CPB Netherlands Bureau for
Economic Policy Analysis

SVARs, the central bank balance sheet and the effects of unconventional monetary policy in the euro area

The way the ECB has implemented unconventional monetary policy means that standard time series techniques will not correctly identify exogenous variation in the ECB's balance sheet. As such, the resulting estimates of the effects of unconventional policies will be unreliable.

Previous published studies don't account for these issues so do not provide evidence that ECB policy has been successful.

Replacing all information about the stance of monetary policy with random numbers produces equally plausible impulse responses and time series of purported monetary policy shocks.

CPB Discussion Paper

Adam Elbourne

December 2019

SVARs, the central bank balance sheet and the effects of unconventional monetary policy in the euro area

Adam Elbourne*

December 8, 2019

Abstract

Institutional features of monetary policy in the euro area likely make correctly identifying unconventional monetary policy shocks in SVAR models using the central bank balance sheet impossible. To illustrate this, I show that equally plausible results to those reported in Burriel & Galesi (2018), Boeckx *et al.* (2017) and Gambacorta *et al.* (2014) arise from models using random numbers rather than the balance sheet as the policy instrument in the euro area. I also show that the shocks identified by these SVAR models show clear evidence of the foresight problem and most have the wrong sign: the SVAR shocks are contractionary in periods when the ECB announced major expansionary policies. As such, SVAR models employing balance sheet data tell us nothing about the efficacy of these policies in the euro area.

JEL Classification: C32; E52
Keywords: Unconventional monetary policy; VAR models; Identification

*I would like to thank Gert Peersman for sharing the original code and data for the Boeckx *et al.* (2017) article, without which much of this analysis would not have been possible. I would also like to thank Klaus Adam, Michele Lenza, other participants at the EABCN conference in Warsaw and Bert Smid for helpful comments. Special thanks to Kan Ji for collaboration on a previous version of this paper.

1 Introduction

With short-term nominal interest rates at or close to their lower bound in the euro area, the ECB has turned to unconventional monetary policy (UMP) to try to fulfill its mandate. Since these policies are new, there is considerable uncertainty surrounding their effectiveness, so understanding the effects of UMP is vital for the practice of macroeconomic stabilisation. Theoretical models (see Arce *et al.* (2019), for example) and event studies focusing on the responses of financial markets (see Krishnamurthy *et al.* (2017) or Szczerbowicz *et al.* (2015), for example) provide evidence consistent with changes in central bank balance sheets having meaningful effects on macroeconomic aggregates. However, evidence consistent with effective UMP is not evidence that it is effective. That requires evidence directly linking changes in the ECB's balance sheet with output and inflation. Recently a number of authors have used structural vector autoregression (SVAR) models to provide such evidence (Boeckx *et al.* (2017), Burriel & Galesi (2018), Gambacorta *et al.* (2014), Lewis & Roth (2019), Haldane *et al.* (2016) and Wieladek & Garcia Pascual (2016)). Many of these studies report similar statistically significant hump-shaped responses for output and prices following balance sheet expansions, which are interpreted as evidence that unconventional monetary policy has been effective at stabilising the euro area economy. This paper, however, shows that this evidence is far from convincing because the UMP shocks the estimates rely on are not well identified.

The contribution of this paper is two-fold. Firstly, I argue that institutional features of ECB monetary policy means it is likely impossible to identify unconventional monetary policy shocks using the aggregate balance sheet. Specifically, many of the major policies were announced in advance, which gives rise to the foresight problem. In principal, this means a constant parameters VAR will never recover the true shocks even in an infinitely long sample (Giannone & Reichlin (2006)). This problem could, theoretically be overcome by modelling the announcements. However, there is an even more fundamental problem with the aggregate balance sheet data. The fixed-rate tender with full allotment (FRTFA) policy employed by the ECB since October 2008 involves announcing in advance that the supply curve of central bank funds will be horizontal. As a consequence, all variation in the balance sheet due to those policies are the endogenous response of the ECB to shifts in the demand for credit. In principal, this variation tells us nothing about exogenous monetary policy supply shocks. Since these policies account for the vast majority of variation in the aggregate balance sheet, it is highly unlikely a VAR model estimated on aggregate balance sheet data will correctly identify exogenous money supply shocks. The second contribution is to highlight that plausible impulse response functions do not imply successful identification. I show that it is easy to generate equally plausible impulse response functions by replacing the size of the ECB's balance sheet with random numbers[1] or by assuming expansionary policy shocks shrink the balance sheet. This clearly demonstrates that just because a model produces plausible IRFs doesn't mean they are responses to monetary policy shocks. I make no claim that this is a groundbreaking insight, but the vast majority of papers in this literature simply report the identifying assumptions and argue the impulse response functions look plausible. They provide no direct evidence that the exogenous variation they claim to isolate are plausibly monetary policy shocks. The only paper from those listed here that provides further evidence to back up the claim that monetary policy shocks have been identified is Boeckx *et al.* (2017), although a closer inspection of their evidence shows it is unconvincing.

The remainder of this paper continues as follows. Section 2 introduces the basic features of identification strategies employed in the literature. Section 3 highlights the fundamental econometric shortcomings with the balance sheet that makes it highly implausible that it will be useful for identifying monetary policy shocks. Section 4 shows how replacing the size of the ECB's balance sheet with random numbers in models from the existing literature produces equally plausible impulse response functions before showing clear evidence that the purported monetary policy shocks from these VAR models suffer from the econometric shortcomings I highlight in Section 3. Finally, Section 5 offers some concluding comments.

2 Identifying monetary policy shocks

As in all empirical economics, in order to be able to make statements about cause and effect, VAR modelling relies on isolating a source of exogenous variation. Exogenous shocks are unpredictable, so VAR modelling starts by controlling for predictable variation in the policy variable in the reduced form

[1] I am not the first to evaluate the credibility of existing estimates by comparing them to models with some randomly generated inputs. Keller (1998) used randomly generated trade patterns to cast doubt on estimates of international R&D spillovers.

$$y_t = c + A(L)\, y_{t-1} + u_t, \tag{1}$$

where y_t is a vector of endogenous variables, c is a vector of constants, $A(L)$ is a lag polynomial containing matrices of regression coefficients and u_t are reduced form errors. A requirement of successful identification is that u_t is unpredictable. Even if this requirement is met, u_t is unsuitable for policy analysis because the reduced form error in the balance sheet equation contains a weighted average of both exogenous variation due to changes in monetary policy and the endogenous response of monetary policy to all other shocks impacting the economy in that period. Since it is a weighted average of many shocks, adding a shock to the reduced form error doesn't allow us to say anything about the effects of changes in monetary policy because the observed responses may have been responses to the other shocks that make up u_t. To analyse the effects of policy requires that we isolate a component of u_t that represents exogenous variation in monetary policy. This is done by the structural form and is given by

$$y_t = c + A(L)\, y_{t-1} + Be_t, \tag{2}$$

where $u_t = Be_t$ contains the mapping from reduced form errors to the structural shocks, e_t, we are interested in.

In summary, successful identification of exogenous balance sheet policy shocks depends on both $A(L)\, y_{t-1}$ and B. The role of $A(L)\, y_{t-1}$ is to control for all predictable variation in the data. Once the unpredictable variation has been isolated, the matrix B isolates the component of the unpredictable variation in the balance sheet that is due to monetary policy shocks from the endogenous response of the balance sheet to all other shocks impacting the economy. If either $A(L)\, y_{t-1}$ or B fails to fulfill its role, then the e_t that results will once again be a weighted average of all of the shocks impacting the economy, just like in the reduced form. Adding a shock to an unsuccessfully identified e_t suffers from exactly the same problem as adding a shock to the reduced form error described above. The core argument of this paper is that a constant parameters VAR estimated using balance sheet data will fail because both $A(L)\, y_{t-1}$ and B will be unable to fulfill their roles.

2.1 Identification in the existing literature

Before detailing the econometric problems with the balance sheet in Section 3, I will first introduce an identification scheme that has been widely adopted in the literature. This identification scheme was introduced by Gambacorta *et al.* (2014), further developed by Boeckx *et al.* (2017) and adapted to a Global VAR setting by Burriel & Galesi (2018). The problems with the balance sheet I highlight are not restricted to this identification scheme, but it is commonly used and will serve to highlight the problems faced by researchers attempting to identify UMP shocks. To recover UMP shocks Burriel & Galesi (2018) use the identification scheme and endogenous variables shown in Table 1.[2] The assumptions employed that are supposed to isolate an unconventional monetary policy shock are, firstly, that the endogenous variables listed in Table 1 are sufficient to control for predictable variation in monetary policy. Secondly, to distinguish between endogenous and exogenous variation in the balance sheet it is assumed that an expansionary unconventional monetary policy shock increases the size of the central bank's balance sheet, lowers financial system stress and lowers the EONIA-MRO spread, whilst having no contemporaneous effect on output and prices. The logic behind this set of restrictions is that they are intended to distinguish monetary policy shocks from financial system shocks, which do not move financial stress indicators and the size of the balance sheet in the opposite direction. Hence, the movement of the ECB's balance sheet is of crucial importance to the identification scheme since other shocks, such as any news about the health of the financial system, would also lower both financial system stress and the EONIA-MRO spread, whilst also having only a lagged effect on output and prices.

2.2 Judging identification strategies

Since this paper is arguing that existing VAR studies haven't successfully identified UMP shocks I first need to introduce the criteria I will apply to

[2]Boeckx *et al.* (2017) and Gambacorta *et al.* (2014) use comparable identification schemes: Boeckx *et al.* (2017) do not include the exchange rate, which is of no importance to the arguments presented here because the exchange rate is unrestricted in the identification scheme of Burriel & Galesi (2018), whilst in Gambacorta *et al.* (2014) there is also no exchange rate or a spread and equity volatility is used in place of the CISS index. The core identifying restriction remains the same across all the papers: unconventional monetary policy shocks lower financial stress indicators and increase the size of the balance sheet.

Table 1: Identification scheme used by Burriel & Galesi (2018)

	Balance sheet shock		Interest rate shock	
	On impact	One month after	On impact	One month after
Total assets	+	+	0	?
EONIA-MRO	-	-	?	?
CISS index	-	-	?	?
MRO rate	0	?	-	-
Output	0	?	0	+
Prices	0	?	0	+
Exchange rate	?	?	?	?

evaluate the estimates reported in the literature. Boeckx *et al.* (2019) argue for two criteria for judging an identification strategy. Firstly, "an identification strategy should be evaluated by the plausibility of the restrictions" and, secondly, "the identified shocks should be consistent with the narrative of the conduct of policy of the period." It is worth noting that neither of these criteria rely on evidence from impulse response functions. That's because, in principle, even if a given model produces very plausible impulse response functions, it doesn't rule out that some other shock or some weighted average of other shocks produces the impulse response functions rather than the shock of interest. Therefore, just observing responses theoretically consistent with UMP shocks doesn't tell us if they are responses to UMP shocks or to something else.

In principle, this is not a new insight. As such, one would expect VAR studies to provide convincing additional evidence to support their claims that they have correctly identified UMP shocks. With the exception of Boeckx *et al.* (2017), that is not the case.

3 Econometric issues

The SVAR model given by Equation (2) requires that the data meets a number of requirements to allow successful isolation of exogeneous variation. This section details various ways in which the conduct of monetary policy in the euro area violates these requirements.

3.1 The foresight problem

Most of the balance sheet policies of the ECB were announced in advance. The clearest example of the balance sheet being announced in advance was the Asset Purchase Programme (APP). When the ECB announced the APP in January 2015 it announced that €60bn of assets would be purchased each month until at least September 2016. Therefore, the balance sheet changes in the months following the January announcement were highly predictable in advance to agents in the real economy. Since agents know policy will be expansionary in future that will change behaviour today.

It's not just the APP that was announced in advance. The Long Term Refinancing Operations (LTROs), the Targeted Long Term Refinancing Operations (TLTROs) and Covered Bond Purchase Programmes (CBPPs) were also announced in advance. For example, the ECB press release of 7 May 2008 announced LTRO allotments in June, September and December 2008 and the ECB press release of 8 December 2011 announced both December 2011 and February 2012 LTRO allotments. There is ample evidence from financial markets that market participants react to the announcements (see Beirne *et al.* (2011), Falagiarda & Reitz (2015), Szczerbowicz *et al.* (2015), Krishnamurthy *et al.* (2017) or Crosignani *et al.* (2019), for examples of financial markets responding immediately to policy announcements), so any empirical approach needs to allocate the shock to the period it was announced rather than the period it was implemented to be able to correctly link the changed behaviour to the shock.

Looking at Equation (2) it is clear to see how unlikely it is that a VAR model without a moving average term in the error will correctly identify a shock that has zero impact on the balance sheet for many months. First off, the immediate impact of a shock is given by the B matrix. Restricting the immediate impact of a shock to be zero is standard practice in VAR modelling and is uncontroversial. However, to achieve a zero response in the second period imposes an extra restriction on the system, namely, that $A_1 \Delta y_t = 0$ for the balance sheet, where Δy_t indicates the response of the system in the period of the shock. Since the A_is are unrestricted reduced form parameters, this requires the coincidence that the restrictions imposed on the B matrix that achieve a zero response in period zero also achieve this in the next period. And the next period too and a number of subsequent periods too until, finally, in the month of implementation, the response has to jump from zero to some large positive number. Suppose, as is the case with the VAR models discussed here, that the number of lags employed is less than the maximum number of months, j, a policy was announced in advance.[3] This requires that $A(L)\Delta y_{t+j-1} = 0$ but $A(L)\Delta y_{t+j} \gg 0$. Since Δy_{t+j-1} and Δy_{t+j} are likely to be very similar (for the balance sheet they are identical, since they are both restricted to be zero if the foresight is correctly modelled), it is intuitively highly unlikely that these conditions will be met. This is not a new issue in econometrics: it is already well-known that a system where movements in the policy variable were predictable in advance creates an equilibrium with a non-fundamental moving average representation (Hansen & Sargent (1991)). In principle, this means a constant parameters VAR will never recover the true shocks even in an infinitely long sample (Giannone & Reichlin (2006)) and failing to take this into account in a VAR framework leads to biased estimates of the effects of policy. In fact, this is essentially the same problem as the fiscal foresight problem in empirical analyses of the effects of fiscal policy (Leeper *et al.* (2013)) and is driven by the information set of the econometric model differing significantly from the information set of economic agents in the economy under investigation. In terms of the discussion of the basic VAR model above, $A(L)y_{t-1}$ is not adequately capturing the predictable movements in the balance sheet.

What makes this particularly problematic in the euro area is that the lag between announcement and implementation varies varies from policy decision to policy decision. Some were announced in the same month as the eventual balance sheet changes took place, some one month in advance and others even six months in advance. Standard approaches to identification through the contemporaneous covariance matrix do not allow the researcher to capture this complex information structure - the best a constant parameters VAR can hope to achieve is to capture the average lag between announcement and implementation. In that case, they will not recover the true unexpected component of policy. Looking at the responses in the papers listed in the introduction, all of the balance sheet responses start with a large statistically significant response in the period of the shock which decays back towards zero over the following months. That's clearly incompatible with these policies being announced in advance since, in that case, they should at least start

[3]The papers discussed here allow, but do not impose, that the balance sheet responds with a lag to a UMP shock. For example, Boeckx *et al.* (2017) impose that the response of the balance sheet is non-negative for the first two months and Lewis & Roth (2019) for four months. Even so, none of these is sufficiently long to be able to accurately model the longest leads for the ECB policy announcements. For example the December 2008 LTRO was announced a full 6 months in advance and the APP announcement was for €60bn each month indefinitely. Even if they were sufficiently long, as the discussion above made clear, simply allowing for a zero response in the first months is not going to be enough to correctly capture the true shocks.

close to zero before increasing on implementation. Therefore, none of these existing studies have adequately dealt with the foresight problem.

3.2 Fixed rate tenders with full allotment

In order for the B matrix to fulfill its role in identification, the reduced form errors must contain some variation due to the shocks of interest. Otherwise, no decomposition of the reduced form errors will recover the true UMP shocks. Unfortunately, this is the case for most of the policies implemented in the euro area prior to the start of the Asset Purchase Programme in 2015. Since 15 October 2008, all liquidity providing operations of the ECB have been undertaken as fixed rate tenders with full allotment (FRTFA). This includes the MROs, the LTROs, the TLTROs and the US dollar liquidity funding provided repeatedly by the ECB. The reduced form errors contain no information about unexpected variation in money supply because fixed rate tenders with full allotment work by:

1. The ECB announces an interest rate at which counterparties can borrow,
2. Counterparties submit bids for how much they want to borrow,
3. The ECB supplies whatever quantities counterparties have asked for, subject to qualifying collateral.

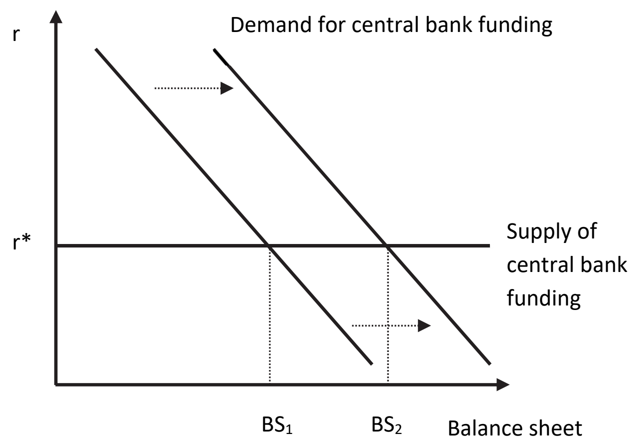
In other words, the ECB allows its balance sheet to respond perfectly elastically to the demand for credit. As Figure 1 shows, that means that all variation in the balance sheet as a consequence of these policies reflects shifts in the demand curve.[4]

Boeckx *et al.* (2017) argue that this variation in demand might be stimulated by changes in monetary policy, so can be labeled as a monetary policy shock. Specifically, euro area banks might demand a lot of LTRO financing

[4]Due to this problem, Lewis & Roth (2019) attempt to identify UMP shocks in a non-FRTFA component of the balance sheet. Nonetheless, they still do not account for foresight or the presumably large structural break that occured when the APP started in 2015 and the IRF they report for their subcomponent of the balance sheet starts close to it's maximum value, which as the previous section argued is incompatible with these policies being announced in advance.

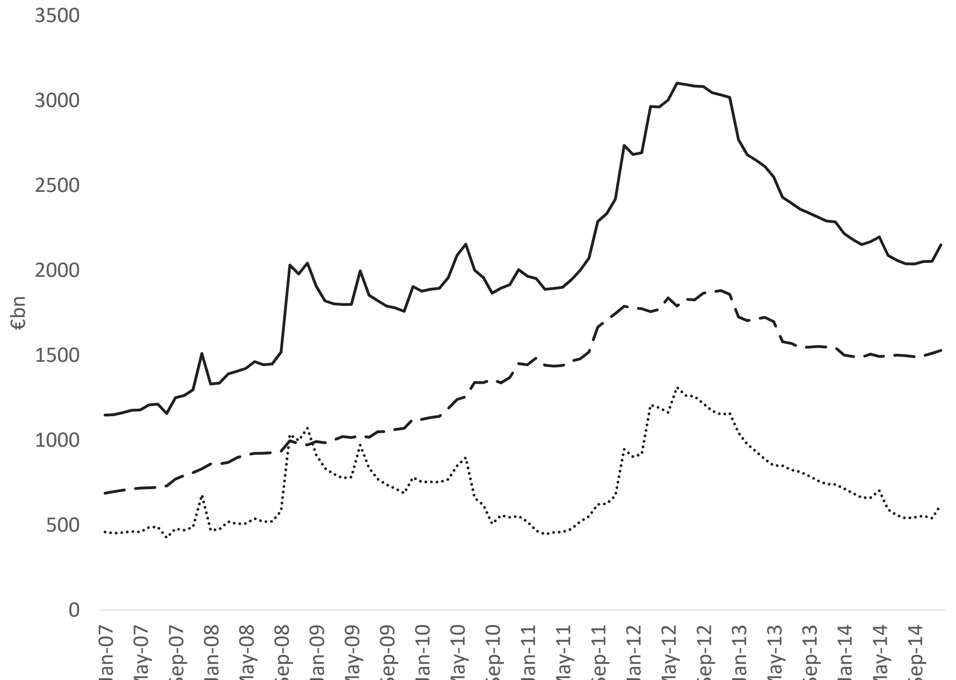
simply because it is available or because the ECB has relaxed collateral requirements. However, that's just another way of saying the announcement is the shock (foresight problem), and that the announcement changes behaviour (structural break). Another way in which these movements in the balance sheet could reflect monetary policy decisions is if the interest rate was itself a surprise. However, for both the LTROs and the TLTROs the policies were announced in advance along with their interest rates (the average of the MRO rates over the lifetime of the loan). In all likelihood, the same can be argued for the MROs: given that the lower bound problem is well known and given the amount of information the ECB communicated about the expected path of short term rates, it's a reasonable approximation to argue that the MRO rate was also known in advance in most of the months in this sample period. As such, the observed variation in the balance sheet due to the FRTFA policies that could be attributed to monetary policy decisions is predictable and subject to the foresight problem, so should not be ending up in the error term anyway. Therefore the unpredictable variation in the size of these balance sheet items contains no information about the supply side for LTROs and TLTROs, and the MROs contain very little, if any. As such, the major UMPs of the ECB do not produce reduced form errors in the balance sheet equation that contain a recoverable structural shock.

Figure 1: Under fixed rate tenders with full allotment variation in the size of the balance sheet reflects demand shocks, not supply shocks



Nonetheless, FRTFA policies are not the only policies of the ECB, so maybe UMP shocks can still be recovered from the aggregate balance sheet that arise in the other components. However, as Figure 2 shows, the FRTFA policies account for the vast majority of variation in the ECB's balance sheet prior to the start of the APP. The variance of the growth rate of the FRTFA component of the balance sheet is 24 times larger than the variance of the non-FRTFA component. Given the short sample period these studies have to work with, adding such a large noise component to the part of the balance sheet where it might be possible to identify policy shocks, means it is highly implausible that an SVAR estimated on the aggregate balance sheet will identify the correct shocks: any exogenous variation in the non-FRTFA component will be swamped by the variation in the FRTFA components. At best, the UMP shocks returned by the VAR will be a very noisy signal of the true shocks: a weighted average of true monetary policy shocks and all of the shocks impacting the economy, with a very low weight on the true shocks.

Figure 2: The vast majority of variation in the aggregate balance sheet is due to the FRTFA components



Note: The FRTFA assets are the sum of the MRO, LTRO and Claims on euro area residents denominated in foreign currency components of the balance sheet. The latter is included because a large amount of the variation in this sample period in this component is the dollar funding the ECB made available to euro area banks in FRTFA auctions.

3.3 Structural breaks

The balance sheet has yet more econometric weaknesses. Over the period 2007-2017, the ECB has undertaken a variety of different unconventional policies. To highlight the differences between them, each policy was given its own name and each was targeting a different aspect of the euro area economy. Some were conducted via interest rates (MROs, LTROs and TLTROs) while some were conducted in quantity space (CBPP, SMP, APP). As such, the true model is that each of these different policies has its own policy rule responding to different variables. Since these policies were undertaken in different proportions at different times throughout the sample period, any policy rule specified at the aggregate balance sheet level will be subject to multiple structural breaks.

ECB press releases provide clear evidence for these structural breaks during this sample period. A particularly important break occurs in mid-2009 when short term interest rates closed in on the lower bound. Monetary policy went from setting the short term interest rate to target inflation, to a situation where short-term rates are essentially fixed. Since the monetary policy rule closes a macroeconomy and ensures stability, going from an active short-term interest rate policy rule to something else can, in principle, induce a structural break in all of the reduced-form parameters of the VAR. For example, Nakata (2017) shows that the dynamic responses of a macroeconomy to any type of shock are different at the zero lower bound than away from the lower bound.

As such, our prior should be that treating the parameters of the VAR as if they were constant throughout this period will result in a misspecified model. As this subsection will show, ignoring these structural breaks will contaminate the SVAR estimate of the unconventional monetary policy shocks with random noise arising from the misspecified policy rule of the constant parameters SVAR. To see the effects of a structural break on the identified shocks, note that the policy rule of the central bank in the SVAR can be written as:

$$bs_t = c_0^{-bs} u_t^{-bs} + \hat{bs}_t + e_t^{bs}. \tag{3}$$

The observed balance sheet in period t, bs_t, is the sum of an endogenous response to economic developments, $c_0^{-bs} u_t^{-bs} + \hat{bs}_t$, and an exogenous policy shock, e_t^{bs}. The endogenous part of the SVAR policy rule is the predicted balance sheet based on information available in the previous period, $\hat{bs}_t$, plus the endogenous reaction of policy to the one-step-ahead forecast errors in the remaining variables, $c_0^{-bs} u_t^{-bs}$, where the superscript $^{-bs}$ indicates all variables except the balance sheet. Now suppose for exposition purposes that there is one single break in c_0^{-bs} in the true DGP. Letting $\hat{c}_0^{-bs}$ denote the estimated coefficients when assuming constant parameters, in the early part of the sample $c_{0,early}^{-bs} < \hat{c}_0^{-bs}$ and in the latter part $c_{0,late}^{-bs} > \hat{c}_0^{-bs}$. That is, the balance sheet didn't respond systematically to macroeconomic conditions at the start of the sample period, but it does at the end. A straightforward substitution shows that under these circumstances the estimated constant parameters SVAR error term becomes

$$e_t^{bs,SVAR} = e_t^{bs} + \left(c_{0,j}^{-bs} - \hat{c}_0^{-bs}\right) u_t^{-bs}, \quad j = [early, \, late]. \tag{4}$$

Since the one-step-ahead forecast errors u_t^{-bs} from a VAR model that satisfies the basic criteria for inference in VAR models are iid random numbers, not taking into account the switch from conventional interest rate policy to a balance sheet policy for macroeconomic stabilisation will put a random component into the 'identified' policy shocks.

Under these circumstances, any constant parameters SVAR employing the balance sheet as the measure of monetary policy will, at best, recover a noisy measure of the true shocks. In a small sample like those employed in the literature it will likely be difficult to extract clear information about the true shocks from the noisy measure we have available. The policy changes the ECB has undertaken also imply structural breaks in the $\hat{bs}_t$ term. Hence, incorrectly assuming that $A(L)$ is constant throughout the sample period will, once again, also ensure that predictable variation in the balance sheet ends up in the supposedly exogenous error term.

3.4 Other econometric issues

These are not the only econometric problems facing the researcher in this period. VAR models are notorious for the curse of dimensionality - even a relatively low dimension VAR requires a large number of parameters to be estimated. Before the Great Recession, monetary policy VARs were estimated on long sample periods with 20-30 years of observations. Not only that, in normal times a reasonable case can be made that only a handful of variables are needed to approximate the information set of a central bank whose policy changes can be described well by a Taylor-type rule: the output gap, inflation and some variable to proxy inflation expectations. In contrast, the sample periods employed in this literature are 1) very short and 2) the number of important economic variables the ECB has reacted to is significantly larger than in normal times. For example, in addition to the euro area output gap and expected inflation, during the period 2007-2017, the ECB has reacted to financial system stress at the euro area level, financial system stress at the country level and worries about the sustainability of the debt of individual euro area governments. So not only is the sample size much smaller than normal, the number of variables that enter the true policy rule of the ECB has increased dramatically. Almost inevitably, any attempt to estimate the policy rule of the ECB will be simulatenously subject to low degrees of freedom and omitted variables, with predictable consequences for the ability of these models to identify true exogenous variation in monetary policy.

Taken together, these issues make it *a priori* highly unlikely that a constant parameters SVAR estimated using balance sheet data will correctly identify the true balance sheet policy shocks in this sample period. To evaluate the practical relevance of these issues, in the following section I will replicate three SVAR models for which I have the original code and data. When discussing the problems with the balance sheet I have argued that many of these problems mean the best that can be hoped of is that the balance sheet contains a very noisy signal of the true UMP shocks. Therefore, in addition to replicating the models, I introduce a large misspecification that guarantees the measure of the stance of monetary policy is all noise: I replace the balance sheet time series with random [0,1] numbers.

4 Evidence presented in the existing literature

The previous section detailed a number of reasons it is highly unlikely to be able to successfully identify UMP shocks in this type of model, this section uses original code and data from three published papers to show they haven't successfully identified UMP shocks. I start by reexaming the impulse response function evidence presented in these papers. With the exception of Boeckx *et al.* (2017), all of the papers listed here report just the identification scheme and the impulse responses. That plausible impulse response functions are a necessary but not sufficient condition for successful identification is not a new idea, but it appears it is one that needs repeating. The first part of this section shows that it is easy to get very similar and equally plausible impulse response functions by doing things no-one would ever claim would identify monetary policy shocks: replacing the balance sheet with random numbers or assuming expansionary policy shrinks the balance sheet. In the second half of this section I take a closer look at the purported UMP shocks reported by Boeckx *et al.* (2017) and show that they contain clear evidence of the foresight problem and that they are consistent with being money demand rather than money supply shocks.

4.1 Impulse response functions: Boeckx *et al.* (2017)

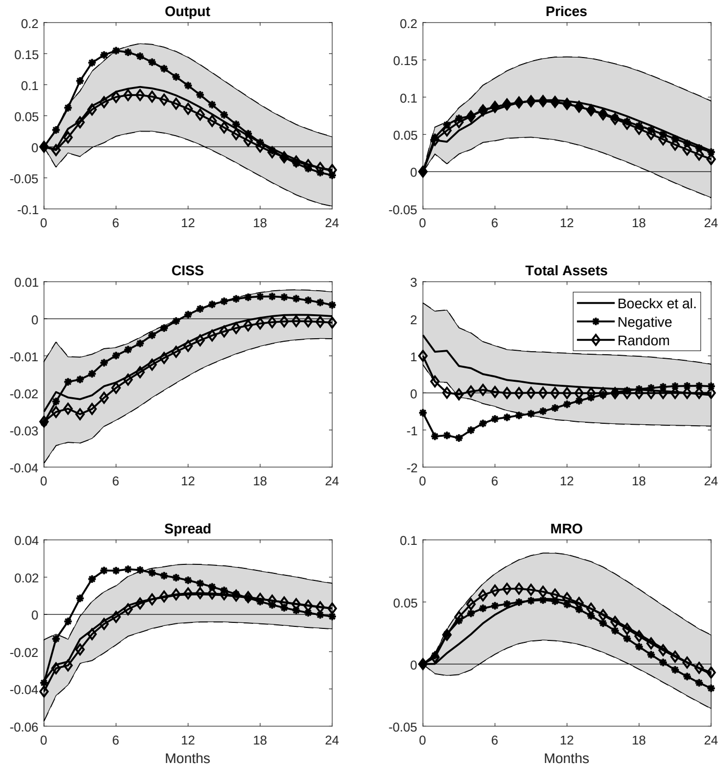
Figure 3 reproduces the impulse response functions from Boeckx *et al.* (2017)[5] estimated over the sample period January 2007 to December 2014. Four of the six responses seem plausible: output and prices follow hump-shaped responses, whilst financial system stress and the spread are affected immediately before slowly returning to baseline. Given that for about half the sample period short term interest rates were at what was thought to be the lower bound, it is hard to judge whether the hump-shaped response of the MRO rate is plausible or not - for half of the sample period our prior should be a flat line. In any case, this is consistent with an unmodeled structural break. We know that in the second half of the sample period the response should be flat - if there was no structural break then it must also be flat in the first half and, therefore, in the whole sample period too. What's even more clear than the MRO response is, given how many of the UMPs were announced in advance, the response of total assets contradicts one of the few things we know for certain: the surprise component of monetary policy should have larger effects after 3 months than on impact. In contrast, the IRF of Boeckx *et al.* (2017) has the largest value on impact, before monotonically returning to the baseline.

Figure 3 also presents the IRFs from two nonsense identification schemes: one were expansionary UMP shocks are assumed to shrink the balance sheet and one where the Total Assets data has been replaced by random [0,1] numbers.[6] These are interesting alternatives. We know for certain that expansionary balance sheet policies expand the size of the balance sheet, so by assuming the opposite there is no single structural shock that will meet these requirements. As such, the zero and sign algorithm is forced to find combinations of the true structural shocks that will produce responses with the restrictions I impose. In other words, an unidentified weighted average of the true structural shocks. Replacing the size of the ECB's balance sheet with

[5]Using the original RATS code and data, kindly made available by the authors.

[6]The random responses in are the mean IRFs from ten different random $[0, 1]$ time series.

Figure 3: Impulse response functions in Boeckx *et al.* (2017) compared to negative and random number identification schemes



Note: The grey bands represent 68% confidence bands from the original specification.

random numbers ensure that the variable measuring the stance of monetary policy contains only noise. That implies that the IRFs returned by the random numbers scheme is driven by the restrictions on the other variables in the system. In principle, there are many shocks (and many combinations of shocks) that could lower financial system stress and the spread, so the IRFs returned are responses to an unidentified weighted average of the true shocks impacting the economy. Both nonsense identification schemes give equally plausible IRFs: output, prices and the MRO rate follow similar hump-shaped responses, whilst the CISS response returns gradually to baseline just like the original IRFs. The major difference is the response of the spread under the negative identification scheme, which returns to zero more quickly than the other two and has a more pronounced overshoot. However, it's hard to argue this is particularly implausible given the width of 90% or 95% confidence intervals.

Reexamining the impulse responses of Boeckx *et al.* (2017) there is clear evidence in the total assets response that the foresight problem hasn't been resolved and evidence of an unmodeled structural break in the MRO rate response. Moreover, the arguments put forward above tell us that a constant parameters SVAR estimated using balance sheet data will not correctly identify exogenous variation in monetary policy - it will return an unidentified weighted average of all the shocks impacting the euro area economy. When I impose nonsense identification schemes that guarantee I get an unidentified weighted average of the true shocks I get essentially the same IRFs.

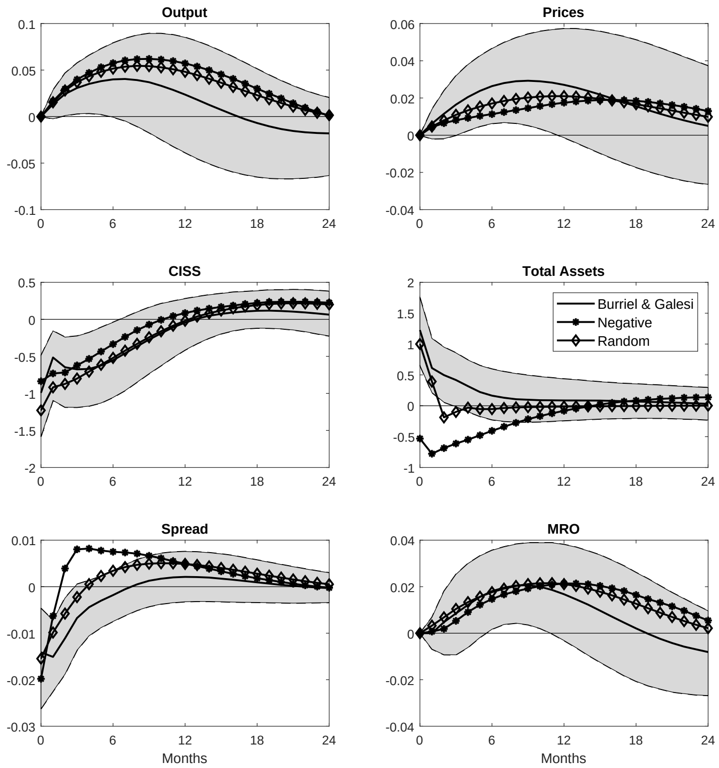
4.2 Impulse response functions: Burriel & Galesi (2018)

Figure 4 reproduces the GVAR responses to an unconventional monetary policy shock reported in Burriel & Galesi (2018)[7], estimated over the sample period January 2007 to September 2015. This sample period extends that of Boeckx *et al.* (2017) to include the start of the APP, which massively increased the size of the ECB's balance sheet and is clearly subject to the foresight problem. Moreover, it is an important structural break since the bulk of monetary policy shifted from interest rate policies to explicitly quantity policies. Nonetheless, these responses are qualitatively similar to the SVAR responses of Boeckx *et al.* (2017): both output and prices follow a hump-shaped increase with peaks occuring 8-12 months after an expansionary shock, whilst the other variables also show qualitatively similar responses. Once again, the total assets response is inconsistent with being announced in advance and the MRO rate response isn't flat, which we know it should be at the end of this sample period.

As with the Boeckx *et al.* (2017) responses, Figure 4 also presents the impulse responses from the two nonsense specifications. Again these nonsense identification schemes produce equally plausible impulse response functions, which is once again consistent with the original responses being responses to an unidentified weighted average of all the structural shocks rather than to correctly identified UMP shocks.

[7]Using their original MATLAB code downloaded from the European Economic Review website.

Figure 4: Impulse response functions in Burriel & Galesi (2018) compared to negative and random number identification schemes

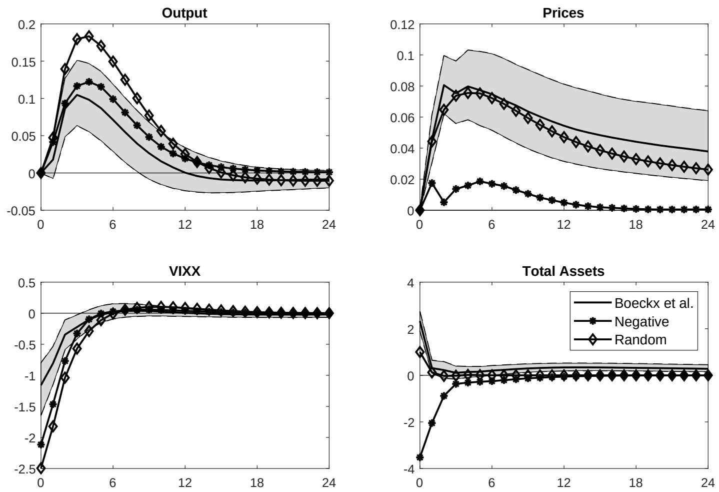


Note: The grey bands represent 68% confidence bands from the original specification.

4.3 Impulse response functions: Gambacorta *et al.* (2014)

Gambacorta *et al.* (2014) employ a very similar identification scheme with a panel of 8 economies over the sample period January 2008 to June 2011. The panel VAR model contains four variables: the natural logarithm of real GDP, the natural logarithm of the CPI, the natural logarithm of the level of central bank assets and the level of equity market volatility in each economy. The identification scheme used by Gambacorta *et al.* (2014) is similar to the schemes used by Boeckx *et al.* (2017) and Burriel & Galesi (2018): an unconventional monetary policy shock increases the size of the central bank balance sheet and lowers financial market volatility, but has no contemporaneous effect on output and prices. Since Gambacorta *et al.* (2014) use the mean group estimator, they actually estimate separate single country VAR models for each economy, before taking the mean across all eight models to produce their panel impulse response functions. As such, they also present results for a single country model of the euro area.

Figure 5: Impulse response functions in Gambacorta *et al.* (2014) compared to negative and random number identification schemes



Note: The grey bands represent 68% confidence bands from the original specification.

Figure 5 reproduces the euro area impulse responses to an unconventional monetary policy shock in Gambacorta *et al.* (2014).[8] The responses are comparable with the results obtained in Burriel & Galesi (2018) and Boeckx *et al.* (2017): both output and prices follow a hump-shaped increase, although the peaks occur at 3 months after an expansionary shock, which is earlier than the results from Burriel & Galesi (2018) and Boeckx *et al.* (2017). Once again, the total assets response is at odds with being announced in advance: it starts at its peak value and monotonically returns to baseline. In Figure 5 I also report responses using the two nonsense identification schemes. Once again, the impulse response functions display the same shapes as the original specification responses. For prices the magnitudes of the random number response are almost identical, whilst for output the random number response lies outside the 68% bands of the original specification. For the negative balance sheet response identification scheme, the response of output

[8]We use the authors original RATS code for replication.

displays the same hump-shaped response as the original and the magnitudes are very similar. In contrast, the price level response of the negative response identification scheme lies a considerable distance below the bands of the original specification. However, given the persistence of low inflation in the euro area it's hard to argue that the response isn't plausible. The ECB has spent much of the last decade with inflation considerably below target and hasn't succeeded in raising it.

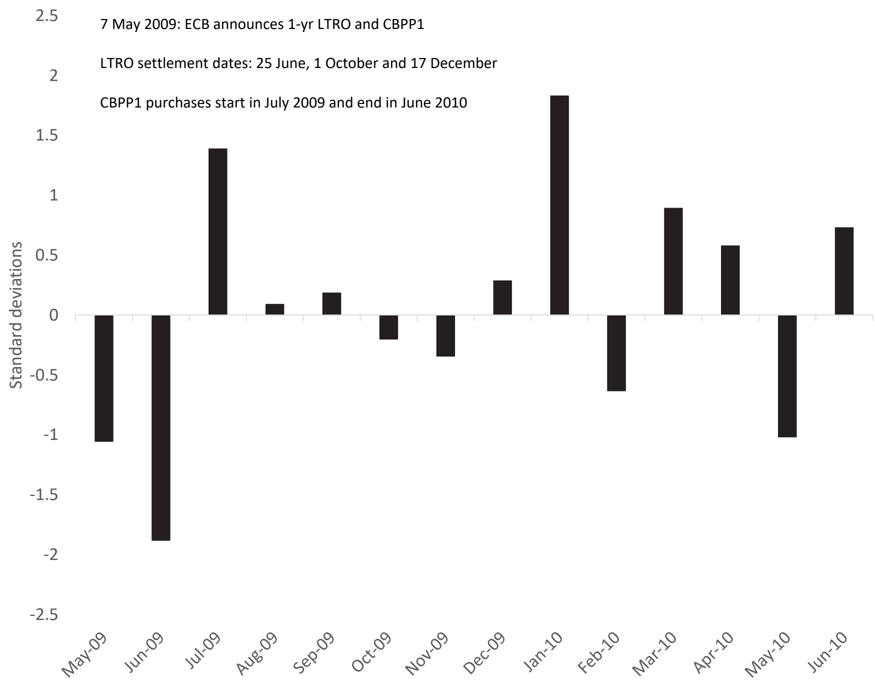
Furthermore, that the Gambacorta *et al.* (2014) IRFs show the biggest differences with the nonsense identification schemes is not a surprise. This paper uses by far the shortest sample period, so is more sensitive to individual data points than the other two.

4.4 Shocks in Boeckx *et al.* (2017)

Only one of the papers listed here attempts to provide evidence that UMP shocks have actually been identified. Boeckx *et al.* (2017) produce a plot of the sum of their purported UMP shocks and compare this to a narrative benchmark of the major monetary policy changes during the period 2007-2014. Boeckx *et al.* (2017) make two questionable choices: they compare their shocks to the implementation of the policies rather than the announcements of these policies and, even though the plausibility of the identified shocks depends on the whether the shocks occur in the correct months, Boeckx *et al.* (2017) present the sum of the shocks rather than shocks themselves. In this section I reevaluate the purported UMP shocks of the Boeckx *et al.* (2017) model by comparing the actual shocks with the announcement dates of various policies. In the appendix I show that even if you ignore the foresight problem and use the implementation dates as the narrative benchmark, the Boeckx *et al.* (2017) model produces purported UMP shocks that have the wrong sign in 11 out of the 13 narrative periods they select and small shocks in the other 2. Even their own benchmark is evidence against these being UMP shocks.

To evaluate whether the identified shocks from Boeckx *et al.* (2017) do accurately capture monetary policy shocks I focus on three key periods. The first period starts in May 2009 and covers the 1 year LTROs and the first Covered Bond Purchase Programme (CBPP), both of which were announced on 7 May 2009 and only started being implemented in June and July. The second period starts in October 2011 and covers the 1 year LTROs and

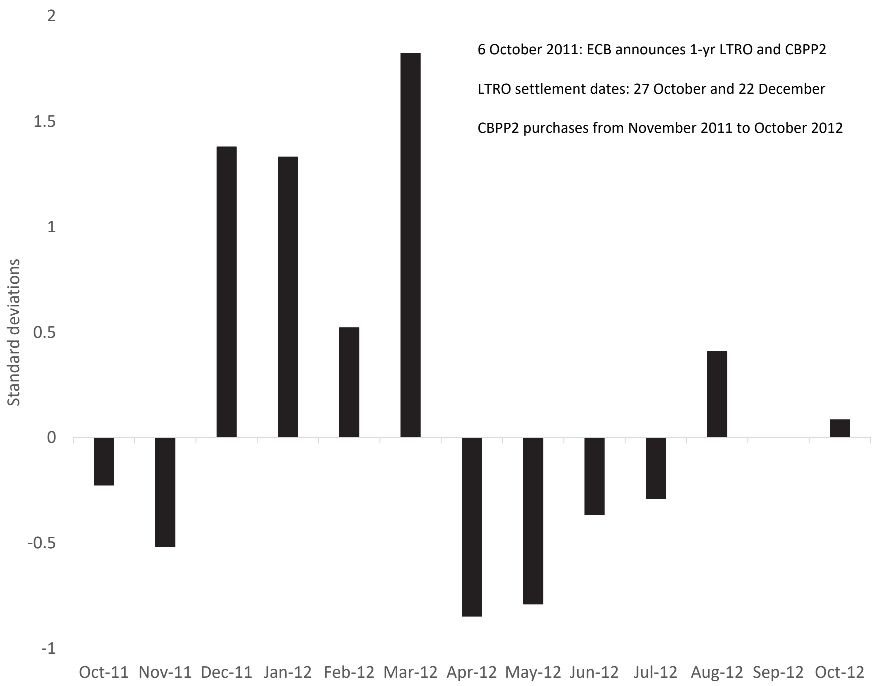
Figure 6: Purported UMP shocks from Boeckx *et al.* (2017) following the 7 May 2009 announcement of 1-yr LTROs and the CBPP1



the second CBPP, both of which were announced on 6 October 2011 to be implemented starting in October and November. The third period starts with the 5 June 2014 announcement of the TLTRO programme and also covers the first months of the third CBPP, which was announced on 4 September 2014. Both of these programmes were also announced in advance: the TLTROs were settled on 18 September and 11 December 2014, whilst the third CPBB purchases started a month after the announcement in October 2014.

As Figures 6 - 8 show, three out of four announcement months are associated with negative UMP shocks in the model of Boeckx *et al.* (2017). The announcements of these major expansionary policies are labeled as contractionary by the SVAR. The only positive UMP shock on announcement is a small positive shock in September 2014 when the third CBPP was announced. However, this announcement occurred in the same month as a TLTRO auction which expanded the balance sheet considerably. As such, this SVAR shock probably reflects the balance sheet changes caused by this previously announced TLTRO auction. As I described above, many authors have found significant financial market reactions in the expected direction to these an-

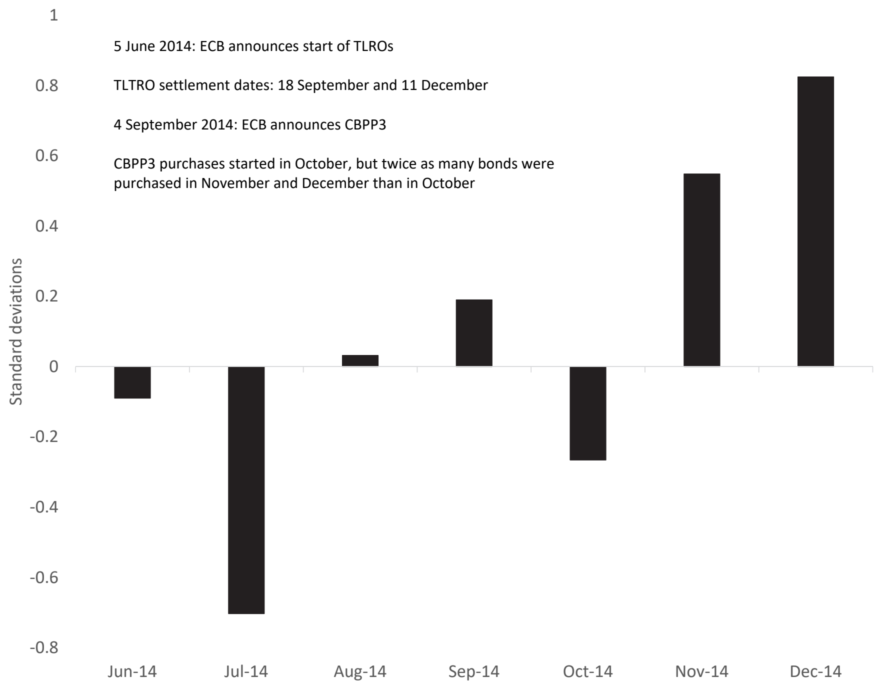
Figure 7: Purported UMP shocks from Boeckx *et al.* (2017) following the 6 October 2011 announcement of 1-yr LTROs and the CBPP2



nouncements. In contrast, because the SVAR has labeled these periods as contractionary UMP shocks, the SVAR is saying that equity volatility and the spread were increased by monetary policy in these periods. Therefore, the reductions in these variables observed on announcement by previous papers must have been attributed to non-monetary policy shocks by the SVAR.

Moving forward from the announcements, the figures often do show large positive shocks in the periods these policies were implemented. That's exactly what you'd expect to see if the model was suffering from the foresight problem: nothing positive on announcement followed by positive shocks on implementation. Moreover, for the LTROs and the TLTRO auctions there is evidence consistent with these shocks being demand driven. In the periods shown here, multiple auctions were announced for each LTRO and TLTRO. Each qualifying counterparty knew that if they didn't ask for enough funding at the first auction, they could always ask for more at the subsequent auctions. Given that more information about the state of the economy arrives as time goes by, there was an option value to waiting until the last of each announced series of auctions. In each of the figures, the final FRTFA auction

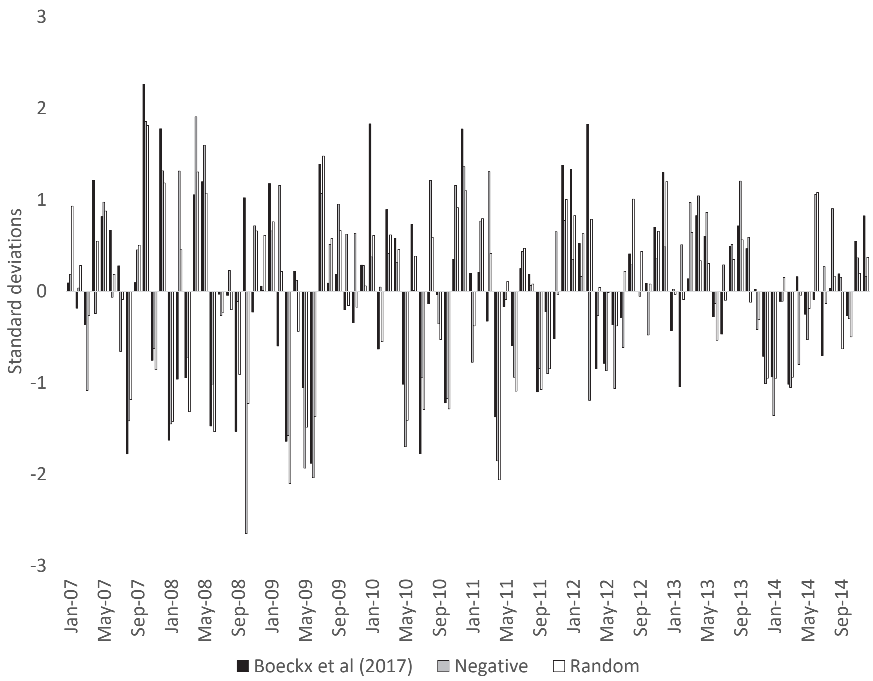
Figure 8: Purported UMP shocks from Boeckx *et al.* (2017) following the 5 June 2014 announcement of the TLTRO programme and the announcement of CBPP3 on 4 September 2014



is the largest positive shock. Although the number of observations is small, this is entirely consistent with these being money demand shocks, not money supply shocks.

Analysis of the shocks also sheds more light on why the impulse responses are so similar. The correlation coefficient between the Boeckx *et al.* (2017) shocks and the shocks from the negative identification scheme is 0.56, whilst with the random shocks it is 0.79. Since the same A and B matrices determine both the mapping from the observables to the structural shocks and the IRFs, the more similar the shocks, the more similar the IRFs. Looking more closely at the shocks, which are shown in Figure 9, there are only 3 months were there are big differences between the shocks: February 2008, October 2008 and February 2013. The largest of these shocks in the Boeckx *et al.* (2017) specification was October 2008. Clearly, October 2008 was an important month, but it is only the 26th largest shock by magnitude identified by the Boeckx *et al.* (2017) SVAR out of 96 shocks. The Boeckx *et al.* (2017) and the nonsense identification schemes largely agree on the 25 most impor-

Figure 9: Purported UMP shocks from Boeckx *et al.* (2017) are highly correlated with negative and random identification scheme shocks



tant shocks in this sample. In any case, we know these differences are not important for the magnitudes and shapes of the IRFs, because the previous section showed how similar the IRFs from the two nonsense identification schemes were to the Boeckx *et al.* (2017) IRFs.

It is also possible to compare the SVAR shocks to other estimates of monetary policy surprises over this sample period. A high correlation between the SVAR shocks and other measures should reassure us that the SVARs have identified exogenous variation in monetary policy. Whilst Sims (1998) shows a low correlation doesn't necessarily imply the SVAR shocks are not UMP shocks, it does open up that possibility. By allowing euro area counterparties to borrow for one and three years at the average MRO rate, the LTROs and TLTROs were effectively interest rate policies aimed at longer maturity interest rates. As such, it would be natural to compare the SVAR UMP shocks with movements observed in longer term interest rates when monetary policy announcement were made in this period. Kerssenfischer (2019) provides a measure of the unexpected movement in 2-year German bond yields due to monetary policy in a small window around ECB announcements. In contrast to the correlations of 0.56 and 0.79 with the negative and random identification scheme shocks, the correlation between the SVAR UMP shocks of Boeckx *et al.* (2017) and the monetary policy surprises identified by Kerssenfischer (2019) over the whole sample period is negative: -0.04. If I limit the sample period to start in July 2009 so that it coincides with the period in which short term interest rate policy has been limited by the lower bound, the correlation is still tiny: 0.05.

To explain the low correlation between VAR shocks, Sims (1998) showed that although the correlations between the identified shocks from different VAR models were low, they did at least agree on the periods when the biggest shocks took place. Furthermore, Sims argued that if VAR models are estimated on periods without large shocks the correlations between the VAR shocks could become very low, without implying they weren't monetary policy shocks. That's not the case for this sample period: it would stretch credibility to argue that the low correlation between the SVAR UMP shocks and the Kerssenfischer (2019) surprises are due to this being a period with unusually small shocks. So do the different estimates of monteray surprises agree at least on the biggest shocks? Of the 20 largest SVAR shocks from Boeckx *et al.* (2017) only four appear in the list of 20 biggest shocks from Kerssenfischer (2019). Since there are 96 months in the Boeckx *et al.* (2017) sample period the 20 largest represent about 20% of those observations. Therefore, four of these also appearing in the list from Kerssenfischer (2019) is what you would expect if there was no agreement whatsoever about when the biggest shocks took place. In reality, this low correlation is perfectly consistent with the foresight problem – 2-year interest rates move on announcement whilst the balance sheet only moves months later when the policies are implemented.

The evidence I have presented in this section shows that the SVAR UMP shocks from the Boeckx *et al.* (2017) model typically have the wrong sign on announcement, but have the correct sign and are large when policies were implemented, even though these effects were highly predictable. I have also shown that there is no correlation between the SVAR UMP shocks and the surprise movements in 2-year German government bond yields identified by Kerssenfischer (2019). The SVAR UMP shocks do, however, agree with estimated shocks from models we know *a priori* will not return the correct shocks. These findings are consistent with the theoretical problems I outlined in Section 3. Hence, these models have not identified exogenous UMP shocks and tell us nothing about the efficacy of these policies in the euro area.

5 Conclusion

In this paper, I have detailed several reasons why VAR models estimated on balance sheet data will not correctly identify UMP shocks. These econometric deficiencies in the balance sheet data imply SVAR shocks will be unidentified weighted averages of all the shocks impacting the euro area economy in this period, not UMP shocks. By replicating the results reported in Burriel & Galesi (2018), Boeckx *et al.* (2017) and Gambacorta *et al.* (2014) I have shown that nonsense identification schemes guaranteed to return an unidentified weighted average of all the shocks produce very similar and equally plausible IRFs to those reported in the literature. That finding alone shows that authors need to do more than simply report plausible IRFs to show that they have correctly identified exogenous variation in monetary policy. As I have highlighted here, very few VAR analyses provide any extra evidence to prove they have plausibly identified exogenous variation in monetary policy. This leaves the reader of these studies unable to determine whether the reported impulse response functions are reactions to monetary policy shocks or are meaningless responses to unidentified errors.

In this paper I have also reexamined the extra evidence provided by Boeckx *et al.* (2017), the only paper that provides any additional evidence concerning the purported exogenous variation they have isolated. The reexamination shows clear signs of the econometric shortcomings I described in this paper. There is no evidence these VAR models have correctly identified UMP shocks and the evidence there is suggests they have not successfully identified them. As such, these papers provide no evidence to support the claim that UMP has been successful in stabilising the euro area economy.

That still leaves an important question in macroeconomics unanswered. There are other variables that can be used to measure the stance of monetary policy, such as longer term interest rates or shadow short rates (see Wu & Xia (2016), for example). However, previous research (Elbourne & Ji (2019)) has shown that using shadow short rates also fail to credibly identify UMP shocks in VAR models. Identifying monetary policy shocks through restrictions on the covariance matrix of the residuals is not the only approach. High frequency identification such as Jarociński & Karadi (2018) or Kerssenfischer (2019) appears to be a fruitful strategy that will not suffer from all of the problems listed here.

Nonetheless, alternative estimates should come with a clear warning that applies to the majority of time series approaches to identifying the effects of unconventional monetary policy. The short sample period involved necessitates using various unattractive econometric choices, such as using interpolated monthly data or assuming the transmission mechanism and central bank policy rule has remained unchanged throughout. All of these issues only make the task of identifying true policy shocks even harder. All we can really conclude is that the evidence from these VAR models does not support the claim that unconventional monetary policy has been successful in affecting real economic activity. If we are uncertain about the efficacy of UMP, policy makers should be prepared to use other tools to manage the euro area business cycle.

Appendices

A Nonsense identification schemes

Tables 2 and 3 show the nonsense identification schemes used in section 4, based on the GVAR of Burriel & Galesi (2018).

Table 2: Nonsense identification scheme 1: iid noise as policy instrument

	Balance sheet shock	
	On impact	One month after
Random [0,1]	+	+
EONIA-MRO	-	-
CISS index	-	-
MRO rate	0	?
Output	0	?
Prices	0	?
Exchange rate	?	?

Table 3: Nonsense identification scheme 2: Negative response of total assets to expansionary shock

	Balance sheet shock	
	On impact	One month after
Total assets	-	-
EONIA-MRO	-	-
CISS index	-	-
MRO rate	0	?
Output	0	?
Prices	0	?
Exchange rate	?	?

B Implementation benchmark in Boeckx *et al.* (2017)

As I have argued in the main text of the paper, comparing SVAR shocks to the implementation dates instead of the announcement dates is incorrect. Nonetheless, the narrative benchmark of Boeckx *et al.* (2017) is based on the implementation dates. For example, the first 1-year LTROs that appear in the narrative benchmark of Boeckx *et al.* (2017) were announced on 7 May 2009 and started being implemented in June. The Boeckx *et al.* (2017) narrative is marked in June 2009. That the narrative benchmark is based on implementation dates can be more clearly seen by the narrative measure for the start of the TLTROs. The TLTROs were announced on 5 June 2014 and started being implemented in September, and it is this latter month when they are marked in the Boeckx *et al.* (2017) narrative benchmark.[9] Hence, if the narrative benchmark of Boeckx *et al.* (2017) aligned well with the SVAR UMP shocks, this would not be evidence that the SVAR has correctly identified the right shocks. It would simply highlight that the SVAR was suffering from the foresight problem.

Furthermore, the list of events that make up the narrative benchmark of Boeckx *et al.* (2017) are sometimes confusing. Firstly, the early repayment of the 3-year TLTROs is explained by Boeckx *et al.* (2017) as demonstrating "a desire by counterparties to avoid stigma attached to using the LTROs by signaling improvements in their funding conditions (p.313)". This is clearly a description of a change in the demand for central bank funding, not a change in the supply. Nonetheless, it still makes the list of the 13 most important monetary policy changes in this period and is one of the periods the UMP shocks captures well, according to Boeckx *et al.* (2017). Similar confusion surrounds whether the start of the TLTROs should be in the narrative benchmark. Boeckx *et al.* (2017) claim their model can accommodate policies announced in advance because they allow (but do not impose) the contemporaneous and following month response of the balance sheet to be zero. This rules out policies announced further in advance. In a follow up article, Boeckx *et al.* (2019) write "balance sheet policies that were announced more than 1-2 months in advance are captured by the other innovations in the VAR model (p.7)", yet the start of the TLTROs are included in the

[9] That the 1-year LTRO is marked on implementation not announcement is just about observable in the original figure because it is just before one of the grey markers on the x-axis, which mark December/January and June/July every year. The TLTRO mark is clearly after the June/July marker in 2014.

narrative benchmark despite being announced 3 months in advance. It is also one of the policies that Boeckx *et al.* (2017) claim is well captured by their UMP shocks, even though they also say it should be captured by other shocks in the SVAR.

Even ignoring the problems with their narrative benchmark, the remainder of this appendix shows that the SVAR shocks of Boeckx *et al.* (2017) don't even capture their own narrative benchmark well. Moreover, by presenting the sum of the shocks instead of the shocks themselves, the evidence looks more convincing than it is.

B.1 The shocks versus the sum of the shocks

The last term in Equation (2) is Be_t: the model and identification hinges on the shocks in period t. Therefore a logical criteria is that the SVAR shocks occur in 1) the correct periods, 2) have the correct signs and 3) are larger than one standard deviation. The last criteria arises naturally from the idea that a narrative benchmark exists - if an episode is noteworthy enough to make the narrative benchmark it should be one of the biggest shocks in the sample.

However, in Boeckx *et al.* (2017) it is not the shocks themselves that are compared to the narrative benchmark but the sum of the shocks. Given that the goal is to evaluate whether the shocks occur in the correct periods with the correct signs, presenting a transformation of the shocks rather than the shocks themselves makes this comparison less easy. As this section shows, by construction the sum of the shocks from any SVAR estimated on balance sheet data will bear some resemblance to the raw balance sheet data and might, therefore, look like a plausible narrative of monetary policy. In fact, the resemblance will be closer the worse the SVAR model is. By the nature of balance sheet policies, after expansionary policies are announced, the size of the balance sheet in the raw data goes up, whilst after the policies are tightened or withdrawn, the balance sheet shrinks. These ups and downs, therefore, align well with the periods in which these policies were undertaken. Hence, if the sum of the shocks resembles the raw data, it will also align nicely with these periods. These ups and downs may have been perfectly predictable or endogenous responses of monetary policy to other shocks, so they do not tell us that exogenous variation has been correctly identified.

To show why the sum of the shocks resembles the raw data, we should first note that any invertible VAR can be rewritten as a vector moving average, $y_t = \mu + C(L)\,e_t$. In other words, the variation in the raw data is a weighted sum of the past shocks. If one compares a weighted sum of all the past shocks to an unweighted sum of one of the shocks, the two will likely bear some resemblance to each other, depending on the weights and the estimated importance of the other shocks. As far as the weights are concerned, since macroeconomic data is typically persistent, it is likely that the weights in the MA representation will be close to one, which increases the *ex ante* probability that the two will be similar. The response of total assets to a purported UMP shock in Boeckx *et al.* (2017) shows exactly this persistence.

That leaves the similarity between the sum of the SVAR shocks and the raw data down to the estimated importance of other varaibles in the SVAR. In fact, the worse the SVAR is at approximating the systematic behaviour of the ECB, the more the sum of the SVAR UMP shocks will look like the raw balance sheet data. For illustration purposes, suppose that the raw balance sheet data contains a unit root and the SVAR model is so misspecified that it is estimated that the ECB does not take any of the other variables into account when setting the balance sheet. This SVAR would describe monetary policy as a series of entirely random shocks with policy makers paying no attention to economic developments, which is unarguably a very poor model of monetary policy. In this hypothetical bad model, all of the endogenous reaction of monetary policy to economic developments will end up in the SVAR UMP shock, which will be an unidentified weighted average of all shocks impacting the economy. Nonetheless, the sum of these SVAR policy shocks will exactly equal the raw balance sheet data. Since they match exactly, the sum of the SVAR shocks will go up following the implementation of expansionary policies and go down after contractionary policies. It will look like a superficially plausible narrative of what monetary policy has done. By getting the reader to focus on the ups and downs of the sum, it distracts the reader from what we are really interested in: have the unexpected changes in monetary policy been correctly identified in the correct periods?

To rectify this potentially misleading presentation, in Figure 10 I compare the underlying shocks rather than the sum to the same narrative benchmark.[10] Of the 10 expansionary episodes highlighted by the narrative mea-

[10] In Figure 3 of Boeckx *et al.* (2017) there is an extra data point for January 2015 (the sample period is January 2007 to December 2014) that has been added to the sum of the shocks, presumably to ensure that "the sum of the shocks is zero over the whole sample period" (p.312). The sum of the shocks that come out of the model is not zero and it would be strange if that were the case. Bayesian estimates are a weighted average

sure of Boeckx *et al.* (2017) 8 of the SVAR balance sheet shocks in these periods are negative. For example, in March 2008 when the ECB supplied unlimited US dollar funding the SVAR 'balance sheet policy shock' contracted the balance sheet by about 1.5% and the implementation of the 1-year LTROs in June 2009 coincides with an SVAR shock that shrunk the balance sheet by nearly 3%. The other 2 expansionary episodes coincide with shocks that are about 0.2 standard deviations. These are clearly not among the most important expansionary episodes according to the SVAR. In addition to the expansionary episodes, Boeckx *et al.* (2017) highlight 3 contractionary episodes in their narrative benchmark. All 3 contractionary narrative episodes coincide with SVAR shocks that expand the balance sheet. Out of 13 episodes deemed important enough by Boeckx *et al.* (2017) to be included in their narrative benchmark, not a single SVAR 'balance sheet shock' meets the criteria of having the correct sign and being bigger than 1 standard deviation.

B.2 The timing criteria of Boeckx *et al.* (2017)

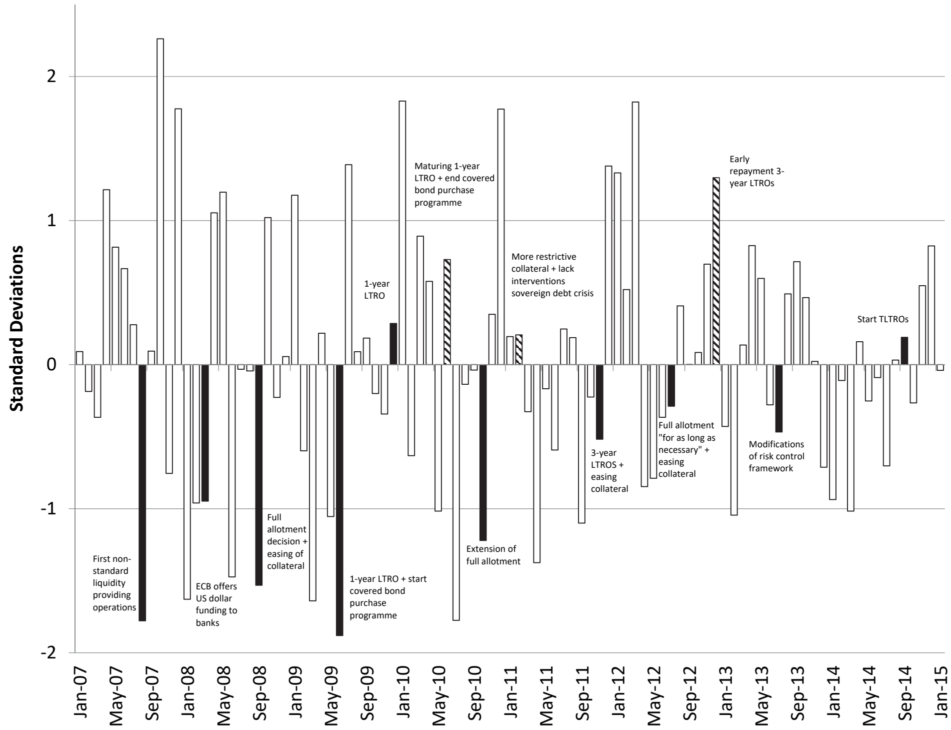
Despite the model producing incorrectly signed shocks, Boeckx *et al.* (2017) argue that their model captures the changes in monetary policy during the period and has plausibly identified exogenous variation. In a follow up article, Boeckx *et al.* (2019) argue that

> "Overall, the chart shows that major balance sheet policies launched by the ECB between 2007 and 2015 were followed by expansionary UMP shocks, while periods of more restrictive balance sheet policies or policy action were associated with contractionary UMP shocks, which indicates that the identification based on a combination of zero and sign restrictions is plausible" (p.18).

This criteria loosens the timing requirement of the narrative benchmark criteria I detailed in the previous section - it allows the SVAR shocks to occur in the months following the policy change rather than the exact month. Loosing the timing requirement is incorrect because it mixes up the predictable component of policy with the unpredictable component. That correct signed

of least squares and the prior. Abstracting from the prior for a moment, minimising the sum of squared errors does not result in errors that sum to zero. Only in the case that the estimation technique is minimising the sum of the absolute errors would the sum equal zero. Even in that case, once the prior has been added, the resultant shocks will not sum to zero since the prior introduces a bias term into the estimates.

Figure 10: The SVAR 'balance sheet shocks' of Boeckx *et al.* (2017) have the wrong sign in narrative periods



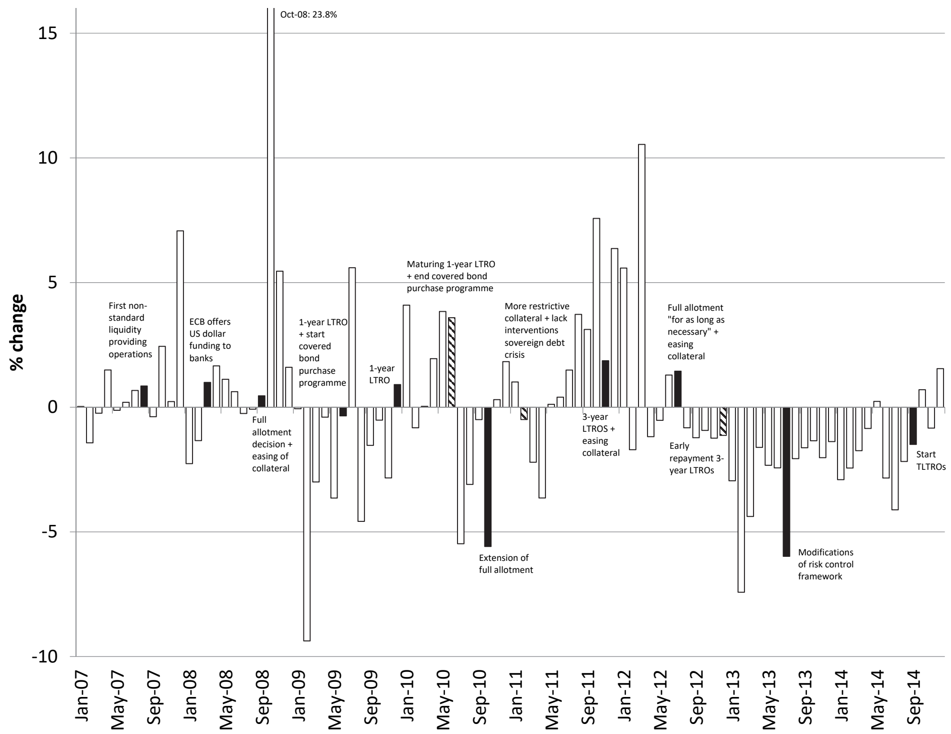
Note: The bars indicate the SVAR 'balance sheet shocks' in each period - a positive shock increases the size of the ECB's balance sheet. The months highlighted as expansionary episodes by Boeckx *et al.* (2017) are solid black, whilst the contractionary episodes are striped.

shocks systematically follow the initial implementation dates is just another way of saying they are predictable. As such, in a successful SVAR model they should be captured the $A(L)\, y_{t-1}$ component in Equation (2).

Moreover, it is a flawed criteria for claiming the model has successfully isolated exogenous variation. We know that the raw balance sheet data is a mix of a systematic endogenous response of policy to economic developments and exogenous deviations from normal policy behaviour. Therefore, we know that using the raw data will suffer from an endogeneity problem and not allow us to distinguish cause and effect. As such, any criteria for comparing a narrative benchmark with the proposed exogenous shocks should, therefore, reject the variation in the raw data. As shown in Figure 11, with the exception of the announcement that full allotment would be continued in

July 2012 and the modifications of the risk control framework in July 2013[11], the raw balance sheet data passes the same criteria applied by Boeckx *et al.* (2019) without any attempt being made to isolate exogenous variation. In fact, a really convincing model should not meet this criteria because it should attribute a large proportion of the variation in the balance sheet to the endogenous component. In contrast, not a single narrative episode coincides with a correct signed change in the raw data that is bigger than one standard deviation.

Figure 11: The detrended growth rate of the raw balance sheet data also has ‘correct signed’ changes in the months following the narrative episodes



Note: The bars are the growth rate of the Boeckx *et al.* (2017) balance sheet data after being detrended with a linear trend so that the growth rate sums to zero. The months highlighted as expansionary episodes by Boeckx *et al.* (2017) are solid black, whilst the contractionary episodes are striped.

[11]These policy changes have a less direct link to changes in the balance sheet than LTROs or TLTROs.

References

Arce, Óscar, Nuño, Galo, Thaler, Dominik, & Thomas, Carlos. 2019. A large central bank balance sheet? Floor vs corridor systems in a New Keynesian environment. *Journal of Monetary Economics.*

Beirne, John, Dalitz, Lars, Ejsing, Jacob, Grothe, Magdalena, Manganelli, Simone, Monar, Fernando, Sahel, Benjamin, Sušec, Matjaž, Tapking, Jens, & Vong, Tana. 2011 (Jan.). *The impact of the Eurosystem's covered bond purchase programme on the primary and secondary markets.* Occasional Paper Series 122. European Central Bank.

Boeckx, Jef, Dossche, Maarten, & Peersman, Gert. 2017. Effectiveness and Transmission of the ECB's Balance Sheet Policies. *International Journal of Central Banking*, **13**(1), 297–333.

Boeckx, Jef, Dossche, Maarten, Galesi, Alessandro, Hofmann, Boris, & Peersman, Gert. 2019 (June). *Do SVARs with sign restrictions not identify unconventional monetary policy shocks?* BIS Working Papers 788. Bank for International Settlements.

Burriel, Pablo, & Galesi, Alessandro. 2018. Uncovering the heterogeneous effects of ECB unconventional monetary policies across euro area countries. *European Economic Review*, **101**(C), 210–229.

Crosignani, Matteo, Faria-e Castro, Miguel, & Fonseca, Luís. 2019. The (unintended?) consequences of the largest liquidity injection ever. *Journal of Monetary Economics.*

Elbourne, Adam, & Ji, Kan. 2019. *Do zero and sign restricted SVARs identify unconventional monetary policy shocks in the euro area?* CPB Discussion Paper 391. CPB Netherlands Bureau for Economic Policy Analysis.

Falagiarda, Matteo, & Reitz, Stefan. 2015. Announcements of ECB unconventional programs: Implications for the sovereign spreads of stressed euro area countries. *Journal of International Money and Finance*, **53**, 276–295.

Gambacorta, Leonardo, Hofmann, Boris, & Peersman, Gert. 2014. The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis. *Journal of Money, Credit and Banking*, **46**(4), 615–642.

Giannone, Domenico, & Reichlin, Lucrezia. 2006. Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, **4**(2-3), 455–465.

Haldane, Andrew, Roberts-Sklar, Matt, Wieladek, Tomasz, & Young, Chris. 2016 (Oct.). *QE: The Story so far.* Bank of England working papers 624. Bank of England.

Hansen, Lars Peter, & Sargent, Thomas J. 1991. Two Difficulties in Interpreting Vector Autoregressions. *Pages 77–119 of:* Hansen, Lars Peter, & Sargent, Thomas J. (eds), *Rational Expectations Econometrics.* Boulder, CO.: Westview Press.

Jarociński, Marek, & Karadi, Peter. 2018 (Feb.). *Deconstructing monetary policy surprises: the role of information shocks.* Working Paper Series 2133. European Central Bank.

Keller, Wolfgang. 1998. Are international R&D spillovers trade-related?: Analyzing spillovers among randomly matched trade partners. *European Economic Review*, **42**(8), 1469–1481.

Kerssenfischer, Mark. 2019. *Information effects of euro area monetary policy: New evidence from high-frequency futures data.* Discussion Paper 07/2019. Deutsche Bundesbank.

Krishnamurthy, Arvind, Nagel, Stefan, & Vissing-Jorgensen, Annette. 2017. ECB policies involving government bond purchases: Impact and channels. *Review of Finance*, **22**(1), 1–44.

Leeper, Eric M., Walker, Todd B., & Yang, Shu-Chun Susan. 2013. Fiscal Foresight and Information Flows. *Econometrica*, **81**(3), 1115–1145.

Lewis, Vivien, & Roth, Markus. 2019. The financial market effects of the ECB's asset purchase programs. *Journal of Financial Stability*, **43**, 40–52.

Nakata, Taisuke. 2017. Uncertainty at the Zero Lower Bound. *American Economic Journal: Macroeconomics*, **9**(3), 186–221.

Sims, Christopher A. 1998. Comment on Glenn Rudebusch's" Do Measures of Monetary Policy in a VAR Make Sense?". *International Economic Review*, **39**(4), 933–941.

Szczerbowicz, Urszula, *et al.* 2015. The ECB unconventional monetary policies: have they lowered market borrowing costs for banks and governments? *International Journal of Central Banking*, **11**(4), 91–127.

Wieladek, Tomasz, & Garcia Pascual, Antonio. 2016 (June). *The European Central Bank's QE: A new hope.* CEPR Discussion Papers 11309. C.E.P.R. Discussion Papers.

Wu, Jing Cynthia, & Xia, Fan Dora. 2016. Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking*, **48**(2-3), 253–291.