Granger-Causal-Priority and Choice of Variables in Vector Autoregressions*

Marek Jarociński
European Central Bank

Bartosz Maćkowiak
European Central Bank and CEPR

First version: September 2011. This version: April 2014

Abstract

We derive a closed-form expression for the posterior probability of Granger-noncausality in a Gaussian vector autoregression with a conjugate prior. We also express in closed form the posterior probability of Granger-causal-priority, a more general relation that accounts for indirect effects between variables and therefore is suitable in a multivariate context. We show how to use these results to choose variables for a vector autoregression, whether the goal is prediction or impulse response analysis.

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

*Jarociński: European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany (e-mail: marek.jarocinski@ecb.int); Maćkowiak: European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany (e-mail: bartosz.mackowiak@ecb.int). 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. This is a revised version of the working paper under the same title circulated in 2013 as the ECB Working Paper no. 1600 and CEPR Discussion Paper no. 9686. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the European Central Bank.

1 Introduction

The vector autoregression (VAR) has been the dominant model in econometrics applied to macroeconomic data starting with Sims (1980). We highlight two features of the vast macroeconomic literature using VARs. First, the most common restriction tested and imposed in VARs is probably Granger-noncausality. There has been an interest in testing Granger-noncausality since Granger (1969) and Sims (1972) and in the context of VARs since Sims (1980). Second, almost all applications in the VAR literature involve small or medium-sized VARs.[1] In other words, many variables that economists have data on are left out. The starting point for this paper is the observation that Granger-causal relations and choice of variables for a VAR are tightly linked. We provide two analytical results useful for researchers wishing to test Granger-causal relations in a Bayesian VAR. Bayesian VARs have been popular following Doan et al. (1984). Furthermore, we show how to choose variables for a VAR using the two analytical results.

To begin, we work out a closed-form expression for the posterior probability of Granger-noncausality in a Gaussian VAR with a conjugate prior. Gaussianity and prior conjugacy are standard assumptions in Bayesian VARs. Up to now, Bayesians wishing to test Granger-noncausality had to rely on the frequentist likelihood ratio test or on cumbersome Monte Carlo.[2] Thanks to our result, a Bayesian test of Granger-noncausality becomes simple.

Building on this result, we also express in closed form the posterior probability of Granger-causal-priority. Although Granger-causal-priority is not a new concept, we believe that it is unfamiliar to most economists.[3] Unlike the well-known Granger-noncausality, Granger-causal-priority takes into account indirect effects between variables and therefore is more suitable than Granger-noncausality in a multivariate context.

Granger-causal-priority is tightly linked to choice of variables. 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 are interested in are: (i)

[1] We point out examples and exceptions below.

[2] We expand on this point below.

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

Can y_j be omitted from the VAR to be used to predict y_i? (ii) Can y_j be omitted from the VAR to be used to compute impulse responses of y_i to structural shocks?

We observe that it is natural to think of choosing variables as imposing a restriction on the data generating process followed by *all variables in the dataset*. Suppose that the set of variables we have data on, y, follows a VAR process. The decisive restriction on the data generating process is Granger-causal-priority. 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. Thus, y_j can be omitted from 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 condition 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. Thus, y_j can be omitted from the VAR to be used to compute the impulse response of y_i to the structural shock of interest. We emphasize that Granger-causal-priority is tightly linked to choice of variables in VARs not only when the objective is prediction, but also when the objective is impulse response analysis.

In practice, we do not *know* if y_i is Granger-causally-prior to y_j. We need tools for inference. 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. At this point, our analytical results turn out to be critical. Thanks to the closed-form expression for the posterior probability that y_i is Granger-causally-prior to y_j, we can evaluate this posterior probability quickly and accurately. Furthermore, after specifying a loss function, we can make an optimal choice of variables, i.e., make the choice of variables minimizing the posterior expected loss.

"Evaluating the posterior probability that y_j does not Granger-cause y_i" means "comparing the marginal likelihoods of VAR models of y with and without the restriction that y_j does not Granger-cause y_i." There are two models on the table: the restricted model and the unrestricted model. Similarly, "evaluating the posterior probability that y_i is Granger-causally-prior to y_j" means "comparing the marginal likelihoods of VAR models of y with and without the restriction that y_i is Granger-causally-prior to y_j." The difference is that there are *multiple restricted models* on the table, because multiple restricted models are consistent with Granger-causal-priority of y_i to y_j, as we will explain. We emphasize that in both cases, the Bayesian approach to testing is conceptually the same and we follow it:

"evaluate the marginal likelihoods, with and without the restriction, and use the marginal likelihoods to obtain the posterior probability of the restriction." As an alternative approach to the choice-of-variables problem, one can imagine eschewing marginal likelihood and computing the predictive density score of y_i implied by VARs with different variables. We compare the two approaches conceptually.

As an application, we investigate which variables belong in a quarterly VAR if the variables of interest (i.e., the elements of y_i) are real GDP, the price level and the short-term interest rate. We perform this exercise twice, for the euro area and the United States. In each exercise, we consider 38 macroeconomic and financial variables. The findings are similar between the euro area and the United States. Both in the euro area and in the United States, 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: survey-based indicators of economic sentiment and activity such as industrial confidence, interest rates on private debt and on government debt, selected components of real GDP such as the change in inventories, the unemployment rate, and the price of oil.

We do not argue that one *must* choose variables formally in each application of VARs. We do suggest that: (i) one *can* choose variables formally in a straightforward way, and (ii) even when the choice of variables is informal, it is useful to know what assumptions are implicit and to be able to check to what extent the 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]

We stated before that almost all papers in the VAR literature involve small or medium-sized VARs. One can give many well-known examples.[5] Furthermore, in the literature the choice of which variables to include in a VAR occurs informally, based on the researcher's prior or an informal specification search. By contrast, we choose variables formally.

An important line of research is concerned with fitting linear time series models (VARs, factor models, and factor-augmented VARs) to large datasets, with more than 100 vari-

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

[5]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, hours worked and labor productivity. One can give many other examples.

ables.[6] 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. One 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." Furthermore, we show that, at least in our application, a VAR with a carefully chosen Granger-noncausality restriction has a much higher posterior probability than the unrestricted, large VAR. Given this finding, it is a good idea to omit some variables from the large VAR (and we determine which variables to omit).

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. This literature tells us 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.

In Bayesian statistics, there is a controversy over whether or not to engage in model choice, and choosing variables is an instance of model choice. There is also a related disagreement about whether testing point-null hypotheses is worthwhile. We agree with Robert (2001) who argues in favor of model choice and choice of variables as well as in favor of point-null hypothesis testing, though we are aware that others, cited by him, are skeptical.[7]

We wrote before that prior to this paper, Bayesians wishing to test Granger-noncausality had to rely on the frequentist likelihood ratio test or on cumbersome Monte Carlo. The properties of the likelihood ratio test of a zero restriction in a Bayesian VAR with an informative prior are unclear. A formal Bayesian test has been possible, though essentially unused in practice, by applying the Gibbs sampler of Waggoner and Zha (2003) to sample from the posterior density of the parameters of a VAR with a Granger-noncausality restriction and then using the method of Chib (1995) to compute from the Gibbs output

[6] 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.

[7] 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 our analytical results.

the marginal likelihood of that VAR. The marginal likelihood of an unrestricted Gaussian VAR with a conjugate prior is standard and available in closed form.

A literature initiated by George et al. (2008) studies Bayesian VARs with zero restrictions 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 choosing variables.[8]

The methodology for the choice of variables that we propose can guide the development of dynamic stochastic general equilibrium (DSGE) models.[9] For example, a large 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 when we apply our methodology to data. The 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 data on sentiment or expectations, interest rates on private debt and on government debt, inventories, unemployment 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.

Section 2 defines Granger-causal-priority and explains the relationship between Granger-causal-priority and choice of variables in VARs. 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 this paper's methodological results to data. Section 6 discusses the concept of marginal likelihood, central to this paper, comparing it with the predictive density score and with two other objects. Section 7 concludes.

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

[9] 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).

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 choice of variables in VARs.

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 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

Let us begin by recalling the 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.

Definition 1 ***Granger-noncausality:*** *Consider y_i and y_j, non-overlapping subvectors of y. 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 $B_{ij}(L) = 0$.[10] 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).

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

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.*

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

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

In the special case when y consists only of y_i and y_j, Granger-causal-priority is the same as 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 $B_{12}(L)$, i.e., Granger-causal-priority requires a *stronger* restriction than Granger-noncausality. 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.[11] 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

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

any dimension. This condition is very complex and thus difficult to test in practice.

Next, we explain the relationship between Granger-causal-priority and choice of variables in VARs.

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.[12] 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.[13]

This result has the following implication for the choice of variables: 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. 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,[14] or (ii) if we want to forecast y_i only one period ahead.

What if y_i is not Granger-causally-prior to y_j? 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.[15] However, testing the necessary-and-sufficient condition is difficult

[12] The proof is straightforward.

[13] Recall that 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.

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

[15] Put differently, the absence of Granger-causal-priority of y_i to y_j does not imply that y_j must affect the

and essentially not done in practice, whereas we show below that evaluating the posterior probability of Granger-causal-priority is straightforward. Taking all this into consideration, we think that a simple and prudent approach is to check if y_i is Granger-causally-prior to y_j and, if it is not, to include y_j in the VAR to be used to forecast y_i.

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

Granger-causal-priority is tightly linked to choice of variables in VARs not only when the objective is prediction, but also when the objective is impulse response analysis. 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: (i) the researcher is interested in the impulse response of y_i to a subset of the structural shocks $\varepsilon_k \subset \varepsilon$, (ii) the researcher has specified an identification scheme, and (iii) if the scheme implies that a variable is necessary for identification, this variable is included in y_i.[16] 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

forecasts of y_i. In particular, it may happen that y_j helps forecast a linear combination of variables that does not help forecast y_i, even though each of these variables individually helps forecast y_i.

[16] Which variables are necessary for identification is application-specific. For example, a researcher interested in the impulse response of hours worked to technology may hold the view that a measure of labor productivity is necessary for identification. The conditions that one needs to check in connection with identification are given in Rubio-Ramírez et al. (2010).

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 $C(L)$. Proposition 1 gives the conditions guaranteeing that 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 $y_i \subseteq y_1$, $y_j \subseteq y_2$, and y_2 does not Granger-cause y_1; (ii) there exists a set of N_q variables $y_q \subseteq y_1$ that respond contemporaneously only to N_q structural shocks $\varepsilon_q \subset \varepsilon$ and $\varepsilon_k \subseteq \varepsilon_q$. Then the impulse response of y_i to ε_k, $C_{ik}(L)$, is a function only of $B_{11}(L)$ and Σ_{11}.*

Proof. See Appendix A. ■

Condition (i) means that y_i is Granger-causally-prior to y_j. Condition (ii) 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* ε_k. Proposition 1 has the following implication for the choice of variables: If conditions (i) and (ii) hold, 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.

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 (condition (i)) and y_1 responds contemporaneously only to N_1 structural shocks, ε_1 (a special case of condition (ii)), y_1 responds only to the N_1 structural shocks ε_1 *at any horizon*. The implication for variable choice is straightforward: 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 can be omitted from the VAR to be used to compute the impulse response of y_i to ε_k.

We find the assumption $N_q = N_1$ natural, and therefore we think that Proposition 1 is applicable in many practical situations. The assumption $N_q = N_1$ 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.[17] 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. In effect, this researcher says: "I assume that N_1 variables matter and I assume that N_1 structural shocks matter." Thanks to Proposition 1, researchers will be able to say: "I use N_1 variables because I checked that my variables of interest are Granger-causally-prior to the other variables I have data on, and I assume that N_1 structural shocks matter contemporaneously." The assumption in the latter statement is weaker than the two assumptions in the former statement.

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_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. Again, this makes us think that Proposition 1 is applicable in many practical situations. 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 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.[18] Consider a variable y_j, $y_j \notin y_i$. Can y_j be omitted from 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,

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

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

Proposition 1 applies: If y_i is Granger-causally-prior to y_j, y_j can be omitted from 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). We emphasize 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.[19]

What if y_i is not Granger-causally-prior to y_j? The literature on fundamentalness tells us 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. This seems like a strong reason to retain y_j. However, 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). Keeping this subtlety in mind, we think that a simple and prudent approach is to check if y_i is Granger-causally-prior to y_j and, if it is not, to include y_j in the VAR to be used to compute the impulse response of y_i to ε_k.

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

Let us summarize the relation between Granger-causal-priority and 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. Thus, y_j can be omitted from the VAR to be used to predict y_i. Furthermore, if y_i is Granger-causally-prior to y_j and an additional condition 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. Thus, y_j can be omitted from the VAR to be used to compute the impulse response of y_i to the structural shock of interest.

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 practice, we do not 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.

[19] It is straightforward to check if condition (ii) in Proposition 1 holds also in the case of a non-recursive identification.

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.[20] 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.

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).

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

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

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.[21]

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

[21] We appealed to this result in footnote 13.

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 $B_{\beta\alpha} = \mathbf{0}$.[22] ■

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 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

[22]Thus, there is no need to evaluate any density implied by the restricted model ω^R. Only 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. Furthermore, recall that the marginal likelihood of an unrestricted Gaussian VAR with a conjugate prior is standard and available in closed form. Multiplying the Bayes factor given in equation (10) by that marginal likelihood yields the marginal likelihood of the restricted VAR ω^R.

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.

Proposition 2 is of independent interest. 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.[23] 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 $y_i \subseteq y_1$.*[24]

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 Ω satisfies condition (9) for appropriate α and β. Furthermore, we assume that all models

[23] This is true in the realistic case when there are variables in y that belong neither to y_i nor to y_j.

[24] In our example in Section 2.1, Ω consists of the following four models:
$\begin{pmatrix} B_{xx} & B_{xw} & B_{xz} \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$, $\begin{pmatrix} B_{xx} & 0 & 0 \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$, $\begin{pmatrix} B_{xx} & B_{xw} & 0 \\ B_{wx} & B_{ww} & 0 \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$, $\begin{pmatrix} B_{xx} & 0 & B_{xz} \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & 0 & B_{zz} \end{pmatrix}$.

in Ω have equal prior probabilities.[25]

Definition 4 *Let Ω^j be the subset of Ω containing all models in which y_i is Granger-causally-prior to y_j.*[26]

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. Second, Ω is defined so as to treat the null hypothesis and the alternative hypothesis *as symmetrically as possible*.[27] In particular, if y_j consists of a single variable (as in the application in the next section), Ω^j and its complement $\Omega \setminus \Omega^j$ have *the same number of elements*. To see this, note that

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

[26] In our example in Section 2.1, Ω^j consists of the following two models:
$\begin{pmatrix} B_{xx} & 0 & 0 \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}, \begin{pmatrix} B_{xx} & B_{xw} & 0 \\ B_{wx} & B_{ww} & 0 \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$.

[27] 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.

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 with the same number of elements.

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.[28] 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 property of posterior probability.[29] 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 for the choice of variables. 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.[30] Furthermore, we think that a simple and prudent approach 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.[31]

Finally, we comment on computation. In principle, one can evaluate $p(\Omega^j|Y,\Omega)$ *exactly*

[28]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 the same number of elements. 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.

[29]See Fernández-Villaverde and Rubio-Ramírez (2004), Section 2.3.

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

[31]We make this argument at the end of Section 2.2 in the case of reduced-form VARs and at the end of Section 2.3 in the case of structural VARs.

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 the application in the next section N_J is a large number: $N_J = 38$. Therefore, 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 six hours on a standard personal computer.

5 Application

This section illustrates how one can apply the paper's methodological results to data. We define y, y_i, and y_J in the data (Section 5.1). We state the prior (Section 5.2). In Section 5.3, we evaluate the posterior probability that y_i is Granger-causally-prior to an individual variable y_j, for each individual variable $y_j \in y_J$ in the data, using the methodology from Section 4.

From Section 4 we know what will happen if we use this methodology in a large sample: The posterior probability that y_i is Granger-causally-prior to y_j will be equal to either one or zero *for each* y_j. We can then use the posterior probabilities and the properties of Granger-causal-priority stated in Section 2 to make a decision about the choice of variables. What are we to do if not all posterior probabilities come out equal to either one or zero? One possibility is to make a decision informally: one can rank y_j's according to the posterior probability that y_i is Granger-causally-prior to y_j; furthermore, one can stipulate that a y_j can be omitted from the VAR with y_i if this y_j is sufficiently poorly placed in the ranking. Another possibility is to make a decision formally following Bayesian decision theory. This approach requires specifying a loss function, i.e., a function that assigns a numerical loss to every combination of actual choice of variables and the correct choice of variables, and minimizing the posterior expected loss.[32] We pursue this route in Section 5.4. We specify a loss function and choose a set of variables by minimizing the posterior expected loss.

[32]By the "correct choice of variables" 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).

Finally, we report root mean squared errors of point forecasts (Section 5.5) and in Section 5.6 we formulate a general lesson 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 3 variables of interest (i.e., in each dataset y_i has 3 elements) and 38 other variables (i.e., in each dataset y_J has 38 elements). Table 1 lists all variables.

The variables of interest (i.e., the elements of y_i) are: real GDP, the price level and the short-term interest rate. We motivated this choice in Section 1. In the euro area exercise, we use the harmonized index of consumer prices (HICP) and the overnight interbank interest rate Eonia. In the U.S. exercise, we use 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. CPI, the federal funds rate and 35 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 35 U.S. variables listed in the next paragraph.

The 35 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, corporate bond spread defined as the 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, 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 (capacity utilization, purchasing managers' index); (xi) survey-based indicators of economic sentiment (industrial confidence, consumer confidence).[33]

[33]Due to data availability, we include hours worked only for the United States and we include M3 only

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	
Purchasing managers' index	index	SA	
Industrial confidence	index	SA	
Consumer confidence	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.

The main sample contains quarterly data from 1999Q1 to 2013Q3. 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 we explain in the next subsection.

5.2 Prior

The prior in the unrestricted model ($\omega^U \in \Omega$) consists of two pieces: (i) an initial prior formulated before seeing 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 below. Given this prior in the unrestricted model, the prior in each restricted model (in each $\omega^R \in \Omega$) is defined by equation (9). The initial prior is the prior of Sims and Zha (1998), the standard prior used in Bayesian VARs, which we implement via dummy observations Y_{SZ} and X_{SZ}. The training sample prior contains 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.[34] The prior of Sims and Zha (1998) is governed by ν_{SZ} and five other hyperparameters. In what we refer to as the baseline specification, we use the hyperparameter values employed by Sims and Zha (1998) and standard in the literature, except that we set $\nu_{SZ} = N + 20$. See Appendix C for the details concerning the prior of Sims and Zha (1998) including the hyperparameter values. In the baseline specification we use four lags ($P = 4$), which is typical of a quarterly VAR in the literature. We also report how our findings change as we move away from the baseline by varying the prior and the

for the euro area. The variables "oil price" and "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.

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

number of lags.

5.3 Evaluating the probability of Granger-causal-priority

Table 2 reports the posterior probability that y_i (real GDP, the price level and the short-term interest rate) is Granger-causally-prior to a variable y_j, for each variable $y_j \in y_J$, in the euro area (left column) and in the United States (right column). We summarize the findings in the following two points. (1) Each y_j falls into one of three groups: The probability of Granger-causal-priority from y_i to y_j is indistinguishable from zero, indistinguishable from one, or in an intermediate range. The first group has roughly the same size in the euro area and in the United States, about 20 variables. The second group is larger in the euro area than in the United States. (2) The y_j's associated with the lowest probabilities, i.e., the most useful variables for modeling y_i, are: (i) industrial confidence, a survey-based indicator of economic sentiment; (ii) survey-based indicators of economic activity (purchasing managers' index in the euro area, capacity utilization and purchasing managers' index in the United States); (iii) interest rates on private debt and on government debt (the lending rate to non-financial corporations, the mortgage interest rate, and the yield on 2-year government bonds); (iv) selected components of real GDP (the change in inventories, investment, exports, and imports); (v) the unemployment rate; and (vi) the price of oil. This list of most useful variables applies to the euro area and to the United States. In general, the posterior probabilities are similar between the euro area and the United States. The rank correlation between the posterior probabilities in the euro area and in United States is 0.6.

How do the findings change as we vary the prior? Inspired by Giannone et al. (2012), we carry out an intuitive exercise in which we loosen or tighten the prior relative to the baseline specification by changing a single scalar. The prior density of B conditional on Σ implied by the prior of Sims and Zha (1998) in the baseline can be written as $\mathcal{N}(\text{vec}\, B_{SZ}, \Sigma \otimes Q_{SZ})$, where $B_{SZ} = (X'_{SZ}X_{SZ})^{-1}X'_{SZ}Y_{SZ}$ and $Q_{SZ} = (X'_{SZ}X_{SZ})^{-1}$. Replace the term $\Sigma \otimes Q_{SZ}$ with $\xi\Sigma \otimes Q_{SZ}$, where ξ is a positive scalar and $\xi = 1$ is the baseline. Setting $\xi > 1$ implies loosening the prior and setting $\xi < 1$ implies tightening the prior relative to the baseline. Table 3 (euro area) and Table 4 (United States) report the results for $\xi = 4$ ("loosest"), $\xi = 2$ ("looser"), $\xi = 1/2$ ("tighter"), and $\xi = 1/4$ ("tightest") in comparison with the

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

Euro area			United States	
Variable	Prob.	Rank	Variable	Prob.
Industrial confidence	0.00	1	**Lending rate to NFCs**	0.00
Mortgage interest rate	0.00	2	**Capacity utilization**	0.00
Real exports	0.00	3	**Unemployment rate**	0.00
Oil price	0.00	4	**Change in real inventories**	0.00
Change in real inventories	0.00	5	**2-year government bond yield**	0.00
Real investment	0.00	6	**Real investment**	0.00
2-year government bond yield	0.00	7	**Oil price**	0.00
Real imports	0.00	8	**10-year government bond yield**	0.00
Lending rate to NFCs	0.00	9	**Industrial confidence**	0.00
Employment	0.00	10	**Mortgage interest rate**	0.00
Real consumption	0.00	11	**Purchasing managers' index**	0.00
Unemployment rate	0.00	12	**Real imports**	0.00
Purchasing managers' index	0.00	13	**Euro area consumer prices**	0.00
Producers price index	0.00	14	**Producers price index**	0.00
Loans to NFCs	0.00	15	**Real exports**	0.01
Loans for house purchase	0.01	16	**Stock index**	0.01
Consumer loans	0.02	17	**Real consumption**	0.01
Unit labor cost	0.06	18	**Corporate bond spread**	0.01
Real housing investment	0.09	19	**Employment**	0.02
Capacity utilization	0.11	20	**M1**	0.02
M3	0.24	21	**Hours worked**	0.03
M2	0.24	22	**Consumer prices excl. energy, food**	0.07
U.S. consumer prices	0.27	23	**Nominal effective exchange rate**	0.15
10-year government bond yield	0.33	24	**Consumer confidence**	0.15
M1	0.42	25	**Commodity prices**	0.25
Federal funds rate	0.49	26	**Real government consumption**	0.36
Real government consumption	0.74	27	**Stock volatility index**	0.39
Commodity prices	0.75	28	Unit labor cost	0.46
House prices	0.95	29	**Loans to NFCs**	0.48
U.S. real GDP	0.96	30	Euro area real GDP	0.53
Consumer prices excl. energy, food	0.96	31	Real housing investment	0.57
Consumer confidence	0.99	32	**Loans for house purchase**	0.59
Stock index	1.00	33	Government debt	0.64
Corporate bond spread	1.00	34	House prices	0.91
Nominal effective exchange rate	1.00	35	Consumer loans	1.00
Government debt	1.00	36	Dollar-euro exchange rate	1.00
Dollar-euro exchange rate	1.00	37	Eonia	1.00
Stock volatility index	1.00	38	M2	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.

baseline ($\xi = 1$).[35] There are two takeaways. First, the looser the prior, the higher the posterior probability of Granger-causal-priority, typically. This is what we expected to find. If the prior is loose, the fit of a VAR is likely to improve a lot after we add a zero restriction. If the prior is tight, the fit of a VAR is likely to improve little after we add a zero restriction. Second, the rank correlation between the posterior probabilities with the baseline prior and with any other prior we consider is high.[36] In other words, the ranking of the y_j's in terms their usefulness for modeling y_i changes little as we vary the prior. Consequently, the answer to the question "Is a variable x more useful than a variable z?" is typically independent of the prior. This finding is an indicator of robustness.

We also studied how the findings change as we vary the number of lags and the sample. The rank correlation between the posterior probabilities in the baseline with four lags ($P = 4$) and with two or three lags is very high (0.97 in the euro area and 0.96 in the United States for $P = 3$, 0.93 in the euro area and 0.92 in the United States for $P = 2$). The fit is about as good with $P = 3$ and $P = 2$ as with $P = 4$.[37] The fit deteriorates and the rank correlations fall somewhat with one lag ($P = 1$). To investigate the sensitivity to the sample we redid the baseline analysis having omitted the last four quarters, the last eight quarters, and the last twelve quarters from the sample. The rank correlation between the posterior probabilities in the baseline sample and in each of the shorter samples is very high.[38] We conclude that the findings from the baseline specification are robust.

5.4 Choosing the best variables

Though some posterior probabilities reported in Table 2 are indistinguishable from one or zero, some lie in an intermediate range. In this subsection, we follow what Bayesian decision theory prescribes for this situation: We specify a loss function and choose a set of variables

[35] See the end of Appendix C for more information concerning the setup of this exercise.

[36] Tables 3-4 report how the marginal likelihood implied by the set of models Ω, $\log p(Y|\Omega)$, and the marginal likelihood implied by the unrestricted VAR, $\log p(Y|\omega^U)$, depend on the prior. The best fit, by far, is achieved in the baseline ($\xi = 1$) and with the tighter prior ($\xi = 1/2$). The rank correlation between the posterior probabilities in the baseline and with the tighter prior is very high, 0.95 both in the euro area and in the United States.

[37] The marginal likelihood implied by the set of models Ω, $\log p(Y|\Omega)$, and the marginal likelihood implied by the unrestricted VAR, $\log p(Y|\omega^U)$, change only moderately whether we assume $P = 2$, $P = 3$ or $P = 4$.

[38] In the euro area, the rank correlation between the posterior probabilities in the baseline sample and in the shorter samples lies between 0.98 (without the last four quarters) and 0.97 (without the last twelve quarters). In the United States, the rank correlation between the posterior probabilities in the baseline sample and in the shorter samples lies between 0.99 (without the last four quarters) and 0.92 (without the last twelve quarters).

Table 3: Posterior probability that real GDP, price level and short-term interest rate are Granger-causally-prior to a variable, sensitivity to the prior, euro area

Variable	loosest $\xi = 4$	looser $\xi = 2$	baseline $\xi = 1$	tighter $\xi = 1/2$	tightest $\xi = 1/4$
Industrial confidence	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Mortgage interest rate	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Real exports	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Oil price	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Change in real inventories	0.99	**0.03**	**0.00**	**0.00**	**0.00**
Real investment	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
2-year government bond yield	**0.00**	**0.00**	**0.00**	**0.02**	**0.35**
Real imports	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Lending rate to NFCs	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Employment	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**
Real consumption	1.00	**0.31**	**0.00**	**0.00**	**0.00**
Unemployment rate	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Purchasing managers' index	**0.01**	**0.00**	**0.00**	**0.01**	**0.01**
Producers price index	**0.00**	**0.00**	**0.00**	**0.03**	**0.04**
Loans to NFCs	**0.00**	**0.00**	**0.00**	**0.01**	**0.01**
Loans for house purchase	**0.00**	**0.01**	**0.01**	**0.01**	**0.01**
Consumer loans	**0.46**	**0.18**	**0.02**	**0.01**	**0.02**
Unit labor cost	**0.00**	**0.09**	**0.06**	**0.05**	**0.03**
Real housing investment	0.54	**0.15**	**0.09**	**0.02**	**0.02**
Capacity utilization	**0.00**	**0.01**	**0.11**	0.67	0.52
M3	**0.00**	**0.38**	**0.24**	**0.13**	**0.03**
M2	**0.00**	**0.38**	**0.24**	**0.13**	**0.03**
U.S. consumer prices	**0.02**	**0.27**	**0.27**	0.45	0.34
10-year government bond yield	1.00	0.91	**0.33**	**0.17**	**0.33**
M1	**0.00**	**0.10**	**0.42**	0.69	0.69
Federal funds rate	0.95	0.44	0.49	**0.04**	**0.08**
Real government consumption	1.00	0.95	0.74	**0.22**	**0.05**
Commodity prices	0.91	0.74	0.75	0.83	0.54
House prices	**0.00**	**0.01**	0.95	1.00	0.98
U.S. real GDP	1.00	0.90	0.96	0.91	0.62
Consumer prices excl. energy, food	1.00	1.00	0.96	**0.55**	**0.03**
Consumer confidence	1.00	1.00	0.99	**0.46**	**0.00**
Stock index	1.00	1.00	1.00	0.99	0.89
Corporate bond spread	1.00	1.00	1.00	0.98	0.57
Nominal effective exchange rate	1.00	1.00	1.00	1.00	0.94
Government debt	1.00	1.00	1.00	1.00	0.99
Dollar-euro exchange rate	1.00	1.00	1.00	1.00	0.96
Stock volatility index	1.00	1.00	1.00	1.00	0.99
rank correlation with the baseline	0.77	0.91	1.00	0.95	0.80
no. of variables in y_1 in the best model	22	25	25	26	26
$\log p(Y\|\Omega)$	4335	4575	4664	4624	4530
$\log p(Y\|\omega^U)$	4006	4384	4561	4583	4525
$\log p(Y\|\omega^*)$	4361	4599	4688	4647	4552
$\log p(Y\|\omega^*) - \log p(Y\|\omega^U)$	354	215	127	63	27

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 4: Posterior probability that real GDP, price level and short-term interest rate are Granger-causally-prior to a variable, sensitivity to the prior, United States

Variable	loosest $\xi = 4$	looser $\xi = 2$	baseline $\xi = 1$	tighter $\xi = 1/2$	tightest $\xi = 1/4$
Lending rate to NFCs	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Capacity utilization	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Unemployment rate	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Change in real inventories	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
2-year government bond yield	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Real investment	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Oil price	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**
10-year government bond yield	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Industrial confidence	**0.21**	**0.03**	**0.00**	**0.00**	**0.00**
Mortgage interest rate	**0.00**	**0.00**	**0.00**	**0.00**	**0.01**
Purchasing managers' index	**0.01**	**0.02**	**0.00**	**0.01**	**0.01**
Real imports	**0.00**	**0.00**	**0.00**	**0.02**	**0.01**
Euro area consumer prices	1.00	0.64	**0.00**	**0.00**	**0.01**
Producers price index	**0.00**	**0.00**	**0.00**	**0.01**	**0.02**
Real exports	**0.00**	**0.00**	**0.01**	**0.01**	**0.01**
Stock index	1.00	**0.29**	**0.01**	**0.00**	**0.01**
Real consumption	**0.21**	**0.06**	**0.01**	**0.00**	**0.00**
Corporate bond spread	0.88	**0.17**	**0.01**	**0.00**	**0.00**
Employment	**0.02**	**0.05**	**0.02**	**0.03**	**0.01**
M1	1.00	0.98	**0.02**	**0.00**	**0.00**
Hours worked	**0.03**	**0.10**	**0.03**	**0.01**	**0.01**
Consumer prices excl. energy, food	1.00	0.99	**0.07**	**0.04**	**0.07**
Nominal effective exchange rate	**0.01**	**0.26**	**0.15**	**0.03**	**0.01**
Consumer confidence	1.00	**0.48**	**0.15**	**0.08**	**0.02**
Commodity prices	**0.02**	0.54	**0.25**	**0.05**	**0.02**
Real government consumption	1.00	0.99	**0.36**	**0.07**	**0.03**
Stock volatility index	1.00	0.79	**0.39**	**0.16**	**0.05**
Unit labor cost	1.00	0.77	0.46	**0.39**	**0.17**
Loans to NFCs	**0.00**	**0.31**	**0.48**	**0.16**	**0.02**
Euro area real GDP	0.93	0.70	0.53	**0.13**	**0.02**
Real housing investment	**0.02**	0.24	0.57	0.40	**0.02**
Loans for house purchase	**0.11**	**0.56**	**0.59**	**0.20**	**0.03**
Government debt	**0.01**	**0.18**	0.64	0.63	**0.11**
House prices	**0.01**	0.35	0.91	0.69	**0.19**
Consumer loans	1.00	1.00	1.00	0.97	**0.23**
Dollar-euro exchange rate	1.00	1.00	1.00	0.55	**0.00**
Eonia	0.99	0.98	1.00	1.00	1.00
M2	1.00	1.00	1.00	1.00	1.00
rank correlation with the baseline	0.68	0.84	1.00	0.95	0.72
no. of variables in y_1 in the best model	24	24	29	31	36
$\log p(Y\|\Omega)$	3239	3460	3536	3507	3423
$\log p(Y\|\omega^U)$	2967	3327	3484	3501	3440
$\log p(Y\|\omega^*)$	3265	3483	3558	3532	3448
$\log p(Y\|\omega^*) - \log p(Y\|\omega^U)$	298	156	75	31	8

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.

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 25 variables ranked at the top of the left column of Table 2 and listed there in bold font, from industrial confidence to M1. The best model in the United States has in y_1 the 27 variables ranked at the top of the right column of Table 2 and listed there in bold font (from the lending rate to non-financial corporations to the stock volatility index) plus two other variables ranked lower and also listed in bold font (loans to non-financial corporations and loans for house purchase).

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 127 in the euro area and 75 in the United States. See the last row in the column "baseline, $\xi = 1$" in Table 3 and Table 4, respectively. A log Bayes factor of 127 corresponds to posterior odds of about 10^{55} to 1. A log Bayes factor of 75 corresponds to posterior odds of about 10^{32} to 1. Thus, there is overwhelming support for omitting some variables from the VAR (and we determine which variables to omit).

The vector y_1 in the best model, with its 28 elements in the euro area and 32 elements in the United States, is longer than usual in the VAR literature. This result is consistent with the following lesson from the literature on linear time series models (VARs, factor models, and factor-augmented VARs) applied to large datasets: Small VARs are likely to leave out valuable information. At the same time, let us emphasize the finding that there is overwhelming support for omitting a non-trivial number of variables from the VAR.

Omitting a subset of the variables *selected wisely* improves fit, whereas omitting *any* subset of the variables can deteriorate fit. Consider 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.[39] The log Bayes factor in favor of this model is -141 in the euro area and -218 in the United States, indicating overwhelming evidence in favor of the unrestricted model.

How do the findings change as we vary the prior? The following conclusions emerge from Tables 3-4. First, the looser the prior, the fewer variables enter y_1 in the best model. This finding is expected and consistent with the intuition given in the previous subsection: The VAR "likes" zero restrictions more if the prior is loose than if the prior is tight. Furthermore, this finding suggests that a researcher can formalize his or her preference for model size by controlling the tightness of the standard prior of Sims and Zha (1998). Second, the best model fits better than the unrestricted model for each prior we consider. Note that all entries are positive in the row in Tables 3-4 reporting the marginal likelihood implied by model ω^* relative to the marginal likelihood of the unrestricted model, $\log p(Y|\omega^*) - \log p(Y|\omega^U)$. Thus, it is a good idea to omit some variables from the VAR no matter which prior we consider. Third, the best fitting model of all models we consider is the best model in the baseline specification (i.e., model ω^* for $\xi = 1$) both in the euro area and in the United States. See the row in Tables 3-4 reporting the marginal likelihood implied by model ω^*, $\log p(Y|\omega^*)$, and notice that $\log p(Y|\omega^*)$ peaks for $\xi = 1$ in both tables. As noted before, this finding suggests that we omit 13 variables in the euro area and 9 variables in the United States. The best model with the tightest prior ($\xi = 1/4$) in the United States recommends omitting only 2 variables, but this model fits much worse than the best model in the baseline. Fourth, the looser the prior, the greater the advantage of the best model vis-à-vis the unrestricted model. In Tables 3-4, note how the difference $\log p(Y|\omega^*) - \log p(Y|\omega^U)$ varies monotonically with the tightness of the prior. Again, this finding is expected and consistent with the intuition that Granger-noncausality restrictions become less valuable as the prior tightens.

We studied how the findings change as we vary the number of lags and the sample. Specifications with fewer lags than in the baseline favor omitting more variables. For example, the specification with two lags ($P = 2$) suggests that we omit 17 variables in the

[39]Specifically, we choose the model in which y_2 consists of "the best variables," y_1 consists of all other variables including y_i, and y_2 does not Granger-cause y_1.

euro area and 16 variables in the United States, while fitting only somewhat worse than the baseline ($P = 4$). This tells us that some variables become useful for modeling y_i only after a sufficient number of their lags are included. Examples are monetary aggregates (M1, M2 and M3) in the euro area and credit aggregates (loans for house purchase in the euro area and in the United States, and loans to non-financial corporations in the United States). Furthermore, the findings in the shorter samples, defined as in the previous subsection, are very similar to the baseline, though the best model in the shortest sample (without the last twelve quarters of the data) recommends that we omit a few more variables than in the baseline, both in the euro area and in the United States. It makes sense that a smaller model is preferable in a shorter sample.[40]

5.5 Root mean squared errors

The paper's methodology is based on marginal likelihood. Marginal likelihood is an out-of-sample predictive density, as we discuss in the next section. Furthermore, knowing the value of the marginal likelihood is necessary and sufficient for making probability statements about models given data. Knowing errors of out-of-sample point forecasts is of no 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 of which includes y_i. The VAR "top 10" in addition to y_i includes the ten variables ranked at the top of Table 2, i.e., the ten variables associated with *lowest* probabilities of Granger-causal-priority. This is a model likely to be chosen by a researcher with a preference for a VAR with about 10-12 variables. We expect the VAR "top 10" to yield good point forecasts. For comparison, we use the VAR "all" that includes all variables we have data on, y, and the VAR "bottom 10" that 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. For each quarter from 1999Q1 to 2011Q3 we fitted the three VARs to the data up

[40]Instead of looking only at the single best model in a given specification, we also 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 the single best model ω^*. In the baseline, the set of best models contains 11 models in the euro area and 15 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 euro area the 20 variables ranked at the top of the left column of Table 2 (i.e., from industrial confidence to capacity utilization) enter y_1 in each model in the set of best models.

to that quarter, and we generated out-of-sample point forecasts one quarter ahead and one year ahead. We used four lags 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 5 reports the root mean squared errors (RMSEs) of the VAR point forecasts 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. We find that the VAR "top 10" yields good point forecasts: First, the VAR "top 10" always does better than the VAR "bottom 10." Second, the VAR "top 10" does better than the VAR "all" in 11 out of the 12 forecast comparisons in Table 5. In the twelfth forecast comparison (real GDP one-quarter-ahead) the VAR "top 10" does about as well as the VAR "all." Third, the VAR "top 10" outperforms the random walk with drift in the cases of real GDP and the short-term interest rate. All the VARs we consider do worse than the random walk with drift in the case of the price level.[41]

Table 5: 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						
Real GDP	0.81	0.79	**0.72**	0.92	**0.85**	0.85
Price level	1.06	**1.05**	1.06	1.09	**1.04**	1.15
Short-term interest rate	0.74	**0.67**	0.68	0.83	**0.69**	0.76
One-year-ahead						
Real GDP	1.34	**0.97**	1.03	1.19	**0.88**	0.98
Price level	1.48	**1.33**	1.62	1.57	**1.32**	1.98
Short-term interest rate	1.36	**0.99**	1.30	1.00	**0.89**	1.16

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

5.6 Summing up

Let us conclude this section with a general lesson. Evaluating the posterior probability of Granger-causal-priority yields a ranking of the variables in one's dataset. Furthermore, we

[41] This result is consistent with the consensus in the literature that, in the words of Stock and Watson (2009), p.101, "inflation is hard to forecast," in particular in the last 25 years when inflation varied little.

find that a choice-of-variables procedure based on a particular loss function selects variables from the top of the ranking. This finding suggests that in practice choice of variables can occur by examining the posterior probabilities of Granger-causal-priority without specifying a loss function explicitly. A researcher with a preference for a small VAR will select only a few variables from the top of the ranking, whereas a researcher who prefers a medium-sized VAR will choose a larger subset of the ranking.

6 Marginal likelihood versus other objects

Throughout the paper, we followed the Bayesian approach to inference and therefore we relied on marginal likelihood to compare models. In this section, we contrast marginal likelihood with three other objects. We think that this will give readers additional insight into what we did in the previous sections.

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:

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

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

We make four observations: (1) Equation (13) shows that the 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. (2) 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, the marginal likelihood measures the *overall* out-of-sample predictive performance of a model, at any horizon. We emphasize points (1) and (2) because some economists seem to believe that "computing the marginal likelihood" and "evaluating the out-of-sample predictive performance" are different activities. (3) Knowing the value of the marginal likelihood is necessary and sufficient for making probability statements about models given data. In other words, knowing the value of the marginal likelihood leads one to "update one's prior" about models via Bayes' theorem. (4) 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 $y = (y_i, y_J)$. Model comparison via marginal likelihood requires that the models under consideration make predictions about the same variables. Furthermore, marginal likelihood keeps track of all predictions of each model, i.e., the predictions about all the variables.

An object with some popularity in the literature is the predictive density score.[42] 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 choosing variables: (1) The predictive density score is a measure of the out-of-sample predictive performance of a model, like the marginal likelihood. (2) The predictive density score is specific to a given forecast horizon, h, *unlike* the marginal likelihood. For instance, expression (14) assumes that the forecast horizon is one period, $h = 1$. One obtains a different expression and 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.[43] (3) Knowing the value of the predictive density score of y_i does *not* justify making probability statements given data. The practical implication for us 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. (4) 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 may seem attractive when y_i is the set of variables of interest. However, this feature must be weighed against features (2) and (3), both of which we see as important.

In addition, one must take into account that it is costly – in some realistic applications prohibitively costly – to compare models based on their predictive density scores. Com-

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

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

puting the predictive density score requires a loop in which for each period the researcher reestimates the model using the data up to that period, forms 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. This difference can be the difference between computational non-feasibility and feasibility if many models are to be compared.

Two other objects may seem to be possible criteria for choosing variables. In the rest of this section, we argue that each of the two objects has a serious flaw.

Consider the following predictive density:

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

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

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 the variables so that each variable in $y_{J\psi}$ and each variable in $y_{J\tilde{\psi}}$ have the same value in the initial period, i.e., $y_{J\psi}(0) = y_{J\tilde{\psi}}(0)$. 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![44]

Next, consider the following predictive density:

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

[44]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.

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

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 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, 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. 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. The analytical results are of independent interest.

We think of the application as an illustration, certainly not the final word on which macroeconomic and financial variables interact most closely with real GDP, 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 short as well as the fact that a financial crisis occurs in the sample. In future, it will be useful to redo this paper's analysis with models other than a VAR with constant variances of the residuals, which may fit better in this sample. In some models, such as a VAR with stochastic volatility and a VAR with Markov-switching variances of the residuals, 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 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 $y_q \cup y_r = y_1$, $y_i \subseteq y_1$, and $y_j \subseteq y_2$. 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, $\varepsilon_q \cup \varepsilon_r = \varepsilon_1$, and $\varepsilon_k \subseteq \varepsilon_q$. 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 condition (i). The restrictions $C_{qr}(0) = 0$ and $C_{q2}(0) = 0$ follow from condition (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 condition (i). Furthermore, condition (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}. Remember that we assume that if a variable is necessary for identification, this variable is included in y_i and thus in y_1. Then the first equation implies that $C_{qq}(0)$ is a function only of Σ_{11}. Furthermore, 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_{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

$$\begin{aligned} C_{qq}(L) &= D_{qq}(L) C_{qq}(0) + D_{qr}(L) C_{rq}(0), \\ C_{rq}(L) &= D_{rq}(L) C_{qq}(0) + D_{rr}(L) C_{rq}(0), \end{aligned}$$

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 response of y_1 to ε_q, $C_{1q}(L)$, is a function only of $B_{11}(L)$ and Σ_{11}. By implication, the impulse response of y_i to ε_k, $C_{ik}(L)$, 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 is interested in 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 $p(\Sigma|\omega^U)$.

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}.[45]

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 the standard deviation of the residuals from the univariate autoregression of order one fit by ordinary least squares to the n'th time series. 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

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

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 $\operatorname{diag}(x)$ 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

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

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 $Z = \sqrt{\nu_{SZ} - N - 1}\,\mathrm{diag}(\sigma)$, 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

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

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

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

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

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.,

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

Hyperparameter values in the baseline. We set: $\lambda_1 = 0.2$, $\lambda_3 = 1$, $\lambda_4 = 1$, $\mu_5 = 1$, $\mu_6 = 1$. These are the values employed by Sims and Zha (1998) and standard in the literature. We set $\nu_{SZ} = N + 20$, about halfway between $\nu_{SZ} = N + 1$ used by Sims and Zha (1998) and $\nu_{SZ} = 2N$ used in the Dynare software package.[46] The reason for this

[46] In Dynare the fact that $\nu_{SZ} = 2N$ is implied by the default Dynare setting $\omega = 1$. See Villemot (2012),

choice of ν_{SZ} is model fit. The marginal likelihood of the unrestricted model ω^U and the marginal likelihood of the best model ω^* are more than one hundred log points higher if $\nu_{SZ} = N + 20$ than if $\nu_{SZ} = N + 1$. See Table 6. In other words, setting $\nu_{SZ} = N + 1$ yields a horrible fit. Furthermore, the marginal likelihood of the unrestricted model ω^U and the marginal likelihood of the best model ω^* are much higher if $\nu_{SZ} = N + 20$ than if $\nu_{SZ} = 2N$. Again, see Table 6. Note that varying ν_{SZ} does not have a straightforward interpretation in terms of "tightening" or "loosening" the prior, since ν_{SZ} affects both the left and the right tails of the prior density of the diagonal elements of Σ.

The choice of ν_{SZ} also affects the odds ratios in favor of restrictions. With $\nu_{SZ} = N+1$ the unrestricted model ω^U is the best model, hence the zeros in the last row of Table 6. Recall, however, that the fit of this model is very bad compared with either ω^U or ω^* in the baseline. With $\nu_{SZ} = 2N$ the support for restrictions is stronger than in the baseline. Again, see the last row of Table 6. Consistent with this finding, the best model with $\nu_{SZ} = 2N$ has fewer variables in y_1 than the best model in the baseline, both in the euro area and in the United States. The ranking of the variables when $\nu_{SZ} = 2N$ is very similar to the baseline.

Table 6: Marginal likelihood given alternative values of ν_{SZ}

	Euro area			United States		
ν_{SZ}	$N+1$	$\boldsymbol{N+20}$	$2N$	$N+1$	$\boldsymbol{N+20}$	$2N$
$\log p(Y\|\omega^U)$	4382	**4561**	4412	3174	**3484**	3361
$\log p(Y\|\omega^*)$	4382	**4688**	4590	3174	**3558**	3495
$\log p(Y\|\omega^*) - \log p(Y\|\omega^U)$	0	**127**	178	0	**75**	134

Tighter and looser priors in the sensitivity analysis. In the main text, we report the findings with the baseline prior and four alternative prior settings. We construct the alternative prior settings by varying the scalar ξ as explained in Section 5.3. Note that introducing $\xi \neq 1$ is equivalent to replacing λ_1, λ_4, μ_5, μ_6 with $\lambda_1\xi$, $\lambda_4\xi$, μ_5/ξ, μ_6/ξ, while the value of λ_3 is unaffected by changing ξ.

pp.3-4, noting that ν_{SZ} is denoted there as df^p.

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 model ω. The process continues until the chain has the desired length. In each exercise (i.e., for each column in Tables 2-4), we ran two chains of one million draws each, starting from random 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 in a statistically significant way. A chain of one million draws runs about one hour on a standard personal computer when the VAR has 1 lag and about six hours when the VAR has 4 lags.

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. We draw the initial point for each chain randomly from Ω, assigning equal probability to each model in Ω.

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 100,000 states of the chain to ensure that the results do not depend on the initialization. That is, we estimated the GCP probability as $\hat{p}(\Omega^j|Y,\Omega) = \left(\sum_{m=100,001}^{1000,000} \theta_m^j\right)/900,000$, 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. The Newey-West standard errors were typically below 0.005 and only in rare cases reached about 0.01.

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 was 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 is the best model in Ω. The ranking of several hundred 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 actual numbers are 809 models in the euro area exercise and 1700 models in the U.S. exercise, in the baseline specification.

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

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.

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., Litterman, R., and Sims, C. (1984). Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews*, 3(1):1–100.

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-Ramírez, 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.

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.

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

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

Rubio-Ramírez, 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.

Stock, J. H. and Watson, M. W. (2009). Phillips curve inflation forecasts. In Fuhrer, J., Little, J. S., Kodrzycki, Y. K., and Olivei, G. P., editors, *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, pages 101–186. MIT Press.

Villemot, S. (2012). BVAR models á la Sims in Dynare. Technical report, Dynare documentation, CEPREMAP.

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.