Bayesian Local Projections

Leonardo N. Ferreira*
Central Bank of Brazil

Silvia Miranda-Agrippino†
Bank of England
CfM(LSE) and CEPR

Giovanni Ricco‡
École Polytechnique, University of Warwick
OFCE-Sciences Po and CEPR

April 14, 2023

Abstract

We propose a Bayesian approach to Local Projections that optimally addresses the empirical bias-variance trade-off intrinsic in the choice between direct and iterative methods. Bayesian Local Projections (BLP) regularise LP regressions via informative priors, and estimate impulse response functions that capture the properties of the data more accurately than iterative VARs. BLPs preserve the flexibility of LPs while retaining a degree of estimation uncertainty comparable to Bayesian VARs with standard macroeconomic priors. As regularised direct forecasts, BLPs are also a valuable alternative to BVARs for multivariate out-of-sample projections.

Keywords: Local Projections, VARs, Bayesian Techniques, Impulse Response Functions, Direct Forecasting

JEL Classification: C32; C11; C14.

*leonardo.ferreira@bcb.gov.br Web: https://sites.google.com/view/leonardonferreira
†silvia.miranda-agrippino@bankofengland.co.uk Web: www.silviamirandaagrippino.com
‡G.Ricco@warwick.ac.uk Web: www.giovanni-ricco.com

We thank Gianni Amisano, Fabio Canova, Matteo Ciccarelli, Marco Del Negro, Filippo Ferroni, Raffaella Giacomini, Domenico Giannone, Oscar Jordà, Dimitri Korobilis, Michele Lenza, Ulrich Müller, Giorgio Primiceri, Giuseppe Ragusa, Frank Schorfheide, seminars and conference participants at the NBER SBIES 2016, the ECB Workshop on New Techniques and Applications of BVARs 2016, ESOBE 2016 for helpful comments and suggestions. We are grateful to Patrick Kehoe and Ellen McGrattan for sharing their code. The views expressed in this paper are those of the authors and do not represent those of the Bank of England, the Central Bank of Brazil, or any of their Committees.

1 Introduction

Local Projections (LP, Jordà, 2005) have rapidly become one of the main tools in empirical macroeconomics to study the propagation of structural shocks (see Ramey, 2016). LPs are closely related to multi-step direct forecasts (DF), and consist of estimating a series of predictive regressions at different horizons of a variable of interest on a set of predictors. The coefficients of the different regressions are then 'collated' across horizons to obtain the Impulse Response Functions (IRFs). Contrary to IRFs from Vector Autoregressions (VARs), LPs are semi-parametric in nature, and do not assume a specific underlying model. Potentially, this allows for more flexibility. This flexibility, however, comes at the cost of higher variance and inefficiency of the estimator, relative to VARs.

For stationary data and infinite samples, responses estimated by LP and VAR estimators coincide (see Plagborg-Møller and Wolf, 2021). For unknown data generating process (DGP) and finite samples, the choice between the two methods entails a trade-off. From a classical perspective, choosing between iterative (VAR) and direct (LP) methods for either structural analysis or forecasting involves an empirical trade-off between bias and estimation variance. Iterative methods are more efficient, but are more prone to bias if the model is misspecified.[1] Conversely, direct methods suffer from higher estimation uncertainty, due to serially correlated residuals, and to over-parametrisation in small samples where degrees of freedom quickly dry up at longer horizons. In macroeconomic applications where time-series are short and strongly autocorrelated, the gains afforded by the flexibility of direct methods can be outweighed by the higher estimation uncertainty both in structural applications (see Kilian and Kim, 2011, Brugnolini, 2018 and Li et al., 2021) and in forecasting (see Marcellino et al., 2006, Pesaran et al., 2011, and Chevillon, 2007 for a literature review).

[1]Misspecification is likely to arise along a number of dimensions, e.g. lag order, omitted variables, unmodelled moving average components, time-varying parameters, heteroscedastic residuals, and non-linearities, among others (see discussion in Braun and Mittnik, 1993; Schorfheide, 2005).

We propose a Bayesian Quasi-Maximum Likelihood approach to local projections, with hierarchical informative priors, that optimally addresses this empirical bias-variance trade-off. Intuitively, this methodology, that we refer to as Bayesian Local Projections (BLP), regularises the estimates of LP coefficients via informative priors, while hierarchical modelling allows the data structure to select the optimal degree of departure from the priors at each horizon. The same approach can be used in reduced-form for Bayesian direct forecasting (BDF).

When conducting Bayesian inference on the LP coefficients, a potential tension emerges between the non-parametric nature of the LP approach, and the parametric view that is inherently Bayesian. In LP, the object of inference is the prediction of the variables of interest, conditional on past realisations and possibly on a measure of a structural shock. Hence, rather than on the true parameters of the DGP, LPs conduct inference on the coefficients of the best h-step-ahead conditional linear predictor, under squared loss. Conversely, Bayesian estimation generally requires the specification of a parametric model, i.e. of a joint probability distribution, for both the observables and the coefficients. The posterior distribution is then obtained as the distribution of the coefficients after having observed the data, and is determined by Bayes' rule. This is proportional to the likelihood times the prior – i.e. to the distribution of the observed data (sampling distribution/likelihood function) times the distribution of the coefficients before any data is observed (prior distribution). In a similar vein to LP, we conduct inference on the BLP coefficients by specifying at each horizon an 'artificial' Gaussian likelihood for the data alongside a prior distribution for the projection coefficients. Hence, also for BLP the object of interest are the pseudo-true autoregressive coefficients of a 'misspecified model', i.e. of the h-step-ahead predictive regression.

Because of the serial correlation in the h-step-ahead projection residuals, specifying a Gaussian likelihood leads to underestimating the true variance. We deal with this problem using a sandwich estimator. This approach is grounded in the results of Huber (1967) and White (1982) who showed that, in these cases, the sampling distribution of the MLE is asymptotically centred at the Kullback-Leibler divergence-minimising pseudo-true parameter value and, to first asymptotic order, it is Gaussian with sandwich covariance matrix.[2] This result extends to the asymptotic behaviour of the posterior in misspecified parametric models. Following this intuition, Müller (2013) shows that a superior mode of inference is obtained by using an artificial Gaussian posterior that is centred at the MLE with a sandwich covariance matrix. We follow this approach, and conduct inference on the BLP coefficients based on an artificial Gaussian posterior with a HAC covariance matrix at each horizon. Interestingly, this also matches the frequentist approach of Jordà (2005).

A central problem in Bayesian inference is how to elicit prior distributions that summarise information on the coefficients that is available before any sample is observed. In general, for BLP as well as for VARs, such prior information can be either contained in samples of past data (data-based prior), or it can be elicited from introspection, casual observation, and theoretical models (nondata-based prior). If no prior information is available, researchers can resort to non-informative, or Jeffreys' priors (Geisser, 1965; Tiao and Zellner, 1964). Under non-informative priors, the BLP and LP estimators coincide.

We discuss BLPs under three different priors specifications. The first two are nondata-based. One generalises the standard 'Minnesota' priors of Litterman (1980); Doan, Litterman and Sims (1983).[3] This is a prior often used for Bayesian VARs, and is based on a statistical stylised description of macroeconomic data as independent random walk (or white

[2] In large samples, and under more stringent regularity conditions, the likelihood function converges to a Gaussian distribution, with mean at the MLE and covariance matrix given by the usual MLE estimator for the covariance matrix. This implies that conditioning on the MLE and using its asymptotic Gaussian distribution is, in large samples, approximately equivalent to conditioning on all the data (see discussion in Sims, 2010).

[3] We consider the commonly adopted version of the Minnesota priors as Normal-Inverse-Wishart (NIW) distributions. This is a natural conjugate prior for the likelihood of a VAR with normally distributed disturbances (see Kadiyala and Karlsson, 1997 and Sims and Zha, 1998). Conjugate priors are such that the posterior distribution belongs to the same family as the prior probability distribution.

noise) processes. We refer to this prior as a random-walk (or RW-based) prior.[4] The second nondata-based prior centres the distribution of the dynamic responses to macro shocks around the IRFs of a theoretical economic model (e.g. DSGE). We refer to this prior as a model-based prior. The third type of prior is data-based. This incorporates the widely held belief that the joint dynamics of economic time series are well described – in first approximation, and especially at short horizons – by a VAR. This prior can be formulated as a Normal-Inverse-Wishart (NIW) prior centred around the coefficients of a VAR that is estimated on a pre-sample and iterated at the relevant horizon (VAR-based BLP prior henceforth).

To determine the informativeness of the priors, we adopt a hierarchical approach, and define a second level of prior distributions for the (hyper)parameters that regulate the tightness of prior beliefs (hyperpriors).[5] In doing so, we extend the methodology of Giannone et al. (2015), and treat the overall informativeness of the priors (whether RW-, model-, or VAR-based) as an additional model parameter that is estimated at each horizon as the maximiser of the marginal data likelihood. That is, of the distribution of the data conditional on the hyperparameters, once the model coefficients have been integrated out. We specify the variance of the hyperprior at each horizon as to reflect the intuition that at longer horizons the DGP is more likely to deviate from the stylised data representations incorporated in the priors. An interesting by-product of this approach is that BLP can be seen as a diagnostic tool for the ability of DSGEs and VARs to summarise the dynamic properties of the data.

We study the behaviour of BLP in three settings. First, we conduct two sets of experiments in simulated environments, and compare BLPs across different priors specifications with (i) standard LPs, (ii) Bayesian VARs, and (iii) the Smooth Local Projections (SLPs)

[4]Minnesota priors incorporate a stylised representations of the DGP that is commonly accepted for macroeconomic variables. Hence they are 'statistical priors' and do not incorporate the investigator's 'subjective' beliefs.

[5]This method is also known in the literature as the Maximum Likelihood Type II (ML-II) approach to prior selection (Berger, 1985; Canova, 2007).

of Barnichon and Brownlees (2019).[6] In the first exercise, we simulate data from a version of the medium-scale DSGE of Justiniano et al. (2010). This model admits a finite VAR(5) representation and therefore offers an ideal setting to study how the different methods deal with moderate misspecification. In the second, we consider data generated from the DGP analysed in Chari et al. (2008). This is a model that does not admit a finite VAR representation, and provides us with a setting in which the presence of misspecification is both severe and unavoidable. Results show that BLP is effective at addressing the bias-variance trade-off. BLP-based inference is more accurate than any other method, at the cost of an intermediate bias. BLP outperforms VAR, is as robust as LP when abstracting from the estimation uncertainty of the latter, and faster and more flexible than SLP. Moreover, BLP allows to flexibly incorporate model-based priors about the objects of interest directly and in a straightforward way. This yields considerable improvements relative to using model-based priors in VARs, both in terms of bias reduction and accuracy of the inference.

Second, we compare different BLP priors and estimation methods for IRFs in an empirical application where we study the response of macro aggregates to a Federal Funds rate innovation. Our analysis finds that BLP IRFs tend to imply richer adjustment dynamics following macroeconomic shocks than VAR IRFs, while retaining comparable estimation uncertainty. Moreover, the BLP estimator improves on the efficiency of both LP and SLP. In the application, the RW-based and VAR-based priors lead to essentially identical results.

Finally, we test the BLP framework as a method for Bayesian Direct Forecasting (BDF). We design a multivariate recursive forecasting exercise for quarterly US variables and compare point and density forecasts obtained with classic direct forecasts (DF), Bayesian VARs, and BDF. Out-of-sample BDFs are as accurate as those of a Bayesian VAR, and produce comparable predictive densities. Overall, our analysis shows that BDFs are competitive in

[6] Li, Plagborg-Møller and Wolf (2022) conduct a comprehensive comparison of the performance of alternative methods for the estimation and identification of dynamic causal responses across a large number of DGPs. They find that, on average, the best performing methods are SLPs and Bayesian VARs.

small samples and misspecified models, and that they outperform DF for what concerns estimation uncertainty while retaining equivalent degrees of flexibility.

The paper is organised as follows. In the reminder of this section we discuss the related literature. In Section 2, we introduce Bayesian Local Projections, and discuss the choice of the priors specifications and estimation in Section 3. Section 4 and 5 contain, respectively, the results of our simulation and empirical application, where we compare BLP IRFs across priors and against other methods. The forecasting exercise is in Section 6, and Section 7 concludes. Additional results are reported in the Appendix.

Related Literature Our paper sits at the intersection between the Bayesian VAR and the Local Projection literatures, and merges the non-parametric LP intuition of Jordà (2005) with the Bayesian parametric framework of BVARs (see e.g. Sims, 1980; Doan et al., 1983; Sims and Zha, 1998, among many other contributions). There are several excellent books and survey articles on BVARs. Canova (2007) provides a book treatment of VARs and BVARs in the context of applied macroeconomic research. Del Negro and Schorfheide (2011) have a deep and insightful discussion of BVAR with a broader focus on Bayesian macroeconometrics and DSGE models. Koop and Korobilis (2010) propose a discussion of Bayesian multivariate time series models with an in-depth discussion of time-varying parameters and stochastic volatility. Geweke and Whiteman (2006) and Karlsson (2013) provide a detailed survey with a focus on forecasting with Bayesian Vector Autoregression. Alternatively, one can refer to Miranda-Agrippino and Ricco (2019) that adopt a similar notation to this paper.

Close to the spirit of this paper is the 'Smooth Local Projection' approach of Barnichon and Brownlees (2019) that proposes an alternative method to LP regularisation based on classic regularisation techniques. While the methodology is different, their approach is motivated by the same intuition as our work. Along similar lines, Barnichon and Matthes (2014) have suggested a method to approximate IRFs using Gaussian basis functions.

While presented in a Bayesian language, our approach can also be understood from the alternative frequentist interpretation provided by the theory of 'regularisation' of statistical regressions (see e.g. Chiuso, 2015). In fact, using priors to inform the estimation is equivalent to penalised regressions, such as e.g. Ridge or Lasso (see discussion in De Mol et al., 2008).

Our methodology also builds on the approach of Giannone et al. (2015) to estimating the optimal priors' tightness, and extends it to regression models estimated at different horizons. In taking a Bayesian approach to address the trade-offs between VARs and LPs, our paper provides a practical solution in finite samples to some of the problems discussed in the literature on Local Projections (see, for example Kilian and Kim, 2011 and Brugnolini, 2018). Plagborg-Møller and Wolf (2021) prove the equivalence of the LP and VAR estimator asymptotically, highlighting the empirical nature of the trade-offs that arise when choosing between the two methods.

Finally, this paper is also related to the forecasting literature, where the distinction between LP and VAR-based response functions corresponds to the dichotomy between direct and iterated forecasts (see Marcellino, Stock and Watson, 2006; Pesaran, Pick and Timmermann, 2011; Chevillon, 2007, among others). While direct forecasts are theoretically more appealing because of the added robustness to misspecification, empirically Marcellino et al. (2006) show that iterated forecasts generally outperform direct ones, particularly when long lag lengths are allowed. Direct forecasts tend to have higher sample MSFEs than iterated forecasts, and become increasingly less desirable as the forecast horizon lengthens.

An early application of BLP to the study of monetary policy shocks has appeared in Miranda-Agrippino and Ricco (2021) together with the replication codes. Ho, Lubik and Matthes (2021) include BLP alongside other models in prediction pools designed for the estimation of robust impulse response functions. The BLP methodology is also distributed within the econometric package of Canova and Ferroni (2020).

2 A Bayesian Approach to Local Projections

In this section we introduce the BLP machinery, discuss our Bayesian (Quasi-)Maximum Likelihood approach to estimation, and derive the BLP estimator under conjugate priors. It is worth stressing that while our discussion is proposed in a multivariate setting, it encompasses univariate specifications as a special case.[7]

2.1 A Likelihood Function for LPs

Let $y_t = (y_t^1, \ldots, y_t^n)$ denote an n-dimensional vector of macroeconomic variables. Linear LP estimate the impulse response functions from the sequence of the coefficients of predictive regressions

$$\begin{aligned}
y_{t+1} &= C^{(1)} + B_1^{(1)} y_t + \ldots + B_{\tilde{p}}^{(1)} y_{t-(\tilde{p}+1)} + \varepsilon_{t+1}^{(1)} , \\
y_{t+2} &= C^{(2)} + B_1^{(2)} y_t + \ldots + B_{\tilde{p}}^{(2)} y_{t-(\tilde{p}+1)} + \varepsilon_{t+2}^{(2)} , \\
&\vdots \\
y_{t+H} &= C^{(H)} + B_1^{(H)} y_t + \ldots + B_{\tilde{p}}^{(H)} y_{t-(\tilde{p}+1)} + \varepsilon_{t+H}^{(H)} , \qquad (1)
\end{aligned}$$

where, in principle, $\tilde{p}$ may vary across horizons, and other controls may be present.[8] In this non-parametric approach, the horizon-h IRFs are the coefficients $\widehat{B}_1^{(h)}$, estimated with OLS. Beyond $h = 1$, the residuals $\varepsilon_{t+h}^{(h)}$ are serially correlated and heteroscedastic, being a combination of one-step-ahead forecast errors. The LP estimation procedure therefore typically adopts a 'sandwich' correction to the variance-covariance matrix of the residuals to compute confidence bands.

[7]While LPs are typically represented as single equations, it can be useful in some circumstances to think about the single-equation LP as being part of a multivariate system. This is for example the case when priors for the long-run and for cointegration can be used to better describe the variables' joint dynamics (Giannone et al., 2019). Using a multivariate framework for BLP allows to import such priors in a direct estimation approach.

[8]For ease of exposition, in what follows we fix $\tilde{p} = p \ \forall \ h = 1, \ldots, H$.

The non-parametric character of LP stands in contrast with the parametric nature of Bayesian estimation that requires the specification of a likelihood function. To motivate our approach, we note that at horizon $h = 1$ the LP OLS regressions coincide with the OLS estimation of a linear model, e.g. a VAR. It is well known that under the assumption of Gaussianity of the projection residuals, i.e. if $\varepsilon_{t+1}^{(1)} \sim i.i.d.\mathcal{N}\left(0, \Sigma_{\varepsilon}^{(1)}\right)$, and conditional on the first p observations, the OLS estimator of the regression model in Eq. (1) coincides with the MLE of the conditional likelihood (see e.g. Hamilton, 1994).[9]

Generalising this observation, we think of the LP estimator as equivalent to the MLE obtained from an artificial likelihood for each horizon, under the assumption of Gaussianity and conditional on the first p observations. This approach allows to write a parametric likelihood, and to introduce Bayesian methods and priors for the estimation of 'direct regressions'. Specifically, we propose to think of the likelihood function of the regression model in Eq. (1) as the likelihood functions of a set of misspecified auxiliary models. In the same spirit of LP, the object of interest are not the 'true parameters' of the DGP, but rather the pseudo-true parameters of a 'misspecified model', i.e. of the h-step ahead predictive regression.

The misspecification in the likelihood arises from the assumption around the innovations, which are instead both serially correlated and heteroscedastic. For this reason, the estimator has to be thought of as the Quasi-Maximum Likelihood estimator of a pseudo-true parameter (see White, 1994).[10] For such misspecified models, Huber (1967) and White (1982) show that, asymptotically, the sampling distribution of the MLE is centred at the Kullback-Leibler divergence-minimising pseudo-true parameter value and, to first asymptotic order, it is Gaussian with sandwich covariance matrix.

[9] In the case in which the DGP were a correctly specified Gaussian linear model for $h = 1$, $\varepsilon_{t+h}^{(h)}$ would be a Gaussian MA, and hence a Gaussian process itself.

[10] For example, if we believed the data generating process to be a VAR of order p, the LP regressions would have to be specified as ARMA$(p, h-1)$ regressions. Their coefficients could be then estimated by combining informative priors with a fully specified likelihood (see Chan et al., 2016). If, however, the VAR(p) were to effectively capture the DGP, it would be wise to discard direct methods altogether.

The key advantage of defining an auxiliary albeit misspecified Gaussian likelihood at each horizon is that we can elicit prior distributions over the pseudo-true parameters, and obtain posterior distributions that summarise the information of the data and of the priors. In fact, the intuition of Huber (1967) and White (1982) extends to the asymptotic behaviour of the posterior in misspecified parametric models. From a frequentist point of view, the key observation is that in large samples the likelihood dominates the prior, leading to a Gaussian posterior centred at the MLE and with covariance matrix equal to the inverse of the second derivative of the log-likelihood. Formalising this intuition, Müller (2013) shows that posterior beliefs constructed from a misspecified likelihood such as the one discussed here are unreasonable, in the sense that they lead to inadmissible decisions about the pseudo-true values, and proposes a superior mode of inference – i.e. of asymptotically uniformly lower risk –, based on an artificial Gaussian posterior centred at the MLE with a sandwich covariance matrix.

We use this approach for BLP, and specify an artificial posterior along with a HAC covariance matrix correction. As noted, this also matches the frequentist approach of Jordà (2005), where a HAC-corrected estimator is used to account for the serial correlation of the LP residuals.

2.2 Conjugate Prior Distributions

While many different prior distributions are possible in principle, having specified a Gaussian likelihood makes the choice of conjugate priors from the Normal-inverse Wishart (NIW) family particularly convenient.[11]

For each horizon-h, the set of equations in Eq. (1) can be rewritten in compact form as

$$\mathbf{y}^{(h)} = x\mathbf{B}^{(h)} + \mathbf{e}^{(h)}, \tag{2}$$

[11] Among others, an advantage of these priors is that they can be implemented in a Theil's 'Mixed Estimation' approach, where 'dummy' or pseudo-observations are appended to the data sample and enforce the prior beliefs on the parameters (see Sims, 2005).

where $\mathbf{B}^{(h)} \equiv [B_1^{(h)}, \dots B_p^{(h)}, C^{(h)}]'$ is a $k \times n$ matrix, with $k = np + 1$, and the $(T-h) \times n$ matrices $\mathbf{y}^{(h)}$ and $\mathbf{e}^{(h)}$ and the $(T-h) \times k$ matrix x are defined as

$$\mathbf{y}^{(h)} = \begin{pmatrix} y'_{1+h} \\ \vdots \\ y'_T \end{pmatrix}, \quad x = \begin{pmatrix} x'_1 \\ \vdots \\ x'_{T-h} \end{pmatrix}, \quad \mathbf{e}^{(h)} = \begin{pmatrix} \varepsilon^{(h)\prime}_{1+h} \\ \vdots \\ \varepsilon^{(h)\prime}_T \end{pmatrix}, \tag{3}$$

where $x'_t \equiv \begin{pmatrix} y'_t & \dots & y'_{t-p+1} & 1 \end{pmatrix}$.

Under the assumption of i.i.d. residuals, i.e. $\varepsilon^{(h)}_{t+h} \sim i.i.d.\ \mathcal{N}(0, \Sigma^{(h)}_\varepsilon)$, the Gaussian likelihood, conditional on the parameters and on the first p observations, takes the form

$$\begin{aligned} p\left(y_{1:(T-h)}|\mathbf{B}^{(h)}, \Sigma^{(h)}_\varepsilon, y_{1-p:0}\right) = & \frac{1}{(2\pi)^{(T-h)n/2}}|\Sigma|^{-(T-h)/2} \\ & \times \exp\left\{-\frac{1}{2}tr\left[\Sigma^{(h)^{-1}}_\varepsilon \widehat{S}^{(h)}\right]\right\} \\ & \times \exp\left\{-\frac{1}{2}tr\left[\Sigma^{(h)^{-1}}_\varepsilon \left(\mathbf{B}^{(h)} - \widehat{\mathbf{B}}^{(h)}\right)' x'x \left(\mathbf{B}^{(h)} - \widehat{\mathbf{B}}^{(h)}\right)\right]\right\} \ , \end{aligned} \tag{4}$$

where tr denotes the trace operator, $\widehat{\mathbf{B}}^{(h)}$ is the maximum-likelihood estimator (MLE) of $\mathbf{B}^{(h)}$, and $\widehat{S}^{(h)}$ the matrix of sums of squared residuals, i.e.

$$\widehat{\mathbf{B}}^{(h)} = (x'x)^{-1}x'\mathbf{y}^{(h)}, \qquad \widehat{S}^{(h)} = \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right)' \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right). \tag{5}$$

For each horizon-h regression model we define a generic Inverse-Wishart prior for the variance of the projection residuals, and a conditionally Gaussian prior for the LP coefficients, as follows

$$\Sigma^{(h)}_\varepsilon \sim \mathcal{IW}\left(\Psi^{(h)}_0, d^{(h)}_0\right) \ , \tag{6}$$

$$\beta^{(h)} \mid \Sigma^{(h)}_\varepsilon \sim \mathcal{N}\left(\beta^{(h)}_0, \Sigma^{(h)}_\varepsilon \otimes \Omega^{(h)}_0\right) \ , \tag{7}$$

where $\left(\Psi_0^{(h)}, d_0^{(h)}, \beta_0^{(h)}, \Omega_0^{(h)}\right)$ are the priors' parameters, typically functions of a lower dimensional vector of hyperparameters. $d_0^{(h)}$ and $\Psi_0^{(h)}$ denote, respectively, the degrees of freedom and the scale of the prior Inverse-Wishart distribution for the variance-covariance matrix of the residuals. $\beta_0^{(h)} \equiv vec\left(\underline{\mathbf{B}}^{(h)}\right)$ where $\underline{\mathbf{B}}^{(h)} \equiv \left[\underline{B}_1^{(h)}, \ldots, \underline{B}_p^{(h)}, \underline{C}^{(h)}\right]'$ is the prior mean of the LP coefficients, and $\Omega_0^{(h)}$ acts as a prior on the variance-covariance matrix of the regressors.

The posterior distribution for the BLP coefficients can then be obtained by multiplying the priors by the likelihood of the auxiliary model in Eq. (4), where the autocorrelation of the projection residuals is not taken into account (see Kadiyala and Karlsson, 1997).

Conditional on the data, the posterior distribution takes the following form

$$\Sigma_\varepsilon^{(h)} \,|\, \mathrm{y} \sim \mathcal{IW}\left(\Psi^{(h)}, d\right) \tag{8}$$

$$\beta^{(h)} \mid \Sigma_\varepsilon^{(h)}, \mathrm{y} \sim \mathcal{N}\left(\tilde{\beta}^{(h)}, \Sigma_\varepsilon^{(h)} \otimes \Omega^{(h)}\right) \ , \tag{9}$$

where $d = d_0^{(h)} + (T - h)$ and

$$\begin{aligned}
\Omega^{(h)} &= \left(\Omega_0^{(h)-1} + x'x\right)^{-1}, \\
\tilde{\beta}^{(h)} &\equiv vec\left(\overline{\mathbf{B}}^{(h)}\right) = vec\left(\Omega^{(h)} \ \left(\left(\Omega_0^{(h)}\right)^{-1} \underline{\mathbf{B}}^{(h)} + x'x\widehat{\mathbf{B}}^{(h)}\right)\right), \\
\Psi^{(h)} &= \widehat{\mathbf{B}}^{(h)\prime} x'x\widehat{\mathbf{B}}^{(h)} + \underline{\mathbf{B}}^{(h)\prime}\left(\Omega_0^{(h)}\right)^{-1} \underline{\mathbf{B}}^{(h)} + \Psi_0^{(h)} \\
&\qquad + \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right)'\left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right) - \overline{\mathbf{B}}^{(h)\prime}\left(\left(\Omega_0^{(h)}\right)^{-1} + x'x\right)\overline{\mathbf{B}}^{(h)} \ ,
\end{aligned} \tag{10}$$

where $\overline{\mathbf{B}}^{(h)} \equiv \left[\overline{B}_1^{(h)}, \ldots, \overline{B}_p^{(h)}, \overline{C}^{(h)}\right]'$.

It is important to observe that not having explicitly modelled the autocorrelation of $\varepsilon_{t+h}^{(h)}$ has two important advantages. First, the NIW priors are conjugate, hence the posterior distribution is of the same Normal inverse-Wishart family as the prior probability distribution. Second, the Kronecker structure of the standard macroeconomic priors that allows for

SURE is preserved.[12] These two important properties make the estimation analytically and computationally tractable.

As noted, however, conducting inference about the horizon-h responses from the misspecified posterior in Eq. (9) leads to underestimating the true variance, while still correctly capturing the mean of the distribution of the regression coefficients. We adopt the solution in Müller (2013), which requires 'correcting' the variance by means of a sandwich estimator equation-by-equation. Specifically, we compute the corrected variance as

$$\hat{V}(\beta_i^{(h)}) = (T-h) \times (x'x)^{-1} \left[\hat{\Gamma}_{0,i} + \sum_{l=1}^{L} w_l (\hat{\Gamma}_{l,i} + \hat{\Gamma}'_{l,i}) \right] (x'x)^{-1}$$

where $i = 1, ..., n$ denotes equation i, and $\hat{\Gamma}_{l,i} = \frac{1}{T} \sum_{l+1}^{T-h} x_t \hat{\varepsilon}_{i,t+h}^{(h)} \hat{\varepsilon}_{i,t+h,-l}^{(h)} x'_{t-l}$. $\hat{\varepsilon}_{t+h}^{(h)}$ are the horizon-h projection residuals when $\mathbf{B}^{(h)}$ are set at the mode of the posterior distribution, $w_l = 1 - \frac{l}{L+1}$, and L is the maximum lag used in the autocorrelation correction.

3 Informative Priors for LP

In this section we provide a general framework for the choice of priors for LP coefficients. We first discuss the map between iterative and direct response functions. This is useful to clarify the intuition around how priors over the behaviour of the variables of interest at different horizons can be elicited. Second, we introduce three types of priors for the mean of the BLP coefficients – statistical, data-based, and model-based. Third, we discuss how to define the prior variance. Finally, we provide a method to conduct inference on the hyperparameters of the prior distribution.

[12]Preserving the symmetric structure that results in the Kronecker product is not strictly necessary, but it is helpful from a computational prospective. Carriero, Clark and Marcellino (2019) and Chan (2019) discuss this point and provide efficient computational approaches to implement asymmetric priors that do not preserve the VAR Kronecker structure. Our approach can be easily generalised to asymmetric priors.

3.1 Direct and Iterated Responses

Differently from LP, iterative methods such as VARs recover forecasts and impulse responses by iterating up to the relevant horizon the coefficients of a system of one-step-ahead reduced-form equations. To fix ideas, and without loss of generality, consider the two alternative approaches when $p = 1$:[13]

VAR:

$$
\begin{aligned}
y_{t+1} &= By_t + \varepsilon_{t+1} \\
y_{t+2} &= B^2 y_t + B\varepsilon_{t+1} + \varepsilon_{t+2} \\
&\vdots \\
y_{t+H} &= B^H y_t + B\varepsilon_{t+1} + B^2\varepsilon_{t+2} + \cdots + \varepsilon_{t+H}
\end{aligned}
$$

LP:

$$
\begin{aligned}
y_{t+1} &= B^{(1)} y_t + \varepsilon^{(1)}_{t+1} \\
y_{t+2} &= B^{(2)} y_t + \varepsilon^{(2)}_{t+2} \\
&\vdots \\
y_{t+H} &= B^{(H)} y_t + \varepsilon^{(H)}_{t+H}
\end{aligned}
$$

where B^h denotes the h-th power of the VAR coefficients, and $B^{(h)}$ the LP coefficients of the projection of y_{t+h} on y_t. For given impact effects of the structural shocks, collected in the matrix A_0, the horizon-h impulse response functions from the two methods are given by[14]

$$
\mathrm{IRF}^{\mathrm{VAR}}_h = B^h A_0 \tag{11}
$$

$$
\mathrm{IRF}^{\mathrm{LP}}_h = B^{(h)} A_0 \ . \tag{12}
$$

Three observations are in order. First, conditional on the underlying DGP actually being the VAR model, and abstracting from estimation uncertainty, the IRFs computed with the two methods should coincide (the equivalence holds in general for an infinite sample and with

[13] To simplify the notation, we omit deterministic components and consider a simple VAR(1). However, this is equivalent to a VAR(p) written in VAR(1) companion form.

[14] A_0 identifies the mapping between the structural shocks u_t and the reduced-form one-step-ahead forecast errors, i.e. $\varepsilon_t = A_0 u_t$. We frame the discussion in terms of impulse response functions, but obviously, aside from considerations relative to the identification of A_0, this is equivalent to comparing forecasts produced under the two methods.

unrestricted lag order, see Plagborg-Møller and Wolf, 2021). In particular, the coefficients and residuals of an iterated VAR can be readily mapped into those of LP, yielding

$$B^{(h)} \longleftrightarrow B^{(\mathrm{VAR},h)} = B^h \ , \tag{13}$$

$$\varepsilon_{t+h}^{(h)} \longleftrightarrow \varepsilon_{t+h}^{(\mathrm{VAR},h)} \ = \sum_{j=1}^{h} B^{h-j} \varepsilon_{t+h} \ . \tag{14}$$

This mapping provides a natural bridge between the two empirical specifications, and a roadmap for the specification of priors for LP. Second, conditional on the linear model being correctly specified, LPs are bound to have higher estimation variance due to (strongly) autocorrelated residuals.[15] Third, given that for $h = 1$ VARs and LPs coincide, the identification problem is identical for the two methods. In other words, given an external instrument or a set of theory-based assumptions, the way in which the A_0 matrix is derived from either VARs or LPs coincides.

3.2 Prior mean for BLP coefficients

3.2.1 Statistical Priors

A possible formulation for the prior mean of the LP coefficients is obtained by generalising the standard Minnesota-type priors commonly used in empirical macroeconomics in the context of Bayesian VARs (Litterman, 1980, 1986; Kadiyala and Karlsson, 1997). While not motivated by economic theory, these are computationally convenient priors, and formalise the intuition that most macroeconomic time series are approximated, at first order, by an independent random walk, possibly with drift.[16] These priors 'centre' the distribution of the coefficients in $\mathbf{B}^{(h)}$ at a value that implies an independent random-walk behaviour for all the

[15]Most macroeconomic variables are close to I(1) and even I(2) processes. Hence LP residuals are likely to be strongly autocorrelated.

[16]Based on a commonly accepted stylised representation of the data, these priors help in making the likelihood-based description of the data communicable across researchers with potentially diverse prior beliefs.

elements in y_t

$$y_{j,t} = c + \delta_j y_{j,t-1} + \varepsilon_{j,t} \quad j = 1, \ldots, n. \tag{15}$$

Banbura et al. (2010) suggested setting δ_j to one or zero, depending on whether the variable is thought to be in first approximation a random walk or a stationary process. To frame the discussion in terms of the map in Eq. (13) one can observe that the h-step ahead conditional expectation of the process in Eq. (15) is given by

$$y_{j,t+h|t} = \mathbb{E}[y_{j,t+h}|y_{j,t}] = c \sum_{k=0}^{h} \delta_j^k + \delta_j^h y_{j,t}. \tag{16}$$

Hence, these priors generalise to the case of local projections in a straightforward way, especially so in the cases in which δ_j is either one ($y_{j,t+h|t} = ch + y_{j,t}$) or zero ($y_{j,t+h|t} = c$).

Specifically, the Minnesota priors can be generalised by assuming that, for each horizon-h regression model, the coefficients $B_1^{(h)}, \ldots, B_p^{(h)}$ are a priori independent and normally distributed. The prior is formulated as follows

$$\beta_0^{(h)} = vec\left(\mathbf{B}_{\mathrm{RW}}^h\right), \tag{17}$$

where $\mathbf{B}_{\mathrm{RW}}^h \equiv \left[B_1^{RW}, \ldots, B_p^{RW}, C^{RW}\right]'$. The matrices B_j^{RW}, for $j = 2, \ldots, p$ and C^{RW} are set to zero, while $B_1^{RW} = diag(\delta_1, \ldots, \delta_n)$ with $\delta_j, j = 1 : n$ either zero or one.[17]

It is worth noting that these priors allow to interpret BLP as a frequentist regularised regression. In fact, when all the variables are assumed to be stationary ($\delta_j = 0\ \forall j$) and both the data and priors are assumed to be normally distributed, the regression model corresponds to a frequentist regularised Ridge regression (see De Mol et al., 2008).[18]

[17] In general, δ_i could be between zero and one but, from a practical perspective, such a fine tuning of the priors has little impact on the estimated coefficients for any reasonable value of the tightness parameter.

[18] In a similar manner, one could implement a Lasso penalty on the coefficients of a potentially rich set of controls. This would be equivalent to the double exponential (Laplace) prior used to perform variable selection rather than shrinkage as is the case in Ridge regressions.

3.2.2 Data-based Priors

An interesting alternative to the statistical priors is motivated by the intuition provided by the map in Eqs. (13-14). Using this notion, we can formulate a prior for BLP coefficients that is centred around the coefficients of a VAR with equivalent set of regressors, estimated over a pre-sample of size T_0, and iterated up to the relevant horizon h, as follows

$$\beta_0^{(h)} = vec\left(\mathbf{B}_{\text{VAR}}^h\right), \tag{18}$$

where $\mathbf{B}_{\text{VAR}}^h$ is the h-th power of the autoregressive coefficients of a VAR(p) in y_t estimated over T_0. Such a prior gives weight to the belief that a VAR provides a plausible description of the joint behaviour of economic time series, at least in first approximation.

An appealing property of this formulation for the priors is that it allows us to interpret BLP as effectively spanning the space between VARs and LPs. To see this, note that given Eq. (10) the posterior mean of BLP coefficients under the VAR-based prior takes the form

$$\mathbf{B}_{\text{BLP}}^{(h)} \propto \left(\Omega_0^{(h)^{-1}} + x'x\right)^{-1} \left(\Omega_0^{(h)^{-1}} \mathbf{B}_{\text{VAR}}^h + x'x\widehat{\mathbf{B}}_{\text{LP}}^{(h)}\right). \tag{19}$$

At each horizon h, the relative weight of VAR and LP responses is regulated by $\Omega_0^{(h)}$. As we discuss below, $\Omega_0^{(h)}$ can be written as a function of a single (hyper)parameter $\lambda(h)$ that regulates the overall informativeness of the prior. As in the case of Minnesota priors, when $\lambda(h) = 0$, BLP IRFs collapse into the prior VAR-based IRFs (estimated over T_0). Conversely, if $\lambda(h) \to \infty$ BLP IRFs coincide with those implied by standard OLS LP.

It is worth observing that, in general, BLP IRFs may not necessarily lie between VAR and LP IRFs for two reasons. First, the VAR prior for the BLP coefficients is drawn over a pre-sample whose properties may differ from the estimation sample. Second, note that Eq.

(19) can be rewritten as

$$\mathbf{B}_{\text{BLP}}^{(h)} \propto \left[\mathbb{I}_k + M^{-1}\right]^{-1} \mathbf{B}_{\text{LP}}^{(h)} + \left[\mathbb{I}_k + M\right]^{-1} \mathbf{B}_{\text{VAR}}^{h}, \tag{20}$$

$$= Q\mathbf{B}_{\text{LP}}^{(h)} + (\mathbb{I}_k - Q)\mathbf{B}_{\text{VAR}}^{h}, \tag{21}$$

where $M \equiv x'x\Omega_0^{(h)}$. Each column of $\mathbf{B}_{\text{BLP}}^{(h)}$ refers to a different equation in the system. Since Q is a full matrix, BLP IRFs for variable j at horizon h are not a simple weighed sum of the LP and VAR IRFs for variable j at horizon h with scalar weights, and hence are not restricted to lie in-between them.

3.2.3 Model-based Priors

In several cases, the researcher may instead want to use a model of choice, for example a linear or non-linear DSGE model, to formulate priors (Del Negro and Schorfheide, 2004). This can be fruitful in many ways: to elicit priors for the reduced-form model, to provide posterior inference for the parameters of the DSGE, and, via a prior-hierarchical approach, to evaluate how reasonable the model is as a description of the data.

Using a theoretical model to formulate priors has a distinctive advantage in a BLP setting relative to a VAR. In VARs the IRFs are non-linear functions of the autoregressive coefficients. Hence, using DSGE-based IRFs to inform the inference is not straightforward, and can rapidly exhaust the degrees of freedom of the model (see Canova et al., 2023). In LPs, on the other hand, the projection coefficients are the IRFs (or the reduced form h-step ahead forecasts). This permits imposing priors directly on the object of interest.

From a practical perspective, one can use the model of choice to compute h-step ahead forecasts, or use the model responses to all or some of the shocks to inform priors. It should be noted that when the theoretical IRFs are used to inform the priors, they will contain combinations of the structural parameters, some of which may live in parts of the parameters space for which data are not informative. Hence, along those dimensions the posterior distribution will not be updated by the likelihood of the data.

In practice, for each horizon-h, we inform priors for the relevant elements of the coefficients $B_1^{(h)}, \dots, B_p^{(h)}$ by centring the prior distribution around the model's IRFs. The prior is formulated as follows

$$\beta_0^{(h)} = vec\left(\mathbf{B}_{\mathrm{DSGE}}^h\right), \tag{22}$$

where $\mathbf{B}_{\mathrm{DSGE}}^h \equiv \left[B_1^{DSGE}, \dots, B_p^{DSGE}, C^{DSGE}\right]'$. The matrices $B_j^{DSGE}, j = 2, \dots, p$ and C^{DSGE} are set to zero, and the row of B_1^{DSGE} associated with the identified shock is set to be equal to the model's IRF. This prior can also be combined with the Minnesota prior by setting the diagonal elements of B_1^{DSGE} that are not related to the shock of interest to either one or zero (see Section 3.2.1).

3.3 Prior Variance for BLP coefficients

We specify the prior variance in the same way for all the specifications of the prior mean of the BLP coefficients. For the prior scale $\Psi_0^{(h)}$ in Eq. (6), we follow Doan et al. (1983) and fix it using sample information, as it is common in the literature.[19] Specifically, we set

$$\Psi_0^{(h)} = diag\left(\left[\left(\sigma_1^{(h)}\right)^2, \dots, \left(\sigma_n^{(h)}\right)^2\right]\right), \tag{23}$$

where $\left(\sigma_i^{(h)}\right)^2$ are HAC-corrected variances of univariate local projection residuals for each variable. Similarly, we set $\Omega_0^{(h)}$ to be

$$\Omega_0^{(h)} = \begin{pmatrix} diag([1, \dots, p])^{-\gamma} \otimes \lambda(h)^2 diag\left(\left[\left(\sigma_1^{(h)}\right)^2, \dots, \left(\sigma_n^{(h)}\right)^2\right]\right)^{-1} & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}, \tag{24}$$

[19] Alternatively, these parameters can be considered hyperparameters and estimated with the approach of Giannone et al. (2015).

where ϵ is a very small number, reflecting a very diffuse prior on the intercepts, and $\lambda(h)$ is the key hyperparameter that controls the overall tightness of the priors at each horizon. The hyperparameter γ makes the prior tighter for the coefficients at more distant lags. We set $\gamma = 2$ with the RW-based and model-based prior, and $\gamma = 0$ with the VAR-based prior to account fully for the autoregressive structure of the model. As in Kadiyala and Karlsson (1997), it is convenient to set the prior degrees of freedom of the Inverse-Wishart distribution to $d_0^{(h)} = n + 2$, in order to guarantee the existence of a prior mean for $\Sigma_\varepsilon^{(h)}$, equal to $\Psi_0^{(h)}/(d_0^{(h)} - n - 1)$.

This specification implies the following prior variance for the BLP coefficients at each lag $\ell = 1, \ldots, p$, conditional on a draw for $\Sigma_\varepsilon^{(h)}$

$$\mathbb{V}ar\left[\mathbf{B}_{\mathrm{BLP},\ell ij}^{(h)} \mid \Sigma_\varepsilon^{(h)}\right] = \frac{\lambda(h)^2}{\ell^\gamma} \frac{\Sigma_{\varepsilon,ij}^{(h)}}{\left(\omega_{0,ij}^{(h)}\right)^2} , \tag{25}$$

where $\mathbf{B}_{\mathrm{BLP},\ell ij}^{(h)}$ is the BLP coefficient of variable i in equation j at lag ℓ and horizon h or, equivalently, the coefficient of the forecast for variable i at horizon h. The factor $\Sigma_{\varepsilon,ij}^{(h)}/\left(\omega_{0,j}^{(h)}\right)^2$ accounts for the different scales of variables i and j, and we use $\omega_{0,ij}^{(h)}$ to denote the entries of $\Omega_0^{(h)}$.

3.4 Optimal Prior Tightness: the Choice of $\lambda(h)$

The hyperparameter $\lambda(h)$ can either be set to a specific value, or estimated following a hierarchical Bayes model approach.[20] Treating $\lambda(h)$ as an additional model parameter provides a way to optimally address the empirical bias-variance trade-off that arises when choosing between iterative (RW, VAR) and direct (LP) methods. This requires specifying a second level of prior distributions (or hyperpriors) for $\lambda(h)$, and estimating it as the maximiser of its marginal distribution, conditional on the data and model, as proposed by Giannone et al.

[20]This approach is also known as a Maximum Likelihood Type II (ML-II) approach to prior selection, see Berger (1985), Canova (2007).

(2015) for Bayesian VARs.

Specifically, given a hyperprior and conditional on the data and model, it is possible to estimate $\lambda(h)$ from its marginal distribution $p(\lambda(h)|\mathbf{y}^{(h)}) = p(\mathbf{y}^{(h)}|\lambda(h)) \cdot p(\lambda(h))$, where $p(\mathbf{y}^{(h)}|\lambda(h))$ is the marginal density of the data as a function of the hyperparameters – i.e. $p(\mathbf{y}^{(h)}|\lambda(h)) = \int p(\mathbf{y}^{(h)}|\lambda(h), \theta)p(\theta|\lambda(h))d\theta \ \forall h$ –, and $p(\theta|\lambda(h))$ is the prior distribution of the remaining model parameters $\left(\mathbf{B}_{\mathrm{BLP}}^{(h)} \text{ and } \Sigma_{\varepsilon}^{(h)}\right)$, conditional on $\lambda(h)$.

Extending the argument in Giannone et al. (2015) we provide the intuition for how this procedure addresses the empirical bias-variance trade-off. As shown in Giannone et al. (2015) – derivations are exactly the same – it is possible to analytically rewrite the likelihood in closed form as a function of $\lambda(h)$,

$$p(\mathbf{y}^{(h)}|\lambda(h)) \propto \underbrace{\left|\left(V_{\varepsilon^{(h)}}^{\text{posterior}}\right)^{-1} V_{\varepsilon^{(h)}}^{\text{prior}}\right|^{\frac{T-(\tilde{p}+h)+d}{2}}}_{\text{Fit}} \underbrace{\prod_{t=\tilde{p}+1}^{T-h} \left|V_{t+h|t}\right|^{-\frac{1}{2}}}_{\text{Penalty}} \qquad \forall h \ , \tag{26}$$

where $V_{t+h|t} = \mathbb{E}_{\Sigma_{\varepsilon}^{(h)}}\left[\mathbb{V}ar(y_{t+h}|\mathbf{y}^{t}, \Sigma_{\varepsilon}^{(h)})\right]$ is the variance (conditional on $\Sigma_{\varepsilon}^{(h)}$) of the h-step-ahead forecast of y_t, averaged across all possible a priori realisations of $\Sigma_{\varepsilon}^{(h)}$, and $V_{\varepsilon^{(h)}}^{\text{posterior}}$ and $V_{\varepsilon^{(h)}}^{\text{prior}}$ are the posterior and prior mean of $\Sigma_{\varepsilon}^{(h)}$. The first term in Eq. (26) relates to the model's in-sample fit, and it increases when $V_{\varepsilon^{(h)}}^{\text{posterior}}$ falls relative to $V_{\varepsilon^{(h)}}^{\text{prior}}$. The second term is related to the model's (pseudo) out-of-sample forecasting performance, and it increases in the risk of overfitting (i.e. with either large uncertainty around the parameters' estimates, or large a-priori residual variance). Hence, an ML approach to estimating the hyperparameters would favour values that generate both smaller forecast errors and low forecast error variance, therefore balancing the trade-off between model fit and variance.

As in Giannone et al. (2015), we suggest choosing the hyperprior distribution $p(\lambda(h))$ from a family of Gamma distributions. In setting the parameters of the hyperprior, it is important to observe that while at short horizons a VAR (or RW) prior is likely to be a reasonable approximation to the DGP, over medium horizons the bias introduced by the model misspecification is compounded and grows due to the iteration. In the long run the coefficients have to decline to zero due to the system's stationarity and, before that, the variance of the LP estimator would balance out the bias of the VAR coefficients. Such a reasoning provides the rationale for choosing the scale and shape parameters of the Gamma distribution such that the mode of the distribution is fixed, and the standard deviation increases at each horizon along an 'S'-shaped curve, i.e. a sigmoid. In other words, the standard deviation increases over the horizons before saturating at a fixed value. This allows for larger deviations of the estimator from the priors at longer horizons, while still allowing for regularisations at medium horizons. Specifically, to regulate the variance of the hyperprior we use a shifted logistic function, specified as $\sigma_{\lambda(h)} = \kappa + \frac{\alpha}{1+e^{-\theta(h-h_0)}}$, where κ is the shift, α is the curve's maximum value, h_0 is the value of the sigmoid's midpoint, and θ is the logistic growth rate, or steepness of the curve.

4 Testing BLP in a Simulated Environment

In this section we put BLP to test and analyse its finite-sample performance, across different priors specifications for the projection coefficients, and against IRFs estimated using three alternative methods – (i) standard LPs,[21] (ii) Bayesian VARs with standard Minnesota NIW priors and prior-tightness optimally set as in Giannone et al. (2015), and (iii) the Smooth Local Projections (SLPs) of Barnichon and Brownlees (2019). This latter method smooths out the LP responses by first approximating the projection coefficients using a linear B-splines basis function expansion in the forecast horizon, and then estimating the B-splines

[21] Relative to the original specification of Jordà (2005), we correct the error bands for both autocorrelation and heteroscedasticity as done in more recent works. The same correction is adopted in the BLP specification. Specifically, the Newey-West correction for the confidence bands of both LP and BLP includes $h + 1$ lags.

parameters using a penalised Ridge estimator.[22]

We conduct two set of experiments in a simulated environment. First, we assume that data are generated by the medium-scale DSGE of Justiniano, Primiceri and Tambalotti (2010) – JPT henceforth. The JPT model admits a VAR(5) representation and therefore offers an ideal setting to test different approaches and 'moderately' misspecified models.[23] Second, we consider data generated by the model that Chari, Kehoe and McGrattan (2008) – CKM henceforth – use to discuss the identification of technology shocks on hours worked. This is a model that does not admit a finite VAR representation, and provides us with a natural environment to study the behaviour of BLP and competing methods in the presence of unavoidable misspecification.

In all our applications throughout the paper, we fix the mode of the Gamma hyperprior for $\lambda(h)$ at 0.4 across horizons. We let the hyperprior become more diffuse as the horizon grows by setting the parameters of the logistic function that regulates its variance, $\sigma_{\lambda(h)}$, to $\kappa = 0.1$, $\alpha = 0.4$, $\theta = 0.3$ and $h_0 = 12$. Under this parametrisation, the variance of the hyperprior reaches its maximum at horizons larger than $h = 36$.[24]

[22]A shrinkage coefficient regulates the bias-variance trade-off of the estimator. When the shrinkage coefficient is zero the estimator coincides with the least square estimator. Barnichon and Brownlees (2019) propose to specify the penalty matrix in the Ridge estimator such that when the degree of shrinkage is high SLP coincides with an Almon's polynomial distributed lag model.

[23]A stationary VAR of finite order with a constant mean has an MA representation of infinite order with absolutely summable coefficient matrices. The Wold theorem implies that any subprocess of such a stationary process, consisting of any subset of its components, also admits an MA representation with absolutely summable coefficient matrices (see Lütkepohl, 2005). Hence, a subvector of the variables of the correctly specified VAR will have an infinite VAR representation, with coefficient matrices that converge rapidly to zero as the lag order increases. Therefore, it can be well approximated by a finite order VAR process, that will have a moderate degree of misspecification.

[24]See Figure B.1 in Section B of the Online Appendix. Alternatively, the parameters of the Logistic function could also be treated as additional hyperparameters.

4.1 Simulations 1: the JPT model

In our first experiment, we compare theoretical and empirical IRFs to a monetary policy shock, adopting the setting of Justiniano et al. (2010). The model is modified, as in Giannone et al. (2015), so that the behaviour of the private sector is predetermined relative to the monetary policy rule. This allows for a recursive identification of monetary policy shocks.[25] This medium-scale DSGE admits a theoretical VAR(5) representation.

We simulate 500 datasets from the JPT model of length T = 80, 120, and 240 quarters. We then recover IRFs using a system that only includes output, prices, and the federal funds rate. This is a misspecification that is likely to materialise in practice when the true DGP is unknown, and the researcher used a small information set that is perceived as sufficient.

We then investigate the properties of the confidence/coverage bands and the bias of the IRFs across different priors for BLP, and different methods: LP, BVAR, and SLP. Table 1 reports the average bias of the response functions, and the average length and coverage for the 68% and 95% confidence bands in all cases.[26] Averages are calculated across both variables and horizons up to 20 quarters ahead.[27] As a metric for the bias we follow Forni et al. (2022) and use the sum of the squared errors divided by the sum of the squared

[25]The model parameters are set at their posterior mode, and estimated using quarterly U.S. data from 1965Q1 to 2019Q4 for the seven endogenous variables – output, consumption and investment growth, hours worked, wage and price inflation, and the federal funds rate. Details on data and transformations are reported in the Online Appendix.

[26]To express uncertainty in the location of a (function of) parameters, the frequentist approach uses a ‘confidence interval’ – a stochastic range of values designed to include, prior to observing the sample, the true value of the parameter with some probability. Ex-post the probability is either zero or one since the true parameters are either inside the realisation of the confidence interval or outside. Bayesian coverage intervals are computed after observing the data and represent the ‘a posteriori probability’ that, conditional on a realisation of the data, the parameters are inside a given interval. Here we use ‘confidence’ and ‘coverage’ interchangeably, having to deal with both classic and Bayesian methods, with a slight abuse of the conceptual difference.

[27]Detailed results across horizons are reported in the Online Appendix.

coefficients of the true IRFs:

$$100\frac{\sum_{n=1}^{N}\sum_{h=0}^{H}(\widehat{IRF}-IRF_{true})^2}{\sum_{n=1}^{N}\sum_{h=0}^{H}IRF_{true}^2},$$

where $\widehat{IRF}$ denotes the average IRF across simulations for each method. This ratio is equal to 100 when the estimated IRFs are equal to zero at all horizons, and decreases as the estimated IRFs approach the true ones, such that smaller values denote a smaller bias.

Let us start by comparing results across different priors for BLP, for samples of different length (Panel A in Table 1). For the RW-based priors we centre the distribution of the coefficients of the first own lag around a vector of zeros, consistent with the system being stationary. VAR-based priors are centred around the coefficients of a trivariate BVAR(5) estimated over a pre-sample of 40 observations.[28] Finally, model-based priors are centred around the true IRFs of the model. For this exercise, we fix the lag length for BLP to $p = 5$. Other than being the true lag order in the correctly specified system, this is also a typical choice with quarterly variables.

[INSERT TABLE 1 ABOUT HERE]

The bands of BLP IRFs across priors have similar average length and coverage accuracy. The DSGE prior, unsurprisingly, tends to produce lower bias. BLP IRFs estimated with a VAR prior, on the other hand, come with somewhat higher bias. The optimal prior tightness, $\lambda(h)$, for the three prior types follows the same shape (Figure 1 reports the case T = 240).[29] The optimal informativeness of the priors increases over the shorter horizon to then plateau at medium-long horizons where the bias induced by the prior balance against the increase in variance (see discussion in Section 3). The small jump in between horizons $h = 1$ and

[28]When T = 80 we only use RW and DSGE priors.

[29]For a given prior, the optimal informativeness parameter evolves in a similar way for different sample sizes.

$h = 2$ in the VAR priors is due to the different specification of the prior variance that does not penalise longer lags for $h > 1$ (see Eq. 25).

[INSERT FIGURE 1 ABOUT HERE]

The same metrics allow for a comparison across alternative methods (Panel B of Table 1). The simulations confirm the underperformance of LP in finite samples (see Kilian and Kim, 2011). LP bands tend to have significantly smaller coverage despite their length, also in large samples. On the other hand, they tend to show somewhat lower bias. At the other end of the spectrum, VAR bands are the narrowest, but this comes with potentially severe size distortions also in large samples.[30] SLP is effective at reducing the average width of the error bands of standard LP. However, this comes at the cost of even smaller coverage and higher bias. It is also worth noticing that SLPs have larger computational costs as compared to the other methods.[31] Overall, BLP tends to produce coverage bands that are more accurate than any other method and an intermediate bias, addressing the empirical bias-variance tradeoff.

[INSERT TABLE 2 ABOUT HERE]

A larger degree of misspecification, i.e. setting $p = 2$, does not change the broad picture (Table 2). BVAR's performances are more affected by lag truncation, in terms of both bias and properties of the confidence bands. BLP produces coverage bands that have slightly

[30] For T=240, the BVAR shows a bias smaller than the classical LP, but this is related to the fact that the DGP admits a VAR(5) representation. The BLP using a VAR prior takes advantage of that information, and produces an in-between bias.

[31] For each draw of the Monte Carlo simulations, BLP took 1.5 seconds to run while SLP took 33.1 seconds when T=240 and p=5. The computations were carried out using MATLAB R2022b on an 4-core Dell Inspiron 14 7000 laptop (11th Gen Intel(R) Core(TM) i7) with a 2.80Ghz processor and 16 Gb of RAM.

bigger length when compared to standard LP. Importantly, when the degree of misspecification is exacerbated, inference based on BLP is more accurate than any other method. If anything, BLP bands tend to have a slightly higher coverage.

4.2 Simulations 2: the CKM model

Chari et al. (2008) propose a stylised business cycle model with two shocks: changes in technology, and an orthogonal tax on labour. In the model, only technology shocks permanently affect labour productivity, and the labour wedge effectively accounts for the convolution of all non-technology shocks. This model does not admit a finite VAR representation. Moreover, conditional on the original parametrisation, the coefficients of this infinite-lag VAR decay very slowly. As a consequence, the lag-truncation bias that occurs when estimating finite-order VARs can be very severe.

Chari et al. (2008) set up a simulation study where the model is used to generate artificial data, and finite-order VARs are then used to estimate IRFs to the technology shocks identified with the same long-run restrictions that are implied by the model. We replicate exactly their original setup, and adopt the specification with the first difference of the log of productivity and hours in log levels (Christiano et al., 2006). From the model, we simulate 1,000 artificial datasets with sample length equal to 200 quarterly observations, and then estimate response functions using LP, BVAR, and BLP (Table 3). We fix the number of lags to 4 throughout, following Chari et al. (2008).

[INSERT TABLE 3 ABOUT HERE]

In a first exercise, we centre the BLP prior around the coefficients of a VAR(4) estimated over the first 50 observations, and use a standard Minnesota prior for the BVAR. In this exercise and throughout the paper, the optimal priors' tightness for the BVAR is chosen following Giannone et al. (2015). Both BLP and BVAR improve significantly and to an equivalent degree over LP in terms of bias, and display similar degrees of coverage accuracy at short horizons. However, the coverage of BVAR bands deteriorates as the horizon grows, as also noted in the context of the JPT simulation exercise. In all cases the lag truncation impairs the ability of the methods to recover the true impact matrix. This is reflected in the large value of the bias across methods, which in this case depends predominantly on the short-horizon responses. As discussed, for $h = 1$ BLP and BVAR coincide. Hence, one should not expect BLP to improve relative to the BVAR for what concerns the estimation of the impact matrix for any identification scheme.

In the second exercise, all the IRFs are estimated conditional on the population CKM impact matrix. Moreover, we centre the prior mean of the coefficients of both the BVAR and BLP around the value of the first 4 matrices of autoregressive coefficients of the (infinite-order) VAR representation of the CKM model. The first thing that emerges is that, conditional on the true impact matrix, the bias in standard LP is drastically reduced. The coverage accuracy however remains an issue. The second important result that emerges is that BLP are more successful than BVAR at effectively incorporating a model-based prior. The bias of BLP IRFs is negligible, and orders of magnitude smaller than that of BVAR IRFs. The large BVAR bias in this case is due to the IRFs reverting to zero much more quickly than the population ones. Similar conclusions as above can be drawn for what concerns the accuracy of the posterior credible sets of the two methods.

The broad picture that emerges from the simulation exercises can be summarised as follows. First, BLP is effective at reducing the large estimation uncertainty that characterises standard linear LP. Second, BLP-based inference is more accurate than that of competing methods, particularly as the projection horizon grows. Third, when large samples are available, the choice of the priors for the BLP coefficients has minimal impact on the resulting IRFs. When only a small number of observations are available the RW-based prior provides a more effective way of disciplining the estimates relative to the VAR-based prior. Fourth, BLP allows to flexibly incorporate model-based information about the objects of interest directly and in a straightforward way. This yields considerable improvements relative to using model-based priors in VAR, both in terms of bias reduction and accuracy of the inference.

5 Empirical BLP Response Functions

Which estimation methodology and priors one should prefer is ultimately an empirical question. In this section, we provide an application to our approach by estimating responses of some key macro aggregates to an innovation in the Federal Funds rate. We first compare empirical BLP responses across different prior specifications, and then compare BLP responses with those of LP, SLP, and BVAR.

In all the systems, the vector of endogenous variables includes real GDP, real consumption, real investment, total hours worked, real wages, the GDP deflator and the FFR. The full sample spans the period 1954Q3:2019Q4.[32] With the exception of the policy rate, all variables are expressed in log levels. All the IRFs are normalised such that the impact response of the FFR is equal to 1%. The FFR is ordered last in all cases to align the treatment with the simulation exercises of Section 4. The number of lags is fixed to 5 throughout. We report IRFs to a selection of variables and include the full set of IRFs in the Online Appendix.

[INSERT FIGURE 2 ABOUT HERE]

We start by comparing BLP IRFs under alternative choices for the priors (Figure 2). The solid lines in the figure denote the responses with the VAR-based priors, while markers are used for the RW-based prior and the model-based prior in the top and bottom row of the figure respectively. The model-based prior corresponds to the population IRFs of the JPT model.

[32]Details of the dataset are reported in Section A in the Online Appendix. The observations from 1954Q3 to 1964Q4 are used to initialise the VAR-based BLP prior.

The empirical IRFs are remarkably robust to the choice of the priors used. This points to all the priors providing a reasonable centre for the priors distributions, in line with our discussion in Section 4. Furthermore, the large sample available and the characteristics of the JPT model both concur to mitigate the differences across alternative priors in this case.

In our next exercise, we focus on the VAR-based prior for two main reasons. First, the RW prior may potentially discard important information in the off-diagonal entries of the matrices of autoregressive coefficients that are relevant for the dynamic responses of correlated variables to a shock. Second, as noted in Section 3, the VAR-based prior allows us to interpret BLP-IRFs as spanning the model space between Bayesian VARs and Local Projections. However, the results in Figure 2 show that if the sample length available in the empirical analysis does not permit setting aside some observations to inform the VAR-based prior, the RW and model-based priors provide valuable alternatives.

[INSERT FIGURE 3 ABOUT HERE]

The comparison across different approaches delivers interesting insights. Figure 3 compares BLP responses with those from BVAR (top row), LP (middle row), and SLP (bottom row).[33] Overall, the shape of the IRFs is qualitatively similar across methods. Following a positive innovation in the Federal Funds rate all real variables contract.[34] As is to be expected, inference based on LP responses appear to be less precise – albeit the length of the sample limits the potentially more erratic nature of LP. BLP is effective at reducing estimation uncertainty, in line with our simulation results. VAR and SLP responses are, by construction, the smoothest. As expected, VAR responses also have tighter bands than LP

[33] Robustness exercises are reported in Section D of the Online Appendix: (i) a version of Figure 3 for a sample ending in 2007Q4 to avoid the zero lower bound; and ii) a version of Figure 3 where LP is estimated with the lag order suggested by AIC: 4 lags.

[34] In all cases a pronounced price puzzle emerges, likely pointing to an inability of the standard Cholesky identification to correctly recover monetary policy shocks.

do. This feature, however, also seems to result in VARs estimating stronger and more persistent effects than BLPs (and LPs) do. Conversely, the uncertainty around SLP responses is comparable to that of LP. Conditional on a very similar path for the policy rate response, BLP-IRFs tend to revert to equilibrium faster than VAR-IRFs do, and tend to imply richer adjustment dynamics. This may indicate that some of the characteristics of the responses of the VAR may depend on the dynamic restrictions imposed by the iterative nature of the VAR, rather than being genuine features of the data. Indeed, in the BLP estimates the VAR-prior is optimally loosened as the horizon grows, suggesting that VAR-informed (or equivalently RW) responses tend to be progressively rejected by the data.

[INSERT FIGURE 4 ABOUT HERE]

An important empirical question concerns the behaviour of different approaches over limited spans of data. To this aim, Figure 4 compares BLP-IRFs informed by RW-priors with BVAR-IRFs (top row) and LP-IRFs (bottom row) computed over a set of fixed-length rolling 30-year samples from 1965Q1 to 2019Q1. Starting from 1965Q1, we use 30 years of data to estimate IRFs with the three methods. Then we move forward by one year and repeat the procedure. This yields a total of 25 different subsamples. In the figure we use shaded areas to highlight the space spanned by all the BLP responses. For each variable these are the same in the top and bottom rows of the figure. In the top row, the dash-dotted lines are used for the BVAR-based IRFs across all the subsamples. In the bottom row the dotted lines trace the corresponding LP-IRFs. (Here we do not report estimation uncertainty.) The broad picture that emerges is that BLP-IRFs are remarkably stable across samples, and especially relative to LP. Hence, the regularisation implicit in BLP allows to reduce the estimation uncertainty that is typical of direct methods, and suggests a lower degree of time-variation in the dynamic interaction among macroeconomic variables than those implied by the alternative methods.

6 Bayesian Direct Forecast

The Bayesian methodology presented in the previous sections can be straightforwardly applied in reduced-form to produce Bayesian direct forecasts (BDF). In this section, we compare Bayesian direct forecasts informed by RW priors, with multivariate direct forecasts (DF), and iterated BVARs, as well as with a naive univariate random walk forecast (RWF) which serves as a benchmark.

The design of the recursive forecasting exercise is as follows. The first estimation sample is 1965Q1 to 1990Q1. Out-of-sample forecasts from all the methods are then produced for three forecast horizons equal to 1, 4 and 8 quarters ahead. Observations for 1990Q2 are then added to the estimation sample and the procedure is repeated. The last forecast origin is 2017Q4. This yields a sequence of 112 out-of-sample forecasts over which the performance of each method is evaluated.

Let y_t denote the n-dimensional vector of endogenous variables at t, and $y_{t+h|t}$ its h-step ahead forecast. For each of the methods considered the forecasts are computed as follows:

$$y^{\mathrm{j}}_{T+h|T} = \widehat{\mathbf{B}}^{(h)}_{\mathrm{j}} \mathbf{y}_T \qquad j = \text{DF, BVAR, BDF} \tag{27}$$

where $\mathbf{y}_T \equiv (1, y'_T, y'_{T-1}, \ldots, y'_{T-p+1})'$, $T = 1990Q1, \ldots, 2017Q4$, $p = 5$, $h = 1, 4, 8$, and each of the estimated $\widehat{\mathbf{B}}$ matrices of coefficients is of dimension $n \times (np + 1)$. The random walk forecast is computed as a naive constant-growth forecast.

We evaluate point forecasts by computing root mean squared forecast errors, while we use log-scores for the predictive densities defined as:

$$\text{RMSFE}^j = \sqrt{\frac{1}{N} \sum_{T=90Q1}^{17Q4} \left(y_T - y^j_{T+h|T} \right)^2}, \qquad \text{LS}^j = \frac{1}{N} \sum_{T=90Q1}^{17Q4} \log p \left(y^j_{T+h|T} \right), \tag{28}$$

where $j = \{\text{RWF, DF, BVAR, BDF}\}$, $N = 112$ is the length of the forecast sequence, and $p\left(y^j_{T+h|T} \right)$ denotes the predictive density.

[INSERT TABLE 4 ABOUT HERE]

[INSERT TABLE 5 ABOUT HERE]

The forecasting exercise suggests that, as expected, BDF yields forecasts which have point accuracy comparable to that of both BVARs and DFs (Table 4).[35] As noted, however, the large variance associated with standard direct forecasts makes the predictive densities in this case very wide, which is visible in the large standard deviations in Table 5. It is worth noting that the design of our forecasting exercise tends to downplay the differences among methods due to the estimation sample increasing in size over time. Rolling forecasts computed over fixed-length estimation windows are likely to make the differences starker, as noted in the context of Figure 4. As a consequence, the numbers reported in this section can be thought of as conservative estimates. Nonetheless, they confirm that a Bayesian direct approach is a valuable method also for forecasting purposes.

7 Conclusions

In this paper we have proposed Bayesian Local Projections (or BLP) as a way to address the empirical bias-variance trade-off that is inherent in the choice between iterative (VAR) and direct (LP) methods for both structural analysis and forecasting. Bayesian techniques allow to resolve the empirical dichotomy between VARs and LPs by optimally resolving the standard bias-variance trade-off that is at the heart of the choice between direct and iterated methods.

In setting up a Bayesian Quasi-Maximum Likelihood approach for LP, we suggest the use of different types of informative conjugate priors that can be statistical in nature, data-based or model-based. Hierarchical modelling allows to optimally select the informativeness of the

[35]Point and density forecasts for all variables at all horizons are reported in the Online Appendix.

priors, and the data to optimally deviate from the priors, at each horizon. Such an approach also delivers a natural diagnostics on the priors.

In simulation and with empirical data our approach proves to be competitive. BLP-estimated IRFs are more robust to model misspecification than VAR-based IRFs, but have smaller estimation uncertainty relative to LP-IRFs. This makes them potentially preferable to both methods. BLP-IRFs are also competitive when compared to other approaches to regularise LPs, such as SLP. In a multivariate out-of-sample forecasting exercise, we show that Bayesian direct methods are also a valuable alternative to Bayesian VARs.

References

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010) "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, Vol. 25, No. 1, pp. 71–92.

Barnichon, Regis and Christian Brownlees (2019) "Impulse Response Estimation by Smooth Local Projections," *The Review of Economics and Statistics*, Vol. 101, No. 3, pp. 522–530, July.

Barnichon, Régis and Christian Matthes (2014) "Gaussian Mixture Approximations of Impulse Responses and the Nonlinear Effects of Monetary Shocks," Working Paper 16-8, Federal Reserve Bank of Richmond.

Berger, James O. (1985) *Statistical Decision Theory and Bayesian Analysis*: Springer.

Braun, Phillip A. and Stefan Mittnik (1993) "Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions," *Journal of Econometrics*, Vol. 59, No. 3, pp. 319–341, October.

Brugnolini, Luca (2018) "About Local Projection Impulse Response Function Reliability," CEIS Research Paper 440, Tor Vergata University, CEIS.

Canova, Fabio (2007) *Methods for Applied Macroeconomic Research*: Princeton University Press.

Canova, Fabio and Filippo Ferroni (2020) "A hitchhiker guide to empirical macro models," mimeo, Reserve Bank of Chicago.

Canova, Fabio, Andrzej Kociecki, and Michele Piffer (2023) "Flexible prior beliefs on impulse responses in Bayesian vector autoregressive models." mimeo.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019) "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, Vol. 212, No. 1, pp. 137–154.

Chan, Joshua C. C. (2019) "Asymmetric conjugate priors for large Bayesian VARs," CAMA Working Papers 2019-51, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Chan, Joshua C.C., Eric Eisenstat, and Gary Koop (2016) "Large Bayesian VARMAs," *Journal of Econometrics*, Vol. 192, No. 2, pp. 374–390.

Chari, V.V., Patrick J. Kehoe, and Ellen R. McGrattan (2008) "Are structural VARs with long-run restrictions useful in developing business cycle theory?," *Journal of Monetary Economics*, Vol. 55, No. 8, pp. 1337–1352, November.

Chevillon, Guillaume (2007) "Direct Multi-Step Estimation and Forecasting," *Journal of Economic Surveys*, Vol. 21, No. 4, pp. 746–785.

Chiuso, Alessandro (2015) "Regularization and Bayesian Learning in Dynamical Systems: Past, Present and Future," *CoRR*, Vol. abs/1511.01543.

Christiano, Lawrence J., Martin S. Eichenbaum, and Robert J. Vigfusson (2006) "Assessing structural VARs," International Finance Discussion Papers 866, Board of Governors of the Federal Reserve System (U.S.).

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008) "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, Vol. 146, No. 2, pp. 318 – 328. Honoring the research contributions of Charles R. Nelson.

Del Negro, Marco and Frank Schorfheide (2004) "Priors from General Equilibrium Models for VARS*," *International Economic Review*, Vol. 45, No. 2, pp. 643–673.

——— (2011) "Bayesian Macroeconometrics," in John Geweke, Gary Koop, and Herman Van Dijk eds. *The Oxford Handbook of Bayesian Econometrics*: Oxford University Press, pp. 293–389.

Diebold, Francis X and Roberto S Mariano (1995) "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, Vol. 13, No. 3, pp. 253–263, July.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1983) "Forecasting and Conditional Projection Using Realistic Prior Distributions," NBER Working Papers 1202, National Bureau of Economic Research, Inc.

Forni, Mario, Luca Gambetti, and Giovanni Ricco (2022) "External Instrument SVAR Analysis for Noninvertible Shocks," the warwick economics research paper series (twerps), University of Warwick, Department of Economics.

Geisser, Seymour (1965) "Bayesian Estimation in Multivariate Analysis," *The Annals of Mathematical Statistics*, Vol. 36, No. 1, pp. 150–159, 02.

Geweke, John and Charles Whiteman (2006) "Bayesian Forecasting," in G. Elliott, C. Granger, and A. Timmermann eds. *Handbook of Economic Forecasting*, Vol. 1: Elsevier, 1st edition, Chap. 01, pp. 3–80.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2015) "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, Vol. 2, No. 97, pp. 436–451, May.

——— (2019) "Priors for the Long Run," *Journal of the American Statistical Association*, Vol. 114, No. 526, pp. 565–580.

Hamilton, James D. (1994) *Time Series Analysis*: Princeton University Press, 1st edition.

Ho, Paul, Thomas Lubik, and Christian Matthes (2021) "Averaging Impulse Responses Using Prediction Pools." Indiana University.

Huber, Peter J. (1967) "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp. 221–233, Berkeley, Calif.: University of California Press.

Jordà, Òscar (2005) "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, Vol. 95, No. 1, pp. 161–182, March.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2010) "Investment shocks and business cycles," *Journal of Monetary Economics*, Vol. 57, No. 2, pp. 132–145, March.

Kadiyala, K Rao and Sune Karlsson (1997) "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, Vol. 12, No. 2, pp. 99–132, March-Apr.

Karlsson, Sune (2013) *Forecasting with Bayesian Vector Autoregression*, Vol. 2 of Handbook of Economic Forecasting, Chap. 0, pp. 791–897: Elsevier.

Kilian, Lutz and Yun Jung Kim (2011) "How Reliable Are Local Projection Estimators of Impulse Responses?," *The Review of Economics and Statistics*, Vol. 93, No. 4, pp. 1460–1466, November.

Koop, Gary and Dimitris Korobilis (2010) "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends(R) in Econometrics*, Vol. 3, No. 4, pp. 267–358, July.

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf (2021) "Local Projections vs. VARs: Lessons From Thousands of DGPs." mimeo.

——— (2022) "Local Projections vs. VARs: Lessons From Thousands of DGPs," NBER Working Papers 30207, National Bureau of Economic Research, Inc.

Litterman, Robert B. (1980) "A Bayesian Procedure for Forecasting with Vector Autoregression," working papers, MIT Department of Economics.

Litterman, Robert B (1986) "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience," *Journal of Business & Economic Statistics*, Vol. 4, No. 1, pp. 25–38, January.

Lütkepohl, Helmut (2005) *New introduction to multiple time series analysis*: Springer.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson (2006) "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, Vol. 135, No. 1-2, pp. 499–526.

Miranda-Agrippino, Silvia and Giovanni Ricco (2019) "Bayesian Vector Autoregressions: Estimation," 04.

——— (2021) "The Transmission of Monetary Policy Shocks," *American Economic Journal: Macroeconomics*, Vol. 13, No. 3, pp. 74–107, July.

Müller, Ulrich K. (2013) "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, Vol. 81, No. 5, pp. 1805–1849.

Pesaran, M. Hashem, Andreas Pick, and Allan Timmermann (2011) "Variable selection, estimation and inference for multi-period forecasting problems," *Journal of Econometrics*, Vol. 164, No. 1, pp. 173–187, September.

Plagborg-Møller, Mikkel and Christian K. Wolf (2021) "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, Vol. 89, No. 2, pp. 955–980.

Ramey, V.A. (2016) "Chapter 2 - Macroeconomic Shocks and Their Propagation," Vol. 2 of Handbook of Macroeconomics: Elsevier, pp. 71 – 162.

Schorfheide, Frank (2005) "VAR forecasting under misspecification," *Journal of Econometrics*, Vol. 128, No. 1, pp. 99–136, September.

Sims, Christopher A. (1980) "Macroeconomics and Reality," *Econometrica*, Vol. 48, No. 1, pp. 1–48, January.

——— (2005) "Conjugate Dummy Observation Priors for VARs,"Technical report, Princeton University, mimeo.

——— (2010) "Understanding Non-Bayesians,"Technical report, Princeton University, mimeo.

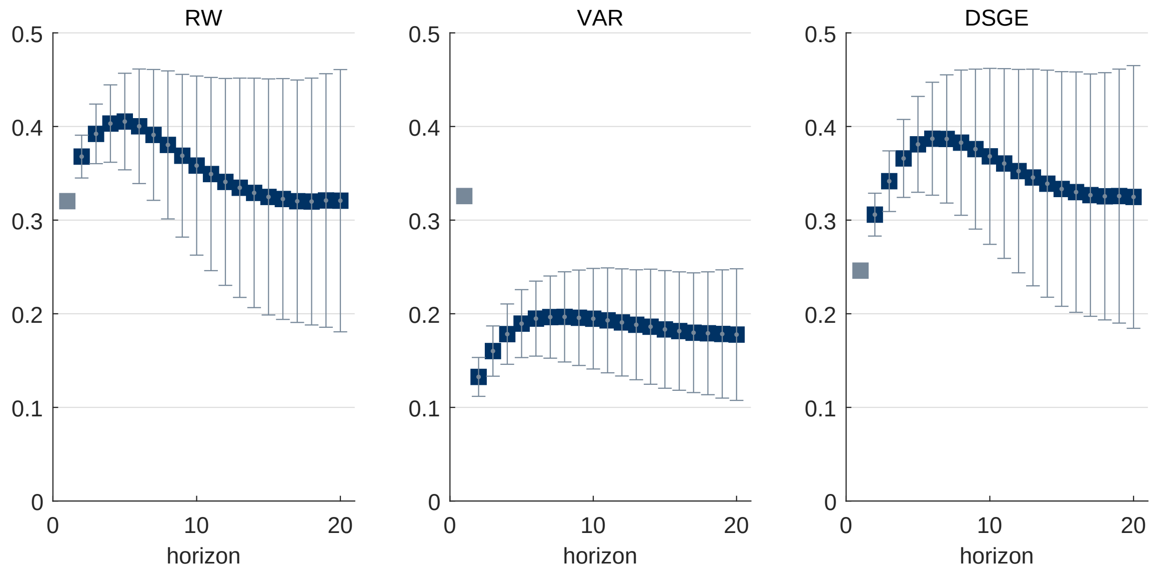
Sims, Christopher A. and Tao Zha (1998) "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, Vol. 39, No. 4, pp. 949–68, November.

Tiao, George C. and Arnold Zellner (1964) "On the Bayesian Estimation of Multivariate Regression," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 26, No. 2, pp. 277–285.

White, Halbert (1982) "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, Vol. 50, No. 1, pp. 1–25.

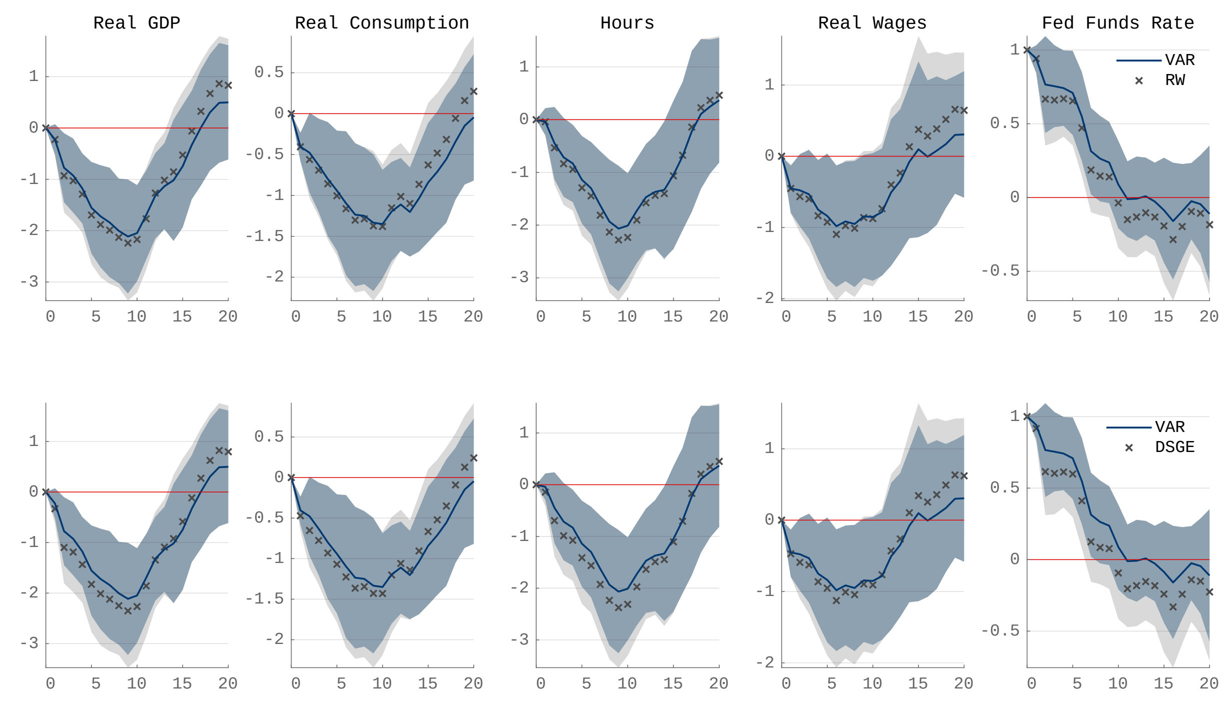
——— (1994) *Estimation, Inference and Specification Analysis*: Cambridge University Press.

Figure 1: Optimal Prior Tightness



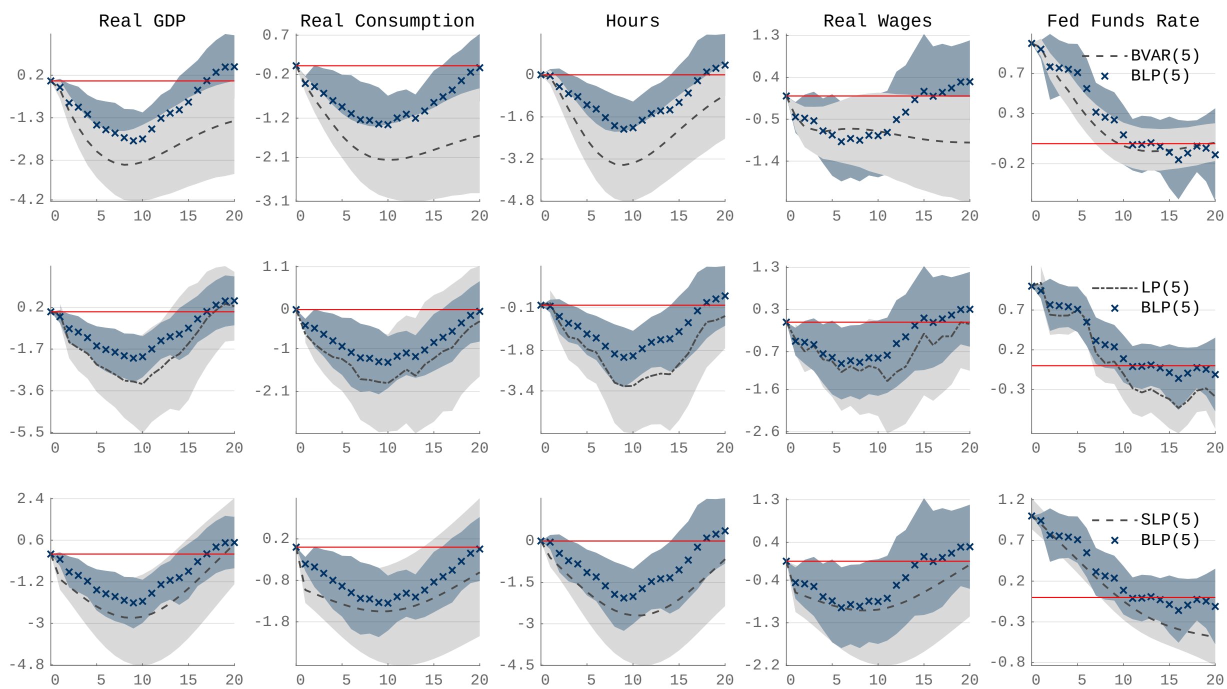
Notes: The first marker is the optimal shrinkage of the Litterman (1986) prior for the BVAR coefficients at $h = 1$ for different prior types. The other markers denote the optimal tightness for different prior types for BLP coefficients for $h > 1$. Average across replications. The error bars are constructed across simulations.

Figure 2: Empirical BLP responses for different prior specifications



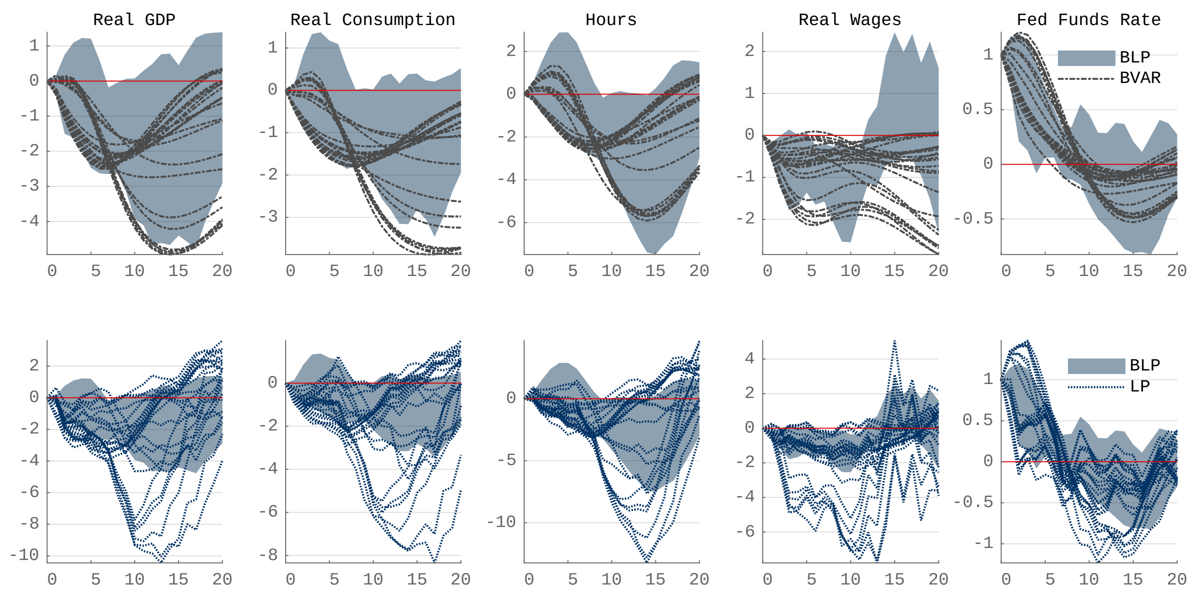
Note: BLP(5) with random walk (RW) prior (markers), and BLP(5) with VAR(5)-based prior (solid line) in the top row. BLP(5) with model-based (DSGE) prior (markers), and BLP(5) with VAR(5) prior (solid line) in the bottom row. Estimation sample: 1965Q1 to 2019Q4. Pre-sample: 1954Q3 to 1964Q4. Shaded areas denote 90% posterior coverage bands.

Figure 3: Empirical IRFs: BVAR, LP and BLP



Note: Impulse response functions to a FFR innovation. Top row: BLP(5) and BVAR(5). Bottom row: BLP(5) and LP(5). Estimation sample: 1954Q3 to 2019Q4. BLP uses 1954Q3 to 1964Q4 as a pre-sample. Shaded areas denote 90% posterior coverage bands.

Figure 4: Stability over Subsamples: BVAR, LP and BLP



Note: BLP(5) with RW prior and BVAR(5) in the top row, and BLP(5) with RW prior and LP(5) in the bottom row. Estimation sample: 1965 to 1995 (first run); 1989 to 2019 (last run).

Table 1: Bias of IRFs and Coverage Accuracy on Simulated Data, JPT

T		Bias	68%		95%	
			L	C	L	C
A. BLP						
Priors						
80	RW	25.4	0.68	0.81	1.34	0.97
	VAR	–	–	–	–	–
	DSGE	17.8	0.68	0.82	1.34	0.97
120	RW	25.6	0.56	0.81	1.11	0.97
	VAR	22.2	0.69	0.86	1.36	0.98
	DSGE	19.4	0.56	0.82	1.11	0.97
240	RW	25.3	0.42	0.77	0.82	0.97
	VAR	22.1	0.46	0.81	0.90	0.97
	DSGE	20.5	0.42	0.78	0.82	0.97
B. Other methods						
80	LP	16.54	0.63	0.49	1.25	0.79
	BVAR	21.01	0.30	0.62	0.71	0.95
	SLP	22.49	0.35	0.36	0.70	0.64
120	LP	19.68	0.55	0.54	1.09	0.84
	BVAR	21.01	0.26	0.58	0.59	0.93
	SLP	26.14	0.35	0.44	0.69	0.74
240	LP	25.44	0.42	0.57	0.82	0.88
	BVAR	21.08	0.21	0.53	0.43	0.87
	SLP	31.69	0.29	0.49	0.58	0.80

Note: The table reports the bias of the IRFs and the average length (L) and coverage (C) across variables and horizons of the 68% and 95% credible sets BLP using different priors (top panel) and methods (bottom panel), estimated with three variables and $p = 5$ for $T = 80$, 120, and 240.

Table 2: Bias of IRFs and Coverage Accuracy on Simulated Data, JPT $p = 2$

T		Bias	68%		95%	
			L	C	L	C
80	LP	22.85	0.64	0.53	1.27	0.83
	BLP	–	–	–	–	–
	BVAR	21.69	0.24	0.49	0.58	0.87
	SLP	28.06	0.38	0.41	0.75	0.71
120	LP	26.73	0.56	0.56	1.11	0.87
	BLP	26.34	0.64	0.84	1.26	0.98
	BVAR	21.41	0.21	0.45	0.47	0.84
	SLP	32.83	0.36	0.45	0.72	0.77
240	LP	33.03	0.43	0.57	0.84	0.88
	BLP	26.72	0.45	0.80	0.89	0.97
	BVAR	21.76	0.16	0.39	0.34	0.74
	SLP	38.87	0.31	0.49	0.61	0.80

Note: The table reports the bias of the IRFs and the average length (L) and coverage (C) across variables and horizons of the 68% and 95% confidence intervals (credible sets) for SLP and LP (BLP and VAR) estimated with three variables and $p = 2$ for $T = 80$, 120, and 240.

Table 3: Bias of IRFs and Coverage Accuracy on Simulated Data, CKM

		Bias	Coverage across horizons						
			1	5	10	15	20	25	30
VAR prior	LP	25.08	0.280	0.634	0.762	0.790	0.804	0.766	0.746
	BLP	17.41	0.954	0.748	0.924	0.960	0.964	0.956	0.954
	BVAR	17.13	0.962	0.960	0.950	0.918	0.908	0.898	0.882
Model prior + True impact	LP	0.72	0.850	0.900	0.846	0.852	0.806	0.774	0.736
	BLP	0.22	0.864	0.988	0.986	0.990	0.972	0.960	0.962
	BVAR	10.01	0.864	0.906	0.808	0.780	0.776	0.776	0.766

Note: The table reports the bias of the IRFs and the average coverage accuracy of the 95% confidence intervals for LP and the 95% credible sets for VAR and BLP for selected horizons.

Table 4: Relative Average RMSFE – Point Forecast

	DF vs RWF			BVAR vs RWF			BDF vs RWF		
	$h = 1$	$h = 4$	$h = 8$	$h = 1$	$h = 4$	$h = 8$	$h = 1$	$h = 4$	$h = 8$
RGDP	1.025	1.249	1.167	0.956	1.030	1.026	0.956	1.074	1.106
	(0.785)	(0.083)	(0.128)	(0.575)	(0.793)	(0.836)	(0.575)	(0.563)	(0.499)
RCON	1.006	0.995	1.163	0.928	0.822	0.974	0.928	0.863	1.070
	(0.928)	(0.969)	(0.141)	(0.301)	(0.137)	(0.834)	(0.301)	(0.291)	(0.681)
RINV	1.061	1.121	0.941	1.031	0.986	0.888	1.031	1.045	0.985
	(0.407)	(0.480)	(0.677)	(0.709)	(0.941)	(0.494)	(0.709)	(0.824)	(0.931)
HOUR	1.134	1.142	0.748	0.995	0.911	0.718	0.995	0.980	0.730
	(0.083)	(0.394)	(0.133)	(0.944)	(0.583)	(0.058)	(0.944)	(0.900)	(0.073)
WAGE	0.685	0.863	0.850	0.658	0.777	0.870	0.658	0.787	0.840
	(0.000)	(0.277)	(0.142)	(0.000)	(0.007)	(0.114)	(0.000)	(0.037)	(0.148)
DEFL	1.112	1.447	1.831	1.023	1.324	1.419	1.023	1.483	1.807
	(0.155)	(0.046)	(0.038)	(0.772)	(0.011)	(0.097)	(0.772)	(0.015)	(0.055)
FFR	2.115	1.503	1.284	1.801	1.078	0.771	1.801	1.088	0.765
	(0.000)	(0.010)	(0.133)	(0.000)	(0.487)	(0.092)	(0.000)	(0.422)	(0.085)

Note: RMSFE. Recursive forecasts for all methods start in 1965Q1, the forecast origins go from 1990Q1 to 2017Q4. DF, BVAR and BDF are all estimated with 5 lags. The p-values of Diebold and Mariano (1995)'s test are reported in parentheses.

Table 5: Log Predictive Scores – Density Forecast

	Relative to DF			Relative to BVAR		
	$h = 1$	$h = 4$	$h = 8$	$h = 1$	$h = 4$	$h = 8$
RGDP	0.042	0.150	0.084	–	0.694	0.967
	(0.283)	(0.543)	(0.81)	–	(0.495)	(0.667)
RCON	0.075	0.141	0.108	–	0.668	0.894
	(0.207)	(0.418)	(0.953)	–	(0.437)	(0.765)
RINV	-0.015	0.070	-0.021	–	0.689	1.074
	(0.256)	(0.639)	(0.709)	–	(0.487)	(0.703)
HOUR	0.095	0.153	0.021	–	0.612	1.089
	(0.248)	(0.555)	(0.778)	–	(0.31)	(0.844)
WAGE	0.040	0.095	0.056	–	0.743	1.008
	(0.221)	(0.592)	(1.252)	–	(0.465)	(0.797)
DEFL	0.073	-0.006	0.014	–	0.572	0.516
	(0.520)	(1.02)	(1.378)	–	(0.487)	(1.133)
FFR	0.109	0.329	0.654	–	0.695	0.765
	(0.223)	(0.615)	(1.219)	–	(0.558)	(0.716)

Note: Log predictive scores. Recursive forecasts for all methods start in 1965Q1, the forecast origins go from 1990Q1 to 2017Q4. DF, BVAR and BDF are all estimated with 5 lags. The standard deviations are reported in parentheses.