EUROPEAN CENTRAL BANK

EUROSYSTEM




WORKING PAPER SERIES

NO 1600 / OCTOBER 2013

GRANGER-CAUSAL-PRIORITY AND CHOICE OF VARIABLES IN VECTOR AUTOREGRESSIONS

Marek Jarociński and Bartosz Maćkowiak

In 2013 all ECB publications feature a motif taken from the €5 banknote.


Acknowledgements
We thank for helpful comments Gianni Amisano, Fabio Canova, Marco Del Negro, John Geweke, Domenico Giannone, David Madigan, Giorgio Primiceri, Frank Schorfheide, Chris Sims, Ellis Tallman, Tao Zha, and seminar and conference participants. We thank Giovanni Nicolo for excellent research assistance. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the European Central Bank.

Marek Jarociński
European Central Bank

Bartosz Maćkowiak
European Central Bank and CEPR; e-mail: bartosz.mackowiak@ecb.europa.eu

Address	Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
Postal address	Postfach 16 03 19, 60066 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Internet	http://www.ecb.europa.eu
Fax	+49 69 1344 6000

ISSN	1725-2806 (online)
EU Catalogue No	QB-AR-13-097-EN-N (online)

Abstract

A researcher is interested in a set of variables that he wants to model with a vector autoregression and he has a dataset with more variables. Which variables from the dataset to include in the VAR, in addition to the variables of interest? This question arises in many applications of VARs, in prediction and impulse response analysis. We develop a Bayesian methodology to answer this question. We rely on the idea of Granger-causal-priority, related to the well-known concept of Granger-noncausality. The methodology is simple to use, because we provide closed-form expressions for the relevant posterior probabilities. Applying the methodology to the case when the variables of interest are output, the price level, and the short-term interest rate, we find remarkably similar results for the United States and the euro area.

Keywords: Vector autoregression, structural vector autoregression, Granger-causal-priority, Granger-noncausality, Bayesian model choice. (*JEL*: C32, C52, E32.)

Non-technical summary

The *theory* of econometrics usually takes a model as given. But an important element in the *practice* of econometrics is model specification, or model choice. In econometrics applied to macroeconomic data the dominant model since Sims (1980) has been the vector autoregression (VAR). VARs have been widely used, including at central banks and other policy institutions, for forecasting as well as for studying the effects of economically interpretable shocks such as changes in monetary policy. The crucial aspect of model specification in the context of VARs is the *choice of variables*. If a modeller wants to forecast or to compute the impulse response of a variable *y* with a VAR, which other variables should the modeller include in the VAR? For example, if the variable *y* is the consumer price level, think of what a large number of variables might matter in forecasting and computing the impulse responses of *y*! Then think of checking all combinations of this large number of variables! Although assessing *how much a given variable matters* taking into account all combinations of variables seems like a gargantuan task, this paper develops a methodology to do precisely that. The methodology is both formal – in the sense that we will explain – and *very simple*. We write "very simple" because the output of the methodology for any variable in the modeller's dataset is *a single number* between zero and one – think of this number as a probability – that summarises how much this particular variable matters in modelling the variable of interest *y*.

The methodology is formal in the sense that it adheres to the following principle of Bayesian statistics: To compare or evaluate models, one needs to examine the *out-of-sample* predictive performance of the models. Model A is better than model B if and only if model A forecasts better out-of-sample than model B. We implement this principle of Bayesian statistics in the following way. First, we specify a restriction on the data generating process that connects to variable choice. This restriction is called Granger-causal-priority and appears in unpublished work by Sims (2010) and Todd and Doan (2010). Granger-causal-priority is related to the well-known idea of Granger-noncausality, but it accounts better for the presence of multiple variables in a VAR. Second, we develop tools for inference concerning Granger-causal-priority. The key ingredient is that we derive an *analytical* expression for a Bayes factor, which is a function of out-of-sample predictive performance, and that lets one evaluate the posterior probability of Granger-noncausality and the posterior probability of Granger-causal-priority.

To see how the methodology that we propose works, we investigate which variables belong in a quarterly VAR with real GDP, the price level, and the short-term interest rate. We perform the exercise twice, for the United States and for the euro area. In each exercise we consider thirty-eight macroeconomic and financial variables. The findings turn out to be remarkably similar between the United States and the euro area. Both in the United States and in the euro area we find that the following variables are most likely to belong in the VAR with real GDP, the price level, and the short-term interest rate: (i) survey-based indicators of economic sentiment and activity, (ii) a component of real GDP, the change in inventories, (iii) interest rates on government debt and private debt, and (iv) the price of oil.

1 Introduction

The theory of econometrics usually takes a model as given. But an important element in the practice of econometrics is model specification, as stressed by Leamer (1978). A key question in the specification of a vector autoregression (VAR) – the dominant model in econometrics applied to macroeconomic data – is which variables to include in the VAR. We highlight two features of the vast literature on VARs in macroeconomics started by Sims (1980). First, almost all applications in the literature involve small or medium-sized VARs. In other words, many variables that economists have data on are left out.[1] Second, the choice of which variables to include in a VAR occurs informally, based on the researcher's prior or an informal specification search. This paper is concerned with formal choice of variables in VARs. We discuss a concept – a restriction on a data generating process – that is tightly linked with variable choice in VARs. Furthermore, we show how to use data to evaluate the plausibility of this restriction in a simple way.

Consider a researcher with an a priori interest in a set of variables y_i. The researcher wants to predict y_i with a VAR ("a reduced-form VAR") or to compute impulse responses of y_i to structural shocks ("a structural VAR"). The researcher has data on a set of variables y that includes the variables of interest, i.e., $y_i \subset y$, but also includes other variables y_J, $y_J \equiv y \setminus y_i$. Let $y_j \subseteq y_J$ be a subset of the other variables. The questions that we study in this paper are: (i) Does y_j belong in the VAR to be used to predict y_i? (ii) Does y_j belong in the VAR to be used to compute impulse responses of y_i to structural shocks? We develop a methodology to answer both questions. The methodology relies on two ingredients: a restriction on a data generating process and tools for inference.

We observe that it is natural to think of variable choice as a restriction on the data generating process followed by *all variables in the dataset*. Suppose that the set of all variables we have data on, y, follows a VAR process. Then the decisive restriction on the data generating process for a researcher asking "Does y_j belong in the VAR with y_i?" is

[1] Let us give some examples. The original version of the classic VAR model for forecasting the U.S. economy developed by Robert Litterman used six variables: the Treasury-bill rate, M1, the GNP deflator, real GNP, real investment, and the unemployment rate. See Sims (1993), who studies a nine-variable extension of Litterman's model. Sims and Zha (2006), p.60, write that they employ six variables "commonly-used" in the VAR literature on the effects of monetary policy: a commodity price index, M2, the federal funds rate, real GDP, the personal consumption expenditure price index, and the unemployment rate. The classic VAR analysis of the effects of technology shocks in Galí (1999) uses two variables, labor productivity and hours worked. One can give many other examples. As we discuss below, recently VARs have been applied to large datasets following Bańbura et al. (2010).

Granger-causal-priority. While Granger-causal-priority is not a new concept, we believe that it is unfamiliar to most economists.[2] Granger-causal-priority is related to the well-known idea of *Granger-noncausality*, but it accounts better for the presence of variables other than y_i and y_j in the VAR. In the paper we define Granger-causal-priority formally and we describe how it differs from Granger-noncausality. Importantly, we explain why it is Granger-causal-priority that is tightly linked with variable choice in VARs.

The key point is that Granger-causal priority is central to sufficiency results for variable choice in VARs. If y_i is Granger-causally-prior to y_j, the forecasts of y_i obtained from a VAR with all variables y are *equal* to the forecasts of y_i obtained from a smaller VAR that omits y_j. In this sense y_j does not belong in the VAR to be used to predict y_i. Furthermore, we show that if y_i is Granger-causally-prior to y_j and an additional assumption holds, the impulse response of y_i to a structural shock of interest obtained from a VAR with all variables y is *equal* to the impulse response of y_i to that shock obtained from a smaller VAR that omits y_j. In this sense y_j does not belong in the VAR to be used to compute impulse responses of y_i to structural shocks.

In practice, based on a finite sample, we cannot *know* if y_i is Granger-causally-prior to y_j. We need tools for inference. We adopt the Bayesian approach to inference, in part because Bayesian VARs are popular in empirical work. As Bayesians, we can *infer the probability* that y_i is Granger-causally-prior to y_j given data. Furthermore, we can make an optimal decision, i.e., choose variables optimally, given data. We begin by working out a *closed-form* expression for the posterior probability of a Granger-noncausality restriction assuming that the VAR is Gaussian and the prior is conjugate.[3] We then derive a *closed-form* expression for the posterior probability that y_i is Granger-causally-prior to y_j. This result is important, because it implies that a researcher can evaluate the posterior probability of Granger-causal-priority quickly and accurately. Finally, we illustrate how after specifying a loss function, a researcher can make an optimal variable choice, i.e., the variable choice minimizing the posterior expected loss.

"Evaluating the posterior probability that y_i is Granger-causally-prior to y_j" is the same thing as "comparing the marginal likelihoods of VAR models of y with and without the re-

[2]The concept of Granger-causal-priority appears in unpublished work by Sims (2010) and Doan and Todd (2010).

[3]We also assume that the prior in the VAR with a Granger-noncausality restriction is consistent with the prior in the unrestricted VAR, in a sense that we make precise.

striction that y_i is Granger-causally-prior to y_j." As an alternative approach to variable choice in VARs, one can imagine eschewing marginal likelihood and computing the predictive density score of y_i implied by different VARs. In the paper we compare the two approaches conceptually.

To see how the methodology that we propose works, we investigate which variables belong in a quarterly VAR with real GDP, the price level, and the short-term interest rate. We perform the exercise twice, for the United States and for the euro area. In each exercise we consider thirty-eight macroeconomic and financial variables. The findings turn out to be remarkably similar between the United States and the euro area. Both in the United States and in the euro area we find that the following variables are most likely to belong in the VAR with real GDP, the price level, and the short-term interest rate: (i) survey-based indicators of economic sentiment and activity, (ii) a component of real GDP, the change in inventories, (iii) interest rates on government debt and private debt, and (iv) the price of oil.

We do not intend to argue that variable choice must occur formally in each application of VARs. We do want to suggest that: (i) variable choice can occur formally in a straightforward way, and (ii) even when choice of variables is informal, it is useful to know what assumptions are implicit and to what extent these assumptions are supported by the data. Moreover, we think that the question which variables to include in a macroeconomic time series model became more important after the financial crisis of 2008-2009.[4]

Our work makes contact with several strands of literature.

An important line of research is concerned with fitting linear time series models (VARs, factor models, and factor-augmented VARs) to *large* datasets.[5] We do not deprecate this line of research and we see our work as complementary. We believe that in some situations it is attractive to fit a VAR to a *subset* of the variables in one's dataset. The reason, we think, is that most economists prefer using the minimal means to get their points across, and most audiences and readers want to understand in simplest possible terms "where results come from." This reason explains why most applications of VARs involve variable choice.

In Bayesian statistics there is a controversy over whether or not to engage in model choice, and variable choice is an instance of model choice. There is also a related disagree-

[4]This point is made by Sims (2013) who argues that "we don't have a standard list of variables" now.

[5]See, e.g., Bańbura et al. (2010) in the case of VARs, Forni et al. (2000) and Stock and Watson (2002) in the case of factor models, and Bernanke et al. (2005) in the case of factor-augmented VARs.

ment about whether point-null hypothesis testing is worthwhile. We agree with Robert (2001) who argues in favor of model choice and variable choice as well as in favor of point-null hypothesis testing, though we are aware that others, cited by him, are skeptical.[6]

The methodology that we propose can guide the development of dynamic stochastic general equilibrium (DSGE) models.[7] For example, a vast literature extends the simple New Keynesian model by adding one variable or multiple variables. By "the simple New Keynesian model" we mean the well-known three-equation DSGE model that makes predictions about output, the price level (or inflation), and the short-term interest rate. We choose the same three variables as the variables of interest (i.e., as the elements of y_i) when we apply our methodology to data. Our findings suggest that if a researcher is interested in explaining the dynamics of output, the price level, and the short-term interest rate, then adding to the simple New Keynesian model survey data on expectations, inventories, interest rates on government debt and on private debt, or the price of oil is most likely to improve that model. The same findings can also be helpful to a researcher who wants to fit a statistical model that is computationally demanding – and therefore must be fit to at most a medium-sized dataset – and asks which variables will be most useful in that model.

The literature on fundamentalness, initiated by Hansen and Sargent (1991) and Lippi and Reichlin (1993), studies if structural shocks are fundamental, i.e., if one can obtain structural shocks from current and past values of data. In this literature it is known that failure to include in a VAR a variable that Granger-causes the included variables is a sufficient condition for nonfundamentalness. See Giannone and Reichlin (2006) and Forni and Gambetti (2012). The methodology that we develop identifies the minimal set of variables sufficient to eliminate this kind of nonfundamentalness.

The analytical expression for the posterior probability of a Granger-noncausality restriction that we derive is of independent interest. Since Granger (1969) and Sims (1972) there has been a significant interest in testing Granger-noncausality and, since Sims (1980), often in VARs. Tests of Granger-noncausality have been performed using the frequentist likelihood ratio test, even in Bayesian VARs, or using the Schwarz criterion which gives

[6] One alternative to model choice is model averaging. While we do not engage in model averaging in this paper, a researcher interested in model averaging will find useful the closed-form expressions for the posterior probability of Granger-noncausality and the posterior probability of Granger-causal-priority that we derive.

[7] An important application of VARs in macroeconomics has been to guide the development of DSGE models. See, e.g., Christiano et al. (2005) and Altig et al. (2011).

only an asymptotic approximation to a Bayesian test.[8] The properties of the likelihood ratio test of a zero restriction in a Bayesian VAR with an informative prior are unclear. Formal Bayesian tests have been possible in principle, though essentially unused in practice because they require cumbersome Monte Carlo.[9] Our analytical result allows to test Granger-noncausality in a standard Bayesian VAR without resorting to Monte Carlo.[10]

A growing literature initiated by George et al. (2008) studies Bayesian VARs with zero restrictions that are a priori independent across coefficients. By contrast, we are concerned with zero restrictions that apply to appropriate sets of coefficients. Furthermore, this literature aims at inference using a set of VARs with many different patterns of zero restrictions, via model averaging, whereas we are interested in variable choice.

Section 2 defines Granger-causal-priority and explains the relationship between Granger-causal-priority and the two questions that we study in this paper. Section 3 derives a closed-form expression for the posterior probability of Granger-noncausality in a Gaussian VAR with a conjugate prior. Section 4 shows how to evaluate the posterior probability of Granger-causal-priority. In Section 5 we apply the methodology to data. Section 6 discusses the concept of marginal likelihood, central to our methodology, comparing it with the predictive density score and with two other objects. Section 7 concludes.

2 Relation between Granger-causal-priority and choice of variables

This section defines Granger-causal-priority and explains the relationship between Granger-causal-priority and the two questions that we study in this paper.

Throughout the paper we assume that the set of variables y follows a VAR:

$$y(t) = \gamma + B(L)\, y(t-1) + u(t), \tag{1}$$

where $y(t)$ denotes y in period $t = 1, ..., T$, γ is a constant term, $B(L)$ is a matrix polynomial

[8] For instance, Cushman and Zha (1997) use the likelihood ratio test and Maćkowiak (2007) uses the Schwarz criterion. Both papers use Bayesian VARs.

[9] For example, one can use the Gibbs sampler developed by Waggoner and Zha (2003) to sample from the posterior density of the parameters of a VAR with a Granger-noncausality restriction and then use the method of Chib (1995) to compute from the Gibbs output the marginal likelihood implied by that VAR.

[10] Zha (1999) studies Bayesian inference in a structural VAR with recursive restrictions, but he does not consider testing such restrictions.

in the lag operator of order $P-1$, $P \geq 1$, and $u(t)$ is a Gaussian vector with mean zero and variance-covariance matrix Σ conditional on $y(t-s)$ for all $s \geq 1$. We denote with N the number of variables in y.

In this section we assume that the parameters of this data generating process are known, i.e., the values of γ, $B(p)$ for all $p = 1, ..., P$, and Σ are known. In the subsequent sections we consider inference. We then assume that a dataset with $T + P$ observations of y is available.

2.1 Granger-noncausality and Granger-causal-priority

Before defining Granger-causal-priority, it is helpful to recall the related, familiar concept of Granger-noncausality. Granger (1969) proposed that a variable z causes a variable x if the variable z helps predict the variable x. He formalized this idea in terms of the variance of the prediction error *one period ahead*. We state Granger's definition in the way in which it has been used in the VAR literature. Consider y_i and y_j, non-overlapping subvectors of y.

Definition 1 ***Granger-noncausality:*** *In the VAR given in equation (1), y_j does not Granger-cause y_i if the coefficients on all lags of y_j in the equations with y_i on the left-hand side are equal to zero, $B_{ij}(L) = 0$.*

The likelihood ratio test, which relies on an asymptotic χ^2 statistic, is a well-known frequentist test of the restriction $B_{ij}(L) = 0$.[11] In Section 3 we show how a Bayesian econometrician can evaluate the posterior probability of this restriction in a simple way.

Let us turn to Granger-causal-priority. The next definition, which we think is unfamiliar to most economists, appears in unpublished work by Sims (2010) and Doan and Todd (2010).

Definition 2 ***Granger-causal-priority:*** *In the VAR given in equation (1), y_i is Granger-causally-prior to y_j if it is possible to partition all the variables in y into two subsets, y_1 and y_2, such that $y_i \subseteq y_1$, $y_j \subseteq y_2$, and y_2 does not Granger-cause y_1.*

[11] See, e.g., Hamilton (1994), Chapter 11.

This definition states that y_i is Granger-causally-prior to y_j in VAR (1) if the VAR has the following recursive form:

$$\begin{matrix} y_i \rightarrow \\ y_j \rightarrow \end{matrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{pmatrix} \begin{pmatrix} y_1(t-1) \\ y_2(t-1) \end{pmatrix} + \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} \tag{2}$$
$$\textit{with } B_{12}(L) = 0.$$

Granger-causal-priority requires a *stronger* restriction than Granger-noncausality. If there are other variables in y in addition to y_i and y_j, the set of coefficients in $B_{ij}(L)$ is a *strict subset* of the set of coefficients in $B_{12}(L)$. In the special case when y consists only of y_i and y_j, the two restrictions are equivalent. Consider an example. Suppose that $y = \{x, w, z\}$, x, w, and z are scalars, $y_i = x$, $y_j = z$, and y follows the VAR

$$\begin{pmatrix} x(t) \\ w(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} B_{xx} & B_{xw} & B_{xz} \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix} \begin{pmatrix} x(t-1) \\ w(t-1) \\ z(t-1) \end{pmatrix} + u(t).$$

If $B_{xz} = 0$, $x(t+1) = B_{xx}x(t) + B_{xw}w(t) + u_1(t+1)$ and thus z does not Granger-cause x. However, $x(t+2) = \ldots + B_{xw}B_{wz}z(t) + \ldots$, i.e., the two-period-ahead forecast of x depends on the current value of z so long as $B_{xw}B_{wz} \neq 0$. If $B_{xw}B_{wz} \neq 0$, z helps predict x two periods ahead *indirectly*, through the effect of z on the third variable in the system, w.[12] To account for indirect effects, Dufour and Renault (1998) refine Granger's definition by defining Granger-noncausality *at a horizon* $h \geq 1$. In their terminology, z does not Granger-cause x at horizon $h = 2$ if $B_{xz} = 0$ *and* $B_{xw}B_{wz} = 0$. Dufour and Renault show that Granger-causal-priority, which they call "the separation condition," is a sufficient condition for Granger-noncausality *at all horizons*. In our example, x is Granger-causally-prior to z if *either* $B_{xz} = B_{xw} = 0$ *or* $B_{xz} = B_{wz} = 0$. In *either* case, our example VAR becomes recursive as in equation (2) with $x \subseteq y_1$ and $z \subseteq y_2$. Dufour and Renault also find a necessary-and-sufficient condition for Granger-noncausality at all horizons in a VAR of any dimension. This condition is very complex and thus difficult to test in practice.

Next, we explain how Granger-causal-priority informs the two questions that we study in this paper.

[12] This point is made by Lütkepohl (1993).

2.2 Granger-causal-priority and forecasting y_i with a VAR that omits y_j

Consider a researcher who wants to predict y_i. The following result is available.[13] Suppose that y_i is Granger-causally-prior to y_j, i.e., there exists an appropriate partition of y into y_1 and y_2 with $B_{12}(L) = 0$. Then the forecasts of y_i obtained with VAR (1) are *equal* at all horizons to the forecasts of y_i obtained with the VAR

$$y_1(t) = \gamma_1 + B_{11}(L)\, y_1(t-1) + u_1(t), \tag{3}$$

where the variance-covariance matrix of $u_1(t)$ is Σ_{11}, the appropriate submatrix of Σ. By "the forecasts are equal" we mean that, for given parameters γ, $B(p)$ for all $p = 1, ..., P$, and Σ and for given data $y(t-P+1), ..., y(t)$, the predictive density of $y_i(t+h)$ for any horizon $h \geq 1$ implied by model (2) is equal to the predictive density of $y_i(t+h)$ implied by model (3). Consequently, any point forecasts are also equal.[14]

This result has the following implication for variable choice. If y_i is Granger-causally-prior to y_j, the researcher can omit y_j (as well as all other variables in y_2) from the VAR to be used to forecast y_i and the forecasts of y_i do not change. Thus if y_i is Granger-causally-prior to y_j, y_j does not belong in the VAR to be used to forecast y_i.

Let us emphasize that to justify omitting y_j we need Granger-causal-priority; Granger-noncausality does not suffice. Granger-noncausality, i.e., $B_{ij}(L) = 0$ in model (1), does *not* imply that the forecasts of y_i obtained with model (1) are equal to the forecasts of y_i obtained with a smaller VAR that omits y_j *except* in the following two special cases: (i) if y consists only of y_i and y_j,[15] or (ii) if we want to forecast y_i only one period ahead.

Suppose that y_i is not Granger-causally-prior to y_j. Recall that Granger-causal-priority of y_i to y_j is a sufficient condition for y_j not to affect the forecasts of y_i at any horizon; it is not a necessary condition. Therefore, the absence of Granger-causal-priority of y_i to y_j does *not* imply that y_j *must* affect the forecasts of y_i. Since testing the necessary condition is difficult, we think that a simple and prudent rule is to include y_j in the VAR to be used to forecast y_i.

[13] The proof is straightforward.

[14] In this section we assume that the parameters of the data generating process are known. Below we discuss intuitive priors implying that the posterior predictive density of $y_i(t+h)$ (i.e., the predictive density that incorporates the uncertainty about the parameters) is the same in model (3) as in model (2). See Section 3.2 and Appendix B.

[15] Recall that in this special case Granger-noncausality is equivalent to Granger-causal-priority.

2.3 Granger-causal-priority and impulse responses of y_i from a VAR that omits y_j

Consider a researcher who wants to compute impulse responses of y_i to structural shocks. Let ε denote the structural shocks that generate the variation in y. We assume that the researcher is interested in the impulse response of y_i to a subset of the structural shocks $\varepsilon_k \subset \varepsilon$. Under what conditions is the impulse response of y_i to ε_k obtained from model (1) equal to the impulse response of y_i to ε_k obtained from model (3)? In a nutshell, the answer is that in addition to Granger-causal-priority we require a particular zero restriction on the contemporaneous impulse response of y to ε. Let us give the details.

Recursive substitution in model (1) implies that $y(t) = \delta + D(L)\,u(t)$, where δ is a constant term and $D(L)$ is a matrix polynomial in the lag operator of order infinity. We assume that there exists a matrix $C(0)$ such that $u(t) = C(0)\,\varepsilon(t)$ and $C(0)C(0)' = \Sigma$, where $\varepsilon(t)$ is a Gaussian vector with mean zero and variance-covariance matrix identity conditional on $y(t-s)$ for all $s \geq 1$. In words, we assume that one can obtain the structural shocks ε from current and past values of the data y. In the language of the literature initiated by Hansen and Sargent (1991) and Lippi and Reichlin (1993), we assume that ε is *fundamental* for y. We return to the issue of fundamentalness below. Given the assumption that ε is fundamental for y, we can write the impulse response of y to ε as $C(L)$, where $C(L)$ is a matrix polynomial in the lag operator of order infinity such that $C(L) = D(L)\,C(0)$. The impulse response of y_i to ε_k is given by $C_{ik}(L)$, i.e., the intersection of rows i and columns k of $C(L)$. Corollary 1 to Proposition 1 shows under what conditions the impulse response $C_{ik}(L)$ is a function *only* of $B_{11}(L)$ and Σ_{11} (i.e., $C_{ik}(L)$ is *not* a function of the other elements of $B(L)$ and Σ).

Proposition 1 *Consider the VAR given in equation (1) and the impulse response of y to ε given by $C(L)$. Suppose that: (i) one can partition y into two subsets, y_1 and y_2, such that y_2 does not Granger-cause y_1, and (ii) there exists a set of N_q variables $y_q \subseteq y_1$ that respond contemporaneously only to N_q structural shocks $\varepsilon_q \subset \varepsilon$. Then the impulse response of y_1 to ε_q, $C_{1q}(L)$, is a function only of $B_{11}(L)$ and Σ_{11}.*

Proof. See Appendix A. ■

Proposition 1 has the following implication for variable choice.

Corollary 1 *If y_i is Granger-causally-prior to y_j and assumption (ii) holds with $\varepsilon_k \subseteq \varepsilon_q$, the researcher can omit y_j (as well as all other variables in y_2) from the VAR to be used to compute the impulse response of y_i to ε_k and the impulse response does not change. Thus if y_i is Granger-causally-prior to y_j and assumption (ii) holds with $\varepsilon_k \subseteq \varepsilon_q$, y_j does not belong in the VAR to be used to compute the impulse response of y_i to ε_k.*

Corollary 1 is obtained simply by setting $y_i \subseteq y_1$, $y_j \subseteq y_2$, and $\varepsilon_k \subseteq \varepsilon_q$ in Proposition 1. Setting $y_i \subseteq y_1$ and $y_j \subseteq y_2$ in Proposition 1 means that y_i is Granger-causally-prior to y_j. Setting $\varepsilon_k \subseteq \varepsilon_q$ in Proposition 1 means that a subset of the variables in y_1 of size N_q respond contemporaneously only to N_q structural shocks *including the structural shocks of interest*. The impulse response of y_i to ε_k, $C_{ik}(L)$, is then a function only of $B_{11}(L)$ and Σ_{11}.

We make the following remarks about Proposition 1 and Corollary 1.

The case of $N_q = N_1$. Let N_1 denote the number of variables in y_1. If $N_q = N_1$, Proposition 1 simplifies. In particular, if $N_q = N_1$, by Proposition 1 we can write $y(t)$ as

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \delta + \begin{pmatrix} C_{11}(L) & 0 \\ C_{21}(L) & C_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}, \tag{4}$$

where $\varepsilon_1 = \varepsilon_q$. The impulse response of y to ε, $C(L)$, has a block of zeros corresponding to the impulse response of y_1 to ε_2. In words, if y_2 does not Granger-cause y_1 (assumption (i)) and y_1 responds contemporaneously only to N_1 structural shocks, ε_1 (a special case of assumption (ii)), y_1 responds only to the N_1 structural shocks ε_1 *at any horizon*. By Corollary 1, if y_i is Granger-causally-prior to y_j and the number of structural shocks that affect y_1 contemporaneously is equal to the number of variables in y_1, y_j does not belong in the VAR to be used to compute the impulse response of y_i to ε_k.

We find the assumption that $N_q = N_1$ natural, because this assumption is related to a standard assumption in the structural VAR literature. In the structural VAR literature, it is standard to assume that the number of structural shocks that affect the variables being modeled is equal to the number of the variables being modeled. For example, the typical researcher who computes impulse responses from a VAR with N_1 variables, like model (3), assumes that the variation in y_1 is generated by N_1 structural shocks.[16]

[16]This assumption is seldom explicit in the structural VAR literature. In a classic paper, Sims (1986) makes this assumption explicit when he writes in footnote 7 that his identification "requires (...) that the

The case of $N_q < N_1$. If $N_q < N_1$, y_1 responds to *all* N structural shocks, i.e., the impulse response of y to ε, $C(L)$, does *not* have a block of zeros as in equation (4). Nevertheless, Proposition 1 guarantees that a VAR with only N_1 variables y_1 suffices to compute the impulse response of y_1 to ε_q, $C_{1q}(L)$. By Corollary 1 this VAR suffices to compute the object of interest, the impulse response of y_i to ε_k $C_{ik}(L)$.

The case $N_q < N_1$ is important because it applies among others to recursiveness, perhaps the most popular identification approach in the structural VAR literature. Consider the recursive identification in the classic analysis of monetary policy in Christiano et al. (1999). Suppose that y_i consists of the variable controlled directly by monetary policy (e.g., the short-term interest rate), all variables that enter the reaction function of the central bank contemporaneously, and any other variables that the researcher is interested in. The variables that enter the reaction function of the central bank contemporaneously are ordered first, the variable controlled by monetary policy is ordered second, and the variables that do not enter the reaction function of the central bank contemporaneously are ordered third. Let y_q denote the variables that enter the reaction function of the central bank contemporaneously and the variable controlled directly by monetary policy.[17] Consider a variable y_j, $y_j \notin y_i$. Does y_j belong in the VAR to be used to compute the impulse response of y_i to monetary policy shocks? The recursive identification has the feature that y_q responds contemporaneously only to N_q structural shocks, where N_q denotes the number of variables in y_q. Therefore Proposition 1 applies. In particular, if y_i is Granger-causally-prior to y_j, y_j does not belong in the VAR to be used to compute the impulse response of y_i to monetary policy shocks under the identification of Christiano et al. (1999). The researcher can omit y_j and the impulse responses to monetary policy shocks are unaffected. Note that Granger-causal-priority plus the recursive identification suffice. One does not have to assume that the number of variables in the VAR match the number of structural shocks.

Identification. Typically a researcher knows a priori, e.g., based on an economic theory, that a particular variable is necessary for identification of a structural shock that the researcher is interested in. For instance, a researcher interested in identification of news shocks may hold the view that an indicator of economic sentiment is necessary for identification. We assume that if a researcher knows a priori that a variable is necessary for identification of a structural shock that the researcher is interested in, the researcher has

number of variables in the model match the number of behavioral disturbances (...)."

[17] Note that y_i consists of y_q and any other variables that the researcher is interested in.

included this variable in y_i. Which variables are necessary for identification is application-specific. The conditions that one needs to check in connection with identification are given in Rubio-Ramirez et al. (2010).

Fundamentalness. Suppose that y_i is not Granger-causally-prior to y_j. In the literature on fundamentalness it is known that failure to include in a VAR a variable that Granger-causes the included variables is a sufficient condition for nonfundamentalness. See, e.g., Giannone and Reichlin (2006), Proposition 1. Therefore, if y_i is not Granger-causally-prior to y_j and one omits y_j from model (3), ε_1 is nonfundamental for y_1. Nonfundamentalness need not affect *all* structural shocks. It may happen that the impulse response of y_i to a subset of ε_1 can still be computed from model (3). Nevertheless, we think that a simple and prudent rule is to include y_j in the VAR to be used to compute impulse responses of y_i.

2.4 Connecting the concept of Granger-causal-priority to data

Let us summarize the relation between Granger-causal-priority and the choice of variables. If y_i is Granger-causally-prior to y_j, the forecasts of y_i obtained from a VAR with all variables y are *equal* to the forecasts of y_i obtained from a smaller VAR that omits y_j. In this sense y_j does not belong in the VAR to be used to forecast y_i. Furthermore, if y_i is Granger-causally-prior to y_j and an additional assumption holds, the impulse response of y_i to a structural shock of interest obtained from a VAR with all variables y is *equal* to the impulse response of y_i to that shock obtained from a smaller VAR that omits y_j. In this sense y_j does not belong in the VAR to be used to compute impulse responses of y_i to structural shocks.

The principles stated in the previous paragraph guide the choice of variables so long as we *know*, for each $y_j \in y_J$, if y_i is Granger-causally-prior to y_j or not. In reality, we cannot know if y_i is Granger-causally-prior to y_j or not. As Bayesians, we can *infer the probability* that y_i is Granger-causally-prior to y_j given data. We show how to do it in Sections 3 and 4, thereby connecting the concept of Granger-causal-priority to data.

3 Granger-noncausality: a closed-form Bayes factor

We turn to inference. In this section we consider a Bayesian econometrician who wants to evaluate the posterior odds in favor of a hypothesis of Granger-noncausality.[18] The posterior odds are equal to the prior odds multiplied by the Bayes factor in favor of the hypothesis. We derive a *closed-form* expression for the Bayes factor in favor of Granger-noncausality in a Gaussian VAR with a conjugate prior. This result is of independent interest, because researchers often test Granger-noncausality. We use this result in the next section to evaluate the posterior probability of Granger-causal-priority.

Suppose that y follows the VAR given in equation (1). Let B be the $K \times N$ matrix of stacked coefficients, $B = (B(1), ..., B(P), \gamma)'$, where $K = NP + 1$ is the number of right-hand side variables in each equation in the VAR. The likelihood of the data implied by this VAR, conditional on initial observations, is

$$p(Y|B, \Sigma) = (2\pi)^{-NT/2} |\Sigma|^{-T/2} \exp\left(-\frac{1}{2} \operatorname{tr}(Y - XB)'(Y - XB)\Sigma^{-1}\right),$$

where

$$\underset{T\times N}{Y} = \begin{pmatrix} y(1)' \\ y(2)' \\ \vdots \\ y(T)' \end{pmatrix} \quad \text{and} \quad \underset{T\times K}{X} = \begin{pmatrix} y(0)' & y(-1)' & \dots & y(1-P)' & 1 \\ y(1)' & y(0)' & \dots & y(2-P)' & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ y(T-1)' & y(T-2)' & \dots & y(T-P)' & 1 \end{pmatrix}.$$

Consider a zero restriction on a subset of coefficients in this VAR. The restriction has the form that in each affected equation the coefficients of the *same* right-hand side variables are restricted. Formally, let α denote the indexes of a subset of the equations. Let β denote the indexes of a subset of the right-hand side variables. Consider the restriction

$$B_{\beta\alpha} = \mathbf{0}, \tag{5}$$

where $B_{\beta\alpha}$ denotes the matrix consisting of the intersection of rows β and columns α of the matrix B and $\mathbf{0}$ denotes the matrix of zeros of the same size as $B_{\beta\alpha}$. Note that a Granger-noncausality restriction is a special case of restriction (5).

[18] As is well known, reporting the posterior odds in favor of a hypothesis is equivalent to reporting the posterior probability of that hypothesis.

We turn to the specification of the prior in the unrestricted VAR and the prior in the restricted VAR.

3.1 Unrestricted VAR: conjugate prior and posterior

Let ω^U denote the unrestricted VAR. We assume that the prior density of B and Σ in the unrestricted VAR, $p(B, \Sigma|\omega^U)$, is conjugate and proper:

$$p(B, \Sigma|\omega^U) \propto |\Sigma|^{-(\tilde{\nu}+K+N+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\tilde{Y} - \tilde{X}B)'(\tilde{Y} - \tilde{X}B)\Sigma^{-1}\right), \tag{6}$$

where $\tilde{\nu} > N - 1$, $\tilde{Y}$ and $\tilde{X}$ are hyperparameters of appropriate dimensions, and $\tilde{X}'\tilde{X}$ and $\tilde{Y}'(I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}')\tilde{Y}$ are nonsingular. Note that the standard prior used in the VAR literature, the prior of Sims and Zha (1998) consisting of a modified Minnesota prior and additional dummy observations, is a special case of prior (6).

Expression (6) is a kernel of a normal-inverted-Wishart density. It is straightforward to show that

$$p(B, \Sigma|\omega^U) = p(B|\Sigma, \omega^U)p(\Sigma|\omega^U) = \mathcal{N}\left(\operatorname{vec}\tilde{B}, \Sigma \otimes \tilde{Q}\right)\mathcal{IW}\left(\tilde{S}, \tilde{\nu}\right), \tag{7}$$

where $\mathcal{N}$ denotes the multivariate normal density, $\mathcal{IW}$ denotes the inverted Wishart density,

$$\tilde{B} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}, \quad \tilde{Q} = (\tilde{X}'\tilde{X})^{-1}, \quad \text{and} \quad \tilde{S} = (\tilde{Y} - \tilde{X}\tilde{B})'(\tilde{Y} - \tilde{X}\tilde{B}).$$

Furthermore, the posterior density of B and Σ, $p(B, \Sigma|Y, \omega^U)$, satisfies

$$p(B, \Sigma|Y, \omega^U) = p(B|\Sigma, Y, \omega^U)p(\Sigma|Y, \omega^U) = \mathcal{N}\left(\operatorname{vec}\bar{B}, \Sigma \otimes \bar{Q}\right)\mathcal{IW}\left(\bar{S}, \bar{\nu}\right), \tag{8}$$

where

$$\bar{B} = \left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\bar{Y}, \quad \bar{Q} = \left(\bar{X}'\bar{X}\right)^{-1}, \quad \bar{S} = (\bar{Y} - \bar{X}\bar{B})'(\bar{Y} - \bar{X}\bar{B}),$$

$$\bar{X} = \begin{pmatrix}\tilde{X}\\ X\end{pmatrix}, \quad \bar{Y} = \begin{pmatrix}\tilde{Y}\\ Y\end{pmatrix}, \quad \text{and} \quad \bar{\nu} = \tilde{\nu} + T.$$

3.2 Prior in the restricted VAR

Let ω^R denote the restricted VAR. Let $B_{(\beta\alpha)}$ denote the vector collecting all coefficients in B other than the coefficients in $B_{\beta\alpha}$. We assume that the prior density of $B_{(\beta\alpha)}$ and Σ in the restricted VAR, $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$, satisfies

$$p(B_{(\beta\alpha)}, \Sigma|\omega^R) = p(B_{(\beta\alpha)}, \Sigma|\omega^U, B_{\beta\alpha} = \mathbf{0}). \tag{9}$$

Equation (9) states that the prior in the restricted model is equal to the prior in the unrestricted model conditional on the restriction. We think that assumption (9) is the most natural assumption one can make concerning the prior in model ω^R given a prior in model ω^U. Consider a researcher who holds prior $p(B, \Sigma|\omega^U)$. Suppose that this researcher obtains a new piece of information: the researcher learns that $B_{\beta\alpha} = 0$. Following the rules of probability, the researcher will update his or her prior precisely using equation (9). In addition to having this intuitive appeal, assumption (9) helps us derive a closed-form expression for the Bayes factor in favor of Granger-noncausality. See Section 3.3. We do not see how one can derive a closed-form expression for this Bayes factor without assumption (9).

The prior density in the restricted VAR defined in equation (9), $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$, is a conditional density of the normal-inverted-Wishart density $p(B, \Sigma|\omega^U)$. While $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$ itself is not normal-inverted-Wishart, some marginal densities of $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$ are normal-inverted-Wishart and have intuitive properties. As an example, Appendix B shows that the marginal density of the parameters of the equations indexed by α (i.e., the restricted equations) is normal-inverted-Wishart. Furthermore, this density's intuitive properties are pointed out. In particular, suppose that model ω^R is a VAR with Granger-causal-priority given in equation (2) with a prior satisfying equation (9). Then the posterior predictive density of y_1 in model (2) is the same as the posterior predictive density of y_1 in model (3) with a standard prior.

3.3 Closed-form Bayes factor

We are ready to state and prove the main result of this section: The Bayes factor in favor of restriction (5) can be expressed in closed-form.

Let $p(Y|\omega^R)$ denote the marginal likelihood of the data implied by the restricted model

ω^R. Let $p(Y|\omega^U)$ denote the marginal likelihood of the data implied by the unrestricted model ω^U.

Proposition 2 *The Bayes factor in favor of model ω^R, defined in expressions (1), (5), and (9), relative to model ω^U, defined in expressions (1) and (6), is given by*

$$\frac{p(Y|\omega^R)}{p(Y|\omega^U)} = \frac{\Gamma_{N_\alpha}\left(\frac{\bar{\nu}-N_{(\alpha)}+K_\beta}{2}\right)}{\Gamma_{N_\alpha}\left(\frac{\bar{\nu}-N_{(\alpha)}}{2}\right)} \frac{\Gamma_{N_\alpha}\left(\frac{\tilde{\nu}-N_{(\alpha)}}{2}\right)}{\Gamma_{N_\alpha}\left(\frac{\tilde{\nu}-N_{(\alpha)}+K_\beta}{2}\right)}$$

$$\times \frac{\left|\bar{S}_{\alpha\alpha}\right|^{\frac{\bar{\nu}-N_{(\alpha)}}{2}} \left|(\bar{Q}_{\beta\beta})^{-1}\right|^{\frac{N_\alpha}{2}} \left|\bar{S}_{\alpha\alpha}+\bar{B}'_{\beta\alpha}(\bar{Q}_{\beta\beta})^{-1}\bar{B}_{\beta\alpha}\right|^{-\frac{\bar{\nu}-N_{(\alpha)}+K_\beta}{2}}}{\left|\tilde{S}_{\alpha\alpha}\right|^{\frac{\tilde{\nu}-N_{(\alpha)}}{2}} \left|(\tilde{Q}_{\beta\beta})^{-1}\right|^{\frac{N_\alpha}{2}} \left|\tilde{S}_{\alpha\alpha}+\tilde{B}'_{\beta\alpha}(\tilde{Q}_{\beta\beta})^{-1}\tilde{B}_{\beta\alpha}\right|^{-\frac{\tilde{\nu}-N_{(\alpha)}+K_\beta}{2}}}, \tag{10}$$

where N_α is the number of the restricted equations, $N_{(\alpha)}$ is the number of the unrestricted equations, K_β is the number of the right-hand side variables whose coefficients are restricted, and $\Gamma_N(\cdot)$ denotes the multivariate gamma function of dimension N, $\Gamma_N(z) = \pi^{N(N-1)/4}\prod_{j=1}^{N}\Gamma\left(z+(1-j)/2\right)$.

Proof. Step 1a: Given equation (7), the marginal prior density of B is matricvariate Student, which implies that the marginal prior density of $B_{\beta\alpha}$ is also matricvariate Student. **Step 1b:** Given equation (8), the marginal posterior density of B is matricvariate Student, which implies that the marginal posterior density of $B_{\beta\alpha}$ is also matricvariate Student. Steps 1a-1b as well as the parameters of the two densities of $B_{\beta\alpha}$ follow directly from Bauwens et al. (1999), Appendix A.2.7. **Step 2:** The Savage-Dickey result of Dickey (1971) implies that if the prior in the restricted model ω^R satisfies condition (9), the Bayes factor in favor of the restricted model ω^R relative to the unrestricted model ω^U is equal to the ratio of the marginal posterior density of $B_{\beta\alpha}$ evaluated at $B_{\beta\alpha} = \mathbf{0}$ to the marginal prior density of $B_{\beta\alpha}$ evaluated at $B_{\beta\alpha} = \mathbf{0}$. Therefore, equation (10) is obtained as the ratio of the marginal posterior density of $B_{\beta\alpha}$ from Step 1b evaluated at $B_{\beta\alpha} = \mathbf{0}$ to the marginal prior density of $B_{\beta\alpha}$ from Step 1a evaluated at $B_{\beta\alpha} = \mathbf{0}$.[19] ■

Given Proposition 2, a researcher who wants to evaluate the posterior odds in favor of the hypothesis of Granger-noncausality can proceed as follows: (i) specify the prior odds

[19]Thus there is no need to evaluate any density implied by the restricted model ω^R; only the two densities associated with the unrestricted model ω^U, the marginal prior and the marginal posterior density of $B_{\beta\alpha}$, need to be evaluated.

in favor of model ω^R relative to model ω^U; it is common to specify the prior odds to be uninformative, i.e., the prior odds equal to one; (ii) use equation (10) to compute the Bayes factor in favor of model ω^R relative to model ω^U; and (iii) multiply the prior odds by the Bayes factor to obtain the posterior odds. The posterior odds in favor of model ω^R relative to model ω^U are the posterior odds in favor of the Granger-noncausality restriction.

In the next section we use Proposition 2 to evaluate the posterior probability of Granger-causal-priority.

4 Posterior probability of Granger-causal-priority

In this section we derive a *closed-form* expression for the posterior probability that y_i is Granger-causally-prior to y_j.

In Section 3 we evaluated the posterior probability of Granger-noncausality conditional on the set of models with two elements, the restricted model ω^R and the unrestricted model ω^U. By contrast, evaluating the posterior probability of Granger-causal-priority is complicated by the fact that there are multiple partitions of y consistent with Granger-causal-priority of y_i to y_j. In other words, there are *multiple restricted models* consistent with Granger-causal-priority of y_i to y_j.[20] Here we propose to evaluate the posterior probability of Granger-causal-priority conditional on the set of models Ω. Let us define Ω, explain how to evaluate the posterior probability of Granger-causal-priority conditional on Ω, and discuss why it is sensible to evaluate the posterior probability of Granger-causal-priority conditional on Ω.

Definition 3 *Let Ω be the set of models such that: (i) each model in Ω is a VAR of the form given in equation (1), (ii) Ω includes the unrestricted VAR, (iii) Ω includes all VARs with the restriction $B_{12}(L) = 0$ for some partition of y into two subsets, y_1 and y_2, such that $y_i \subseteq y_1$.*

We continue to assume as in Section 3 that the prior in the unrestricted model in Ω is conjugate and proper, i.e., satisfies expression (6), and the prior in each restricted model in

[20] This is true in the realistic case when there are variables in y that belong neither to y_i nor to y_j. In our example in Section 2.1 there are two partitions of y consistent with Granger-causal-priority of y_i to y_j, i.e., there are two restricted models consistent with Granger-causal-priority of y_i to y_j.

Ω satisfies condition (9) for appropriate α and β. Furthermore, we assume that all models in Ω have equal prior probabilities.[21]

Definition 4 *Let Ω^j be the subset of Ω containing all models in which y_i is Granger-causally-prior to y_j.*

We are ready to make the main point of this section: Evaluating the posterior probability that y_i is Granger-causally-prior to y_j conditional on Ω is equivalent to evaluating the posterior probability of Ω^j conditional on Ω, $p(\Omega^j|Y,\Omega)$. Furthermore, $p(\Omega^j|Y,\Omega)$ can be expressed in *closed-form*. Namely,

$$p(\Omega^j|Y,\Omega) = \frac{p(\Omega^j|Y)}{p(\Omega|Y)} = \frac{\sum_{\omega_k\in\Omega^j} p(\omega_k|Y)}{\sum_{\omega_l\in\Omega} p(\omega_l|Y)} = \frac{\sum_{\omega_k\in\Omega^j} p(Y|\omega_k)p(\omega_k)/p(Y)}{\sum_{\omega_l\in\Omega} p(Y|\omega_l)p(\omega_l)/p(Y)}$$

$$= \frac{\sum_{\omega_k\in\Omega^j} p(Y|\omega_k)}{\sum_{\omega_l\in\Omega} p(Y|\omega_l)} = \frac{\sum_{\omega_k\in\Omega^j} p(Y|\omega_k)/p(Y|\omega^U)}{\sum_{\omega_l\in\Omega} p(Y|\omega_l)/p(Y|\omega^U)}. \tag{11}$$

The first equality follows from the definition of conditional probability and the fact that $\Omega^j \subset \Omega$. The second equality follows from the definitions of Ω and Ω^j. The third equality follows from Bayes' law. The fourth equality follows from the assumption that the prior probability $p(\omega)$ is equal for all models; thus the terms $p(\omega_k)/p(Y)$ and $p(\omega_l)/p(Y)$ are all equal to one another. The fifth equality follows after we divide the numerator and the denominator by $p(Y|\omega^U)$. The final expression is a ratio of two sums of Bayes factors, where each Bayes factor has the form given in Proposition 2. See equation (10). Thus, the posterior probability that y_i is Granger-causally-prior to y_j can be expressed in closed-form.

There are two advantages of evaluating the posterior probability of Granger-causal-priority conditional on Ω, as proposed here. First, the posterior probability of Granger-causal-priority conditional on Ω can be expressed in closed-form. By contrast, the posterior probability of Granger-causal-priority conditional on some other set of models may be difficult to evaluate. For example, in Section 5.4 we study one generalization of Ω as a robustness check and this exercise is computationally very demanding. Second, Ω is defined so as to treat the null hypothesis and the alternative hypothesis *as symmetrically as possible*.[22] In

[21] It is a straightforward extension to consider the case when different models in Ω have different prior probabilities.

[22] It is always important to treat the null and the alternative symmetrically. For example, it is unappealing to specify that many models are consistent with the null, while few models are consistent with the alternative. This amounts to tilting the inference in favor of the null.

particular, if y_j consists of a single variable (as in the empirical application in this paper), Ω^j and its complement $\Omega \setminus \Omega^j$ have *equal size*. To see this, note that Ω contains 2^{N_J} models, where N_J denotes the number of variables in y_J. Furthermore, if y_j consists of a single variable, Ω^j contains 2^{N_J-1} models and $\Omega \setminus \Omega^j$ also contains 2^{N_J-1} models. Thus, if y_j consists of a single variable, evaluating the posterior probability of Granger-causal-priority amounts to evaluating the posterior odds in favor of a subset of models against the alternative of a subset of models of *equal size*.

Given equation (11) and given that y_j contains a single variable, a researcher who wants to evaluate the posterior probability that y_i is Granger-causally-prior to y_j proceeds as follows. The researcher begins attaching a prior probability of 0.5 to y_i being Granger-causally-prior to y_j.[23] The researcher then revises the prior belief in light of the data, via equation (11), arriving at the posterior probability of y_i being Granger-causally-prior to y_j.

Let us emphasize the following standard property of posterior probability.[24] Asymptotically, the posterior probability that y_i is Granger-causally-prior to y_j converges to *one* if y_i is Granger-causally-prior to y_j and converges to *zero* if y_i is not Granger-causally-prior to y_j. In other words, (i) a single model in Ω has a posterior probability of one asymptotically, and (ii) this model has in y_2 all variables that y_i is Granger-causally-prior to and has in y_1 all variables that y_i is not Granger-causally-prior to.

This property of posterior probability has an important implication. A researcher who evaluates the posterior probability that y_i is Granger-causally-prior to y_j, for each $y_j \in y_J$, asymptotically zeros in on the best model in Ω, the model with a posterior probability of one. Let y_1^* denote y_1 in this model and let y_2^* denote y_2 in this model. By the properties of Granger-causal-priority given in Section 2, the researcher can omit all variables in y_2^* from the VAR to be used to predict y_i or to compute impulse responses of y_i to structural shocks.[25] Furthermore, we think that a simple and prudent rule is to retain all variables in y_1^* in the VAR to be used to predict y_i or to compute impulse responses of y_i to structural shocks.[26]

[23]The prior probability is equal to 0.5 because: (i) all models in Ω are assumed to have equal prior probabilities, and (ii) Ω^j and its complement $\Omega \setminus \Omega^j$ have equal size. If either (i) or (ii) fails to hold, the prior probability will in general be some number between one and zero other than 0.5. The prior probability then gets updated in the same way, except that if (i) fails to hold, a trivial modification of equation (11) is required.

[24]See, e.g., Fernández-Villaverde and Rubio-Ramirez (2004).

[25]More precisely, this decision is justified in the case of a structural VAR if an additional assumption, assumption (ii) in Proposition 1, holds.

[26]We make this argument at the end of Section 2.2 in the case of reduced-form VARs and at the end of

Finally, we comment on computation. In principle, one can evaluate $p(\Omega^j|Y,\Omega)$ *exactly* using expression (11). However, if N_J is a large number, a present-day computer may be too slow to calculate all sums in expression (11). In our application N_J is a large number: $N_J = 38$. See Section 5. Therefore, in our application we approximate $p(\Omega^j|Y,\Omega)$ using the Markov chain Monte Carlo model composition algorithm of Madigan and York (1995). In Appendix D we explain how we implement this algorithm and assess its convergence. Here we emphasize that this algorithm is simple, fast and converges reliably. In our application, generating a Markov chain sufficient to produce reliable numerical approximations of all posterior probabilities, i.e., $p(\Omega^j|Y,\Omega)$ for all $y_j \in y_J$, takes about one hour on a standard personal computer.

5 Application

This section applies our methodology to data. In Section 5.1 we define y, y_i, and y_J in the data. We then state our prior (Section 5.2). In Section 5.3 we evaluate the posterior probability that y_i is Granger-causally-prior to y_j, for each $y_j \in y_J$ in the data, using the methodology from Section 4. From the previous discussion, we know what would happen if we used this methodology *asymptotically*: The posterior probability that y_i is Granger-causally-prior to y_j would be equal to either one or zero *for each* y_j. We could then use the posterior probabilities and the properties of Granger-causal-priority stated in Section 2 to make a decision about variable choice. What are we to do if we use this methodology *in a finite sample*? One possibility is to make a decision informally using the results of Section 5.3. For instance, one can rank y_j's according to the posterior probability that y_i is Granger-causally-prior to y_j and one can stipulate that a given y_j belongs in the VAR with y_i so long as this y_j is sufficiently well placed in the ranking; otherwise this y_j is to be omitted. Another possibility is to make a decision formally following Bayesian decision theory. This requires specifying a loss function, i.e., a function that assigns a numerical loss to every combination of actual variable choice and the correct variable choice, and minimizing the posterior expected loss.[27] We pursue this approach in Section 5.4. We specify a loss function and choose a set of variables by minimizing the posterior expected

Section 2.3 in the case of structural VARs.

[27] By the "correct variable choice" we mean the choice one would make in the case when one knew if y_i is Granger-causally-prior to y_j or not, for each $y_j \in y_J$ (as opposed to having to infer from data the probability that y_i is Granger-causally-prior to y_j).

loss. Section 5.5 reports root mean squared errors of point forecasts and gives an example of impulse responses. Section 5.6 summarizes the general lessons from Section 5.

5.1 Data: defining y, y_i, and y_J

We put together two datasets, one for the euro area and one for the United States. Each dataset has three variables of interest (i.e., in each dataset y_i has three elements) and thirty-eight other variables (i.e., in each dataset y_J has thirty-eight elements). Table 1 lists all variables.

The variables of interest (i.e., the elements of y_i) are: a measure of real output, a measure of the price level, and a measure of the short-term interest rate. We motivated this choice in Section 1. In particular, in the euro area exercise we use real gross domestic product (GDP), the harmonized index of consumer prices (HICP), and the overnight interbank interest rate Eonia. In the U.S. exercise we use real GDP, the consumer price index (CPI), and the federal funds rate.

In the euro area exercise the other variables (i.e., the elements of y_J) are: U.S. real GDP, the U.S. consumer price index, the federal funds rate, and thirty-five euro area variables listed in the next paragraph. In the U.S. exercise the other variables are: euro area real GDP, the HICP for the euro area, the Eonia, and thirty-five U.S. variables listed in the next paragraph.

The thirty-five variables are: (i) components of real GDP (consumption, government consumption, investment, exports, imports, change in inventories); (ii) labor market variables (unit labor cost, employment, unemployment rate, hours worked); (iii) interest rates (2-year and 10-year government bond yields, spread between corporate bonds rated BBB with maturity 7-10 years and government bonds with the same maturity, lending rate to non-financial corporations, mortgage interest rate); (iv) monetary aggregates (M1, M2, M3); (v) credit aggregates (government debt, loans for house purchase, consumer loans, loans to non-financial corporations); (vi) exchange rates (nominal exchange rate between the U.S. dollar and the euro, nominal effective exchange rate); (vii) commodity prices and other price indexes (oil price, index of commodity prices, consumer price index excluding energy and food, producer price index); (viii) housing market variables (house price index, real housing investment); (ix) stock market variables (stock index, stock volatility index); (x) survey-based indicators of economic activity and sentiment (capacity utilization, consumer

Table 1: Variable names, units and transformations

Variable	Units	Transformations	
Euro area real GDP	real currency units*	SA	log
Euro area HICP	index	SA	log
Eonia	percent		
U.S. real GDP	real currency units*	SA	log
U.S. CPI	index	SA	log
Federal funds rate	percent		
Real consumption	real currency units*	SA	log
Real government consumption	real currency units*	SA	log
Real investment	real currency units*	SA	log
Real exports	real currency units*	SA	log
Real imports	real currency units*	SA	log
Change in real inventories	percent of real GDP	SA	
Unit labor cost	index	SA	
Employment	thousands of people	SA	log
Unemployment rate	percent	SA	
Hours worked (U.S. only)	hours	SA	log
2-year government bond yield	percent		
10-year government bond yield	percent		
Corporate bond spread	percent		
Lending rate to NFCs	percent		
Mortgage interest rate	percent		
M1	nominal currency units†	SA	log
M2	nominal currency units†	SA	log
M3 (euro area only)	nominal currency units†	SA	log
Government debt	nominal currency units‡	SA	log
Loans for house purchase	nominal currency units†	SA	log
Consumer loans	nominal currency units†	SA	log
Loans to NFCs	nominal currency units†	SA	log
Dollar-euro exchange rate	dollars per euro		log
Nominal effective exchange rate	index		log
Oil price	dollar per barrel		log
Commodity prices	index		log
Consumer prices excl. energy, food	index	SA	log
Producer prices	index	SA	log
House prices	index	SA	log
Real housing investment	real currency units*	SA	log
Stock index	index		log
Stock volatility index	percent		log
Capacity utilization	percent	SA	
Consumer confidence	index	SA	
Industrial confidence	index	SA	
Purchasing managers' index	index	SA	

Notes: *Euro area: chained 2005 euros; United States: chained 2000 dollars. † Euro area: nominal index; United States: dollars. ‡Euro area: euros; United States: dollars.

confidence, industrial confidence, purchasing managers' index).[28]

The main sample contains quarterly data from 1999Q1 to 2012Q4. In the euro area exercise we decided to use data from 1999Q1, because this is when the monetary union started operating. We then decided to use the same period in the U.S. exercise, for the sake of comparability. In addition, we used a training sample in each exercise, as discussed in the next subsection.

5.2 Prior

The prior in the unrestricted model, i.e., in model $\omega^U \in \Omega$, consists of two pieces: (i) an *initial prior* formulated before seeing any data, and (ii) a *training sample prior*. Formally, matrices $\tilde{Y}$ and $\tilde{X}$ in expression (6) have the form

$$\tilde{Y} = \begin{pmatrix} Y_{SZ} \\ Y_{ts} \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} X_{SZ} \\ X_{ts} \end{pmatrix}, \tag{12}$$

and $\tilde{\nu} = \nu_{SZ} + T_{ts}$, where Y_{SZ}, Y_{ts}, X_{SZ}, X_{ts}, ν_{SZ}, and T_{ts} are defined next. Given this prior, the prior in each restricted model, i.e., in each model $\omega^R \in \Omega$, is defined by equation (9). The initial prior is the prior of Sims and Zha (1998), the standard prior used in the VAR literature, implemented by means of dummy observations Y_{SZ} and X_{SZ}. The training sample prior is implemented with data from the training sample, 1989Q1-1998Q4, collected in matrices Y_{ts} and X_{ts}. T_{ts} is equal to the number of periods in the training sample minus the number of lags, P.[29]

To choose the values of the hyperparameters controlling the prior of Sims and Zha (1998) we pursued an approach common in Bayesian econometrics: We searched for the hyperparameter values that maximize the marginal likelihood implied by model ω^U in the training sample. The hyperparameter values we found were close to the hyperparameter

[28] Due to data availability, we include hours worked only for the United States and we include M3 only for the euro area. The variables oil price and index of commodity prices are the same variables in the euro area exercise and in the U.S. exercise. We use the Dow Jones Euro STOXX index and the Dow Jones Industrial Average as the variable stock index in the euro area and in the United States, respectively. We use the VSTOXX (spliced with the VIX before the year 2000) and the VIX as the variable stock volatility index in the euro area and in the United States, respectively. The source of all data is the database of the ECB. The data are available from the authors upon request.

[29] We found that adding this training sample improved the marginal likelihood implied by the unrestricted model in the sample 1999Q1-2012Q4 compared with using only the prior of Sims and Zha (1998), both in the euro area exercise and in the U.S. exercise.

values used by Sims and Zha and common in the VAR literature. Appendix C discusses the prior in detail and reports the hyperparameter values. Below we refer to the prior with these hyperparameter values as the "baseline." We also report how our findings change as we vary the hyperparameter values.

5.3 Evaluating the probability of Granger-causal-priority

Table 2 reports the posterior probability that y_i (output, the price level, and the short-term interest rate) is Granger-causally-prior to a given y_j, for each $y_j \in y_J$, in the euro area (left column) and in the United States (right column).[30] Three main findings are evident:

(1) The posterior probability of Granger-causal-priority is close to zero for: (i) the change in inventories, (ii) survey-based indicators of economic sentiment (industrial confidence and consumer confidence), (iii) survey-based indicators of economic activity (purchasing managers' index in the euro area; purchasing managers' index and capacity utilization in the United States), (iv) interest rates on government debt and on private debt (the yield on 2-year government bonds and the lending rate to non-financial corporations in the euro area; the corporate bond spread, the yield on 2-year government bonds, and the lending rate to non-financial corporations in the United States), and (v) the price of oil.[31]

(2) The posterior probability of Granger-causal-priority is close to one for the dollar-euro exchange rate, house prices, loans for house purchase, and government debt.[32]

(3) The findings are remarkably similar between the euro area and the United States. The variables at the top of Table 2 are essentially the same variables in the euro area and in the United States. The variables at the bottom of Table 2 are essentially the same variables in the euro area and in the United States. The rank correlation between the posterior probabilities in the euro area exercise and in the U.S. exercise is 0.73.

Let us note two additional findings apparent from Table 2. First, no monetary aggregate

[30] Table 2 reports the findings obtained with one lag, i.e., $P = 1$. We chose to focus on the case $P = 1$ because the marginal likelihood implied by models in Ω was much higher with $P = 1$ than with any $P > 1$. See also Appendix C. Below we report how our findings change as we vary the number of lags, i.e., as we vary the value of P.

[31] The posterior probability of Granger-causal-priority is also close to zero for investment, exports, and imports in the euro area and for the unemployment rate and the Eonia in the United States. By "close to zero" we mean "smaller than 0.1."

[32] The posterior probability of Granger-causal-priority is also close to one for the nominal effective exchange rate, the consumer price index excluding energy and food, and the stock index in the euro area and for consumer loans, housing investment, and M2 in the United States. By "close to one" we mean "larger than 0.9."

Table 2: Posterior probability that output, price level, and short-term interest rate are Granger-causally-prior to a variable

Euro area			United States	
Variable	Prob.	Rank	Variable	Prob.
Change in inventories	0.00	1	**Oil price**	0.00
Industrial confidence	0.00	2	**Industrial confidence**	0.00
Purchasing managers' index	0.00	3	**Corporate bond spread**	0.00
2-year government bond yield	0.02	4	**Change in inventories**	0.01
Oil price	0.02	5	**2-year government bond yield**	0.01
Lending rate to NFCs	0.04	6	**Purchasing managers' index**	0.02
Investment	0.06	7	**Capacity utilization**	0.02
Exports	0.06	8	**Unemployment rate**	0.02
Imports	0.09	9	**Eonia**	0.04
Consumer confidence	0.10	10	**Consumer confidence**	0.04
Corporate bond spread	0.15	11	**Lending rate to NFCs**	0.06
Consumption	0.17	12	**Hours worked**	0.11
Mortgage interest rate	0.19	13	**Mortgage interest rate**	0.15
Unit labor cost	0.20	14	**10-year government bond yield**	0.15
Housing investment	0.21	15	**Consumption**	0.19
Unemployment rate	0.26	16	**Investment**	0.21
Employment	0.27	17	**Euro area GDP**	0.22
Federal funds rate	0.34	18	**Producer prices**	0.24
Capacity utilization	0.38	19	**Employment**	0.34
Producer prices	0.48	20	Imports	0.38
U.S. consumer prices	0.57	21	Stock index	0.49
10-year government bond yield	0.63	22	Nominal effective exchange rate	0.58
Government consumption	0.72	23	Commodity prices	0.64
Consumer loans	0.78	24	Exports	0.68
U.S. GDP	0.80	25	Euro area consumer prices	0.70
Loans to NFCs	0.81	26	Stock volatility index	0.70
M1	0.81	27	Government consumption	0.74
M2	0.83	28	Unit labor cost	0.77
M3	0.86	29	M1	0.77
Stock volatility index	0.87	30	Consumer prices excl. energy, food	0.84
Commodity prices	0.88	31	Loans to NFCs	0.87
Stock index	0.94	32	M2	0.92
Consumer prices excl. energy, food	0.96	33	Government debt	0.93
Nominal effective exchange rate	0.96	34	Housing investment	0.96
Government debt	0.99	35	Loans for house purchase	0.98
Loans for house purchase	0.99	36	Consumer loans	0.99
House prices	1.00	37	Dollar-euro exchange rate	0.99
Dollar-euro exchange rate	1.00	38	House prices	1.00

Note: Bold font indicates that a variable enters y_1 in the best model, where "the best model" is defined in Section 5.4.

and no credit aggregate ranks at the top of the table, either in the euro area or in the United States. This finding stands in contrast to the result that multiple interest rates, on government debt and on private debt, are important and rank at the top of the table. Second, both stock market variables (the stock index and the stock volatility index) rank in the lower half of Table 2 in the case of the euro area and in the case of the United States.

To study sensitivity of the findings reported in Table 2, we repeated the analysis in subsamples, varied the values of the hyperparameters of the prior, and changed the number of lags. Appendix E gives the details. Here we emphasize the main finding: The rank correlation between the posterior probabilities in the "baseline" case (in a given column of Table 2) and in any other specification that we considered is *high*. In particular, (i) this rank correlation is always higher than the rank correlation between the posterior probabilities in the euro area and in the United States in the "baseline" case, and (ii) the rank correlation between the posterior probabilities in the "baseline" case and in any subsample considered in Appendix E is higher than 0.9. In other words, the ranking of the variables is similar across the different specifications. Consequently, the answer to the question "Is a variable x more useful than a variable z?" is likely to be the same in the different specifications. We think that this finding is an important indicator of robustness.

5.4 Choosing the best variables

In this subsection we specify a loss function and choose a set of variables by minimizing the posterior expected loss. The main finding is that the variables chosen via this formal procedure are the variables ranked at the top of Table 2, i.e., the variables associated with lowest posterior probabilities of Granger-causal-priority. In other words, the formal approach leads us to make the same decision about variable choice that we can make based on the posterior probabilities of Granger-causal-priority alone, without specifying a loss function. We think that this finding is comforting.

We assume the commonly used *zero-one* loss function. This loss function is simple computationally and intuitive: A researcher with the zero-one loss function selects the model implying the highest marginal likelihood of the data. Let $\omega^* \in \Omega$ denote the model with the highest marginal likelihood. We refer to this model as "the best model." Note that model ω^* will have in y_1 the variables of interest (i.e., y_i) and possibly one or more other variables (i.e., possibly one or more y_j's). We refer to any y_j's in y_1 in model ω^* as "the

best variables."

We find that the best model in the euro area has in y_1 the twenty variables ranked at the top of the left column of Table 2 and listed there in bold font, from the change in inventories to producer prices. The best model in the United States has in y_1 the nineteen variables ranked at the top of the right column of Table 2 and listed there in bold font, from the price of oil to employment.

The best model fits the data very well. The log Bayes factor in the favor of the best model (model ω^*) relative to the unrestricted model (model ω^U) is 42 in the euro area and 21 in the United States. See Table 3. A log Bayes factor of 42 corresponds to a posterior odds of about 10^{18} to 1. A log Bayes factor of 21 corresponds to a posterior odds of about 10^9 to 1.

Choosing a subset of the available variables *wisely* improves fit, whereas choosing *any* subset of the available variables can deteriorate fit. Table 3 reports the log Bayes factor, relative to the unrestricted model, in favor of an arbitrarily chosen model with a Granger-noncausality restriction that is very different from the Granger-noncausality restriction in the best model. This log Bayes factor is -87 in the euro area and -91 in the United States, indicating overwhelming evidence in favor of the unrestricted model.

Table 3: Log Bayes factors in favor of individual models in Ω relative to the unrestricted model ω^U

	Euro area	United States
Best model in Ω, ω^*	42	21
Model "opposite" to ω^*	-87	-91

Note: The model "opposite" to ω^* is constructed so that y_1 in this model consists of y_i and all variables entering y_2 in model ω^*.

How sensitive are the findings concerning which variables are the best variables? We examined sensitivity in three ways. First, instead of looking only at the single best model we considered "the set of best models" defined as all models $\omega \in \Omega$ such that $\ln p(Y|\omega) > \ln p(Y|\omega^*) - 1$, i.e., the marginal likelihood of a model in the set of best models is within one log point of the marginal likelihood of model ω^*. The set of best models contains six models in the euro area and fourteen models in the United States. The models in the set of best models are very similar to each other and to model ω^*. For example, in the case of the

United States the seventeen variables ranked at the top of the right column of Table 2 (i.e., from the price of oil to euro area GDP) enter y_1 in *each* model in the set of best models.

Second, we searched for the best model in a set of models larger than Ω. Specifically, we studied the set of models $\tilde{\Omega}$ defined as the union of Ω and the set of models with *multiple* Granger-noncausality relations imposed simultaneously.[33] We were curious whether "allowing for more zero restrictions" than in the set Ω would lead to much better fit and a very different choice of variables. This turned out not to be the case. The marginal likelihood of the best model in $\tilde{\Omega}$ exceeds the marginal likelihood of the best model in Ω by only about 2 log points, in the euro area and in the United States. Furthermore, in the euro area the best model in $\tilde{\Omega}$ has in y_1 the same variables as the best model in Ω; in the United States, the best model in $\tilde{\Omega}$ has in y_1 the same variables as the best model in Ω except for one variable.[34]

Third, we searched for the best model in subsamples, assuming different values of the hyperparameters of the prior, and changing the number of lags. Appendix E gives the details. Here we summarize the findings as follows: A) In each specification that we considered the best model has in y_1 only or mainly the top ranking variables according to the posterior probability of Granger-causal-priority. The ranking of the variables is similar across the specifications (see the previous subsection), and thus the best model tends to include in y_1 the same variables in the different specifications. B) Assuming a prior tighter than the "baseline" does not reverse the conclusion from the "baseline" specification that the best model is not the unrestricted model, i.e., the best model has in y_1 only a subset of the variables in the dataset and y_2 is nonempty. C) The number of variables entering y_1 in the best model varies across the specifications, from eleven to thirty-two. At the same time, we do not find that assuming a prior tighter than the "baseline" systematically favors models with more variables entering y_1.

[33] Consider our example from Section 2.1. In this example, Ω includes models with a single Granger-noncausality relation, e.g., $\begin{pmatrix} B_{xx} & 0 & 0 \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$ and $\begin{pmatrix} B_{xx} & B_{xw} & 0 \\ B_{wx} & B_{ww} & 0 \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$. $\tilde{\Omega}$ in addition includes models with multiple Granger-noncausality relations, e.g., $\begin{pmatrix} B_{xx} & 0 & 0 \\ B_{wx} & B_{ww} & 0 \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$.

[34] Studying $\tilde{\Omega}$ has a serious disadvantage compared with studying Ω, because the Bayes factor for comparing models in $\tilde{\Omega}$ is not available in closed-form. It turns out that the computational burden increases enormously as one switches to $\tilde{\Omega}$ from Ω. Our exercise with $\tilde{\Omega}$ included only seventeen variables and took several days of computing time. The details of the exercise are in an online appendix available at http://www2.wiwi.hu-berlin.de/bartosz/VariableChoiceOA.pdf.

5.5 Root mean squared errors and impulse responses

This subsection reports root mean squared errors of point forecasts and gives an example of impulse responses.

The methodology developed in this paper relies on marginal likelihood. Marginal likelihood is an out-of-sample predictive density, as we discuss in the next section. Furthermore, knowing the value of marginal likelihood is necessary and sufficient for making probability statements about models given data. Therefore, knowing errors of out-of-sample point forecasts is of little additional value to Bayesians. However, reporting errors of out-of-sample point forecasts is a common practice. This being the case, we think it is worthwhile to see what the findings reported so far imply for errors of out-of-sample point forecasts.

We generated out-of-sample point forecasts from three different VARs. Each VAR includes y_i (i.e., output, the price level, and the short-term interest rate). The VAR that we refer to as "bottom 10" in addition to y_i includes the ten variables ranked at the bottom of Table 2, i.e., the ten variables associated with *highest* probabilities of Granger-causal-priority. We expect the VAR "bottom 10" to yield bad point forecasts. The VAR "top 10" in addition to y_i in includes the ten variables ranked at the top of Table 2, i.e., the ten variables associated with *lowest* probabilities of Granger-causal-priority. We expect the VAR "top 10" to yield good point forecasts. For comparison, we use the VAR "all" that includes all variables, i.e., y. For each quarter from 1999Q1 to 2010Q4 we fitted the three VARs to the data up to that quarter, and we generated out-of-sample point forecasts one quarter ahead and one year ahead. We used a single lag and the "baseline" prior including the training sample prior. We computed the point forecast as the median of the posterior predictive density. As a benchmark, we produced analogous forecasts from the random walk model with drift.

Table 4 reports the root mean squared errors (RMSEs) of the VAR point forecasts of output and the short-term interest rate, relative to the RMSEs of the random-walk-with-drift point forecasts. A number less than one indicates that a given VAR outperforms the random walk with drift. The VAR "top 10" yields good forecasts: (i) the VAR "top 10" typically outperforms the random walk with drift, (ii) the VAR "top 10" does about as well as or better than the VAR "all," and (iii) the VAR "top 10" always does better than the VAR "bottom 10."

We do not report RMSEs of the point forecasts of inflation. We found that the random

walk with drift outperformed each of the VARs in the case of inflation. This is a common finding. It is well known that during the Great Moderation the random walk with drift beats VARs in terms of RMSEs of point forecasts of inflation. Furthermore, in the case of inflation, the rankings of the three VARs based on the RMSEs of the point forecasts were inconsistent across forecast horizons as well as between the euro area and the United States. This result makes sense. When any VAR is beaten by the random walk with drift, the least bad VAR (i.e., the VAR closest to the random walk with drift) can be a VAR with "irrelevant" variables, like the variables in the VAR "bottom 10."

Table 4: RMSEs of VARs relative to RMSE of random-walk-with-drift

	Euro area			United States		
	"bottom 10"	"top 10"	"all"	"bottom 10"	"top 10"	"all"
One-quarter-ahead						
Output	0.93	0.85	**0.82**	1.02	**0.74**	0.77
Short-term interest rate	0.97	**0.70**	0.75	1.04	**0.79**	0.79
One-year-ahead						
Output	1.02	**0.99**	1.05	1.08	**0.76**	0.81
Short-term interest rate	1.14	**1.04**	1.36	1.09	**0.95**	0.97

Note: Bold font indicates the lowest RMSE for a given variable-horizon pair.

We now give an example of impulse responses. We consider the recursive identification of monetary policy used by Christiano et al. (1999) and popular in the literature. The variables that enter the reaction function of the central bank contemporaneously are ordered first, the short-term interest rate (the variable controlled by monetary policy) is ordered second, and the variables that do not enter the reaction function of the central bank contemporaneously are ordered third. As is common, we divide our variables into "slow-moving" and "fast-moving." We assume that the "slow-moving" variables enter the reaction function of the central bank contemporaneously, but do not respond to monetary policy shocks contemporaneously, and that "fast-moving" variables do not enter the reaction function of the central bank contemporaneously, but can respond to monetary policy shocks contemporaneously. The "fast-moving" variables in our dataset are interest rates, monetary aggregates, credit aggregates, exchange rates, and stock market variables. All other variables that we have data on are "slow-moving". We are not interested in whether

this identification yields reasonable impulse responses in our sample. We do not expect this to be the case, and searching for an identification of monetary policy that does produce reasonable impulse responses is beyond the scope of this paper.[35] We focus on the following question. Suppose that one uses the same identification in a small VAR (i.e., a VAR with a small subset of the variables that one has data on), a large VAR (i.e., a VAR with all variables that one has data on), and different medium-sized VARs. Suppose also that the impulse responses to the identified shock differ notably between the small VAR and the large VAR. Is it true that the impulse responses in a medium-sized VAR with a *wisely* chosen subset of the variables are very similar to the impulse responses in the large VAR? We show that the answer is "yes."

Figure 1 displays the impulse responses of selected euro area variables to a recursively identified shock to the Eonia from four different VARs. The VAR that we refer to as "small" includes in addition to y_i commodity prices, the nominal effective exchange rate, U.S. GDP, and the federal funds rate.[36] The VAR "worst variables" is a medium-sized VAR that includes in addition to y_i the eighteen variables ranked at the bottom of Table 2 (left column), i.e., the eighteen variables associated with *highest* probabilities of Granger-causal-priority. The VAR "best variables" is a medium-sized VAR that includes in addition to y_i the twenty variables ranked at the top of Table 2 (left column), i.e., the twenty variables associated with *lowest* probabilities of Granger-causal-priority. The VAR "large" includes all variables, i.e., y. As Figure 1 shows, the impulse responses differ notably between the VAR "small" and the VAR "large." In particular, the VAR "small" displays the price puzzle and permanent effects of the Eonia shock on output.[37] The impulse responses in the VAR "worst variables" are very similar to the impulse responses in the VAR "small." Crucially, the impulse responses in the VAR "best variables" are very similar to the impulse responses in the VAR "large."[38] In particular, the price puzzle disappears and the effects of the Eonia shock on output become transitory. We conclude that choosing a subset of the available variables wisely can yield impulse responses very similar to the impulse responses from a VAR that includes all available variables. This is an important result for a researcher with

[35] By "reasonable impulse responses" we mean impulse responses consistent with the intuition economists have about how the economy reacts to a monetary policy shock.

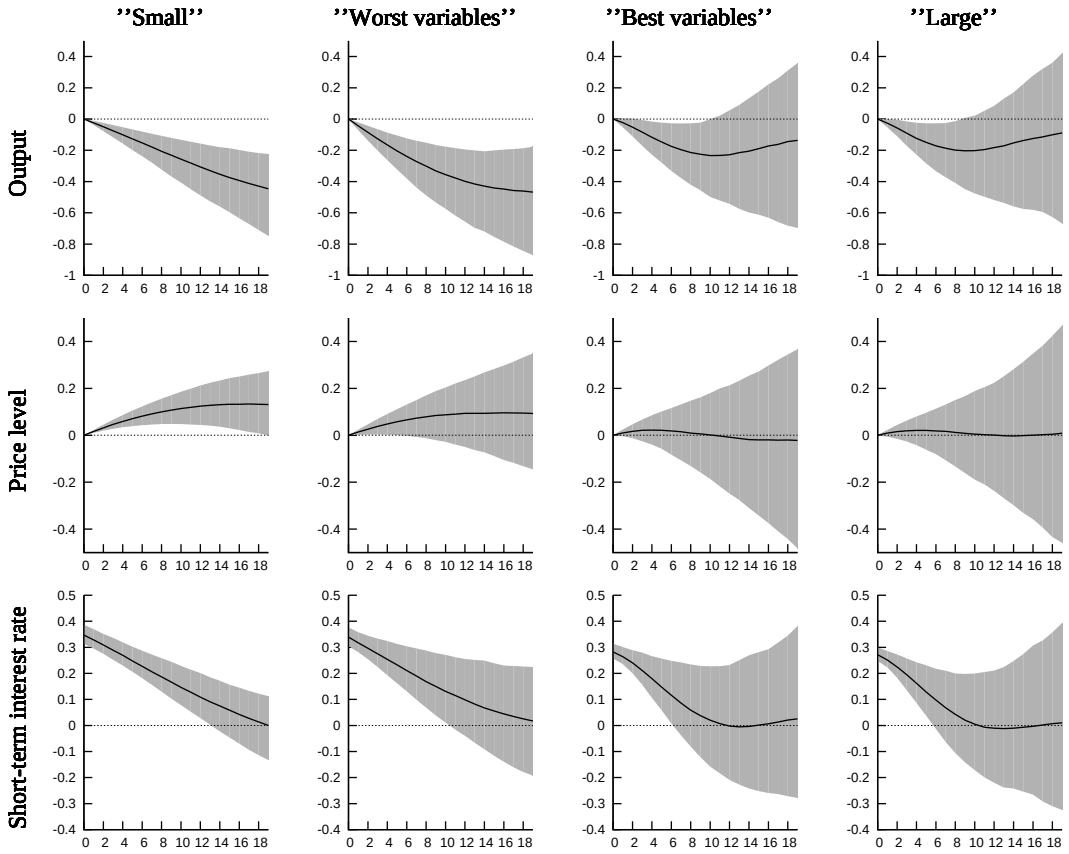
[36] This choice of variables is motivated by Peersman and Smets (2003), a well-known study of the effects of monetary policy shocks in the euro area.

[37] The "price puzzle," well known in the VAR literature on monetary policy, is the tendency for the price level to rise after what otherwise appears to be a monetary contraction.

[38] We obtained the same finding in the case of the United States.

a preference for fitting a structural VAR to a subset of the variables in his or her dataset.

Figure 1: Impulse responses to a recursively identified shock to short-term interest rate, euro area (posterior median, posterior 5th and 95th percentiles)



5.6 Summing up

Let us conclude this section with general lessons. Evaluating the posterior probability of Granger-causal-priority yields a ranking of the variables in one's dataset. Furthermore, a formal choice-of-variables procedure, based on a particular loss function, selects variables from the top of the ranking. This finding suggests that in practice variable choice can also occur informally, without specifying a loss function explicitly, simply by examining the posterior probabilities of Granger-causal-priority. A researcher with a preference for a small VAR will want to select only a few variables from the top of the ranking, whereas a researcher who is less constrained will choose a larger subset of the ranking. Another,

formal possibility is to specify a loss function, different from the zero-one loss function, that explicitly incorporates a preference for a VAR of a particular size, or preferences about other aspects of VAR specification.

6 Marginal likelihood versus other objects

The methodology developed in this paper relies on marginal likelihood. In particular, our methodology involves comparisons of the marginal likelihood of an unrestricted VAR with the marginal likelihoods of VARs with Granger-noncausality restrictions. See, in particular, Proposition 2 and equation (11). In this section we compare marginal likelihood with three other objects. We think that this will help readers understand our methodology.

As is well known, the marginal likelihood of the data Y implied by a model $\omega \in \Omega$ can be written as the product of one-period-ahead out-of-sample predictive densities:

$$p(Y|\omega) = p(y_i(1,...,T), y_J(1,...,T)|y_i(-P+1,...,0), y_J(-P+1,...,0), \omega)$$

$$= \prod_{t=0}^{T-1} p(y_i(t+1), y_J(t+1)|y_i(-P+1,...,t), y_J(-P+1,...,t), \omega). \quad (13)$$

First, equation (13) shows that marginal likelihood measures the *out-of-sample* predictive performance of a model. Hence, a high marginal likelihood of a VAR with a particular Granger-noncausality restriction constitutes evidence that this restriction improves *out-of-sample* forecasts. Second, one can rewrite equation (13) in terms of two-period-ahead out-of-sample predictive densities, three-period-ahead out-of-sample predictive densities, and so on. See, e.g., Geweke (2005), Section 2.6.2. Therefore marginal likelihood measures the *overall* out-of-sample predictive performance of a model, at any horizon. Third, the marginal likelihood of a model ω measures how well model ω fits *all* the data available to the researcher, Y. To stress this fact, we write (y_i, y_J) in equation (13) remembering that $y = (y_i, y_J)$. We think that confronting each model with *all* the data is consistent with the principle that "no information must be wasted" when one evaluates models. Fourth, knowing the value of marginal likelihood is necessary and sufficient for making probability statements about models given data. In other words, knowing the value of marginal likelihood leads one to "update one's prior" about models via Bayes' theorem.

An object with some popularity in the literature is the predictive density score.[39] The one-period-ahead predictive density score of y_i implied by a model $\omega \in \Omega$ is given by

$$g(Y_i|\omega, h=1) = \prod_{t=0}^{T-1} p(y_i(t+1)|y_i(-P+1,...,t), y_J(-P+1,...,t), \omega). \tag{14}$$

Consider using the predictive density score of y_i as a criterion for variable choice, instead of marginal likelihood. First, the predictive density score measures the out-of-sample predictive performance of a model, like marginal likelihood. Second, the predictive density score is specific to a given forecast horizon, h, *unlike* marginal likelihood. For instance, expression (14) assumes that the forecast horizon is one period, $h = 1$. One obtains a different expression and therefore a different value of the predictive density score for any other forecast horizon $h \neq 1$. Hence, a methodology based on the predictive density score of y_i will yield different choices of variables for different forecast horizons. We think that this will be a disadvantage in most contexts, because in most contexts – possibly always in structural VARs – researchers seek findings that are general and not specific to a particular forecast horizon.[40] Third, one can compute the predictive density score of a subset of the variables being modeled, here y_i, paying no attention to prediction of the other variables, here y_J. This feature can be seen as an advantage if the interest is in the models' fit only to Y_i. On the other hand, we see this feature as inconsistent with the principle that "no information must be wasted" when one evaluates models. Fourth, knowing the value of the predictive density score of y_i does *not* justify making probability statements given data. The practical implication is that when one uses the predictive density score of y_i, one cannot summarize the evidence in the data by reporting posterior probabilities, as we do in Section 5.3. Furthermore, one cannot establish that a particular model is most probable given the data, as we do in Section 5.4.

An additional important consideration is that the predictive density score is costly to compute. Computing the predictive density score requires a loop in which for every period t the researcher reestimates the model using the data up to period t, computes the predictive density and evaluates it at the actually observed data point. By contrast, the marginal likelihood of any model $\omega \in \Omega$ can be evaluated analytically given Proposition 2.[41]

[39] See, e.g., Amisano and Giacomini (2007) and Geweke and Amisano (2011).

[40] Of course, sometimes researchers are interested in prediction only for a particular horizon.

[41] A closed-form expression for the Bayes factor in favor of any restricted model $\omega^R \in \Omega$ is given in Proposition 2. The marginal likelihood of an unrestricted Gaussian VAR with a conjugate prior is standard

Two other objects may seem to be possible criteria for variable choice. In the rest of this section we argue that each object has a serious flaw.

Consider the following predictive density:

$$p(Y_i|\omega) = \int p(Y_i, Y_J|\omega) dY_J$$

$$= \prod_{t=0}^{T-1} p(y_i(t+1)|y_i(-P+1,...,t), y_J(-P+1,...,0), \omega). \quad (15)$$

One can think of $p(Y_i|\omega)$ as the marginal likelihood of the data $Y = (Y_i, Y_J)$ implied by a model $\omega \in \Omega$ "marginalized" with respect to Y_J. The predictive density $p(Y_i|\omega)$ measures the out-of-sample fit to the data on y_i *assuming that no data on* y_J *have been observed except for the initial observations.* Note the term $y_J(-P+1,...,0)$ in equation (15). Consider the following example. Suppose that we want to compare a VAR model ψ of y_i and $y_{J\psi}$ with another VAR model $\tilde{\psi}$ of y_i and another set of variables $y_{J\tilde{\psi}}$. Let $y_{J\tilde{\psi}}$ have the same number of variables as $y_{J\psi}$. Furthermore, let each VAR have one lag and the same prior, e.g., the prior of Sims and Zha (1998). Suppose that we rescale variables such that each variable in $y_{J\psi}$ and each variable in $y_{J\tilde{\psi}}$ have the same value in the initial period, $t = 0$. Then it is straightforward to show that $p(Y_i|\psi) = p(Y_i|\tilde{\psi})$. The implication is strong. If one uses the predictive density $p(Y_i|\psi)$ as a criterion to decide whether to include $y_{J\psi}$ or $y_{J\tilde{\psi}}$ in a VAR model with y_i, one ends up *indifferent.* Even if $y_{J\psi}$ is strongly related to y_i and $y_{J\tilde{\psi}}$ follows an independent white noise process![42]

Next, consider the following predictive density:

$$p(Y_i|Y_J, \omega) = \frac{p(Y_i, Y_J|\omega)}{\int p(Y_i, Y_J|\omega) dY_i}$$

$$= \prod_{t=0}^{T-1} p(y_i(t+1)|y_i(-P+1,...,t), y_J(-P+1,...,T), \omega). \quad (16)$$

The density $p(Y_i|Y_J, \omega)$ is the predictive density of Y_i conditional on the actually observed Y_J. This density measures the fit to the data on y_i *assuming that data on* y_J *have been*

and available in closed-form. Multiplying the Bayes factor by the marginal likelihood of the unrestricted VAR yields the marginal likelihood of the restricted VAR ω^R.

[42] If one uses a training sample prior in addition to the prior of Sims and Zha (1998), in this example it is no longer true that $p(Y_i|\psi)$ is literally equal to $p(Y_i|\tilde{\psi})$. In our application, we evaluated $p(Y_i|\omega)$ for many VARs using a training sample prior in addition to the prior of Sims and Zha (1998). We found that the differences between the values of $p(Y_i|\omega)$ across different VARs were very small rather than literally zero.

observed through the end of the sample, period T. Note the term $y_J(-P+1,...,T)$ in equation (16). The implication is strong. The density $p(Y_i|Y_J,\omega)$ is *not* a measure of out-of-sample predictive performance. The density $p(Y_i|Y_J,\omega)$ tells us how well model ω captures the relation between y_J and y_i given a particular Y_J, namely the actually observed Y_J. The density $p(Y_i|Y_J,\omega)$ can attain a high value for a given model ω, even if that model predicts *both* y_J *and* y_i poorly out-of-sample.

7 Conclusions

We develop a Bayesian methodology to choose variables to include in a reduced-form or structural VAR with an a priori given set of variables of interest. We rely on the idea of Granger-causal-priority, related to the well-known concept of Granger-noncausality. Our methodology is based on analytical results and thus it is simple to use. Applying the methodology to the case when the variables of interest are output, the price level, and the short-term interest rate, we find remarkably similar results for the euro area and the United States.

We think of the application shown in this paper as an illustration, certainly not the final word on which macroeconomic and financial variables interact most closely with output, the price level, and the short-term interest rate. We are interested in modeling the euro area economy and therefore we wanted to apply the methodology to euro area data. We then had to accept the fact that the euro area data sample is fairly short as well as the fact that a financial crisis occurs in the sample. In the future, it will be useful to redo this paper's analysis with models other than a VAR with a constant variance-covariance matrix which may fit better in this particular period, also in the United States. In some models, such as a VAR with stochastic volatility and a VAR with Markov-switching, the principle behind the choice of variables will be the same as the principle used in this paper. However, the computation of marginal likelihood will be much more complex than shown here.

A Proof of Proposition 1

Consider the VAR given in equation (1) and the impulse response of y to ε given by $C(L) = D(L)\,C(0)$. Partition y into subsets y_q, y_r, and y_2, with $y_q \cup y_r = y_1$. Partition ε into subsets ε_q, ε_r, and ε_2, where ε_q is of the same size as y_q, ε_r is of the same size as y_r, ε_2

is of the same size as y_2, and $\varepsilon_q \cup \varepsilon_r = \varepsilon_1$. The elements of ε can be ordered arbitrarily. Therefore, the subsets of ε are ordered ε_q, ε_r, ε_2 without loss of generality.

We can write the VAR given in equation (1) as

$$\begin{pmatrix} y_q(t) \\ y_r(t) \\ y_2(t) \end{pmatrix} = \gamma + \begin{pmatrix} B_{qq}(L) & B_{qr}(L) & 0 \\ B_{rq}(L) & B_{rr}(L) & 0 \\ B_{2q}(L) & B_{2r}(L) & B_{22}(L) \end{pmatrix} \begin{pmatrix} y_q(t-1) \\ y_r(t-1) \\ y_2(t-1) \end{pmatrix} + \begin{pmatrix} C_{qq}(0) & 0 & 0 \\ C_{rq}(0) & C_{rr}(0) & C_{r2}(0) \\ C_{2q}(0) & C_{2r}(0) & C_{22}(0) \end{pmatrix} \begin{pmatrix} \varepsilon_q(t) \\ \varepsilon_r(t) \\ \varepsilon_2(t) \end{pmatrix}.$$

The restrictions $B_{q2}(L) = 0$ and $B_{r2}(L) = 0$, i.e., $B_{12}(L) = 0$, follow from assumption (i). The restrictions $C_{qr}(0) = 0$ and $C_{q2}(0) = 0$ follow from assumption (ii). Recursive substitution yields

$$\begin{pmatrix} y_q(t) \\ y_r(t) \\ y_2(t) \end{pmatrix} = \delta + \begin{pmatrix} D_{qq}(L) & D_{qr}(L) & 0 \\ D_{rq}(L) & D_{rr}(L) & 0 \\ D_{2q}(L) & D_{2r}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} C_{qq}(0) & 0 & 0 \\ C_{rq}(0) & C_{rr}(0) & C_{r2}(0) \\ C_{2q}(0) & C_{2r}(0) & C_{22}(0) \end{pmatrix} \begin{pmatrix} \varepsilon_q(t) \\ \varepsilon_r(t) \\ \varepsilon_2(t) \end{pmatrix}.$$

The restrictions $D_{q2}(L) = 0$ and $D_{r2}(L) = 0$, i.e., $D_{12}(L) = 0$, follow from assumption (i). Furthermore, assumption (i) implies that $D_{qq}(L)$, $D_{qr}(L)$, $D_{rq}(L)$, and $D_{rr}(L)$ each is a function only of $B_{11}(L)$, i.e., $D_{11}(L)$ is a function only of $B_{11}(L)$.

We now argue that $C_{qq}(0)$ is a function only of Σ_{11} and $C_{rq}(0)$ is a function only of Σ_{11}. Recall that $C(0)C(0)' = \Sigma$. Therefore,

$$C_{qq}(0)C_{qq}(0)' = \Sigma_{qq},$$
$$C_{rq}(0)C_{qq}(0)' = \Sigma_{rq},$$

where Σ_{qq} is a submatrix of Σ_{11} and Σ_{rq} a is submatrix of Σ_{11}. The first equation implies that $C_{qq}(0)$ is a function only of Σ_{qq}. The second equation implies that $C_{rq}(0) = \Sigma_{rq}C_{qq}(0)'^{-1}$, i.e., $C_{rq}(0)$ is a function only of Σ_{qq} and Σ_{rq}. Hence, $C_{qq}(0)$ is a function only of Σ_{11} and $C_{rq}(0)$ is a function only of Σ_{11}.

Finally, matrix multiplication yields

$$\begin{pmatrix} y_q(t) \\ y_r(t) \\ y_2(t) \end{pmatrix} = \delta + \begin{pmatrix} C_{qq}(L) & C_{qr}(L) & C_{q2}(L) \\ C_{rq}(L) & C_{rr}(L) & C_{r2}(L) \\ C_{2q}(L) & C_{2r}(L) & C_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_q(t) \\ \varepsilon_r(t) \\ \varepsilon_2(t) \end{pmatrix}.$$

Consider $C_{qq}(L)$ and $C_{rq}(L)$, i.e., the impulse response of $y_1 = \{y_q, y_r\}$ to ε_q. We have

$$C_{qq}(L) = D_{qq}(L)C_{qq}(0) + D_{qr}(L)C_{rq}(0),$$
$$C_{rq}(L) = D_{rq}(L)C_{qq}(0) + D_{rr}(L)C_{rq}(0),$$

where all terms on the right-hand side of each of the two equations have been shown to be functions only of $B_{11}(L)$ and Σ_{11}. Thus the impulse responses of y_1 to ε_q, $C_{1q}(L)$, is a function only of $B_{11}(L)$ and Σ_{11}. ■

B Prior in the restricted VAR

Consider the prior density of $B_{(\beta\alpha)}$ and Σ in the restricted VAR, $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$. Recall that this density is defined in equation (9). In this appendix we focus on a particular marginal density of $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$. We define this density, state a proposition about it, and point out this density's intuitive properties.

Consider the coefficients in the equations indexed by α, i.e., the equations in which we impose zero restrictions. Let $B_{(\beta)\alpha}$ denote the matrix collecting the unrestricted coefficients in these equations. Recall that the remaining coefficients in these equations, collected in matrix $B_{\beta\alpha}$, are set to zero in equation (5). Let $\Sigma_{\alpha\alpha}$ denote the variance-covariance matrix of the innovations in these equations. Consider the prior density of $B_{(\beta)\alpha}$ and $\Sigma_{\alpha\alpha}$ in the restricted VAR, $p(B_{(\beta)\alpha}, \Sigma_{\alpha\alpha}|\omega^R)$. Note that $p(B_{(\beta)\alpha}, \Sigma_{\alpha\alpha}|\omega^R)$ is a marginal density of $p(B_{(\beta\alpha)}, \Sigma|\omega^R)$.

Proposition 3 *Consider model ω^R, defined in expressions (1), (5), and (9). The marginal*

prior density $p(B_{(\beta)\alpha}, \Sigma_{\alpha\alpha}|\omega^R)$ *is normal-inverted-Wishart and satisfies*

$$p(B_{(\beta)\alpha}, \Sigma_{\alpha\alpha}|\omega^R) \propto |\Sigma_{\alpha\alpha}|^{-(\tilde{\nu}_{(\beta)\alpha}+K_{(\beta)}+N_\alpha+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\tilde{Y}_\alpha - \tilde{X}_{(\beta)}B_{(\beta)\alpha})'(\tilde{Y}_\alpha - \tilde{X}_{(\beta)}B_{(\beta)\alpha})\Sigma_{\alpha\alpha}^{-1}\right). \quad (17)$$

Furthermore, the marginal posterior density $p(B_{(\beta)\alpha}, \Sigma_{\alpha\alpha}|Y, \omega^R)$ *is normal-inverted-Wishart and satisfies*

$$p(B_{(\beta)\alpha}, \Sigma_{\alpha\alpha}|Y, \omega^R) \propto |\Sigma_{\alpha\alpha}|^{-(\bar{\nu}_{(\beta)\alpha}+K_{(\beta)}+N_\alpha+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\bar{Y}_\alpha - \bar{X}_{(\beta)}B_{(\beta)\alpha})'(\bar{Y}_\alpha - \bar{X}_{(\beta)}B_{(\beta)\alpha})\Sigma_{\alpha\alpha}^{-1}\right). \quad (18)$$

The following notation appears in expressions (17)-(18): $\tilde{\nu}_{(\beta)\alpha} = \tilde{\nu} + K_\beta - N_{(\alpha)}$, $\bar{\nu}_{(\beta)\alpha} = \tilde{\nu}_{(\beta)\alpha} + T$, K_β *is the number of the right-hand side variables whose coefficients are restricted,* $K_{(\beta)}$ *is the number of the remaining right-hand side variables,* N_α *is the number of the restricted equations,* $N_{(\alpha)}$ *is the number of the unrestricted equations,* $\tilde{Y}_\alpha$ *denotes the columns of* $\tilde{Y}$ *corresponding to the equations indexed by* α*,* $\tilde{X}_{(\beta)}$ *denotes the columns of* $\tilde{X}$ *corresponding to the right-hand side variables indexed by* (β)*, i.e., the right-hand side variables whose coefficients are unrestricted,* $\bar{Y}_\alpha$ *denotes the columns of* $\bar{Y}$ *corresponding to the equations indexed by* α*, and* $\bar{X}_{(\beta)}$ *denotes the columns of* $\bar{X}$ *corresponding to the right-hand side variables indexed by* (β)*.*

Proof. Follows from Theorems A.17 and A.20 in Bauwens et al. (1999). ■

Consider a researcher whose prior in the unrestricted VAR, $p(B, \Sigma|\omega^U)$, is the prior of Sims and Zha (1998) with given hyperparameter values. Suppose that the researcher considers the restriction that y_2 does not Granger-cause y_1, i.e., $B_{12}(L) = 0$. Consider the marginal prior density of the parameters of the equations with y_1 on the left-hand side, $p(B_{(2)1}, \Sigma_{11}|\omega^R)$. By Proposition 3, $p(B_{(2)1}, \Sigma_{11}|\omega^R)$ is *the prior of Sims and Zha (1998) with the same hyperparameter values* as $p(B, \Sigma|\omega^U)$, *except* that the degrees of freedom of the density $p(\Sigma_{11}|\omega^R)$ are greater by $K_\beta - N_{(\alpha)}$ than the degrees of freedom of the density $p(\Sigma|\omega^U)$. Here $K_\beta - N_{(\alpha)} = N_2P - N_2$. Thus $K_\beta - N_{(\alpha)} = 0$ if $P = 1$, i.e., the hyperparameter values are exactly the same if the VAR has one lag. In our empirical application we use the prior of Sims and Zha (1998) and we have $P = 1$ in the baseline case and, therefore, the hyperparameter values in $p(B, \Sigma|\omega^U)$ and $p(B_{(2)1}, \Sigma_{11}|\omega^R)$ are exactly the same.

Next, suppose that the researcher forecasts y_1 using two models: (i) model ω^R, the VAR given in equation (2), and (ii) the VAR given in equation (3) with the prior of Sims and Zha (1998) with the same hyperparameter values as $p(B, \Sigma|\omega^U)$, except for the degrees of freedom correction. By Proposition 3, the posterior densities of $B_{11}(L)$ and Σ_{11} are identical in the two models. Consequently, the posterior predictive densities of y_1 are identical in the two models.

C Prior of Sims and Zha (1998)

This appendix contains the details of the prior of Sims and Zha (1998) including our choice of the hyperparameter values. The prior of Sims and Zha consists of four components controlled by scalar hyperparameters λ_1, λ_3, λ_4, μ_5, μ_6, and ν_{SZ}.[43] As explained below, we extend this set of hyperparameters with another scalar, μ_σ.

The first component is the Minnesota prior for B given by

$$p(\operatorname{vec} B|\Sigma) = \mathcal{N}\left(\operatorname{vec}\begin{pmatrix} I_N \\ 0_{K-N\times N}\end{pmatrix}, \Sigma \otimes WW'\right),$$

where W is a diagonal matrix of size $K \times K$ such that: (i) the diagonal entry corresponding to variable n and lag p is equal to $\lambda_1/(\sigma_n p^{\lambda_3})$, and (ii) the last diagonal entry, corresponding to the constant term, is equal to λ_4. We set σ_n equal to $\hat{\sigma}_n\mu_\sigma$, where $\hat{\sigma}_n$ is the standard deviation of the residuals from the univariate autoregression with P lags fit by ordinary least squares to the n'th time series, and μ_σ is a scaling factor common for all n. (Typically, in the literature $\mu_\sigma = 1$.) We implement the Minnesota prior with K dummy observations collected in the matrices

$$Y_{Minnesota} = W^{-1}\begin{pmatrix} I_N \\ 0_{K-N\times N}\end{pmatrix}, \quad X_{Minnesota} = W^{-1}.$$

The second component is the no-cointegration prior. We implement the no-cointegration prior with N dummy observations collected in the matrices

$$Y_{no-cointegration} = \mu_5 \operatorname{diag}(\bar{y}), \quad X_{no-cointegration} = \mu_5(\operatorname{diag}(\bar{y}), ..., \operatorname{diag}(\bar{y}), 0),$$

[43] We use the notation of Sims and Zha (1998) for the hyperparameters. Historically, there was also a hyperparameter λ_2, but this hyperparameter is always equal to 1 in the conjugate framework.

where $\bar{y} = (1/P)\sum_{t+P-1}^{0} y(t)$, i.e., $\bar{y}$ is equal to the average of the initial values of y, and $\text{diag}(x)$ denotes a diagonal matrix with vector x on the diagonal.

The third component is the one-unit-root prior. We implement the one-unit-root prior with the single dummy observation

$$Y_{one-unit-root} = \mu_6 \bar{y}', \quad X_{one-unit-root} = \mu_6(\bar{y}', ..., \bar{y}', 1).$$

The fourth component is the marginal prior about Σ, $p(\Sigma) = \mathcal{IW}(ZZ', \nu_{SZ})$, where Z is an $N \times N$ matrix and ν_{SZ} is a scalar hyperparameter. We set $Z = \sqrt{\nu_{SZ} - N - 1}\,\text{diag}(\sigma)$, where $\sigma = (\sigma_1, ..., \sigma_N)$. This choice of Z implies that, so long as $\nu_{SZ} > N + 1$, the prior expectation of Σ is

$$E(\Sigma) = \frac{ZZ'}{\nu_{SZ} - N - 1} = \text{diag}(\sigma^2).$$

Note that the density $p(\Sigma)$ satisfies

$$p(\Sigma) \propto |\Sigma|^{-(\nu_{SZ}+N+1)/2} \exp\left(-\frac{1}{2}\,\text{tr}\left(ZZ'\Sigma^{-1}\right)\right)$$

$$= |\Sigma|^{-(\nu_{SZ}+1)/2} |\Sigma|^{-N/2} \exp\left(-\frac{1}{2}\,\text{tr}\left(Z' - 0B\right)'\left(Z' - 0B\right)\Sigma^{-1}\right),$$

i.e., $p(\Sigma)$ is proportional to a likelihood of N observations with Z' on the left-hand side and $0_{N\times K}$ on the right-hand side multiplied by the factor $|\Sigma|^{-(\nu_{SZ}+1)/2}$. Therefore, we implement the marginal prior about Σ with N dummy observations given in the matrices

$$Y_\Sigma = Z', \quad X_\Sigma = 0_{N\times K}.$$

Collecting all dummy observations introduced here yields the matrices Y_{SZ} and X_{SZ} appearing in the main text in expression (12), i.e.,

$$Y_{SZ} = \begin{pmatrix} Y_{Minnesota} \\ Y_{one-unit-root} \\ Y_{no-cointegration} \\ Y_\Sigma \end{pmatrix}, \quad X_{SZ} = \begin{pmatrix} X_{Minnesota} \\ X_{one-unit-root} \\ X_{no-cointegration} \\ X_\Sigma \end{pmatrix}.$$

Let us explain how we selected the values of the hyperparameters λ_1, λ_3, λ_4, μ_5, μ_6, ν_{SZ}, and μ_σ. We searched for the hyperparameter values that maximize the marginal likelihood

Table 5: Hyperparameter values in the "baseline" prior and in alternative priors

Priors	λ_1	λ_3	λ_4	μ_5	μ_6	ν_{SZ}	μ_σ
"Baseline"	0.1	1	1	0.5	0.5	N+20	1
"Fixed ν_{SZ}"	0.2	1	1	0.5	0.5	N+2	3
"Tighter"	0.1	1	1	1	2	N+20	1
Sims and Zha (1998)	0.2	1	1	1	1	N+1	1

implied by the unrestricted VAR, i.e., by model $\omega^U \in \Omega$, in the training sample. Simultaneously, we searched for the marginal-likelihood-maximizing lag length P. We used a grid of values for each hyperparameter and lag length: $\lambda_1 \in \{0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4\}$, $\lambda_3 \in \{0.5, 1, 2\}$, $\lambda_4 \in \{0.5, 1, 2, 5\}$, $\mu_5 \in \{0, 0.5, 1, 2, 4\}$, $\mu_6 \in \{0, 0.5, 1, 2, 4\}$, $\nu_{SZ} \in \{N+2, N+10, N+20, N+30, N+40\}$, $\mu_\sigma \in \{0.25, 0.5, 1, 2, 3, 4\}$, $P = \{1, 2, 3, 4\}$. We found that the hyperparameter values that maximized the marginal likelihood in the euro area were different from those that maximized the marginal likelihood in the United States. However, the second-best hyperparameter values were the same in the euro area and in the United States. The marginal likelihood associated with the second-best hyperparameter values was approximately 1 log point lower than the maximum in the euro area and less than 2.5 log points lower than the maximum in the United States. We decided to use the second-best hyperparameter values as our "baseline" prior. Table 5 reports these hyperparameter values in the first row ("baseline"). These hyperparameter values are close to those used by Sims and Zha (1998) and common in the literature. We report the hyperparameter values used by Sims and Zha in the last row of Table 5. Furthermore, we found that in both the euro area and the United States the marginal likelihood in the training sample was maximized at lag length $P = 1$.

In the paper we report the findings obtained with the "baseline" prior and two alternative prior settings, referred to as "fixed ν_{SZ}" prior and "tighter" prior, respectively. To define the "fixed ν_{SZ}" prior, we repeated the maximization of the marginal likelihood in the training sample keeping the value of ν_{SZ} fixed at $\nu_{SZ} = N + 2$. Fixing ν_{SZ} at $N + 2$ is a common practice in the literature. We found that with ν_{SZ} fixed at $N + 2$ setting $\mu_\sigma = 3$ (i.e., $\hat{\sigma}$ scaled by a factor of 3) dominated all other values of μ_σ and that λ_1 should be higher than 0.1, while the best values of the other hyperparameters were not far from the "baseline"

prior. Therefore, in the "fixed ν_{SZ}" prior we decided to use $\nu_{SZ} = N + 2$, $\mu_\sigma = 3$, $\lambda_1 = 0.2$ and the same values of the hyperparameters λ_3, λ_4, μ_5, μ_6 as in the "baseline" prior. Note that, again, we found that the marginal likelihood in the training sample was maximized at $P = 1$. In the "tighter" prior, we set $\mu_5 = 2$ (weight of the "no-cointegration" dummy) and $\mu_6 = 1$ (weight of the "one-unit-root" dummy), while keeping all the other hyperparameters at their "baseline" values. This prior is tighter than the "baseline."

D Computational details

To study the set of models Ω we employ the Markov chain Monte Carlo model composition (MC^3) algorithm of Madigan and York (1995). MC^3 is used when: (i) one wants to obtain posterior results conditional on a set of models, and (ii) the set is too large to evaluate the posterior results of interest in each model in the set in reasonable time. MC^3 generates a Markov chain that moves through the set of models visiting any given model with a probability equal to that model's posterior probability. The researcher computes posterior results of interest based on the visited sample of models. By construction, the visited sample of models contains many models with high posterior probabilities and few models with low posterior probabilities. Since models with low posterior probabilities have little effect on the posterior results of interest, the approximation error caused by using only a subset of models quickly converges to zero.

For each model $\omega \in \Omega$, we define a set of models called the *neighborhood* of model ω, $nbr(\omega)$. Suppose that the chain is at some model ω. We randomly draw a candidate model ω' from $nbr(\omega)$ attaching the same probability to each model in $nbr(\omega)$. The chain moves from ω to ω' with probability

$$\min\left\{1, \frac{\#nbr(\omega)p(Y|\omega')}{\#nbr(\omega')p(Y|\omega)}\right\},$$

where $\#nbr(\omega)$ denotes the number of models in $nbr(\omega)$. With the complementary probability the chain stays at model ω, i.e., we record another occurrence of state ω. The process continues until the chain has the desired length. In each exercise (i.e., for each column in Tables 2, 6, and 7), we ran two chains of one million draws each, starting from maximally dispersed initial points. We estimated the posterior probabilities of Granger-causal-priority (henceforth, "GCP probabilities") in the way described below. In the tables we report the

average of the two estimates. We established convergence by verifying that the estimates from the two independent chains do not differ significantly. A chain of one million draws runs about one hour on a standard personal computer.

Definition of a neighborhood in the set Ω. The neighborhood of a model $\omega \in \Omega$ is the set of all models that differ from ω only by the position of one variable, i.e., models where one variable from y_1 of ω is in y_2 and models where one variable from y_2 of ω is in y_1. For each ω, we have $\#nbr(\omega) = N_J$.

Initial points for the two chains. The first chain starts at the unrestricted model, $y_1 = y$. The second chain starts at the model where y_i is Granger-causally-prior to all other variables, $y_1 = y_i$ and $y_2 = y_J$. These two initial points are maximally dispersed, because the number of moves necessary to get from one of the initial points to the other is weakly larger than for any other pair of models in Ω. We also experimented with random initial points, and the results were unaffected.

Estimator of GCP probability. In each chain we estimated the posterior probability that y_i is Granger-causally-prior to y_j, $p(\Omega^j|Y,\Omega)$, using as an estimator the frequency of visits of the chain in Ω^j. We discarded the first half of the chain (i.e., the first 500,000 states of the chain) to ensure that the results do not depend on the initialization. That is, we estimated the GCP probability as $\hat{p}(\Omega^j|Y,\Omega) = \left(\sum_{m=500,001}^{1000,000} \theta_m^j\right)/500,000$, where θ_m^j is the value of the indicator function taking the value of 1 if the m'th model in the chain belongs to Ω^j and 0 otherwise, $\theta_m^j \equiv I(\omega_m \in \Omega^j)$. We computed numerical standard deviations of $\hat{p}(\Omega^j|Y,\Omega)$ using the Newey-West estimator that accounts for the autocorrelation of θ_m^j up to order 500. Most of the autocorrelations go to zero long before 500, but for few variables the autocorrelation of the order 500 is equal to about 0.1. On average, the Newey-West standard error is equal to about 0.005.

Convergence diagnostics. We tested whether the GCP probabilities differ significantly between the two chains using the test statistic in Geweke (2005), Section 4.7. The joint chi-squared test for the equality of the two vectors of GCP probabilities was never rejected at the 5 percent level of significance. We also tested the equality of GCP probabilities associated with individual variables. The test statistic has the asymptotic standard normal distribution. The overwhelming majority of the test statistics lay below 2. However, in individual cases the test statistic exceeded 2. In these cases, we ran more simulations confirming that we kept obtaining probabilities within 0.01 of the reported numbers (i.e.,

within 0.01 of the numbers like those in Table 2).

Finding the best model. The following piece of evidence makes us confident that the best model visited by the chains that we ran is the best model in Ω: The ranking of more than two thousand top models, i.e., the best model and the models directly below the best model in terms of marginal likelihood, is *the same* in the two chains. The exact numbers are 2646 models in the euro area exercise and 4025 models in the U.S. exercise, in the "baseline" case.

We conclude that the findings reported in Section 5 and Appendix E are robust to Monte Carlo error.

E Sensitivity analysis

We describe how the results reported in Table 2 change as we redo the analysis in subsamples, vary the values of the hyperparameters of the prior, and change the number of lags.

We begin by repeating the analysis in subsamples: (i) we drop the last four quarters from the sample, i.e., the sample ends in 2011Q4, (ii) we drop the last eight quarters from the sample, i.e., the sample ends in 2010Q4, and (iii) we drop the last twelve quarters from the sample, i.e., the sample ends in 2009Q4. Note that dropping the last twelve quarters from the sample amounts to omitting as much as one-fifth of the data (in essence, the entire period after the financial crisis of 2008-2009). The results are in Table 6 (the United States) and in Table 7 (euro area). See the columns labeled "Dropping last..." 4Q, 8Q, and 12Q, respectively. For comparison, each of the two tables reproduces (column "baseline") the results from Table 2. Consider the case of the United States (Table 6). The rank correlation between the posterior probabilities in the "baseline" specification and in any subsample is high, 0.94 or more. Furthermore, in each subsample the best variables (in bold font) are the same or nearly the same as in the "baseline" specification. Recall that the best variables are the variables than enter y_1 in the best model, i.e., in the model attaining highest marginal likelihood in a given specification. Consider the case of the euro area (Table 7). The rank correlation between the posterior probabilities in the "baseline" specification and in any subsample is high, 0.9 or more, though somewhat lower than in the case of the United States. Furthermore, in each subsample the best variables are the same or nearly the same

as in the "baseline" specification, though there is somewhat less overlap compared with in the case of the United States.

Next, we compare the results obtained with the "baseline" prior with the results obtained with two alternative prior settings, referred to in Tables 6-7 as "fixed ν_{SZ}" and "tighter." The details of each of the prior settings are in Appendix C. We note the following three findings. First, the rank correlation between the posterior probabilities in the "baseline" specification and under each alternative prior setting is high, 0.92-0.93 and 0.82 in one case. Second, the number of variables entering y_1 in the best model varies across the different prior settings, from eleven to twenty-one. At the same time, we do not find that assuming a prior tighter than the "baseline" systematically favors models with more variables entering y_1. Third, Tables 6-7 report the log marginal likelihood of the data implied by the best model in Ω, $\log p(Y|\omega^*)$, and the log marginal likelihood of the data implied by the unrestricted model, $\log p(Y|\omega^U)$, in each specification. For each prior setting we find that $\log p(Y|\omega^*) > \log p(Y|\omega^U)$, in the United States and in the euro area. In particular, assuming a prior tighter than the "baseline" does not reverse the conclusion from the "baseline" specification that the best model is not the unrestricted model.[44]

Finally, we consider the effects of changing the number of lags. Tables 6-7 show the results with two lags, $P = 2$, assuming the "baseline" prior. The rank correlation between the posterior probabilities obtained with one lag and obtained with two lags is fairly high, 0.81 in the United States and 0.83 in the euro area. In principle, we could have defined Ω to include models with different values of P, instead of studying Ω conditional on $P = 1$ (Section 5) and redoing the study conditional on $P = 2$ (this appendix). However, it turns out that adding lags deteriorates the fit greatly and thus including in Ω models with different values of P would have had a negligible effect on the results reported in Section 5. To see this, note that Tables 6-7 report the log marginal likelihood of the data in a given specification implied by the set of models Ω, $\log p(Y|\Omega)$. The difference of 22 log points or more (compare the column "baseline" with the column "2 lags" in each table) corresponds to a posterior odds of 10^9 to 1 or higher, in favor of the specification with $P = 1$ relative to

[44] We condition on values of the hyperparameters. A different approach would be to specify a prior about the hyperparameters, as in Giannone et al. (2012), and report posterior probabilities of Granger-causal-priority having integrated over the hyperparameters. To implement this approach, we would need to run a Markov Chain alternating between a step in the space of VAR restrictions and a step in the space of the hyperparameters. We conjecture that this approach is feasible computationally, but we chose not to pursue it for fear that attention would have been diverted from the novelty in this paper.

Table 6: Sensitivity analysis, United States, posterior probability that output, price level, and short-term interest rate are Granger-causally-prior to a variable

Variable	"Baseline"	Dropping last...			Changing prior		2 lags
		4Q	8Q	12Q	"fixed ν_{SZ}"	"tighter"	
Oil price	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**
Industrial confidence	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
Corporate bond spread	**0.00**	**0.00**	**0.00**	**0.00**	0.19	**0.00**	**0.00**
Change in real inventories	**0.01**	**0.04**	**0.05**	**0.07**	**0.05**	**0.00**	**0.00**
2-year government bond yield	**0.01**	**0.02**	**0.03**	**0.05**	**0.05**	**0.31**	**0.00**
Purchasing managers' index	**0.02**	**0.02**	**0.01**	**0.01**	**0.00**	**0.10**	**0.00**
Capacity utilization	**0.02**	**0.04**	**0.06**	**0.06**	0.10	**0.04**	**0.01**
Unemployment rate	**0.02**	**0.04**	**0.00**	**0.00**	0.21	**0.00**	**0.02**
Eonia	**0.04**	**0.06**	**0.01**	**0.00**	**0.03**	0.85	0.88
Consumer confidence	**0.04**	**0.06**	**0.10**	**0.11**	**0.10**	**0.13**	**0.00**
Lending rate to NFCs	**0.06**	**0.09**	**0.14**	**0.21**	**0.03**	**0.05**	**0.00**
Hours worked	**0.11**	**0.12**	**0.10**	**0.10**	0.19	**0.10**	**0.02**
Mortgage interest rate	**0.15**	**0.17**	**0.27**	**0.35**	**0.16**	**0.23**	**0.01**
10-year government bond yield	**0.15**	**0.16**	**0.24**	**0.33**	**0.11**	**0.29**	**0.02**
Real consumption	**0.19**	**0.26**	**0.35**	**0.36**	0.21	**0.35**	**0.01**
Real investment	**0.21**	**0.18**	**0.16**	**0.17**	0.38	**0.07**	**0.02**
Euro area real GDP	**0.22**	**0.24**	**0.23**	**0.21**	0.21	0.69	0.78
Producers price index	**0.24**	**0.18**	**0.26**	**0.27**	**0.18**	**0.38**	**0.03**
Employment	**0.34**	**0.26**	**0.28**	**0.28**	0.48	**0.27**	**0.04**
Real imports	0.38	0.37	**0.38**	**0.38**	0.47	**0.23**	**0.01**
Stock index	0.49	0.54	0.58	0.58	0.57	0.80	**0.04**
Nominal effective exchange rate	0.58	0.56	0.58	0.48	0.36	**0.31**	**0.03**
Commodity prices	0.64	0.62	0.55	0.52	0.64	**0.41**	**0.07**
Real exports	0.68	0.69	0.64	0.59	0.44	0.74	**0.04**
Euro area consumer prices	0.70	0.62	0.62	0.51	0.41	0.67	**0.17**
Stock volatility index	0.70	0.59	0.43	**0.44**	0.56	0.51	**0.41**
Real government consumption	0.74	0.72	0.76	0.75	0.79	0.80	**0.21**
Unit labor cost	0.77	0.78	0.81	0.77	0.82	0.72	**0.07**
M1	0.77	0.73	0.46	**0.31**	0.81	**0.14**	**0.04**
Consumer prices excl. energy, food	0.84	0.63	0.47	0.48	0.69	0.77	**0.09**
Loans to NFCs	0.87	0.84	0.83	0.78	0.94	0.65	**0.18**
M2	0.92	0.89	0.74	0.57	0.89	0.98	0.92
Government debt	0.93	0.92	0.91	0.90	0.94	0.66	**0.23**
Real housing investment	0.96	0.86	0.84	0.82	0.94	0.90	**0.36**
Loans for house purchase	0.98	0.97	0.97	0.96	0.98	0.93	**0.31**
Consumer loans	0.99	0.99	1.00	0.96	0.96	0.89	0.74
Dollar-euro exchange rate	0.99	1.00	1.00	1.00	0.60	0.81	0.73
House prices	1.00	1.00	1.00	1.00	0.99	0.99	0.77
correlation with the baseline		0.99	0.96	0.93	0.95	0.80	0.48
rank correlation with the baseline		0.99	0.96	0.94	0.93	0.82	0.81
$\log p(Y\|\Omega)$	3489	-	-	-	3468	3464	3438
$\log p(Y\|\omega^*)$	3509	-	-	-	3489	3484	3461
$\log p(Y\|\omega^U)$	3489	-	-	-	3484	3464	3456

Notes: The column "baseline" reproduces the results from Table 2. Bold font indicates that a variable enters y_1 in the best model, where "the best model" is defined in Section 5.4.

Table 7: Sensitivity analysis, euro area, posterior probability that output, price level, and short-term interest rate are Granger-causally-prior to a variable

Variable	"Baseline"	Dropping last...			Changing prior		2 lags
		4Q	8Q	12Q	"fixed ν_{SZ}"	"tighter"	
Change in real inventories	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Industrial confidence	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Purchasing managers' index	**0.00**	**0.00**	**0.01**	**0.01**	**0.01**	**0.10**	**0.02**
2-year government bond yield	**0.02**	**0.02**	**0.03**	**0.03**	**0.28**	**0.28**	**0.00**
Oil price	**0.02**	**0.12**	0.42	0.34	**0.00**	**0.00**	**0.00**
Lending rate to NFCs	**0.04**	**0.05**	**0.02**	**0.03**	**0.04**	**0.07**	**0.00**
Real investment	**0.06**	**0.08**	**0.07**	**0.08**	**0.01**	**0.06**	**0.01**
Real exports	**0.06**	**0.07**	**0.06**	**0.08**	**0.01**	**0.07**	**0.00**
Real imports	**0.09**	**0.12**	**0.09**	**0.10**	**0.01**	**0.09**	**0.00**
Consumer confidence	**0.10**	**0.13**	0.30	0.24	**0.04**	**0.02**	0.96
Corporate bond spread	**0.15**	**0.04**	**0.10**	**0.15**	0.97	**0.00**	1.00
Real consumption	**0.17**	**0.21**	**0.10**	**0.11**	**0.03**	**0.19**	**0.00**
Mortgage interest rate	**0.19**	**0.18**	**0.07**	**0.12**	**0.26**	**0.15**	**0.04**
Unit labor cost	**0.20**	**0.26**	**0.18**	**0.25**	**0.06**	**0.29**	**0.01**
Real housing investment	**0.21**	**0.23**	**0.24**	**0.30**	**0.11**	**0.24**	0.61
Unemployment rate	**0.26**	**0.10**	**0.09**	**0.11**	0.81	**0.09**	0.99
Employment	**0.27**	**0.23**	0.30	**0.16**	0.63	**0.22**	**0.53**
Fed funds rate	**0.34**	**0.19**	0.95	0.97	0.96	**0.01**	1.00
Capacity utilization	**0.38**	**0.29**	0.41	**0.49**	0.99	0.63	1.00
Producers price index	**0.48**	**0.50**	0.75	0.76	0.93	0.62	0.90
U.S. consumer prices	0.57	0.60	0.60	0.56	0.90	0.67	0.56
10-year government bond yield	0.63	0.39	**0.21**	**0.30**	0.90	0.35	**0.48**
Real government consumption	0.72	0.58	**0.31**	**0.30**	0.98	0.58	0.96
Consumer loans	0.78	0.68	0.45	**0.24**	0.99	0.71	0.98
U.S. real GDP	0.80	0.79	0.83	0.82	1.00	0.70	1.00
Loans to NFCs	0.81	0.61	0.72	0.69	1.00	0.71	1.00
M1	0.81	0.85	0.84	0.81	0.84	0.90	0.96
M2	0.83	0.75	**0.34**	0.40	1.00	0.87	1.00
M3	0.86	0.78	**0.35**	0.42	1.00	0.90	1.00
Stock volatility index	0.87	0.88	0.98	0.97	1.00	0.89	1.00
Commodity prices	0.88	0.87	0.95	0.94	0.98	0.71	0.93
Stock index	0.94	0.95	0.98	0.97	1.00	0.91	1.00
Consumer prices excl. energy, food	0.96	0.93	0.99	0.99	1.00	0.82	1.00
Nominal effective exchange rate	0.96	0.96	0.97	0.97	0.99	0.96	1.00
Government debt	0.99	0.99	1.00	0.98	1.00	0.99	1.00
Loans for house purchase	0.99	0.99	0.99	0.97	1.00	0.99	1.00
House prices	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dollar-euro exchange rate	1.00	1.00	1.00	1.00	1.00	1.00	1.00
correlation with the baseline		0.98	0.85	0.85	0.86	0.95	0.79
rank correlation with the baseline		0.98	0.90	0.91	0.92	0.93	0.83
$\log p(Y\|\Omega)$	4852	-	-	-	4849	4814	4816
$\log p(Y\|\omega^*)$	4873	-	-	-	4873	4834	4840
$\log p(Y\|\omega^U)$	4831	-	-	-	4837	4797	4808

Notes: The column "baseline" reproduces the results from Table 2. Bold font indicates that a variable enters y_1 in the best model, where "the best model" is defined in Section 5.4.

the specification with $P = 2$. We conclude that in our application it is reasonable to focus on the findings conditional on $P = 1$.

References

Altig, D., Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2011). Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics*, 14(2):225–247.

Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, 25(2):177–190.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Bauwens, L., Lubrano, M., and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

Bernanke, B., Boivin, J., and Eliasz, P. S. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics 1A*, pages 65–148. North-Holland, Amsterdam.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Cushman, D. O. and Zha, T. (1997). Identifying monetary policy in a small open economy under flexible exchange rates. *Journal of Monetary Economics*, 39(3):433–448.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1):204–223.

Doan, T. A. and Todd, R. (2010). Causal orderings for multivariate linear systems. Unpublished note, Estima.

Dufour, J.-M. and Renault, E. (1998). Short run and long run causality in time series: theory. *Econometrica*, 66(5):1099–1125.

Fernández-Villaverde, J. and Rubio-Ramirez, J. F. (2004). Comparing dynamic equilibrium models to data: A Bayesian approach. *Journal of Econometrics*, 123(1):153–187.

Forni, M. and Gambetti, L. (2012). Sufficient information in structural VARs. Discussion paper, Universita di Modena e Reggio Emilia and Universitat Autonoma de Barcelona.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review*, 89(1):249–271.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley and Sons, Hoboken, New Jersey.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130–141.

Giannone, D., Lenza, M., and Primiceri, G. E. (2012). Prior selection for vector autoregressions. Discussion paper, Université Libre de Bruxelles, European Central Bank and Northwestern University.

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

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–38.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hansen, L. P. and Sargent, T. J. (1991). Two difficulties in interpreting vector autoregressions. In Hansen, L. P. and Sargent, T. J., editors, *Rational Expectations Econometrics*, pages 77–119. Westview, Boulder, Colorado.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Non Experimental Data*. John Wiley and Sons, New York.

Lippi, M. and Reichlin, L. (1993). The dynamic effects of aggregate demand and supply disturbances: Comment. *American Economic Review*, 83(3):644–52.

Lütkepohl, H. (1993). Testing for causation between two variables in higher-dimensional VAR models. In Schneeweiss, H. and Zimmermann, K. F., editors, *Studies in Applied Econometrics*, pages 75–91. Physica-Verlag, Heidelberg, Germany.

Maćkowiak, B. (2007). External shocks, U.S. monetary policy and macroeconomic fluctuations in emerging markets. *Journal of Monetary Economics*, 54(8):2512–2520.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232.

Peersman, G. and Smets, F. (2003). The monetary transmission mechanism in the euro area: Evidence from VAR analysis. In Angeloni, I., Kashyap, A., and Mojon, B., editors, *Monetary Policy Transmission in the Euro Area*, pages 36–55. Cambridge University Press, Cambridge.

Robert, C. P. (2001). *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*. Springer, New York.

Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77(2):665–696.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review*, 62(4):540–52.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter:2–16.

Sims, C. A. (1993). A nine-variable probabilistic macroeconomic forecasting model. In Stock, J. H. and Watson, M. W., editors, *Business Cycles, Indicators and Forecasting*, pages 179–212. University of Chicago Press, Chicago.

Sims, C. A. (2010). Lecture notes: Causal orderings and exogeneity. Available at http://sims.princeton.ftp.edu/yftp/times09/addednotes/gcp09.pdf.

Sims, C. A. (2013). Comment on "A macroeconomic framework for quantifying systemic risk" by He and Krishnamurthy. Available at http://www.mfmgroup.org/may-2013-66.html.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–68.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy? *American Economic Review*, 96(1):54–81.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–62.

Waggoner, D. F. and Zha, T. (2003). A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control*, 28(2):349–366.

Zha, T. (1999). Block recursion and structural vector autoregressions. *Journal of Econometrics*, 90(2):291–316.