capital reduces TBTF premiums for D-SIBs whereas improvement in profitability lowers them for both G-SIBs and D-SIBs.

Finally, the Covid-19 crisis appears to have increased TBTF premiums in North America and Europe – but not in the Nordic countries, and not in Asia.

References

Antill, Samuel and Asani Sarkar (2018), 'Is Size Everything?', FRB of New York, Staff Report no. 864.

BCBS (2011), 'Global systemically important banks: assessment methodology and the additional loss absorbency requirement', Rules text, November 2011, Basel Committee on Banking Supervision.

BCBS (2012), 'A framework for dealing with domestic systemically important banks', October 2012, Basel Committee on Banking Supervision.

BCBS (2013), 'Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement', July 2013, Basel Committee on Banking Supervision.

BCBS (2016), 'Systemically important banks: Basel Committee publishes implementation assessments on frameworks', June 2016, Basel Committee on Banking Supervision.

BCBS (2017a), 'Basel III leverage ratio framework – Executive summary', FSI Executive Summaries, October 2017, Basel Committee on Banking Supervision.

BCBS (2017b), 'Basel III: Finalising post-crisis reforms', December 2017, Basel Committee on Banking Supervision.

BCBS (2018), 'Global systemically important banks: revised assessment methodology and the higher loss absorbency requirement', July 2018, Basel Committee on Banking Supervision.

BCBS (2019), 'An examination of initial experience with the global systemically important bank framework', BCBS Working paper 34, Basel Committee on Banking Supervision.

BCBS (2020), 'Governors and Heads of Supervision announce deferral of Basel III implementation to increase operational capacity of banks and supervisors to respond to Covid-19', Press release, 27 March 2020, Basel Committee on Banking Supervision.

Beccalli, Elena and Ludovico Rossi (2017), 'Economies or Diseconomies of Scope in the EU Banking Industry?', mimeo, Università Cattolica del Sacro Cuore (Italy) and University of Reading (UK).

Berger, Allen N., Rebecca S. Demsetz and Philip E. Strahan (1999), 'The consolidation of the financial services industry: Causes, consequences and implications for the future', *Journal of Banking and Finance*, vol. 23 no. 2–4, pp. 135–194.

Black, Fischer and Myron Scholes (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654.

Boot, Arnoud (2017), 'The Future of Banking: From Scale & Scope Economies to Fintech', *European Economy – Banks, Regulation and the Real Sector*, no. 2.

Boyd, John and Gianni De Nicolo (2005), 'The theory of bank risk taking competition revisited', *Journal of Finance*, vol. 60, no. 3, pp. 1329–1343.

Boyd, John, Gianni de Nicolo and Abu M. Jalal (2006), 'Bank risk taking and competition revisited: New Theory and Evidence', IMF Working paper, WP/06/297.

Dwyer, Douglas, Zan Li, Shisheng Qu, Heather Russell and Jing Zhang (2010), 'CDS-implied EDF™ credit measures and fair-value spreads', Moody's Analytics.

ECB (2020), 'Guide on the supervisory approach to consolidation in the banking sector', Public Consultation ECB Guide on the supervisory approach to consolidation in the banking sector (europa.eu).

Engle, Robert and Tianyue Ruan (2019), 'Measuring the probability of financial crisis', PNAS September 10, 2019 116 (37) 18341–18346; https://doi.org/10.1073/pnas.1903879116.

FSB (2014), 'Key Attributes of Effective Resolution Regimes for Financial Institutions', October, Financial Stability Board.

FSB (2015), 'Principles on Loss-absorbing and Recapitalisation Capacity of G-SIBs in Resolution, Total Loss-absorbing Capacity (TLAC) Term Sheet', November 2015, Financial Stability Board.

FSB (2020), '2020 list of global systemically important banks (G-SIBs)', November 2020, Financial Stability Board.

FSB (2021), 'Evaluation of the effects of too-big-to-fail reforms', April 2021, Financial Stability Board.

Finger, Christopher, Vladimir Finkelstein, Jean-Pierre Lardy, George Pan, Thomas Ta and John Tierney (2002), CreditGrades technical document, Risk Metrics Group.

Goel, Tirupam, Ulf Lewrich and Aakriti Mathur (2021), 'Does regulation only bite the less profitable? Evidence from the too-big-to-fail reforms', BIS working paper, no. 922.

G-20 (2009), 'Leaders statement – The Pittsburgh summit', 24–25 September.

Holston, Kathryn, Thomas Laubach, and John C. Williams (2017), 'Measuring the Natural Rate of Interest: International Trends and Determinants', *Journal of International Economics*, vol. 108, pp. 39–75.

IMF (2009), 'Detecting Systemic Risk', Global Financial Stability Report, ch. 3.

IMF (2014), 'How Big Is The Implicit Subsidy For Banks Considered Too Important To Fail?', Global Financial Stability Report, ch. 3.

IMF (2021), *World Economic Outlook Update*, January 2021.

Jensen, Michael C. (1986), 'Agency Cost of Free Cash Flow, Corporate Finance, and Takeovers', *American Economic Review*, vol. 76, no. 2, pp. 323–329.

Jensen, Michael C. and William H. Meckling (1976), 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure', *Journal of Financial Economics*, vol. 3, no. 4, pp. 305–360.

Laeven Luc, Lev Ratnovski and Hui Tong (2014), 'Bank Size and Systemic Risk', IMF Staff Discussion Note, no. 1404.

Merton, Robert C. (1974), 'On the pricing of corporate debt: the risk structure of interest rates', *Journal of Finance*, vol. 29, no. 2, pp. 449–470.

Sarkar, Asani (2020), 'Did Too-Big-To-Fail Reforms Work Globally?', Federal Reserve Bank of New York, Liberty Street Economics, https://libertystreeteconomics.newyorkfed.org/2020/09/did-too-big-to-fail-reforms-work-globally.html.

Schaeck, Klaus, Martin Čihák and Simon Wolfe (2009), 'Are Competitive Banking Systems More Stable?', *Journal of Money, Credit and Banking* vol. 41, no. 4, pp. 711–734.

Schweikhard, Frederic A. and Zoe Tsesmelidakis (2012), 'The impact of government interventions on CDS and equity markets', Working paper, University of Oxford.

Tsesmelidakis, Zoe and Robert C. Merton (2012), 'The Value of Implicit Guarantees', https://ssrn.com/abstract=2231317 or http://dx.doi.org/10.2139/ssrn.2231317

Appendix A

Model description

The estimation of the equity-implied CDS spreads is based on a Merton-type structural credit pricing model developed by Finger et al. (2002) and applied, among others, in Schweikhard and Tsesmelidakis (2012) and Tsemelidakis and Merton (2012). It shares many key characteristics with the original Merton (1974) model.

As in Merton (1974), in this Merton-type model, equity and debt are valued as contingent claims on the firm's value and the distribution of the risk and return of the debt is based only on the firm's fundamentals, i.e. its liabilities structure, equity prices and equity volatility. Default occurs when the asset value falls below a certain default barrier. Notably, the level of the barrier is uncertain.

The basic assumption is that the asset value evolves accordingly as a geometric Brownian motion:

(6) $$\frac{dVt}{Vt} = \mu_v dt + \sigma_v dW_t$$

where W_t is a standard Brownian motion, σ_v is the asset volatility and μ_v is the asset drift. μ_v is set to zero for simplicity.

The default barrier is defined as the amount of the firm's assets that remain after default, i.e. is the amount of asset value recovered by debt holders, *LD*. *L* is the average recovery on the debt-per-share, *D*.

The uncertainty in the barrier is related to *L*, which is assumed to follow a lognormal distribution with average $\bar{L}$ and standard deviation λ. With an uncertain recovery rate, the default barrier can be reached unexpectedly, resulting in a default event.

The survival probability of the firm at time *t*, is then given by the probability that the asset value does not reach the default barrier before time *t* and can be expressed in closed form as:

(7) $$P(t) = \Phi\left(-\frac{At}{2} + \frac{log(d)}{At}\right) - d * \Phi\left(-\frac{At}{2} - \frac{log(d)}{At}\right)$$

where $d = \frac{V_0}{LD} exp\lambda^2$ and $A_t^2 = \sigma_v^2 t + \lambda^2$. Φ denotes the cumulative normal distribution function and σ_v the asset volatility. Note that the mean default barrier changes over time with the capital structure of the firm.

Finally, to convert the probability of survival into a equity-implied CDS or fair-value CDS *(FVCDS)* spread, two more parameters are specified, i.e. the risk-free interest rate *r*, and the probability of recovery of unsecured debt, *R*. Given the metrics above, *FVCDS* can be written as:

(8) $$FVCDS = r(1-R)\frac{1-P(0)+e^{r\xi}(G(t+\xi)-G(\xi))}{P(0)-P(t)e^{-rt}-e^{r\xi}(G(t+\xi)-G(\xi))}$$

where $\xi = \frac{\lambda^2}{\sigma_v^2}$, *r* is the deterministic risk-free interest rate, and *R* is the bond-specific expected recovery rate. The function $G(u)$ is as in Finger et al. (2002).

A key variable of the *FVCDS* is the asset volatility. Finger et al. (2002) show that it can be approximated by:

(9) $$\sigma_v = \sigma_S \frac{S}{S+\bar{L}D}$$

where *S* is the equity price, *D* is the debt per share and σ_S the equity volatility.

The input data and the calibration of the model

The computation of the *FVCDS* requires the following eight input variables: the equity price S, the debt per share D, the interest rate r, the average default threshold $\overline{L}$, the default threshold uncertainty λ, the bond recovery R, the time to expiration T, and the equity volatility σ_S. In short notation, the *FVCDS* can be written as follows:

(10) $FVCDS = f(S_t, D_t, r_t, T - t, \sigma_t; R, \overline{L}, \lambda)$

The three parameters $(R, \overline{L}, \lambda)$ are unobservable and were estimated by Finger et al. (2002). Following Schweikhard and Tsesmelidakis (2012), we set

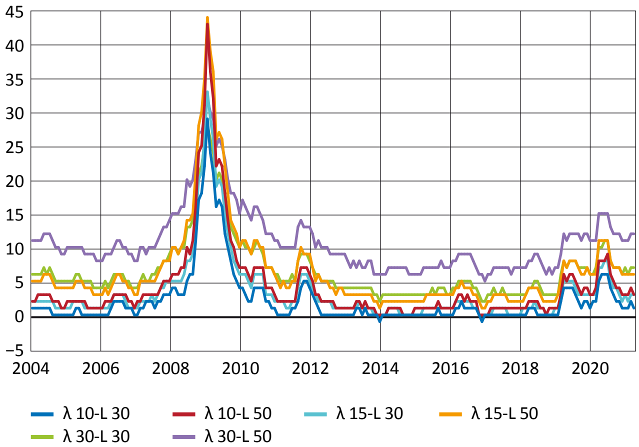
- λ, the standard deviation of the recovery rate of the firm's debt, to 0.3
- R, the recovery rate of unsecured debt, to 0.5
- $\overline{L}$, the recovery rate averaged over all debt classes, to 0.5, as in Finger et al. (2002)
- The debt per share D is calculated as the balance sheet total liabilities over the number of outstanding shares
- The risk-free interest rate r is the one-year government bond yield
- The equity volatility σ_t is the historical volatility of equity returns
- S is equity prices

To estimate the *FVCDS*, all model input data are collected from Bloomberg whereas observed CDS data that are subtracted from *FVCDS* to compute TBTF premiums are collected from Markit CDS data.

To avoid unduly noise, we compute the *FVCDS* at monthly frequency from daily market data and quarterly balance sheet data. The balance sheet data are linearly interpolated. One advantage of interpolating quarterly balance sheet data compared to using the last observable as fixed value for the months through the quarter is that it mitigates potential problems related to balance sheet management, such as window dressing practices.

Figure A1. TBTF premiums for Swedbank

Per cent



Source: Authors' own calculations

We have also done some robustness checks to assess how sensitive the model is to the parameters λ and $\overline{L}$. Figure A1 shows some different combinations of λ and $\overline{L}$ for Swedbank and how they affect the TBTF premium. Both parameters change the level of the TBTF premium, but not the pattern. There is also a trade-off between them. A higher value of λ can be off-set with a lower value of $\overline{L}$. The empirical regularities in Figure A1 are similar for all banks.

Appendix B

Table B1. List of banks included in the empirical analysis

Bank name	D-SIB / G-SIB	Country (headquarters)	Regional group
Banca Monte dei Paschi di Siena	D-SIB	Italy	Other Europe
Banco BPI	D-SIB	Portugal	Other Europe
Banco de Sabadell	D-SIB	Spain	Other Europe
Banco Popular Español	D-SIB	Spain	Other Europe
Bank of America	G-SIB	US	North America
Bank of Montreal	D-SIB	Canada	North America
Bank of New York Mellon	G-SIB	US	North America
Bank of Nova Scotia	D-SIB	Canada	North America
Barclays	G-SIB	UK	Other Europe
BBVA	D-SIB	Spain	Other Europe
BFA (Bankia)	D-SIB	Spain	Other Europe
BNP Paribas	G-SIB	France	Other Europe
Caixia	D-SIB	Spain	Other Europe
CIBC	D-SIB	Canada	North America
Citigroup	G-SIB	US	North America
Commerzbank	D-SIB	Germany	Other Europe
Crédit Agricole	G-SIB	France	Other Europe
Credit Suisse	G-SIB	Switzerland	Other Europe
Daiwa Securities Group, Inc.	D-SIB	Japan	Asia
Danske Bank	D-SIB	Denmark	Nordic
Deutsche Bank	G-SIB	Germany	Other Europe
DNB	D-SIB	Norway	Nordic
Erste Bank	D-SIB	Austria	Other Europe
Goldman Sachs	G-SIB	US	North America
Handelsbanken	D-SIB	Sweden	Nordic
HSBC	G-SIB	UK	Other Europe
ING	G-SIB	the Netherlands	Other Europe
Intesa Sanpaolo	D-SIB	Italy	Other Europe
JP Morgan Chase	G-SIB	US	North America
KBC	D-SIB	Belgium	Other Europe
Lloyds	D-SIB	UK	Other Europe
Mitsubishi UFJ Financial Group, Inc.	G-SIB	Japan	Asia
Mizuho Financial Group, Inc.	D-SIB	Japan	Asia
Mizuho Financial Group, Inc.	G-SIB	Japan	Asia
Morgan Stanley	G-SIB	US	North America
National Bank of Canada	D-SIB	Canada	North America
Nomura Holdings, Inc.	D-SIB	Japan	Asia
Nordea*	G-SIB	Sweden/Finland	Nordic
Raiffeisen Bank	D-SIB	Austria	Other Europe
RBS	G-SIB	UK	Other Europe
Royal Bank of Canada	G-SIB	Canada	North America
Santander	G-SIB	Spain	Other Europe
SEB	D-SIB	Sweden	Nordic
Société Générale	G-SIB	France	Other Europe
Standard Chartered	G-SIB	UK	Other Europe
State Street	G-SIB	US	North America
Sumitomo Mitsui Financial Group, Inc.	G-SIB	Japan	Asia
Sumitomo Mitsui Trust Holdings, Inc.	D-SIB	Japan	Asia
Swedbank	D-SIB	Sweden	Nordic
The Norinchukin Bank	D-SIB	Japan	Asia
Toronto-Dominion Bank	G-SIB	Canada	North America
UBS	G-SIB	Switzerland	Other Europe
Unicredit	G-SIB	Italy	Other Europe

Note. * In October 2018 the Nordea bank moved its headquarters from Sweden to Finland, becoming in 2019 the largest D-SIB bank in Finland.

Table B2. Summary statistics of key data used in the regression analysis

Variables	No. Obs.	mean	sd	p1	p25	p50	p75	p99
CDS spreads	9,772	95.85	106.33	6.61	33.62	69.62	117.94	542.45
TBTF	9,772	9.51	6.30	0.68	4.61	8.02	13.36	27.81
Total Assets	10,516	910,056	767,208	34,202	292,036	672,866	1,423,731	2,922,798
TCR	9,898	15.27	3.79	9.29	12.3	15	17.30	25.96
ROE	10,486	13.33	19.39	−38.95	8.41	13.62	19.48	36.61
ProbCr	10,764	26.66	28.94	0.01	1.75	14.75	47.68	96.93
IntRate	10,557	1.30	1.64	−0.79	0.09	0.80	2.39	5.12
r*	10,764	1.13	0.73	−0.09	0.50	0.94	1.72	2.61
SovDebt	10,764	91.81	44.79	30.66	59.38	85.71	111.53	215.97
VIX	10,764	18.92	8.76	10.26	13.44	16.10	21.84	57.74

Note. CDS spreads are the quoted CDS spreads in basis points, TBTF is the too-big-to-fail premium (%), Total Assets are bank total assets (USD million), TCR is total capital ratio (%), ROE denotes return-on-equity (%), ProbCr is the probability of financial crisis (%), IntRate is the 2-year government bond yield (%), r* is the natural interest rate, i.e. the real interest rate expected to prevail when an economy is at full strength and inflation is stable (%), SovDebt denotes the sovereign debt-to-GDP ratio (%), VIX is the Chicago Board Options Exchange Volatility Index (%-points). sd denotes standard deviation, p25 the 25-percentile, etc.

Appendix C

We show the coefficients on size – as measured by total assets (in log terms), bank profitability as measured by the return on equity – and bank capital as measured by total capital ratio, by year. A higher coefficient is associated with a higher TBTF premium.

Table C1. Coefficients by year

	TA (log)	TCR	ROE
2004	−1.05*** (0.01)	0.89*** (0.00)	−0.02 (0.12)
2005	0.10 (0.69)	1.18*** (0.00)	−0.04*** (0.00)
2006	0.43** (0.04)	0.89*** (0.00)	−0.09*** (0.00)
2007	0.94*** (0.00)	0.87*** (0.00)	−0.06*** (0.00)
2008	1.05*** (0.00)	0.51*** (0.00)	−0.03*** (0.00)
2009	1.92*** (0.00)	0.04 (0.77)	−0.06*** (0.00)
2010	1.74*** (0.00)	−0.19*** (0.00)	−0.03 (0.32)
2011	1.91*** (0.00)	−0.14*** (0.00)	−0.13*** (0.00)
2012	1.62*** (0.00)	−0.14*** (0.00)	−0.10*** (0.00)
2013	0.55** (0.03)	−0.33*** (0.00)	−0.09*** (0.00)
2014	1.12*** (0.00)	−0.09*** (0.00)	−0.10*** (0.00)
2015	0.57** (0.01)	−0.22*** (0.00)	−0.14*** (0.00)
2016	0.41* (0.09)	−0.32*** (0.00)	−0.23*** (0.00)
2017	0.12 (0.59)	−0.48*** (0.00)	−0.15*** (0.00)
2018	0.00 (0.99)	−0.33*** (0.00)	−0.21*** (0.00)
2019	−0.12 (0.59)	−0.16*** (0.00)	−0.20*** (0.00)
2020	0.07 (0.78)	−0.26*** (0.00)	−0.21*** (0.00)

Note. TA (log) denotes bank total assets in log terms, ROE is returns on equity and TCR refers to total capital ratio. We present robust p-values in parenthesis, *** p<0.01, ** p<0.05, * p <0.1.

The development of the Swedish real exchange rate over a longer perspective

Carl-Johan Belfrage*

The author works in the Riksbank's Monetary Policy Department

The real exchange rate of the Swedish krona has shown a weakening trend since the 1970s in a way that is unique among advanced economies in terms of scope and duration. However, other measures of the real exchange rate indicate that the krona has not depreciated quite as much as the usual measure indicates, among other reasons due to differences in how price levels are calculated in different countries. The real exchange rate movement can be explained, at least partly, by the relatively weak development of Swedish productivity and the trend decline in Sweden's terms of trade. The article places the development of the real krona exchange rate in a longer historical perspective, describes how alternative measures modulate the picture, discusses possible explanations and presents a model estimate of the trend and what has driven it.

1 A long-term trend towards a weaker real krona exchange rate

When the Swedish krona exchange rate depreciated by up to 20 per cent over the years 2014–2019, it caused an intensive debate.[1] In this, it was frequently pointed out that a Swedish krona bought fewer and fewer goods and services abroad compared with in Sweden. In other words, we experienced a weakening of the real krona exchange rate. The real exchange rate is usually expressed as

$$\text{real exchange rate} = \frac{\text{nominal exchange rate} \times \text{foreign price level}}{\text{domestic price level}}$$

The real exchange rate for the krona thus compares the price level abroad, translated into kronor, with the price level in Sweden. Using the definition above, a higher value for the real exchange rate corresponds to a lower price level in Sweden compared to abroad, which is to say a weaker real exchange rate. The weakening of the real krona exchange rate that I mentioned above thus means that the price level in Sweden has risen more slowly than it has abroad, expressed in the same currency.

One important detail in this context is what is more specifically meant by price level. This is because there are a number of conceivable alternatives. Sometimes price levels are measured using narrow baskets of goods to make the calculations easy to understand.[2] However, gaining a more reliable picture requires broader indices of the prices of goods and services. In general, therefore, the consumer price index is used to calculate real exchange rates. In this article, I will use consumer price index-based measures of the real exchange rate

* I would like to thank Hanna Armelius, Vesna Corbo, Paola Di Casola, Jesper Hansson, Marianne Nessén, Ulf Söderström and Magnus Åhl for valuable comments. The opinions expressed here are those of the author and are not necessarily shared by the Riksbank.

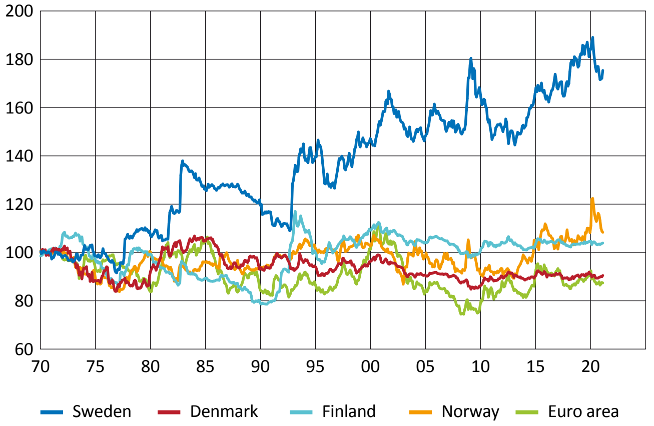
1 Measured in terms of the krona index (KIX).

2 The so-called Big Mac index is such a price level measure; see The Economist (2020).

for the discussion, but, to complement the picture and shed light on the causes underlying changes in the real exchange rate, I will also employ calculations using other price level measures.

In general, major exchange rate fluctuations awaken both curiosity and discussion. However, this is far from being a unique phenomenon; instead, it is something that most countries have been through. Even so, seen over longer time periods, real exchange rates largely tend to fluctuate around the same level. In statistical terms, they are described as stationary.[3] But Sweden's real exchange rate deviates clearly from this pattern. Figure 1 shows what are known as effective real exchange rates, which is to say the trade-weighted average of bilateral real exchange rates, for Sweden and our neighbours since 1970.[4] We can see that, unlike developments in neighbouring countries, the Swedish effective real exchange rate has shown a weakening trend since the 1970s. In a broader international perspective too, this is a unique development. With the exception of when the real exchange rate more than doubled for the Japanese yen between 1970 and 1995, none of the countries classified as an industrialised country in 1970 has since then been through a greater change in its currency's real exchange rate than Sweden has.[5]

Figure 1. Effective real exchange rates
Index 1970 = 100



Note. Monthly data. Calculated using the consumer price index as measure of price levels. A higher value corresponds to a weaker effective real exchange rate.
Source: BIS

In this article, I will investigate what may lie behind this remarkable development. In section 2, I explain what generally suggests that real exchange rates are stationary. In section 3, I calculate the effective real exchange rate of the krona even further back in time to see whether the development shown in Figure 1 forms a continuation of an even longer trend or whether it is perhaps an adjustment of relative price levels to an earlier appreciation. In section 4, I discuss the contribution made by various bilateral real exchange rates to this development. In section 5, I show how the choice of index to represent the

3 Ca' Zorzi et al. (2017) point out that the usual assumption in modern research into exchange rates is that real exchange rates tend to return to their mean values. This fact is being used to do what Meese and Rogoff (1983) found to be far too difficult (at any rate over the time frame in which their investigation was conducted, which is to say up to one year ahead), namely to make better forecasts for exchange rates than a random walk (which simply means assuming that the future exchange rate will be the same as today's exchange rate).

4 The aggregation into an effective real exchange rate for Sweden in this diagram, which comes from the BIS, differs slightly from later diagrams showing effective real exchange rates for Sweden, which use what are known as KIX weights. The reason for this is a desire here to show effective real exchange rates for several countries from the same source using the same principles for weighting.

5 This has been measured as the ratio between the highest and lowest monthly listings for the CPI-based effective real exchange rate over the period 1970 to 2020 according to the exchange rate indices published by the BIS and JP Morgan. According to the World Bank (1970), the industrialised countries were Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom and the United States. The appreciation of the yen was partly reversed after 1995 and the Swedish krona is the currency that has had the greatest absolute percentage change in its effective real exchange rate between 1970 and 2020.

price level affects the view of the development of the real exchange rate – could the trend depreciation possibly be an effect of how prices are measured in Sweden, as compared to abroad? In section 6, I explain how the trend development of the real exchange rate could be connected to the development of fundamental variables. In section 7, I present an empirical estimate of the development of the long-run real exchange rate since 1995, together with what this says about how the gap between the actual real exchange rate and its long-run level has developed. In section 8, I draw a few conclusions.

2 What is there to suggest that real exchange rates are stationary?

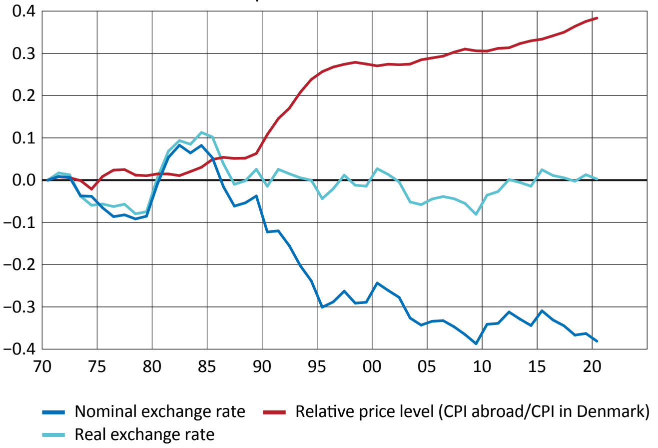
Before I go into more detail as to why the trend development of the krona's real effective exchange rate has been different from those of its more stationary equivalents in neighbouring countries, it may be helpful to remind ourselves of what there is to suggest that real exchange rates are stationary.

When analysing real exchange rates, the general assumption is that the purchasing power of a currency is the same in different countries, which is to say that absolute purchasing power parity prevails, or at least that the relationship between purchasing power in one country and purchasing power in another country is constant, which is to say that relative purchasing power parity prevails. One person who made this argument early on was the Swedish economist Gustav Cassel, who coined the actual expression purchasing power parity over a hundred years ago (Cassel 1918, p. 413). Cassel argued that, if the price level in one country were to increase more than in another, trade between the countries would lead to the differences being evened out over time, either by the actual prices changing or by the nominal exchange rate being adjusted. Somewhat simplified, we can imagine that if goods in Denmark, for example, are cheaper, then it will be worth exporting them to other countries. Eventually, however, the higher demand for Danish goods should lead either to higher prices in Denmark too or to the Danish krone rising in value.

Figure 2 shows the development of prices and the exchange rate in Denmark since 1970. From this, we can see that the red line that shows the ratio between world prices and Danish prices rises until the present day, meaning that Danish prices have risen more slowly than prices abroad. At the same time, however, the dark blue line shows that the nominal exchange rate of the Danish krone has appreciated so much that the real exchange rate has become relatively stable. The Danish real exchange rate thus seems to be stationary.

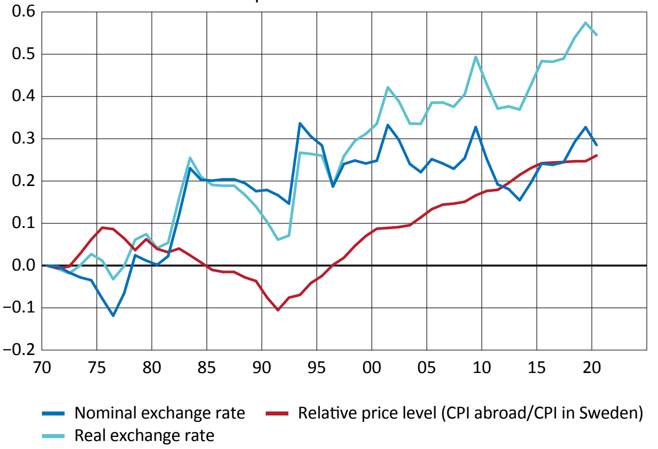
Developments in Sweden since 1970 look different, however. Figure 3 shows that, between 1970 and 1996, the development of prices in Sweden was approximately the same as abroad in that the red line is relatively stable until 1996. However, at the same time, we see that the nominal exchange rate depreciated in that the dark blue line rises. The light blue line thereby rises over this period and shows a weakening of the krona's real exchange rate. It has also continued to depreciate after 1996, but then for a different reason. While the nominal exchange rate has varied around one and the same level, the Swedish price level has risen more slowly than it has abroad, which is to say that the rate of inflation has been lower in Sweden than abroad. Seen over the entire period since 1970, an initial nominal exchange rate depreciation followed by relatively weak domestic price growth have thus contributed to a weakening trend for the real exchange rate of the krona.

Figure 2. Denmark's real and nominal exchange rates, and the country's domestic price level in relation to the world level
Exchange rates and relative domestic price level expressed as log deviations from their respective levels in 1970



Note. Annual data. The real exchange rate development for Denmark differs slightly from that in Figure 1, as it is based on different sources. The difference may be due to the sample and weighting of countries in the calculation of effective exchange rates.
Sources: JP Morgan and own calculations

Figure 3. Sweden's real and nominal exchange rate and the country's domestic price level in relation to the world level
Exchange rates and relative domestic price level expressed as log deviations from their respective levels in 1970



Note. Annual data.
Sources: BIS, Macrobond and own calculations

3 A hundred-year perspective over the real exchange rate of the krona

Despite the trend depreciation of the krona's real exchange rate in Figures 1 and 3, a tendency towards purchasing power parity cannot be ruled out if developments are viewed over an even longer period. This is because the empirical research literature shows that purchasing power parity only holds up over long periods of time.[6] There are many possible explanations as to why it should be this way. When it takes time for production and prices to adjust to new conditions, changes in real exchange rates can be long-lasting, even if they are

6 See Taylor and Taylor (2004) for a review of empirical studies of purchasing power parity.

transitory. In addition, there are factors that may give rise to lasting tends in real exchange rates. Above all, these concern differences in productivity growth and trends in world market prices for the country's exports and imports respectively, factors examined in more depth in section 6 below. However, even such factors can be neutralised in the very long term through ideas and technology, like capital and labour, moving across borders.[7]

However, investigations of a possible tendency towards purchasing power parity over long time periods come up against the problem that there is a lack of published real effective exchange rate series stretching further back in time. For most countries in the Riksbank's official exchange rate index (the krona index or KIX), however, there is access to both consumer price indices and nominal exchange rates against the US dollar from the 1950s or even earlier for some countries. This makes it possible to calculate long time series for bilateral real exchange rates.[8] The trade-based weights included in the calculation of KIX or the similar index I showed in Figure 1 are harder to get at, however. The KIX weights vary over time but have not been calculated for the period prior to 1994. The bilateral real exchange rates, aggregated using 1994's KIX weights, may, however, give an indication of how the krona's real effective exchange rate developed further back in time.[9]

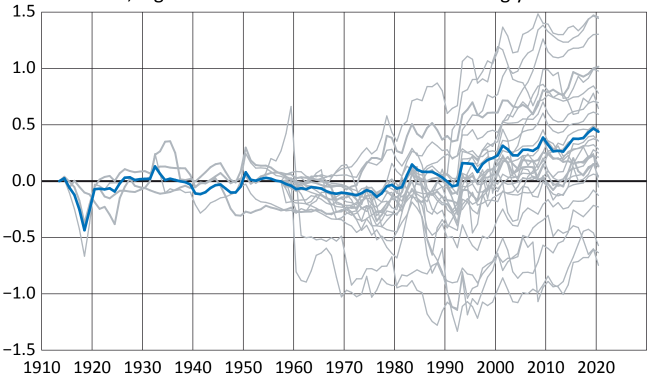
Figure 4 shows the results of these calculations. Here, we have both a number of bilateral real exchange rates (some going all the way back to 1913) and an aggregate produced using KIX weights from 1994, shown by the blue line. For the first decades, the aggregated, i.e. effective, real exchange rate only includes bilateral real exchange rates against the United States, United Kingdom, Norway and Switzerland. However, from 1957, the aggregate includes 97 per cent or more of the currencies included in the current KIX index.

7 Froot and Rogoff (1995, p. 1674) express it like this: 'It is arguable whether one should expect to detect a Balassa-Samuelson effect in really long-run data. Even though technology can differ across countries for extended periods, the free flow of ideas together with human and physical capital produces a tendency towards long-run convergence of incomes.' Another approach, which also leads to the conclusion that real exchange rates should be stationary over the long perspective, is based on variations in the price indices being dominated by monetary factors that, in turn, are neutral in the long term in the sense that they do not affect relative prices between goods or between goods and foreign currency; see Dornbusch (1985).

8 KIX aggregates the bilateral exchange rates for the 32 countries that are most relevant to Sweden's foreign trade. The weights are so-called trade weights and are updated annually. See Alsterlind (2006) for a discussion of fundamental issues around the construction of effective exchange rate indices and Erlandsson and Markowski (2006) on the theory and practice behind the construction of KIX.

9 Using fixed instead of time-varying weights is not unique in itself for effective exchange rate indices. The most commonly used Swedish effective exchange rate index was, until a few years ago, the so-called TCW index, which uses weights set at the start of the 1990s. The Riksbank now publishes both the TCW index and the KIX index on its website.

Figure 4. Sweden's effective real exchange rate and its component bilateral real exchange rates since 1913
CPI-based, log deviations from the level of the starting year of 1913



Note. Annual data. The blue effective real exchange rate has been calculated as the KIX-weighted sum of the annual log changes in the bilateral real exchange rates at any given time, shown by grey curves. Up to and including 1994, KIX weights from 1994 have been used. In the figure, the bilateral rates have been indexed at the level of the KIX-weighted rate at their respective starting points. For the blue curve, the scale means that a level of 0.1 corresponds to a real exchange rate that is approximately 10 per cent weaker than at the starting point of 1913, but this approximative translation to percentage deviations becomes worse as the deviation becomes larger.
Sources: BIS, Macrobond, Sveriges Riksbank and own calculations

Just like Figure 1, Figure 4 shows that the real exchange rate of the krona has weakened in recent decades. But what Figure 4 also shows is that, with a longer perspective, the krona's effective real exchange rate appears somewhat stationary until the start of the 1990s. When the real exchange rate weakened between the mid-1970s and the start of the 1990s, this could be seen as a return to the same level that the real exchange rate had fluctuated around over most of the 1900s. This was a period in which Sweden entered into various different fixed exchange rate systems. When prices then increased more rapidly in Sweden than in other countries, Swedish competitiveness worsened, which is to say that the real exchange rate appreciated. At the end of the 1970s and start of the 1980s, Sweden therefore carried out several devaluations to restore competitiveness.[10] The real exchange rate then appreciated again and, in 1992, the year in which a floating exchange rate was adopted, it was at the same level that it had fluctuated around over most of the 1900s. However, the krona has subsequently depreciated in real terms against all currencies included in the KIX index.

4 A broad depreciation but mostly against rapidly growing countries' currencies

One way of examining in more detail how an effective real exchange rate has developed is to study the bilateral real exchange rates it consists of. This makes it possible to see if the development is being driven by certain countries and, if so, what these countries have in common.

The krona's real exchange rate appreciated against most currencies between the end of the 1950s and the start of the 1990s. This can be anticipated in Figure 4, where most of the grey lines are at or above zero at the end of the 1950s but below zero at the start of the 1990s.[11] However, since 1992, the krona has depreciated clearly against all KIX currencies with available data. This can be seen by the grey lines finishing at a higher level than they

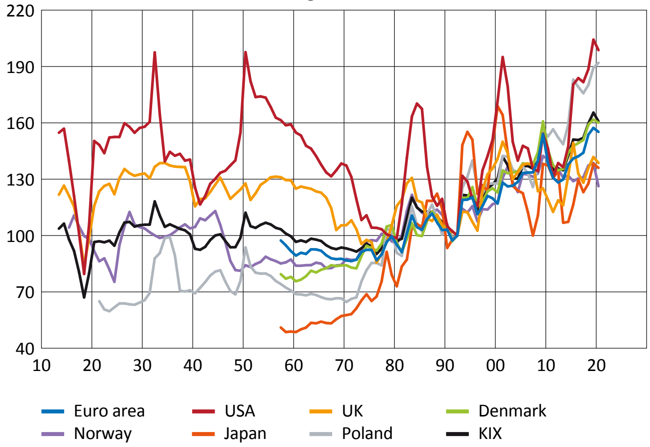
10 See Bordo et al. (2017) for a historical overview of economic policy regimes that have tended to affect the global development of exchange rates since 1880 and Sveriges Riksbank (2000) for a description of Sweden's participation in various fixed exchange rate systems.

11 In most cases, data that allows calculations of bilateral exchange rates is available from 1957 on, which, as the figure is constructed, means that the data started on the blue KIX line's level in 1957 when this was very close to zero.

had in 1992.[12] Figure 5 shows the bilateral real exchange rates that affected the KIX-weighted real exchange rate the most.[13] The krona's effective real exchange rate with the KIX countries and its most important bilateral component, the euro area, follow each other closely, while the pattern for other bilateral real exchange rates varies both among themselves and over time. The bilateral real exchange rate with the United States has fluctuated heavily. A gradual appreciation between 1950 and 1980 was followed by several decades with large changes in value. However, since 1980, the krona's real exchange rate against the US dollar has depreciated and the current level is approximately the same as it was 70 years ago.

Figure 5. Bilateral real exchange rates against the currencies that have contributed most to the depreciation of the KIX-weighted real exchange rate

CPI-based bilateral real exchange rates, index 1992 = 100



Note. Annual data. The bilateral real exchange rates shown are those that have made a contribution of more than 2 percentage points to the depreciation of the KIX-weighted index since 1957, 1970 or 1992. Since 1994, KIX weights from 1994 have been used.

Sources: BIS, Macrobond, Sveriges Riksbank and own calculations

The picture is, to some extent, compatible with the Balassa-Samuelson hypothesis – that the real exchange rate can be expected to appreciate in countries with strong productivity growth against the rest of the world. The hypothesis is based on the assumption that productivity primarily changes in the production of goods and services that are traded internationally and whose prices will develop in the same way in different countries precisely because they are traded internationally. When productivity rises in this part of the economy, wages rise. In turn, the higher wages drive up prices for the goods and services that are not traded internationally, so-called domestic market goods. Prices for domestic market goods (and therefore also the price level in a common currency as a whole) thereby rise faster in countries with stronger productivity growth. I write more about this in section 6.

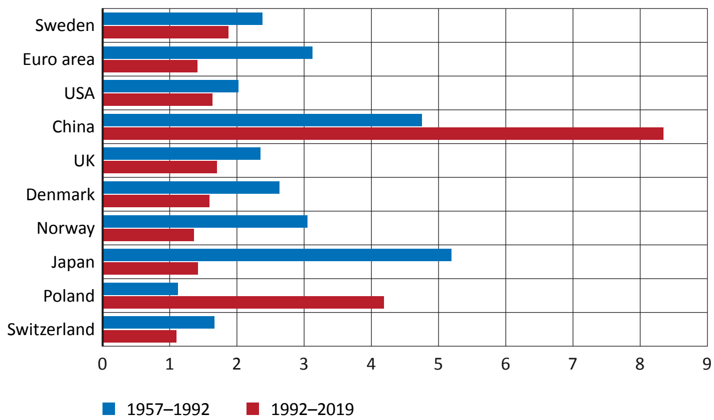
With a few isolated exceptions, over the period 1957–1992, the Balassa-Samuelson hypothesis works well to explain the development of the bilateral real exchange rates in Figure 5. The real exchange rate depreciated heavily against Japan, which then had significantly faster productivity growth than Sweden. The difference is shown by the blue bars in Figure 6. The real exchange rate also weakened against the euro area, Denmark and Norway, which all had higher productivity growth, at the same time as it strengthened against the United States, whose productivity growth was weaker than Sweden's. The exception in this period is that the real exchange rate weakened against Poland, even though productivity growth there was significantly weaker than in Sweden.

12 There is no data for Slovakia prior to 1994 and none for Russia prior to 2001.

13 In this figure, China and Switzerland have been excluded, however, as the differences between the highest and lowest listings in their cases would require a scale that would make the figure hard to read. See the appendix for a version of this figure that also includes China and Switzerland.

Between 1992 and 2019, the Balassa-Samuelson hypothesis does not hold up as well as an explanation for how the bilateral real exchange rates that I presented in Figure 5 developed. The krona's real exchange rate weakened overall, even though Sweden had higher productivity growth than all these countries and regions apart from Poland and China. Within the group advanced economies, the real exchange rate depreciation has certainly been least against Norway and Japan, which have had the weakest productivity growth over this period, but, at the same time, the depreciation has been relatively strong against Switzerland, which had the weakest productivity growth over this period.

Figure 6. Productivity growth in the economies whose currencies have contributed most to the depreciation of the KIX-weighted real exchange rate
Average annual percentage change in GDP per inhabitant aged 15–64 years

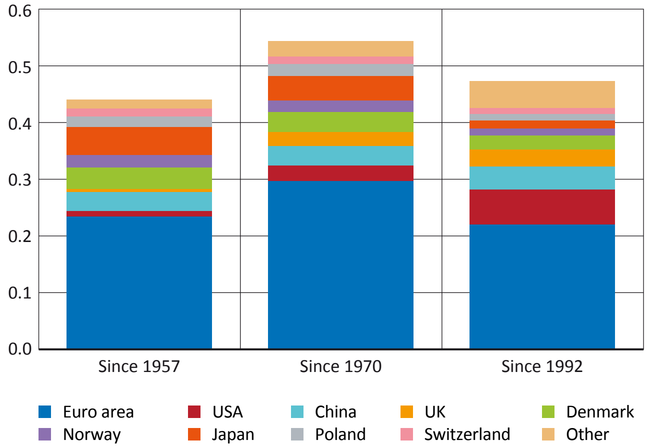


Note. Due to a lack of data, the change in the entire population has been used when calculating productivity growth in the first three years. For the euro area, the blue bar represents the period 1960–1992.
Sources: OECD, Penn World Table and World Bank

The fact that the krona has depreciated in real terms against as good as all other currencies also means, of course, that no single bilateral relationship lies behind the weakening of the effective real exchange rate. This is also illustrated by Figure 7, which shows contributions to the weakening of the KIX-weighted real exchange rate since 1957, 1970 and 1992. The euro is responsible for the single greatest contribution but this is primarily due to the euro's great weight in the KIX index, which is around 50 per cent. The fact is that the bilateral real exchange rate against the euro does not deviate significantly from the KIX rate (see Figure 5). The relatively heavy depreciation against the US dollar in recent years has also clearly contributed to the weakening of the KIX-weighted index, not least since 1992. Alongside the currencies of these two major western economies, the Chinese yuan has made the largest contribution to the weakening of the KIX-weighted real exchange rate since 1992.[14] Alongside the yuan and the euro, the currencies of our Scandinavian neighbours and the Japanese yen have made the greatest contribution to the weakening of the krona's real exchange rate since 1957.

14 The bilateral real exchange rate against China is included in Figure A1 in the appendix. This also makes clear that the bilateral real exchange rate against China appreciated heavily in the years prior to 1992, at least according to available data on nominal exchange rates and inflation. However, this did not significantly affect the KIX-weighted real exchange rate, as China's weight in KIX was very small then, one half of one per cent.

Figure 7. Contribution to the weakening of the KIX-weighted real exchange rate
Contribution to log deviation in KIX-weighted CPI-based real exchange rate



Note. The bilateral real exchange rates mentioned are those that have made a contribution of more than 2 percentage points to the depreciation of the KIX-weighted index since 1957, 1970 or 1992. Each country's contribution has been calculated as the sum of its KIX-weighted log changes over the period for which data has been available for the country in question. Up to and including 1994, KIX weights from 1994 have been used.
Sources: BIS, Macrobond, Sveriges Riksbank and own calculations

To sum up, the krona's effective real exchange rate seems to have been stationary over the greater part of the 1900s. A weakening trend started in the 1970s, the first part of which, up until 1992, can be seen as a return to the level around which this exchange rate fluctuated since the start of the 1900s. Regardless of whether we start from 1970 or 1992, the real exchange rate of the krona has developed in a way that has been unique among industrial countries. A depreciation has taken place against all trading partners of significance, albeit to varying degrees.

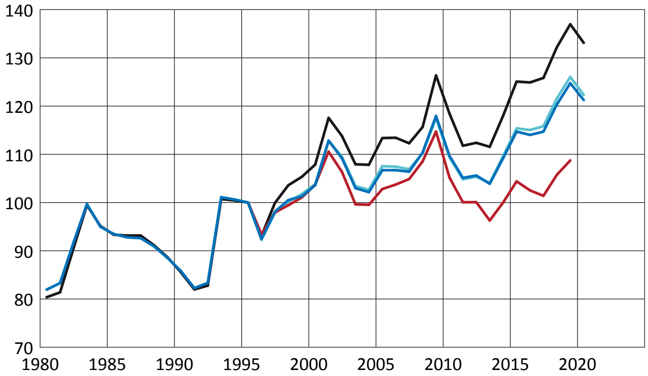
5 Differences in CPI content and price measurement methods may give an exaggerated view of changes in the real exchange rate

So far, my analysis of the real krona exchange rate has focused on measures in which domestic price levels are represented by the consumer price index (CPI). But the choice of price index plays a part. Using other measures of price levels gives us a slightly less dramatic image of the development of the real exchange rate. Among other reasons, this is because different price indices are constructed in different ways. One example is the CPI, where one known difference between Sweden and other countries is that mortgage interest costs are included in the calculation of CPI for Sweden as they are treated as part of households' housing costs here. The trend decrease in interest rates in recent decades has had a clear impact on CPI inflation. Other price indices do not include these effects of interest rate adjustments, for example the European harmonised index of consumer prices (HICP) or the Swedish CPIF, which is the CPI with a fixed interest rate.[15] If the CPI is exchanged for the HICP or CPIF, the measured real depreciation of the krona since 1992 becomes about 15 percentage points less; see Figure 8. This reflects the considerable fall in interest rates since

15 Statistics Sweden (2017) describes the main differences between these three consumer price indices.

the mid-1990s.[16] It is this KIX-weighted effective real exchange rate for the krona, calculated using the CPI for other countries and the CPIF for Sweden, that the Riksbank usually uses and sometimes shows in its Monetary Policy Reports. Hereinafter, it will be referred to as the 'real KIX'.

Figure 8. KIX-weighted real exchange rate with different measures of the consumer price level
Index 1995 = 100



CPI abroad, CPIF in Sweden (Real KIX)
PPP-based real exchange rate
CPI abroad, HICP in Sweden
CPI abroad and in Sweden

Note. Annual data. The difference between the three upper curves is exclusively due to different choices of price index for Sweden, which is to say the denominator in equation (1). The black curve corresponds largely with the curve for Sweden in Figure 1 but differs slightly due to differences in the selection of trading partner and their weights. Up to and including 1994, KIX weights from 1994 have been used.
Sources: BIS, National Institute of Economic Research, Macrobond, Statistics Sweden, Sveriges Riksbank and own calculations

Figure 8 also shows an effective real exchange rate for the krona, constructed using statistics on price levels from the international comparison programme used to calculate purchasing power parity adjusted (PPP-adjusted) GDP levels.[17] This statistic is based on comparisons of prices for identical or very similar products in all countries. However, the comparisons are based on a significantly smaller sample of products than normal price statistics and price data is collected less frequently. Short-term fluctuations in price levels are thereby not captured so well but, seen over a number of years, the international comparisons should give a picture of the development of the real exchange rate that can be compared with other measures of the real exchange rate.[18]

As the short-term variations of the real exchange rate, regardless of measure, are dominated by variations in the nominal exchange rate, the variations in the effective real exchange rate for the krona that can be extracted from the PPP statistics are clearly reminiscent of the variations we see in the CPI/CPIF-based real exchange rate. On the other hand, the PPP-based real exchange rate does not demonstrate the same clear depreciation trend. This is primarily due to developments between 2008 and 2019. During this period, the real KIX weakened by 13 per cent, while the PPP-based real exchange rate was unchanged. This pattern for the difference between CPI/CPIF-based and PPP-based real exchange rates can also be found in the most important bilateral relationships, with the exception of Norway; see Figure 9. The difference is particularly clear for how the krona's bilateral real

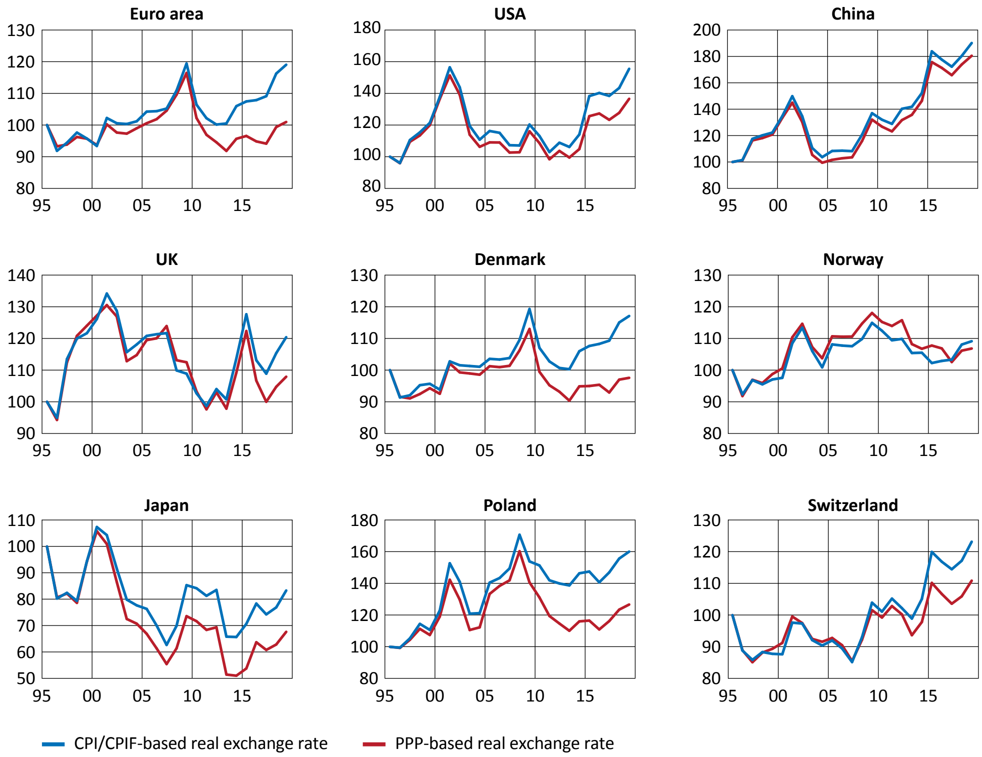
16 The development in the figure is based on data available from 1980. Even before that, trend changes in the level of interest rates have taken place that may have affected how measured inflation in Sweden related to the rest of the world. However, these historical trend interest rate fluctuations appear to be smaller than the one taking place since the mid-1990s.

17 PPP stands for purchasing power parity. See EU and OECD (2012) for details of how this statistic is calculated.

18 The comparison does not go back further in time than 1995 because published PPP data from before that point are backdates based on the consumption deflators of the countries included (see OECD, 2016).

exchange rate has developed against the euro. It weakened by 19 per cent between 2008 and 2019 according to the CPI/CPIF-based measure but only by 1 per cent according to the PPP-based measure.

Figure 9. Bilateral real exchange rates – PPP vs. CPI/CPIF
Bilateral real exchange rates based on CPI/CPIF and comparative price level indices, index 1995 = 100



Note. Annual data. CPI/CPIF-based real exchange rate means that the calculation of the real exchange rate has used the CPI for other countries and the CPIF for Sweden. The PPP-based real exchange rates are the comparative price level indices calculated as the ratio of the PPP exchange rate and the actual nominal exchange rate.
Sources: BIS, Eurostat, Macrobond, OECD, Statistics Sweden and own calculations

One possible contributing explanation for the difference between these measures of the real exchange rate may be differences between how statistical authorities make quality adjustments when producing various price indices, which does not have to be done at all when producing PPP statistics. For example, Tysklind (2020) shows that price growth measured for quality-adjusted product groups differs significantly between different countries in Europe, of which the measured price growth in Sweden is one of the slowest. This is despite these largely being products that are similar and can easily be traded between countries. This indicates that the prices for these goods are adjusted downwards for quality improvements more in Sweden than in other countries.

As I mentioned above, the PPP-based real exchange rate measure does not risk being disrupted by differences in quality adjustment. At the same time, there may be a tendency for prices included in the PPP statistics to comply with the law of one price to a greater extent than consumer prices in general.[19] The desire for comparability in the products included may push the sample towards those products that are easiest to compare from country to country. Consequently, their prices can be expected to be smoothed out by

19 The law of one price means that identical goods are sold for the same price on all markets expressed in a common currency.

trade from country to country to a greater extent than other products.[20] This could be an explanation for why the PPP-based measure conveys a less clear trend in the real krona exchange rate.

The differences between the various measures of the real exchange rate that we have compared so far suggest that it would be appropriate to complement the picture with other measures. One way of proceeding would be to use costs instead of prices. We can do this by calculating the real exchange rate based on unit labour cost (ULC).[21] This uses ULC for the entire economy, or some sector such as the manufacturing sector, instead of the consumer price index. Instead of being a measure of how many baskets of Swedish goods and services Swedish consumers would have to give up to purchase a similar basket of goods and services abroad, the ULC-based real exchange rate indicates how many units of Swedish labour would be needed to produce what would be required to buy what a unit of labour produces abroad. Use of the ULC-based method avoids variations stemming from changes in price mark-ups as well as the difficulties in comparing data to which consumer price indices can give rise. At the same time, the unit labour cost captures the labour cost and not the total cost of a produced unit, which could affect the picture of the relative price level development if the labour cost's share of the total cost were to develop differently in Sweden and abroad. Another difference from the consumer price index is that ULC refers to what is produced and not what is consumed in each country. Consequently, it does not include costs for producing whatever is imported but does include costs for producing export goods, investment goods and public services, which is to say goods and services that are not included in the consumer price index.

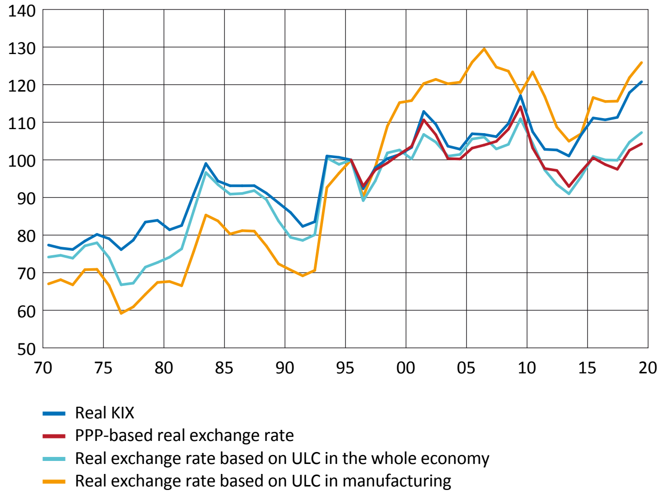
Available statistics make it possible to create ULC-based real exchange rates from 1970 on. Figure 10 shows that, despite the not inconsiderable differences between what is included in the measures, the development of the real krona exchange rate based on ULC for the entire economy is similar to the development of the real KIX until 2006. On the other hand, there arises a deviation after the global financial crisis when real KIX develops considerably more weakly.[22] The development of the ULC-based measure is then more consistent with the PPP-based measure. The figure also shows the real exchange rate based on ULC in the manufacturing sector, which shows a much more powerful depreciation than the other measures of the real exchange rate between 1992 and 2006. This measure only captures costs associated with the production of goods that are traded internationally to a very great extent. The depreciation therefore probably reflects the deterioration of Sweden's terms of trade over the same period – a development that I will discuss in more detail in the next section.

20 The manual for data collection for PPP statistics explicitly specifies how the selection of products is to be carried out and that it is desirable to select products that are available for price determination in many countries; see EU and OECD (2012). Ravaillon (2018) finds indirect support for the hypothesis of an implicit preference for internationally comparable traded goods in the international price comparison programme. In addition, it can be demonstrated that, even if all prices are measured perfectly, the development of relative CPI and relative PPP will differ due to relative price changes as long as the consumption baskets differ among the countries compared (see Deaton and Aten, 2017).

21 Comparisons of ULC in common currency are also made with the aim of studying competitiveness; see, for example, Sveriges Riksbank (2019).

22 It is possible that the differences in the development of the different measures can, to some extent, be explained by differences in the development of the labour share in Sweden and abroad. According to the Penn World Table, the share of labour cost in GDP fell more in Sweden than abroad between 1977 and 1998, which coincides well with a period in which the krona's ULC-based real exchange rate depreciated more than the CPI/CPIF-based one. After that, the labour share again increased somewhat in Sweden, at the same time as it fell or remained unchanged in the most important KIX countries. The ULC-based real exchange rate then showed less weak development than the CPI/CPIF-based one.

Figure 10. KIX-weighted real exchange rate with different measures of the price level
Index 1995 = 100



Note. Annual data. Up to and including 1994, KIX weights from 1994 have been used in the aggregation. For real KIX, the CPIF has been used as price index for Sweden from 1980 on, and the CPI before this. Data access is more limited for ULC in the manufacturing sector than for other measures. To promote comparability between the series, the index for other series has therefore only been calculated with the aid of the observations (for countries and dates) that are also available for ULC in the manufacturing sector. One implication of this is that data for the largest emerging market economies (Brazil, India, China and Russia) is completely absent and data for certain other countries only affects the index over part of the period. This makes the weakening of the real KIX and the PPP-based real exchange rates somewhat smaller than in the other figures.
Sources: BIS, European Commission (DG ECFIN AMECO), National Institute of Economic Research, Macrobond, Statistics Sweden, Sveriges Riksbank and own calculations

6 Terms of trade and relative productivity are key variables for the real krona exchange rate

As was mentioned above, normally, a trend weakening of a country's real exchange rate is traditionally explained by the country having weaker productivity growth than other countries.[23] In addition, it is often pointed out that worsened terms of trade can explain a weakening of the real exchange rate.[24] The basic idea behind this is that international trade evens out prices of goods and services that can be traded internationally. The more a country can produce with the available resources (its productivity) or obtain through trade (its terms of trade), the higher the wages in that country can be expected to be. In turn, these determine prices for domestic market goods, which is to say the only prices that can differ more permanently from country to country, and thereby the strength of the country's real exchange rate.

6.1 The trend of relative productivity has changed

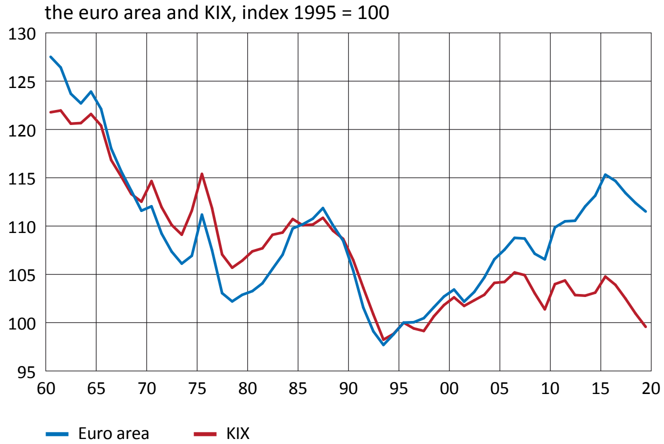
Relative productivity growth has had a prominent role in the empirical research literature that aims, with the Balassa-Samuelson hypothesis as a starting point, to find explanations for lasting changes to real exchange rates. In this context, the way productivity is defined and measured is crucial. The Balassa-Samuelson hypothesis basically suggests that the real exchange rate is determined by how productivity in internationally-traded production relative to productivity in the production of domestic market goods can be compared to

23 See Balassa (1964) and Samuelson (1964).
24 See, for example, Neary (1988), Chen and Rogoff (2003), and Berka et al. (2018).

the same relationship in other countries. This means that, if we wish to know to which extent productivity growth can explain how a real exchange rate has developed, we need to have access to comparable sector-specific productivity data for the home country as well as for all countries included in the index calculation. This is a high demand. Nevertheless, a number of attempts of this kind have been made and, in most cases, they provide support for the hypothesis.[25] However, if we are interested, as we are in our case, in explaining the development of an effective real exchange rate that includes emerging market economies and, furthermore, wish to be able to illustrate developments over a longer period, there are limited possibilities for obtaining the necessary data. In practice, one must instead rely on one measure of productivity for the entire economy in the hope that productivity growth mainly takes place within internationally-traded production, which is an assumption used and also empirically supported already by Balassa (1964).

Figure 11 uses GDP per inhabitant aged 15–64 to measure productivity. This lets us create a KIX-weighted index starting in 1960.[26] In the figure, we can see that productivity measured in this way increased more slowly in Sweden than abroad in the years between 1960 and 1993. This can be expected to have contributed to weakening the real krona exchange rate during this period. After this, there followed a period of relatively strong Swedish productivity growth until 2006, which has, however, fallen again in recent years. Relative to the euro area, Sweden's productivity growth has followed a similar pattern but it weakened slightly more until 1993, followed by a slightly larger strengthening afterwards. One conclusion of this is that it is possible to explain, to a certain extent, the depreciation of the real krona exchange rate until about 1993 through productivity growth, but not afterwards.

Figure 11. Sweden's productivity relative to other countries
Sweden's GDP per inhabitant between the ages of 15–64, relative to the euro area and KIX, index 1995 = 100



Note. Annual data.
Sources: OECD, Penn World Table, World Bank

6.2 Worsened terms of trade may have contributed to the depreciation

Terms of trade have also frequently been included in empirical studies that have attempted to explain real exchange rate movements. More favourable terms of trade are not just equivalent to higher productivity in the production of internationally-traded goods and services (as this means that one and the same amount of work in the production of export

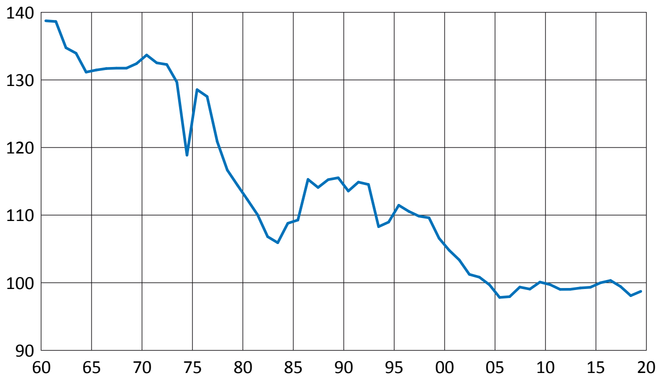
25 See Gubler and Sax (2019) for an overview.

26 A smaller investigation of how Sweden's productivity in internationally-traded production in relation to productivity in production of domestic market goods compares to the corresponding relative productivity in the euro area over the period 1995–2018 points to a development closely resembling that for relative GDP per inhabitant aged 15–64.

goods allows a larger volume of imports), which may be expected to affect the real exchange rate positively by pushing up prices for domestic market goods. They may, in addition, have a direct effect on relative consumer price levels and thereby the real exchange rate if there is a so-called home bias in consumption.[27]

In Figure 12, we can see a negative trend in the Swedish terms of trade, measured as the relationship between export prices and import prices in the national accounts, until about 2005. This can be presumed to have contributed to the trend depreciation of the real krona exchange rate over the same period. Among the product groups playing an important role within Swedish international trade, price growth for petroleum products, as well as pulp and paper, seem to have contributed to worsened terms of trade.[28] After 2005, it becomes more difficult to explain a depreciation of the real exchange rate on the basis of the development of the terms of trade.

Figure 12. Sweden's terms of trade
Index 2015 = 100



Note. Annual data.
Sources: European Commission, DG ECFIN, AMECO

6.3 Other possible factors behind the development of the real exchange rate

Alongside productivity in relation to other countries and the terms of trade, there are a number of other factors that are mentioned in the research literature and that could contribute towards explaining the depreciation trend in the real krona exchange rate. These can be divided into factors that, like productivity in relation to other countries, affect how rich a country is in relation to other countries, and into factors that more directly affect the relationship between prices for domestic market goods and internationally-traded goods.

The net external position, which is to say the difference between claims on and liabilities to other countries, reflects a country's wealth in comparison to the rest of the world and has therefore been highlighted as potentially important for a country's real exchange rate.[29] The net external position depends partly on the historical development of the current account and partly on changes in the market values of the securities that make up the country's assets and liabilities. Unfortunately, there are significant measurement problems here, which make it difficult to determine how this variable has developed for Sweden's part.[30] However, a dataset that is often used in this context and which was developed by Lane and Milesi-Ferretti (2018), the current level lies very close to the 1970 level. Even if Sweden were to have undergone a major change in its net external position, Christopoulos et al. (2012)

27 See Berka et al. (2018).
28 According to the export and import price indices published by Statistics Sweden.
29 See, for example, Lane and Milesi-Ferretti (2004).
30 See Blomberg and Östberg (1999) and Blomberg and Falk (2006) for discussions of measurement problems and difficulties in making accurate comparisons further back in time.

show that this should not affect the real krona exchange rate as there are no restrictions on Sweden's access to foreign capital.

Factors that more directly affect the relationship between prices on the domestic market and internationally-traded goods should also be able to affect the real exchange rate. In a comprehensive analysis of 48 countries' effective real exchange rates, Ricci et al. (2008) find that three such factors exert a certain influence on real exchange rates: public consumption since it can affect total demand in the direction of domestic market production, protectionism since it can drive up domestic prices above world market prices and price controls since they can keep prices below their market levels. However, a review of how these variables have developed in Sweden and among our most important trading partners indicates that the relative changes have been so minor that they could only have had a marginal influence on the trend development of the real krona exchange rate.

In summary, it seems that, for a number of decades, there have been trends in certain key variables, more specifically in Sweden's productivity relative to other countries and in the terms of trade, that could explain a trend depreciation of the real krona exchange rate, at least up until 2005. However, how much these variables may have contributed to the trend of the real exchange rate and what can explain the variations around this trend, remain to be determined. In the next section, I show how this can be done using an empirical model.

7 A model estimate and the development and current level of the long-run real exchange rate

Belfrage et al. (2020) describe an empirical model to explain the trend in the real krona exchange rate and how this relates to the central explanatory variables that I have presented above. This is a time series model with time-varying equilibrium levels. We interpret the long-run real exchange rate as the trend level obtained from the model and which we assume is a linear combination of estimated trends in the measures of relative productivity and terms of trade that I discussed in section 6.[31] In one and the same estimate, we can then use so-called Bayesian methods to calculate the long-run real exchange rate and the variations in the difference between the actual and long-run real exchange rate, known as the real exchange rate gap. In turn, the short-term variations in the real exchange rate gap are explained by the variations around the trends in relative productivity, the terms of trade and the current account, variations in the interest rate differential against the rest of the world (based on the hypothesis that the exchange rate is affected by how the return on investments differs from the rest of the world) and the VIX index (which reflects uncertainty over developments on the financial markets and thus demand for assets in currencies such as the krona that are considered less liquid), as well as a real exchange rate shock.

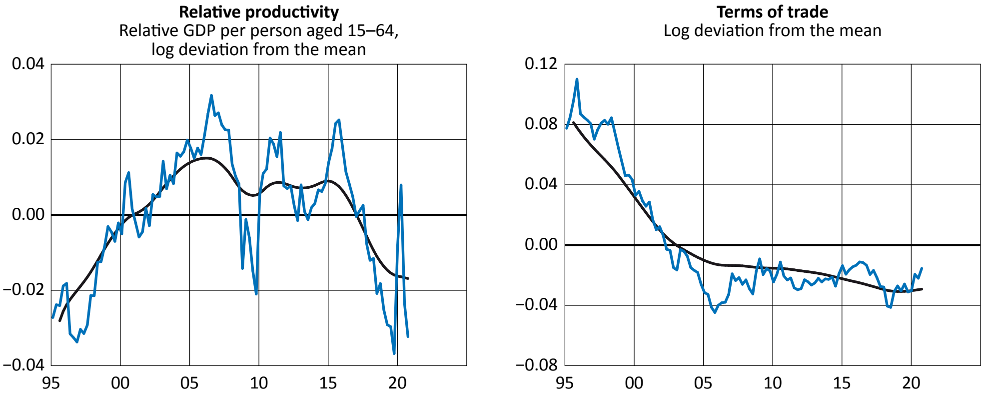
Figure 13 shows outcomes and estimated trend levels for Sweden's relative productivity and terms of trade, while Figure 14 shows the model's estimate of the long-run level of the real KIX since 1995.[32] We can see that, between 1995 and 2005, there is a trend increase in Sweden's productivity compared with the rest of the world, which itself suggests an appreciation of the long-run real exchange rate. At the same time, however, there is a powerful trend deterioration of the terms of trade, which itself suggests a depreciation of the long-run real exchange rate. In the model estimate, we can say that the effect of the worsened terms of trade dominates, as the estimated long-run real exchange rate depreciates over the period in which the trends of the explanatory variables move in different directions. After this, it is almost unchanged until 2015, after which a certain

31 See Lane and Milesi-Ferretti (2004) for a theoretical model that gives rise to such a link.

32 Limited access to more frequent data (here, quarterly data) and the fact that the dynamics of the real exchange rate also changed in conjunction with the transition from fixed rate to floating exchange rate at the end of 1992 explain why the model estimate restricts itself to the period from 1995.

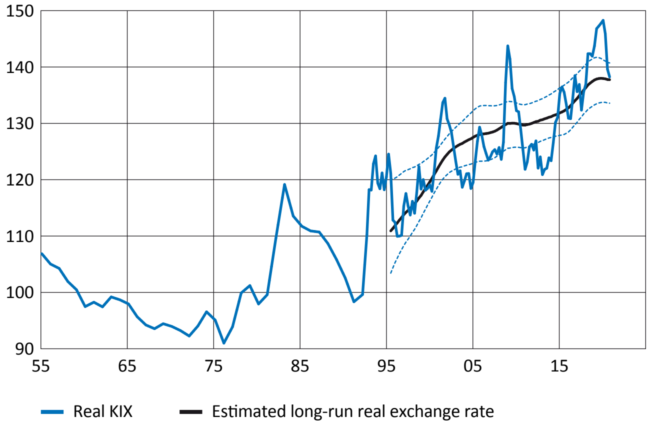
depreciation takes place as the trend levels fall for both relative productivity and the terms of trade.

Figure 13. Outcomes and estimated trend levels for central explanatory variables



Note. Quarterly data.
Sources: BIS, National Institute of Economic Research, Macrobond, national sources, Statistics Sweden, Sveriges Riksbank and own calculations

Figure 14. Long-run real KIX according to estimate using the TVE-VAR model
Index 18 Nov 1992 = 100



Note. Annual data until end of 1992, quarterly data from 1993 to 2020. Up to and including 1994, KIX weights from 1994 have been used in aggregation. The CPIF has been used as price index for Sweden from 1980 on, and the CPI before this. The dotted lines show the 95 per cent posterior coverage interval for the model estimate.
Sources: BIS, National Institute of Economic Research, Macrobond, national sources, Statistics Sweden, Sveriges Riksbank and own calculations

The krona can be said to have been weak over the periods in which the actual real exchange rate exceeded the estimated long-run level, and strong when the opposite was true. For example, the Swedish krona was weak during the most intensive phase of the global financial crisis 2009–2010, which was a period of clearly elevated uncertainty surrounding the world economy in general and developments on the financial markets in particular. The krona was subsequently strong over the years 2011–2014, when Swedish monetary policy was less expansionary than it was abroad, and may therefore have contributed to a stronger krona exchange rate by affecting the interest rate differential against the rest of the world.[33] Between 2018 and 2020, the krona has been weak again. Bacchetta and Chikani (2021) also

33 The model explains the real exchange rate gap through the included variables' deviations from their respective trend levels, but without making further assumptions it is unfortunately impossible to identify the contributions made by each variable.

attempt to explain the development of the real krona exchange rate with largely the same set of variables but with a different method of estimation. They obtain a similar picture of the development of the real exchange rate gap.[34] Unfortunately, the reasons for the deviations from the long-run level cannot be identified, either in their model or in the model presented here. Further analysis is needed to gain deeper insights into this.

8 Conclusions

The effective real krona exchange rate – measured using the consumer price index – seems to have been stationary over most of the 1900s. However, in recent decades, it has undergone a clear trend depreciation that is almost unique in an international comparison in terms of size and duration. The real krona depreciation has taken place against all trading partners of significance, albeit to varying degrees. Even when the real exchange rate is calculated using alternative measures of price levels, such as the OECD and Eurostat's comparative price level indices and unit labour costs, a depreciation trend is visible. However, these calculations also give a slightly less dramatic picture of the development of the real krona exchange rate, particularly after 2008. The trend depreciation can, at least partially, be explained by weaker productivity growth in Sweden than abroad and a trend weaking of Sweden's terms of trade.

34 They use dynamic OLS regressions, in one variant on quarterly data from 1975 to 2018 and in another variant on annual data from 1970 to 2018. Their analysis does not relate trend levels but the variables' actual levels. In their model, the exchange rate gap is therefore simply the residuals from the regression.

References

Bacchetta, Phillippe and Pauline Chikhani (2021), 'On the weakness of the Swedish krona', *Sveriges Riksbank Economic Review*, no. 1.

Balassa, Bela (1964), 'The Purchasing-Power Parity Doctrine: A Reappraisal', *Journal of Political Economy*, vol. 72, no. 6, pp. 584–596.

Belfrage Carl-Johan, Paolo Bonomolo and Pär Stockhammar (2020), 'A time-varying equilibrium VAR model of the long-run real exchange rate', Staff Memo, Sveriges Riksbank.

Berka, Martin, Michael B. Devereux and Charles Engel (2018), 'Real Exchange Rates and Sectoral Productivity in the Eurozone', *American Economic Review*, Vol. 108, no. 6, pp. 1543–1581.

Blomberg, Gunnar and Maria Falk (2006), 'How do large current-account surpluses co-exist with a weak international investment position?', *Sveriges Riksbank Economic Review*, no. 1.

Blomberg, Gunnar and Johan Östberg (1999), 'Market evaluation of external position – a new picture of Sweden's international dependence', *Sveriges Riksbank Economic Review*, no. 2.

Bordo, Michael D., Ehsan U. Choudri, Giorgio Fazio and Ronald MacDonald (2017), 'The real exchange rate in the long run: Balassa-Samuelson effects reconsidered', *Journal of International Money and Finance*, vol. 75, pp. 69–92.

Cassel, Gustav (1918), 'Abnormal Deviations in International Exchanges', *The Economic Journal*, vol. 28, no. 112, pp. 413–415.

Ca' Zorzi, Michele, Marcin Kolasa and Michał Rubaszek (2017), 'Exchange rate forecasting with DSGE models', *Journal of International Economics*, vol. 107, pp. 127–146.

Chen, Yu-chin and Kenneth Rogoff (2003), 'Commodity Currencies', *Journal of International Economics*, vol. 60, no. 1, pp. 133–160.

Christopoulos, Dimitris K., Karine Gente and Miguel A. León-Ledesma, 'Net foreign assets, productivity and real exchange rates in constrained economies', *European Economic Review*, vol. 56, pp. 295–316.

Deaton, Angus and Bettina Aten (2017), 'Trying to Understand the PPPs in ICP 2011 Why Are the Results So Different?', *American Economic Journal: Macroeconomics*, vol. 9, no. 1, pp. 243–264.

Dornbusch, Rudiger (1985), 'Purchasing power parity', *NBER Working Paper Series*, no. 1591.

Erlandsson, Mattias and Alek Markowski (2006), 'The Effective Exchange Rate Index KIX – Theory and Practice', Working Paper No. 95, National Institute of Economic Research.

EU and OECD (2012), *Eurostat-OECD Methodological Manual on Purchasing Power Parities*.

Froot, Kenneth A. and Kenneth Rogoff (1995), 'Perspectives on PPP and long-run real exchange rates', chapter 32 of *Handbook of International Economics* vol. 3, G. Grossman and K. Rogoff (eds.), Elsevier Science.

Gubler, Matthias and Christoph Sax (2019), 'The Balassa-Samuelson effect reversed: new evidence from OECD countries', *Swiss Journal of Economics and Statistics*, vol. 155, no. 3.

Lane, Philip R. and Gian M. Milesi-Ferretti (2004), 'The transfer problem revisited: net foreign assets and real exchange rates', *The Review of Economics and Statistics*, vol. 86, no. 4, pp. 841–857.

Lane, Philip R. and Gian M. Milesi-Ferretti (2018), 'The External Wealth of Nations Revisited: International Financial Integration in the Aftermath of the Global Financial Crisis', *IMF Economic Review*, vol. 66, pp. 189–222.

Meese, Richard A. and Kenneth Rogoff (1983), 'Empirical Exchange Rate Models of the Seventies: Do they fit out of sample?', *Journal of International Economics*, vol. 14, pp. 3–24.

Neary, Peter (1988), 'Determinants of the Equilibrium Real Exchange Rate', *American Economic Review*, vol. 78, no. 1, pp. 210–215.

OECD (2016), 'Estimation of annual PPPs before 2013 in the Eurostat-OECD PPP Programme', www.oecd.org/sdd/prices-ppp/Estimation-of-annual-PPPs-before-2013-in-the-Eurostat-OECD-PPP-Prog.pdf, accessed 16 February 2021.

Ravaillon, Martin (2018), 'An exploration of the changes in the international comparison program's global economic landscape', *World Development*, vol. 105, pp. 201–216.

Samuelson, Paul A. (1964), 'Theoretical Notes on Trade Problems', *The Review of Economics and Statistics*, vol. 46, no. 2, pp. 145–154.

Statistics Sweden (2017),'Statistics Sweden measures inflation – the difference between the CPI, the CPIF and the HICP', www.scb.se/hitta-statistik/artiklar/2017/Sa-mater-SCB-inflation--skillnaden-mellan-KPI-KPIF-och-HIKP), accessed 18 May 2021.

Sveriges Riksbank (2019), 'Development of Swedish labour costs in an international perspective', article in *Monetary Policy Report*, February.

Sveriges Riksbank (2020), 'Historical time line', www.riksbank.se/sv/om-riksbanken/historia/historisk-tidslinje, accessed 4 December 2020.

Taylor, Alan M. and Mark P. Taylor (2004), 'The Purchasing Power Parity Debate', *Journal of Economic Perspectives*, vol. 18, no. 4, pp. 135–158.

The Economist (2020), 'Burgernomics: The Big Mac index', www.economist.com/news/2020/07/15/the-big-mac-index, accessed 21 December 2020.

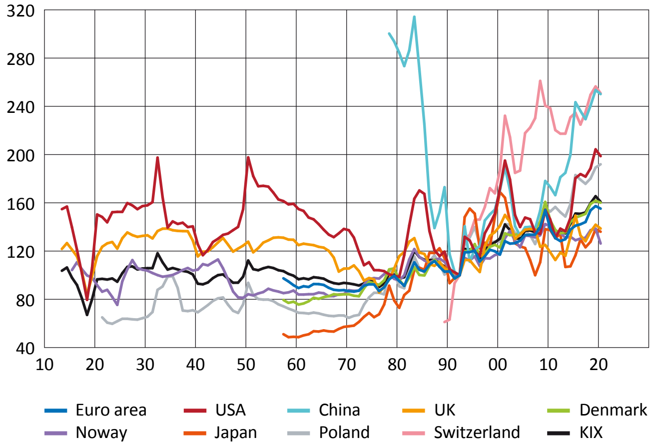
Tysklind, Oskar (2020), 'Quality adjustments and international price comparisons', Staff Memo, Sveriges Riksbank.

World Bank (1970), *Annual Report 1970*, pubdocs.worldbank.org/en/815631586875431072/IBRD-Financial-Statements-June-1970.pdf, accessed 6 January 2021.

Appendix

Figure A1. Bilateral real exchange rates against the currencies that have contributed most to the depreciation of the KIX-weighted real exchange rate

CPI-based bilateral real exchange rates, index 1992 = 100



Note. Annual data. The bilateral real exchange rates shown are those that have made a contribution of more than 2 percentage points to the depreciation of the KIX-weighted index since 1957, 1970 or 1992. Up to and including 1994, KIX weights from 1994 have been used.

Sources: BIS, Macrobond, Sveriges Riksbank and own calculations

How lasting are the economic consequences of pandemics? 220 years of Swedish experiences

Stefan Laséen*

The author works in the Riksbank's Monetary Policy Department

In this article I use the Riksbank's historical monetary statistics to analyse what effects pandemics have had on demographic and economic variables in Sweden since the start of the 19th century. The results show that pandemics have had negative effects on birth rates, death rates and family formation. Pandemics have also adversely affected the Swedish economy in the short term. The longer term effects are less clear. The effects on foreign trade and investment have, on the other hand, tended to be more long lasting. Going forward, this could imply that it will be important to be aware of potential protectionist tendencies, such as export restrictions and tariffs.

The COVID-19 crisis is in many respects unique, and therefore it is difficult to draw conclusions regarding the current situation on the basis of earlier pandemics. Furthermore, society has developed quite dramatically over the past 220 years with regard to knowledge, statistics, amount and spread of information, supply of media, technology and medical care. But even if one can discuss what conclusions can be drawn on the basis of earlier pandemics, the historical perspective is interesting in itself. Thanks to this, one can identify structures and mechanisms that can help today's decision-makers and authorities to better plan for and manage future threats.

1 Crises can have long-lasting effects

The COVID-19 pandemic is causing tremendous human and economic hardship around the world. The question that many people are now asking, is how long-lasting its effects might be.[1] In this article, I use long time series from the Riksbank's historical monetary statistics to study what effects earlier pandemics have had on variables such as gross domestic product (GDP) and inflation.[2] The Riksbank's statistics extend back to the early 17th century, and include information that, as far as I know, has not been studied in this particular context. Sweden's unique historical statistics also contain important demographic information such as the number of deaths, births, marriages and population. Based on pandemics in the period 1800–2020, I show that the long-term effects are uncertain. Pandemics have had effects in the shorter term and they have affected the entire fabric of society: from death rates to

* The author would like to thank Mikael Apel, Meredith Beechey Österholm, Jesper Hansson, Ulf Söderström and Anders Vredin as well as participants at a seminar at the Riksbank for their valuable comments. The opinions expressed in this paper are the sole

1 See, for instance, Blanchard and Pisani-Ferry (2021), Bodnár et al. (2020), Cerra et al. (2020a), Cerra et al. (2020b), Kozlowski et al. (2020), Martín Fuentes and Moder (2021), Moghadam et al. (2021), Sveriges Riksbank (2020).

2 With regard to the effects on inflation Bordo and Levy (2020), Goodhart (2020) and Goodhart and Pradhan (2020) contribute both theoretical and historical perspectives on how inflation can be lastingly affected by changes in the interplay between fiscal and monetary policy after wars, crises and demographic changes. Bordo and Levy (2020) discuss the connection between fiscal and monetary policy, while Goodhart (2020) and Goodhart and Pradhan (2020) argue that an ageing population and a decline in the rate of globalisation may entail higher inflation. They claim that a rising future dependency burden (total population compared with employed) will raise inflation. The impact of pandemics on demographics and sovereign debt could thus have lasting effects on inflation. Note that the analysis refers to lasting and not permanent effects on inflation. Blanchard (2020) discusses different potential scenarios for inflation after the pandemic.

family formation, from external markets to internal ones, from supply to demand. Overall, I find that there tends to be a more persistently negative effect on foreign trade, investment and real sovereign debt.

What repercussions a pandemic has depends on virulence, that is to say the ability of a microorganism to cause disease in the host. But it also depends on the economic, political and medical responses and on how much households and companies change their behaviour and for how long.[3] Pandemics can have long-lasting effects on the labour supply if, for instance, the working age population is affected. People who have been unemployed for a longer period risk losing competence and skills, increasing the risk that they will get caught in long-term unemployment. Even those who enter the labour market during a deep crisis can be affected through persistently lower wages. Pandemics can also have lasting effects on demand through increased precautionary savings and lower investment. Moreover, international trade may be affected by protectionism or changes in value chains and trade patterns.[4]

2 Effects of pandemics on Swedish demographic and economic conditions between 1800 and 2020

Pandemics have not only caused considerable human hardship, but also major economic strains. Figures 1 and 2 show the percentage change in the number of deaths and GDP per capita, respectively. The red dots show which year various epidemic diseases spread among the Swedish population. The pattern is relatively clear and unfortunately familiar from the coronavirus crisis. We can see in the figures that GDP fell by 7.8 per cent during the second cholera pandemic in 1834 and by 6.8 per cent during the 'Spanish flu' in 1918–1920. However, in the years when the 'Russian flu' (1889), the 'Asian flu' (1957) and the 'Hong Kong flu' (1968) affected Sweden and the world, growth was not negative. Figure 3 confirms the general picture that GDP growth was more clearly negatively affected in years when more serious epidemics, in terms of the number of deaths, affected Sweden.[5] In other periods, the correlation between mortality rate and GDP growth is not significant.

I study pandemics that have taken place since the start of the 19th century and limit my analysis to pandemics that have costed more than 100,000 lives in Europe.[6] Pandemics are per definition global, but I also control for epidemics that have at least partly only affected Sweden, as shown in Figures 1–3.[7] The reason for studying pandemics, and not just Swedish epidemics, is that their origins can be regarded as independent of Swedish economic and political conditions.[8] It is in this way possible to derive the effects from the pandemic and not

3 See, for instance, the article 'Long-term effects of the pandemic on the Swedish economy' in the Account of Monetary Policy 2020 (Sveriges Riksbank 2020).

4 Structural transformation and investment in new technology and new methods of working or organising companies and societies may on the other hand ultimately have positive effects on technological advances and growth rates of economies (Dieppe (ed.) 2020). With regard to pandemics' effects on protectionism, Boberg-Fazlic et al. (2020) find, for instance, that the Spanish flu 1918–1920 had a significant effect on trade policy and that tariffs increased as a consequence of the pandemic.

5 Barro et al. (2020) shows similar results for 43 countries affected by the Spanish flu during 1918–1920. They draw the conclusion that a higher influenza mortality rate led to lower GDP growth. The traditional perception that cyclical variations do not affect long-term growth has to some extent been questioned in academic research. For instance, the existence of hysteresis effects on the labour market (that is, a very lasting or permanent effect of shocks in unemployment) is a phenomenon that has been investigated and discussed to a large degree (see for example Blanchard and Summers 1986). Moreover, academic research, motivated by the slow recovery after the global financial crisis, has shown that economic recessions can cause lasting ('scarring') effects on the level of GDP, as more cyclical phenomena and events can affect the supply side of the economy (see, for example, Cerra et al. 2020a, Jordà et al. 2020 and Bluedorn and Leigh 2018).

6 One alternative is to study more detailed statistics and information about a specific pandemic. See, for instance, Karlsson et al. (2014) for a regional analysis of the Spanish flu in Sweden.

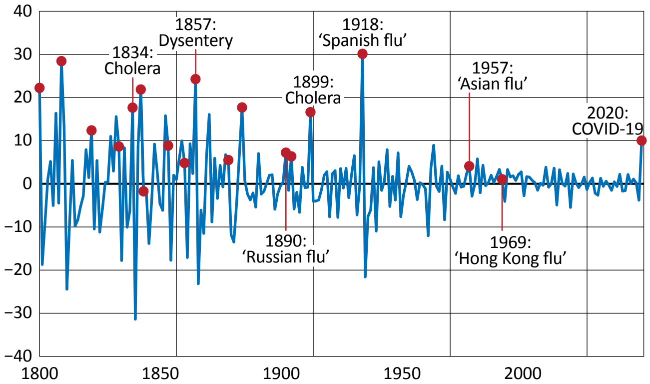
7 See, for instance, Kelly (2011) for a discussion of how a pandemic is defined.

8 Historians and various social commentators have discussed the causes of these crises. Suggestions include food shortages, lack of hygiene, war or overpopulation, but the results are often contradictory. Here I therefore take a more global perspective and study the larger European pandemics that have not originated in Sweden.

from other conditions. The pandemics I study are the same as those studied by Jordà et al. (2020) and my analysis follows their approach to a large degree. My focus is however on Sweden, where we have access to more detailed macroeconomic and demographic statistics. Although I have statistics for several variables from the start of the 17th century, I limit the sample to 1800–2020. The reason is that I can control for and analyse more variables after 1800. I thus obtain a better picture of how pandemics have affected the economy.[9]

Figure 1. Change in number of deaths and different epidemics in Sweden

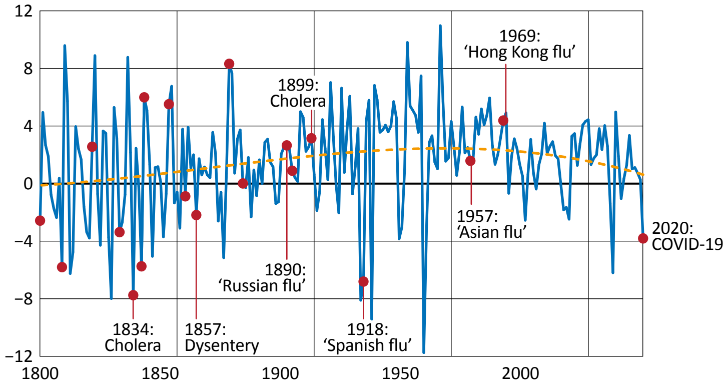
Per cent



Note. See Table A2 in Appendix A for a compilation of epidemics.
Source: Statistics Sweden

Figure 2. GDP growth per capita and different epidemics in Sweden

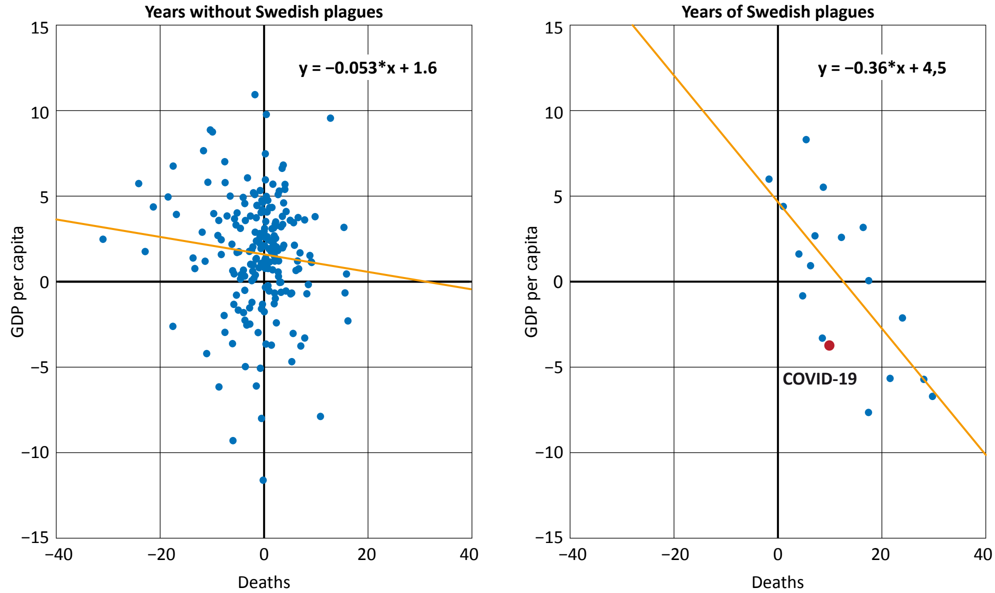
Per cent



Note. See Table A2 in Appendix A for a compilation of epidemics. The broken line shows a trend estimated using a third degree polynomial.
Source: Statistics Sweden

9 However, the choice of sample period implies that plague pandemics are not studied, as the final plague occurred between 1720 and 1722.

Figure 3. Correlation between GDP growth per capita and change in number of deaths
Per cent



Note. See Table A2 in Appendix A for a compilation of epidemics. The solid line represents the best linear fit.
Source: Statistics Sweden

To study the effects of pandemics, I use a time series regression model that in the academic literature is termed a local projection model.[10] Under certain conditions one can use such a model to estimate the causal effects of, for instance, pandemics on different demographic or economic variables. In the regressions I control for a number of demographic and economic variables for up to eight years prior to the pandemics. (See Appendix B for a more detailed description.) Figures 4, 5 and 6 show the estimated average effects during a period of up to 20 years after pandemics between 1800 and 2020.[11] In Figure 4 we can see that the number of deaths has risen on average by several per cent in the years following a pandemic. In the longer run, the number tends to fall. The number of deaths has not only risen directly, but also again after four to five years. Some of the pandemics, such as the second cholera pandemic, took several years to reach Sweden, which was affected in 1834.[12] The most obvious effect on the demographic variables I study is that the population has declined by on average around 2 per cent in the longer run after a pandemic. The number of marriages and the number of births have on average declined in the short and medium term by 1 and 2 per 1,000 inhabitants. To summarise, the number of births, deaths and marriages have all been adversely affected by the pandemics.[13]

In Figure 5 we can see that the average effect on GDP per capita is relatively uncertain. In the short term, GDP has fallen after a pandemic, but the results are not significantly different from zero. This is in line with Figure 3, which also shows that Swedish epidemics are not always correlated with a negative GDP growth. However, concealed behind the effects on GDP are relatively noticeable changes in the components of GDP. Consumption as a percentage of GDP, both private and public sector, has tended to rise while investments, exports and imports have declined as a percentage of GDP in the shorter term (up to almost 10 years). Investments and exports have declined and are the components that

10 See, for example, Jordà (2005) and Montiel Olea and Plagborg-Møller (under publication).

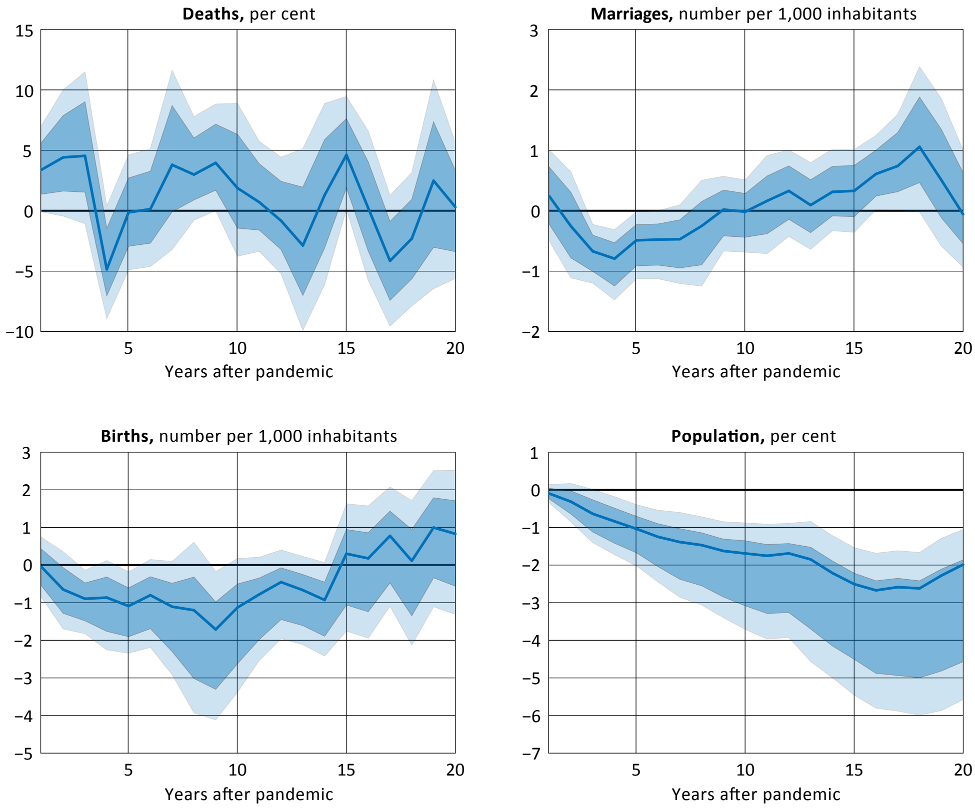
11 If the assumption that the pandemics are exogenous and random, and that the residuals in equation 2 in Appendix B are independent of information going forwards and backwards in time (assumption 1 in Montiel Olea and Plagborg-Møller) then the effects can be interpreted as causal relationships.

12 See, for instance, the descriptions on pages 182–183 in Lundin and Strindberg (1882).

13 See Boberg-Fazlic et al. (2017) and Bloom-Feshbach et al. (2011) for similar results, and Ullah et al. (2020) for a discussion of potential effects of COVID-19 on future birth rates.

have primarily contributed to GDP falling. Gross national saving (investments and net exports) in the economy has thus fallen over a period, which one would not expect if it was precautionary saving that was primarily affected, as savings would in this case have increased.

Figure 4. Demographic effects of pandemics between 1800 and 2020
Per cent and number per 1,000 inhabitants



Note. The different figures show the average historical effects on demographic variables up to 20 years after the pandemics that are compiled in Table A2 in Appendix A. The shaded areas show 1 and 2 standard deviation confidence intervals. If the responses (including the shaded areas) are different from zero (they do not include the x-axis), one can conclude that the pandemics on average have had a statistically significant effect on the variable in question.

The fact that the population declines after a pandemic means, all else being equal, a lower supply of labour (see Figure 6). This is compatible with higher real wages, for which there have been tendencies. However, in the long term, real wages have fallen back. The effects on inflation are uncertain and not significant. This may be because pandemics have negative effects on both supply, which increases costs and prices, and demand, which reduced prices. Economic policy has been relatively passive or even tighter in the short/medium term, interpreted in terms of the effects on the money supply and sovereign debt. The effect on the money supply is negative and there are no significant effects on sovereign debt for up to ten years. In the longer run, monetary and fiscal policy have been more expansionary, with both rising sovereign debt and money supply.

All in all, the average effects in the short and medium term are relatively clear. They are also by and large in line with what one can expect according to models that integrate spread of infection and the macro economy, which show negative effects on both demand for and supply of labour.[14] This has contradictory effects on prices, which are in line with the relatively moderate, but primarily uncertain effects that pandemics have on inflation. In the

14 See, for example, Eichenbaum et al. (2021).

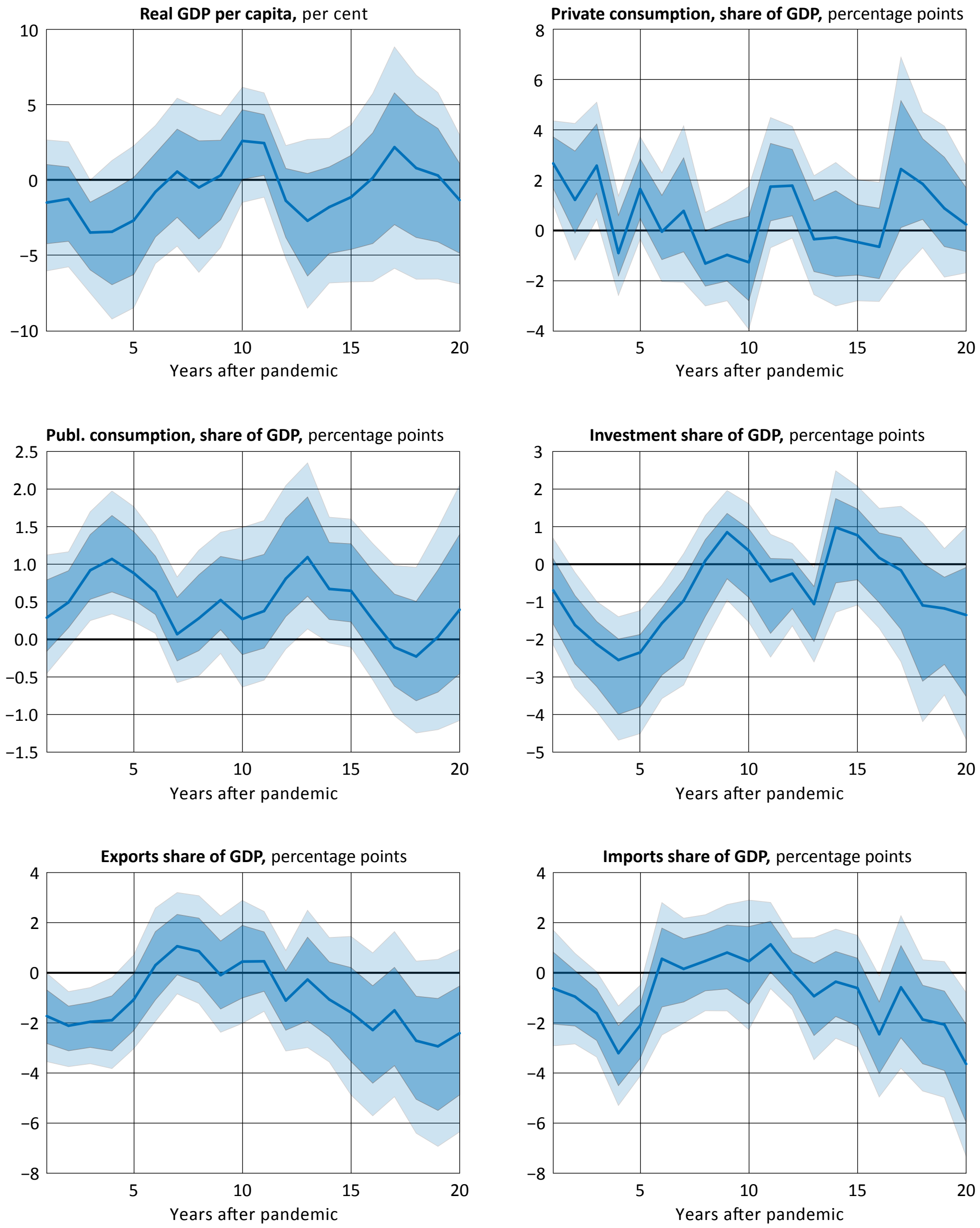
longer term, the effects are uncertain and mostly not significant. Foreign trade, investment and real sovereign debt are exceptions, however. They have on average tended to be more persistently affected by pandemics.

Other research articles show both similar and different effects of pandemics. Jordà et al. (2020) study what effects pandemics have on, for instance, real interest rates and find that interest rates fall persistently. Their historical perspective and method are similar to my own approach, but they lack detailed information on demographics, GDP and inflation. For the United Kingdom, they show that GDP per capita and real wages have on average risen after pandemics between 1311 and 2016.

Ma et al. (2020) analyse the effects on the basis of a much smaller sample of pandemics between 1968 and 2018, but for a larger number of countries.[15] Their results indicate significant effects of pandemics in the countries affected. Real GDP falls by around three per cent and unemployment rises by around one percentage point. Moreover, the effects last for up to five years. The growth rate bounces back relatively quickly, but the level of GDP remains low even after five years. Public consumption rises, however, and counteracts to some extent the effects of the health crisis. They also show a negative impact on trade. The epidemics included in the analysis were mostly local events that are not comparable with a global pandemic. But on the whole, the results are concordant with those I find for Sweden.

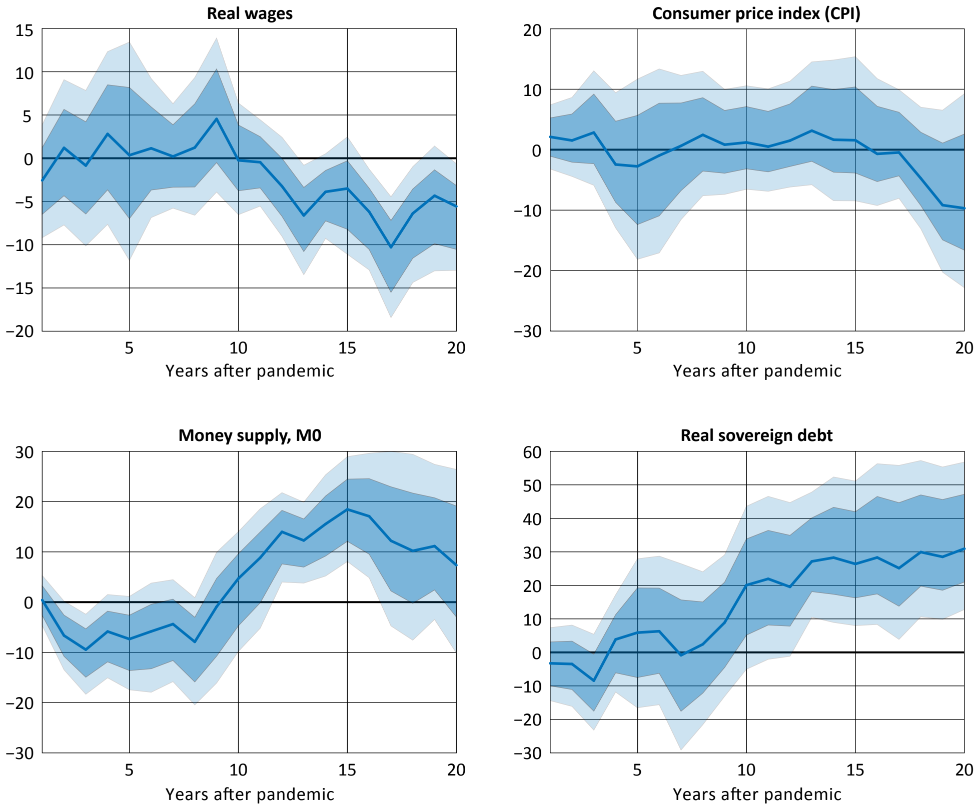
15 An article that is very similar in many ways to that of Ma et al. (2020) is Martín Fuentes and Moder (2020). They analyse what effects pandemic have on potential growth, as well as the effects of other crisis-like events such as wars, oil embargos and financial crises, since 1970. The results imply that the initial effect on the level of potential production is relatively short-lived and tends to disappear two years after the end of the epidemic. Financial crises, on the other hand, are linked to very lasting negative effects on potential production levels.

Figure 5. Economic effects of pandemics between 1800 and 2020
Per cent and percentage points



Note. See note under Figure 4.

Figure 6. Economic effects of pandemics between 1800 and 2020
Per cent



Note. See note under Figure 4.

3 Closing comments

The results show that pandemics have had negative effects on birth rates, death rates and family formation. They have also had a negative impact on the Swedish economy in the short term. The longer term effects are however less clear. Foreign trade, investment and real sovereign debt have tended to be negatively impacted more persistently. This could possibly imply that it will be important to be watchful of potential protectionist tendencies, such as export restrictions and tariffs which can in turn negatively affect foreign trade going forward.

What conclusions can we actually draw from earlier experiences of pandemics?[16] It is difficult to respond to this for several reasons. Society has developed dramatically over the past 220 years, and the situation with regard to overall knowledge, public access to media, technological progress and medical care is very different today. At the same time, diseases can now spread rapidly, both within and between countries, which means that one must act quickly in response to the initial outbreak. Climate change can also play a role.[17] All of this means that the consequences of the coronavirus pandemic may differ from previous pandemics. Another aspect is that the national lockdowns have no counterpart in history, even if regional travel bans have been used in earlier pandemics.[18] In addition, the effects may depend on which part of the population is mostly affected. Almost all deaths related to COVID-19 were among the elderly. During the pandemic 1918–1919, deaths occurred

16 See Conley and Johnson (2021) for a discussion.
17 Apart from being exacerbated by globalisation, epidemic potential is elevated by climate change and urbanisation (Bloom et al. 2018). Climate change extends life environments for various common disease vectors, such as the Aedes aegypti mosquito, which can spread dengue, chikungunya, Zika and yellow fever. Urbanisation means that more people live close together, which amplify transmission of contagious diseases.
18 See Mateus et al. (2014).

instead mainly among younger people.[19] Gagnon et al. (2020) show in a theoretical model that if COVID-19 would have affected the population according to the same age pattern as the Spanish flue, the effects on the supply side of the economy would be stronger and more lasting.

Even if one can discuss what conclusions can be drawn on the basis of earlier pandemics, the historical perspective is interesting in itself. Thanks to this, one can identify structures and mechanisms that can help today's decision-makers and authorities to better plan for and manage future threats.[20]

19 See, for example, Simonsen et al. (1998)
20 See, for example Elgh (2007) for a discussion.

References

Barro, Robert, José Ursúa, and Joanna Weng (2020), 'The Coronavirus and the great influenza pandemic: lessons from the "Spanish Flu" for the Coronavirus's potential effects on mortality and economic activity', Working Paper Series 26866, NBER working paper.

Blanchard, Olivier (2020), 'Is there deflation or inflation in our future?', VoxEU.org, 24 April.

Blanchard, Olivier and Jean Pisani-Ferry (2021), 'Persistent COVID-19: Exploring potential economic implications', PIIE series on Economic Policy for a Pandemic Age: How the World Must Prepare.

Bloom-Feshbach, Kimberley, Lone Simonsen, Cécile Viboud, Kåre Mølbak, Mark Miller, Magnus Gottfredsson and Viggo Andreasen (2011), 'Natality decline and miscarriages associated with the 1918 influenza pandemic: the Scandinavian and United States experiences', *The Journal of Infectious Diseases*, vol. 204, no. 8, pp. 1157–1164.

Bloom, David, Daniel Cadarette and JP Sevilla (2018), 'The economic risks and impacts of epidemics', *International Monetary Fund, F&D Magazine*, June 2018.

Bluedorn, John and Daniel Leigh (2018), 'Is the cycle the trend? Evidence from the views of international forecasters', IMF Working Papers, No. 18/163.

Boberg-Fazlic, Nina, Maryna Ivets, Martin Karlsson and Therese Nilsson (2017), 'Disease and fertility: evidence from the 1918 influenza pandemic in Sweden', Research Institute of Industrial Economics Working Paper Series, No. 1179.

Boberg-Fazlic, Nina, Markus Lampe, Maja Uhre Pedersen and Paul Sharp (2020), 'Pandemics and protectionism: evidence from the "Spanish flu"', CAGE Online Working Paper Series 479, Competitive Advantage in the Global Economy (CAGE).

Bodnár, Katalin, Julien Le Roux, Paloma Lopez-Garcia and Béla Szörfi (2020), 'The impact of COVID-19 on potential output in the euro area', *Economic Bulletin*, No. 7, ECB.

Cerra, Valerie, Antonio Fatás, and Sweta Saxena (2020a), 'The persistence of a COVID-induced global recession', VoxEU.org, 14 May.

Cerra, Valerie, Antonio Fatás, and Sweta Saxena (2020b), 'Hysteresis and business cycles,' CEPR Discussion Papers 14531, C.E.P.R. Discussion Papers.

Conley, Dalton and Tim Johnson (2021), 'Opinion: Past is future for the era of COVID-19 research in the social sciences', *Proceedings of the National Academy of Sciences* Mar 2021, vol. 118, no. 13.

Dieppe, Alistair (ed.), (2020), *Global productivity: trends, drivers, and policies*, World Bank.

Edvinsson, Rodney, Tor Jacobson and Daniel Waldenström (ed.), *Historical Monetary and Financial Statistics for Sweden*, Vol. 1 and 2 , Stockholm: Sveriges Riksbank and Ekerlids förlag, 2014.

Eichenbaum, Martin, Sergio Rebelo and Mathias Trabandt (2021), 'The macroeconomics of epidemics', *The Review of Financial Studies*, 2021.

Elgh, Fredrik, (2007), 'Pandemiers påverkan på samhället historisk erfarenhet nödvändig för dagens beredskapsåtgärder', *(The effects of pandemics on society, historical experience necessary for today's emergency preparedness)*, *Läkartidningen*, vol. 104, No. 8, pp. 615–619.

Goodhart, Charles (2020), 'Inflation after the pandemic: theory and practice', voxEU.org, 13 June.

Goodhart, Charles and Manoj Pradhan (2020), *The great demographic reversal: ageing societies, waning inequality, and the inflation revival*, Palgrave Macmillan.

Herbst, Edward and Benjamin Johannsen (2021), 'Bias in local projections,' Finance and Economics Discussion Series 2020-010r1. Washington: Board of Governors of the Federal Reserve System.

Jordà, Òscar (2005), 'Estimation and inference of impulse responses by local projections', *American Economic Review*, vol. 95, no. 1, pp. 161–182.

Jordà, Òscar, Sanjay Singh and Alan Taylor (2020), 'Longer-run economic consequences of pandemics', Federal Reserve Bank of San Francisco Working Papers No 2020-09.

Karlsson, Martin, Therese Nilsson and Stefan Pichler (2014), 'The impact of the 1918 Spanish flu epidemic on economic performance in Sweden: an investigation into the consequences of an extraordinary mortality shock,' *Journal of Health Economics*, vol. 36, pp. 1–19.

Kelly, Heath (2011), 'The classical definition of a pandemic is not elusive', *Bulletin of the World Health Organization*, vol. 89, pp. 540–541.

Kozlowski, Julian, Laura Veldkamp and Venky Venkateswaran (2020), 'Scarring body and mind: The long-term belief-scarring effects of COVID-19', NBER Working Paper No. 27439.

Lundin, Claes and August Strindberg (1882), *Gamla Stockholm: anteckningar ur tryckta och otryckta källor*,(Old Stockholm: notes from printed and unprinted sources) Jos Seligmann & Co förlag.

Ma, Chang, John Rogers and Sili Zhou (2020), 'Modern pandemics: Recession and recovery,' International Finance Discussion Papers 1295, Washington: Board of Governors of the Federal Reserve System.

Martín Fuentes, Natalia and Isabella Moder (2020), 'The scarring effects of past crises on the global economy', *Economic Bulletin*, No. 8, ECB.

Mateus, Ana, Harmony Otete, Charles Beck, Gayle Dolan, and Jonathan Nguyen-Van-Tam (2014), 'Effectiveness of travel restrictions in the rapid containment of human influenza: a systematic review', *Bulletin of the World Health Organization*, vol. 92, no. 12, pp. 868–880D.

Moghadam, Reza, Markus Guetschow and Camille White (2021), 'Scarring in Europe', SUERF Policy Note, no. 227.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (under publication), 'Local projection inference is simpler and more robust than you think' *Econometrica*.

Statistics Sweden (2020), 'Kohortdödligheten i Sverige' (Cohort mortality in Sweden), Mortality developments up to 2019, Demographic reports 2020:3.

Schmelzing, Paul (2020), 'Eight centuries of global real interest rates, R–G, and the "suprasecular" decline, 1311–2018', Bank of England Staff Working Paper 845.

Simonsen, Lone, Matthew Clarke, Lawrence Schonberger, Nancy Arden, Nancy Cox, and Keiji Fukuda (1998), 'Pandemic versus epidemic influenza mortality: a pattern of changing age distribution', *The Journal of Infectious Diseases*, vol. 178, no. 1, pp. 53–60.

Sveriges Riksbank (2020), 'The long-term economic effects of the pandemic are uncertain' Article in Monetary Policy Report, November 2020.

Ullah, Asad, Abu Tayab Moin, Yusha Araf, Atiqur Rahman Bhuiyan, Mark Griffiths, David Gozal (2020), 'Potential effects of the COVID-19 pandemic on future birth rate', *Frontiers in Public Health*, vol. 8.

Appendix A

Data

Table A1. Compilation of pandemics with more than 100,000 deaths in Europe
Year and illness.

Year	Pandemic	Year	Pandemic
1596–1602	Plague in Spain	1915–1926	European sleeping sickness
1629–1631	Plague in Italy	1918–1920	'Spanish flu'
1647–1652	Plague in Seville	1957–1958	'Asian flu'
1656–1658	Plague in Naples	1968–1969	'Hong Kong flu'
1665–1666	Plague in London	2009	Swine flu
1700–1721	Plague in Nordic countries	2020–	COVID-19
1720–1722	Plague in Marseilles		
1816–1824	First cholera pandemic		
1826–1851	Second cholera pandemic		
1852–1860	Third cholera pandemic		
1863–1875	Fourth cholera pandemic		
1889–1890	'Russian flu'		
1899–1923	Sixth cholera pandemic		

Source: Jordà et al. (2020)

Table A2. Compilation of epidemics and pandemics in Sweden
Year and illness.

Year	Epidemic / Pandemic	Year	Epidemic / Pandemic
1757	Smallpox	1838–1839	Smallpox
1763	Smallpox	1847	Cholera
1772–1773	Dysentery	1853	Cholera
1779	Smallpox	1857	Dysentery
1783	Dysentery	1869	Smallpox
1784	Smallpox	1874–1876	Smallpox
1795	Typhus; smallpox	1889 (December)	'Russian flu'
1800	Smallpox	1892	Diphtheria
1808–1809	Dysentery	1899	Sixth cholera pandemic
1819	Dysentery	1918–1919	'Spanish flu'
1829	Measles	1957	'Asian flu'
1834	Cholera; smallpox	1969	'Hong Kong flu'
1837	Smallpox		

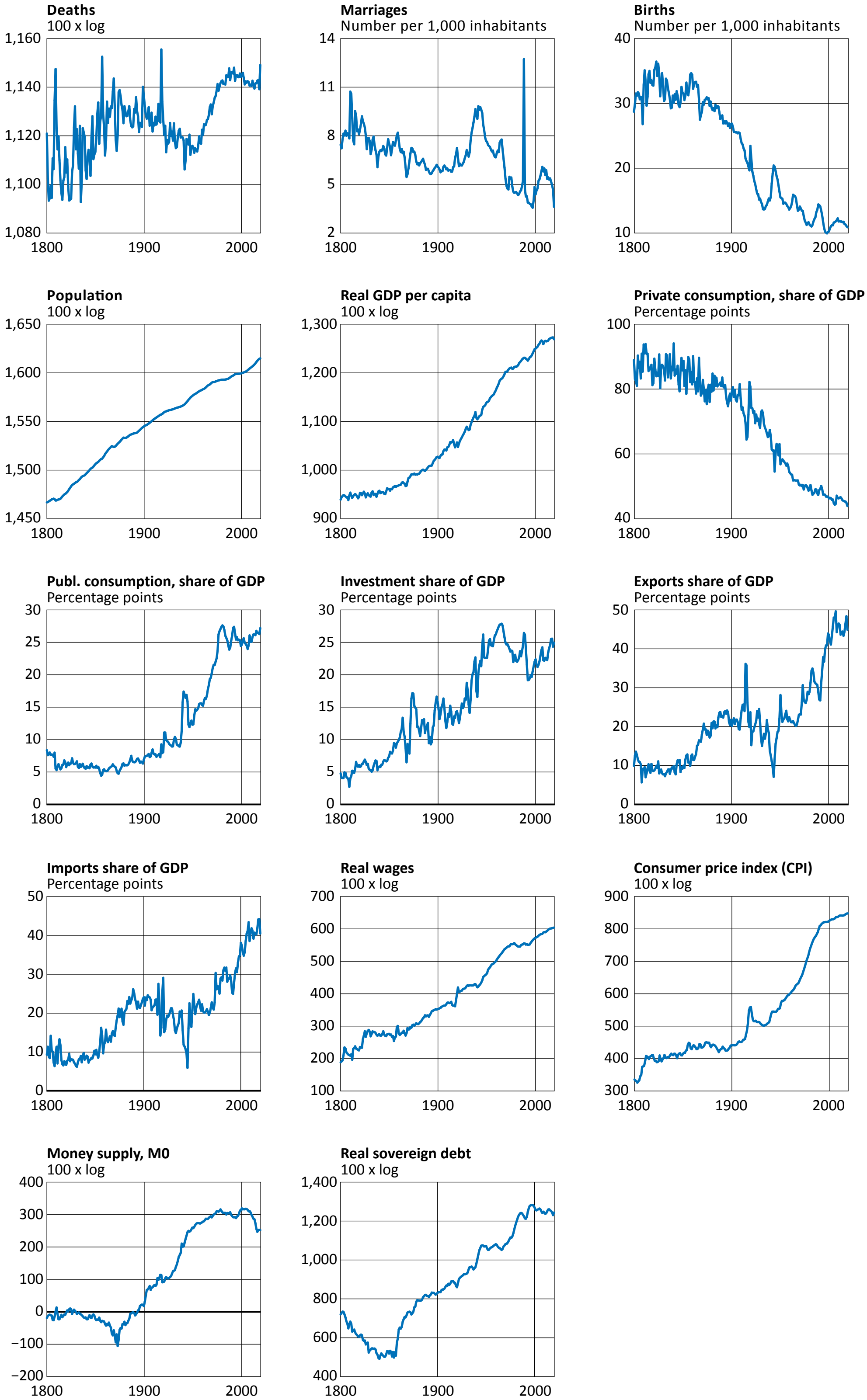
Source: Statistics Sweden (2020)

Table A3. Compilation of statistics, transformations and sources.

Variable	Unit	Years available	Transformation	Source
Pandemic	Dummy	1600–2020	None	Jordà et al. (2020)
Epidemics	Dummy	1757–2020	None	Statistics Sweden (2020)
Deaths	Number	1749–2020	100 x log	Statistics Sweden
Marriages	Number	1749–2020	1000 x (marriages / population)	Statistics Sweden
Births	Number	1749–2020	1000 x (births / population)	Statistics Sweden
Population	Number	1749–2020	100 x log	Statistics Sweden
Real GDP per capita	SEK, year 2000 prices	1620–2020	100 x log	Sveriges Riksbank and Statistics Sweden
Private consumption, share of GDP	Percentage	1800–2020	100 x (nom C / nom GDP)	Sveriges Riksbank and Statistics Sweden
Publ. consumption, share of GDP	Percentage	1800–2020	100 x (nom G / nom GDP)	Sveriges Riksbank and Statistics Sweden
Investment share of GDP	Percentage	1800–2020	100 x (nom I / nom GDP)	Sveriges Riksbank and Statistics Sweden
Exports share of GDP	Percentage	1800–2020	100 x (nom Exp / nom GDP)	Sveriges Riksbank and Statistics Sweden
Imports share of GDP	Percentage	1800–2020	100 x (nom Imp / nom GDP)	Sveriges Riksbank and Statistics Sweden
Real wages	SEK	1600–2020	100 x log (wages / CPI)	Sveriges Riksbank, Statistics Sweden and National Mediation Office
Consumer price index (CPI)	1914 = 100	1600–2020	100 x log	Sveriges Riksbank and Statistics Sweden
Money supply, M0	SEK	1620–2020	100 x log (M0 / CPI)	Sveriges Riksbank and Statistics Sweden
Real sovereign debt	SEK	1670–2020	100 x log (debt / CPI)	Sveriges Riksbank and Statistics Sweden

Note. The source Sveriges Riksbank refers to historical monetary statistics for Sweden. A description can be found in Edvinsson et al. (2014).

Figure A1. Compilation of statistics described in Table A3



Source: See Table A3 for transformations and sources.

Appendix B

Econometric method

One can compare a pandemic to a randomised controlled study, but on a larger scale, where the spread of infection determines when a population is affected.[21] If one treats a pandemic as this type of controlled randomised experiment, it is statistically possible to construct a counterfactual expected economic development, given what has been observed and the historical statistical sample. Similarly, one can calculate the expected development if a pandemic occurs. This is the background to the statistical method I use (Jordà 2005). The effect of a pandemic that occurs in the period *t* on, for example, GDP in period *t* + *h* can more formally be expressed as follows:

(1) $$\tau(h) = E(y_{t+h} | P_t = 1; \Omega_t) - E(y_{t+h} | P_t = 0; \Omega_t), \qquad h = 0{:}H,$$

where y_{t+h} refers the logarithm of the dependent variables studied for year t+h, for instance, Swedish GDP. P_t is a dummy variable[22] that assumes the value 1 in the last year of a pandemic (see Table A1) and the value 0 if there is no pandemic in a particular year, and the operator $E(.|.)$ states the best forecast in terms of lowest mean square value of the forecast errors. The information set in year *t* that I control for, that is, the variables included as independent variables in the regression below (X_t), is given by Ω_t. I estimate $\tau(h)$ in the same way as Jordà et al. (2020), namely with a so-called local projection which means that *H* regressions are estimated where $h = 20$:

(2) $$y_{t+h} = \alpha^h + \beta^h P_t + \sum_{l=1}^{L} \gamma_l^h X_{t-l} + \varepsilon_{t+h}^h; \qquad h = 0,\dots,H,$$

where $\beta^h = \tau(h)$. I choose 8 lags and thereby set $L = 8$.[23] In X_t the basic specification includes the following variables: lagged dependent variable, epidemics that spread among the Swedish population different years, number of deaths during a year, marriages, births, population number, real GDP per capita, real money supply, real sovereign debt, inflation index (consumer price index), private consumption as a percentage of GDP, public consumption as a percentage of GDP, fixed gross investment as a percentage of GDP, exports as a percentage of GDP, imports as a percentage of GDP and real wages.[24] All variables are multiplied by 100. The variables that are not measured as a percentage of GDP are expressed in logarithms, which means that the effects are expressed in per cent or percentage points (the exception is marriages and births, which are expressed in numbers per 1,000 individuals in the population). The pandemics are a dummy variable that is not transformed. See Table A3 above for a description of sources and transformations.

21 A randomised controlled study is a study where the participants are randomly selected either to the group that has the intervention or treatment to be studied, or to a control group. This in theory creates groups that are similar to one another on average. The only thing that separates the groups systematically is what intervention they receive.

22 A dummy variable is a variable that indicates the absence or presence of a particular property. A dummy variable assumes the values 0 and 1, where 0 indicates the absence of properties and 1 indicates the presence of the properties.

23 The results are robust for L=4. See Montiel Olea and Plagborg-Møller (under publication) for the importance of including the lagged dependent variable (y_{t-1}) as control variable in local projections where the variables are persistent. See also Herbst and Johannsen (2021) for a discussion of the bias (systematic errors) in β^h which can arise in short samples (they focus on 50 and 100 periods). The results are almost the same if I instead estimate (2) with $y_{t+h} - y_{t-1}$ as dependent variable and where the control variables are $X_t - X_{t-1}$.

24 In my application, I thus follow Montiel Olea and Plagborg-Møller (under publication) in two different ways. I check both for lagged variables of the dependent variable and for a large number of control variables. This is important for consistent inference of long-term impulse responses when data are persistent and to ensure that the standard deviation of the regression is conditionally average independent (their first assumption).

SVERIGES RIKSBANK
SE-103 37 Stockholm, Sweden
(Brunkebergstorg 11)

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