SVERIGES RIKSBANK
WORKING PAPER SERIES

451

Long-Lag VARs

Ferre De Graeve and Andreas Westermark

May 2025

WORKING PAPERS ARE OBTAINABLE FROM

www.riksbank.se/en/research
Sveriges Riksbank • SE-103 37 Stockholm
Fax international: +46 8 21 05 31
Telephone international: +46 8 787 00 00

Long-Lag VARs*

Ferre De Graeve and Andreas Westermark

Sveriges Riksbank Working Paper Series No. 451

May 2025

Abstract

Macroeconomic research often relies on structural vector autoregressions, (S)VARs, to uncover empirical regularities. Critics argue the method goes awry due to lag truncation: short lag-lengths imply a poor approximation to important data-generating processes (e.g. DSGE-models). Empirically, short lag-length is deemed necessary as increased parametrization induces excessive uncertainty. The paper shows that this argument is incomplete. Longer lag-length simultaneously reduces misspecification, which in turn reduces variance. For data generated by frontier DSGE-models long-lag VARs are feasible, reduce bias and variance, and have better coverage. Long-lag VARs are also viable in common macroeconomic data and applications. Thus, contrary to conventional wisdom, the trivial solution to the critique actually works.

Keywords: VAR, SVAR, Lag-length, Lag truncation

JEL: C18, E37

*This paper has benefited from discussions with Paul Beaudry, Fabio Canova, Yongsung Chang, Martin Eichenbaum, Jesús Fernández-Villaverde, Jonas Fisher, Jordi Galí, Karl Harmenberg, Marco Lippi, Jean-Paul L'Huillier, Per Krusell, Ellen McGrattan, Juan Rubio-Ramirez, Federico Ravenna, Barbara Rossi, Chris Sims, Dan Waggoner, Tao Zha and seminar participants at the Atlanta Fed, Cleveland Fed, Dallas Fed, EIEF, Universitat Pompeu Fabra, KU Leuven and NORMAC. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of Sveriges Riksbank. This working paper supersedes and replaces WP No. 271. Correspondence: ferre.degraeve@kuleuven.be, andreas.westermark@riksbank.se.

1 Introduction

Structural Vector Autoregressions (SVARs) have proven to be an important tool for measuring macroeconomic regularities. Following Sims' (1980) seminal contribution Bernanke (1983), Blanchard and Quah (1989), Sims (1989, 1992), Eichenbaum and Evans (1995), Galí (1999), Fisher (2006), Beaudry and Portier (2006) and many others since have provided SVAR-based evidence for a variety of shocks and their macroeconomic effects.

Yet the SVAR method is not without its critics. Many critiques of SVARs boil down to the problem of lag truncation. In particular, while DSGE models tend to imply reduced form VAR representations with long lag-length (often infinity), when going to the data, macroeconomists invariably settle on using a very small number of lags (typically one to four quarters). Because lags are truncated, the critics show, impulse response functions (IRFs) computed using the SVAR may not correspond to those of the underlying DSGE model. Chari, Kehoe and McGrattan (2008, henceforth CKM) is perhaps the most well-known elicitation of that critique.[1]

The trivial solution to lag truncation, i.e., dramatically increasing lag-length, is unexplored. What keeps macroeconomists from using long lag-lengths is the intuition that uncertainty becomes pervasive. That is, increasing lag-length increases the number of parameters rapidly, thereby reducing the degrees of freedom and making confidence interval width prohibitively large.[2]

We show that this standard intuition is only part of the story. In the face of misspecification due to lag truncation, increasing lag-length can actually reduce uncertainty. The reason is that as truncation reduces, misspecification reduces. The reduction in misspecifcation not only leads to the well-known bias reduction, but it also reduces variance. This reduction in

[1] Others include Braun and Mittnik (1993), Faust and Leeper (1997), Cooley and Dwyer (1998) and Ravenna (2007).

[2] For instance, the literature comparing Local Projections and VARs does not consider VARs with long lags a viable approach (e.g. Li et al. (2024a)) due to high variance (e.g. Li et al. (2024b)).

variance will work against the imprecision resulting from increased parametrization. This trade-off is general: it applies to all truncated SVARs, no matter whether they are identified with short-run, long-run or other restrictions.

We show that when increasing lag-length in standard SVARs on small samples of data generated by standard DSGE models, the variance-effect of misspecification reduction often dominates the increased imprecision due to increased parametrization. The result is then almost unequivocally in favor of long-lag VARs: reduced truncation bias, more precise inference, reduced MSE, better coverage rates.

The implication is, contrary to conventional wisdom, that it is possible to estimate VARs with long lags, and hence reduce truncation bias, and still derive precise structural predictions from them.

These conclusions are not particular to simulated data generated from DSGE models. We show that long-lag VARs are also feasible in data frequently studied in macroeconomics. Specifically, we show that various well-known short-lag SVAR studies allow much longer lag representations, without uncertainty becoming prohibitively large.

The paper is organized as follows. We start by laying out a standard single-equation omitted variables argument. This provides the intuition for the effect of reducing truncation in SVAR impulse responses, where analytics are not tractable. We then assess long-lag VARs on the basis of a series of Monte Carlo experiments. We draw data from a variety of DSGE models, estimate SVARs of different (and possibly very long) lag-length and evaluate their performance. We then turn to the data and re-evaluate some well-known SVAR results on the effect of technology and monetary policy shocks. Finally, we assess the implications of our results and discuss some possible avenues for future research.

2 Misspecification

We first briefly re-state a textbook omitted variables argument, which facilitates understanding the intuition behind the general VAR results.

2.1 Some useful single-equation intuition

Consider a data-generating process

$$y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + \epsilon_t,\ V(\epsilon_t) = \sigma^2 \tag{1}$$

where a variable y is determined by two (sets of) exogenous variables, X_1 and X_2 and a shock ϵ. Now run the regression

$$y_t = X_{1t}b_1 + e_t,\ V(e_t) = s^2. \tag{2}$$

It is well-known that omission of the relevant variable X_2 leads to biased point estimates (unless $X_1 \perp X_2$):

$$E(b_1) \neq \beta_1$$

as well as an upwardly biased variance estimate (always):

$$s^2 > \sigma^2.$$

2.2 Omitted variables and truncation in VARs

The single-equation textbook result straightforwardly generalizes to VARs. It suffices to think of y as a vector of variables, X_1 as the lags the researcher includes, and X_2 as the lags not included, or truncated.

It is then immediate that a VAR, denoted by

$$\begin{aligned} Y_t &= B_1 Y_{t-1} + ... + B_p Y_{t-p} + u_t,\ E(u_t u_t') = \Sigma \\ B(L) &= B_1 L + ... + B_p L^p, \end{aligned}$$

which has $p \ll p^*$ (where p^* denotes the true lag-length) will suffer from truncation bias. The omitted variables argument above highlights why: lag truncation (or omitting relevant variables) results in a bias in the reduced form coefficients $B(L)$ *and* in the reduced form covariance matrix Σ. Any SVAR analysis has impulse responses as a function of both these reduced form objects; let

$$IRF = f(B(L), \Sigma). \tag{3}$$

Because impulse responses are a function of both $B(L)$ and Σ they will tend to become less biased if both arguments become less biased. In other words, reducing truncation reduces bias.[3]

But what do we know about variance? Recall that the intuition that keeps macroeconomics from considering long lag-lengths is that the increased parametrization (dimension of $B(L)$) leads to increased imprecision.

Though conceptually simple, equation (3) helps formalize that standard intuition. Essentially, recalling that $V(.)$ denotes variance, the intuition simply states that $V(B(L)) \uparrow \Longrightarrow V(IRF) \uparrow$ as lag-length increases. But (3) also makes clear that this argument is incomplete. In particular, it neglects that there is a second argument, Σ. Therefore, any claims about $V(IRF)$ solely based on $V(B(L))$ are only partial. Importantly, the omitted variables argument suggests a reduction in bias of the estimate of Σ, which may well contribute to a reduction in variance of impulse responses.

Equation (3) also makes clear why general statements about $V(IRF)$ are hard to make: the non-linearity of f (also across horizons) interacts with the multi-dimensionality of both its arguments, $B(L)$ and Σ. Therefore, we ascertain the balance of this trade-off by means of

[3] We merely refer to a documented tendency in DSGE models analyzed in the literature (see, for instance, CKM). From a theoretical perspective, this reduction in bias is not a certitude. Generally, bias reduction in its arguments does not guarantee bias reduction in the impulse response function. See Sims (1972) for an elicitation of a related point in terms of reduced form objects: convergence in individual point estimates (i.e. function arguments) may imply divergence of the sum of coefficients (i.e. the function itself).

a series of Monte Carlo experiments based on frequently studied models in macroeconomics.

3 Monte Carlo evidence

For each DSGE model considered, we sample data of length equal to that available in typical macro data samples ($T = 200$).[4] Given one such draw of data, we estimate VARs of different lag-lengths, calculate impulse response functions and construct confidence bands using standard methods.[5] We repeat that exercise 1000 times for each model and subsequently investigate bias, confidence interval width, mean-squared error and coverage rates.

3.1 Setup

We consider a range of models, both real and nominal, and identified with both short and long-run restrictions. More precisely, we consider estimating IRFs using long-run restrictions on data generated from CKM's RBC model as well as the short-run restriction version in Christiano, Eichenbaum and Vigfusson (2007), henceforth CEV, of that same model (in which agents do not observe the productivity shock at the time of making the labor decision). We consider both these models because they have taken center stage in much of the debate on the use of SVARs. In addition, we also consider the Smets and Wouters (2007) model, henceforth SW, because it nests many shocks and frictions frequently discussed in macro and arguably captures dynamics deemed important in the data. As a simple way of building in a short-run restriction in that model, we assume that monetary policy responds only to lagged macroeconomic aggregates. The identifying restriction is then that only the monetary

[4] When comparing VARs of different lag-length, we ensure each VAR has the same number of effective observations, equal to $T = 170$. That is, lag initialization does not affect sample size. That said, our results do not hinge on this implementation detail.

[5] See Christiano, Eichenbaum and Vigfusson (2007) for a discussion of why this is the appropriate way to evaluate SVARs. Essentially, one takes an econometrician's perspective - who has only one draw of data and faces a question of inference on the basis of just that data.

policy shock affects the interest rate contemporaneously. Because each of these models is well-known, we refer the reader to the respective papers for a precise description of model equations and parameter calibration/estimation.[6]

We work under a number of maintained simplifications. First, the identification assumptions are invariably correct (i.e., the long or short-run restrictions hold true in the DGP). Second, invertibility is never a problem; all the models we consider are fundamental. Third, all our experiments are based on two-shock models and two-variable VARs. Both RBC models fit that framework by construction, but the SW model does not. For the latter, we consider the model with only monetary policy and preference shocks, and a VAR on GDP-growth and the short term interest rate (in that order). Finally, inference is standard. Uncertainty bands are computed as in Sims and Zha (1999), Canova (2007) and Uhlig (2005). In particular, given a weak conjugate prior, VARs have a posterior distribution of the Normal-Inverse Wishart form, where the distributions are centered around their OLS estimates.[7] Importantly, these priors do not put different weight on short vs. long lags, as one would in e.g. a Minnesota prior.

3.2 Results

Figure 1 contains, for each model, the median bias across all replications for VARs of different lag-length. The figure resembles those found in the literature and shows how short lag-length can imply substantial bias. Particularly, the short-lag VAR ($p = 4$) frequently exhibits the maximum bias at multiple horizons for the different models considered. Long lag-length, or

[6]For CKM and CEV, we follow the CKM baseline calibration. For SW we modify the policy rule to

$$r_t = \rho r_{t-1} + (1-\rho)\left\{r_\pi \pi_{t-1} + r_y\left(y_{t-1} - y^p_{t-1}\right)\right\} + \varepsilon^r_t$$

and calibrate the model at the median of SW's posterior distribution.

[7]Sims and Zha (1999) show these provide good approximations to frequentist intervals. We later also consider a simple bootstrap and show our results do not hinge on the exact inference procedure.

reduced truncation, can induce substantial bias reduction, most notably in CKM and, from intermediate horizons onward, in CEV and SW. To evaluate if such biases are of concern, we now turn to measures of uncertainty.

Result 1: Uncertainty does not explode for long-lag VARs Figure 2 plots a standard measure of uncertainty about IRF: the median width of the confidence bands across all draws.[8] A first glance at that figure reveals that, contrary to common wisdom, CI width does not explode. Instead, even for VARs with very long lags uncertainty bands are roughly in the same ballpark as those of short-lag VARs.

For longer horizons, short-lag VARs trivially attain minimum CI width. The reason is that a VAR(p) cannot propagate much beyond horizon p. As a result, uncertainty cannot propagate much beyond that horizon either. The consequence is, as apparent from Figure 2, that CI width mechanically converges to zero soon after horizon p.

Result 2: Short-lag VARs have maximal uncertainty for horizons where uncertainty is not mechanically low For short horizons short-lag VARs have maximal CI width. This holds true for each of the models considered. A possible reason for that to occur is that misspecification error is maximal for short-lag VARs. Individual reduced-form coefficients may be estimated more precisely for a given draw, but across draws short-lag VARs have increased variance due to the misspecification of the VAR. Long-lag VARs, by contrast, may have individually imprecise reduced form coefficients, but they suffer much less from misspecification.

Result 3: Long-lag VARs have comparable coverage and comparable or better MSE than short-lag VARs Combined with a tendency to produce smaller biases, long-lag VARs have favorable properties compared to more standard short-lag VARs. Figure 3

[8] That is, for each sample draw we subtract the 5th percentile from the 95th, and then take the median across all draws. Results are similar for 68% credible intervals.

documents how long-lag VARs attain coverage rates that are 1) reasonably good overall, 2) comparable to those for short-lag VARs for the CKM and CEV models, 3) much better for the SW model, where short-lag VARs with short run restrictions go astray entirely.[9]

Figure 4 combines bias and variance in a different way, by plotting mean-squared errors (MSE) across horizons. The message is very much the same: at short horizons - where uncertainty does not mechanically shrink - short-lag VARs are either comparable or considerably worse than long-lag VARs.

3.3 Decomposing uncertainty effects

From the above results it may not be obvious that standard intuition - increased parametrization leading to increased uncertainty - holds at all. We here provide a decomposition to measure the impact of the standard intuition on the total variance effect.

Figure 5 plots the Monte Carlo distribution of CI width for three types of impulse responses. Specifically, for each draw of data from the DSGE model, we measure the CI width around the contemporaneous impulse response for CKM and SW, and the second horizon for CEV.[10] The medians of these two distributions are already contained in Figure 2. For short-lag VARs, the dashed line (B_4, Σ_4) plots the distribution of CI widths across all 1000 draws, based on the lag polynomial $B_4(L)$ and covariance matrix Σ_4. Similarly, the solid line (B_{30}, Σ_{30}) plots the distribution of CI widths for a long-lag VAR, based on the lag polynomial $B_{30}(L)$ and covariance matrix Σ_{30}. Comparing these two distributions confirms the earlier results: long-lag VARs do not necessarily imply overwhelmingly dispersed uncertainty

[9] The huge swings in coverage for short-lag VARs in SW arise as the combination of substantial bias and mechanically low uncertainty. As a result, from intermediate horizons onward, the econometrician becomes relatively certain about the wrong point.

[10] While similar effects are at work at longer horizons for all models considered, they are harder to disentangle due to the mechanical reduction in uncertainty for short-lag VARs, as apparent in Figure 2. For CEV the contemporaneous ($h = 0$) response of hours to technology shocks is subject to a zero restriction and is thus uninformative. The figure therefore contains the IRF uncertainty distribution for the horizon $h = 1$.

bands.

To understand why, and to relate our results to the standard intuition, we construct the following counterfactual impulse responses:

$$IRF = f(B_{30}(L), \Sigma_4).$$

These hypothetical IRFs are constructed using the (many) reduced form coefficients of a long-lag VAR, $B_{30}(L)$, combined with the reduced form covariance matrix of a short-lag VAR. Such IRFs can be interpreted as isolating the effect of increased parametrization. They shut down the effect of misspecification reduction by ignoring the reduced bias in Σ. The dotted (B_{30}, Σ_4) distributions in Figure 5 show the CI width associated with these counterfactual impulse responses. Standard intuition dictates that long lag-length makes the entire distribution shift outward, through the additional uncertainty created by the strong increase in number of parameters.

It is immediately apparent that, across models, the dotted distribution does not unequivocally lie to the right of the dashed distribution. In other words, the strong increase in number of parameters need not imply an increase in uncertainty. For the SW model, there is no effect at all from increased parametrization, since the short-run restriction implies that the contemporaneous IRF only depends on Σ and not on $B(L)$. For the CEV and CKM models the right tail of the CI width distribution becomes fatter, as standard intuition would suggest. However, two observations stand out. First, the increase in CI width is not overwhelming. Second, a significant portion of the mass is shifting to the left of the dashed, short-lag distribution, indicating reduced uncertainty.

The fact that increased parametrization does not invariably increase uncertainty is at odds both with standard intuition (less degrees of freedom) and with a well-known omitted variables result. Particularly, coefficient estimates b_1 in (2) are not only biased, but also have too low variance. Intuitively, to the extent that omitted variables correlate with included ones, the explanatory power of those included will appear to be larger than it really is.

Analytically, if we denote the coefficients on X_1 in the correct regression (which does include X_2) by $b_{1.2}$, then

$$Var(b_1) < Var(b_{1.2}). \tag{4}$$

This suggests that by including additional relevant variables one increases the variance of coefficients. We now provide detail on the effects in each of the individual models, which will lay bare the reasons for these seemingly counterintuitive results.

Let us start with the SW model in Figure 5. As mentioned above, since identification is based on short run restrictions, contemporaneous IRFs are not a function of $B(L)$, only of Σ. Hence, the dashed and dotted lines overlap. The effect of misspecification reduction, on the other hand, substantially reduces uncertainty, as can be seen by the shift to the solid distribution.

Now consider the CI width distribution for the CEV model. Here, taking into account the long-lag polynomial clearly only partially results in an increase in uncertainty measures. To see the reason for this, note that IRFs are functions involving multiple coefficients. As a result, covariance between coefficients becomes an issue. For the sake of argument, consider the simplest possible function involving two parameters in (1), their sum. Let $X_1 = [X_{1a}, X_{1b}]$ and denote the corresponding point estimates by b_{1a} and b_{1b}. Then the variance of the sum of the two coefficients in b_1 in the equation that omits X_2 is

$$V(b_{1a} + b_{1b}) = V(b_{1a}) + V(b_{1b}) + 2Cov(b_{1a}, b_{1b}). \tag{5}$$

Similarly, the variance of the sum in the correct regression (which includes X_2) is

$$V(b_{1a.2} + b_{1b.2}) = V(b_{1a.2}) + V(b_{1b.2}) + 2Cov(b_{1a.2}, b_{1b.2}). \tag{6}$$

While we know that each of the first two terms is smaller in (5) than the corresponding terms in (6), the presence of the covariances prevents any automatic conclusion on whether $V(b_{1a} + b_{1b}) \lesseqgtr V(b_{1a.2} + b_{1b.2})$.

Thus, as soon as one considers functions that combine coefficients of a regression subject to omitted variables, the usual variance relation in (4) can break down. This explains the shift from the dashed to the dotted distribution in the CEV model, and particularly why there can be significant mass shifting towards lower uncertainty despite having a big increase in the number of parameters.

The quantitatively more important effect on uncertainty is not due to the big increase in parametrization, however, but rather the effect of the reduction in misspecification. This is illustrated by the shift from the dotted to the solid distribution.

Finally, consider the CKM model in Figure 5. The dotted line in the figure shows how increased parametrization, along the lines of standard intuition, tends to shift the distribution of uncertainty outward compared to the short-lag VAR. Here, too, there is some mass that shifts leftward. As in the case of the CEV model, this can occur because IRFs involve a combination of parameters.[11] Despite the push toward increased uncertainty following the increase in number of parameters, once the misspecification effect through Σ is incorporated long-lag VARs are associated with smaller, not larger uncertainty.

Thus, the figures show the uncertainty trade-off: increased parametrization ($B_4 \rightarrow B_{30}$) which can - but need not - push the distribution outward (from dashed to dotted) vs. reduced misspecification ($\Sigma_4 \rightarrow \Sigma_{30}$) which shrinks uncertainty and thus pulls the distribution to the left (from dotted to solid). In sum, while standard intuition on increased parametrization is partially correct and clearly part of the story, misspecification reduction tends to have more substantial variance effects. As a result, for VARs on data generated by standard DSGE models, the total effect of increasing lag-length can easily imply a reduction in variance. While the simulation setup provides ample insight into VAR inference in small samples, the

[11] The reduction in uncertainty in the dotted distribution can also be the result of reduced misspecification in $B(1)$, documented by Sims (1972), in combination with long-run identifying restrictions. This effect exists because $B(1)$ enters the identification procedure in the case of long-run restrictions. For more on the importance of $B(1)$, see Christiano, Eichenbaum and Vigfusson (2004).

empirical macroeconomist does not have the luxury of knowing the DGP. We now ask if long-lag VARs are also viable in the data by revisiting a few well-known empirical SVAR studies.

4 Empirical applications

4.1 Application 1: Technology shocks and hours

Much of the debate about the use and pitfalls of SVARs centered on the question of the hours response to technology shocks. Galí's (1999) finding that hours fall after a positive technology shock was met with severe backlash. On the theoretical front, CKM stressed that lag truncation bias makes short-lag SVARs with long-run identification restrictions unable to recover true IRF. On the empirical front, Chang and Hong (2006) applied Galí's approach to 458 US manufacturing sectors and found that most sectors exhibit a positive rather than a negative hours response. With the knowledge that long-lag SVARs overcome the truncation problem and may still allow precise inference in realistic samples we now revisit that debate. Specifically, we ask whether long-lag versions of the SVARs of Galí (1999) and Chang and Hong (2006) imply excessive uncertainty and whether their conclusions are different from the original short-lag versions. We use the same samples as the original studies: quarterly US data 1948:1-1994:4 for Galí (1999; 4 quarterly lags) and annual sectoral US data 1958-1996 for Chang and Hong (2006; 1 annual lag). We use the same Bayesian methods as in the Monte Carlo evaluations.

Figure 6 shows the Galí (1999) result for SVARs of increasing lag-length. The estimated impact effect of technology shocks on hours is invariably negative, regardless of lag-length. This effect is significant for VARs with (quarterly) lag-lengths up to $p = 20$ (5 years). At lag-lengths of around $p = 30$ (7.5 years) the effect turns insignificant (a result of both a less negative point estimate and a larger variance). The bottom row shows that CI width does not obviously increase with lag-length. While the VAR(4) has minimal variance at horizons $h = 0, 1$, the VAR(30) has smaller variance at horizons $h = 2, 3, 4$. At longer horizons $h > p$ mechanical effects start kicking in for short-lag VARs. Broadly speaking, however, for horizons $h < p$ VARs of different lag-length have largely comparable CI width.

Figure 7 (top panel, dashed distribution) contains our replication of Chang and Hong's (2006) result: the majority of sectors see an increase in hours in the wake of a positive technology shock. The bottom panel shows the associated cross-sectional distribution of CI width. Combined, the short-lag VAR finds 203 significantly positive sectors and only 46 significantly negative (Table 1).

The results for the long-lag SVAR are striking. First, the point estimates differ dramatically. The cross-sectional distribution of IRFs shifts substantially to the left in Figure 7 (top panel). While short-lag VARs estimate 70% of sectors (Table 1: 317/458) respond positively, long-lag VARs find 50% of sectors respond positively and the other half negatively. Second, variance does not become excessively large. The bottom panel of Figure 7 shows that long-lag VARs have mostly lower variance than their short-lag counterpart: the cross-sectional distribution of CI width predominantly shifts to the left for long lags. Combined, the number of industries that differ significantly from zero is almost the same for short and long-lag VARs (249 and 244 respectively). Hence, once again, long-lag VARs do not exhibit excessive variance. While there is a stark contrast between the short-lag results of Galí (1999) and Chang and Hong (2006), long-lag results seem more in line with one another.

We take from this evidence that long-lag VARs are a viable, feasible tool in the macro empiricist's toolkit. Long-lag VARs do not imply excessive variance even in realistic samples. The evidence also suggests that there are substantial effects on point estimates, confirming the potential importance of truncation bias also in practice.

Table 1: Short-run hours response to technology shock - Chang and Hong (2006)

	1 lag			10 lags		
	Number of industries			Number of industries		
	Positive	Negative	Total	Positive	Negative	Total
Estimate (median)	317	141	458	229	229	458
Significant at 10%	203	46	249	128	116	244

4.2 Application 2: Monetary policy shocks

Perhaps most of the development of SVAR methods involved the study of monetary policy shocks. We here revisit a prototype monetary policy VAR and again ask to what extent long-lag VARs change inference. We take this opportunity to depart from a number of maintained assumptions in our analysis thus far. Particularly, the reader may wonder whether the viability of long-lag VARs hinges on two specific factors: Bayesian inference or two-variable SVARs. To address these questions our monetary application uses frequentist inference methods (a simple bootstrap) and a three-variable SVAR, both as in Stock and Watson (2001) who use quarterly data 1960:1-2000:4 and a VAR(4).

Figure 8 shows that the researcher using long-lag SVARs, just like the literature based on short-lag VARs summarized by Stock and Watson (2001), finds that an exogenous increase in the policy rate leads to a significant delayed increase in unemployment and an even further delayed drop in inflation. The last column of the figure shows that here too VARs of different lag-length have comparable variances (at all horizons where uncertainty is not mechanically low).

At this juncture we do not want to read much more into detailed differences between the estimated short and long-lag SVARs (e.g. long-lag VARs do not feature a price puzzle, and exhibit an unemployment reversal in the monetary application), nor draw strong conclusions on the basis of them (e.g. long-lag VARs show equal support for salt vs. freshwater camps in the technology application). This partly because there are perhaps newer and better ways to identify shocks than the recursive approaches adopted in the above papers. More importantly, the variance trade-off we document does not hinge on the particular identification assumptions made - it applies to all SVAR approaches that rely on a correct (untruncated) reduced form. Equally importantly, our point is not that long-lag SVARs are always and everywhere the better approach. Our simulations and applications do prove that long-lag VARs are a viable tool, that no longer should be dismissed on the grounds of presumed excessively high variance. Whether the researcher decides on using short or long lag-length is a model-selection question that will obviously be application-specific. But long-lag VARs should be in the set of models considered.

5 Discussion

5.1 On choosing lag-length

All of the above results are in terms of structural inference. None of our results imply that long-lag VARs ought to be used for matters such as forecasting. For instance, the large dimensionality of the lag polynomial in long-lag VARs prohibits any success in forecasting due to the lack of parsimony. While one can certainly envisage ways to reduce the dimensionality, that is not the issue here. Rather, if one wants to draw structural conclusions, e.g. by means of IRFs, then misspecification concerns are essential. Therefore, if forecasting is not the main purpose of the model, it may be ill-advised to trust lag-selection criteria which purely focus on forecasting/parsimony.

A potential drawback of including longer lags is that it induces overfitting. We have extensively investigated this possibility. For the models and the lag-lengths considered here, we find it not be a major problem. One way to see this is as follows. If present, overfitting should have a first-order effect on bias. In other words, one would expect bias to increase when overfitting sets in. This does not generally occur in our simulations. An important avenue of future research lies in the development of information criteria that move beyond forecasting/parsimony, and take into account that the purpose of the model is structural inference, while also avoiding issues of overfitting.

5.2 On the maintained simplifications

Throughout our analysis, the only modification as compared to the standard approach is an increase in lag-length. No additional degree of complexity is introduced and only standard tools are used. But let us briefly reflect on some of the choices made in our analysis.

First, while the inference method is Bayesian, our priors are weak and do not put different prior weights on short vs. long lags.[12] Our general result goes through for frequentist methods, as shown in our monetary application.

Second, our simulations and empirical applications are based on two and three-variable VARs. Considering small VARs serves to keep the number of parameters limited. As lags increase, the number of parameters increases faster the more variables in the system. Since much of the influential SVAR evidence in the literature is based on small VARs, with two or three variables, it seems reasonable to focus on small VARs.

Moreover, many developments in empirical macro enable dealing with larger systems, both in terms of longer lags and more variables. For instance, variants on Minnesota-type priors can allow inclusion of long lags in VARs with many variables. Alternatively, factor dynamics with potentially long lags may well improve structural inference without a large increase in parameters relative to the size of the data. Smoothness priors are yet another

[12] While Antolin-Diaz and Surico (2025) also consider VARs with long lag-length, their setup is very different because 1) they use priors that strongly favor short lags and shrink long lags toward zero, and 2) they use a sample that is much longer than what is typically available in macroeconomic data. In such a setting it is not surprising that long lag-length does not induce larger variance.

available alternative. In short, there are potentially many ways of dealing with larger systems. Irrespective of the particular approach, the variance trade-off we document will be at work in larger systems, too.

6 Conclusion

We document a general trade-off. Of course, it is possible to design models or find data for which the balance of the trade-off leans toward short lag-lengths. However, contrary to common wisdom, long lag-length need not imply prohibitively large imprecision. While increased parametrization in itself may increase uncertainty, this effect is counteracted by a reduction in misspecification. For SVARs estimated on data generated by frequently used DSGE models, longer lag-length tends to imply less bias and more precise inference. In empirical applications we find that the variance trade-off in VARs is not particular to data generated by DSGE models. For long-lag versions of prominent SVARs in the literature, the balance of uncertainty effects seems to favor misspecification reduction over parametrization concerns. In particular, we find that results can be substantially different from their short-lag counterparts and that uncertainty does not necessarily become excessively large. Long-lag VARs are therefore a viable instrument in the empirical macroeconomist's toolkit.

References

[1] Antolin-Diaz, J., Surico, P., 2025. "The long-run effects of government spending", *American Economic Review*, forthcoming.

[2] Beaudry, P., Portier, F., 2006. "Stock prices, news, and economic fluctuations", *American Economic Review* 96, 1293-1307.

[3] Bernanke, B.S., 1983. "Nonmonetary effects of the financial crisis in the propagation of the Great Depression", *American Economic Review* 73, 257-76.

[4] Blanchard, O.J., Quah, D., 1989. "The dynamic effects of aggregate demand and supply disturbances", *American Economic Review* 79, 655-73.

[5] Braun, P.A., Mittnik, S., 1993. "Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions", *Journal of Econometrics* 59, 319-341.

[6] Canova, F., 2007. Applied Macroeconomic Research, Princeton University Press, Princeton, New Jersey.

[7] Chang, Y., Hong, J.H., 2006. "Do technological improvements in the manufacturing sector raise or lower employment?", *American Economic Review* 96, 352-368.

[8] Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2008. "Are structural VARs with long-run restrictions useful in developing business cycle theory?", *Journal of Monetary Economics* 55, 1337-52.

[9] Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2004. "The response of hours to a technology shock: Evidence based on direct measures of technology", *Journal of the European Economic Association* 2, 381-95.

[10] Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2007. "Assessing structural VARs", In: Acemoglu, D., Rogoff, K.S., Woodford, M. (Eds.), NBER Macroeconomics Annual 2006. MIT Press, Cambridge, 1-106.

[11] Cooley, T.F., Dwyer, M., 1998. "Business cycle analysis without much theory. A look at structural VARs", *Journal of Econometrics* 83, 57-88.

[12] Eichenbaum, M., Evans, C., 1995. "Some empirical evidence on the effects of shocks to monetary policy on exchange rates", *Quarterly Journal of Economics* 110, 975-1009.

[13] Faust, J., Leeper, E.M., 1997. "When do long-run identifying restrictions give reliable results?", *Journal of Business and Economic Statistics* 15, 345-53.

[14] Fisher, J.D.M., 2006. "The dynamic effects of neutral and investment-specific technology shocks", *Journal of Political Economy* 114, 413-51.

[15] Galí, J., 1999. "Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?", *American Economic Review* 89, 249-71.

[16] Li, D., Plagborg-Møller, M., Wolf, C.K., 2024a. "Local projections vs. VARs: Lessons from thousands of DGPs", *Journal of Econometrics* 244, 105722, 1-21.

[17] Li, D., Plagborg-Møller, M., Wolf, C.K., 2024b. "Local projections vs. VARs: Lessons from thousands of DGPs", slides, available at www.mikkelpm.com/research/.

[18] Ravenna, F., 2007. "Vector autoregressions and reduced form representations of DSGE models", *Journal of Monetary Economics* 54, 2048-64.

[19] Sims, C., 1972. "The role of approximate prior restrictions in distributed lag estimation", *Journal of the American Statistical Association* 67, 169-75.

[20] Sims, C., 1980. "Macroeconomics and reality", *Econometrica* 48, 1-48.

[21] Sims, C., 1989. “Models and their uses”, *American Journal of Agricultural Economics* 71, 489-94.

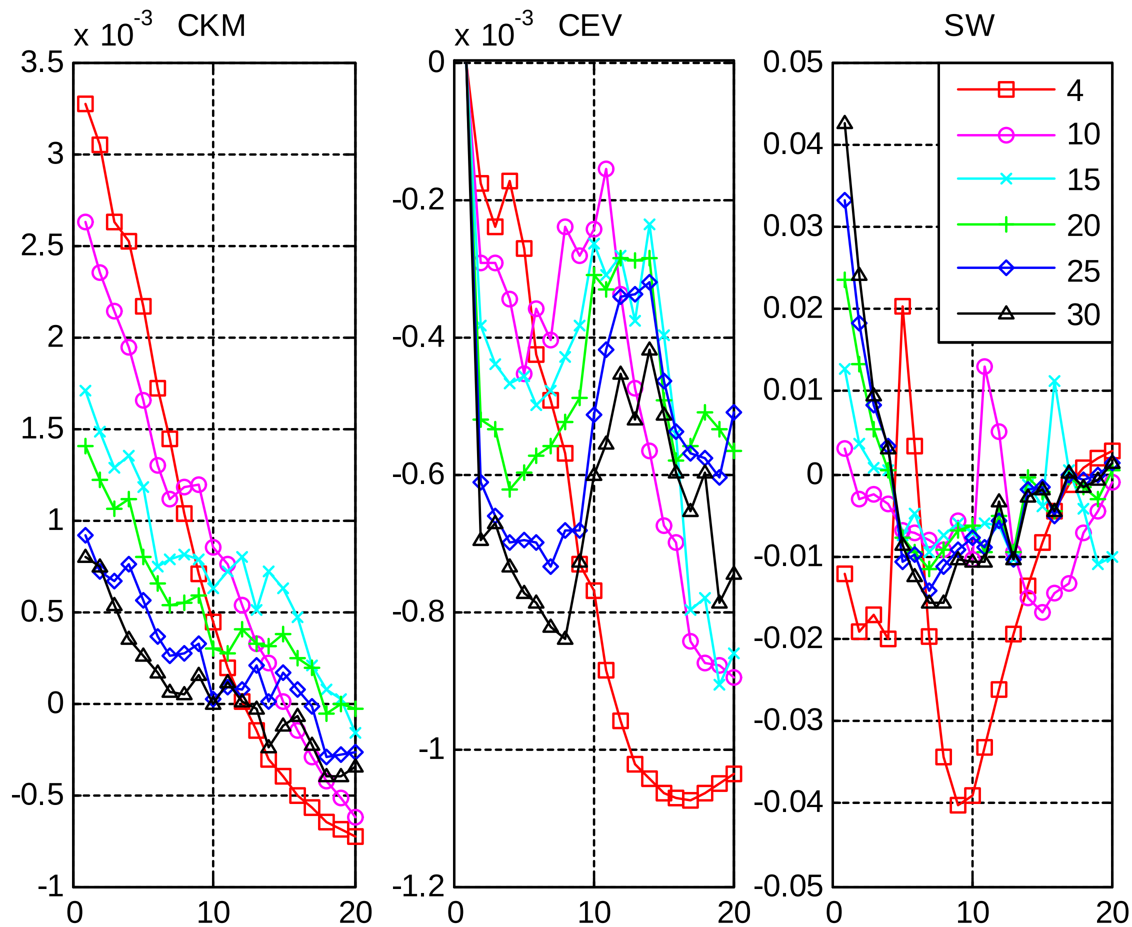
[22] Sims, C., Zha, T., 1999. “Error bands for impulse responses”, *Econometrica* 67, 1113-1155.

[23] Smets, F., Wouters, R., 2007. “Shocks and frictions in US business cycles: a Bayesian DSGE approach”, *American Economic Review* 97, 586-606.

[24] Stock, J.H., Watson, M.W., 2001. “Vector autoregressions”, *Journal of Economic Perspectives* 15, 101-115.

[25] Uhlig, H., 2005. “What are the effects of monetary policy on output? Results from an agnostic identification procedure”, *Journal of Monetary Economics* 52, 381-419.

Figure 1: Bias



Note: Bias calculated as $IRF(VAR(p)) - IRF(DSGE)$. Horizontal axis is horizon in quarters.

Figure 2: Confidence interval width (95th-5th percentile)

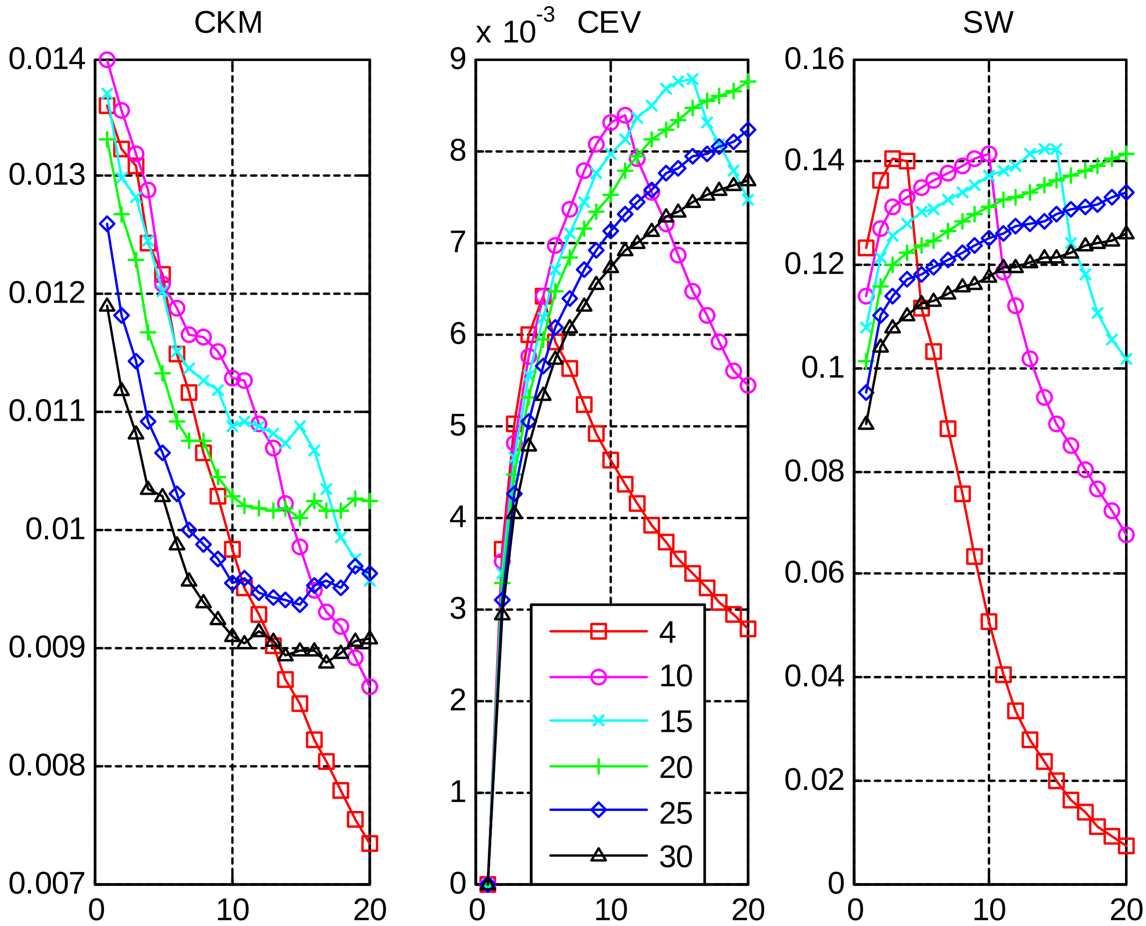


Figure 3: Coverage (90 percent)

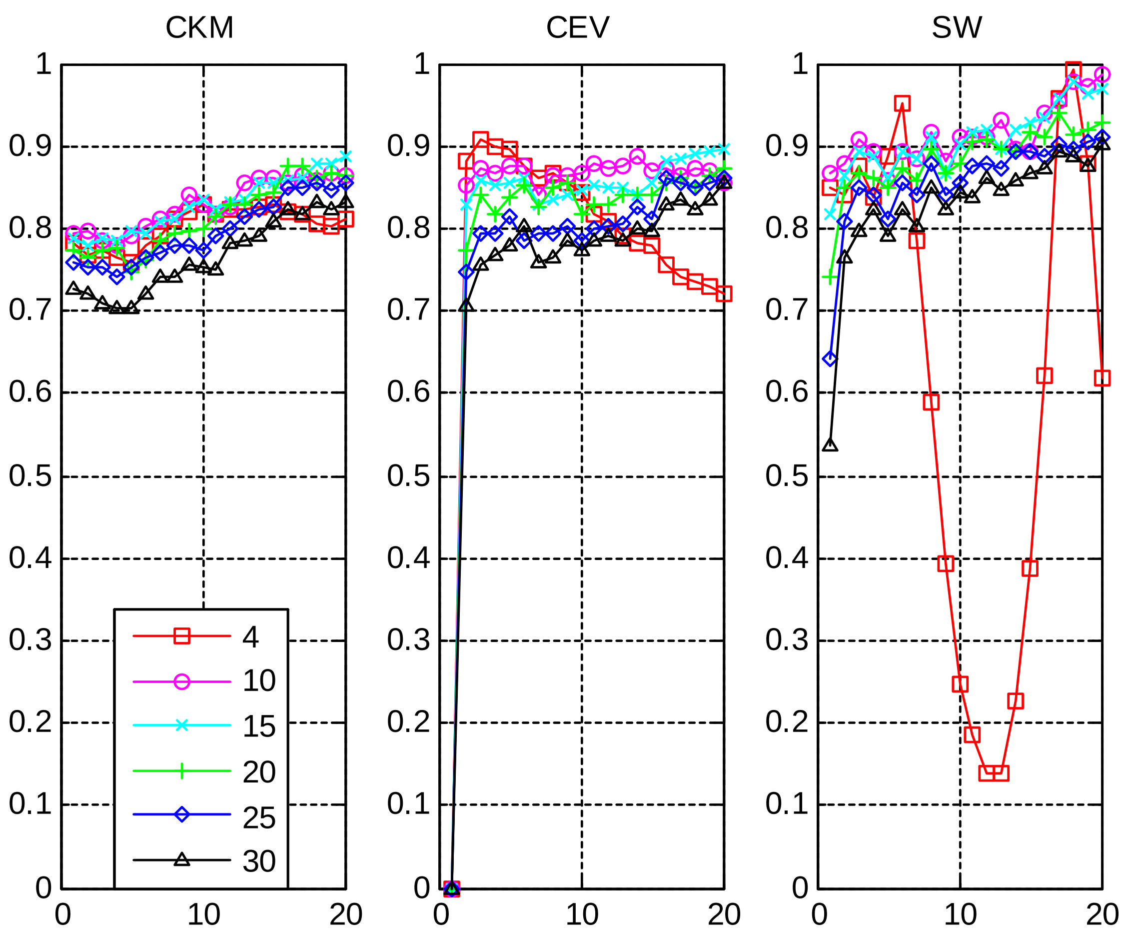


Figure 4: Mean-squared error

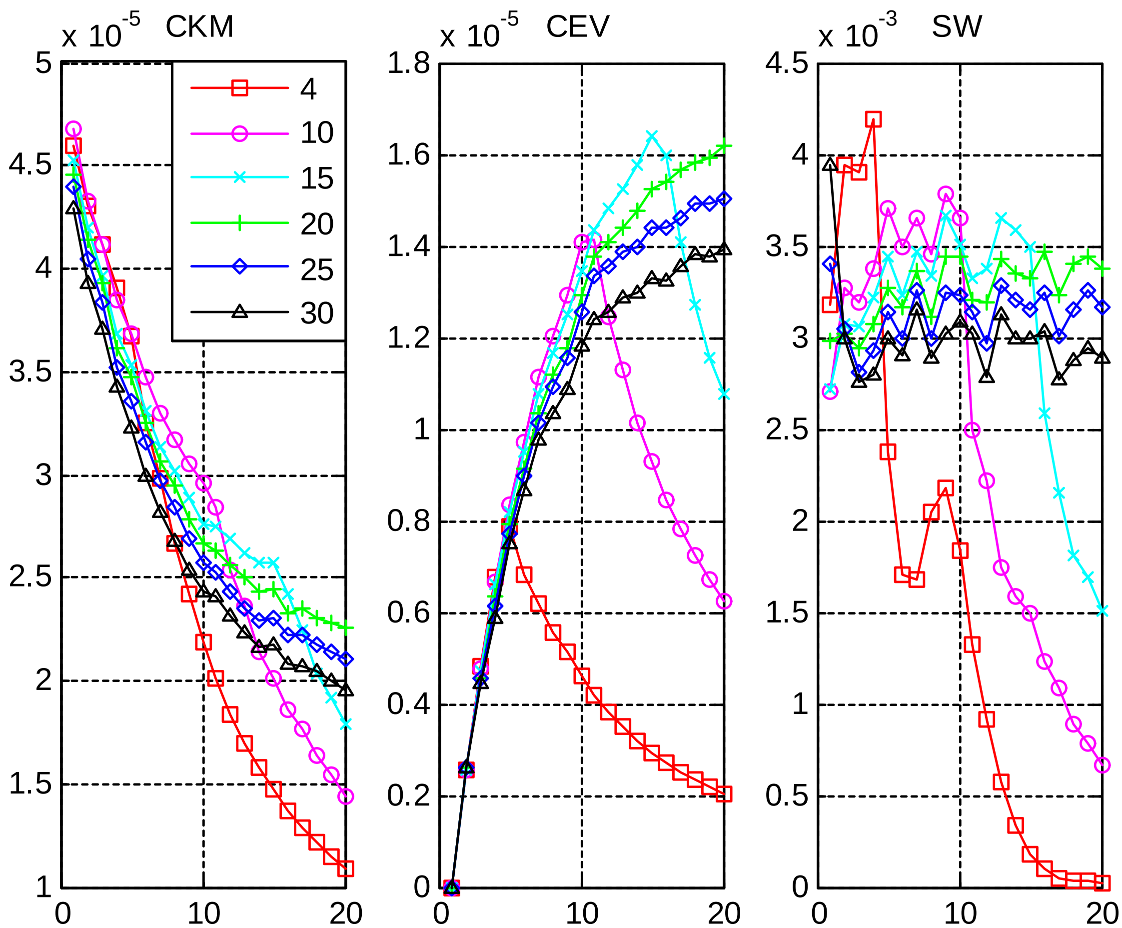
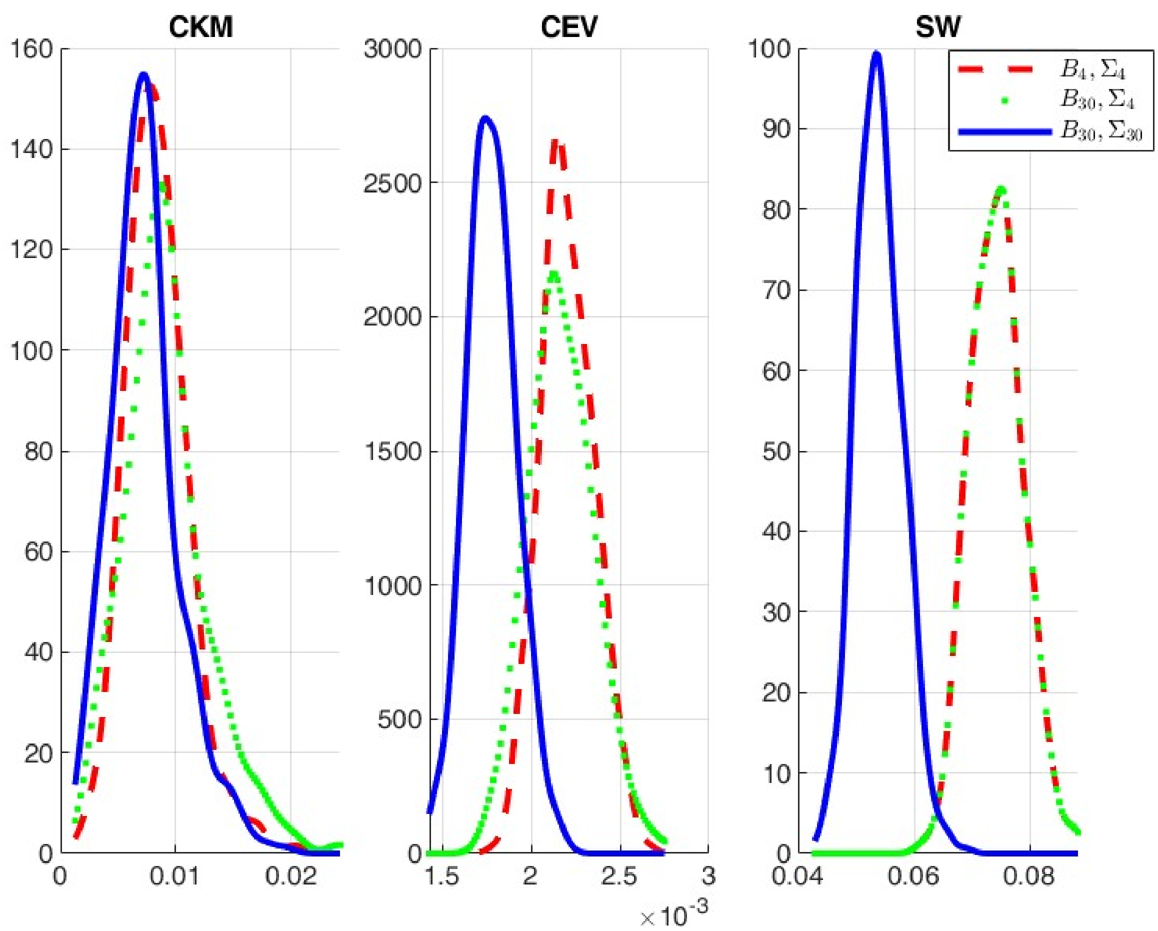
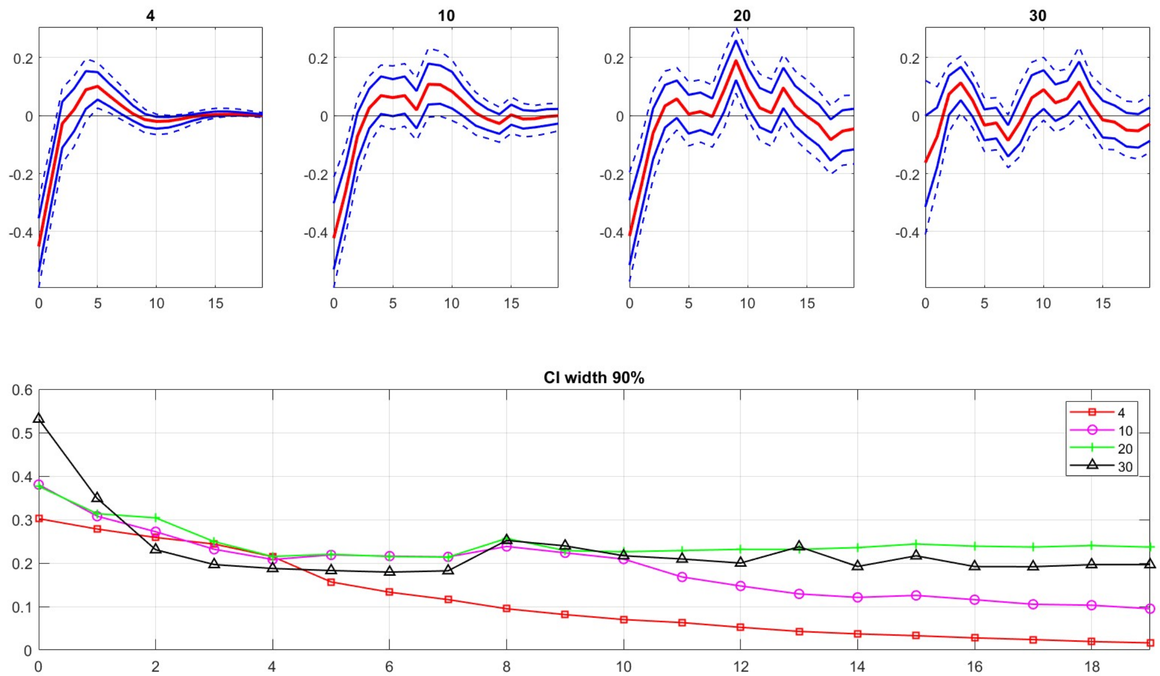


Figure 5: CI width distribution



Note: CI width (90%) of IRF. For CKM and SW horizon $h = 0$. For CEV horizon $h = 1$.

Figure 6: Hours response to technology shock: Galí (1999)



Note: The top row shows IRF and bootstrapped CI (68% and 90%) for VARs of increasing lag-length. The bottom row overlays CI width (90%) for the different VARs.

Figure 7: Hours response to technology shock: Chang and Hong (2006)

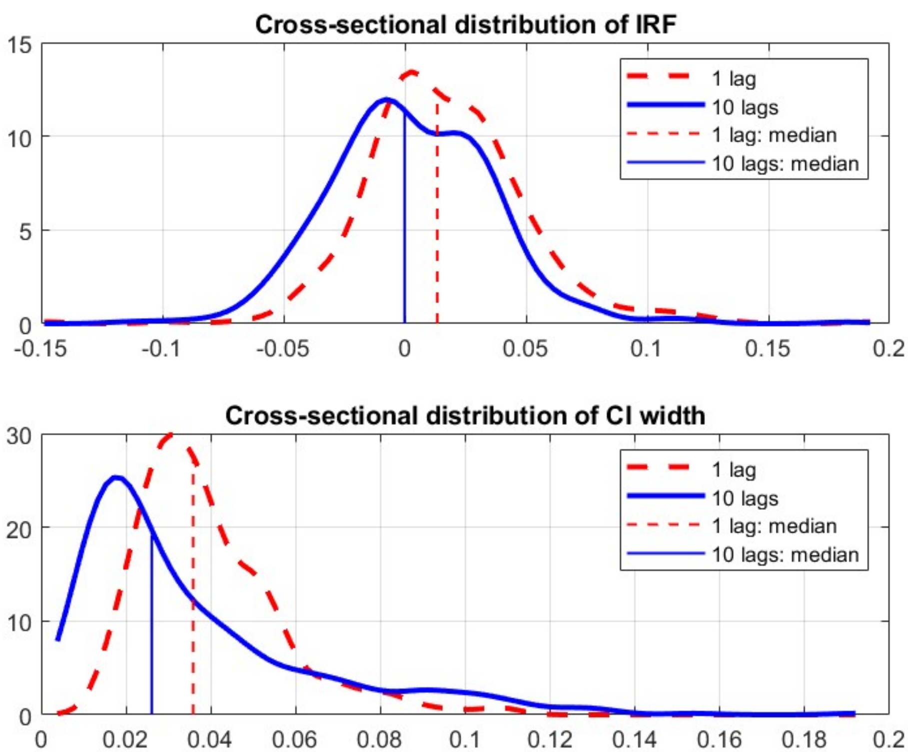
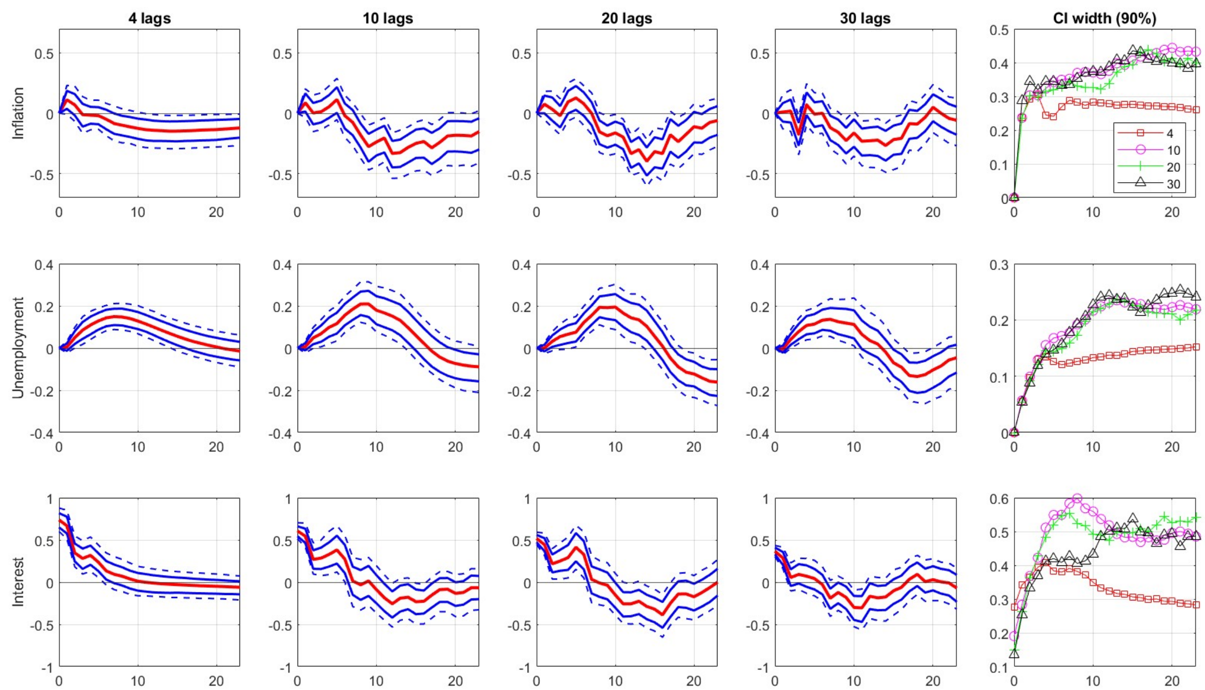


Figure 8: Monetary policy shock: Stock and Watson (2001)



Note: The first four columns show IRF and bootstrapped CI (68% and 90%) for VARs of increasing lag-length. The last column overlays CI width (90%) for the different VARs.

Recent Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

The Macroeconomic Effects of Trade Tariffs: Revisiting the Lerner Symmetry Result *by Jesper Lindé and Andrea Pescatori*	2019:363
Biased Forecasts to Affect Voting Decisions? The Brexit Case *by Davide Cipullo and André Reslow*	2019:364
The Interaction Between Fiscal and Monetary Policies: Evidence from Sweden *by Sebastian Ankargren and Hovick Shahnazarian*	2019:365
Designing a Simple Loss Function for Central Banks: Does a Dual Mandate Make Sense? *by Davide Debortoli, Jinill Kim and Jesper Lindé*	2019:366
Gains from Wage Flexibility and the Zero Lower Bound *by Roberto M. Billi and Jordi Galí*	2019:367
Fixed Wage Contracts and Monetary Non-Neutrality *by Maria Björklund, Mikael Carlsson and Oskar Nordström Skans*	2019:368
The Consequences of Uncertainty: Climate Sensitivity and Economic Sensitivity to the Climate *by John Hassler, Per Krusell and Conny Olovsson*	2019:369
Does Inflation Targeting Reduce the Dispersion of Price Setters' Inflation Expectations? *by Charlotte Paulie*	2019:370
Subsampling Sequential Monte Carlo for Static Bayesian Models *by David Gunawan, Khue-Dung Dang, Matias Quiroz, Robert Kohn and Minh-Ngoc Tran*	2019:371
Hamiltonian Monte Carlo with Energy Conserving Subsampling *by Khue-Dung Dang, Matias Quiroz, Robert Kohn, Minh-Ngoc Tran and Mattias Villani*	2019:372
Institutional Investors and Corporate Investment *by Cristina Cella*	2019:373
The Impact of Local Taxes and Public Services on Property Values *by Anna Grodecka and Isaiah Hull*	2019:374
Directed technical change as a response to natural-resource scarcity *by John Hassler, Per Krusell and Conny Olovsson*	2019:375
A Tale of Two Countries: Cash Demand in Canada and Sweden *by Walter Engert, Ben Fung and Björn Segendorf*	2019:376
Tax and spending shocks in the open economy: are the deficits twins? *by Mathias Klein and Ludger Linnemann*	2019:377
Mind the gap! Stylized dynamic facts and structural models *by Fabio Canova and Filippo Ferroni*	2019:378
Financial Buffers, Unemployment Duration and Replacement Labor Income *by Mats Levander*	2019:379
Inefficient Use of Competitors' Forecasts? *by André Reslow*	2019:380
How Much Information Do Monetary Policy Committees Disclose? Evidence from the FOMC's Minutes and Transcripts *by Mikael Apel, Marianna Blix Grimaldi and Isaiah Hull*	2019:381
Risk endogeneity at the lender/investor-of-last-resort *by Diego Caballero, André Lucas, Bernd Schwaab and Xin Zhang*	2019:382
Heterogeneity in Households' Expectations of Housing Prices – Evidence from Micro Data *by Erik Hjalmarsson and Pär Österholm*	2019:383
Big Broad Banks: How Does Cross-Selling A Affect Lending? *by Yingjie Qi*	2020:384
Unemployment Fluctuations and Nominal GDP Targeting *by Roberto Billi*	2020:385
FAQ: How do I extract the output gap? *by Fabio Canova*	2020:386

Drivers of consumer prices and exchange rates in small open economies *by Vesna Corbo and Paola Di Casola*	2020:387
TFP news, stock market booms and the business cycle: Revisiting the evidence with VEC models *by Paola Di Casola and Spyridon Sichlimiris*	2020:388
The costs of macroprudential deleveraging in a liquidity trap *by Jiaqian Chen, Daria Finocchiaro, Jesper Lindé and Karl Walentin*	2020:389
The Role of Money in Monetary Policy at the Lower Bound *by Roberto M. Billi, Ulf Söderström and Carl E. Walsh*	2020:390
MAJA: A two-region DSGE model for Sweden and its main trading partners *by Vesna Corbo and Ingvar Strid*	2020:391
The interaction between macroprudential and monetary policies: The cases of Norway and Sweden *by Jin Cao, Valeriya Dinger, Anna Grodecka-Messi, Ragnar Juelsrud and Xin Zhang*	2020:392
Withering Cash: Is Sweden ahead of the curve or just special? *by Hanna Armelius, Carl Andreas Claussen and André Reslow*	2020:393
Labor shortages and wage growth *by Erik Frohm*	2020:394
Macro Uncertainty and Unemployment Risk *by Joonseok Oh and Anna Rogantini Picco*	2020:395
Monetary Policy Surprises, Central Bank Information Shocks, and Economic Activity in a Small Open Economy *by Stefan Laséen*	2020:396
Econometric issues with Laubach and Williams' estimates of the natural rate of interest *by Daniel Buncic*	2020:397
Quantum Technology for Economists *by Isaiah Hull, Or Sattath, Eleni Diamanti and Göran Wendin*	2020:398
Modeling extreme events: time-varying extreme tail shape *by Bernd Schwaab, Xin Zhang and André Lucas*	2020:399
The Effects of Government Spending in the Eurozone *by Ricardo Duque Gabriel, Mathias Klein and Ana Sofia Pessoa*	2020:400
Narrative Fragmentation and the Business Cycle *by Christoph Bertsch, Isaiah Hull and Xin Zhang*	2021:401
The Liquidity of the Government Bond Market – What Impact Does Quantitative Easing Have? Evidence from Sweden *by Marianna Blix Grimaldi, Alberto Crosta and Dong Zhang*	2021:402
Five Facts about the Distributional Income Effects of Monetary Policy *by Niklas Amberg, Thomas Jansson, Mathias Klein and Anna Rogantini Picco*	2021:403
When domestic and foreign QE overlap: evidence from Sweden *by Paola Di Casola and Pär Stockhammar*	2021:404
Dynamic Macroeconomic Implications of Immigration *by Conny Olovsson, Karl Walentin, and Andreas Westermark*	2021:405
Revisiting the Properties of Money *by Isaiah Hull and Or Sattath*	2021:406
The cost of disinflation in a small open economy vis-à-vis a closed economy *by Oleksandr Faryna, Magnus Jonsson and Nadiia Shapovalenko*	2021:407
On the Performance of Cryptocurrency Funds *by Daniele Bianchi and Mykola Babiak*	2021:408
The low-carbon transition, climate commitments and firm credit risk *by Sante Carbone, Margherita Giuzio, Sujit Kapadia, Johannes Sebastian Krämer, Ken Nyholm and Katia Vozian*	2022:409
Seemingly Irresponsible but Welfare Improving Fiscal Policy at the Lower Bound *by Roberto M. Billi and Carl E. Walsh*	2022:410
Pension Reform and Wealth Inequality: Evidence from Denmark *by Torben M. Andersen, Joydeep Bhattacharya, Anna Grodecka-Messi and Katja Mann*	2022:411

Inflation Targeting or Fiscal Activism? *by Roberto M. Billi*	2022:412
Trading volume and liquidity provision in cryptocurrency markets *by Daniele Bianchi, Mykola Babiak and Alexander Dickerson*	2022:413
DISPERSION OVER THE BUSINESS CYCLE: PASSTHROUGH, PRODUCTIVITY, AND DEMAND *by Mikael Carlsson, Alex Clymo and Knut-Eric Joslin*	2022:414
Electoral Cycles in Macroeconomic Forecasts *by Davide Cipullo and André Reslow*	2022:415
The Curious Incidence of Monetary Policy Across the Income Distribution *by Tobias Broer, John Kramer and Kurt Mitman*	2022:416
Central Bank Mandates and Monetary Policy Stances: through the Lens of Federal Reserve Speeches *by Christoph Bertsch, Isaiah Hull, Robin L. Lumsdaine, and Xin Zhang*	2022:417
The Political Costs of Austerity *by Ricardo Duque Gabriel, Mathias Klein and Ana Sofia Pessoa*	2022:418
Central bank asset purchases: Insights from quantitative easing auctions of government bonds *by Stefan Laséen*	2023:419
Greenflation? *by Conny Olovsson and David Vestin*	2023:420
Effects of foreign and domestic central bank government bond purchases in a small open economy DSGE model: Evidence from Sweden before and during the coronavirus pandemic *by Yildiz Akkaya, Carl-Johan Belfrage, Paola Di Casola and Ingvar Strid*	2023:421
Dynamic Credit Constraints: Theory and Evidence from Credit Lines* *by Niklas Amberg, Tor Jacobson, Vincenzo Quadrini and Anna Rogantini Picco*	2023:422
Stablecoins: Adoption and Fragility *by Christoph Bertsch*	2023:423
CBDC: Lesson from a Historical Experience *by Anna Grodecka-Messi and Xin Zhang*	2023:424
Do Credit Lines Provide Reliable Liquidity Insurance? Evidence from Commercial-Paper Backup Lines *by Niklas Amberg*	2023:425
Price Pass-Through Along the Supply Chain: Evidence from PPI and CPI Microdata *by Edvin Ahlander, Mikael Carlsson and Mathias Klein*	2023:426
Cash for Transactions or Store-of-Value? A comparative study on Sweden and peer countries *by Carl Andreas Claussen, Björn Segendorf and Franz Seitz*	2023:427
Fed QE and bank lending behaviour: a heterogeneity analysis of asset purchases *by Marianna Blix Grimaldi and Supriya Kapoor*	2023:428
Monetary policy in Sweden after the end of Bretton Woods *by Emma Bylund, Jens Iversen and Anders Vredin*	2023:429
Banking Without Branches *by Niklas Amberg and Bo Becker*	2024:430
Climate impact assessment of retail payment services *by Niklas Arvidsson, Fumi Harahap, Frauke Urban and Anissa Nurdiawati*	2024:431
Four Facts about International Central Bank Communication b*y Christoph Bertsch, Isaiah Hull, Robin L. Lumsdaine, and Xin Zhang*	2024:432
Optimal Monetary Policy with r* < 0 *by Roberto Billi, Jordi Galí, and Anton Nakov*	2024:433
Quantitative Easing, Bond Risk Premia and the Exchange Rate in a Small Open Economy *by Jens H. E. Christensen and Xin Zhang*	2024:434
Supply-Chain Finance: An Empirical Evaluation of Supplier Outcomes *by Niklas Amberg, Tor Jacobson and Yingjie Qi*	2024:435
Optimal Contracts and Inflation Targets Revisited *by Torsten Persson and Guido Tabellini*	2024:436
Potential Climate Impact of Retail CBDC Models *by Niklas Arvidsson, Fumi Harahap, Frauke Urban and Anissa Nurdiawati*	2024:437

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se