EUROPEAN SUMMER SYMPOSIUM IN INTERNATIONAL MACROECONOMICS (ESSIM) 2013

Hosted by the
Central Bank of the Republic of Turkey (CBRT)

Izmir, Turkey; 21-24 May 2013

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

Marek Jarocinski and Bartosz Mackowiak

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: December 2012

Abstract

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

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, Marco Del Negro, John Geweke, Domenico Giannone, David Madigan, Frank Schorfheide, Chris Sims, Ellis Tallman, Tao Zha and seminar and conference participants at ECB, ESEM 2012, Institute for Advanced Studies Vienna, ETH Zurich, National Bank of Poland, Oesterreichische Nationalbank, Humboldt University Berlin and Université Libre de Bruxelles. We thank Giovanni Nicolò for excellent research assistance. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the European Central Bank.

1 Introduction

Vector autoregression is perhaps the most standard tool of empirical macroeconomics. When we set out to estimate a VAR, we rarely know a priori which variables to include in the VAR. Often, we have a small number of variables we are interested in and we know that, in principle, we can include many other variables in the VAR.

This paper studies formally the choice of variables in VARs. Consider a researcher who is interested in a set of variables y_i. The researcher wants to forecast y_i with a VAR or to infer impulse responses of a subset of y_i to structural shocks with a structural VAR.[1] The researcher has data on a larger set of variables y that includes y_i, i.e., $y_i \subset y$, and also includes other variables $y_J \equiv y \setminus y_i$. Let $y_j \subseteq y_J$ be a subset of the other variables. The questions that we study in this paper are: (i) Does y_j belong in the VAR to be used to forecast y_i? (ii) Does y_j belong in the structural VAR to be used to infer impulse responses of y_i to structural shocks?

We show that Granger-causal-priority is the relevant econometric hypothesis to study in order to answer both questions. Granger-causal-priority is a concept derived from the well-known idea of Granger-noncausality. We demonstrate that an appropriate Granger-causal-priority restriction implies that the forecasts of y_i obtained from a VAR with all variables y are *identical* to the forecasts of y_i obtained from a smaller VAR that omits y_j. In this sense, if Granger-causal-priority holds, y_j does not belong in the VAR to be used to forecast y_i. Furthermore, under an additional natural assumption, we show that an appropriate Granger-causal-priority restriction implies that the impulse responses of y_i obtained from a structural VAR with all variables y are *identical* to the impulse responses of y_i obtained with a smaller structural VAR that omits y_j. In this sense, if Granger-causal-priority holds, y_j does not belong in the structural VAR to be used to infer impulse responses of y_i to structural shocks.

We develop new, simple Bayesian tests of Granger-noncausality and Granger-causal-priority in a VAR. We work out an analytical expression for the posterior probability of Granger-noncausality. Furthermore, we provide a simple Monte Carlo procedure to compute the posterior probability of Granger-causal-priority. Our results rely on the assumption that the prior in each restricted VAR has to be consistent with the prior in the unrestricted VAR,

[1] In the latter case, we assume that y_i includes the variables whose impulse responses the researcher wants to infer and the variables used in identification of shocks.

in a sense that we will make precise. This assumption is simple and natural. Furthermore, the prior in the unrestricted VAR is assumed to be conjugate and proper. In an empirical application, we use the conventional prior employed in Bayesian VARs due to Sims and Zha (1998).

In an empirical application, we study which variables belong in the VAR to forecast or infer impulse responses of real GDP, the consumer price level, and the short-term interest rate, in the euro area and in the United States. These are the variables that enter the simple New-Keynesian dynamic stochastic general equilibrium (DSGE) model. We consider a quarterly dataset with the three variables of interest and thirty-eight other macroeconomic and financial variables.

We find that GDP, the consumer price level, and the short-term interest rate are least likely to be Granger-causally-prior to: (i) survey-based leading indicators, (ii) the change in inventories, (iii) the corporate bond spread, (iv) the unemployment rate, and (v) the price of oil. The posterior probabilities of Granger-causal-priority associated with these variables are smaller than or about equal to 0.1. We also find that GDP, the consumer price level, and the short-term interest rate are most likely to be Granger-causally-prior to: (i) house prices, and (ii) broad money. The posterior probabilities of Granger-causal-priority associated with these variables are greater than or about equal to 0.95. Remarkably, these findings obtain both in the euro area *and* in the United States.

We also ask what the single best set of variables is that belongs in the VAR with GDP, the consumer price level, and the short-term interest rate. The answer to this question in the euro area is: industrial confidence (a survey-based leading indicator), the change in inventories, the corporate bond spread, the unemployment rate, and the federal funds rate. The answer to this question in the United States is: industrial confidence, the change in inventories, the corporate bond spread, and the price of oil.

Let us return to the motivation for what we do. Empirical macroeconomic research often uses small or medium-size VARs, i.e., VARs with 3 to 15 variables. These variables are chosen out of many other plausible candidate variables. In principle, one could avoid choosing variables and just include all candidate variables, since there exist techniques for estimating VARs or related models using large datasets.[2] However, small or medium-size VARs remain attractive, for the following reasons. First, researchers have a preference for

[2]These techniques include large VARs (Bańbura et al. (2010)), FAVARs (Bernanke et al. (2005)) and factor models (Forni et al. (2000), Stock and Watson (2002)).

using the minimal means for getting their point across. Second, with fewer variables it may be easier to explain to an audience "where results come from." Third, VARs serve as benchmarks that are compared with or guide the development of DSGE models or sophisticated statistical models that are difficult to estimate with a large dataset. Fourth, small and medium-size VARs are preferred when practical or institutional constraints preclude the use of larger or more sophisticated models.

Our findings about which variables belong in a VAR with output, prices and short-term interest rates have implications for the development of DSGE models. This is so because DSGE models are VARs to an approximation. Our findings suggest that when researchers are interested in modeling the dynamics of output, prices and short-term interest rates, then incorporating in the DSGE model corporate bond spreads, inventories, survey-based measures of confidence, the unemployment rate and the oil price has a higher potential for improving the model than incorporating house prices and monetary aggregates.

[Literature review to be added]

Section 2 defines Granger-causal-priority and explains the relationship between Granger-causal-priority and the two questions that we study in this paper. Section 3 derives a closed-form expression for the posterior probability of Granger-noncausality in a Gaussian VAR with a conjugate prior. Section 4 shows how to evaluate the posterior probability of Granger-causal-priority. We then turn to the empirical application. In Section 5 we report posterior probabilities of Granger-causal-priority and in Section 6 we search for the best single set of variables to be included in a VAR. Section 7 discusses three alternatives to the methodology laid out in this paper. Section 8 concludes.

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

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

The main takeaways from this section are as follows. If y_i is Granger-causally-prior to y_j, the forecasts of y_i obtained from a VAR with all variables y are *identical* to the forecasts of y_i obtained from a smaller VAR that omits y_j. In this sense, if Granger-causal-priority holds, y_j does not belong in the VAR to be used to forecast y_i. Furthermore, if

y_i is Granger-causally-prior to y_j and an additional natural assumption holds, the impulse responses of y_i obtained from a structural VAR with all variables y are *identical* to the impulse responses of y_i obtained from a smaller structural VAR that omits y_j. In this sense, if Granger-causal-priority holds, y_j does not belong in the structural VAR to be used to infer impulse responses of y_i to structural shocks.

We begin by defining Granger-causal-priority. Granger-causal-priority is related to the well-known concept of Granger-noncausality.

2.1 Granger-noncausality and Granger-causal-priority

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 and assume that a dataset with $T+P$ observations of y is available.

The following definition is familiar from Granger (1969).

Definition 1 ***Granger-noncausality:*** *y_j does not Granger-cause y_i if prediction of y_i based on the set Δ of predictors is no better than prediction based on $\Delta \setminus y_j$, that is, on the set with y_j omitted.*[3]

In the VAR literature Granger's definition is essentially always applied to one-step-ahead forecasting.[4] The definition then specializes to stating that 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 $B_{ij}(L) = 0$.[5] In Section 3 we show how a Bayesian econometrician can evaluate the posterior probability of this restriction in a simple way.

[3] In a VAR, Δ consists of past values of y.

[4] The exceptions that we are aware of are Dufour and Renault (1998) and a small number of related papers. We discuss Dufour and Renault (1998) below.

[5] See for example Hamilton (1994), Chapter 11.

The next definition, which we think is unfamiliar to most economists, derives from old unpublished work by Thomas A. Doan and appears recently in unpublished work by Doan and Todd (2010) and Sims (2010).

Definition 2 ***Granger-causal-priority:*** *y_i is Granger-causally-prior to y_j in the VAR (1) 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 says that y_i is Granger-causally-prior to y_j in the VAR (1) if this VAR has the following recursive form

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

with $B_{12}(L) = 0$.

Granger-causal-priority requires a stronger restriction than Granger-noncausality. If there are other variables in y, in addition to y_i and y_j, the set of coefficients in $B_{ij}(L)$ is a *strict subset* of the set of coefficients in $B_{12}(L)$. In the special case when y consists of y_i and y_j only, the two restrictions are equivalent.

Let us emphasize the point that Granger-causal-priority requires a stronger restriction than Granger-noncausality. Consider an example. Suppose that $y = \{x, w, z\}$, $y_i = x$, $y_j = z$, and y follows a VAR[6]

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

If $B_{xz} = 0$, $x(t+1) = B_{xx}x(t) + B_{xw}w(t) + u_1(t+1)$ and therefore z does not Granger-cause x in the sense essentially always used in the literature. However, $x(t+2) = \ldots + B_{xw}B_{wz}z(t) + \ldots$, i.e. the two-step-ahead forecast of x depends on the current value of z so long as $B_{xw}B_{wz} \neq 0$. 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

[6] For simplicity, in this example we set $P = 1$ and drop the constant term.

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

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

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

Consider a researcher who wants to forecast y_i.

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 $B_{12}(L) = 0$. Since $B_{12}(L) = 0$, the forecasts of y_i at all horizons obtained with the VAR (2) are equal to the forecasts of y_i obtained with the VAR

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

Therefore, 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. If y_i is Granger-causally-prior to y_j, y_j does not belong in the VAR to be used to forecast y_i.

Let us emphasize that to justify omitting y_j we need Granger-causal-priority; Granger-noncausality does not suffice. Granger-noncausality, i.e., $B_{ij}(L) = 0$ in the VAR (1), does *not* imply that the forecasts of y_i obtained with the VAR (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 of y_i and y_j only,[7] or (ii) if we want to forecast y_i only one step ahead.

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

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

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

Consider a researcher who wants to infer impulse responses of y_i to structural shocks.

We assume that y follows a structural VAR

$$A\left(L\right)y\left(t\right)=\delta+\varepsilon\left(t\right), \tag{4}$$

where $A\left(L\right)$ is a matrix polynomial in the lag operator of order P, $A\left(0\right)$ is nonsingular, δ is a constant term, and $\varepsilon\left(t\right)$ is a Gaussian vector with mean zero and variance-covariance matrix identity conditional on $y\left(t-s\right)$ for all $s\geq 1$. As is familiar, ε contains the structural shocks that generate the data.[8]

The following proposition establishes the connection between Granger-causal-priority and the choice of variables in a structural VAR.

Proposition 1 *Consider the structural VAR (4). Suppose that: (i) y_i is Granger-causally-prior to y_j, and (ii) the number of structural shocks that affect y_1 contemporaneously is equal to the number of variables in y_1. Then the structural VAR (4) can be partitioned*

$$\begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix}\begin{pmatrix} y_1\left(t\right) \\ y_2\left(t\right) \end{pmatrix}=\begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}+\begin{pmatrix} \varepsilon_1\left(t\right) \\ \varepsilon_2\left(t\right) \end{pmatrix} \tag{5}$$

with $A_{12}(L)=0$.

Proof. **Step 1:** Since it is arbitrary how we order the structural shocks in ε, we order them so that the shocks that affect y_1 contemporaneously, ε_1, come first. **Step 2:** We show that assumption (ii) implies that $A_{12}(0)=0$. Let C denote the matrix of contemporaneous impulse responses of y to ε, i.e., $C=A(0)^{-1}$. Assumption (ii) implies that the contemporaneous impulse response of y_1 to shocks other than ε_1 is zero, $C_{12}=0$. Since $A(0)=C^{-1}$, we have that $A_{12}(0)=0$ (the inverse of a block-triangular matrix is a block-triangular matrix). **Step 3:** We show that $A_{12}(p)=0$ for $p=1,...,P$. Recall from footnote 8 that $A(p)=-A(0)B(p)$. By step 2, $A_{12}(0)=0$. By assumption (i), $B_{12}(p)=0$

[8] The following relationships hold: $B\left(p\right)=-A\left(0\right)^{-1}A\left(p\right)$ for $p=1,...,P$, $\Sigma=A\left(0\right)^{-1}\left(A\left(0\right)^{-1}\right)'$, and $\gamma=A\left(0\right)^{-1}\delta$. Note that since we assume that $A\left(0\right)$ is nonsingular and thus invertible, we assume that $\varepsilon\left(t\right)$ is a linear combination of $u\left(t\right)$.

for $p = 1, ..., P$. Hence, $A_{12}(p) = 0$ for $p = 1, ..., P$ (the product of two block-triangular matrices is a block-triangular matrix). ■

Since $A_{12}(L) = 0$, y_1 is affected only by ε_1 and the impulse responses of y_1 obtained with the structural VAR (5) are equal to the impulse responses of y_1 obtained with the structural VAR

$$A_{11}(L)y_1(t) = \delta_1 + \varepsilon_1(t). \tag{6}$$

Therefore, the researcher can omit y_j (as well as all other variables in y_2) from the structural VAR to be used to infer the impulse responses of y_i and the impulse responses of y_i do not change. If y_i is Granger-causally-prior to y_j and the number of structural shocks that affect y_1 contemporaneously is equal to the number of variables in y_1, y_j does not belong in the structural VAR to be used to infer impulse responses of y_i to structural shocks.

Let us discuss assumption (ii) in Proposition 1. This assumption is natural in the context of choosing variables in a structural VAR. This is so because in structural VARs the number of shocks that affect (contemporaneously or with a lag) the variables being modeled is *always* tacitly assumed to be equal to the number of the variables. Therefore, any researcher who considers omitting a variable from a structural VAR also implicitly considers reducing the number of structural shocks, which is equivalent to or stronger than assumption (ii).

Granger-causal-priority is related to the issue of fundamentalness of structural shocks. The literature on fundamentalness, initiated by Hansen and Sargent (1991) and Lippi and Reichlin (1993), assumes that a vector $\tilde{y}$ follows a linear state-space model driven by structural shocks $\tilde{\varepsilon}$. That literature asks whether the structural shocks $\varepsilon_1 \subseteq \tilde{\varepsilon}$ are fundamental with respect to the variables $y_1 \subseteq \tilde{y}$, i.e. whether the structural VAR (6) exists. Failing to include in a structural VAR a variable that Granger-causes the included variables is a sufficient condition for nonfundamentalness (Giannone and Reichlin, 2006, Proposition 1). Therefore, if y_i is not Granger-causally-prior to y_j, omitting y_j gives rise to nonfundamentalness. Nonfundamentalness need not affect *all* structural shocks. It may happen that the impulse responses of y_i to a subset of structural shocks can still be recovered from the structural VAR (6).[9] However, a simple and prudent rule is to include y_j in the structural VAR to be used to infer impulse responses of y_i.

[9]See Forni and Gambetti (2011), Section 2.6.

3 Granger-noncausality: a closed-form Bayes factor

Consider a Bayesian econometrician who wants to evaluate the posterior odds in favor of the hypothesis of Granger-noncausality.[10] The posterior odds are equal to the prior odds multiplied by the Bayes factor in favor of the hypothesis. In this section, 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 (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 VAR equation. The likelihood of the data implied by this VAR, conditional on initial observations, is

$$p(Y|B,\Sigma) = (2\pi)^{-NT/2} |\Sigma|^{-T/2} \exp\left(-\frac{1}{2}\operatorname{tr}(Y - XB)'(Y - XB)\Sigma^{-1}\right), \tag{7}$$

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 a subset of equations 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

$$B_{\beta,\alpha} = \mathbf{0}, \tag{8}$$

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 (8).

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

[10] 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:

$$p(B, \Sigma|\omega^U) \propto |\Sigma|^{-(\tilde{\nu}+K+N+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\tilde{Y} - \tilde{X}B)'(\tilde{Y} - \tilde{X}B)\Sigma^{-1}\right), \tag{9}$$

where $\tilde{\nu} > 0$, $\tilde{Y}$ and $\tilde{X}$ are hyperparameters of appropriate dimensions, and $\tilde{X}'\tilde{X}$ is non-singular. Note that the standard prior used in VARs, the prior in Sims and Zha (1998) consisting of a modified Minnesota prior and additional dummy observations, is a special case of prior (9).

It is straightforward to show that

$$p(B, \Sigma|\omega^U) = p(B|\Sigma, \omega^U)p(\Sigma|\omega^U) = \mathcal{N}\left(\operatorname{vec}\tilde{B}, \Sigma \otimes \tilde{Q}\right)\mathcal{IW}\left(\tilde{S}, \tilde{\nu}\right), \tag{10}$$

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, it is straightforward to show that the posterior density of B and Σ, $p(B, \Sigma|Y, \omega^U)$, satisfies

$$p(B, \Sigma|Y, \omega^U) = p(B|\Sigma, Y, \omega^U)p(\Sigma|Y, \omega^U) = \mathcal{N}\left(\operatorname{vec}\bar{B}, \Sigma \otimes \bar{Q}\right)\mathcal{IW}\left(\bar{S}, \bar{\nu}\right), \tag{11}$$

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

$$p(B_{(\beta,\alpha)}, \Sigma|\omega^R) = p(B_{(\beta,\alpha)}, \Sigma|\omega^U, B_{\beta,\alpha} = \mathbf{0}). \tag{12}$$

Equation (12) 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 (12) 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 the prior $p(B, \Sigma|\omega^U)$. Suppose that this researcher obtains a new piece of information: he learns that $B_{\beta,\alpha} = 0$. Following the rules of probability, the researcher will update his prior precisely using equation (12). In addition to having this intuitive appeal, assumption (12) 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 (12).

The density $p(B_{(\beta,\alpha)}, \Sigma|\omega^R)$ is a conditional density of a normal-inverted-Wishart density $p(B, \Sigma|\omega^U)$ given in equation (10). It turns out that $p(B_{(\beta,\alpha)}, \Sigma|\omega^R)$ is a non-standard density. We can learn about $p(B_{(\beta,\alpha)}, \Sigma|\omega^R)$ by Monte Carlo simulation. We reach the following conclusions after comparing draws from $p(B_{(\beta,\alpha)}, \Sigma|\omega^R)$ of individual parameters with draws of the same parameters from $p(B_{(\beta,\alpha)}, \Sigma|\omega^U)$.[11] First, the densities of the same parameters have approximately the same shape. Second, the densities of the same parameters overlap substantially.[12] This Monte Carlo simulation supports the view that any researcher comfortable with $p(B_{(\beta,\alpha)}, \Sigma|\omega^U)$, which is a standard prior, should also be comfortable with $p(B_{(\beta,\alpha)}, \Sigma|\omega^R)$.

3.3 Closed-form Bayes factor

We are ready to state and prove the result of Section 3: the Bayes factor in favor of restriction (8) can be expressed in closed-form.

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

[11] We use a few of the restricted models from the empirical application of this paper.

[12] We measure the overlap with the frequently used "overlapping coefficient", defined as $\int \min(p_1(x), p_2(x))\, dx$, and taking values between 0 (no overlap) and 1 (identical distributions). For elements of $B_{(\beta,\alpha)}$ the overlap is on average 0.88 (and at least 0.19). For elements of Σ the overlap is on average 0.63 (and at least 0.09).

Proposition 2 *The Bayes factor in favor of model ω^R, defined in expressions (1), (8), and (12), relative to model ω^U, defined in expressions (1) and (9), is given by*

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

$$\times \frac{|\bar{S}_{\alpha,\alpha}|^{\frac{\bar{\nu}-N_{\bar{\alpha}}}{2}} |(\bar{Q}_{\beta,\beta})^{-1}|^{\frac{N_\alpha}{2}} |\bar{S}_{\alpha,\alpha} + \bar{B}'_{\beta,\alpha}(\bar{Q}_{\beta,\beta})^{-1}\bar{B}_{\beta,\alpha}|^{-\frac{\bar{\nu}-N_{\bar{\alpha}}+K_\beta}{2}}}{|\tilde{S}_{\alpha,\alpha}|^{\frac{\tilde{\nu}-N_{\bar{\alpha}}}{2}} |(\tilde{Q}_{\beta,\beta})^{-1}|^{\frac{N_\alpha}{2}} |\tilde{S}_{\alpha,\alpha} + \tilde{B}'_{\beta,\alpha}(\tilde{Q}_{\beta,\beta})^{-1}\tilde{B}_{\beta,\alpha}|^{-\frac{\tilde{\nu}-N_{\bar{\alpha}}+K_\beta}{2}}}, \tag{13}$$

where N_α denotes the number of elements in the set α, K_β denotes the number of elements in the set β, $N_{\bar{\alpha}} \equiv N - N_a$ denotes the number of elements in the complement of the set α, and $\Gamma_N(\cdot)$ denotes the multivariate gamma function of dimension N, $\Gamma_N(z) = \pi^{N(N-1)/4} \prod_{j=1}^N \Gamma(z + (1-j)/2)$.

Proof. **Step 1a:** Given expression (10), the marginal prior density of B is of matricvariate Student form, which implies that the marginal prior density of $B_{\beta,\alpha}$ is also of matricvariate Student form. **Step 1b:** Given expression (11), the marginal posterior density of B is of matricvariate Student form, which implies that the marginal posterior density of $B_{\beta,\alpha}$ is also of matricvariate Student form. Steps 1a-1b together with 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 satisfies (12), the Bayes factor in favor of the restriction $B_{\beta,\alpha} = \mathbf{0}$ against the alternative $B_{\beta,\alpha} \neq \mathbf{0}$ 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, expression (13) 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 $B_{\beta,\alpha} = \mathbf{0}$.[13] ∎

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 (13) to compute the Bayes

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

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.

Reporting the posterior odds in favor of model ω^R relative to model ω^U gives the same information as reporting the posterior probabilities of models ω^R and ω^U conditional on the set of models with two elements, ω^R and ω^U. The researcher begins attaching a prior probability to the model consistent with Granger-noncausality, ω^R, and a prior probability to the unrestricted model, ω^U. The researcher then revises his prior belief in light of the data, via the Bayes factor given in equation (13), arriving at the posterior probability of ω^R and the posterior probability of ω^U.

In the next section, we use Proposition 2 as a building block for evaluating 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, ω^R and ω^U. In 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.[14] 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*

[14] This is true in the realistic case when there are variables in y that belong neither to y_i nor to y_j. In our example in Section 2.1 with $y = \{x, w, z\}$, $y_i = x$ and $y_j = z$, there are two partitions of y consistent with Granger-causal-priority of y_i to y_j. In other words, there are two restricted models consistent with Granger-causal-priority of y_i to y_j.

that $y_i \subseteq y_1$.

We continue to assume as in Section 3 that the prior in the unrestricted model in Ω is conjugate and proper, i.e. satisfies expression (9), and the prior in each restricted model in Ω satisfies condition (12) for appropriate α and β. Furthermore, we assume that all models in Ω have equal prior probabilities.[15]

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

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

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

$$= \frac{\sum_{\omega_k \in \Omega^j} p(Y|\omega_k)}{\sum_{\omega_l \in \Omega} p(Y|\omega_l)} = \frac{\sum_{\omega_k \in \Omega^j} p(Y|\omega_k)/p(Y|\omega^U)}{\sum_{\omega_l \in \Omega} p(Y|\omega_l)/p(Y|\omega^U)}. \tag{14}$$

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 (13). 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. In contrast, the posterior probability of Granger-causal-priority conditional on some other set of models may be difficult to evaluate. Second, Ω is defined so as to treat the null hypothesis and the alternative

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

hypothesis as *symmetrically* as possible.[16] In particular, if y_j contains a single variable (as in the empirical application in this paper), Ω^j and its complement $\Omega \setminus \Omega^j$ have *equal size*. To see this, note that Ω contains 2^{N_J} models, where N_J denotes the number of variables in y_J. Furthermore, if y_j consists of a single variable, Ω^j contains 2^{N_J-1} models and $\Omega \setminus \Omega^j$ also contains 2^{N_J-1} models. Thus, if y_j contains a single variable, evaluating the posterior probability of Granger-causal-priority amounts to evaluating the posterior odds in favor of a subset of models against the alternative of a subset of models of *equal size*.[17]

Given expression (14) 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. The researcher then revises his prior belief in light of the data, via expression (14), arriving at the posterior probability of y_i being Granger-causally-prior to y_j.[18]

Let us emphasize the following standard property of posterior probability.[19] 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, a single model in Ω has a posterior probability of one asymptotically and all other models in Ω have a posterior probability of zero asymptotically. The single model with the asymptotic posterior probability of one 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.

Finally, we comment on computation. Although in principle one can evaluate expression (14) exactly, when N_J is large our computers may be too slow to calculate all sums in this expression. In the empirical application in Section 5, N_J is large and we approximate $p(\Omega^j|Y,\Omega)$ using the Markov chain Monte Carlo model composition algorithm of Madigan

[16] 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.

[17] While it is sensible to evaluate the posterior probability of Granger-causal-priority conditional on Ω, this is not the only possible approach. For example, one could evaluate the posterior probability of Granger-causal-priority conditional on the set of models $\tilde{\Omega}$ defined in Section 6.

[18] The prior probability is equal to 0.5 because (i) all models in Ω are assumed to have equal prior probabilities, and (ii) Ω^j and its complement $\Omega \setminus \Omega^j$ have equal size. If either (i) or (ii) fails to hold, the prior probability will in general be some other number between one and zero. The prior probability then gets updated in exactly the same way, except that if (i) fails to hold, a trivial modification of expression (14) is required.

[19] See, for example, Fernandez-Villaverde and Rubio-Ramirez (2004).

and York (1995). This algorithm is simple and converges reliably. In Appendix A we explain how we implement this algorithm and assess its convergence.

5 Empirical application: probability of Granger-causal-priority

We turn to the empirical application of this paper. The empirical application is presented in this section and in Section 6.

In this section, we define y, y_i, and y_J and, for each variable $y_j \in y_J$, we evaluate the posterior probability that y_i is Granger-causally-prior to y_j using the methodology from Section 4. Asymptotically, the posterior probabilities lead to a unique choice of variables that belong in the VAR and in the structural VAR with y_i. In our finite sample, we find some y_j's associated with a posterior probability of Granger-causal-priority *either* close to zero *or* close to one. However, we also find some y_j's associated with a posterior probability of Granger-causal-priority *neither* close to zero *nor* close to one. See Section 5.3. This is as expected, because in any finite sample the posterior probabilities *per se* generally do not lead to a unique choice of variables. One can then make a decision informally. For example, one can stipulate that a y_j belongs in the VAR with y_i so long as the posterior probability that y_i is Granger-causally-prior to y_j is smaller than some number close to zero.

Another possibility is to make a decision formally, as in Bayesian decision theory. This requires specifying a loss function, i.e., a function that assigns a numerical loss to every combination of actual variable choice and the correct variable choice, and minimizing the posterior expected loss. In Section 6 we specify a loss function and choose a set of variables by minimizing the posterior expected loss. This formal decision turns out to lead to a choice of variables very similar to the choice of variables that we could make informally based on the findings in Section 5.3, if we stipulated that "a y_j belongs in the VAR with y_i so long as the posterior probability that y_i is Granger-causally-prior to y_j is smaller than 0.1."[20]

5.1 Data: defining y, y_i, and y_J

We put together two datasets, one for the euro area and one for the United States. Each dataset has three variables of interest (i.e., in each dataset there are three variables in y_i)

[20]To minimize a posterior expected loss, in general one needs to use posterior probabilities of individual models in the set of models under consideration (and not only posterior probabilities of Granger-causal-priority). See Section 6.

and thirty-eight remaining variables (i.e., in each dataset there are thirty-eight variables in y_J). Table 1 lists the variables in both datasets.[21]

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

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

The thirty-five variables are: (i) real GDP components (consumption, government consumption, investment, exports, imports, change in inventories); (ii) the level of government debt; (iii) labor market variables (unit labor costs, employment, unemployment rate, hours worked);[22] (iv) interest rates (2-year and 10-year government bond yields, the spread between corporate bonds rated BBB with maturity 7-10 years and government bonds with the same maturity, lending rate for non-financial corporations, mortgage interest rate); (v) monetary aggregates (M1, M2, M3);[23] (vi) credit aggregates (real estate loans, consumer credit, bank credit to non-financial corporations); (vii) exchange rates (nominal exchange rate between the euro and the U.S. dollar, nominal effective exchange rate); (viii) commodity prices and other price indices (price of oil, index of commodity prices, both measured in U.S. dollars, producer price index, consumer price index excluding energy and food);[24] (ix) housing market variables (index of house prices, real housing investment); (x) stock market variables (stock market index, implied volatility index); (xi) survey-based leading indicators of economic activity (consumer confidence, industrial confidence, purchasing managers' index, capacity utilization in manufacturing).

[21] The source of the data is the database of the ECB. The data are available from the authors upon request.

[22] Due to data availability we include hours worked only for the United States.

[23] Due to data availability we include M3 only for the euro area.

[24] The variables price of oil and index of commodity prices are the same variables in the euro area exercise and in the U.S. exercise.

Table 1: Variable names, units and transformations.

Variable	Units	Transformation	
Euro Area Real GDP	real currency units†	SA	log
HICP (Consumer Prices)	index	SA	log
Eonia (Overnight Interbank Rate)	percent		none
U.S. Real GDP	real currency units†	SA	log
U.S. CPI	index	SA	log
U.S. Federal Funds Rate	percent		none
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 of Real Inventories	percent of Real GDP	SA	none
Government Debt	nominal currency units ‡	SA	log
Unit Labor Cost	-	SA	none
Total Employment	thousands of people	SA	log
Unemployment Rate	percent	SA	none
Hours Worked (U.S. only)	hours	SA	log
2 Year Bond Yield	percent		none
10 Year Bond Yield	percent		none
BBB Bond Spread 7-10 Years	percent		none
Lending Rate to NFC	percent		none
Mortgage Interest Rate	percent		none
M1	nominal currency units ‡	SA	log
M2	nominal currency units ‡	SA	log
M3 (euro area only)	nominal currency units ‡	SA	log
Real Estate Loans Outstanding	nominal currency units ‡		log
Consumer Credit Outstanding	nominal currency units ‡		log
Bank Credit to NFC Outstanding	nominal currency units ‡		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 Price Index (PPI)	index	SA	log
Residential Property Prices	index	SA	log
Real Housing Investment	real currency units†	SA	log
Stock Market Index*	index		log
Stock Market Volatility Index**	percent		log
Capacity Utilization	percent	SA	none
Consumer Confidence	index	SA	none
Industrial Confidence	index	SA	none
PMI	index	SA	none

Notes: † Euro area: millions of chained 2005 euros; U.S.: billions of chained 2000 dollars. ‡ Euro area: millions of euros; U.S.: billions of dollars. * Euro area: Dow Jones Euro Stoxx; U.S.: Dow Jones Industrial Average. ** Euro area: VSTOXX, before the year 2000 extended back with VIX; U.S.: VIX.

The main sample contains quarterly data from 1999Q1 to 2011Q2. 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.

5.2 Prior

We use the standard prior employed in VARs following Sims and Zha (1998). The Sims-Zha prior is controlled by several hyperparameters. We choose values of the hyperparameters implying a *tighter* prior than in Sims and Zha (1998). We like the idea of using a relatively tight prior, because a tighter prior makes it *harder* to find support for zero restrictions in a VAR. Furthermore, our choice is consistent with the suggestion of Giannone et al. (2012) to use a tighter prior when the number of variables in a VAR is larger than in Sims and Zha (1998).

Let us give the details. The prior consists of two pieces: (i) an *initial prior* formulated before seeing any data, and (ii) a *training sample prior*. Formally, matrices $\tilde{Y}$ and $\tilde{X}$ in expression (9) have the form

$$\tilde{Y} = \begin{pmatrix} Y_{SZ} \\ Y_{ts} \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} X_{SZ} \\ X_{ts} \end{pmatrix}, \tag{15}$$

where Y_{SZ}, Y_{ts}, X_{SZ}, and X_{ts} are defined next. The initial prior is the Sims-Zha prior implemented by creating dummy observations Y_{SZ} and X_{SZ}.[25] We choose the values of the hyperparameters controlling the Sims-Zha prior based on an approach that is common in Bayesian econometrics. We select the values of the hyperparameters that maximize the marginal likelihood of the VAR (1) in the training sample: the "overall tightness" of 0.05, the weight of the "one-unit-root" dummy of 1, and the weight of the "no-cointegration" dummies of 2.[26] These values imply a tighter prior than in Sims and Zha (1998).[27] Below, we refer to the prior with these hyperparameter values as the "baseline" and we report how our findings change as we vary the values of the hyperparameters. The matrices Y_{ts} and X_{ts} in expression (15) contain data from the training sample, 1989Q1-1998Q4. We found that adding this training sample improved the marginal likelihood implied by the VAR (1)

[25]The term $\tilde{\nu}$ in expression (9) also belongs to the initial prior.
[26]See Appendix B for more details concerning the Sims-Zha prior and our choice of hyperparameter values.
[27]Sims and Zha (1998) use 0.2, 1, and 1, respectively.

in the sample 1999Q1-2011Q2 compared with using the Sims-Zha prior only, both in the euro area exercise and in the U.S. exercise.

5.3 Main findings

Table 2 reports the posterior probability that y_i (GDP, the consumer price level, and the short-term interest rate) is Granger-causally-prior to each variable $y_j \in y_J$, in the euro area exercise (left column) and in the U.S. exercise (right column).[28]

Three main findings are evident from Table 2:

(1) The posterior probability of Granger-causal-priority is close to zero for: (i) survey-based leading indicators, (ii) the change in inventories, (iii) the corporate bond spread, (iv) the unemployment rate, and (v) the price of oil. The posterior probabilities of Granger-causal-priority associated with these variables are smaller than or about equal to 0.1.[29]

(2) The posterior probability of Granger-causal-priority is close to one for: (i) house prices, and (ii) broad money. The posterior probability of Granger-causal-priority with respect to house prices is 0.99 (euro area) and 0.98 (United States). The posterior probability of Granger-causal-priority with respect to broad money is 0.93-0.94 (euro area M2 and M3) and 0.92 (U.S. M2).

(3) The first main finding and the second main finding are *the same* in the euro area exercise and in the U.S. exercise. In general, the findings are remarkably similar in the euro area and in the United States. The correlation between the posterior probabilities in the euro area exercise and in the U.S. exercise is 0.66.

Let us comment on some additional findings apparent from Table 2.

The GDP component with the second lowest posterior probability, after the change in inventories, is investment (0.15 in the euro area and 0.21 in the United States).

The federal funds rate emerges as the U.S. variable with the lowest posterior probability in the euro area exercise (0.17). The other two U.S. variables have fairly high posterior probabilities in the euro area exercise. In the U.S. exercise, all three euro area variables have fairly high posterior probabilities.

[28] Table 2 reports the findings obtained with one lag, i.e. $P = 1$. Below we discuss the effects of changing the value of P.

[29] In the euro area exercise, two survey-based leading indicators are associated with a posterior probability of Granger-causal-priority smaller than or about equal to 0.1. The same is true in the U.S. exercise. In the euro area exercise the two indicators are industrial confidence and consumer confidence. In the U.S. exercise the two indicators are PMI and industrial confidence.

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

Euro area			U.S.	
Variable	prob.	rank	Variable	prob.
Industrial Confidence	0.00	1	BBB Bond Spread 7-10 Years	0.01
Consumer Confidence	0.03	2	PMI	0.06
Unemployment Rate	0.03	3	Change of Real Inventories	0.10
Change of Real Inventories	0.06	4	Oil Price	0.11
BBB Bond Spread 7-10 Years	0.10	5	Industrial Confidence	0.11
Oil Price	0.11	6	Unemployment Rate	0.14
Real Investment	0.15	7	Capacity Utilization	0.17
PMI	0.17	8	Real Investment	0.21
U.S. Federal Funds Rate	0.17	9	Lending Rate to NFC	0.22
Real Exports	0.18	10	Consumer Confidence	0.22
Real Imports	0.20	11	Hours Worked (U.S. only)	0.32
Lending Rate to NFC	0.24	12	Stock Market Volatility Index**	0.36
Real Consumption	0.32	13	Real Imports	0.37
Unit Labor Cost	0.36	14	Total Employment	0.38
Real Housing Investment	0.37	15	Producer Price Index (PPI)	0.42
2 Year Bond Yield	0.39	16	10 Year Bond Yield	0.42
Bank Credit to NFC Outstanding	0.44	17	2 Year Bond Yield	0.42
Total Employment	0.44	18	Mortgage Interest Rate	0.44
Mortgage Interest Rate	0.45	19	Nominal Effective Exchange Rate	0.45
Producer Price Index (PPI)	0.50	20	Commodity Prices	0.45
Real Government Consumption	0.53	21	Real Consumption	0.46
U.S. CPI	0.56	22	Euro Area Real GDP	0.48
Commodity Prices	0.61	23	M1	0.56
Capacity Utilization	0.64	24	Real Exports	0.59
U.S. Real GDP	0.69	25	HICP (Consumer Prices in Europe)	0.59
Real Estate Loans Outstanding	0.69	26	Eonia (Euro Overnight Interbank Rate)	0.60
Consumer Prices Excl. Energy, Food	0.73	27	Stock Market Index*	0.64
10 Year Bond Yield	0.73	28	Consumer Prices Excl. Energy, Food	0.75
Stock Market Index*	0.82	29	Bank Credit to NFC Outstanding	0.77
M1	0.86	30	Unit Labor Cost	0.78
Stock Market Volatility Index**	0.89	31	Dollar-Euro Exchange Rate	0.79
M3 (euro area only)	0.93	32	Government Debt	0.82
Nominal Effective Exchange Rate	0.94	33	Consumer Credit Outstanding	0.82
M2	0.94	34	Real Government Consumption	0.85
Consumer Credit Outstanding	0.94	35	Real Housing Investment	0.86
Government Debt	0.99	36	M2	0.92
Residental Property Prices	0.99	37	Real Estate Loans Outstanding	0.93
Dollar-Euro Exchange Rate	1.00	38	Residental Property Prices	0.98

Notes: See the notes under Table 1

The evidence in favor of Granger-causal-priority with respect to government debt is strong in the euro area (a posterior probability of 0.99) and moderate-to-strong in the United States (a posterior probability of 0.82).

Real estate loans and consumer credit are associated with fairly high posterior probabilities (0.69 and 0.94, respectively, in the euro area; 0.93 and 0.82, respectively, in the United States). The analogous number for the third credit aggregate, bank credit to non-financial corporations, is also fairly high in the United States (0.77), but moderate in the euro area (0.44).

Both stock market variables and both measures of the exchange rate are associated with high posterior probabilities in the euro area (between 0.82 and 1). So is narrow money, M1 (0.86).

5.4 Robustness

We studied robustness of the findings shown in Table 2. In particular, we repeated the analysis in subsamples, varied the values of the hyperparameters of the Sims-Zha prior, and added lags. The details are in Appendix C. Let us emphasize here what we see as the most important findings concerning robustness.

When we drop the last four quarters and the last eight quarters, respectively, from the sample, the results change hardly at all. Note that omitting eight quarters from the sample amounts to dropping as much as one-sixth of the sample. We conclude that the findings are not sensitive to non-trivial perturbations of the sample.

When we drop the last twelve quarters from the sample, i.e. when we drop the *entire* crisis period (as much as one-fourth of the sample), the findings change somewhat. For example, the posterior probability of Granger-causal-priority with respect to the corporate bond spread rises notably in the euro area exercise (but *not* in the U.S. exercise). We are not surprised that the results change between the full sample and the sample excluding the entire crisis period, but we emphasize that the findings change only *somewhat*. See Appendix C.

Finally, the results are robust to reasonable variation in the values of the hyperparameters. The results change notably only when we assume values of the hyperparameters implying that the models fit the data very poorly. Again, see Appendix C.[30]

[30]We always condition on fixed values of the hyperparameters. In Section 5 we report the main findings

6 Empirical application: choosing the single best set of variables

In this section we specify a loss function and choose a set of variables of minimizing the posterior expected loss. In particular, we assume the zero-one loss function. A researcher with the zero-one loss function seeks a model implying the highest marginal likelihood of the data. This model will have in y_1 the variables of interest and possibly one or more remaining variables. We declare any remaining variables in y_1 in this model "the best variables." We have to compute marginal likelihood by Monte Carlo in this section. To be robust against Monte Carlo error, we do not literally search for the single model with the highest marginal likelihood. Instead, we find a set of models with highest marginal likelihoods. It turns out that the models in this set have the same, or nearly the same, variables in y_1.

We work in this section with a set of models larger than the set Ω. The reason for extending the set of models is the following. We evaluated the marginal likelihood of models in the set Ω. We found that the models with high marginal likelihood typically have about the same number of variables in y_1 and in y_2. It is simple to see that when y_1 and y_2 have equal size, the number of zero restrictions in a model in Ω is maximized. The fact that the best models in Ω have the largest possible number of zero restrictions suggests that the constraint on the number of zero restrictions implied in the definition of Ω is binding. We work with a larger set of models in order to relax that constraint.

Below we define the extended set of models $\tilde{\Omega}$, give the methodological details, comment on the data, and present the findings.

6.1 Multiple Granger-noncausality relations

The difference between $\tilde{\Omega}$ and Ω is that $\tilde{\Omega}$ includes models with multiple Granger-noncausality relations whereas Ω includes models with one Granger-noncausality relation.

Definition 5 *Let $\tilde{\Omega}$ be the set of models such that: (i) each model in $\tilde{\Omega}$ is a VAR of the form*

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

given in equation (1), (ii) $\tilde{\Omega}$ includes the unrestricted VAR, (iii) $\tilde{\Omega}$ includes all VARs with a sequence of $r = 1, ..., R \geq 1$ restrictions where each restriction has the form $B_{12}^r(L) = 0$ for some partition of y into two subsets, y_1^r and y_2^r, such that $y_i \subseteq y_1^r$ and for $r > 1$ y_1^{r-1} is a strict subset of y_1^r.

To understand the definition of $\tilde{\Omega}$ consider again the trivariate example from Section 2.1. In this example, Ω contains: (i) the unrestricted VAR, (ii) the VAR with the single restriction that (w, z) do not Granger-cause x, (iii) the VAR with the single restriction that z does not Granger-cause (x, w), and (iv) the VAR with the single restriction that w does not Granger-cause (x, z). $\tilde{\Omega}$ contains Ω *as well as*: (v) the VAR with two restrictions, (ii) and (iii), imposed simultaneously and (vi) the VAR with two restrictions, (ii) and (iv), imposed simultaneously. Cases (ii), (iii) and (v) are illustrated below.

$$\text{case (ii), in } \Omega \text{ and } \tilde{\Omega}: \begin{pmatrix} B_{xx} & 0 & 0 \\ B_{wx} & B_{ww} & B_{wz} \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix} \quad \text{case (iii), in } \Omega \text{ and } \tilde{\Omega}: \begin{pmatrix} B_{xx} & B_{xw} & 0 \\ B_{wx} & B_{ww} & 0 \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix} \quad \text{case (v), only in } \tilde{\Omega}: \begin{pmatrix} B_{xx} & 0 & 0 \\ B_{wx} & B_{ww} & 0 \\ B_{zx} & B_{zw} & B_{zz} \end{pmatrix}$$

When we test multiple restrictions $B_{12}^r(L) = 0$ simultaneously, we restrict different sets of coefficients in different VAR equations. Therefore, the analytical results of Section 3 do not apply. Instead, we approximate the Savage-Dickey ratio numerically with a simple Monte Carlo. Let us explain what we do by focusing on the case of two restrictions. A generalization is straightforward.

Consider a partition of y into three subsets, $y = \{y_1, y_2, y_3\}$. Suppose that we want to test simultaneously two restrictions: (1) y_1 is Granger-causally-prior to (y_2, y_3) and (2) (y_1, y_2) is Granger-causally-prior to y_3. Let α_1 denote the numbers of VAR equations with variables y_1 on the left-hand side. Let β_1 denote the positions of the lags of the variables (y_2, y_3) among the right-hand-side variables. Let α_2 denote the numbers of VAR equations with variables y_2 on the left-hand side. Let β_2 denote the positions of the lags of the variables y_3 among the right-hand-side variables. We want to test zero restrictions for (1) the coefficients of variables β_1 in the equations α_1, i.e. $B_{\beta_1\alpha_1}$ and (2) the coefficients of variables β_2 in the equations α_2, i.e. $B_{\beta_2\alpha_2}$. (Note that restriction (2) requires also that we restrict the coefficients of variables β_2 in the equations α_1, $B_{\beta_2\alpha_1}$. However, since $\beta_2 \subset \beta_1$, we have that $B_{\beta_2\alpha_1} \subset B_{\beta_1\alpha_1}$ and so the coefficients $B_{\beta_2\alpha_1}$ are already included in the first restriction, so we only need to restrict $B_{\beta_2\alpha_2}$ in addition.) The two restrictions are $B_{\beta_1\alpha_1} = \mathbf{0}$ and $B_{\beta_2\alpha_2} = \mathbf{0}$ or, vectorizing and stacking together, $((\operatorname{vec} B_{\beta_1\alpha_1})', (\operatorname{vec} B_{\beta_2\alpha_2})')' = 0$.

The priors are the same as the priors in Section 4. Namely, the prior about the parameters of the unrestricted VAR, ω^U, is still given by expression (9) and the prior about the parameters of the restricted VAR, ω^R, is analogous to expression (12), specifically

$$p(B_{(\alpha_1\beta_1;\alpha_2\beta_2)}, \Sigma|\omega^R) = p(B_{(\alpha_1\beta_1;\alpha_2\beta_2)}, \Sigma|\omega^U, B_{\beta_1\alpha_1} = \mathbf{0}, B_{\beta_2\alpha_2} = \mathbf{0}).$$

The marginal densities of $\operatorname{vec} B_{\beta_1\alpha_1}$ and $\operatorname{vec} B_{\beta_2\alpha_2}$ in the in the unrestricted VAR are each matricvariate Student, but the joint density of $((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')'$, which enters the Savage-Dickey ratio, is not a standard density. In particular, this joint density is not matricvariate Student because the variance-covariance matrix of $((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')'$ does not have the Kronecker structure.[31]

However, the density of $((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')'$ *conditional on* Σ is multivariate normal. The reason is that $((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')'$ is a subvector of $\operatorname{vec} B$ which, conditionally on Σ, is multivariate normal. See equations (10) and (11). Therefore, the marginal prior density at zero can be approximated from M Monte Carlo draws of Σ from its prior $p(\Sigma|\omega^U)$ as

$$p\left(((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')' = 0|\omega^U\right)$$
$$= \frac{1}{M}\sum\nolimits_{m=1}^{M} p\left(((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')' = 0|\Sigma^m, \omega^U\right).$$

Analogously, the marginal posterior density at zero can be approximated from M Monte Carlo draws of Σ from its posterior $p(\Sigma|Y, \omega^U)$ as

$$p\left(((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')' = 0|Y, \omega^U\right)$$
$$= \frac{1}{M}\sum\nolimits_{m=1}^{M} p\left(((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')' = 0|\Sigma^m, Y, \omega^U\right).$$

Recall that Ω contains 2^{N_J} elements. The number of elements in $\tilde{\Omega}$ increases with N_J much faster than does the number of elements in Ω. It turns out that the number of elements

[31] The joint density of $((\operatorname{vec} B_{\beta_1,\alpha_1})', (\operatorname{vec} B_{\beta_2,\alpha_2})')'$ is not multivariate Student, either. See Appendix A of Bauwens et al. (1999) for the definitions and properties of matricvariate and multivariate Student densities.

in $\tilde{\Omega}$, denoted $K(N_J)$, is given by $K(N_J) = 2C(N_J)$, where

$$C(N_J) = \sum_{k=0}^{N_J - 1} \binom{N_J}{k} C(k). \tag{16}$$

For example: $K(N_J = 2) = 6$, $K(N_J = 3) = 26$, $K(N_J = 4) = 150$, and so on.[32] In the empirical application that follows we have $N_J = 14$ and thus $K(N_J = 14) \approx 2.13 \times 10^{13}$.

6.2 Data: defining y, y_i, and y_J

In this section we use a smaller dataset than in Section 5 because the computational burden in this section is high, for three reasons. First, the number of models in $\tilde{\Omega}$ grows much faster with the number of variables than does the number of models in Ω. Second, in this section we use a Monte Carlo to evaluate the Bayes factor in favor of multiple Granger-causal-priority restrictions, while in the case of a single Granger-causal-priority restriction we used the analytical result of Section 3 to evaluate the Bayes factor. Third, we need to explore $\tilde{\Omega}$ more thoroughly than we explored Ω, since we are looking for the best model(s) and not merely approximating posterior probabilities of large sets of models, as in the computation of Granger-causal-priority probabilities in Section 5.

We use a subset of the data employed in Section 5, with seventeen variables out of the forty-one variables used in Section 5. We have the same three variables of interest and we have fourteen remaining variables. We include five variables associated with low Granger-causal-priority probabilities: industrial confidence, unemployment rate, change in inventories, corporate bond spread, and oil price. We include three variables associated with high Granger-causal-priority probabilities: housing investment, stock market index, and nominal effective exchange rate. We also include six variables associated with medium Granger-causal-priority probabilities: consumption, investment, commodity prices, the federal funds rate (in the euro area exercise) and Eonia (in the U.S. exercise), 10-year government bond yield, and M1. We aim to achieve two objectives with this choice of variables. We want

[32] To see where equation (16) comes from, consider the number of Granger-causal-priority relations among N_J variables, denoted $C(N_J)$. Since Granger-causal-priority is a transitive relation, $C(N_J)$ is equal to the number of weak orders of N_J elements. One can show that $C(N_J)$ is given by (16) and, for a large N_J, $C(N_J) \approx N_J!/\left(2(\ln 2)^{N_J+1}\right)$. See OEIS (2011). The number of elements in $\tilde{\Omega}$ is equal to $2C(N_J)$, i.e. twice the number of Granger-causal-priority relations among the N_J variables in $y \setminus y_i$. The reason why multiplication by two is necessary is that, given each pattern of Granger-causal-priority relations within $y \setminus y_i$, we can either have or not have y_i Granger-causally-prior to all of $y \setminus y_i$.

to include variables that are likely, based on the results in Section 5, to be most relevant for the VAR with the variables of interest. Furthermore, we want to include some variables that are unlikely, based on the results in Section 5, to be most relevant because we wish to check whether the conclusions of Section 5 are robust to redefining the set of models.

6.3 Main findings

What is the single best set of variables that belongs in the VAR with GDP, the consumer price level, and the short-term interest rate? The answer to this question in the euro area exercise is: industrial confidence, change in inventories, federal funds rate, corporate bond spread, and unemployment rate. The answer to this question in the U.S. exercise is: industrial confidence, change in inventories, corporate bond spread, and oil price. Let us explain how we obtain these answers.

We search for the set of best models in $\tilde{\Omega}$, i.e., the set of all models with marginal likelihood not materially different from the marginal likelihood of the single best model in $\tilde{\Omega}$. We define the set of best models in $\tilde{\Omega}$ as the set of all $\omega \in \tilde{\Omega}$ such that $\ln p(Y|\tilde{\omega}^*) - \ln p(Y|\omega) < 1$, where $\tilde{\omega}^*$ is the model in $\tilde{\Omega}$ with the highest marginal likelihood.[33,34]

In the euro area exercise, the set of best models contains fifty-three models. The first column in Table 3 summarizes which variables enter the first blocks, y_1, of the best models in the euro area exercise. Industrial confidence, change in inventories, and the federal funds rate enter y_1 in *every one* of the best models. The corporate bond spread enters y_1 in 98 percent of the best models. The unemployment rate enters y_1 in 85 percent of the best models. No other variable enters y_1 in more than 4 percent of the best models. Let us also report two findings not apparent from Table 3. In 77 percent of the best models y_1 consists *only* of industrial confidence, change in inventories, the federal funds rate, the corporate bond spread, and the unemployment rate – in addition to GDP, consumer prices, and short-term interest rate. In the other 23 percent of the best models y_1 includes one more or one less variable.

[33] We use one log point difference in marginal likelihood as the threshold following Kass and Raftery (1995, p.77), who advise to interpret differences in marginal likelihood of up to one log point as "not worth more than a bare mention". The results are very similar when we enlarge the set of best models by increasing the threshold to two log points.

[34] In both the euro area exercise and the U.S. exercise, we run seven independent MC^3 chains. We start each chain at a different randomly chosen model and in each chain we draw one million models. We present here the results based on all the seven chains stacked together. All the reported results are very similar across the seven chains.

Table 3: Percent of best models with a given variable in the first block.

Variable	Euro Area	US
Industrial Confidence	100	100
Unemployment Rate	85	0
Change of Real Inventories	100	100
BBB Bond Spread 7-10 Years	98	100
Oil Price	4	80
Real Investment	2	0
U.S. Federal Funds Rate/Eonia†	100	0
Real Consumption	0	10
Real Housing Investment	2	0
Commodity Prices	0	0
10 Year Bond Yield	0	0
Stock Market Index*	0	0
M1	0	0
Nominal Effective Exchange Rate	0	10

Notes: †Euro area exercise: Fed Funds Rate, U.S. exercise: Eonia. ∗ See the notes under Table 1

In the U.S. exercise, the set of best models contains ten models. The second column in Table 3 summarizes which variables enter the first blocks, y_1, of the best models in the U.S. exercise. Industrial confidence, change in inventories, and the corporate bond spread enter y_1 in *every one* of the best models. The oil price enters y_1 in 80 percent of the best models. No other variable enters y_1 in more than one of the best models. In particular, in one of the best models consumption rather than the oil price enters y_1; and in one of the best models the exchange rate rather than the oil price enters y_1.

Table 4 reports the log marginal likelihood for a few benchmark models from Ω and $\tilde{\Omega}$.[35] We draw the following lessons from Table 4. The marginal likelihood of the best model in $\tilde{\Omega}$ is higher than the marginal likelihood of the best model in Ω. We conclude that it is important to consider multiple Granger-causal-priority restrictions when searching for the best model. Furthermore, the marginal likelihood of the best model in Ω and the marginal likelihood of the best model in $\tilde{\Omega}$ are higher than the marginal likelihood of the unrestricted model. We conclude that it is pays off to impose Granger-causal-priority restrictions in a large VAR.

In the last two rows of Table 4 we report the marginal likelihood of selected models with

[35] One can compare the log marginal likelihoods across alternative models and priors, but one cannot compare euro area models with U.S. models.

Table 4: Log marginal likelihoods of individual models.

	Euro area	U.S.
The best model in $\tilde{\Omega}$: $\tilde{\omega}^*$	1166.3	753.5
The best model in Ω: ω^*	1161.4	747.0
The unrestricted model	1152.9	740.0
The model opposite† to ω^*	1138.6	729.2
The model opposite‡ to $\tilde{\omega}^*$	1136.2	718.8

Notes: †The first block of this model consists of y_i and the second block of ω^*. ‡The first block of this model consists of y_i and the last block of $\tilde{\omega}^*$. The second block of this model consists of the last-but-one block of $\tilde{\omega}^*$, etc...

restrictions that are very different from the restrictions in the best models. See the notes under Table 4 for the details. The marginal likelihood of these selected models is much lower than the marginal likelihood of the unrestricted model. We conclude that using a large unrestricted VAR is a better option than using a small VAR with incorrectly chosen variables.

6.4 Robustness

We repeat the analysis in the full sample with the standard Sims-Zha prior instead of the baseline prior. We also repeat the analysis in the same subsamples we considered in Section 5.3. The results are in Table 5 (euro area exercise) and Table 6 (U.S. exercise). The takeaways are as follows. When we change the prior, we get exactly the same answer to the question "what is the single best set of variables?", in the euro area exercise and in the U.S. exercise. The answer to this question changes only slightly when we drop the last four quarters from the sample, since in the U.S. exercise an additional variable, the unemployment rate, now enters the single best set of variables. When we drop the last twelve quarters, i.e. the entire crisis period, from the sample, the evidence in favor of some variables being in the single best set of variables weakens. In particular, this is true for the unemployment rate, the corporate bond spread and the federal funds rate in the euro area exercise; and industrial confidence and the oil price in the U.S. exercise. Furthermore, the oil price now enters the single best set of variables in the euro area exercise. In other words, in the subsample without the crisis period the single best set of variables consists of the following variables. In the euro area exercise: industrial confidence, change in inventories and oil price; in the U.S. exercise: change in inventories and corporate bond spread, and possibly also industrial confidence, unemployment rate, oil price and M1.

Table 5: Percent of best models with a given variable in the first block - sensitivity analysis for the Euro area.

Variable	baseline	standard	drop 4Q	drop 12Q
Industrial Confidence	100	100	100	100
Unemployment Rate	85	100	90	0
Change of Real Inventories	100	100	100	55
BBB Bond Spread 7-10 Years	98	86	70	0
Oil Price	4	0	0	93
Real Investment	2	0	0	2
U.S. Federal Funds Rate	100	100	100	0
Real Consumption	0	0	0	0
Real Housing Investment	2	0	0	0
Commodity Prices	0	0	0	0
10 Year Bond Yield	0	0	0	0
Stock Market Index*	0	0	0	0
M1	0	0	0	0
Nominal Effective Exchange Rate	0	0	0	0
number of best models	53	7	10	44

Notes: ∗ - see notes under table 1.

Table 6: Percent of best models with a given variable in the first block - sensitivity analysis for the U.S.

Variable	baseline	standard	drop 4Q	drop 12Q
Industrial Confidence	100	100	†100	20
Unemployment Rate	0	0	†63	20
Change of Real Inventories	100	100	†100	100
BBB Bond Spread 7-10 Years	100	100	†100	80
Oil Price	80	100	†100	20
Real Investment	0	0	0	0
Eonia	0	0	0	0
Real Consumption	10	0	13	0
Real Housing Investment	0	0	0	0
Commodity Prices	0	8	0	0
10 Year Bond Yield	0	0	0	0
Stock Market Index*	0	0	0	0
M1	0	0	0	20
Nominal Effective Exchange Rate	10	0	13	0
number of best models	10	12	†8	5

Notes: ∗ See notes under table 1. †In the 'drop 4Q' exercise the set of best models consisted of only one model. Therefore, we signal the variables entering that one model with † and present percentages for a wider definition of the best models that uses the threshold of 2 log points.

7 Alternatives

In this section, we compare the methodology laid out in this paper with three alternatives. The methodology laid out in this paper involves evaluating the marginal likelihood, i.e., the prior predictive density, of all data $Y = (Y_i, Y_J)$ and comparing the value of the marginal likelihood in models with different restrictions. In contrast, each of the alternatives relies on evaluation of a prior predictive density of the variables of interest Y_i only.

Let ψ denote a model of y_i and y_j^{ψ}, where $y_j^{\psi} \subseteq y \setminus y_i$. Consider the following three statistics. The first statistic is the marginal predictive density of Y_i, that is, the marginal likelihood of $\left(Y_i, Y_j^{\psi}\right)$ marginalized with respect to Y_j^{ψ}

$$p(Y_i|\psi) = \int p(Y_i, Y_j^{\psi}|\psi) dY_j^{\psi}. \tag{17}$$

The second statistic is the predictive density of Y_i conditional on the actually observed Y_j^{ψ}

$$p(Y_i|Y_j^{\psi}, \psi) = \frac{p(Y_i, Y_j^{\psi}|\psi)}{\int p(Y_i, Y_j^{\psi}|\psi) dY_i}. \tag{18}$$

The third statistic is the predictive density score[36] of Y_i at horizon $h > 0$, typically computed as

$$g(Y_i, h|\psi) = \prod_{t=1}^{T-h} p(y_i(t+h)|y(\tau : \tau \leq t), \psi). \tag{19}$$

Let us understand what each of these statistics tells us and compare these statistics to the statistic that we use. We first present four useful expressions, and then we discuss these expressions.

For the marginal predictive density of Y_i, we have

$$p(Y_i|\psi) = p(y_i(1, ..., T)|y_i(-P+1, ..., 0), y_j^{\psi}(-P+1, ..., 0), \psi)$$

$$= \prod_{q=1}^{Q} p(y_i(s_{q-1}+1, ..., s_q)|y_i(-P+1, ..., s_{q-1}), y_j^{\psi}(-P+1, ..., 0), \psi). \tag{20}$$

In equation (20) we partition the sequence of dates $0, 1, ..., T$ using a strictly increasing sequence of integers $\{s_q\}_{q=0}^{Q}$ with $s_0 = 0$ and $s_Q = T$.[37] Furthermore, we make explicit the

[36] The predictive density score is used in many papers. See Geweke and Amisano (2011) for a discussion.

[37] This partitioning follows Geweke (2005), p. 67.

conditioning on the P initial observations $y_i(-P+1,...,0)$ and $y_j^{\psi}(-P+1,...,0)$.

For the predictive density of Y_i conditional on the actually observed Y_j^{ψ}, we have

$$p(Y_i|Y_j^{\psi},\psi) = p(y_i(1,...,T)|y_i(-P+1,...,0), y_j^{\psi}(-P+1,...,T),\psi)$$

$$= \prod\nolimits_{q=1}^{Q} p(y_i(s_{q-1}+1,...,s_q)|y_i(-P+1,...,s_{q-1}), y_j^{\psi}(-P+1,...,T),\psi). \tag{21}$$

For the predictive density score of Y_i, we have[38]

$$g(Y_i,\{s_q\}_{q=0}^{Q}|\psi) = \prod\nolimits_{q=1}^{Q} p(y_i(s_{q-1}+1,...,s_q)|y_i(-P+1,...,s_{q-1}), y_j^{\psi}(-P+1,...,s_{q-1}),\psi). \tag{22}$$

For the statistic that we use, the marginal likelihood of Y implied by a model in the set Ω, we have

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

$$= \prod\nolimits_{q=1}^{Q} p(y_i(s_{q-1}+1,...,s_q), y_J(s_{q-1}+1,...,s_q)|y_i(-P+1,...,s_{q-1}), y_J(-P+1,...,s_{q-1}),\omega) \tag{23}$$

The following lessons emerge from comparing equations (20)-(22) with equation (23).

The marginal predictive density of Y_i, $p(Y_i|\psi)$, measures the out-of-sample fit to the data on y_i *assuming that no data on* y_j^{ψ} *have been observed except for the initial observations.* Note the term $y_j^{\psi}(-P+1,...,0)$ in expression (20). The fact that this statistic discards all available data on y_j^{ψ} except for the initial observations makes this statistic unattractive as a criterion for model choice. To see why this statistic is unattractive, consider the following fairly common case. Suppose that we want to compare a VAR model ψ of y_i and y_j^{ψ} with another VAR model $\tilde{\psi}$ of y_i and another set of variables, $y_j^{\tilde{\psi}}$. Assume $y_j^{\tilde{\psi}}$ has the same number of variables as y_j^{ψ}. Each VAR has one lag and the same prior, e.g. the Sims-Zha prior. Suppose that we rescale variables such that each variable in y_j^{ψ} and each variable in $y_j^{\tilde{\psi}}$ have the same value in period $t=0$.[39] Then it is straightforward to show that

[38]Expression (22) is a valid way to define the predictive density score, alternative to (19). Expressions (19) and (22) coincide when $h=1$ in (19) and $\{s_q\}_{q=0}^{Q} = \{0,1,...T\}$ in (22). We think that expression (22) makes more transparent the comparison between the predictive density score and the other statistics we consider here.

[39]For example, suppose that y_j^{ψ} consists of a single variable, $y_j^{\tilde{\psi}}$ consists of a single variable, $y_j^{\psi}(0)=5$, and $y_j^{\tilde{\psi}}(0)=10$. Then multiplication of y_j^{ψ} by 2 yields $y_j^{\psi}(0)=y_j^{\tilde{\psi}}(0)=10$, that is, y_j^{ψ} and $y_j^{\tilde{\psi}}$ have the same value in period $t=0$.

$p(Y_i|\psi) = p(Y_i|\tilde{\psi})$, that is, the marginal predictive density of Y_i implied by model ψ *is equal to* the marginal predictive density of Y_i implied by model $\tilde{\psi}$. The implication is strong. If we used $p(Y_i|\psi)$ to decide whether to include y_j^{ψ} or $y_j^{\tilde{\psi}}$ in the VAR with y_i, we would end up indifferent. Even if y_j^{ψ} were strongly related to y_i and $y_j^{\tilde{\psi}}$ followed an independent white noise process.[40]

The predictive density of Y_i conditional on the actually observed Y_j^{ψ}, $p(Y_i|Y_j^{\psi}, \psi)$, measures the fit to the data on y_i *assuming that data on* y_j^{ψ} *have been observed through the end of the sample, period* T. Note the term $y_j^{\psi}(-P+1, ..., T)$ in expression (21). Thus $p(Y_i|Y_j^{\psi}, \psi)$ is not a measure of out-of-sample fit. This statistic tells us how well a model ψ captures the relation between y_j^{ψ} and y_i, for a particular Y_j^{ψ}, namely for the actually observed Y_j^{ψ}. A model ψ could attain a high value of $p(Y_i|Y_j^{\psi}, \psi)$ while yielding poor out-of-sample fit to the data on y_j^{ψ} and, therefore, also poor out-of-sample fit to the data on y_i. This feature makes $p(Y_i|Y_j^{\psi}, \psi)$ unattractive as a criterion for model choice.

The predictive density score of Y_i, $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$, measures the out-of-sample fit to the data on y_i. By comparison, the predictive prior density of Y used in this paper, $p(Y|\omega)$, measures the out-of-sample fit to the data on y. Both statistics condition on all data available at the time when a density is being evaluated. Note the term $y_i(-P+1, ..., s_{q-1}), y_j^{\psi}(-P+1, ..., s_{q-1})$ in expression (22) and the term $y_i(-P+1, ..., s_{q-1}), y_J(-P+1, ..., s_{q-1})$ in expression (23). This feature is attractive and distinguishes both statistics from $p(Y_i|\psi)$ and $p(Y_i|Y_j^{\psi}, \psi)$.

However, as a criterion for model choice $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ has the following drawbacks compared with $p(Y|\omega)$.

First, computation of $p(Y|\omega)$ leads directly to computation of posterior odds on models. In contrast, one cannot assign probabilities to models based on $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$.

Second, different partitions $\{s_q\}_{q=0}^{Q}$ lead to the *same* value of $p(Y|\omega)$ for a given model. In contrast, different partitions $\{s_q\}_{q=0}^{Q}$ in general lead to *different* values of $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ for a given model.[41] The implication is that according to $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ the best model

[40]If one used a training sample prior in addition to the Sims-Zha prior, the marginal predictive densities of Y_i in this example would not be exactly equal to each other. In our application, we computed the marginal predictive density of Y_i implied by many VARs, with a training sample prior in addition to the Sims-Zha prior. We found that the differences between the marginal predictive densities of Y_i across different VARs were very small.

[41]For example, consider the partition $\{0, 1, 2, ...T\}$ that decomposes $p(Y|\omega)$ and $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ into one-step-ahead predictive densities; and consider the partition $\{0, 4, 8, ...T\}$ that decomposes $p(Y|\omega)$ and $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ into one-to-four-steps-ahead predictive densities. Both partitions yield the same value of

for forecasting one period ahead in general differs from the best model for forecasting one-to-four periods ahead. Thus model choice with this statistic requires an arbitrary weighting of forecast horizons.[42]

Third, computation of $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ requires looping over $q = 0, ...Q$ and, for each s_q, evaluating a predictive density of $y_i(s_{q-1}+1, ..., s_q)$. This computation can be time consuming, in particular when T is large or $s_q - s_{q-1}$ is large. In contrast, in the case of a single Granger-causal-priority restriction $p(Y|\omega)$ is available analytically and in the case of multiple Granger-causal-priority restrictions only a simple Monte Carlo is required to evaluate $p(Y|\omega)$. See Section 3 and Section 6. The computational burden of this Monte Carlo increases with N, whereas the computational burden of evaluating $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$ increases with T. Therefore, in large samples computation of $p(Y|\omega)$ is guaranteed to be cheaper than computation of $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$.[43]

8 Conclusions

We develop a Bayesian methodology to choose variables that belong in a VAR or a structural VAR with an a priori given set of variables of interest. We rely on the idea of Granger-causal-priority, related to the well-known concept of Granger-noncausality. Applying the methodology to the case when the variables of interest are GDP, the price level and the short-term interest rate, we find remarkably similar results for the euro area and the United States.

Our findings about which variables belong in a VAR with output, prices and short-term interest rates have implications for the development of DSGE models. This is so because DSGE models are VARs to an approximation. Our findings suggest that when researchers are interested in modeling the dynamics of output, prices and short-term interest rates, then incorporating in the DSGE model corporate bond spreads, inventories, survey-based measures of confidence, the unemployment rate and the oil price has a higher potential for improving the model than incorporating house prices and monetary aggregates.

$p(Y|\omega)$ but, in general, different values of $g(Y_i, \{s_q\}_{q=0}^{Q}|\psi)$.

[42]In practice, researchers compute predictive density scores not only for partitions like $\{s_q\}_{q=0}^{Q}$, but also for different horizons, like in equation (19) when $h > 1$. Each horizon h leads to a different value of $g(Y_i, h \mid \psi)$.

[43]Computation of $p(Y_i|\psi)$ and $p(Y_i|Y_j^{\psi}, \psi)$ requires evaluating the marginal likelihood of an unobserved components model. This computation is demanding when there are many unobservable state variables. Furthermore, the computational burden of evaluating $p(Y_i|\psi)$ and $p(Y_i|Y_j^{\psi}, \psi)$ increases with T.

A Computational details

This appendix explains implementation and convergence diagnostics for the Monte Carlo algorithms that we use in sections 5 and 6.

A.1 MC^3

In the empirical applications in sections 5 and 6 we work with sets of models that are too large to evaluate all models in these sets - it would take too much time. Therefore, we resort to the Monte Carlo technique designed especially for approximating results conditional on a large set of models: Markov chain Monte Carlo model composition (MC^3) of Madigan and York (1995). The MC^3 algorithm generates a stochastic process that moves through the set of models and visits models with the probability equal to their posterior probability. Then the posterior results of interest are approximated based on the visited sample of models. By construction, the visited sample contains many models with high posterior probabilities and few models with low posterior probabilities. Working with such a sample is efficient because models with low posterior probabilities contribute little to the posterior probabilities of subsets of Ω that we compute in section 5 and they are irrelevant when we look for the best models in $\tilde{\Omega}$ in section 6. Of course, we need to monitor convergence to ensure that the MC^3 sample of models is large enough to approximate the posterior results conditional on the whole set of models with desired accuracy.

The MC^3 algorithm works as follows. First, for each model ω we define a set of models called the *neighborhood* of this model and denoted $nbr(\omega)$. The specific definitions of the neighborhoods that we use follow in the next subsections. Second, suppose the chain is at a model ω. We attach equal probability to each model in $nbr(\omega)$ and randomly draw a candidate model ω' from $nbr(\omega)$. The chain moves 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 in model ω, i.e. we record another occurrence of state ω. The process continues until the chain has the desired length.

A.2 Probabilities of Granger-causal-priority in Section 5

For each specification (i.e. for each column in tables 2, 9, 10, 7 and 8) we run two chains of one million draws each, starting from maximally dispersed starting points. We estimate GCP probabilities from each chain and in the tables we report the average of the two estimates. We establish convergence by veryfying that the estimates from independent chains do not differ significantly. A typical chain of one million draws takes about 1 hour on a standard personal computer.

Definition of the neighborhood in the set Ω: the neighborhood of a model ω is the set of all models that differ from ω by the position of one variable: models where one variable from the first block of ω is in the second block and models where one variable from the second block of ω is in the first block. Obviously, for each model $\omega \in \Omega$ we have $\#nbr(\omega) = N_J$.

Starting points for the two chains. The first chain starts with the unrestricted model. In this model the first block consists of all 41 variables, $y_1 = y$. The second chain starts with the model where y_i is Granger-causally-prior to all other variables. In this model the first block is the smallest possible - it consists of the 3 variables of interest $y_1 = y_i$, and the remaining 38 variables are in the second block, $y_2 = y_J$. These starting points are maximally dispersed because the number of moves of the chain required to go from one starting point to the other is larger than for any other two models in Ω. We also experimented with random starting points and the results were unaffected.

Estimator of GCP probability. In each chain we estimate the probability that y_i is Granger-causally-prior to y_j, $p(\Omega^j|\Omega)$, for each $j = 1...N_J$, using as the estimator the frequency of visits of the chain in Ω^j. For this computation we discard the first half of the chain (the first 500,000 states of the chain) to ensure that our results do not depend on the starting point. That is, we estimate the GCP probability as $\hat{p}(\Omega^j|\Omega) = \left(\sum_{m=500{,}001}^{1000{,}000} \theta_m^j\right)/500,000$, where θ_m^j is the value of the indicator function taking the value of 1 when the m-th model in the chain belongs to Ω^j and 0 otherwise, $\theta_m^j \equiv I(\omega_m \in \Omega^j)$.

Numerical standard error. We compute numerical standard deviations of $\hat{p}(\Omega^j|\Omega)$ using the Newey-West estimator that accounts for the autocorrelation of θ_m^j up to order 500. Most autocorrelations go to zero long before 500, but for individual variables j the autocorrelation of the order 500 is still about 0.1. The Newey-West standard errors are on average about half a percentage point, 10 times larger than standard errors assuming independent θ_m^j, but for some variables they are slightly above 1 percentage point.

We conduct formal convergence diagnostics following the approach of e.g. Geweke (2005): we test whether GCP probabilities differ significantly across chains. The joint chi-squared test for the equality the two vectors of GCP probabilities from the two chains is never rejected at 5% significance. The lowest p-values are 0.09 in the euro area and 0.06 in the U.S., both in the case of the standard Minnesota prior. We also test separately the equality of GCP probabilities of individual variables. The overwhelming majority of test statistics, which have asymptotic standard normal distribution, are below 2. However, in some individual cases they exceed 2 and sometimes even 3, signalling a significant discrepancy. This happens for the tight prior and two lags specification in the euro area and in the baseline and tight prior specification in the U.S. The differences flagged as significant are not large from the point of view of our discussion: e.g. in the U.S. baseline simulation the GCP probability of M1 is 0.540 in one chain and 0.569 in another, leading to the test statistic of 3.5. In these cases we run more simulations to convince ourselves that we keep getting similar numbers for the GCP probabilities in question.

For the baseline specifications in the euro area and in the U.S. we explore convergence further and we run 20 chains of length 1 million, starting from random starting points. The standard deviations of GCP probabilities across chains are very similar to the Newey-West standard deviations inferred from individual chains (their correlation is 0.94). We make two observations about these standard deviations. First, on average across variables the GCP standard deviations in 20 chains are below 1 percentage point: the average is 0.4 percentage points in the euro area and 0.6 percentage points in the U.S. Second, the standard deviations tend to be larger for GCP probabilities in the vicinity of 0.5, and smaller for GCP probabilities close to 0 or 1, which are of interest in this paper.

In the euro area the largest standard deviations are found for the Mortgage Interest Rate (standard deviation 0.6ppt, GCP probability 0.45), Unit Labor Cost (also 0.6ppt, 0.36), Total Employment (0.6ppt, 0.44) and the Producer Price Index (0.6ppt, 0.50). Clearly, in most of these cases the GCP probabilities are in the vicinity of 0.5. The smallest standard deviations are for the Dollar-Euro Exchange Rate (0.04ppt, 0.998), Industrial confidence (0.1ppt, 0.003) and Residential Property Prices (0.1ppt, 0.99). In all these cases the GCP probabilities are near 0 or 1. Across all 38 variables the correlation between the squared deviation of the GCP probabilities from 0.5 and their standard deviations is -0.84.

In the U.S. case there are 5 standard deviations that exceed 1 percentage point: for the Eonia (standard deviation 1.4ppt, GCP probability 0.66), 10 Year Bond Yield (1.3ppt, 0.48), 2 Year Bond Yield (1ppt, 0.47), Mortgage Interest Rate (1ppt, 0.50) and Consumer Prices Excl. Energy, Food (1ppt, 0.78). Clearly, in most cases the GCP probabilities are in the vicinity of 0.5. The smallest standard deviations are for the BBB Bond Spread (0.1ppt, GCP probability 0.01) and Residential Property Prices (0.2ppt, 0.98). Again, in all these cases the GCP probabilities are near 0 or 1. Across all 38 variables the correlation between the squared deviation of the GCP probabilities from 0.5 and their standard deviations across chains is -0.6. Therefore, if we wanted to draw conclusions about GCP probabilities in the vicinity of 0.5 we might need longer simulations, especially in the U.S. However, the simulation lengths that we use are appropriate for drawing robust lessons about GCP probabilities near 0 or 1.

Overall, we conclude that the lessons we draw in Section 5 are robust to the Monte Carlo error of our procedures.

A.3 Searching for the best models in the set $\tilde{\Omega}$ in Section 6

In the extended set of models $\tilde{\Omega}$ we need to also extend the definition of the neighborhood. We maintain the rule that the *neighborhood* of ω is the set the set of all models that differ from the model ω by the position of only one variable in the pattern of GCP restrictions. However, the difference is that with multiple GCP restrictions the position of a variable can differ in one of four possible ways: (i) the variable may join the previous block, (ii) the variable may joint the next block, (iii) the variable may become a block on its own prior to its current block, (iv) the variable may become a block on its own posterior to its current block.

We began by examining the properties of the Monte Carlo procedure for evaluating the Bayes factor in favor of multiple GCP restrictions, described in Section 6. We verified that when a model has only one GCP restriction – so that we know the marginal likelihood analytically – the Monte Carlo converges to the true value as the number of draws increases. Next, we repeatedly evaluate Bayes factors in favor of 20 randomly selected models with multiple GCP restrictions. With $M = 1000$ draws we recovered the log marginal likelihoods with the mean absolute deviation of 0.6. With $M = 10$ draws we recovered the log marginal likelihood with the mean absolute deviation of 1.9, which is a significant difference, but with

$M = 10$ we can run much longer chains in reasonable time. As a compromise, we proceed in two steps. We first run an MC^3 chain with one million draws using $M = 10$. We then collect the best 100,000 models visited by this chain and we recompute the marginal likelihood implied by each of those models using $M = 1000$.

To assess robustness of the results we repeat this procedure 7 times with different random starting point for the Markov chain and we compare the results. The lessons we draw about the optimal composition of the first block are the same across 7 chains, so we conclude that the procedure is reliable enough for studying the composition of the first block. However, we would not recommend to use this procedure to draw lessons about precise composition of all the other blocks, because the set $\tilde{\Omega}$ is extremely large and contains many models that are very similar to each other. For example, consider two models with several blocks. Moving one variable from the last block to the last-but-one block typically produces only a negligible change in the marginal likelihood.

In our empirical exercise with $N_J = 14$ the search for the best model in the set $\tilde{\Omega}$ takes four days and a half: about 12 hours for the chain and about 100 hours for recomputing the Bayes factors for the best 100,000 models. We run the seven procedures in paralell on seven different processors. Therefore such exercises are feasible today so long as the number of remaining variables in the dataset, N_J, is not much larger than in our study and will become more and more reliable in larger datasets as computers improve.

B Sims-Zha prior

The prior used in this paper consists of two components: (i) an initial prior formulated before seeing any data, and (ii) a training sample prior. See Sections 3.1 and 5.2. This appendix gives the details concerning the initial prior. See Section 5.2 concerning the training sample prior.

The initial prior follows Sims and Zha (1998). This prior depends on scalar hyperparameters λ_1, λ_3, λ_4, μ_5, μ_6 whose values we discuss below. The notation for hyperparameters follows Sims and Zha (1998).[44] The initial prior consists of the following four components.

[44] Historically there was also a hyperparameter λ_2 but this hyperparameter is always equal to 1 in the conjugate framework, see e.g. Sims and Zha (1998).

The first component of the initial prior is the Minnesota prior for B, given by

$$p(\mathrm{vec}\, B|\Sigma) = \mathcal{N}\left(\mathrm{vec}\begin{pmatrix} I_N \\ 0_{K-N\times N}\end{pmatrix}, \Sigma \otimes WW'\right), \tag{24}$$

where W is a diagonal matrix of size $K \times K$ such that the diagonal entry corresponding to variable n and lag p equals $\lambda_1/(\sigma_n p^{\lambda_3})$ and the last diagonal entry, corresponding to the constant term, equals λ_4. As is standard practice, we set σ_n, $n = 1, ...N$, equal to the standard deviation of residuals from the univariate autoregressive model with P lags fit by OLS to the n-th series in the VAR.

The Minnesota prior is implemented with K dummy observations given in the matrices

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

The second component of the initial prior is the no-cointegration prior. The no-cointegration prior is implemented with N dummy observations given in the matrices

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

where we set $\bar{y} = (1/P)\sum_{t=0}^{P-1} y_{-t}$, the average of initial values of y.

The third component of the initial prior is the one-unit-root prior. The one-unit-root prior is implemented 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 of the initial prior is the prior for Σ, given by

$$p(\Sigma) = \mathcal{IW}(ZZ', \nu_0), \tag{25}$$

where $\mathcal{IW}$ denotes the Inverted Wishart density, Z is an $N \times N$ matrix and ν_0 is a scalar hyperparameter. We set $Z = \sqrt{\nu_0 - N - 1}\,\mathrm{diag}(\sigma)$, where $\sigma = (\sigma_1, ..., \sigma_N)$. This choice of Z implies that the mean of this prior is

$$E(\Sigma) = \frac{ZZ'}{\nu_0 - N - 1} = \mathrm{diag}(\sigma^2).$$

We set $\nu_0 = K + N$. The reason for this choice for the value of ν_0 is as follows. The inverted Wishart density is proper when $\nu_0 > N - 1$. The Sims-Zha prior is proper when $\nu_0 > K + N - 1$, because K degrees of freedom are "used up" by the normal density of B. Therefore, as a rule of thumb we use the next integer after $K + N - 1$ setting $\nu_0 = K + N$.

Note that prior (25) satisfies

$$\begin{aligned} p(\Sigma) &\propto |\Sigma|^{-(\nu_0+N+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(ZZ'\Sigma^{-1}\right)\right) \\ &= |\Sigma|^{-(\nu_0+1)/2}|\Sigma|^{-N/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(Z' - 0B\right)'\left(Z' - 0B\right)\Sigma^{-1}\right). \end{aligned}$$

The above expression 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 $|\Sigma|^{-(\nu_0+1)/2}$. Therefore, we implement the prior for 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

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

The matrices Y_{SZ} and X_{SZ} appear in expression (15).

In the main text we use the following names of the hyperparameters: λ_1 - overall tightness, μ_5 - weight of the no-cointegration dummy observations, μ_6 - weight of the one-unit-root dummy observation. The remaining parameters, λ_3 and λ_4 are not discussed in the main text.

We choose hyperparameter values and different values of P by maximizing the marginal likelihood of the data implied by the unrestricted VAR on the training sample 1989Q1-1998Q4. We used a grid of values for each of the hyperparameters and we used a grid of values for the number of lags. We found that $\lambda_1 = 0.05$, $\lambda_3 = 1$, $\lambda_4 = 1$, $\mu_5 = 2$, $\mu_6 = 1$ and $P = 1$ yield the highest value of the marginal likelihood. The values of λ_3 and λ_4 are the same as in Sims and Zha (1998). The value of λ_1 is much smaller than of the value of 0.2

used by Sims and Zha (1998) and the value of μ_5 is larger than the value of 1 used by Sims and Zha (1998). Both lower λ_1 and higher μ_5 imply that the prior is tighter. Our VAR has 41 variables, compared with 6 variables in Sims and Zha (1998), so our choice of a tighter prior is consistent with the suggestion of Giannone et al. (2012) to use tighter priors as the number of variables in the VAR increases.

C Robustness of the results in Table 2

We examine how the results reported in Table 2 change as we vary the values of the hyperparameters, add lags, and repeat the analysis in subsamples.

To begin, we compare the results obtained with the baseline prior defined in Section 5.2 with the results obtained with, alternatively, the standard Sims-Zha prior (which is looser than the baseline prior) and a tighter version of the Sims-Zha prior. In the standard Sims-Zha prior, we use hyperparameter values typically employed in the literature on medium-sized VARs following Sims and Zha (1998): the "overall tightness" of 0.1, the weight of the "one-unit-root" dummy of 1, and the weight of the "no-cointegration" dummies of 1. In the tighter prior, we specify the "overall tightness" of 0.01, the weight of the "one-unit-root" dummy of 1, and the weight of the "no-cointegration" dummies of 3.

The results with the standard prior are very similar to the results with the baseline prior. See Table 7 for the euro area exercise and Table 8 for the U.S. exercise. The findings that we emphasized in the discussion of Table 2 remain unchanged, with two exceptions: (i) in the euro area exercise, the posterior probability with respect to the price of oil rises, and (ii) in the U.S. exercise, the posterior probability with respect to the unemployment rate rises.[45] Furthermore, the correlation between the posterior probabilities with the baseline prior and the posterior probabilities with the standard prior is high: 0.9 in the euro area exercise and 0.89 in the U.S. exercise.

The results with the tighter prior show some notable differences compared with the results with the baseline prior. Again, see Tables 7-8. However, these differences do not worry us because the models with the tighter prior fit the data very poorly. We compute the marginal likelihood of the data implied by the set of models Ω, $p(y|\Omega)$, with a given

[45] In the euro area exercise, the posterior probability with respect to consumer confidence increases. However, it remains true that two survey-based leading indicators are associated with a posterior probability smaller than 0.1: industrial confidence (as with the baseline prior) and PMI.

prior. We find that the marginal likelihood with the tighter prior is extremely low.

We also study the effects of dropping the training sample prior. The marginal likelihood falls significantly and therefore we pay little attention to the results without the training sample prior.

Table 7: Posterior probability that output, prices and short-term interest rates are Granger-causally prior to a variable: sensitivity to the prior, euro area.

	baseline	standard	tight	2 lags
Industrial Confidence	0.00	0.01	0.06	0.00
Consumer Confidence	0.03	0.56	0.18	0.24
Unemployment Rate	0.03	0.03	0.07	0.34
Change of Real Inventories	0.06	0.01	0.46	0.04
BBB Bond Spread 7-10 Years	0.10	0.09	0.78	0.21
Oil Price	0.11	0.58	0.45	0.05
Real Investment	0.15	0.10	0.14	0.05
PMI	0.17	0.05	0.33	0.09
U.S. Federal Funds Rate	0.17	0.24	0.35	0.62
Real Exports	0.18	0.13	0.17	0.06
Real Imports	0.20	0.15	0.17	0.07
Lending Rate to NFC	0.24	0.08	0.12	0.05
Real Consumption	0.32	0.25	0.33	0.15
Unit Labor Cost	0.36	0.29	0.35	0.14
Real Housing Investment	0.37	0.31	0.32	0.26
2 Year Bond Yield	0.39	0.07	0.31	0.08
Bank Credit to NFC Outstanding	0.44	0.51	0.34	0.62
Total Employment	0.44	0.45	0.31	0.63
Mortgage Interest Rate	0.45	0.24	0.27	0.19
Producer Price Index (PPI)	0.50	0.49	0.54	0.48
Real Government Consumption	0.53	0.49	0.53	0.50
U.S. CPI	0.56	0.56	0.60	0.34
Commodity Prices	0.61	0.66	0.67	0.36
Capacity Utilization	0.64	0.52	0.64	0.85
U.S. Real GDP	0.69	0.69	0.75	0.58
Real Estate Loans Outstanding	0.69	0.76	0.55	0.67
Consumer Prices Excl. Energy, Food	0.73	0.90	0.90	0.89
10 Year Bond Yield	0.73	0.65	0.67	0.35
Stock Market Index*	0.82	0.88	0.74	0.81
M1	0.86	0.83	0.88	0.92
Stock Market Volatility Index**	0.89	0.92	0.93	0.85
M3 (euro area only)	0.93	0.86	0.85	1.00
Nominal Effective Exchange Rate	0.94	0.93	0.81	0.95
M2	0.94	0.89	0.91	1.00
Consumer Credit Outstanding	0.94	0.93	0.96	0.93
Government Debt	0.99	0.99	0.96	0.98
Residental Property Prices	0.99	0.99	0.90	1.00
Dollar-Euro Exchange Rate	1.00	1.00	0.77	1.00
correlation with the baseline		0.90	0.86	0.88

Next, we consider the effects of adding lags. Tables 7-8 show the results obtained with two lags, $P = 2$, and the baseline prior. We find that the marginal likelihood of the data implied by models with two lags is much lower than the analogous statistic implied by models with a single lag. Since the models with more than a single lag fit the data very

Table 8: Posterior probability that output, prices and short-term interest rates are Granger-causally prior to a variable: sensitivity to the prior, U.S.

	baseline	standard	tight	2 lags
BBB Bond Spread 7-10 Years	0.01	0.01	0.75	0.01
PMI	0.06	0.06	0.39	0.05
Change of Real Inventories	0.10	0.13	0.58	0.06
Oil Price	0.11	0.01	0.23	0.06
Industrial Confidence	0.11	0.05	0.54	0.09
Unemployment Rate	0.14	0.48	0.22	0.20
Capacity Utilization	0.17	0.23	0.20	0.32
Real Investment	0.21	0.23	0.27	0.25
Lending Rate to NFC	0.22	0.25	0.08	0.12
Consumer Confidence	0.22	0.31	0.34	0.11
Hours Worked (U.S. only)	0.32	0.38	0.36	0.29
Stock Market Volatility Index**	0.36	0.45	0.76	0.30
Real Imports	0.37	0.40	0.40	0.28
Total Employment	0.38	0.35	0.51	0.31
Producer Price Index (PPI)	0.42	0.32	0.40	0.23
10 Year Bond Yield	0.42	0.20	0.12	0.17
2 Year Bond Yield	0.42	0.12	0.11	0.15
Mortgage Interest Rate	0.44	0.32	0.17	0.17
Nominal Effective Exchange Rate	0.45	0.51	0.31	0.43
Commodity Prices	0.45	0.54	0.48	0.31
Real Consumption	0.46	0.48	0.43	0.32
Euro Area Real GDP	0.48	0.32	0.30	0.67
M1	0.56	0.61	0.62	0.44
Real Exports	0.59	0.68	0.37	0.54
HICP (Consumer Prices in Europe)	0.59	0.52	0.42	0.37
Eonia (Euro Overnight Interbank Rate)	0.60	0.11	0.22	0.87
Stock Market Index*	0.64	0.73	0.67	0.52
Consumer Prices Excl. Energy, Food	0.75	0.50	0.83	0.86
Bank Credit to NFC Outstanding	0.77	0.85	0.90	0.85
Unit Labor Cost	0.78	0.76	0.85	0.69
Dollar-Euro Exchange Rate	0.79	1.00	0.34	0.95
Government Debt	0.82	0.91	0.93	0.86
Consumer Credit Outstanding	0.82	0.94	0.85	0.98
Real Government Consumption	0.85	0.88	0.76	0.85
Real Housing Investment	0.86	0.85	0.73	0.89
M2	0.92	0.85	0.84	0.92
Real Estate Loans Outstanding	0.93	0.97	0.95	0.98
Residental Property Prices	0.98	1.00	0.97	1.00
correlation with the baseline		0.89	0.62	0.93

poorly, we pay little attention to the results obtained with more than a single lag.[46]

Finally, we repeat the analysis in three subsamples. First, we drop the last four quarters, i.e., the sample ends in 2010Q2. Second, we drop the last eight quarters, i.e., the sample ends in 2009Q2. Third, we drop the last twelve quarters, i.e., the sample ends in 2008Q2. Omitting the last eight quarters amounts to dropping as much as one-sixth of the sample. Omitting the last twelve quarters amounts to dropping as much as one-fourth of sample and the entire crisis period. The results are in Table 9 for the euro area exercise and in Table 10 for the U.S. exercise. The messages of Tables 9-10 are as follows. When we drop four and eight quarters, the results change hardly at all. In particular, the correlation between the posterior probabilities in the full sample and the posterior probabilities in the shorter samples is very close to one. When we drop the entire crisis period, the results change somewhat. In particular, the posterior probability of Granger-causal-priority with respect to the corporate bond spread rises notably in the euro area exercise (but *not* in the U.S. exercise). The same is true in the case of the change in inventories in the euro area exercise (but not in the U.S. exercise) and in the case of the price of oil in the U.S. exercise (but not in the euro area exercise). The correlation between the posterior probabilities in the full sample and the posterior probabilities in the sample excluding the entire crisis period is fairly high: 0.91 in the euro area exercise and 0.85 in the U.S. exercise.

[46] We condition on the lag length, i.e., we condition on the value of P. In Section 5.3 we report the main findings based on $P = 1$ and in this appendix we verify robustness of the main findings by setting $P = 2$. An alternative approach would be to redefine Ω so that it includes models with different lag lengths and report posterior probabilities of Granger-causal-priority having taken into account the uncertainty about the lag length. We know that this alternative approach would yield very similar conclusions, because we know that the models with more than a single lag fit the data very poorly. However, the same conclusion need not hold in other empirical applications.

Table 9: Posterior probability that output, prices and short-term interest rates are Granger-causally prior to a variable: results for subsamples, euro area.

	baseline	drop 4Q	drop 8Q	drop 12Q
Industrial Confidence	0.00	0.01	0.01	0.01
Consumer Confidence	0.03	0.03	0.03	0.10
Unemployment Rate	0.03	0.02	0.03	0.05
Change of Real Inventories	0.06	0.07	0.14	0.23
BBB Bond Spread 7-10 Years	0.10	0.12	0.06	0.53
Oil Price	0.11	0.17	0.11	0.16
Real Investment	0.15	0.16	0.16	0.23
PMI	0.17	0.14	0.18	0.31
U.S. Federal Funds Rate	0.17	0.33	0.35	0.15
Real Exports	0.18	0.19	0.20	0.24
Real Imports	0.20	0.21	0.19	0.28
Lending Rate to NFC	0.24	0.29	0.39	0.32
Real Consumption	0.32	0.32	0.35	0.31
Unit Labor Cost	0.36	0.34	0.41	0.49
Real Housing Investment	0.37	0.37	0.34	0.25
2 Year Bond Yield	0.39	0.47	0.55	0.46
Bank Credit to NFC Outstanding	0.44	0.42	0.49	0.48
Total Employment	0.44	0.42	0.29	0.30
Mortgage Interest Rate	0.45	0.53	0.63	0.41
Producer Price Index (PPI)	0.50	0.51	0.49	0.82
Real Government Consumption	0.53	0.49	0.47	0.49
U.S. CPI	0.56	0.56	0.48	0.67
Commodity Prices	0.61	0.63	0.61	0.87
Capacity Utilization	0.64	0.61	0.59	0.48
U.S. Real GDP	0.69	0.68	0.65	0.68
Real Estate Loans Outstanding	0.69	0.65	0.44	0.41
Consumer Prices Excl. Energy, Food	0.73	0.63	0.62	0.61
10 Year Bond Yield	0.73	0.75	0.85	0.66
Stock Market Index*	0.82	0.80	0.80	0.90
M1	0.86	0.78	0.80	0.80
Stock Market Volatility Index**	0.89	0.87	0.86	0.89
M3 (euro area only)	0.93	0.89	0.93	0.98
Nominal Effective Exchange Rate	0.94	0.93	0.90	0.85
M2	0.94	0.92	0.96	0.99
Consumer Credit Outstanding	0.94	0.93	0.92	0.89
Government Debt	0.99	0.97	0.88	0.77
Residental Property Prices	0.99	0.97	0.96	1.00
Dollar-Euro Exchange Rate	1.00	1.00	1.00	1.00
correlation with the baseline		0.99	0.97	0.91

Table 10: Posterior probability that output, prices and short-term interest rates are Granger-causally prior to a variable: results for subsamples, U.S.

	baseline	drop 4Q	drop 8Q	drop 12Q
BBB Bond Spread 7-10 Years	0.01	0.01	0.01	0.05
PMI	0.06	0.05	0.07	0.07
Change of Real Inventories	0.10	0.12	0.14	0.10
Oil Price	0.11	0.12	0.11	0.57
Industrial Confidence	0.11	0.13	0.14	0.26
Unemployment Rate	0.14	0.07	0.13	0.05
Capacity Utilization	0.17	0.18	0.18	0.18
Real Investment	0.21	0.22	0.22	0.33
Lending Rate to NFC	0.22	0.26	0.29	0.16
Consumer Confidence	0.22	0.25	0.22	0.13
Hours Worked (U.S. only)	0.32	0.30	0.30	0.36
Stock Market Volatility Index**	0.36	0.37	0.39	0.41
Real Imports	0.37	0.39	0.37	0.49
Total Employment	0.38	0.39	0.33	0.41
Producer Price Index (PPI)	0.42	0.44	0.40	0.72
10 Year Bond Yield	0.42	0.40	0.46	0.55
2 Year Bond Yield	0.42	0.37	0.43	0.44
Mortgage Interest Rate	0.44	0.44	0.51	0.50
Nominal Effective Exchange Rate	0.45	0.51	0.48	0.37
Commodity Prices	0.45	0.48	0.47	0.82
Real Consumption	0.46	0.52	0.53	0.36
Euro Area Real GDP	0.48	0.44	0.38	0.40
M1	0.56	0.56	0.47	0.58
Real Exports	0.59	0.61	0.57	0.66
HICP (Consumer Prices in Europe)	0.59	0.63	0.55	0.20
Eonia (Euro Overnight Interbank Rate)	0.60	0.46	0.55	0.94
Stock Market Index*	0.64	0.64	0.58	0.70
Consumer Prices Excl. Energy, Food	0.75	0.66	0.73	0.95
Bank Credit to NFC Outstanding	0.77	0.74	0.79	0.92
Unit Labor Cost	0.78	0.79	0.77	0.68
Dollar-Euro Exchange Rate	0.79	0.81	0.80	0.57
Government Debt	0.82	0.77	0.77	0.81
Consumer Credit Outstanding	0.82	0.82	0.75	0.72
Real Government Consumption	0.85	0.85	0.83	0.75
Real Housing Investment	0.86	0.83	0.84	0.76
M2	0.92	0.86	0.88	0.98
Real Estate Loans Outstanding	0.93	0.92	0.95	0.98
Residental Property Prices	0.98	0.98	0.99	0.99
correlation with the baseline		0.99	0.99	0.84

References

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.

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

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

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

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

Forni, M. and Gambetti, L. (2011). Testing for sufficient information in structural VARs. CEPR Discussion Papers 8209, C.E.P.R. Discussion Papers.

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.

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. CEPR Discussion Papers 8755, C.E.P.R. Discussion Papers.

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*, chapter 4, pages 77–119. Boulder: Westview.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

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

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

OEIS (2011). On-line encyclopedia of integer sequences: A000670. Available at http://oeis.org.

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

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

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