

# PRIORS FOR THE LONG RUN

DOMENICO GIANNONE, MICHELE LENZA, AND GIORGIO E. PRIMICERI

ABSTRACT. We propose a class of prior distributions that discipline the long-run behavior of Vector Autoregressions (VARs). These priors can be naturally elicited using economic theory, which provides guidance on the joint dynamics of macroeconomic time series in the long run. Our priors for the long run are conjugate, and can thus be trivially implemented using dummy observations and combined with other popular priors. In VARs with standard macroeconomic variables, a prior based on the long-run predictions of a wide class of dynamic stochastic general equilibrium models yields substantial improvements in the forecasting performance.

## 1. INTRODUCTION

In this paper we propose a class of prior distributions that disciplines the long-run behavior of economic time series in Vector Autoregressions (VARs). Bayesian inference with informative priors has a long tradition for VARs, given that these models typically include many free parameters to accommodate general forms of autocorrelations and cross-correlations among variables. Therefore, with flat priors, such flexibility is likely to lead to in-sample overfitting, at the cost of poor out-of-sample forecasting accuracy.

Our prior is motivated by a specific form of overfitting of flat-prior VARs, which is their tendency to attribute an implausibly large share of the variation in observed time series to a deterministic—and thus entirely predictable—component (Sims, 1996 and 2000). In these models, inference is conducted by taking the initial observations of the variables as non-random. Therefore, the likelihood does not penalize parameter values implying that the variables' steady state (for stationary series, their trend for non-stationary ones) is distant from their initial observations. Complex transient dynamics from these initial conditions to the steady state are thus implicitly regarded as reasonable. As a consequence, they end up

---

*Date:* December 2014.

We thank seminar participants at the European Central Bank, ECARES and the 25th (EC)<sup>2</sup> conference for comments and suggestions. The views expressed in this paper are those of the authors and are not necessarily reflective of views at the European Central Bank, the Eurosystem, the Federal Reserve Bank of New York, or the Federal Reserve System.

explaining an implausibly large share of the low frequency variation of the data, yielding inaccurate out-of-sample forecasts.

Modifying inference to explicitly incorporate the density of the initial conditions may not be the right solution to this problem, since most macroeconomic time series are non-stationary and it is not obvious how to choose the distribution of their initial observations. An alternative route is to formulate a prior that expresses disbelief in an excessive explanatory power of the deterministic component of the model, by specifying that initial conditions should not be an important predictor of the subsequent evolution of the series.

There are a variety of specific ways to implement this idea. Our main insight is that economic theory should play a central role for the elicitation of such a prior. Consider, for example, a simple bivariate VAR with (log-)GDP and (log-)investment. A wide class of theoretical models in macroeconomics predicts that these two variables should share a common stochastic trend, while the (log-)investment-to-GDP ratio should be stationary. As a consequence, we might want to formulate a prior according to which initial conditions should explain very little of the dynamics of the common stochastic component, but might have a higher predicting power for the evolution of the investment-to-GDP ratio. In fact, if this variable is really mean reverting, then it is reasonable that initial conditions should shape the low frequency dynamics in the early part of the sample, while the variable converges back to its equilibrium value.

Our prior for the long run (PLR) is a formalization of this general concept. Its key ingredient is the choice of two orthogonal vector spaces, corresponding to the set of linear combinations of the model variables that are a-priori likely to be stationary and non-stationary, i.e. the a-priori cointegrating space and the space orthogonal to it. It is exactly for the identification of these two orthogonal spaces that economic theory plays a crucial role. The PLR essentially consists of shrinking the VAR coefficients towards values that imply little predictive power of initial conditions for the subsequent behavior of all linear combinations of the variables in the system, but particularly so for those that are likely to be non-stationary. In this respect, our prior builds on the original insight of cointegration. However, our fully probabilistic approach does not involve the pre-testing and hard restrictions that typically plague error correction models.

While we postpone the detailed description of our proposal to the main body of the paper, here we stress that our PLR is conjugate, and can thus be trivially implemented using

dummy observations and combined with existing popular priors, such as the Minnesota prior (Litterman, 1979). Moreover, conjugacy allows the closed-form computation of the marginal likelihood, which can be used to select the tightness of our PLR.

We apply these ideas to the estimation of two VARs with standard macroeconomic variables. The first is a small-scale model with real variables such as output, consumption and investment. The second, larger-scale VAR includes also some nominal variables, such as inflation and the short-term interest rate. In both cases, we set up our PLR based on the robust lessons of a wide class of dynamic stochastic general equilibrium (DSGE) models. Roughly speaking these theories typically predict the existence of a common stochastic trend for the real variables, and possibly another trend for the nominal variables, while the ratios are likely to be stationary. We show that a PLR set up in accordance with these theoretical predictions is successful in reducing the spurious explanatory power of the deterministic component implied by flat-prior VARs. The PLR also yields substantial improvements over more traditional BVARs in terms of out-of-sample forecasting performance, especially at long horizons.

The rest of the paper is organized as follows. Section 2 explains in what sense flat-prior VARs attribute too much explanatory power to initial conditions and deterministic trends. Section 3 illustrates our approach to solve this problem, i.e. our PLR. Section 4 puts our contribution in the context of a vast related literature, which is easier to do after having discussed the details of our procedure. Section 5 describes the results of our empirical application, and section 6 concludes by summarizing the advantages and potential limitations of our approach.

## 2. INITIAL CONDITIONS AND DETERMINISTIC TRENDS

In this section, we show that flat-prior VARs tend to attribute an implausibly large share of the variation in observed time series to a deterministic—and thus entirely predictable—component. This problem motivates the specific prior distribution proposed in this paper. Most of the discussion in this section is based on the work of Sims (1996 and 2000), although our recipe to address this pathology differs from his, as we will see in section 4.

To illustrate the problem, let us begin by considering the simple example of an AR(1) model,

$$(2.1) \quad y_t = c + \rho y_{t-1} + \varepsilon_t.$$

Equation (2.1) can be iterated backward to obtain

$$(2.2) \quad y_t = \underbrace{\rho^{t-1}y_1 + \sum_{j=0}^{t-2} \rho^j c}_{DC_t} + \underbrace{\sum_{j=0}^{t-2} \rho^j \varepsilon_{t-j}}_{SC_t},$$

which shows that the model separates the observed variation of the data into two parts. The first component of (2.2)—denoted by  $DC_t$ —represents the counterfactual evolution of  $y_t$  in absence of shocks, starting from the initial observation  $y_1$ . Given that AR and VAR models are typically estimated treating the initial observation as given and non-random,  $DC_t$  corresponds to the deterministic component of  $y_t$ . The second component of (2.2)—denoted by  $SC_t$ —depends instead on the realization of all the shocks between time 2 and  $t$ , and thus corresponds to the unpredictable or stochastic component of  $y_t$ .

To analyze the properties of the deterministic component of  $y_t$ , it is useful to solve the summation in (2.2) and rewrite  $DC_t$  as

$$DC_t = \begin{cases} y_1 + (t-1)c & \text{if } \rho = 1 \\ \frac{c}{1-\rho} + \rho^{t-1} \left( y_1 - \frac{c}{1-\rho} \right) & \text{if } \rho \neq 1 \end{cases}.$$

If  $\rho = 1$ , the deterministic component is a simple linear trend. If instead  $\rho \neq 1$ ,  $DC_t$  is an exponential, and has a potentially more complex shape as a function of time. The problem is that, when conducting inference, this potentially complex deterministic dynamics arising from estimates of  $\rho \neq 1$  can be exploited to fit the low frequency variation of  $y_t$ , even when such variation is mostly stochastic. This peculiar “overfitting” behavior of the deterministic component is clearly undesirable. According to Sims (2000), it is due to two main reasons. First, the treatment of initial observations as non-stochastic removes any penalization in the likelihood for parameter estimates that imply a large distance between  $y_1$  and  $\frac{c}{1-\rho}$  (the unconditional mean of the process in the stationary case) and, as such, magnify the effect of the  $\rho^{t-1}$  term in  $DC_t$ . Second, the use of a flat prior on  $(c, \rho)$  implies an informative prior on  $\left( \frac{c}{1-\rho}, \rho \right)$ , with little density in the proximity of  $\rho = 1$ , and thus on approximately linear behaviors of  $DC_t$ .

Sims (2000) illustrates this pathology by simulating artificial data from a random walk process, and analyzing the deterministic component implied by the flat-prior parameter estimates of an AR(1) model. By construction, all the variation in the simulated data

is stochastic. Nevertheless, the estimated model has the tendency to attribute a large fraction of the low frequency behavior of the series to the deterministic component, i.e. to a path of convergence from unlikely initial observations to the unconditional mean of the process. In addition, Sims (2000) proves that the fraction of the sample variation due to the deterministic component converges to a non-zero distribution, if the data-generating process is a random walk without drift. As a consequence, this pathology may persist even in arbitrarily large samples.

The problem is much worse in VARs with more variables and lags, since these models imply a potentially much more complex behavior of the deterministic trends. For example, the deterministic component of an  $n$ -variable VAR with  $p$  lags is a linear combination of  $n \cdot p$  exponential functions plus a constant term. As a result, it can reproduce rather complicated low-frequency dynamics of economic time series.

To illustrate the severity of the problem in real applications, consider the popular 7-variable VAR with log-real GDP, log-real consumption, log-real investment, log-real wages, log hours worked, inflation and a short-term interest rate (Smets and Wouters, 2007, Del Negro et al., 2007, Justiniano et al., 2010). Suppose that a researcher is estimating this model at the end of 1994, using forty years of quarterly data and 5 lags. Figure 2.1 plots the deterministic components implied by the flat-prior (OLS) estimates for six representative time series, along with their actual realization between 1955:I and 1994:IV. First of all, notice that these deterministic trends are more complex at the beginning of the sample. For example, the predictable component of the investment-to-GDP ratio fluctuates substantially between 1955 and 1970, more so than in the rest of the sample.

In addition to exhibiting this marked temporal heterogeneity (Sims, 2000), the deterministic component also seems to explain a large share of the variation of these time series. Consistent with theory, this feature is most evident for the case of persistent series without (or with little) drift, such as hours, inflation, the interest rate or the investment-to-GDP ratio. For instance, the estimated model implies that most of the hump-shaped low-frequency behavior of the federal funds rate was due to deterministic factors, and was thus predictable since as far as 1955 for a person with the knowledge of the VAR coefficients. And so was the fact that interest rates would hit the zero lower bound around 2010.

Most economists would be skeptical of this likely spurious explanatory power of deterministic trends, and may want to downplay it when conducting inference. In principle,

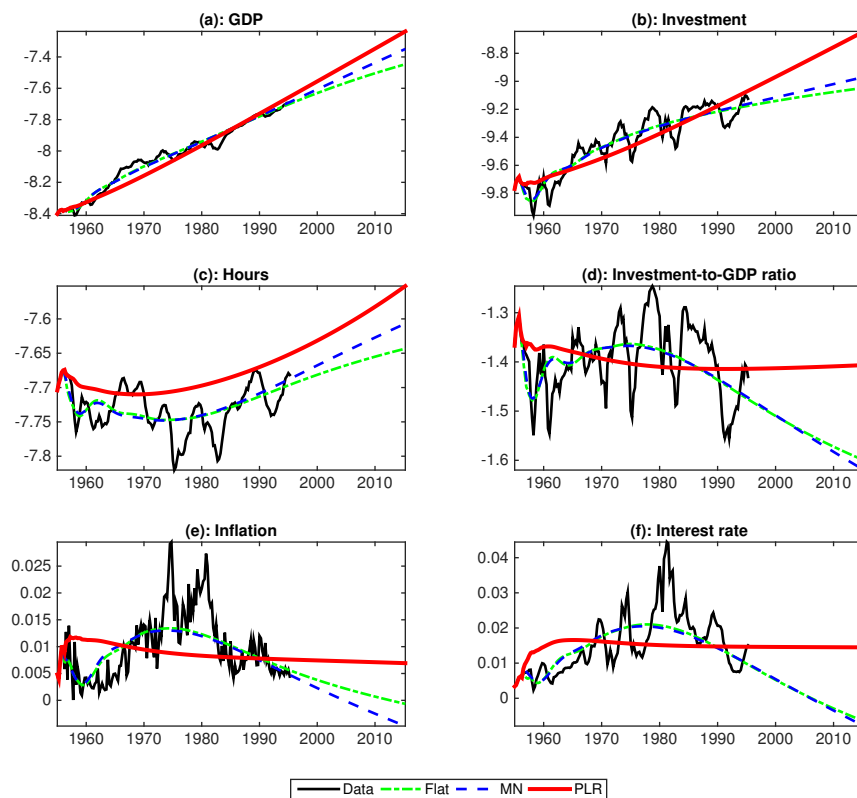


FIGURE 2.1. Deterministic component implied by 7-variable VAR

“one way to accomplish this is to use priors favoring pure unit-root low frequency behavior” (Sims, 2000, pp. 451), according to which implausibly precise long-term forecasts are unlikely. However, it is not obvious how to formulate such a prior. For example, the undesirable properties of the deterministic component persist even when using the popular Minnesota prior with conventional tightness (Litterman, 1979, Sims and Zha, 1998), as shown in figure 2.1. In the next section we detail our specific proposal regarding how to address this problem.

### 3. ELICITATION OF A PRIOR FOR THE LONG RUN

Consider the VAR model

$$(3.1) \quad y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + \varepsilon_t$$

$$\varepsilon_t \sim \text{i.i.d. } N(0, \Sigma),$$

where  $y_t$  is an  $n \times 1$  vector of endogenous variables,  $\varepsilon_t$  is an  $n \times 1$  vector of exogenous shocks, and  $c$ ,  $B_1, \dots, B_p$  and  $\Sigma$  are matrices of suitable dimensions containing the model's unknown parameters. The model can be rewritten in terms of levels and differences

$$(3.2) \quad \Delta y_t = c + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

where  $\Pi = (B_1 + \dots + B_p) - I_n$  and  $\Gamma_j = -(B_{j+1} + \dots + B_p)$ , with  $j = 1, \dots, p-1$ .

The aim of this paper is to elicit a prior for  $\Pi$ . To address the problems described in the previous section, we consider priors that are centered on zero. As for the prior covariance matrix on the elements of  $\Pi$ , our main insight is that its choice must be guided by economic theory, and that alternative—automated or “theory-free”—approaches are likely to lead to a prior specification with undesirable features.

To develop this argument, let  $H$  be any invertible  $n$ -dimensional matrix, and rewrite (3.2) as

$$(3.3) \quad \Delta y_t = c + \Lambda \tilde{y}_{t-1} + \Gamma_1 \Delta y_{t-1} \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

where  $\tilde{y}_{t-1} = Hy_{t-1}$  is an  $n \times 1$  vector containing  $n$  linear combinations of the variables  $y_{t-1}$ , and  $\Lambda = \Pi H^{-1}$  is an  $n \times n$  matrix of coefficients capturing the effect of these linear combinations on  $\Delta y_t$ . In this transformed model, the problem of setting up a prior on  $\Pi$  corresponds to choosing a prior for  $\Lambda$ , conditional on the selection of a specific matrix  $H$ . What is a reasonable prior for  $\Lambda$  will then obviously depend on the choice of  $H$ . For example, consider an  $H$  matrix whose  $i$ -th row contains the coefficients of a linear combination of  $y$  that is a priori likely to be mean reverting. Then it would surely be unwise to place a prior on the elements of the  $i$ -th column of  $\Lambda$  that is excessively tight around zero. In fact, following the standard logic of cointegration, if the elements of the  $i$ -th column of  $\Lambda$  were all zero, there would not be any “error-correction” mechanism at play to preserve the stationarity of this linear combination of  $y$ . A similar logic would suggest that, if a row of  $H$  contains the coefficients of an a-priori likely non-stationary linear combination of  $y$ , one can afford more shrinkage on the elements of the corresponding column of  $\Lambda$ .

This simple argument suggests that is important to set up different priors on the loadings associated with linear combinations of  $y$  with different degree of stationarity. This objective can be achieved by formulating a prior on  $\Lambda$  conditional on a choice of  $H$  that combines

the data in a way that a-priori likely stationary combinations are separated from the non-stationary ones.

Interestingly, in many contexts, economic theory can provide useful information for choosing a matrix  $H$  with these characteristics. For example, according to the workhorse macroeconomic model, output, consumption and investment are likely to share a stochastic trend, while both the consumption-to-output and the investment-to-output ratios should be stationary variables. Similarly, standard economic theory would predict that the price of different goods might be trending, while relative prices should be mean reverting. If these statements were literally true, the corresponding VARs would have an exact “error-correction” representation, as in Engle and Granger (1987), with a reduced-rank  $\Pi$  matrix. In practice, it is very difficult to say with absolute certainty whether certain linear combinations of the data are stationary or integrated. It might therefore be helpful to work with a prior density that is based on some robust insights of economic theory, while also allowing the posterior estimates to deviate from them based on the likelihood information.

We operationalize these ideas by specifying the following prior distribution on the loadings  $\Lambda$  (as opposed to  $\Pi$ ), conditional on a specific choice of the matrix  $H$ :

$$(3.4) \quad \Lambda_{\cdot i} | H_{i\cdot}, \Sigma \sim N \left( 0, \tilde{\phi}_i (H_{i\cdot}) \Sigma \right), \quad i = 1, \dots, n,$$

where  $\Lambda_{\cdot i}$  denotes the  $i$ -th column of  $\Lambda$ ,  $H_{i\cdot}$  is the  $i$ -th row of  $H$ , and  $\tilde{\phi}_i (H_{i\cdot})$  is a scalar that controls the overall tightness. For tractability, we also assume that these priors are scaled by the variance of the error  $\Sigma$ , are independent across  $i$ 's and Gaussian, which guarantees conjugacy. Notice, however, that the assumption that the priors on the columns of  $\Lambda$  are independent from each other does not rule out (and will in general imply) that the priors on the columns of  $\Pi$  are correlated, with a correlation structure that depends on the choice of  $H$  and  $\tilde{\phi}$ .

**3.1. Reference value for  $\tilde{\phi}_i$ .** The crucial element of the density specified in (3.4) is the fact that  $\tilde{\phi}_i$  is a function of  $H_{i\cdot}$ , which is consistent with the intuition that the tightness of the prior on the loadings  $\Lambda_{\cdot i}$  should depend on whether these loadings multiply a likely stationary or non-stationary linear combination of  $y$  from an a-priori perspective. To capture this important insight, we parameterize  $\tilde{\phi}_i$  as follows:

$$(3.5) \quad \tilde{\phi}_i (H_{i\cdot}) = \frac{\phi_i^2}{(H_{i\cdot} \bar{y}_0)^2},$$



where  $\phi_i$  is a scalar hyperparameter (controlling the standard deviation of the prior on the elements of  $\Lambda_{\cdot i}$ ), and  $\bar{y}_0$  is a column vector containing the average of the initial  $p$  observations of each variable of the model (these are the observations taken as given in the computation of the likelihood function). Therefore, the denominator of (3.5) corresponds to the square of the initial value of the linear combination at hand.

There are various reasons why this formulation is appealing. First of all, substituting (3.5) into (3.4) makes it clear that the prior variance of  $\Lambda_{\cdot i}$  has a scaling that is similar to that of the likelihood, with the variance of the error at the numerator, and the (sum of) squared regressor(s) at the denominator (recall the form of the variance of the OLS estimator). Second, expression (3.5) captures the insight that tighter priors are more desirable for the loadings of non-stationary linear combinations of  $y$ , which are likely to have larger initial values (assuming that the data generating process has been in place for a long enough period of time before the observed sample). Third, scaling the prior variance by  $1/(H_i \cdot \bar{y}_0)^2$  is more attractive than any alternative scale meant to capture the same idea, because in this way the prior setup does not rely on any information that is also used to construct the likelihood function, avoiding any type of “double counting” of the data. Fourth, the presence of the hyperparameter  $\phi_i$  gives additional flexibility in the setting of the prior tightness. Moreover, the fact that our prior is conjugate, allows us to easily use the marginal likelihood (or the hierarchical interpretation of the model) for the selection of these hyperparameters, as in Giannone et al. (2014).

In the next subsection, we will explain how to implement this prior in a simple way using dummy observations, which not only provides some additional insights into its interpretation, but has also a clear connection with the the problem of the initial conditions and deterministic trends that we have highlighted in section 2.

**3.2. Implementation with dummy observations.** The prior in (3.4) can be rewritten in a more compact form as

$$(3.6) \quad \text{vec}(\Lambda) | H, \Sigma \sim N\left(0, \tilde{\Phi}_H \otimes \Sigma\right),$$

with  $\tilde{\Phi}_H = \text{diag}\left(\left[\tilde{\phi}_1(H_1), \dots, \tilde{\phi}_n(H_n)\right]\right)$ , where  $\text{vec}(\cdot)$  is the vectorization operator, and  $\text{diag}(x)$  denotes a diagonal matrix with the vector  $x$  on the main diagonal. Since  $\Pi = \Lambda H$ , the implied prior on the columns of  $\Pi$  is given by

$$(3.7) \quad \text{vec}(\Pi) | H, \Sigma \sim N\left(0, H' \tilde{\Phi}_H H \otimes \Sigma\right).$$

Being conjugate, this prior can be easily incorporated into inference using Theil mixed estimation, by adding a set of  $n$  artificial (or dummy) observations to the original sample. Each of these  $n$  dummy observations consists of a value of the variables on the left- and right-hand side of (3.1), at an artificial time  $t_i^*$ . In particular, the implementation of the prior in (3.7) requires the following set of artificial observations:

$$(3.8) \quad y_{t_i^*} = y_{t_i^*-1} = \dots = y_{t_i^*-p} = \frac{H_i \bar{y}_0}{\phi_i} [H^{-1}]_{\cdot i}, \quad i = 1, \dots, n,$$

where the corresponding observation multiplying the constant term is set to zero, and  $[H^{-1}]_{\cdot i}$  denotes the  $i$ -th column of  $H^{-1}$ . We prove this result in appendix A, where we also derive the posterior distribution of the model's unknown coefficients.

To provide yet another interpretation of our prior, it is useful to substitute the dummy observations (3.8) into the level-difference representation of the model (3.2), obtaining

$$0 = \Pi (H_i \bar{y}_0) [H^{-1}]_{\cdot i} + \phi_i \varepsilon_{t_i^*}, \quad i = 1, \dots, n.$$

This expression suggests that the prior is effectively limiting the extent to which the linear combinations  $H_i \cdot y$  help forecasting  $\Delta y$  at the beginning of the sample. This feature reduces the importance of the error correction mechanisms of the model, which are responsible for the complex dynamics and excessive explanatory power of the deterministic component that we have analyzed in section 2. However, given that the value of  $H_i \bar{y}_0$  is typically lower (in absolute value) for mean-reverting combinations of  $y$ , our prior reduces more gently the mechanisms that correct the deviations from equilibrium of likely stationary combinations of the variables, consistent with the idea of cointegration.

**3.3. Simple bivariate example.** Before turning to a more comprehensive comparison with some of the existing literature, it is useful to contrast our prior for the long run to the more standard sum-of-coefficients (SOC) prior, first proposed by Doan et al. (1984), and routinely used for the estimation of BVARs (Sims and Zha, 1998). The SOC prior is a prior on the sum of coefficients on the own lags of each equation of the VAR, and corresponds to

mechanically setting  $H$  equal to the identity matrix, even when there might be some linear combinations of the variables in the system that should be stationary.

For the sake of concreteness, consider the simple example of a bivariate VAR(1) with log-output ( $x_t$ ) and log-investment ( $i_t$ ). The SOC prior corresponds to

$$vec(\Pi) | \Sigma \sim N \left( 0, \begin{bmatrix} \frac{\mu^2}{x_0^2} & 0 \\ 0 & \frac{\mu^2}{i_0^2} \end{bmatrix} \otimes \Sigma \right),$$

where  $\mu$  is an hyperparameter controlling its overall tightness. Economic theory, however, does suggest that output and investment are likely to share a common trend ( $x_t + i_t$ ), while the log-investment-to-output ratio ( $i_t - x_t$ ) is expected to be stationary. Based on this insight, we can form the matrix  $H$ , whose rows correspond to the coefficients of these two different linear combinations of the variables:

$$(3.9) \quad H = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

One can now ask what is the prior implied by SOC on the coefficients capturing the effect of these two linear combinations on  $\Delta x_t$  and  $\Delta i_t$  (i.e. on the error correction coefficients, using the cointegration terminology). To answer this question, recall that  $\Lambda = \Pi H^{-1}$ , which implies that  $vec(\Lambda) = \left( (H^{-1})' \otimes I_n \right) vec(\Pi)$ , and thus

$$vec(\Lambda) | H, \Sigma \sim N \left( 0, \frac{1}{4} \begin{bmatrix} \frac{\mu^2}{x_0^2} + \frac{\mu^2}{i_0^2} & \frac{\mu^2}{i_0^2} - \frac{\mu^2}{x_0^2} \\ \frac{\mu^2}{i_0^2} - \frac{\mu^2}{x_0^2} & \frac{\mu^2}{x_0^2} + \frac{\mu^2}{i_0^2} \end{bmatrix} \otimes \Sigma \right).$$

Notice that the prior on the loadings of the common trend is as tight as that on the loadings of the investment ratio, which is in contrast with the predictions of most theoretical models and with the main insights of cointegration.

On the contrary, our prior for the long run with the choice of  $H$  in (3.9) corresponds to the following prior density on the error correction coefficients:

$$vec(\Lambda) | H, \Sigma \sim N \left( 0, \begin{bmatrix} \frac{\phi_1^2}{(x_0 + i_0)^2} & 0 \\ 0 & \frac{\phi_2^2}{(i_0 - x_0)^2} \end{bmatrix} \otimes \Sigma \right).$$

Clearly, even if  $\phi_1 \approx \phi_2$ , and given that  $(i_0 - x_0)^2$  is a much smaller number than  $(x_0 + i_0)^2$ , this prior performs much less shrinkage on the coefficients that correct the deviations of the investment ratio from its equilibrium value.

## 4. RELATIONSHIP WITH THE LITERATURE

Before turning to the empirical application, it is useful to relate our approach more precisely to the literature on cointegration (Engle and Granger, 1987) and error-correction models (for a comprehensive review, see Watson, 1986). For the purpose of making this comparison as concrete as possible, suppose that the specific model at hand entails a natural choice of  $H$  with the following two blocks of rows:

$$H = \begin{bmatrix} \beta'_{\perp} \\ (n-r) \times n \\ \beta' \\ r \times n \end{bmatrix},$$

where the columns of  $\beta_{\perp}$  are  $(n-r)$  linear combinations of  $y$  that are likely to exhibit a stochastic trend, while the columns of  $\beta$  are  $r$  linear combinations of  $y$  that are more likely to represent stationary deviations from long-run equilibria, i.e. that are likely to correspond to cointegrating vectors. Using this notation, we can rewrite (3.3) as

$$(4.1) \quad \Delta y_t = c + \Lambda_1 (\beta'_{\perp} y_{t-1}) + \Lambda_2 (\beta' y_{t-1}) + \Gamma_1 \Delta y_{t-1} \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

where  $\Lambda_1$  are the first  $n-r$  columns of  $\Lambda$ , and  $\Lambda_2$  are the remaining  $r$  columns.

As described in the previous section, our approach consists of placing priors on the columns of  $\Lambda_1$  and  $\Lambda_2$ . These priors are centered on zero, and are tighter for the elements of  $\Lambda_1$  than for those of  $\Lambda_2$ . The error-correction representation corresponds to an extreme case of our general model, obtained by enforcing a dogmatic prior belief that  $\Lambda_1 = 0$ . As a result,  $\Pi$  would equal  $\Lambda_2 \beta'$ , and would be rank deficient. If, in addition, the prior belief that  $\Lambda_2 = 0$  is also dogmatically imposed, the VAR admits a representation in first differences.

For what concerns the cointegrating vectors  $\beta$ , the literature has proceeded by either fixing or estimating them. The approach that is closer to ours selects the cointegrating vectors  $\beta$  a priori, mostly based on economic theory. This strategy is appealing since the theoretical cointegrating relations are typically quite simple and robust across a wide class of economic models. Conditional on a specific choice of  $\beta$ , one popular approach is to include all the theoretical cointegrating vectors in the error correction representation, and conduct likelihood-based inference (i.e. OLS), as in King et al. (1991) or Altig et al. (2011). In our model, this is equivalent to placing a flat prior on  $\Lambda_2$ . An alternative strategy, however, has been to conduct some pre-testing and to include in the error correction only those deviations

from equilibria for which the adjustment coefficients are statistically significantly different from zero, as in Horvath and Watson (1995). Conditional on the pre-testing results, this approach is equivalent to setting a dogmatic prior that certain columns of  $\Lambda_2$  are equal zero, and a flat prior on the remaining elements.

The other strand of the literature is more agnostic about both the cointegrating rank ( $r$ ) and the cointegrating vectors ( $\beta$ ). In these cases, classical inference is typically conducted using a multi-step methodology. The first step of these procedures requires testing for the cointegrating rank. Conditional on the results of these tests, the second step consists of the estimation of the cointegrating vectors, which are then treated as known in step three for the estimation of the remaining model parameters (Engle and Granger, 1987). Alternatively, the second and third steps can be combined to jointly estimate  $\beta$  and the other parameters with likelihood-based methods, as in Johansen (1995).

The Bayesian approach to cointegration is similar in spirit to the likelihood-based inference (for recent surveys, see Koop et al., 2006, Del Negro and Schorfheide, 2011, and Karlsson, 2013). This literature has also concentrated on conducting inference on the cointegrating rank and the cointegrating space. For example, the number of cointegrating relationships is typically selected using the marginal likelihood, or related Bayesian model comparison methods (Chao and Phillips, 1999, Kleibergen and Paap, 2002, Corander and Villani, 2004, Villani, 2005). In practical applications, this methodology ends up being similar to pre-testing because the uncertainty on the cointegrating rank is seldom formally incorporated into the analysis, despite the fact that the Bayesian approach would allow to do so (for an exception, see Villani, 2001).

Conditional on the rank, the early Bayesian cointegration literature was concerned with formulating priors on the cointegrating vectors, and with deriving and simulating their posterior (Bauwens and Lubrano, 1996, Geweke, 1996). Standard priors, however, have been shown to be problematic, in light of the pervasive local and global identification issues of error-correction models (Kleibergen and van Dijk, 1994, Strachan and van Dijk, 2005). To avoid these problems, a better strategy is to place a prior on the cointegrating space, which is the only object the data are informative about (Villani, 2000). Such priors are studied in Strachan and Inder (2004) and Villani (2005), who also develop methods for inference and posterior simulations. In particular, Villani (2005) proposes a diffuse prior on the cointegrating space, trying to provide a Bayesian interpretation to some popular

likelihood-based procedures. In general, little attention has been given to the elicitation of informative priors on the adjustment coefficients, which is instead the main focus of our paper.

It is well known that maximum-likelihood (or flat-prior) inference in the context of error-correction models can be tricky (Stock, 2010). This is not only because of the practice to condition on initial conditions, as we have stressed earlier, but also because inference is extremely sensitive to the value of non-estimable nuisance parameters characterizing small deviations from non-stationarity of some variables (Elliott, 1998, Mueller and Watson, 2008). Pretesting is clearly plagued by the same problems. The selection of models based on pre-testing or Bayesian model comparison can be thought as limiting cases of our approach, in which the support of the distributions of the hyperparameters controlling the tightness of the prior on specific adjustment coefficients can only take values equal to zero or to a large number. One advantage of our flexible modeling approach, instead, is that it removes such an extreme sparsity of the model space, as recommended by Gelman et al. (2004) and Sims (2003).

Finally, our paper is also related to the methodology of Del Negro and Schorfheide (2004) and Del Negro et al. (2007), who also use a theoretical DSGE model to set up a prior for the VAR coefficients. Their work, however, differs from ours in two important ways. First of all, the prior of Del Negro et al. (2007) is centered on the error-correction representation of the VAR, given that such a prior pushes towards a DSGE model featuring a balance growth path. On the contrary, for the reasons highlighted in section 2, our PLR shrinks the VAR towards the representation in first differences, albeit it does so more gently for the linear combinations of the variables that are supposed to be stationary according to theory. In addition, the approach of Del Negro and Schorfheide (2004) requires the complete specification of a DSGE model, including its short-run dynamics. Instead, we use only the long-run predictions of a wide class of theoretical models to guide the set up of our PLR. Among other things, this strategy allows us to work with a conjugate prior and simplify inference.

## 5. EMPIRICAL RESULTS

In this section we use our prior to conduct inference in VARs with standard macroeconomic variables, whose joint long-run dynamics is sharply pinned down by economic theory.

In particular, we perform two related, but distinct exercises. We begin by re-estimating the 7-variable VAR of section 2, to show that our PLR serves the purpose of reducing the excessive explanatory power of the deterministic components implied by the model with flat or Minnesota priors. Second, we evaluate the forecasting performance of a 3- and a 7-variable VAR, and demonstrate that our prior yields substantial gains over more standard BVARs, especially when forecasting at long horizons. Before turning to the detailed illustration of these results, we begin by describing more precisely the 3- and 7-variable VARs and the priors that we adopt.

The 3-variable VAR includes data on log-real output ( $Y_t$ ), log-real consumption ( $C_t$ ) and log-real investment ( $I_t$ ) for the US economy, and is similar to the VAR estimated by King et al. (1991) in their influential analysis of the sources of business cycles.<sup>1</sup> This model is appealing because of its simplicity and because, in this context, a prior for the long run can be easily elicited based on standard economic theory, which has robust implications about the long-run behavior of these three time series. Specifically, a wide class of macroeconomic models predicts the existence of a balanced growth path, along which output, consumption and investment share a common trend, while the great ratios (the consumption- and investment-to-output ratios) should be stationary. Since our variables are transformed in logarithms, these theoretical predictions suggest the following natural choice of the vector  $\tilde{y}_t$  and the matrix  $H$ :

$$\tilde{y}_t = \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}}_H \underbrace{\begin{pmatrix} Y_t \\ C_t \\ I_t \end{pmatrix}}_{y_t} \rightarrow \begin{array}{l} \text{(log-) common trend} \\ \text{(log-) consumption-to-output ratio} \\ \text{(log-) investment-to-output ratio} \end{array}$$

The 7-variable VAR includes instead data on US log-real output ( $Y_t$ ), log-real consumption ( $C_t$ ), log-real investment ( $I_t$ ), log-real wages ( $W_t$ ), log-hours worked ( $H_t$ ), inflation ( $\pi_t$ ), and the federal funds rate ( $R_t$ ). These are the same time series used to estimate the DSGE model of Smets and Wouters (2007), which builds on the Real Business Cycle (RBC) literature by adding a number of nominal and real frictions. It is representative of modern medium-scale macroeconomic models, and can thus be used as a guide to set up our prior in this context. Similar to the RBC framework, this class of models typically predicts that

<sup>1</sup>The data used in the empirical applications are described in details Appendix C.

real output, consumption, investment, and real wages share a common stochastic trend, while the great ratios (the consumption- and investment- and real wage-to-output ratios) and hours worked should be stationary.

In addition, some New-Keynesian models (e.g. Ireland, 2007) also include a stochastic nominal trend, common to the interest and inflation rates. While the existence of such a stochastic nominal trend is not a robust feature of this class of models, most of them do imply that the low-frequency behavior of inflation and interest rates is tightly related. This is exactly the type of situation in which it might be beneficial to formulate a prior that is centered on the existence of a common nominal trend, without imposing it dogmatically. All in all, these considerations lead to the following choice of  $\tilde{y}_t$  and  $H$ :

$$\tilde{y}_t = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}}_H \underbrace{\begin{pmatrix} Y_t \\ C_t \\ I_t \\ W_t \\ H_t \\ \pi_t \\ R_t \end{pmatrix}}_{y_t} \begin{array}{l} \rightarrow \text{real trend} \\ \rightarrow \text{nominal trend} \\ \rightarrow \text{consumption-to-GDP ratio} \\ \rightarrow \text{investment-to-GDP ratio} \\ \rightarrow \text{real wage-to-GDP ratio} \\ \rightarrow \text{hours} \\ \rightarrow \text{real interest rate} \end{array}$$

We now turn to the description of the specific exercises that we conduct and the empirical results that we obtain.

**5.1. Deterministic trends.** In section 2, we have argued that a serious pathology of flat-prior VARs is that they imply rather complex dynamics and excessive explanatory power of the deterministic component of economic time series (Sims, 1996 and 2000). In addition, the use of the standard Minnesota prior (with conventional hyperparameter values) does very little, if nothing at all, to solve the problem (figure 2.1). In this subsection, we analyze the extent to which our PLR eliminates or reduces this pathology.

To this end, we re-estimate the 7-variable VAR of section 2 using our prior for the long run (PLR-BVAR), and compare the deterministic trends implied by this model to those of section 2, obtained using a BVAR with flat or Minnesota priors (flat-BVAR and



MN-BVAR respectively).<sup>2</sup> In this experiment, for simplicity, we simply set the value of the hyperparameters  $\{\phi_i\}_{i=1}^n$  equal to one, which provides a good reference value (it corresponds to adding one dummy observation).

The results of this experiment are depicted in figure 2.1. In the case of GDP and hours, the difference between the deterministic component of the PLR-BVAR and the flat or MN-BVAR is limited. For the other variables, notice that the shape of the deterministic component implied by the PLR-BVAR is simpler, and explains much less of the low frequency variation of the time series. For example, in the case of investment, the deterministic trend implied by the PLR-BVAR resembles a straight line, implying that the long-run growth rate of investment in the next decades is expected to be similar to the past. Similarly, in the case of inflation and the interest rate, the deterministic trend of the PLR-BVAR does not have the unpleasant property that somebody with the knowledge of the VAR coefficients would have perfectly predicted the hump shape of these two variables already in 1955.

So, overall, our PLR is quite successful in correcting the pathology that we have illustrated in section 2. In the next section, we will demonstrate that this is not simply a theoretical curiosity, but that it is extremely important for the forecasting performance of the model.

**5.2. Forecasting performance.** In this subsection, we compare the forecasting performance of our BVAR to a number of benchmark BVARs. More specifically, we consider the following models:

- MN-BVAR: BVAR with a Minnesota prior
- SZ-BVAR: BVAR with a Minnesota prior and a sum-of-coefficients (also known as no-cointegration) prior, as in the work of Doan et al. (1984) and Sims and Zha (1998). The latter corresponds to our prior for the long run with a mechanic choice of  $H$  equal to the identity matrix (and the same hyperparameter for each dummy observation). It has the effect of pushing the VAR parameter estimates towards the existence of a separate stochastic trend for each variable.
- Naive: BVAR with an infinitely tight Minnesota prior, which results in all the variables following independent random walks with drifts.
- PLR-BVAR: BVAR with a Minnesota prior and our prior for the long run.

---

<sup>2</sup>We conduct this experiment with the 7-variable VAR as opposed to the 3-variable VAR because the problem is more severe in this case.

This comparison of forecasting performances is interesting because the first three models are considered valid benchmarks in the literature. For example, it is well known that MN-BVARs yield substantial forecasting improvements over classical or flat-prior VARs (Litterman, 1979) and that further improvements can be achieved by adding the sum-of-coefficients prior of Doan et al. (1984). In fact, Giannone et al. (2014) show that the predictive ability of the model of Sims and Zha (1998) is comparable to that of factor models. Finally, a number of papers have demonstrated that the naive random walk model forecasts quite well, especially after the overall decline in the predictability of macroeconomic time series in 1985 (Atkeson and Ohanian., 2001, Stock and Watson, 2007, D’Agostino et al., 2007, Rossi and Sekhposyan, 2010).

In what follows, our measure of forecasting accuracy is the out-of-sample mean squared forecast error (MSFE). In particular, for each of the four models, we produce the 1- to 40-quarters-ahead forecasts, starting with the estimation sample that ranges from 1955Q1 to 1974Q4. We then iterate the same procedure updating the estimation sample, one quarter at a time, until the end of the sample in 2013Q1. At each iteration, we select the tightness of the priors by maximizing the posterior of the models’ hyperparameters, using the procedure proposed by Giannone et al. (2014) and summarized in appendix B. Conditional on a specific value of these hyperparameters, we then produce the out-of-sample forecasts by setting the VAR coefficients to their posterior mode. All models are estimated using 5 lags. For all the forecast horizons, the evaluation sample for the computation of the MSFEs ranges from 1985Q1 to 2013Q1.

5.2.1. *3-variable VARs.* We start by focusing on the small-scale model with three variables. Figure 5.1 reports the MSFEs of the level of each variable, the common trend, and the so-called great ratios at horizons 1 to 40. Notice that the PLR-BVAR improves uniformly over the MN-BVAR, especially at long horizons, reflecting the fact that the Minnesota prior alone is not enough to reduce the spurious explanatory power of the deterministic component typical of flat-prior VARs. According to the existing literature, one way to reduce this pathology is to augment the MN-BVAR with a sum-of-coefficients prior. The resulting SZ-BVAR does outperform the MN-BVAR, but is still substantially less accurate than the PLR-BVAR for predicting investment and the investment-to-GDP ratio. Finally, observe that the forecasting performance of the naive model is very similar to that of the SZ-BVAR, which suggests that the sum-of-coefficients prior strongly shrinks the VAR

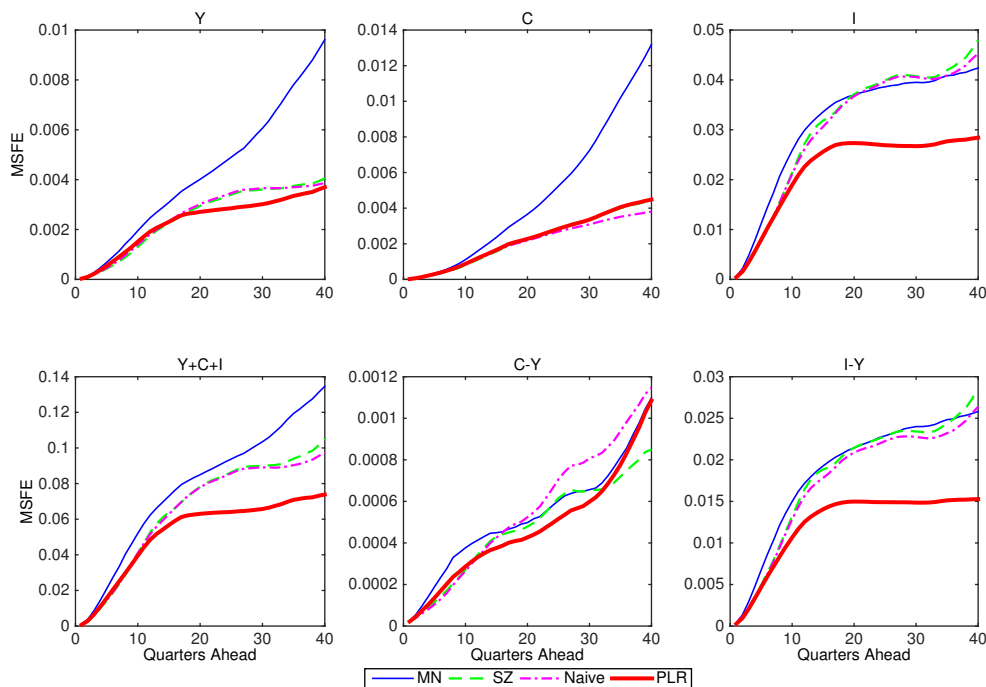


FIGURE 5.1. Mean squared forecast error in models with three variables

coefficients toward values consistent with the existence of independent random walks for all three variables.

The key question for us is understanding why the PLR-BVAR outperforms the SZ-BVAR and the naive model. We address this question in figure 5.2, which plots the realized value of the consumption- and investment-to-GDP ratios, and the 5-year-ahead forecasts obtained with the PLR-BVAR and the naive model (the SZ-BVAR forecasts are very close to those of the naive model, so we do not report them to avoid clogging the figure). Since all the variables follow separate random walks in the naive model, the difference between log-consumption and log-output, and the difference between log-investment and log-output are also random walks. As figure 5.2a makes clear, a random walk is a pretty good predictor of the consumption-to-GDP ratio because this variable displays a (close to) non-stationary behavior in the data. The no-change forecasts of a random walk, however, are poor predictors of the investment-to-GDP ratio at long horizons, because this series looks mean reverting (figure 5.2b).

The strength of the PLR-BVAR is the ability to push the common trend towards a unit root approximately as intensely as the SZ-BVAR or the naive model, while performing

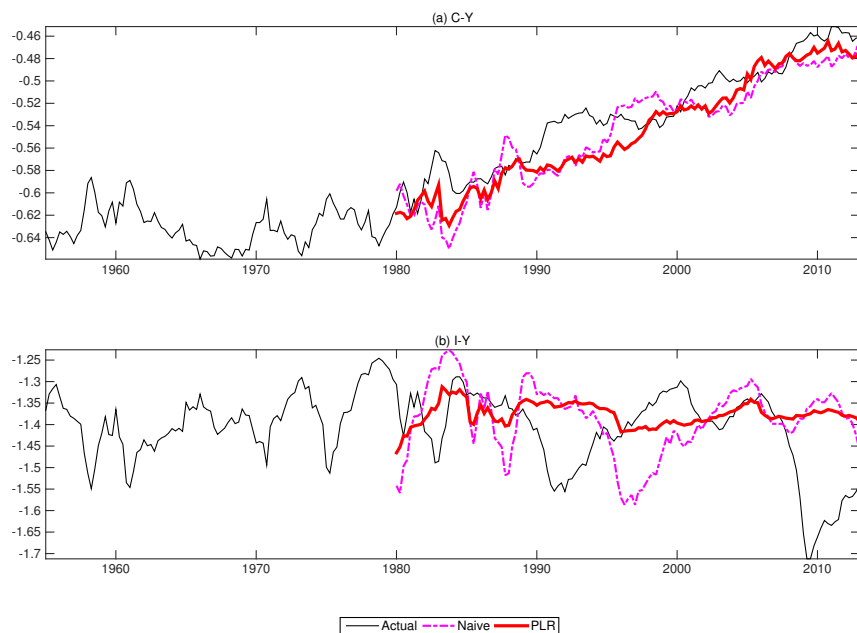


FIGURE 5.2. Forecasts of the great ratios 5-years ahead in models with three variables

substantially less shrinkage on the consumption- and investment-to-GDP ratios. Therefore, this more sophisticated prior does not outweigh the likelihood information about the mean reversion of the investment ratio, while being consistent with the trending behavior of the consumption ratio. Finally, notice that the PLR-BVAR would also outperform the theory-based predictions of constant ratios, which is particularly at odds with the observed pattern of consumption relative to GDP.

5.2.2. *7-variable VARs.* We now turn to the case of the 7-variable VAR. Figures 5.3 and 5.4 plot the MSFEs for various forecasting horizons for the level of all the variables included in the VAR and for the linear combinations of the variables obtained by multiplying the matrix  $H$  by the vector  $y$  (i.e. the common trends, the great ratios and the real rate).

The forecasting comparison summarized in figures 5.3 and 5.4 provides similar insight to the 3-variable case. With the exception of hours and the real wage-to-GDP ratio, the PLR-BVAR improves over the MN-BVAR for all variables and horizons. The most substantial gains are evident for the nominal variables, consumption and investment (and their ratio to GDP). The PLR-BVAR also outperforms the SZ-BVAR uniformly over variables and

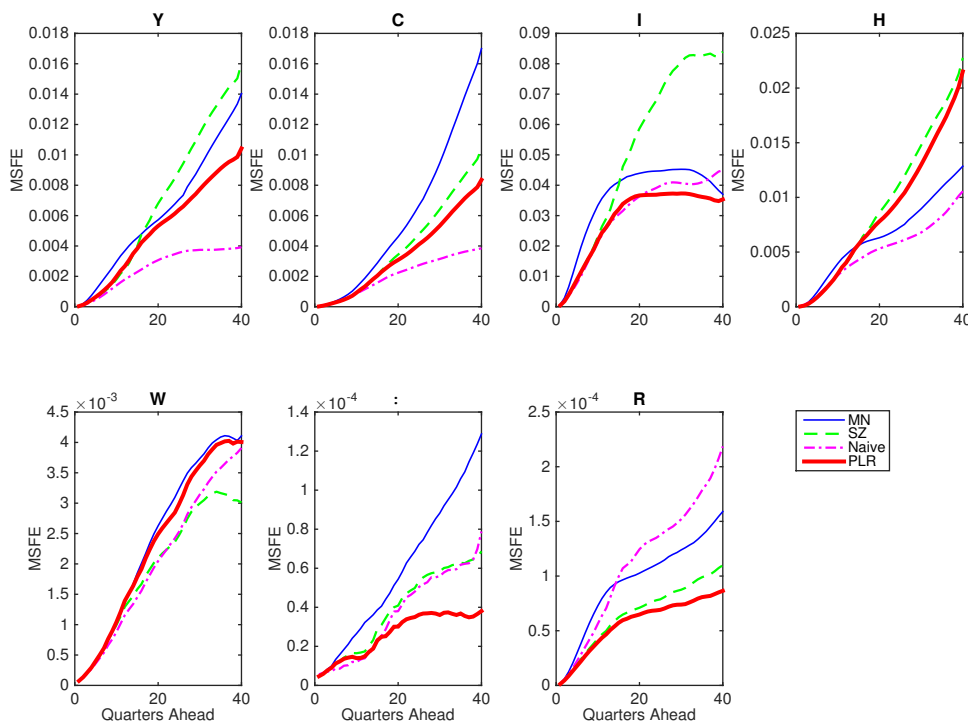


FIGURE 5.3. Mean squared forecast error in models with seven variables (linear combinations)

horizons, once again with the exception of wages. The forecasting accuracy of the PLR-BVAR is also generally good relative to the naive model. What is interesting of the 7-variable case, however, is that the performance of all the BVARs deteriorates relative to the naive model for output, consumption and the wage ratio.

## 6. CONCLUDING REMARKS

In this paper, we have introduced a new class of prior distributions for VARs that (i) impose discipline on long-run behavior of the model; (ii) are based on robust lessons of theoretical macroeconomic models; (iii) can be thought of as a full probabilistic approach to cointegration, and include the error-correction representation of a VAR as a special case; and (iv) perform well in forecasting, especially at long horizons.

While our priors for the long run present a number of appealing features, it is important to also point out their potential limitations. In particular, the procedure to set up these priors is not automated, and requires thinking about the economics behind the determination

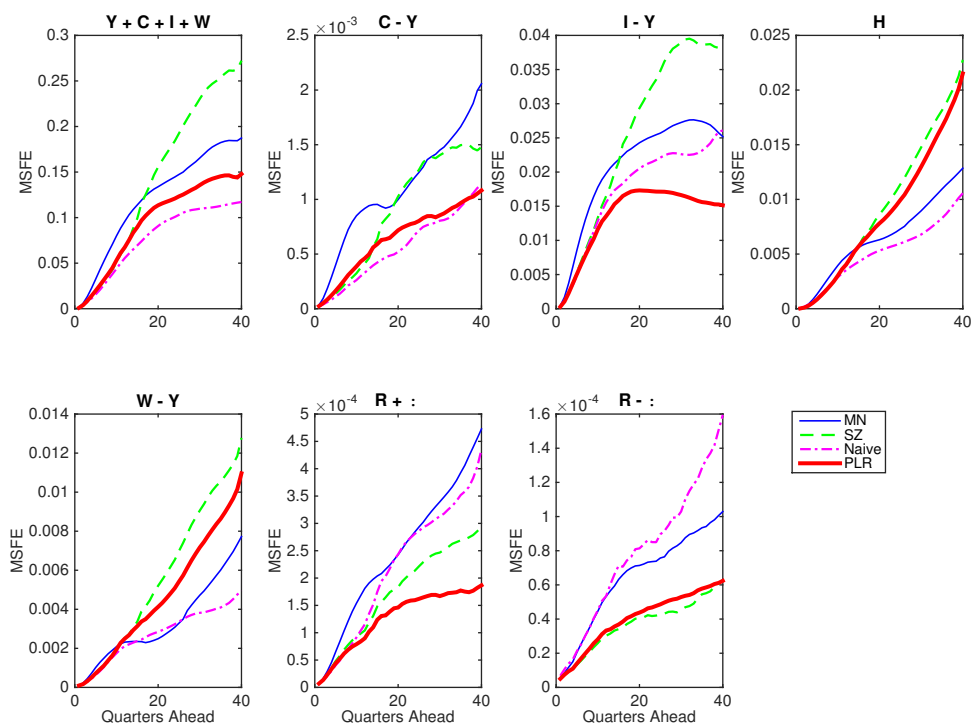


FIGURE 5.4. Mean squared forecast error in models with seven variables

of the variables included in the model, even if such a model is a reduced-form one that normally does not require economic theory. An implication of this statement is that setting up a PLR in large-scale models with dozens of variables may prove difficult. Finally, the implementation of our prior requires the choice of the matrix  $H$ , whose rows contain the coefficients of the linear combinations of the variables that are likely to be stationary and non-stationary. This degree of freedom might of course open the door to data mining, which should be avoided by adhering as much as possible to the basic lessons of standard DSGEs.

#### APPENDIX A. POSTERIOR DISTRIBUTIONS

In this appendix we describe the posterior distribution of the VAR coefficients under the various prior densities that we have used in the paper. Except for the derivations related to the PLR, the other results are standard, and we report them only to make the paper self-contained.

Consider the VAR model of section 3

$$y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + \varepsilon_t$$

$$\varepsilon_t \sim \text{i.i.d. } N(0, \Sigma),$$

and rewrite it as

$$Y = X\beta + \epsilon$$

$$\epsilon \sim N(0, \Sigma \otimes I_{T-p}),$$

where  $y \equiv [y_{p+1}, \dots, y_T]'$ ,  $Y \equiv \text{vec}(y)$ ,  $x_t \equiv [1, y'_{t-1}, \dots, y'_{t-p}]'$ ,  $X_t \equiv I_n \otimes x'_t$ ,  $x \equiv [x_{p+1}, \dots, x_T]'$ ,  $X \equiv I_n \otimes x$ ,  $\varepsilon \equiv [\varepsilon_{p+1}, \dots, \varepsilon_T]'$ ,  $\epsilon \equiv \text{vec}(\varepsilon)$ ,  $B \equiv [C, B_1, \dots, B_p]'$  and  $\beta \equiv \text{vec}(B)$ . Finally, define the number of regressors for each equation by  $k \equiv np + 1$ .

**A.1. Flat prior.** With a flat prior, the posterior of  $(\beta, \Sigma)$  belongs to the usual Normal-Inverse-Wishart family:

$$\Sigma|Y \sim IW(\hat{\varepsilon}'_{ols} \hat{\varepsilon}_{ols}, T - p - n - k - 1)$$

$$\beta|\Sigma, Y \sim N(\hat{\beta}_{ols}, \Sigma \otimes (x'x)^{-1}),$$

where  $\hat{B}_{ols} \equiv (x'x)^{-1}(x'y)$ ,  $\hat{\beta}_{ols} \equiv \text{vec}(\hat{B}_{ols})$ ,  $\hat{\varepsilon}_{ols} \equiv y - x\hat{B}_{ols}$ .

**A.2. Minnesota prior.** The so-called Minnesota prior, first introduced in ?, is centered on the assumption that each variable follows a random walk process, possibly with drift. More precisely, this prior is characterized by the following first and second moments:

$$E[(B_s)_{ij} | \Sigma] = \begin{cases} 1 & \text{if } i = j \text{ and } s = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{cov}((B_s)_{ij}, (B_r)_{hm} | \Sigma) = \begin{cases} \lambda^2 \frac{1}{s^2} \frac{\Sigma_{ih}}{\hat{\sigma}_j^2 / (d-n-1)} & \text{if } m = j \text{ and } r = s \\ 0 & \text{otherwise} \end{cases},$$

where the hyperparameter  $\lambda$  controls the overall tightness of this prior and, as customary, the  $\hat{\sigma}_j^2$ 's are set equal to the residual variance of an AR(1) estimated using the available data for variable  $j$ . To obtain a proper prior, we also specify a standard Inverse-Wishart prior on  $\Sigma$ , as in Kadiyala and Karlsson (1997):

$$\Sigma|Y \sim IW(\text{diag}([\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2]), n + 2).$$

Through an appropriate choice of  $\Psi$ ,  $d$ ,  $b$  and  $\Omega$ , such a prior can be easily cast into the Normal-Inverse-Wishart form

$$\begin{aligned}\Sigma &\sim IW(\Psi; d) \\ \beta|\Sigma, &\sim N(b, \Sigma \otimes \Omega),\end{aligned}$$

and leads to the following posterior distribution for the VAR coefficients

$$\begin{aligned}\Sigma|Y &\sim IW\left(\Psi + \varepsilon'\varepsilon + (\hat{B} - b)'\Omega^{-1}(\hat{B} - b), T - p + d\right) \\ \beta|\Sigma, Y, &\sim N\left(\hat{\beta}, \Sigma \otimes (x'x + \Omega^{-1})^{-1}\right),\end{aligned}$$

where  $\hat{B} \equiv (x'x + \Omega^{-1})^{-1}(x'y + \Omega^{-1}b)$ ,  $\hat{\beta} \equiv \text{vec}(\hat{B})$ ,  $\varepsilon \equiv y - x\hat{B}$ , and  $b$  is a  $k \times n$  matrix obtained by reshaping the vector  $b$  in such a way that each column corresponds to the prior mean of the coefficients of each equation (i.e.  $b \equiv \text{vec}(b)$ ).

**A.3. Prior for the long run.** In the main text, we have stated that the implementation of the PLR requires the following set of dummy observations for the artificial times  $t_1^*, \dots, t_n^*$ :

$$(A.1) \quad yt_i^* = yt_{i-1}^* = \dots = yt_{i-p}^* = \frac{H_i \cdot \bar{y}_0}{\phi_i} [H^{-1}]_{.i}, \quad i = 1, \dots, n.$$

The left- and right-hand side of these artificial observations can be collected into the matrices

$$\begin{aligned}y^+ &= \text{diag}\left(\left[\frac{H_1 \cdot \bar{y}_0}{\phi_1}, \dots, \frac{H_n \cdot \bar{y}_0}{\phi_n}\right]\right) [H^{-1}]' \\ x^+ &= \begin{bmatrix} 0 & y^+ & \dots & y^+ \\ n \times (1+np) & n \times 1 & & \end{bmatrix},\end{aligned}$$

which can then be added on top of the data matrices  $y$  and  $x$  to conduct inference as if they were part of the actual sample.

To proof that these dummy observations imply the density in (3.6), substitute the observations in (A.1) into the level-difference representation of the VAR (3.2), obtaining

$$\Pi \frac{H_i \cdot \bar{y}_0}{\phi_i} [H^{-1}]_{.i} = -\varepsilon_{t_i^*}, \quad i = 1, \dots, n.$$

Grouping the columns on the left- and right-hand side of this expression for each  $i$ , we obtain

$$\left[ \Pi \frac{H_1 \cdot \bar{y}_0}{\phi_1} [H^{-1}]_{.1}, \dots, \Pi \frac{H_n \cdot \bar{y}_0}{\phi_n} [H^{-1}]_{.n} \right] = -[\varepsilon_{t_1^*}, \dots, \varepsilon_{t_n^*}],$$



which can be rewritten as

$$\Pi H^{-1} \text{diag} \left( \left[ \frac{H_1 \cdot \bar{y}_0}{\phi_1}, \dots, \frac{H_n \cdot \bar{y}_0}{\phi_n} \right] \right) = - [\varepsilon_{t_1^*}, \dots, \varepsilon_{t_n^*}].$$

Post-multiplying both sides by  $\text{diag} \left( \left[ \frac{\phi_1}{H_1 \cdot \bar{y}_0}, \dots, \frac{\phi_n}{H_n \cdot \bar{y}_0} \right] \right)$ , recalling that  $\Lambda = \Pi H^{-1}$ , and applying the *vec* operator to both sides, we obtain

$$\text{vec}(\Lambda) | H, \Sigma \sim N \left( 0, \text{diag} \left( \left[ \frac{\phi_1^2}{(H_1 \cdot \bar{y}_0)^2}, \dots, \frac{\phi_n^2}{(H_n \cdot \bar{y}_0)^2} \right] \right) \otimes \Sigma \right),$$

which corresponds to the expression in (3.6).

**A.4. Sum-of-coefficients prior.** The SOC prior of Doan et al. (1984) and Sims and Zha (1998) corresponds to a special case of the PLR, with a mechanical choice of  $H = I_n$ , and hyperparameters  $\phi_1 = \dots = \phi_n = \mu$ .

## APPENDIX B. SETTING OF THE HYPERPARAMETERS

In this appendix we briefly describe the setting of the hyperparameters to generate our empirical results.

We have performed the exercise of section 2 and 5.1 about the shape of the deterministic component using some reference hyperparameter values. In particular, the hyperparameter  $\lambda$  controlling the tightness of the Minnesota prior has been set to 0.2, which is standard. For the hyperparameters  $\{\phi_i\}_{i=1}^n$  of the PLR, we have chosen a value equal to 1, which corresponds to using a single set of dummy observations, with error variance approximately similar to error variance in the actual sample.

For the forecasting exercise of section 5.2, we have adopted a hierarchical interpretation of the model as in Giannone et al. (2014), and set the hyperparameters by maximizing their posterior. The posterior of the hyperparameters is given by the product of the marginal likelihood and the hyperpriors (the prior density on the hyperparameters). Given that our priors are conjugate, the marginal likelihood is available in closed form (see Giannone et al., 2014, and the derivations in their appendix). As priors for the hyperparameters  $\lambda$  and  $\mu$ , we have chosen Gamma densities with mode equal to 0.2 and 1—the values recommended by Sims and Zha (1998)—and standard deviations equal to 0.4 and 1 respectively, as in Giannone et al. (2014). For the hyperparameters of the PLR,  $\{\phi_i\}_{i=1}^n$ , we have also used Gamma densities with mode and standard deviation equal to 1. As argued by Giannone

et al. (2014), an appealing feature of non-flat hyperpriors is that they help stabilize inference when the marginal likelihood happens to have little curvature with respect to some hyperparameters.

### APPENDIX C. DATA

This appendix describes the data series used for the estimation of the 3- and 7-variable VARs. The source of our data is the FRED dataset available on the website of the Federal Reserve Bank of St. Louis, and the sample ranges from 1955Q1 to 2013Q1. The variables entering the 3-variable VARs correspond to the following definitions (series acronym in parenthesis):

- log-real GDP per capita:

$$Y = \log \left[ \frac{\text{Gross Domestic Product (GDP)}}{\text{Population} \cdot \text{GDP Implicit Price Deflator (GDPDEF)}} \right]$$

- log-real consumption per capita:

$$C = \log \left[ \frac{\text{Personal Consumption Expenditure: Nondurable Goods (PCND)} + \text{Services (PCESV)}}{\text{Population} \cdot \text{GDP Implicit Price Deflator (GDPDEF)}} \right]$$

- log-real investment per capita:

$$I = \log \left[ \frac{\text{Gross Private Domestic Investment (GPDI)} + \text{Personal Consumption Expenditures: Durable Goods (PCDG)}}{\text{Population} \cdot \text{GDP Implicit Price Deflator (GDPDEF)}} \right],$$

where the population series used to compute the quantities per capita is the Hodrick-Prescott trend (estimated with smoothing parameter equal to 1600) of the logarithm of the Civilian Noninstitutional Population (CNP16OV) series. The reason to use this smoother population series is to avoid the spikes in the original series that correspond to the census years.

The 7-variable VARs also includes:

- log-hours per capita:

$$H = \log \left[ \frac{\text{Nonfarm Business Sector: Hours of All Persons (HOANBS)}}{\text{Population}} \right]$$

- log-real wages per hour:

$$W = \log \left[ \frac{\text{Nonfarm Business Sector: Compensation Per Hour (COMPNFB)}}{\text{GDP Implicit Price Deflator (GDPDEF)}} \right]$$

- inflation:

$$\pi = \Delta \log [\text{GDP Implicit Price Deflator (GDPDEF)}]$$

- short-term nominal interest rate:

$$R = \frac{\text{Effective Federal Funds Rate (FEDFUNDS)}}{400}.$$

## REFERENCES

- ALTIG, D., L. CHRISTIANO, M. EICHENBAUM, AND J. LINDE (2011): “Firm-Specific Capital, Nominal Rigidities and the Business Cycle,” *Review of Economic Dynamics*, 14, 225–247.
- ATKESON, A. AND L. E. OHANIAN. (2001): “Are Phillips curves useful for forecasting inflation?” *Quarterly Review*, 2–11.
- BAUWENS, L. AND M. LUBRANO (1996): “Identification restrictions and posterior densities in cointegrated Gaussian VAR systems,” in *Advances in Econometrics*, ed. by T. Fomby, JAI Press, vol. 11B, 3–28.
- CHAO, J. C. AND P. C. B. PHILLIPS (1999): “Model selection in partially nonstationary vector autoregressive processes with reduced rank structure,” *Journal of Econometrics*, 91, 227–271.
- CORANDER, J. AND M. VILLANI (2004): “Bayesian assessment of dimensionality in reduced rank regression,” *Statistica Neerlandica*, 58, 255–270.
- D’AGOSTINO, A., D. GIANNONE, AND P. SURICO (2007): “(Un)Predictability and Macroeconomic Stability,” CEPR Discussion Papers 6594, C.E.P.R. Discussion Papers.
- DEL NEGRO, M. AND F. SCHORFHEIDE (2004): “Priors from General Equilibrium Models for VARS,” *International Economic Review*, 45, 643–673.
- (2011): “Bayesian Macroeconometrics,” in *The Oxford Handbook of Bayesian Econometrics*, ed. by G. K. J. Geweke and H. van Dijk, Oxford University Press, vol. 1, chap. 7, 293–389.
- DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, AND R. WOUTERS (2007): “On the Fit of New Keynesian Models,” *Journal of Business & Economic Statistics*, 25, 123–143.

- DOAN, T., R. LITTERMAN, AND C. A. SIMS (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.
- ELLIOTT, G. (1998): "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158.
- ENGLE, R. F. AND C. W. J. GRANGER (1987): "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–76.
- GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN (2004): *Bayesian Data Analysis: Second Edition*, Boca Raton: Chapman and Hall CRC.
- GEWEKE, J. (1996): "Bayesian reduced rank regression in econometrics," *Journal of Econometrics*, 75, 121–146.
- GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2014): "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, forthcoming.
- HORVATH, M. T. AND M. W. WATSON (1995): "Testing for Cointegration When Some of the Cointegrating Vectors are Prespecified," *Econometric Theory*, 11, 984–1014.
- IRELAND, P. N. (2007): "Changes in the Federal Reserve's Inflation Target: Causes and Consequences," *Journal of Money, Credit and Banking*, 39, 1851–1882.
- JOHANSEN, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, no. 9780198774501 in OUP Catalogue, Oxford University Press.
- JUSTINIANO, A., G. E. PRIMICERI, AND A. TAMBALOTTI (2010): "Investment Shocks and Business Cycles," *Journal of Monetary Economics*, 57, 132–145.
- KADIYALA, K. R. AND S. KARLSSON (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, 12, 99–132.
- KARLSSON, S. (2013): "Forecasting with Bayesian Vector Autoregressions," in *Handbook of Economic Forecasting*, ed. by G. Elliot and A. Timmermann, Elsevier, vol. 2 of *Handbook of Econometrics*, chap. 15, 791–897.
- KING, R. G., C. I. PLOSSER, J. H. STOCK, AND M. W. WATSON (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–40.
- KLEIBERGEN, F. AND R. PAAP (2002): "Priors, posteriors and bayes factors for a Bayesian analysis of cointegration," *Journal of Econometrics*, 111, 223–249.
- KLEIBERGEN, F. AND H. K. VAN DIJK (1994): "On the Shape of the Likelihood/Posterior in Cointegration Models," *Econometric Theory*, 10, 514–551.

- KOOP, G., R. STRACHAN, H. VAN DIJK, AND M. VILLANI (2006): “Bayesian approaches to cointegration,” in *The Palgrave Handbook of Theoretical Econometrics*, ed. by T. C. Mills and P. K. Palgrave MacMillan, vol. 1 of *Handbook of Econometrics*, chap. 25, 871–898.
- LITTERMAN, R. B. (1979): “Techniques of forecasting using vector autoregressions,” Working Papers 115, Federal Reserve Bank of Minneapolis.
- MUELLER, U. K. AND M. W. WATSON (2008): “Testing Models of Low-Frequency Variability,” *Econometrica*, 76, 979–1016.
- ROSSI, B. AND T. SEKHOSYAN (2010): “Have economic models’ forecasting performance for US output growth and inflation changed over time, and when?” *International Journal of Forecasting*, 26, 808–835.
- SIMS, C. A. (1996): “Inference For Multivariate Time Series Models With Trend,” Princeton University, mimeo.
- (2000): “Using a likelihood perspective to sharpen econometric discourse: Three examples,” *Journal of Econometrics*, 95, 443–462.
- (2003): “Probability Models for Monetary Policy Decisions,” Mimeo, Princeton University.
- SIMS, C. A. AND T. ZHA (1998): “Bayesian Methods for Dynamic Multivariate Models,” *International Economic Review*, 39, 949–68.
- SMETS, F. AND R. WOUTERS (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 97, 586–606.
- STOCK, J. (2010): “Cointegration in Theory and Practice: A Tribute to Clive Granger,” Tech. rep., ASSA Meeting, presentation Slides.
- STOCK, J. H. AND M. W. WATSON (2007): “Why Has U.S. Inflation Become Harder to Forecast?” *Journal of Money, Credit and Banking*, 39, 3–33.
- STRACHAN, R. W. AND B. INDER (2004): “Bayesian analysis of the error correction model,” *Journal of Econometrics*, 123, 307–325.
- STRACHAN, R. W. AND H. K. VAN DIJK (2005): “Valuing Structure, Model Uncertainty and Model Averaging in Vector Autoregressive Process,” Money Macro and Finance (MMF) Research Group Conference 2005 30, Money Macro and Finance Research Group.
- VILLANI, M. (2000): “Aspects of Bayesian cointegration,” Ph.D. thesis, Stockholm University.

- (2001): “Bayesian prediction with cointegrated vector autoregressions,” *International Journal of Forecasting*, 17, 585–605.
- (2005): “Bayesian Reference Analysis Of Cointegration,” *Econometric Theory*, 21, 326–357.
- WATSON, M. W. (1986): “Vector autoregressions and cointegration,” in *Handbook of Econometrics*, ed. by R. F. Engle and D. McFadden, Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 47, 2843–2915.

FEDERAL RESERVE BANK OF NEW YORK, LUISS UNIVERSITY OF ROME, ECARES AND CEPR  
*E-mail address:* dgiannon2@gmail.com

EUROPEAN CENTRAL BANK AND ECARES  
*E-mail address:* michele.lenza@ecb.int

NORTHWESTERN UNIVERSITY, CEPR, AND NBER  
*E-mail address:* g-primiceri@northwestern.edu