WARWICK

ECONOMICS

Bayesian Local Projections

Silvia Miranda-Agrippin & Giovanni Ricco

April 2021 No: 1348

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Bayesian Local Projections

Silvia Miranda-Agrippino*
Bank of England, CfM and CEPR

Giovanni Ricco†
University of Warwick, OFCE-Sciences Po and CEPR

First version: September 30, 2015
Revised: April 19, 2021

Abstract

We propose a Bayesian approach to Local Projections that optimally addresses the empirical bias-variance tradeoff inherent in the choice between VARs and LPs. Bayesian Local Projections (BLP) regularise the LP regression models by using informative priors, thus estimating impulse response functions potentially better able to capture the properties of the data as compared to iterative VARs. In doing so, BLP preserve the flexibility of LPs to empirical model misspecifications while retaining a degree of estimation uncertainty comparable to a Bayesian VAR with standard macroeconomic priors. As a regularised direct forecast, this framework is also a valuable alternative to BVARs for multivariate out-of-sample projections.

Keywords: Local Projections, VARs
JEL Classification: C11; C14.

*silvia.miranda-agrippino@bankofengland.co.uk, www.silviamirandaagrippino.com
†G.Ricco@warwick.ac.uk, www.giovanni-ricco.com

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

1 Introduction

Local Projections (LP, Jordà, 2005), have rapidly become one of the main tools in macroeconomics to study the propagation of structural shocks (see discussion in Ramey, 2016). LPs are closely related to direct multi-step estimation in forecasting, 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 the horizons to form Impulse Response Functions (IRFs). Compared to IRFs obtained from estimated Vector Autoregressions (VARs), LPs are semi-parametric in nature, and do not assume a specific model. As a consequence, they potentially allow for more flexibility. This flexibility, however, comes at the cost of higher variance and inefficiency of the estimator, relative to VARs.

From a classical perspective, choosing between iterated (VAR) and direct (LP) methods for either structural analysis or forecasting involves an empirical trade-off between bias and estimation variance: iterated 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).

In this paper 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 through informative priors, while hierarchical modelling allows the data structure to select the optimal degree

[1] A VAR model is likely to be misspecified 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).

of departure form the priors at each horizon.

When conducting Bayesian inference on the local projection coefficients there is potentially a tension 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 their past realisations, and possibly on a measure of a structural shock. Hence, rather than on the true parameters of the process generating the data, LPs conduct inference on the coefficients of the best h-step ahead conditional linear predictor, under squared loss.

Bayesian estimation generally requires the specification of a parametric model, i.e. of a joint probability distribution – often Gaussian in macroeconomic applications –, for both the observables and the parameters. The posterior distribution is then obtained as the distribution of the parameters after having observed the data, and is determined by the Bayes' rule. This is proportional to the likelihood times the prior – i.e. the product of the distribution of the observed data (sampling distribution/likelihood function) times the distribution of the parameters before any data is observed (prior distribution).

In a similar vein to LP, to conduct inference on the BLP coefficients we follow a quasi-maximum likelihood approach, and at each horizon specify a Gaussian likelihood function for the data, in conjunction with different specifications for the priors. In other words, also for BLP the object of interest are the pseudo-true autoregressive coefficients of a 'misspecified model', i.e. a of an h-step ahead regression model.

Because of the serial correlation in the residuals of the h-step-ahead regressions, specifying a Gaussian likelihood leads to underestimating the true variance; a problem we deal with using a sandwich estimator for the variance. This approach is grounded in the results of Huber (1967) and White (1982) who showed that in in these cases, the sampling distribution of the MLE is asymptotically centred around 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

[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). Similarly in Bayesian contexts, in large samples the likelihood dominates

behaviour of the posterior in misspecified parametric models. Following this intuition, Müller (2013) shows that a superior mode of inference is obtained in these cases by using an 'artificial' Gaussian posterior that is centred at the MLE with a sandwich covariance matrix.[3] In this work, we follow this approach, and conduct inference on the BLP coefficients based on artificial Gaussian posteriors 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 probability distributions that summarise information on the parameters 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, a researcher can resort to 'non-informative', or Jeffreys' priors (Geisser, 1965; Tiao and Zellner, 1964).[4] Under non-informative priors, the BLP and LP estimators coincide.

We discuss BLPs under two different priors specifications. The first one is a nondata-based statistical prior based on the 'Minnesota' priors of Sims and Zha (1998). Minnesota priors, widely used for Bayesian VARs, assume that macroeconomic time series can be a-priori represented as independent random walks or white noise processes.[5] These can be readily generalised to h-step-ahead regression models, since for both random walk and white noise processes the h-step-ahead and 1-step-ahead conditional expectations coincide. We refer to this prior as a random-walk (or RW-based) BLP prior. The second type of prior is instead data-based, and follows from the widely held belief

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.

[3] Müller (2013) shows that, conversely, 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.

[4] Jeffreys priors are proportional to the square root of the determinant of the Fisher information matrix, and are derived from Jeffreys' 'invariance principle', meaning that the prior is invariant to re-parameterization (see Zellner, 1971). These priors are designed to extract the maximum amount of expected information from the data. They maximise the difference (measured by Kullback-Leibler distance) between the posterior and the prior when the number of samples drawn goes to infinity.

[5] Minnesota priors incorporate a stylised representations of the DGP that is commonly accepted for economic variables. Hence they are 'statistical priors' and do not incorporate the investigator's 'subjective' beliefs. A such, they help in making the likelihood-based description of the data communicable across researchers with potentially diverse prior beliefs.

that the joint dynamic properties 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).

In determining the informativeness of the priors, we adopt a hierarchical approach, and define a second level of prior distributions for the parameters that regulate the tightness of prior beliefs (hyperpriors).[6] In doing so, we extend the methodology of Giannone et al. (2015), and treat the overall informativeness of the priors (either RW- or VAR-based) as an additional model parameter that is estimated at each horizon as the maximiser of the marginal data likelihood, i.e. 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 true DGP is more likely to deviate from the stylised data representation incorporated in the priors. An interesting by-product of this approach is that, in the case of the VAR-based priors, the posterior mean of BLP coefficients can be seen as an optimally weighted combination of VAR and LP coefficients at each forecast horizon/projection lag.

We study the behaviour of BLP in three settings. First, we compare empirical IRFs estimated on quarterly US data with LP, a Bayesian VAR with standard Normal-Inverse Wishart priors, and BLP with both a RW-based and a VAR-based prior. Our analysis finds that the data tends to deviate from the stylised priors at longer horizons, resulting in an optimal level of prior shrinkage that is a monotonic non-decreasing function of the forecast horizon, or projection lag. BLP IRFs tend to imply richer adjustment dynamics following macroeconomic shocks than VAR IRFs, while retaining comparable estimation uncertainty. BLP-IRFs estimated with RW-based and VAR-based priors lead to very similar results.

Second, we evaluate BLP's ability to recover accurate response functions in a simu-

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

lated environment where we can control the degree of empirical model misspecification relative to a well defined benchmark. The reference model is a variant of the DSGE in Justiniano et al. (2010). We compare DSGE-implied dynamic responses to a monetary policy shock with those estimated with standard LP, BLP with a VAR-based prior, and a Bayesian VAR with standard NIW priors, all misspecified to varying degrees. In dealing with misspecification, BLP outperform VARs, and are as robust as LP when abstracting from the estimation uncertainty of the latter.

Finally, we test the framework designed to estimate BLP as a direct forecasting method. We design a multivariate recursive forecasting exercise for quarterly US variables and compare the three methods in terms of both point and density forecasts. BLP out-of-sample forecasts are as accurate as those of a Bayesian VAR, and produce comparable predictive densities. Overall, our analysis shows that BLPs are competitive in small samples and misspecified models, and that they outperform LPs 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. Section 3 discusses the choice of the priors specifications and estimation. In Section 4, we present the BLP IRFs in an empirical setting using quarterly US data under different priors, and we compare them with those estimated with frequentist LP and a Bayesian VAR with standard priors. We evaluate the the BLP method in a simulated environment in Section 5, and in a forecasting exercise is in Section 6. 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, among many others contribution, Sims, 1980; Doan et al., 1983; Sims and Zha, 1998). There are several excellent books and survey articles on BVARs. Canova (2007) provides a book treatment of VARs and BVARs in the context of the methods for 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 models. 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 propose an alternative method to LP regularisation based on classic regularisation techniques. While the methodology is different, their approach is motivated by the the same intuition as our work. Whether one approach or the other may be used would depend on the application at hand and the researcher's preference. Along similar lines, Barnichon and Matthes (2014) have suggested a method to approximate IRFs using Gaussian basis functions.

Our approach, while presented in a Bayesian language, can also be understood from the alternative frequentist interpretation provided by the theory of 'regularisation' of statistical regressions (see, for example, Chiuso, 2015). In fact, the use of priors to inform estimation is equivalent to a penalised regression, as it would be the case in a Ridge or Lasso regression (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 (forthcoming) 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. First we provide the intuition behind our approach by analysing the relationship between VAR-IRFs and LP-IRFs in a simplified setting. We then 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.

2.1 Direct vs Iterated Forecasts and Response Functions

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

$$y_{t+1} = By_t + \varepsilon_{t+1} \qquad \varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon) \ , \tag{1}$$

where $y_t = (y_t^1, \ldots, y_t^n)'$ is a $(n \times 1)$ random vector of macroeconomic variables, B is an n-dimensional matrix of coefficients, and ε_t is an $(n \times 1)$ vector of reduced-form

innovations, or one-step-ahead forecast errors.[7]

Conversely, direct methods such as LPs recover these objects from the coefficients of a set of linear regressions estimated independently at each horizon and of the form

$$y_{t+h} = B^{(h)} y_t + \varepsilon_{t+h}^{(h)} \ , \qquad \varepsilon_{t+h}^{(h)} \sim \mathcal{N}\left(0, \Sigma_{\varepsilon}^{(h)}\right) \qquad \forall \ h = 1, \ldots, H \ . \tag{2}$$

Being a combination of one-step-ahead forecast errors, the projection residuals $\varepsilon_{t+h}^{(h)}$ are serially correlated and heteroscedastic (Jordà, 2005).

The horizon-h impulse response functions from the two methods are given by

$$\mathrm{IRF}_h^{\mathrm{VAR}} = B^h A_0 \ , \tag{3}$$

$$\mathrm{IRF}_h^{\mathrm{LP}} \ = B^{(h)} A_0 \ , \tag{4}$$

where A_0 identifies the mapping between the structural shocks u_t and the reduced-form one-step-ahead forecast errors, i.e. $\varepsilon_t = A_0 u_t$.[8] Assuming the VAR to be the true description of the data generating process, the coefficients and residuals of an iterated VAR can be readily mapped into those of LP, yielding

$$B^{(h)} \longleftrightarrow B^{(\mathrm{VAR},h)} = B^h \ , \tag{5}$$

$$\varepsilon_{t+h}^{(h)} \longleftrightarrow \varepsilon_{t+h}^{(\mathrm{VAR},h)} \ = \sum_{j=1}^{h} B^{h-j} \varepsilon_{t+h} \ . \tag{6}$$

Three observations are in order. First, conditional on the underlying DGP being the linear model in Eq. (1), and abstracting from estimation uncertainty, the IRFs computed with the two methods should coincide (Eq. 5, see also Plagborg-Møller and Wolf, 2021). Second, as shown by Eq. (6), conditional on the linear model being correctly specified, LPs are bound to have higher estimation variance due to (strongly) autocorrelated

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

[8] We frame the discussion in terms of impulse response functions, but it is understood that aside from considerations relative to the identification of A_0, this is equivalent to comparing forecasts produced under the two methods.

residuals.[9] 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. The map in Eqs. (5-6) provides a natural bridge between the two empirical specifications.

2.2 A Likelihood Function for LPs

Horizon-h LP-IRFs obtain from the OLS estimates, denoted $\widehat{B}_1^{(h)}$, of the coefficients of the linear regression

$$y_{t+h} = C^{(h)} + B_1^{(h)} y_t + ... + B_{\tilde{p}}^{(h)} y_{t-(\tilde{p}+1)} + \varepsilon_{t+h}^{(h)} \ , \qquad \forall \ h = 1, \ldots, H \ , \tag{7}$$

where, in principle, $\tilde{p}$ may vary across horizons. For ease of exposition, in what follows we fix $\tilde{p} = p \ \forall \ h = 1, \ldots, H$. It is well known that under the assumption of Gaussianity of the projection residuals, i.e. if $\varepsilon_{t+h}^{(h)} \sim i.i.d.\mathcal{N}\left(0, \Sigma_\varepsilon^{(h)}\right)$, and conditional on the first p observations, the OLS estimator of the regression model in Eq. (7) coincides with the MLE of the conditional likelihood (see e.g. Hamilton, 1994).[10] This observation allows us to think of the OLS LP estimator as equivalent to the MLE for Gaussian likelihood function.

As noted, however, the residuals $\varepsilon_{t+h}^{(h)}$ in Eq. (7) are a combination of one-step-ahead forecast errors, and are thus serially correlated and heteroscedastic. Therefore, a Gaussian likelihood function with i.i.d. errors for the horizon h regression model is misspecified, and the estimator should instead be thought of as a Quasi-Maximum Likelihood estimator (see White, 1994).[11] Huber (1967) and White (1982) show that, asymptotically, in such misspecified models the sampling distribution of the MLE is

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

[10] 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.

[11] 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.

centred on the Kullback-Leibler divergence-minimising pseudo-true parameter value and, to first asymptotic order, it is Gaussian with sandwich covariance matrix. We use this observation to characterise our proposed Bayesian framework for Local Projections. In fact, we will be thinking of the likelihood function of the regression model in Eq. (7) as the likelihood of a misspecified auxiliary model. Hence, and in the spirit of LP, the object of interest will not be the 'true parameters' of the DGP, but rather the pseudo-true parameters of a 'misspecified model', i.e. of the h-step ahead regression model.

This observation is important because it allows us to formally introduce priors for the LP coefficients. The key advantage of defining an auxiliary (albeit misspecified) Gaussian likelihood at each horizon is that, as it is well known, the intuition of Huber (1967) and White (1982) extends to the asymptotic behaviour of the posterior in misspecified parametric models. 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 as 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 posteriors with a HAC covariance matrix. As noted, this also matches the frequentist approach of Jordà (2005) where an HAC-corrected estimator is used to account for the serial correlation of the LP residuals.

2.3 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.

For each horizon-h, the model in Eq. (7) can be rewritten in compact form as

$$\mathbf{y}^{(h)} = x\mathbf{B}^{(h)} + \mathbf{e}^{(h)}, \tag{8}$$

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

$$\mathbf{y}^{(h)} = \begin{pmatrix} y'_{1+h} \\ \vdots \\ y'_T \end{pmatrix}, \quad x = \begin{pmatrix} x'_1 \\ \vdots \\ x'_{T-h} \end{pmatrix}, \quad \mathbf{e}^{(h)} = \begin{pmatrix} \varepsilon^{(h)\prime}_{1+h} \\ \vdots \\ \varepsilon^{(h)\prime}_T \end{pmatrix}, \tag{9}$$

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 following form

$$\begin{aligned} p\left(y_{1:(T-h)} | \mathbf{B}^{(h)}, \Sigma^{(h)}_\varepsilon, y_{1-p:0}\right) = & \frac{1}{(2\pi)^{(T-h)n/2}} |\Sigma|^{-(T-h)/2} \\ & \times \exp\left\{-\frac{1}{2} tr\left[\Sigma^{(h)-1}_\varepsilon \widehat{S}^{(h)}\right]\right\} \\ & \times \exp\left\{-\frac{1}{2} tr\left[\Sigma^{(h)-1}_\varepsilon \left(\mathbf{B}^{(h)} - \widehat{\mathbf{B}}^{(h)}\right)' x'x \left(\mathbf{B}^{(h)} - \widehat{\mathbf{B}}^{(h)}\right)\right]\right\} , \end{aligned} \tag{10}$$

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.

$$\widehat{\mathbf{B}}^{(h)} = (x'x)^{-1} x' \mathbf{y}^{(h)}, \qquad \widehat{S}^{(h)} = \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right)' \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right). \tag{11}$$

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

$$\Sigma^{(h)}_\varepsilon \sim \mathcal{IW}\left(\Psi^{(h)}_0, d^{(h)}_0\right) , \tag{12}$$

$$\beta^{(h)} \mid \Sigma^{(h)}_\varepsilon \sim \mathcal{N}\left(\beta^{(h)}_0, \Sigma^{(h)}_\varepsilon \otimes \Omega^{(h)}_0\right) , \tag{13}$$

where $\left(\Psi^{(h)}_0, d^{(h)}_0, \beta^{(h)}_0, \Omega^{(h)}_0\right)$ are the priors' parameters, typically functions of a lower dimensional vector of hyperparameters $\gamma^{(h)}$. $d^{(h)}_0$ and $\Psi^{(h)}_0$ denote, respectively, the de-

grees 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. (10), 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

$$\Sigma_\varepsilon^{(h)} \mid \mathrm{y} \sim \mathcal{IW}\left(\Psi^{(h)}, d\right) \tag{14}$$

$$\beta^{(h)} \mid \Sigma_\varepsilon^{(h)}, \mathrm{y} \sim \mathcal{N}\left(\tilde{\beta}^{(h)}, \Sigma_\varepsilon^{(h)} \otimes \Omega^{(h)}\right) \ , \tag{15}$$

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

$$\begin{aligned}
\Omega^{(h)} &= \left(\Omega_0^{(h)^{-1}} + x'x\right)^{-1}, \\
\tilde{\beta}^{(h)} &\equiv vec\left(\overline{\mathbf{B}}^{(h)}\right) = vec\left(\Omega^{(h)} \ \left(\left(\Omega_0^{(h)}\right)^{-1} \underline{\mathbf{B}}^{(h)} + x'x\widehat{\mathbf{B}}^{(h)}\right)\right), \\
\Psi^{(h)} &= \widehat{\mathbf{B}}^{(h)'} x'x\widehat{\mathbf{B}}^{(h)} + \underline{\mathbf{B}}^{(h)'} \left(\Omega_0^{(h)}\right)^{-1} \underline{\mathbf{B}}^{(h)} + \Psi_0^{(h)} \\
&\quad + \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right)' \left(\mathbf{y}^{(h)} - x\widehat{\mathbf{B}}^{(h)}\right) - \overline{\mathbf{B}}^{(h)'} \left(\left(\Omega_0^{(h)}\right)^{-1} + x'x\right) \overline{\mathbf{B}}^{(h)} \ ,
\end{aligned} \tag{16}$$

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

[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.

estimation analytically and computationally tractable.

However, as noted, this comes at the cost of underestimating the true variance. The shape of the true likelihood is asymptotically Gaussian and centred at the MLE, but has a different (larger) variance than the misspecified posterior distribution in Eqs. (14-15). This implies that if one were to conduct inference about the horizon-h responses from the misspecified posterior distribution in Eq. (15), one would be underestimating the variance albeit correctly capturing the mean of the distribution of the regression coefficients. As discussed, the solution to this issue is provided by Müller (2013), and requires 'correcting' the variance by means of a sandwich estimator. Hence, and similarly to the frequentist practice, we conduct inference on $\beta^{(h)}$ by replacing the original posterior with an artificial Gaussian posterior that is centred at the MLE but with a HAC-corrected covariance matrix, as follows:

$$\Sigma^{(h)}_{\varepsilon,\text{HAC}} \,|\text{y} \sim \mathcal{IW}\left(\Psi^{(h)}_{\text{HAC}}, d\right) , \tag{17}$$

$$\beta^{(h)} \mid \Sigma^{(h)}_{\varepsilon,\text{HAC}}, \text{y} \sim \mathcal{N}\left(\tilde{\beta}^{(h)}, \Sigma^{(h)}_{\varepsilon,\text{HAC}} \otimes \Omega^{(h)}\right) , \tag{18}$$

where $\Psi^{(h)}_{\text{HAC}} = \Psi^{(h)} + \sum_{j=1}^{h-1}\left[1 - \frac{j}{h}\right]\left(\widehat{\Gamma}^{(h)}_j + \widehat{\Gamma}^{(h)\prime}_j\right)$ and $\widehat{\Gamma}^{(h)}_j$ is the sample autocovariance of the projection residuals at horizon h. This allows us to remain agnostic about the source of model misspecification as in Jordà (2005). A similar adjustment is used for the prior scale of the Inverse-Wishart distribution in Eq. (12) (see Section 3.3).

3 Informative Priors for LPs

3.1 Nondata-based Minnesota-type 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 well approximated

by an independent random walk with drift. Hence, this prior 'centres' the distribution of the coefficients in $\mathbf{B}^{(h)}$ at a value that implies an independent random-walk behaviour for all the elements in y_t

$$y_{j,t} = c + \delta_j y_{j,t-1} + \varepsilon_{j,t} \quad j = 1, \ldots, n. \tag{19}$$

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. The h-step ahead conditional expectation of the process in Eq. (19) is given by

$$y_{j,t+h|t} = \mathbb{E}[y_{j,t+h}|y_{j,t}] = c\sum_{k=0}^{h} \delta_j^k + \delta_j^h y_{j,t}, \tag{20}$$

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 ($y_{j,t+h|t} = c$).

For LPs, the Minnesota priors can be generalised to assume 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

$$\beta_0^{(h)} = vec\left(\mathbf{B}_{\text{RW}}^h\right), \tag{21}$$

where $\mathbf{B}_{\text{RW}}^h \equiv \left[B_1^{RW}, \ldots, B_p^{RW}, C^{RW}\right]$. The matrices $B_j^{RW}, j = 2, \ldots, p$ and C^{RW} are set to zero, whereas $B_1^{RW} = diag(\delta_1, \ldots, \delta_n)$ where $\delta_j, j = 1 : n$ are either zero or one.[13]

The priors' tightness depends on $\Omega_0^{(h)}$, which we in turn specify to be a function of a hyperparameter $\lambda(h)$ that regulates the overall informativeness of this prior. If $\lambda(h) = 0$ the prior information dominates, and system reduces to a vector of univariate models. Conversely, as $\lambda(h) \to \infty$ the prior becomes less informative, and the posterior mostly mirrors sample information. As it is well known, these priors can be implemented with 'dummy' or pseudo-observations with properties specified by the prior beliefs on the

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

VAR parameters (see Sims, 2005).

It is important to stress that the use of these priors allows for an interpretation in terms of a frequentist regularised regression. In fact, when all 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 toa frequentist regularised Ridge regression(see De Mol et al., 2008).[14]

3.2 Data-based Priors

An interesting alternative to the statistical priors discussed so far is motivated by the intuition provided by the map in Eqs. (5-6). 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, and iterated up to the relevant horizon h, as follows

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

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 local projections. To see this, note that given Eq. (16) the posterior mean of BLP coefficients under the VAR-based prior takes the form

$$\mathbf{B}_{\text{BLP}}^{(h)} \propto \left(\Omega_0^{(h)^{-1}} + x'x\right)^{-1} \left(\Omega_0^{(h)^{-1}} \mathbf{B}_{\text{VAR}}^h + x'x\widehat{\mathbf{B}}_{\text{LP}}^{(h)}\right). \tag{23}$$

At each horizon h, the relative weight of VAR and LP responses is regulated by $\Omega_0^{(h)}$, that as we discuss below, can be made a function of a single parameter that regulates the overall level of informativeness of the prior, $\lambda(h)$. As in the case of Minnesota priors,

[14] In a similar manner, one could implement a Lasso penalty on the coefficients of a potentially rich set of controls, and that would be equivalent to the double exponential (Laplace) prior. Such a prior would perform variable selection rather than shrinkage as a Ridge regression.

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. (23) can be rewritten as

$$\mathbf{B}_{\text{BLP}}^{(h)} \propto \left[\mathbb{I}_k + M^{-1}\right]^{-1} \mathbf{B}_{\text{LP}}^{(h)} + \left[\mathbb{I}_k + M\right]^{-1} \mathbf{B}_{\text{VAR}}^{h}, \tag{24}$$

$$= Q\mathbf{B}_{\text{LP}}^{(h)} + (\mathbb{I}_k - Q)\mathbf{B}_{\text{VAR}}^{h}, \tag{25}$$

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.

3.3 Prior Variance

Under the two specifications of the prior for the mean of the BLP coefficients, the prior variance is specified in the same way. For the prior scale $\Psi_0^{(h)}$ in Eq. (12) we follow Doan et al. (1983) and fix it using sample information, as it is common in the literature.[15] Specifically, we set

$$\Psi_0^{(h)} = diag\left(\left[\left(\sigma_1^{(h)}\right)^2, \ldots, \left(\sigma_n^{(h)}\right)^2\right]\right) , \tag{26}$$

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

$$\Omega_0^{(h)} = \begin{pmatrix} \mathbb{I}_p \otimes \lambda(h)^2 diag\left(\left[\left(\sigma_1^{(h)}\right)^2, \ldots, \left(\sigma_n^{(h)}\right)^2\right]\right)^{-1} & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}, \tag{27}$$

[15] 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 $\lambda(h)$ controls the overall tightness of the priors at each horizon h.

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, conditional on a draw for $\Sigma_\varepsilon^{(h)}$

$$\mathbb{V}ar\left[\mathbf{B}_{\text{BLP},ij}^{(h)} \mid \Sigma_\varepsilon^{(h)}\right] = \lambda(h)^2 \frac{\Sigma_{\varepsilon,ij}^{(h)}}{\left(\omega_{0,ij}^{(h)}\right)^2} , \tag{28}$$

where $\mathbf{B}_{\text{BLP},ij}^{(h)}$ is the response of variable i to shock j at 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.[16] 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. (2015) for Bayesian VARs.

Specifically, given an hyperprior distribution, it is possible to estimate $\lambda(h)$ from its marginal distribution, conditional on the data and the model

$$p(\lambda(h)|\mathbf{y}^{(h)}) = p(\mathbf{y}^{(h)}|\lambda(h)) \cdot p(\lambda(h)) , \tag{29}$$

where $p(\mathbf{y}^{(h)}|\lambda(h))$ is the marginal density of the data as a function of the hyperparam-

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

eters

$$p(\mathbf{y}^{(h)}|\lambda(h)) = \int p(\mathbf{y}^{(h)}|\lambda(h),\theta)p(\theta|\lambda(h))d\theta \ \forall h \ , \tag{30}$$

and $p(\theta|\lambda(h))$ is the prior distribution of the remaining model's parameters $\left(\mathbf{B}_{\text{BLP}}^{(h)} \text{ and } \Sigma_{\varepsilon}^{(h)}\right)$ 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 $\lambda(h)$,

$$p(\mathbf{y}^{(h)}|\lambda(h)) \propto \underbrace{\left|\left(V_{\varepsilon^{(h)}}^{posterior}\right)^{-1} V_{\varepsilon^{(h)}}^{prior}\right|^{\frac{T-(\tilde{p}+h)+d}{2}}}_{\text{Fit}} \underbrace{\prod_{t=\tilde{p}+1}^{T-h} \left|V_{t+h|t}\right|^{-\frac{1}{2}}}_{\text{Penalty}} \qquad \forall h \ , \tag{31}$$

where $V_{\varepsilon^{(h)}}^{posterior}$ and $V_{\varepsilon^{(h)}}^{prior}$ are the posterior and prior mean of $\Sigma_{\varepsilon}^{(h)}$, and

$$V_{t+h|t} = \mathbb{E}_{\Sigma_{\varepsilon}^{(h)}}\left[\mathbb{V}ar(y_{t+h}|\mathbf{y}^{t},\Sigma_{\varepsilon}^{(h)})\right]$$

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 $\Sigma_{\varepsilon}^{(h)}$. The first term in Eq. (31) relates to the model's in-sample fit, and it increases when $V_{\varepsilon^{(h)}}^{posterior}$ falls relative to $V_{\varepsilon^{(h)}}^{prior}$. 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 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 distribution, it is important to observe that at short horizons a VAR (or a RW) is likely to be a good approximation to the DGP, while over medium horizons the bias in the

coefficients of the VAR due to model misspecification is compounded and grows due the iteration. In the long run coefficients have to decline to zero due to 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 to 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, in the empirical application, we adopt a shifted Logistic function of h specified as follows:

$$f\lambda(h) = \kappa + \frac{\alpha}{1 + e^{-\theta(h-h_0)}} \ , \tag{32}$$

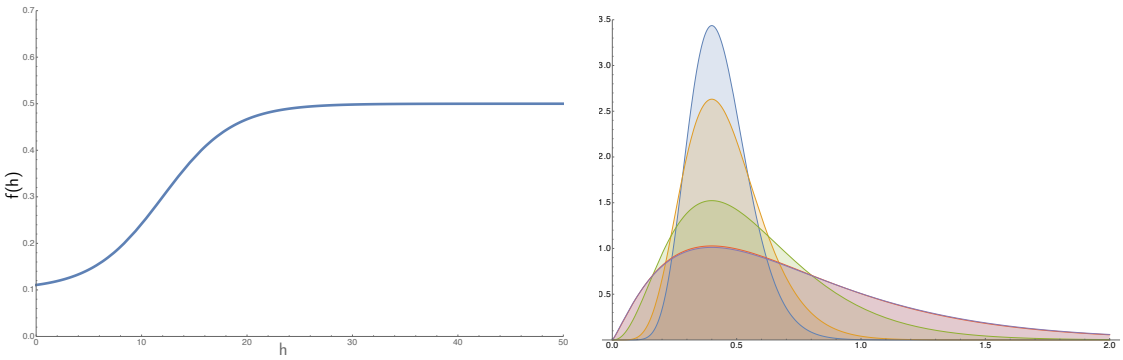
where κ is the shift, α the curve's maximum value, h_0 is is the value of the sigmoid's midpoint, and θ is the logistic growth rate or steepness of the curve.

4 BLP Impulse Response Functions

In this section we evaluate empirical IRFs to an innovation in the Federal Funds Rate estimated using different methods on quarterly US data. In all our empirical applications throughout the paper, we parametrise the Logistic function for the $\lambda(h)$ hyperprior (see Eq. 32)such that it reaches its maximum at horizons larger than $h = 36$, and fix $\kappa = 0.1$, $\alpha = 0.4$ and $\theta = 0.3$ (Figure 1 panel (a)). Panel (b) of Figure 1 illustrates the evolution of the hyperprior for $\lambda(h)$ as a function of the horizon. The mode of the hyperprior is fixed at 0.4, while the prior becomes more diffuse the larger the forecast horizon. Alternatively, the parameters of the Logistic function could also be treated as additional hyperparameters.

We use the variables in Giannone et al. (2015), namely, real GDP, real consumption, real investment, total hours worked, real wages, the GDP deflator and the FFR. With the exception of the policy rate, all variables are expressed in log levels (see Table A.1 in

Figure 1: Hyperprior for BLP Coefficients



(a) (b)

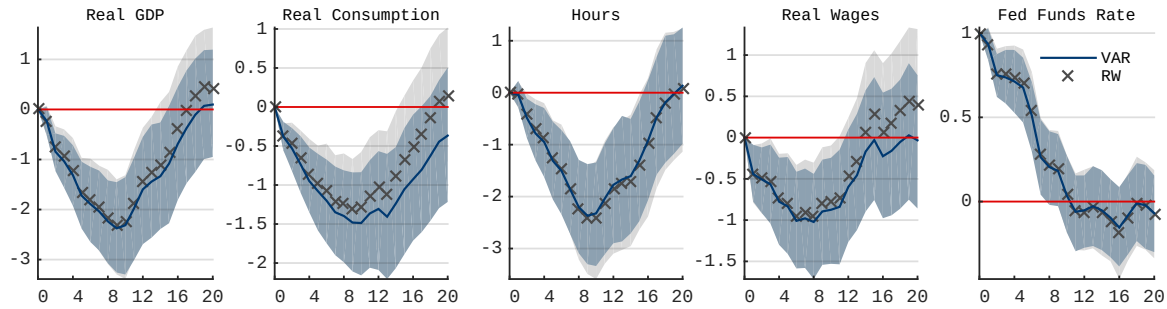
Note: (a) Shifted Logistic function that regulates the variance of the hyperprior for $\lambda(h)$. (b) Hyperprior for $\lambda(h)$ at different horizons. At $h = 1$, the hyperprior has mode equal to 0.4 and standard deviation equal to 0.12 (blue line). The standard deviation increases to 0.16 at $h = 6$ (orange), to 0.30 at $h = 12$ (green), to 0.49 at $h = 24$ (red), and to 0.5 at $h = 36$ (purple).

the Appendix). The sample for the estimation is fixed across all the methods considered and runs from 1965Q1 to 2017Q1. The observations from 1954Q3 to 1964Q4 are used to initialise the BLP priors. We report IRFs to a FFR innovation normalised such that the impact response of the FFR is equal to 1% throughout. The FFR is ordered last in all cases to align the treatment with the simulation exercises reported in Section 5.

VAR-based and RW-based BLP Priors. We start our empirical exploration by evaluating the effects of the choice of the priors for the BLP coefficients in Figure 2 for a selection of variables (full IRFs are reported in Figure A.1 in the Appendix).

In the figure, the markers trace out BLP responses obtained with the RW-based prior, while the solide lines are obtained using a VAR(5)-based prior. In both cases the BLP is specified with 5 lags. We note that BLP responses are remarkably robust to the choice of the prior. In what follows, we prefer to use 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, the VAR-based prior allows us to interpret BLP-IRFs as essentially spanning the model space between Bayesian VARs and Local Projections. That said, the results in Figure 2 show that

Figure 2: BLP responses: VAR vs RW prior



Note: BLP(5) with random walk (RW) prior (markers), and BLP(5) with VAR(5) prior (solid line). Estimation sample: 1965Q1 to 2017Q1. Pre-sample: 1954Q3 to 1964Q4. Shaded areas denote 90% posterior coverage bands.

if the sample length available in the empirical analysis does not permit setting aside some observations to inform the VAR-based prior, the RW prior remains a valuable alternative.

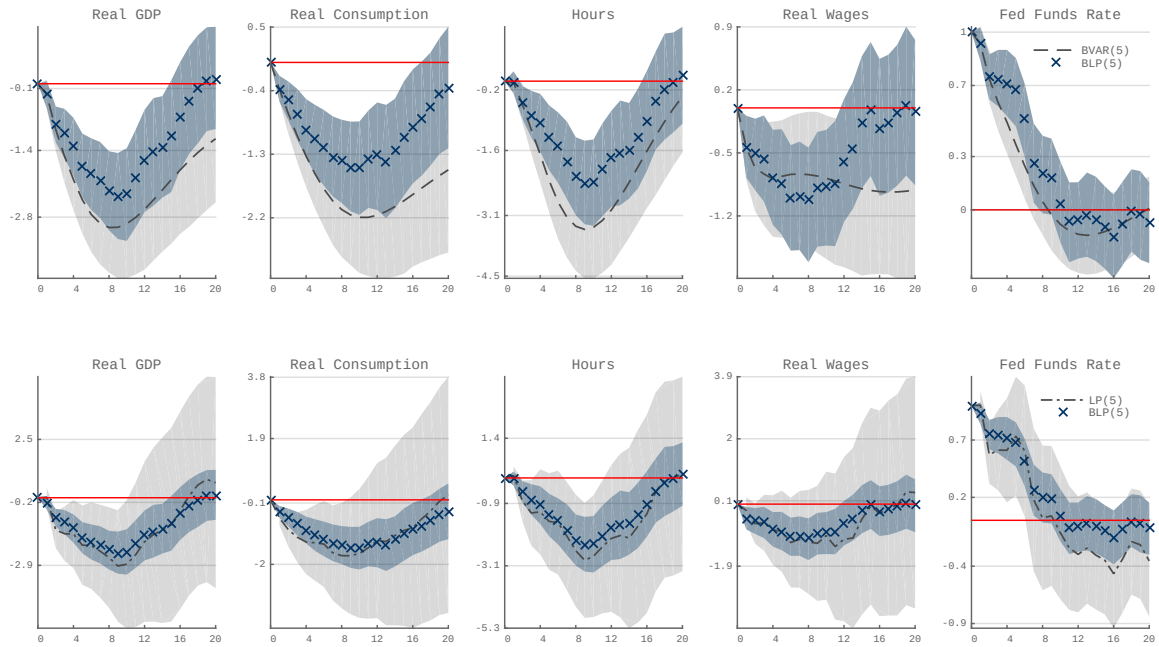
LP, BVAR and BLP IRFs. Figure 3 reports the responses to a FFR innovation estimated using three different methods and for a selection of variables.[17] In the top row, BLP responses are compared with those estimated with a Bayesian VAR with standard NIW priors. In the bottom row, the same BLP responses are compared with those from standard linear local projections. In all cases the number of lags is set to 5 and a VAR(5)-based prior is used for the BLP-IRFs.

A few features emerging from this comparison are worth noticing. Overall, the shape of the IRFs is qualitatively similar across methods. Following a positive innovation in the Federal Funds rate all real variables contract.[18] The length of the sample used, combined with the relatively small number of variables included, limits the erratic nature of LPs. Because many sample observations are available at each horizon, the estimates of projection coefficients are relatively well behaved in this instance. However, notwithstanding the relatively long sample available for the analysis, LP responses quickly become non-significant after the first few horizons. The width of 90% LP confidence

[17]Full IRFs are in Figure A.2 in the Appendix.

[18]In all cases a pronounced price puzzle emerges, likely pointing to an inability of the standard Cholesky identification to recover monetary policy shocks. See Figure A.2.

Figure 3: Empirical IRFs: BVAR, LP and BLP

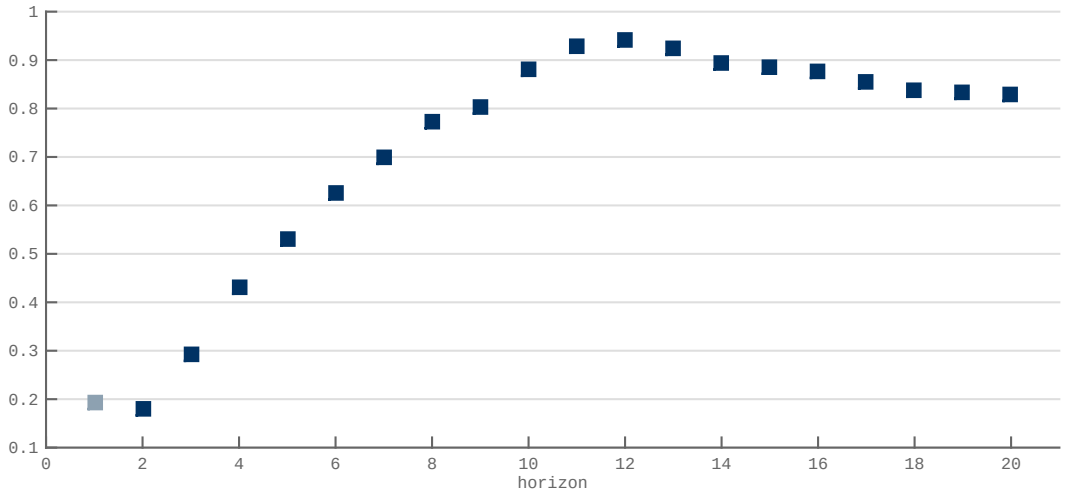


Note: Impulse response functions to a FFR innovation. Top row: BLP(5) and BVAR(5). Bottom row: BLP(5) and LP(5). Estimation sample: 1965Q1 to 2017Q1. Pre-sample: 1954Q3 to 1964Q4. Shaded areas denote 90% posterior coverage bands.

bands dwarfs those of BLP responses, which are instead comparable to those of the VAR (BLP responses are the same in the top and bottom row of the figure).

VAR responses are, by construction, the smoothest. Based on the same one-step-ahead model iterated forward, VAR responses naturally also have tighter bands than LP do (Eq. 6). This feature, however, also results in VARs implying stronger and more persistent effects than BLPs (and LPs) do. 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. The blue markers in Figure 4 display the estimated optimal prior shrinkage hyper-parameters $\lambda(h)$ that maximise the posterior likelihood in the BLP responses in Figure 3. Interestingly, the prior is optimally loosened as the horizon increases, suggesting that

Figure 4: empirical optimal prior tightness



Note: The grey marker is the optimal shrinkage of the Litterman (1986) prior for the BVAR coefficients at $h = 1$, estimated as in Giannone et al. (2015). Blue markers denote the optimal tightness of the VAR prior for BLP coefficients for $h > 1$.

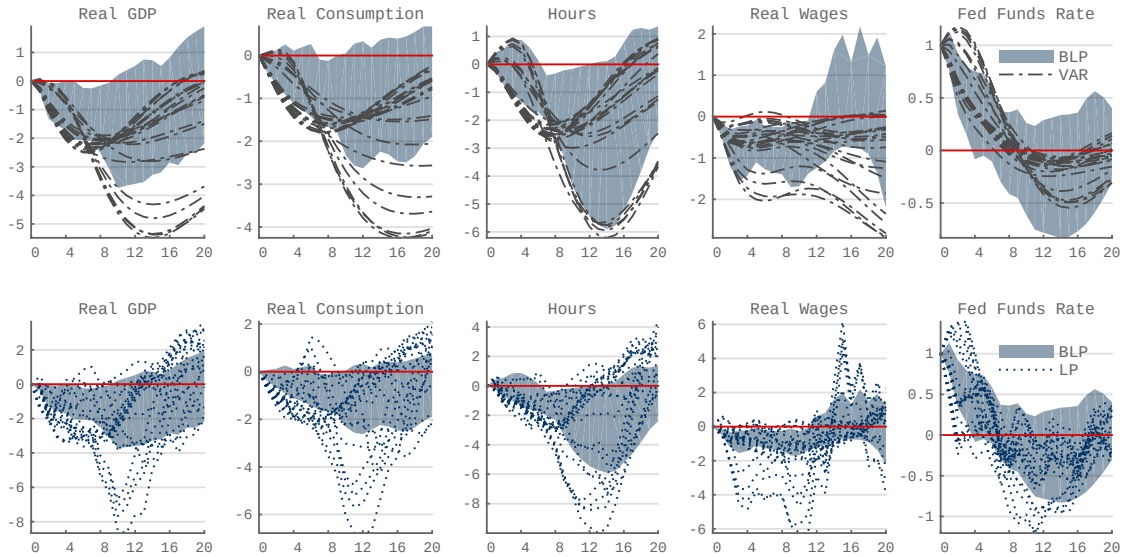
VAR (or equivalently RW) responses tend to be progressively rejected by the data.

Finally, in Figure 5 we evaluate the robustness of BLP estimates over different subsamples.[19] The figure compares BLP-IRFs with BVAR-IRFs (top row) and LP-IRFs (bottom row) computed over a set of fixed-length rolling 30-year samples from 1965Q1 to 2017Q1. Starting from 1965Q1, we use the preceding 10 years to inform the VAR-prior for the BLP coefficients, and the 30 years following to estimate IRFs with the three methods. Then we move forward by one quarter and repeat the procedure. This yields a total of 23 different subsamples. In each case the number of lags is set to 5. 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 abstract from estimation uncertainty.

The broad picture that emerges from Figure 5 is that BLP-IRFs are remarkably stable across samples, relative to both the VAR and the LP. Hence, the regularisation

[19] Full IRFs are reported in Figure A.3 in the Appendix.

Figure 5: Stability over Subsamples: VAR, LP and BLP



Note: BLP(12) with *RW* prior (teal, dashed), and BLP(12) with VAR(12) prior (blue, solid). 1979:2014. Pre-Sample 1969:1979.

implicit in the BLP method not only allows to reduce the estimation uncertainty that is typical of direct methods, but also suggests a lower degree of time-variation in the dynamic interaction among macroeconomic variables that those implied by the alternative methods. Clearly, these results are silent on the ability of BLP to trace out and approximate the true dynamics. We turn to this important point in the next subsection.

5 BLP in a Simulated Environment

We set up a controlled Monte Carlo experiment to evaluate the robustness of BLP to model misspecification, and compare it to that of LPs and VARs. Specifically, we simulate artificial data sets from a medium-scale DSGE model that admits a VAR(p) representation in n endogenous variables. We then recover the IRFs using the three methods estimated with $\tilde{p} < p$ lags and $\tilde{n} < n$ variables. Since medium-scale DSGE models are known to produce reasonably good fit of the data (see e.g. Smets and Wouters, 2007), data simulated from them provide a sensible benchmark to assess the performance

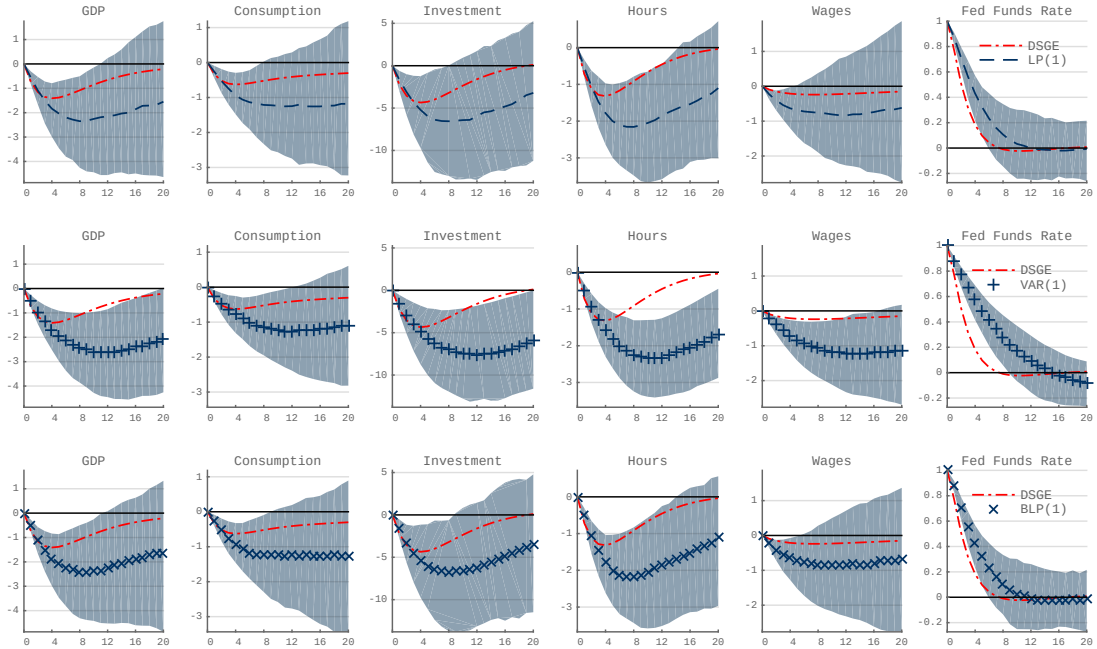
of the different empirical methods.

We use the model in Giannone et al. (2015), in itself a variation of the one in Justiniano et al. (2010). Relative to the original framework in Justiniano et al. (2010), the model we use here assumes that the behaviour of the private sector is predetermined relative to the monetary policy rule, which allows for a recursive identification of monetary policy shocks, same as in Giannone et al. (2015). The model counts seven endogenous variables: output (Y), consumption (C), investment (I), hours worked (H), wages (W), prices (P), and the short-term interest rate (R). The model's dynamics are well approximated by a VAR with five lags.

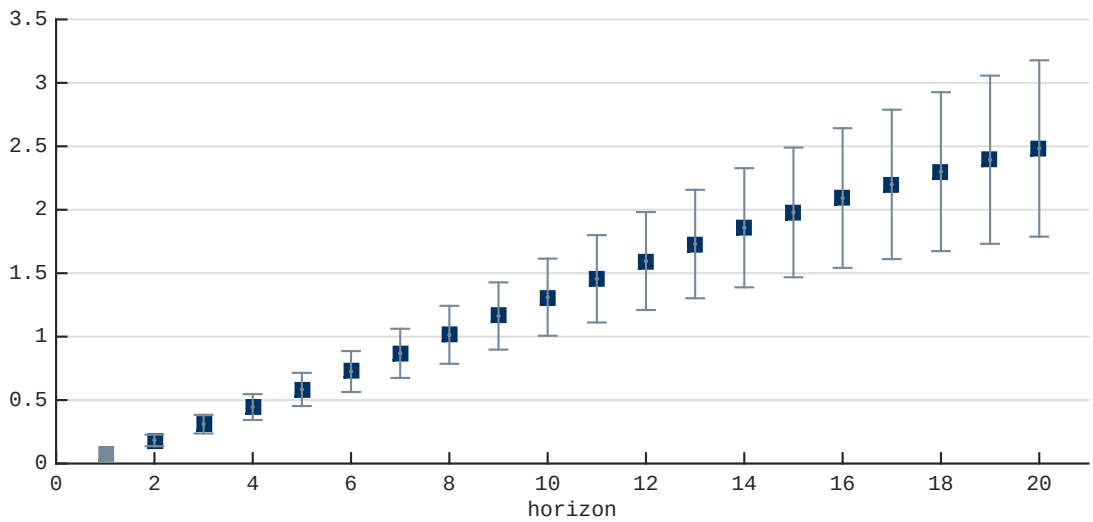
From the model, we simulate 500 artificial time series with 200 data points each for all the endogenous variables. The first 75 data points are used to initialise the BLP prior. For each data set, we estimate the impulse responses to a monetary policy shock obtained with BLP, a VAR with standard Normal-Inverse Wishart priors, and the standard Local Projections. The DSGE is estimated using quarterly U.S. data on output, consumption, and investment growth, hours worked, wage and price inflation, and the federal funds rate from 1965Q1 to 2017Q1. Details on data and transformations are reported in the Appendix.

In our first exercise the misspecification amounts to (i) including only 1 lag instead of 5, and (ii) omitting the price variable. This corresponds to a misspecification that is likely to materialise in practice when the lag order and the information sets are unknown. This is a situation in which LP are typically thought to be more flexible than standard VARs. We use this scenarios to assess how BLP compares with the two methods. Results are in Figure 6a. In each subplot, the red dash-dotted lines are the model-based IRFs evaluated at the posterior mode of the distribution of the parameters, as in Giannone et al. (2015). These are compared with the average of the median responses across simulations for LPs (top row), BVAR (middle row), and BLP (bottom row). The shaded areas in each subplot are the 95% quantiles of the distribution of the IRFs across replications. Figure 6b reports the average optimal prior tightness for the BLP-IRFs across horizons, together with its distribution across the 500 replications. In Figures 7a and 7b the degree of misspecification is exacerbated, and only 3 variables are included,

Figure 6: Simulations
True model: $n = 7$, $p = 5$. Estimated models: $n = 6$, $p = 1$

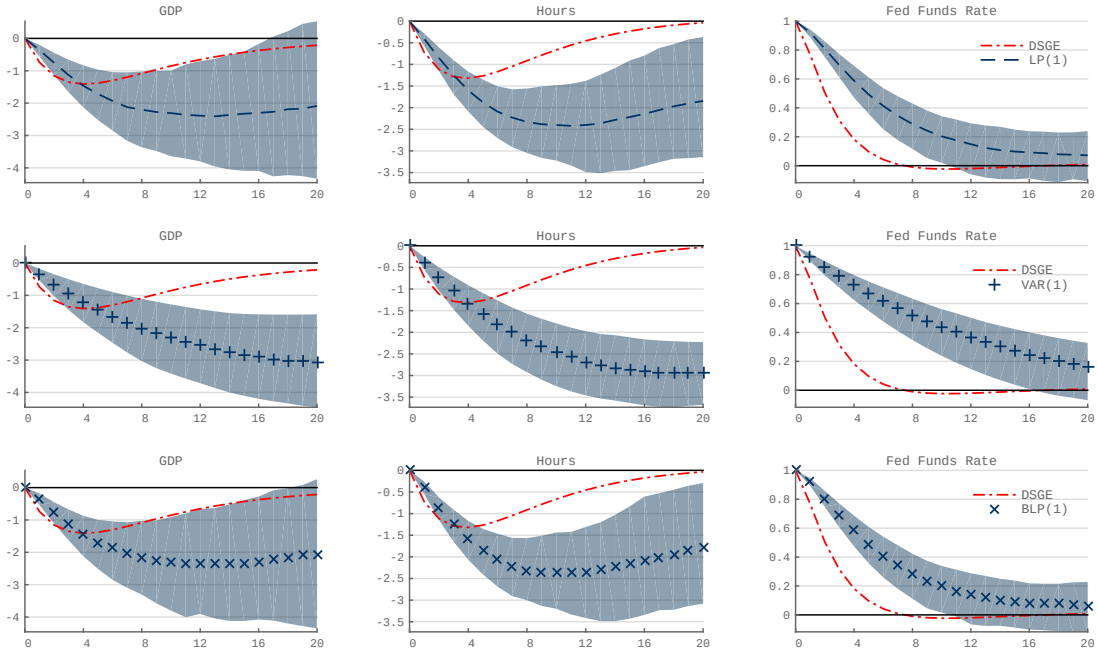


(A) *Notes:* IRFs. In all panels the red dash-dotted lines depict the 'true' DSGE responses. Top row: average of median LP responses across simulations. Middle row: average of median BVAR responses across simulations. Bottom row: average of median BLP responses across simulations. Grey areas are 95% quantiles of the distribution of IRFs across simulations for each method.

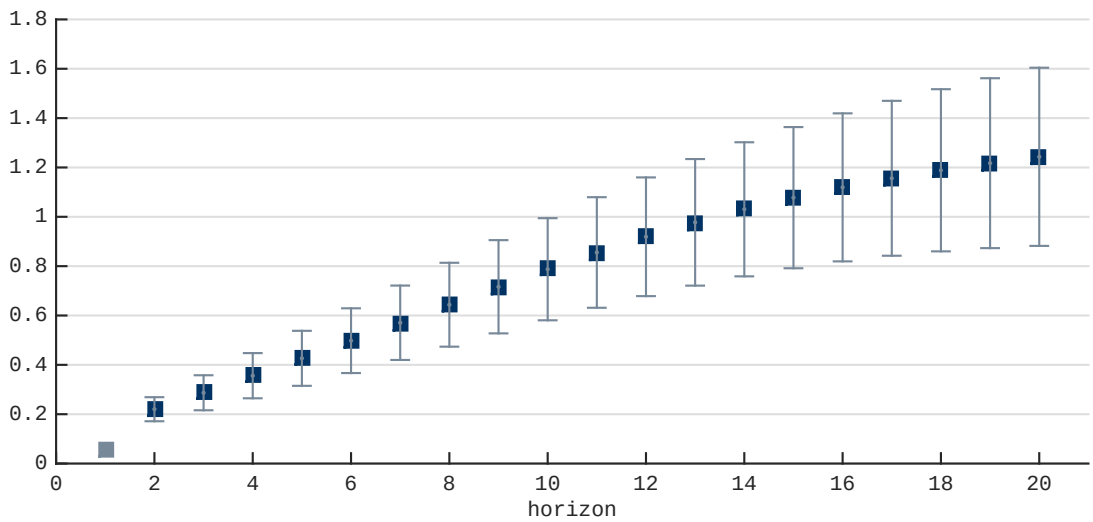


(B) *Notes:* Optimal Prior Tightness. The grey marker is the optimal shrinkage of the Litterman (1986) prior for the BVAR coefficients at $h = 1$, estimated as in Giannone et al. (2015). Average across replications. The blue markers denote the optimal tightness of the VAR prior for BLP coefficients for $h > 1$. Average across replications. The grey error bars are constructed across simulations.

Figure 7: Simulations
True model: $n = 7$, $p = 5$. Estimated models: $n = 3$, $p = 1$



(A) *Notes:* IRFs. In all panels the red dash-dotted lines depict the 'true' DSGE responses. Top row: average of median LP responses across simulations. Middle row: average of median BVAR responses across simulations. Bottom row: average of median BLP responses across simulations. Grey areas are 95% quantiles of the distribution of IRFs across simulations for each method.



(B) *Notes:* Optimal Prior Tightness. The grey marker is the optimal shrinkage of the Litterman (1986) prior for the BVAR coefficients at $h = 1$, estimated as in Giannone et al. (2015). Average across replications. The blue markers denote the optimal tightness of the VAR prior for BLP coefficients for $h > 1$. Average across replications. The grey error bars are constructed across simulations.

again with 1 lag.

The broad picture that emerges from the comparison is that BLPs can be as accurate as LPs, and hence a valid alternative to VARs for what concerns robustness to model misspecification. In Figure 6a the degree of misspecification is relatively mild. Yet, VAR impulse response functions deviate systematically from the true ones virtually in all cases. The difference between BLP/LP and VAR responses becomes even starker in Figure 7a. The evolution of the optimal BLP priors' tightness in Figures 6b and 7b confirms the pattern. As the horizon grows, the data tend to deviate more from the VAR-based BLP prior.

6 Forecasting with BLP

In the forecasting literature, the difference between VARs and LPs has an obvious interpretation as the difference between iterative and direct forecasts. Marcellino et al. (2006) were the first to address the issue of which of the two approaches to forecasting performed better from a purely empirical perspective. The results of that exercise did not return a clear winner. Iterative forecasts were found to be mostly preferable to direct ones, but ultimately the choice between the two methods had to depend on the dynamic properties of the series, on the sample length available, on the empirical specification and the numbers of lags allowed. Ultimately, the problem is always framed in terms of balancing the bias-variance tradeoff. BLPs are designed with this objective; therefore, the same framework can equally be used for forecasting purposes.

In this section, we compare BLP forecasts with multivariate forecasts from both direct methods (LPs) and iterated BVARs, and a naive univariate random walk forecast which serves as a benchmark. The design of the recursive forecasting exercise is as follows. The first estimation sample is 1965Q1 to 1990Q1, where the preceding 10 years are used to inform the prior for the BLP coefficients. Out-of-sample forecasts from all the methods are produced for three forecast horizons equal to 1 quarter, 1 year and 2 years ahead. Observations for 1990Q2 are then added to the estimation sample and the procedure is repeated. The last forecast origin is 2015Q2. This yields a sequence

of 102 out-of-sample forecasts over which the performance of each method is evaluated. The variables used are the seven variables introduced in the previous section, with transformations as in Section **??** (see Table A.1 in the Appendix).

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:

$$y^{\text{LP}}_{T+h|T} = \widehat{\mathbf{B}}^{(h)}_{\text{LP}}\mathbf{y}_T \tag{33}$$

$$y^{\text{VAR}}_{T+h|T} = \widehat{\mathbf{B}}^{h}_{\text{VAR}}\mathbf{y}_T \tag{34}$$

$$y^{\text{BLP}}_{T+h|T} = \widehat{\mathbf{B}}^{(h)}_{\text{BLP}}\mathbf{y}_T, \tag{35}$$

where $\mathbf{y}_T \equiv (1, y'_T, y'_{T-1}, \ldots, y'_{T-p+1})'$, $T = 1990Q1, \ldots, 2015Q2$, $p = 5$, $h = 1, 4, 8$, and each of the estimated $\widehat{\mathbf{B}}$ matrices of coefficients is of dimension $n \times (np+1)$. As noted, the estimation sample always starts in 1965Q1. The random walk forecast is computed as a naive constant-growth forecast.

We evaluate point and density forecasts using standard metrics. For point forecasts, we rely on root mean squared forecast errors, computed as:

$$\text{RMSFE}^j = \sqrt{\frac{1}{N}\sum_{T=90Q1}^{15Q2}\left(y_T - y^j_{T+h|T}\right)^2}, \tag{36}$$

where $j = \{\text{RW}, \text{LP}, \text{VAR}, \text{BLP}\}$, and $N = 102$ is the length of the forecast sequence. For the density forecasts we make use of Log-Scores, computed as:

$$\text{LS}^j = \frac{1}{N}\sum_{T=90Q1}^{15Q2}\log p\left(y^j_{T+h|T}\right), \tag{37}$$

where $p\left(y^j_{T+h|T}\right)$ denotes the predictive density, $j = \{\text{LP}, \text{VAR}, \text{BLP}\}$, and $N = 102$ is the length of the forecast sequence.

The point forecasts for all the variables and all horizons considered are reported in Figure A.4 in the Appendix, while Figure A.5 reports predictive distributions at

Table 1: Average RMSFE – Point Forecast

	$h=1$				$h=4$				$h=8$			
	RW	LP	VAR	BLP	RW	LP	VAR	BLP	RW	LP	VAR	BLP
RGDP	0.671	0.682	0.641	0.641	0.523	0.651	0.530	0.593	0.448	0.503	0.445	0.530
RCON	0.497	0.509	0.469	0.469	0.448	0.456	0.370	0.419	0.341	0.379	0.317	0.379
RINV	2.155	2.298	2.264	2.264	2.136	2.350	2.091	2.414	1.976	1.811	1.743	2.100
HOUR	0.528	0.586	0.517	0.517	0.655	0.727	0.575	0.657	0.747	0.538	0.517	0.587
WAGE	1.577	1.048	1.008	1.008	0.590	0.526	0.466	0.479	0.431	0.353	0.362	0.338
DEFL	0.210	0.238	0.216	0.216	0.160	0.230	0.215	0.251	0.185	0.347	0.274	0.364
FFR	0.088	0.183	0.156	0.156	0.106	0.159	0.111	0.121	0.113	0.141	0.084	0.078

Note: RMSFE. Recursive forecasts for all methods start in 1965Q1, the forecast origins go from 1990Q1 to 2015Q2. LP, VAR and BLP are all estimated with 5 lags.

the three forecast horizons considered for BLP and VAR. The results of the forecast evaluation are reported in Table 1 (RMSFE) and Table 2 (LS).

Table 2: Log Predictive Scores – Density Forecast

	Relative to LP			Relative to VAR		
	$h=1$	$h=4$	$h=8$	$h=1$	$h=4$	$h=8$
RGDP	0.086	-0.445	-1.091	–	-0.124	-0.328
	(0.357)	(1.498)	(1.68)	–	(0.94)	(1.407)
RCON	0.126	-0.762	-1.162	–	0.070	-0.352
	(0.305)	(1.685)	(1.834)	–	(0.741)	(1.359)
RINV	1.089	0.621	-0.118	–	-0.233	-0.329
	(0.676)	(1.443)	(1.744)	–	(0.828)	(1.479)
HOUR	0.304	-0.425	-1.473	–	-0.245	-0.140
	(0.451)	(1.31)	(1.357)	–	(0.688)	(1.56)
WAGE	-0.019	-1.052	-1.442	–	-0.120	-0.384
	(0.65)	(1.772)	(1.796)	–	(0.775)	(0.891)
DEFL	4.031	-0.434	-1.230	–	-0.125	-0.271
	(9.996)	(2.602)	(1.074)	–	(0.683)	(1.096)
FFR	2.413	1.077	5.772	–	-0.069	0.177
	(7.332)	(4.03)	(7.902)	–	(0.491)	(0.677)

Note: Log predictive scores. Recursive forecasts for all methods start in 1965Q1, the forecast origins go from 1990Q1 to 2015Q2. LP, VAR and BLP are all estimated with 5 lags.

The forecasting exercise suggests that, as expected, BLP yields forecasts which have accuracy comparable to that of both VARs and LPs (Table 1). As noted, however, the large variance associated to standard LP-based forecasts makes the predictive densities in this case very wide, which is visible in the large standard deviations in Table 2.

It is worth noting that the design of our forecasting exercise tends to downplay the differences among methods due to the sample being used for the estimation of the coefficient 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 5. As a consequence, the numbers reported in this section can be thought of as conservative estimates. Nonetheless, they confirm that BLP are 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. BLPs resolve the empirical dichotomy between VARs and LPs by framing the choice in terms of the standard bias-variance trade-off that is at the heart of Bayesian estimation.

In setting up BLP we formulate informative priors that give weight to the idea that VARs typically provide a good approximation of the joint dynamics of economic time series. But at each horizon we allow the data to optimally deviate from the prior by making the overall prior tightness a function of the forecast horizon/projection lag, and estimating it in the spirit of hierarchical modelling.

We show that BLP-IRF 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. 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.

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.

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.

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 (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.

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.

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.

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.

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.

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.

——— (forthcoming) "The Transmission of Monetary Policy Shocks," *American Economic Journal: Macroeconomics*.

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.

Smets, Frank and Rafael Wouters (2007) "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, Vol. 97, No. 3, pp. 586–606, June.

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.

Zellner, Arnold (1971) *An introduction to Bayesian inference in econometrics*, Wiley Classics Library: Wiley-Interscience.

A Appendix

A.1 Data Construction and Transformations

Variables Construction (link to download page behind FRED-code)

- **Real GDP: RGDP**$\equiv \log\left(\frac{GDPC1}{POP}\right)$

  GDPC1 Real Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Seasonally Adjusted Annual Rate

  POP Total Population: All Ages including Armed Forces Overseas, Thousands, Quarterly, Not Seasonally Adjusted

- **Real Consumption: RCON**$\equiv \log\left(\frac{PCND+PCESV}{GDPDEF\times POP}\right)$

  PCND Personal Consumption Expenditures: Nondurable Goods, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate

  PCESV Personal Consumption Expenditures: Services, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate

  GDPDEF Gross Domestic Product: Implicit Price Deflator, Index 2009=100, Quarterly, Seasonally Adjusted

- **Real Investment: RINV**$\equiv \log\left(\frac{PCDG+GPDI}{GDPDEF\times POP}\right)$

  PCDG Personal Consumption Expenditures: Durable Goods, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate

  GPDI Gross Private Domestic Investment, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate

- **Total Hours Worked: HOUR**$\equiv \log\left(\frac{HOANBS}{POP}\right)$

  HOANBS Nonfarm Business Sector: Hours of All Persons, Index 2009=100, Quarterly, Seasonally Adjusted

- **Real Compensation per Hour: WAGE**$\equiv \log\left(COMPRNFB\right)$

  COMPRNFB Nonfarm Business Sector: Real Compensation Per Hour, Index 2009=100, Quarterly, Seasonally Adjusted

- **Federal Funds Rate: FFR**$\equiv \frac{FEDFUNDS}{4}$

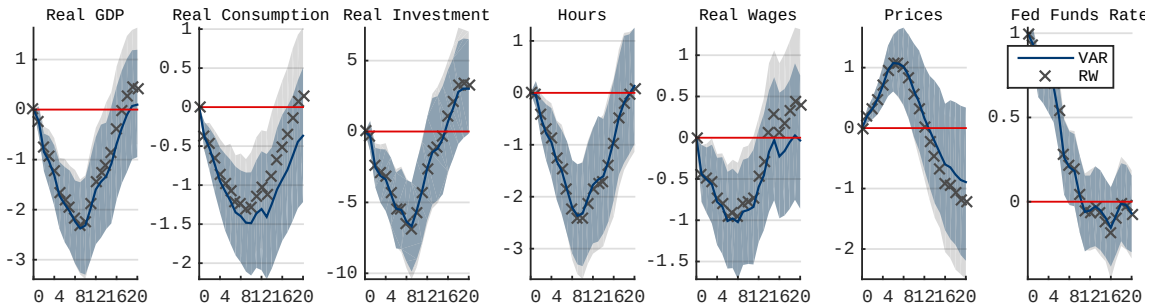
  FEDFUNDS Effective Federal Funds Rate, Percent, Quarterly, Not Seasonally Adjusted

Table A.1: Data and Transformations

		Transformations			
		Per Capita	DSGE	IRFs	FORECASTS
RGDP	Real Gross Domestic Product	•	log-diff	log	log
RCON	Real Consumption	•	log-diff	log	log
RINV	Real Investment	•	log-diff	log	log
HOUR	Hours Worked in Non-Farm Sector	•	log	log	log
WAGE	Real Compensation per Hour		log-diff	log	log
DEFL	GDP Deflator		log-diff	log	log
FFR	Federal Funds Rate		level/4	level/4	level/4

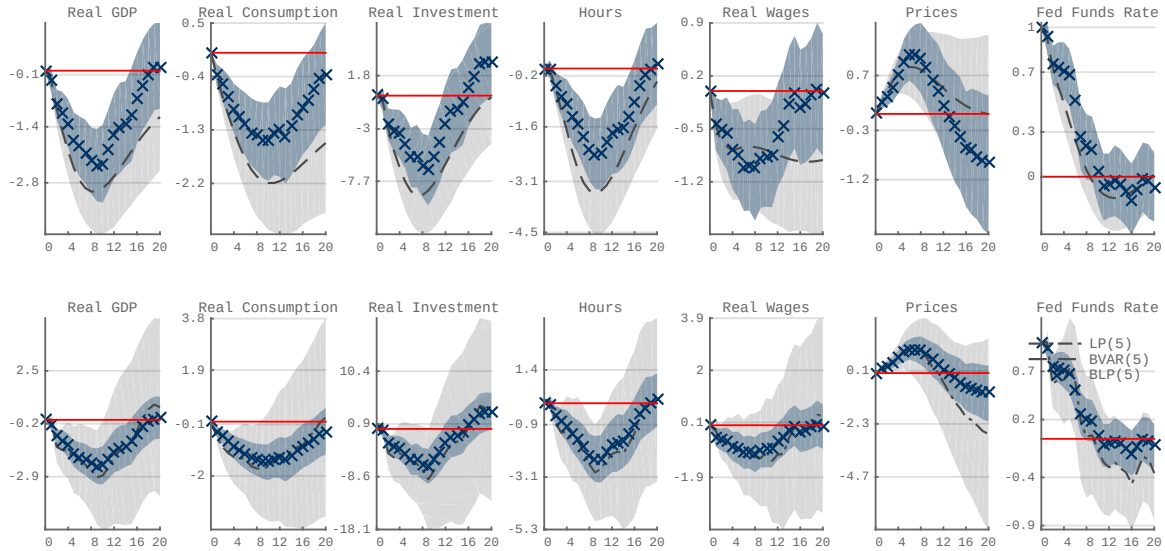
Note: Original data series are retrieved from FRED.

Figure A.1: blp responses: VAR vs RW prior



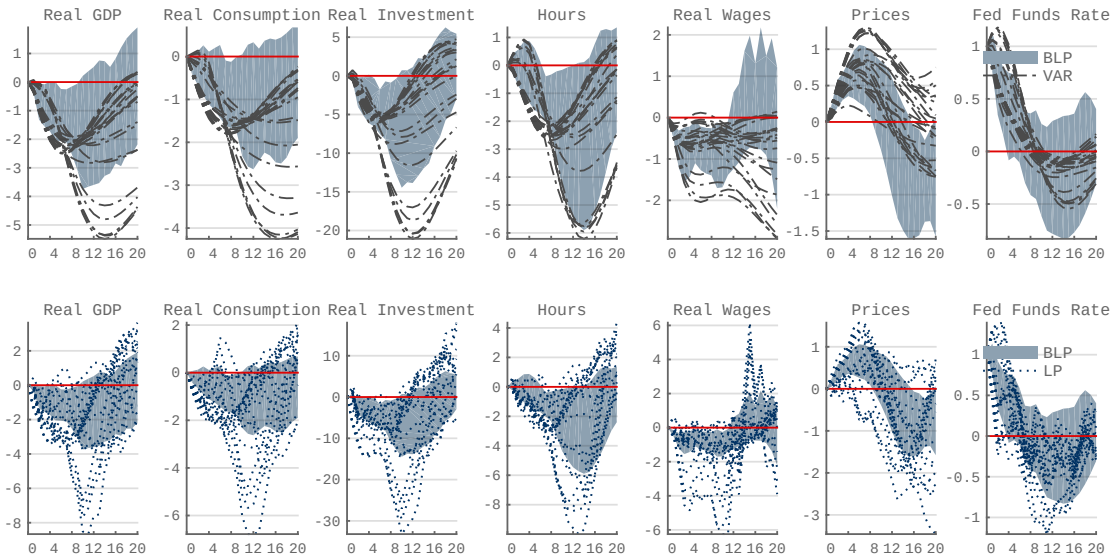
Note: BLP(5) with random walk (RW) prior (markers), and BLP(5) with VAR(5) prior (solid line). Estimation sample: 1965Q1 to 2017Q1. Pre-sample: 1954Q3 to 1964Q4. Shaded areas denote 90% posterior coverage bands.

Figure A.2: Empirical IRFs: BVAR, LP and BLP



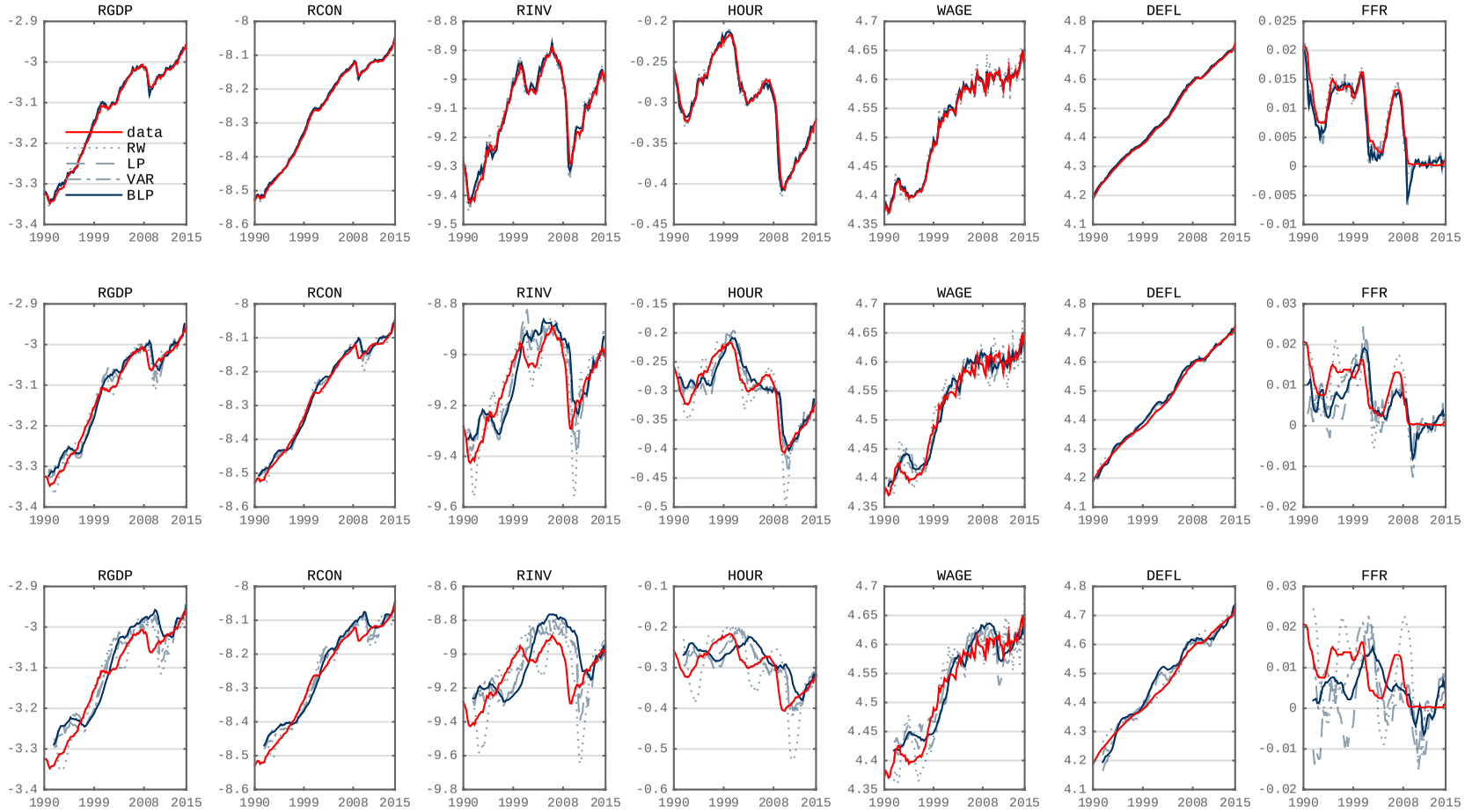
Note: Impulse response functions to a FFR innovation. Top row: BLP(5) and BVAR(5). Bottom row: BLP(5) and LP(5). Estimation sample: 1965Q1 to 2017Q1. Pre-sample: 1954Q3 to 1964Q4. Shaded areas denote 90% posterior coverage bands.

Figure A.3: Stability over Subsamples: VAR, LP and BLP



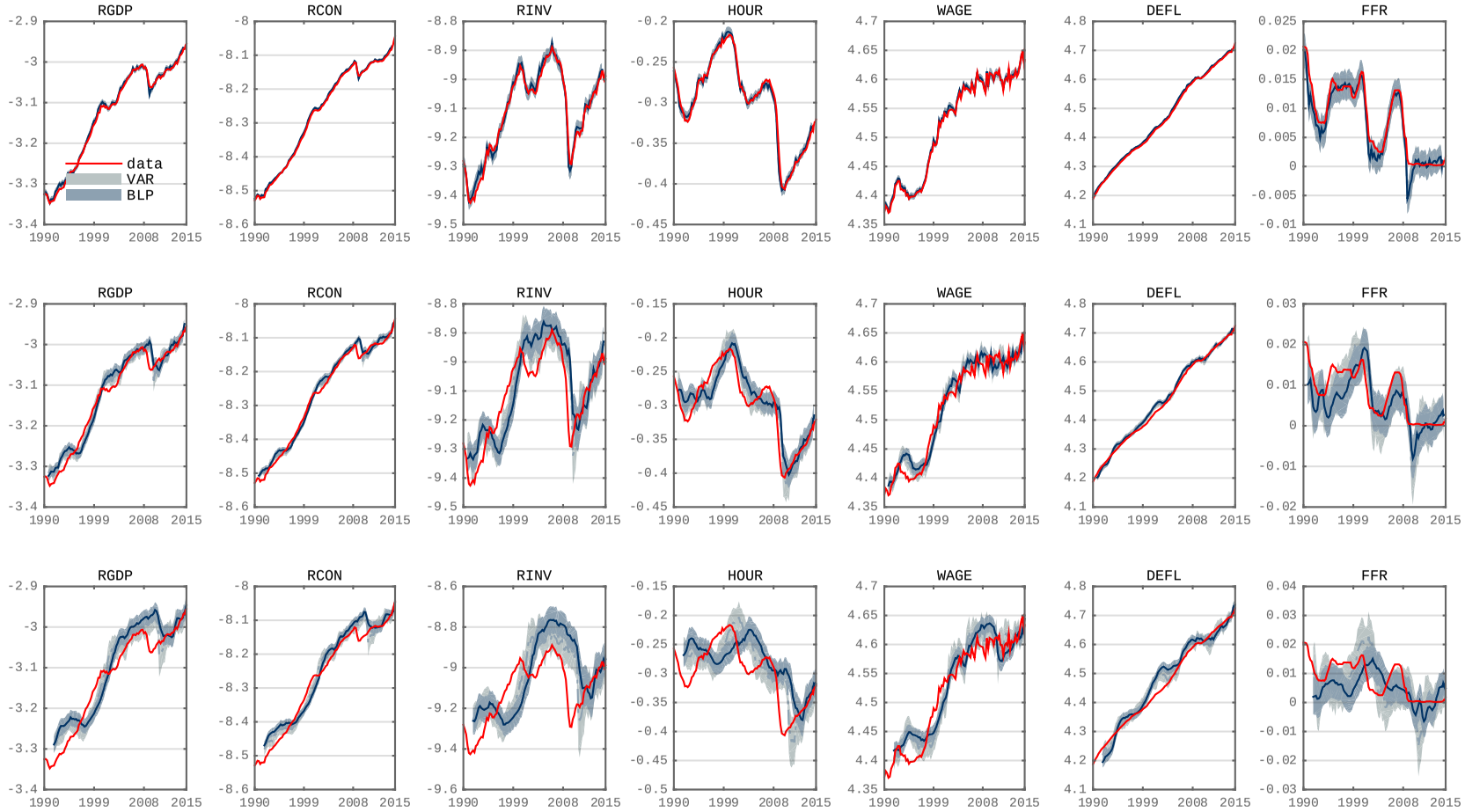
Note: BLP(12) with *RW* prior (teal, dashed), and BLP(12) with VAR(12) prior (blue, solid). 1979:2014. Pre-Sample 1969:1979.

Figure A.4: Point Forecasts



Note: Point forecasts obtained with RW, LP(5), BVAR(5), BLP(5). Recursive forecasts for all methods start in 1965Q1, the forecast origins go from 1990Q1 to 2015Q2. Top panel: $h = 1$; middle panel: $h = 4$; bottom panel: $h = 8$. Forecast horizon expressed in quarters.

Figure A.5: Density Forecasts



Note: Density forecasts obtained with BVAR(5), BLP(5). Recursive forecasts for all methods start in 1965Q1, the forecast origins go from 1990Q1 to 2015Q2. Top panel: $h = 1$; middle panel: $h = 4$; bottom panel: $h = 8$. Forecast horizon expressed in quarters.