Comparison of Local Projection Estimators for Proxy Vector Autoregressions

Martin Bruns[1]
University of East Anglia, School of Economics,
Norwich Research Park, NR4 7TJ, Norwich, United Kingdom
email: martin.j.bruns@gmail.com

and

Helmut Lütkepohl
DIW Berlin and Freie Universität Berlin, Mohrenstr. 58, 10117 Berlin, Germany
email: hluetkepohl@diw.de

September 28, 2021

Abstract. Different local projection (LP) estimators for structural impulse responses of proxy vector autoregressions are reviewed and compared algebraically and with respect to their small sample suitability for inference. Conditions for numerical equivalence and similarities of some estimators are provided. Two generalized least squares (GLS) projection estimators are found to be more accurate than the other LP estimators in small samples. In particular, a lag-augmented GLS estimator tends to be superior to its competitors and to perform as well as a standard VAR estimator for sufficiently large samples.

Key Words: Structural vector autoregression, local projection, impulse responses, instrumental variable

JEL classification: C32

[1]We thank Jörg Breitung, Ralf Brüggemann, Dake Li, Mikkel Plagborg-Møller, Christian Wolf, Ke-Li Xu, Amaze Lusompa and two anonymous referees for helpful comments on an earlier version of this paper. The paper was presented at the XI_t Workshop in Time Series Econometrics, the International Association for Applied Econometrics Annual Conference, the European Meeting of the Econometric Society, and at University of East Anglia School of Economics internal seminars. We thank the participants for helpful comments.

1 Introduction

In structural macroeconometric analysis, local projection (LP) estimators of impulse responses, as proposed by Jordà (2005), have become increasingly popular despite some evidence that they may be inefficient in small samples if the true underlying data generating process (DGP) is a vector autoregression (VAR) (e.g., Meier (2005), Kilian and Kim (2011), Choi and Chudik (2019)). LP estimators are based on linear regressions only, while VAR based impulse responses are nonlinear functions of the VAR slope coefficients. Thereby LP estimators can be defended as nonparametric estimators of impulse responses (Angrist, Jordà and Kuersteiner (2018), Stock and Watson (2018)). They are sometimes regarded as more robust to model deficiencies, which can excuse their small sample inefficiency in standard scenarios. Also, there has been evidence that, in some small sample situations, the loss in efficiency may be quite limited, depending on the choice of the VAR lag order (Brugnolini, 2017). However, based on a large number of DGPs that do not have a finite-order VAR representation, Li, Plagborg-Møller and Wolf (2021) conclude that impulse response estimates based on approximating VARs tend to have much smaller variances than LP estimates but the latter may have smaller bias in small samples. Plagborg-Møller and Wolf (2021) present general conditions for VAR and LP methods to be equivalent tools for impulse response analysis in population.

In this study we focus on a structural VAR setup where the true DGP is a finite-order VAR process and the structural shocks are linear transformations of the reduced-form errors. We also assume that an external instrument or proxy is used to estimate the impact effects of a shock and, thus, the structural parameters (see Stock and Watson (2012), Mertens and Ravn (2013), Gertler and Karadi (2015)). In other words, we focus on a conventional proxy VAR framework. If a suitable external instrument exists, it is also possible to use LP estimators for the corresponding structural impulse responses (e.g., Breitung and Brüggemann (2019), Plagborg-Møller and Wolf (2021)).

The potential small sample inefficiencies of LP estimators have motivated research in modifications with better small sample properties. By now, a number of alternative LP estimators have been proposed (e.g., Plagborg-Møller and Wolf (2017), Stock and Watson (2018), Breitung and Brüggemann (2019), and Lusompa (2021)). The objective of this study is to review and compare the different LP estimators for proxy VAR models in our framework. We derive similarities between the different estimators and even provide conditions for some of them to be numerically equivalent. Some of these results are not apparent from the previous literature. We also compare the small sample properties of the various estimators in a Monte Carlo study.

Anticipating the results, we find that two generalized least squares (GLS) projection estimators dominate the other LP estimators in terms of root mean squared error (RMSE). A lag-augmented GLS version proposed by Breitung and Brüggemann (2019) is the best performing estimator for smaller processes and it is about as accurate in terms of RMSE as a competing LP GLS estimator proposed by Lusompa (2021) for larger VAR models. The lag-augmented GLS estimator also yields small confidence intervals which may, however, have coverage below the desired nominal coverage in small samples if they are constructed with a moving-block bootstrap. For moderately large samples, the estimator has similar properties to the standard VAR estimator if the true DGP is a finite-order VAR process, as assumed in the following.

The study is structured as follows. In the next section the proxy VAR model is presented which is the basis for the LP estimators included in our comparison. In Section 3, the alternative estimators for structural impulse responses are discussed. Section 4 reports small sample results and Section 5 concludes. A number of additional details and results are collected in the Online Supplement which accompanies this paper.

2 Proxy VAR Models

2.1 The General VAR Setup

Consider a K-dimensional reduced-form VAR process of order p (VAR(p)),

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{2.1}$$

The reduced-form error, $u_t = (u_{1t}, \ldots, u_{Kt})'$, is a serially uncorrelated, zero mean white noise process with covariance matrix Σ_u, i.e., $u_t \sim (0, \Sigma_u)$. The VAR(p) model can be written alternatively in the form used for LP estimation,

$$y_{t+h} = \nu_h + A^{(h+1)} Y_{t-1} + v_{t+h}^{(h)}, \qquad \text{(LP form)} \tag{2.2}$$

where ν_h is a constant vector which depends on the integer h, $Y_{t-1}' = (y_{t-1}', \ldots, y_{t-p}')$ is a Kp-dimensional vector of lagged dependent variables,

$$v_t^{(h)} = u_t + \Phi_1 u_{t-1} + \cdots + \Phi_h u_{t-h} \qquad \text{(LP error)} \tag{2.3}$$

is a weighted sum of the reduced-form errors $u_t, \ldots, u_{t-h}$ and

$$A^{(h)} = [A_1^{(h)}, \ldots, A_p^{(h)}]$$

is the $(K \times Kp)$ dimensional matrix consisting of the first K rows of the h^{th} power of the companion matrix

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix}.$$

The $(K \times K)$ weighting matrices Φ_i in (2.3) are equal to the first K columns of $A^{(i)}$, i.e., $\Phi_i = A_1^{(i)}$. They can be computed equivalently as $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$ from the VAR slope coefficients, using $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$ (Lütkepohl (2005, Section 2.1.2) or Kilian and Lütkepohl (2017, Section 12.8)).

The vector of structural errors, $w_t = (w_{1t}, \dots, w_{Kt})'$, is assumed to have instantaneously uncorrelated components, i.e., its covariance matrix, Σ_w, is diagonal. The structural errors are obtained from the reduced-form errors, u_t, by a linear transformation, $u_t = Bw_t$, where B is the matrix of impact effects of the shocks on the observed variables y_t. If the nonsingular $(K \times K)$ matrix B is known, the structural impulse responses can be computed as $\Phi_h B$ for $h = 0, 1, \dots, H$.

Note that for stable, stationary VAR processes satisfying the condition

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1, \tag{2.4}$$

i.e., the determinantal polynomial has no roots in and on the complex unit circle, y_t has moving average (MA) representations,

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \mu + \sum_{i=0}^{\infty} \Phi_i B w_{t-i}. \tag{2.5}$$

In the following we assume that the impulse responses of the first structural shock, w_{1t}, are of primary interest. If only one structural shock is of interest, this choice does not entail a loss of generality because the shocks can be reordered freely. Hence, we only need the first column, denoted by $b = (b_1, \dots, b_K)'$, of the structural matrix B to compute the impulse responses to the first shock as $\Phi_h b$, $h = 0, 1, \dots$.

In line with much of the proxy VAR literature, we also assume that the size of the shock is such that it increases one of the variables on impact by one unit.[2] Without loss of generality, we assume that the first shock has a unit

[2] In the structural VAR literature, shocks of size one standard error are also quite common. As the standard deviation of the shock of interest is typically unknown and has to be estimated, that complicates inference and is perhaps one reason for considering a shock of unit size in much of the proxy VAR literature. An exception is, e.g., the paper by Mertens and Ravn (2013) who use shocks of size one standard deviation.

impact effect on the first variable, possibly after rearranging the variables. In other words, the first element of b is unity, $b_1 = 1$. If structural impulse responses up to a propagation horizon of H periods are of interest, we collect them in the $(K \times (H+1))$ matrix

$$\Theta = [\theta_0, \theta_1, \dots, \theta_H] = [b, \Phi_1 b, \dots, \Phi_H b], \tag{2.6}$$

where the first element of θ_0 is $\theta_{10} = b_1 = 1$. In this study, we will consider alternative estimators of Θ.

2.2 The Proxies

Suppose there are N external instrumental variables in the $(N \times 1)$ vector z_t (called proxies in the following) satisfying

$$\mathbb{E}(w_{1t} z_t') = c' \neq 0 \quad \text{(relevance)}, \tag{2.7}$$

$$\mathbb{E}(w_{kt} z_t') = 0, \quad k = 2, \dots, K \quad \text{(exogeneity)}. \tag{2.8}$$

Here c is a fixed N-dimensional vector. These conditions imply that

$$\mathbb{E}(u_t z_t') = B\mathbb{E}(w_t z_t') = bc'.$$

In other words, the proxies z_t identify multiples of the first column of B.

If proxies are available that satisfy the relevance and exogeneity conditions, then the impact effects can be obtained as

$$b = \theta_0 = \mathbb{E}(u_t z_t') Q \mathbb{E}(z_t u_{1t}) / \mathbb{E}(u_{1t} z_t') Q \mathbb{E}(z_t u_{1t}) \tag{2.9}$$

for any positive definite $(N \times N)$ matrix Q. Once b has been recovered, the matrix of structural impulse responses of interest, Θ, can be determined using the Φ_h of the reduced-form VAR(p) model.

Stock and Watson (2018) also require the following lead-lag exogeneity condition for some of their theoretical results to hold:

$$\mathbb{E}(w_{t+i} z_t') = 0 \quad \text{for} \quad i \neq 0 \quad \text{(lead-lag exogeneity)}. \tag{2.10}$$

Although this condition does not exclude serially correlated z_t, the proxy has to be such that it is not predictable from past y_t because y_t and its lags depend on the shocks w_t. However, what we have in mind in the following are proxies that mimic the shock of interest and, hence, have no serial correlation. This will also be reflected in the way we generate the proxies in the simulations in Section 4. In some studies, serially correlated proxies have been considered (e.g., Gertler and Karadi (2015), Angelini and Fanelli

(2019)). Such proxies qualify as well for our purposes if they satisfy the lead-lag exogeneity condition, which is, of course, possible (see Stock and Watson (2018, p. 924)). Alternatively, one may construct a new proxy by prewhitening the proxy, e.g., by using as proxy the residual of an AR process fitted to the autocorrelated proxy series or the residuals of a regression on own lags and lags of the VAR variables, as in Angelini and Fanelli (2019). We have not explored these options in our simulations, however.

In applications it is not uncommon that there is only one proxy such that $N = 1$ and z_t is a scalar variable. In that case, the impact effects of the first shock are obtained as

$$b = \theta_0 = \mathbb{E}(u_t z_t)/\mathbb{E}(u_{1t} z_t).$$

Although $N = 1$ is a common case in practice, we allow for the general case of $N > 1$ proxies in our theoretical framework. However, we will present the estimators of interest also for the case $N = 1$ to make the formulas easier to digest.

One could also extend the framework such that more than one shock is identified by a set of proxies. Typically that requires additional assumptions for separately identifying the individual shocks and their impulse responses. Once those assumptions are imposed, the shocks and their impulse responses can be considered one by one. Therefore we focus on inference for impulse responses of a single shock in the following.

In the next section, estimation of Θ is discussed. A standard estimator based on estimating the Φ_h from reduced-form estimators of the VAR slope coefficients is presented in addition to alternative LP estimators.

3 Estimators of Structural Impulse Responses

We first present the standard proxy VAR approach for estimating structural impulse responses. It is our benchmark against which we compare the alternative projection approaches discussed subsequently. The following projection estimators will be covered: Section 3.2 extends the standard LP approach proposed by Jordà (2005) to the proxy VAR context and also presents a GLS version developed by Lusompa (2021). In Section 3.3, the lag-augmented LP estimator of Montiel Olea and Plagborg-Møller (2021) and a numerically equivalent projection estimator of Breitung and Brüggemann (2019) are presented. Moreover, an asymptotically more efficient GLS extension proposed by the latter authors is also discussed. Section 3.4 is devoted to instrumental variables (IV) versions of the estimators with and without control variables as suggested by Stock and Watson (2018). For completeness, we show in Section 3.5 how to estimate the impulse responses directly from the LP errors. In Section 3.6 the various estimators are summarized and compared systematically. A range of further proposals is briefly mentioned in Section 3.7.

To simplify the exposition, we will present the different estimators for a scalar proxy first and mention the necessary modifications for a vector of proxies at the end of each section. Thus, z_t is now a scalar proxy variable if not explicitly stated otherwise.

It is assumed that for estimation a gross sample size T is available, including all required presample and lead values of the observable variables. For a fair small sample assessment of the different estimators, it is important to consider the same gross sample size for each of them because the estimators differ also in the number of presample and lead values needed in their calculations and, hence, they differ in the net sample size they are using.

3.1 The Standard VAR Approach

Estimators of the Φ_i matrices may be obtained from estimators, $\hat{A}_i$, of the reduced-form VAR model in equation (2.1) using the recursions

$$\hat{\Phi}_i = \sum_{j=1}^{i} \hat{\Phi}_{i-j}\hat{A}_j, \quad i = 1, \ldots, H.$$

Thus, the reduced-form impulse responses are nonlinear functions of the VAR slope coefficients. Nonlinear functions may magnify estimation errors due to model misspecification. The estimators $\hat{A}_i$ may simply be ordinary least squares (OLS) estimators. Alternatively, one may want to use bias-corrected OLS estimators as suggested by Kilian (1998), to improve inference for impulse responses (see Section S2.1 of the Online Supplement for the precise implementation). Such estimators were found to be superior for stable, stationary VAR processes in a number of small-sample investigations (e.g., Kilian (1998), Lütkepohl, Staszewska-Bystrova and Winker (2015a, 2015b)).

The first column of Θ, θ_0, can be estimated using the proxy z_t. Let

$$\hat{\theta}_0 = \sum_{t=p+1}^{T} \hat{u}_t z_t \Bigg/ \sum_{t=p+1}^{T} \hat{u}_{1t} z_t \ , \tag{3.1}$$

where the $\hat{u}_t$ are the estimated residuals of the reduced-form VAR(p). The estimator $\frac{1}{T-p}\sum_{t=p+1}^{T} \hat{u}_t z_t$ is a standard method-of-moments estimator which, under general conditions, is asymptotically normal (see Newey and McFadden

(1994)) and the last $K-1$ components of $\hat{\theta}_0$ are a differentiable function of that estimator. Thus, they are also consistent and asymptotically normal. More details on asymptotics are given in the Online Supplement.

Combining the proxy VAR estimator $\hat{\theta}_0$ with the reduced-form impulse response estimators gives a conventional VAR based estimator

$$\widehat{\Theta}_{VAR} = [\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_H] = [\hat{\theta}_0, \hat{\Phi}_1\hat{\theta}_0, \dots, \hat{\Phi}_H\hat{\theta}_0] \tag{3.2}$$

of the structural impulse responses Θ.

In some of the related literature, the proxy is turned into an internal variable of the VAR by adding it to the set of observed variables y_t and a VAR model for the augmented vector $(z_t, y_t')'$ is considered (see also Section S1.1 of the Online Supplement). Plagborg-Møller and Wolf (2021) show that an advantage of internalizing the proxy is that asymptotically valid impulse response analysis becomes possible even if the shock of interest is 'noninvertible', that is, the shock cannot be recovered from past and present forecast errors. As we assume that the shocks are linear transformations of the reduced-form VAR errors, they are invertible and do not pose the 'noninvertibility' problem. If the proxy mimics the properties of w_{1t}, so that it is white noise and there are no lags in the proxy equation in the VAR process and also the y_t equations contain no lags of the proxy, then the impact effects of the first shock may be estimated by considering the Cholesky decomposition of

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\begin{pmatrix} z_t \\ \hat{u}_t \end{pmatrix}(z_t, \hat{u}_t').$$

Dividing the first column of this matrix by the second element in that column, the last K elements are an estimator $\hat{\theta}_0$ of θ_0 which is numerically identical to the estimator in expression (3.1) (see Section S1.1 of the Online Supplement). In other words, if we fully take into account the more restrictive assumptions for the proxy, we get the same estimator $\widehat{\Theta}_{VAR}$ as from our standard setup. Therefore we consider the latter setup in the following.

If z_t is an N-dimensional vector of proxies, a possible estimator of θ_0 is

$$\hat{\theta}_0 = \sum_{t=p+1}^{T}\hat{u}_t z_t' Q_z \sum_{t=p+1}^{T} z_t \hat{u}_{1t} \Bigg/ \sum_{t=p+1}^{T}\hat{u}_{1t} z_t' Q_z \sum_{t=p+1}^{T} z_t \hat{u}_{1t}, \tag{3.3}$$

where Q_z is a positive definite matrix. We choose $Q_z = (\sum_{t=1}^{T} z_t z_t')^{-1}$ to improve the estimation efficiency. The other quantities are not affected by using a vector of proxies.

3.2 The Standard Local Projection Estimator

Jordà's (2005) LP estimator is based on the system of KH equations

$$y_{t+h} = \nu_h + A^{(h+1)}Y_{t-1} + v_{t+h}^{(h)}, \quad h = 0, 1, \dots, H-1. \tag{3.4}$$

Estimating this set of equations by OLS gives estimators $\hat{\Phi}_i^{LP} = \hat{A}_1^{(i)}$, where $\hat{A}_1^{(i)}$ denotes the first K columns of the estimator $\hat{A}^{(i)}$. Thus, the reduced-form impulse responses, Φ_i, are estimated by linear regression techniques which is sometimes regarded as an advantage because such estimators are robust to some assumptions underlying the VAR model setup. The drawback is that up to H lead values are needed which reduce the effective sample size and many redundant parameters have to be estimated which may compromise the efficiency of the LP estimators of the relevant parameters in the first K columns of $A^{(i)}$. The estimated Φ_i can be used to estimate the structural impulse responses, Θ, as

$$\widehat{\Theta}_{LP} = [\hat{\theta}_0, \hat{\Phi}_1^{LP}\hat{\theta}_0, \dots, \hat{\Phi}_H^{LP}\hat{\theta}_0], \tag{3.5}$$

where $\hat{\theta}_0$ is the same estimator for the impact effects as in $\widehat{\Theta}_{VAR}$. Note also that, if OLS estimation is used, $\hat{\Phi}_1^{LP} = \hat{\Phi}_1$ and is, hence, identical to the standard VAR estimator. Thus, $\hat{\Phi}_1^{LP}\hat{\theta}_0 = \hat{\theta}_1$ as in $\widehat{\Theta}_{VAR}$.

Jordà (2005) points out that, given that the error term $v_{t+h}^{(h)}$ is autocorrelated and heteroskedastic, GLS estimation can be used for inference. GLS estimation is possible because the stochastic structure of the error term is known if the DGP is a VAR process. Thus, the error covariance matrix can be constructed and estimated from the VAR parameters. GLS estimation can also be used for point estimation to improve the estimation efficiency. A feasible GLS procedure has been proposed by Lusompa (2021) who uses an iterative procedure which pre-cleans the left-hand side of the LP form of the VAR using estimates $\hat{u}_{t+h-1}, \dots, \hat{u}_{t+1}$. More precisely, reduced-form impulse responses are obtained from

$$\tilde{y}_{t+h} = \nu_h + A^{(h+1)}Y_{t-1} + e_t^{(h)}, \tag{3.6}$$

where $\tilde{y}_{t+h} = y_{t+h} - \hat{\Phi}_1^{GLS}\hat{u}_{t+h-1} - \cdots - \hat{\Phi}_{h-1}^{GLS}\hat{u}_{t+1}$ and $\hat{\Phi}_1^{GLS}, \dots, \hat{\Phi}_{h-1}^{GLS}$ are obtained from the regressions at horizons 1 through $h-1$ as the first K columns of the estimator $\hat{A}^{(i)}$. In our estimations, $\hat{u}_{t+h-1}, \dots, \hat{u}_{t+1}$ are OLS or bias-corrected OLS reduced-form errors, depending on the estimation method used for the VAR. The full estimator of Θ corresponding to this GLS procedure is

$$\widehat{\Theta}_{LP}^{GLS} = [\hat{\theta}_0, \hat{\Phi}_1^{LP}\hat{\theta}_0, \hat{\Phi}_2^{GLS}\hat{\theta}_0, \dots, \hat{\Phi}_H^{GLS}\hat{\theta}_0], \tag{3.7}$$

where $\hat{\theta}_0$ is again the estimator of the impact effects in $\widehat{\Theta}_{VAR}$. In other words, the first two columns of $\widehat{\Theta}_{LP}^{GLS}$, $\widehat{\Theta}_{LP}$, and $\widehat{\Theta}_{VAR}$ are identical.

3.3 Lag-augmented Local Projection

Lag-augmentation to fix unit root asymptotics was proposed earlier in other contexts by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) and in the context of impulse response analysis by Dufour, Pelletier and Renault (2006). Breitung and Brüggemann (2019) and Montiel Olea and Plagborg-Møller (2021) propose to use that device in the present context as well. They add an additional lag to the LP form of the VAR(p) process,

$$y_{t+h} = \nu_h + A^{(h+1)}Y_{t-1} + A_{p+1}^{(h+1)}y_{t-p-1} + v_{t+h}^{(h)}, \quad h = 0, 1, \ldots, H-1. \tag{3.8}$$

Note that $A^{(h+1)}$ is $(K \times Kp)$ dimensional while $A_{p+1}^{(h+1)}$ is a $(K \times K)$ matrix. If the true DGP is a VAR(p), then the coefficient matrices of the additional lag are known to be zero, i.e., $A_{p+1}^{(h+1)} = 0$, $h = 0, 1, \ldots, H-1$, and estimating the lag-augmented model by OLS implies an inefficiency. However, Montiel Olea and Plagborg-Møller (2021) show that the resulting lag-augmented LP estimator is more robust to unit roots and near unit roots and therefore has advantages for inference. We denote the corresponding impulse response estimator by $\hat{\Phi}_i^{aug} = \hat{A}_1^{(i)aug}$. The resulting estimator of Θ is

$$\widehat{\Theta}_{LP}^{aug} = [\hat{\theta}_0, \hat{\Phi}_1^{aug}\hat{\theta}_0, \ldots, \hat{\Phi}_H^{aug}\hat{\theta}_0], \tag{3.9}$$

where $\hat{\theta}_0$ is again the same estimator as in $\widehat{\Theta}_{VAR}$.

The model (3.8) can be reparameterized as

$$y_{t+h} = \nu_{h-1} + \Theta_h w_t + A_*^{(h)}Y_{t-1} + v_{t+h}^{(h-1)},$$

where $A_*^{(h)}$ is a $(K \times Kp)$ matrix (see Appendix A.1). To use this model for estimation, the w_t have to be replaced by estimates. As the components of w_t are uncorrelated, it is plausible to consider orthogonal regressors for them. In that case, $w_{2t}, \ldots, w_{Kt}$ can be dropped from the equations for OLS estimation without affecting the estimators of the other parameters. Hence, θ_h, the first column of Θ_h, can be estimated by OLS using the model

$$y_{t+h} = \nu_h + \theta_h\hat{w}_{1t} + A_*^{(h)}Y_{t-1} + v_{t+h}^{(h-1)} \quad \text{for} \quad h = 1, \ldots, H, \tag{3.10}$$

where $\hat{w}_{1t}$ is an estimate of w_{1t}. If we were to choose $\hat{w}_{1t}$ such that the OLS estimator of θ_0, which is obtained from the model (3.10) for $h = 0$, is identical to the estimator $\hat{\theta}_0$, the first column of $\widehat{\Theta}_{VAR}$, we could compute the columns of the estimator $\widehat{\Theta}_{LP}^{aug}$ by OLS estimation of model (3.10). In fact, the shocks w_{1t} corresponding to a given $\theta_0 = b$ can be determined as

$$w_{1t} = b'\Sigma_u^{-1} u_t / b'\Sigma_u^{-1} b \tag{3.11}$$

(see Appendix A.1). Thus, using $b = \hat{\theta}_0$ and substituting the OLS estimator for Σ_u based on a VAR(p) gives a series of shocks $\hat{w}_{1t}$ which yield the estimator $\widehat{\Theta}_{LP}^{aug}$. Of course, the extra step of computing $\hat{w}_{1t}$ is not needed if an estimate of θ_0 is already available and, hence, we do not use it for computing $\widehat{\Theta}_{LP}^{aug}$.

Breitung and Brüggemann (2019) proceed in a different way. They consider the model (3.10) and propose to estimate w_{1t} directly based on estimated reduced-form errors, $\hat{u}_1, \dots, \hat{u}_T$, such that the first element of $\hat{\theta}_0^{BB}$ is 1. In other words, the impulse responses are by construction standardized such that the first shock has a unit impact effect on the first variable. They first estimate the structural errors $w_{2t}, \dots, w_{Kt}$ recursively for $k = 2, \dots, K$, from the system of $K - 1$ equations

$$\begin{aligned} \hat{u}_{2t} &= \gamma_{21}\hat{u}_{1t} + w_{2t}, \\ \hat{u}_{kt} &= \gamma_{k1}\hat{u}_{1t} + \gamma_{k2}\hat{w}_{2t} + \cdots + \gamma_{k,k-1}\hat{w}_{k-1,t} + w_{kt}, \quad k = 3, \dots, K, \end{aligned} \tag{3.12}$$

by the instrumental variables (IV) method using z_t as an instrument for $\hat{u}_{1t}$. The estimated errors are denoted by $\hat{w}_t^{(2)} = (\hat{w}_{2t}, \dots, \hat{w}_{Kt})'$. In the next step, the w_{1t} are estimated as the errors of the OLS regression

$$\hat{u}_{1t} = \eta \hat{w}_t^{(2)} + w_{1t}, \tag{3.13}$$

where η is a $(K-1)$-dimensional row vector. Substituting the estimator $\hat{w}_{1t}$ computed in this way for w_{1t} in (3.10), an estimator of Θ is obtained which we denote as

$$\widehat{\Theta}_{BB} = [\hat{\theta}_{BB,0}, \dots, \hat{\theta}_{BB,H}]. \tag{3.14}$$

Despite the differences in the computations, the estimator $\hat{\theta}_{BB,0}$ precisely matches the estimator $\hat{\theta}_0$ in $\widehat{\Theta}_{VAR}$, if OLS reduced-form VAR errors $\hat{u}_t$ are used, because both estimators fully exploit the information in the reduced form errors and the proxy. Both estimators can be interpreted as generalized method of moments (GMM) estimators based on equivalent moment conditions. This implies that $\widehat{\Theta}_{BB}$ is identical to $\widehat{\Theta}_{LP}^{aug}$ if OLS residuals are used for $\hat{u}_t$ in the Breitung-Brüggeman approach for estimating w_{1t}. In the simulations reported in Section 4, we also consider the residuals of bias-corrected OLS instead, which yields differences in the two estimators and, hence, the different notation is needed when the estimator is computed via the Breitung-Brüggemann approach.

As the estimation equations in (3.10) have an autocorrelated error term, $v_{t+h}^{(h-1)} = u_{t+h} + \Phi_1 u_{t+h-1} + \cdots + \Phi_{h-2} u_{t+2} + \Phi_{h-1} u_{t+1}$, Breitung and Brüggemann (2019) also propose a GLS estimator obtained by replacing the unobserved quantities $u_{t+h}, u_{t+h-1}, \ldots, u_{t+2}$ in the error term of (3.10) by estimates and using the system of equations

$$y_{t+h} - \hat{u}_{t+h} = \nu_h + \theta_h \hat{w}_{1t} + A^{(h+1)} Y_{t-1} + \Phi_1 \hat{u}_{t+h-1} + \cdots + \Phi_{h-2} \hat{u}_{t+2} + e_t^{(h)} \quad (3.15)$$

to estimate θ_h for $h = 3, \ldots, H$. The $\hat{u}_{t+h}, \ldots, \hat{u}_{t+2}$ are estimated reduced-form VAR errors and the estimates of w_{1t} are obtained as in (3.13). We denote the estimator of θ_h based on (3.15) by $\hat{\theta}_{BB,h}^{GLS}$ for $h = 3, \ldots, H$. For $h = 2$, the estimator $\hat{\theta}_{BB,2}^{GLS}$ is determined by OLS estimation of

$$y_{t+2} - \hat{u}_{t+2} = \nu_1 + \theta_2 \hat{w}_{1t} + A^{(3)} Y_{t-1} + e_t^{(2)}.$$

The full estimator of Θ corresponding to the GLS procedure is

$$\widehat{\Theta}_{BB}^{GLS} = [\hat{\theta}_{BB,0}, \hat{\theta}_{BB,1}, \hat{\theta}_{BB,2}^{GLS}, \ldots, \hat{\theta}_{BB,H}^{GLS}], \quad (3.16)$$

where for $h = 0, 1$, the estimators $\hat{\theta}_{BB,0}$ and $\hat{\theta}_{BB,1}$ based on (3.10) are used. In other words, the first two columns of $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{BB}$ are identical. Since the GLS estimator accounts for the autocorrelation in the error term, it is asymptotically more efficient than the lag-augmented LP estimator based on OLS estimation of (3.8) for columns θ_h, $h > 1$.

Breitung and Brüggemann (2019) discuss also other variants of their estimators. We do not consider them in our comparison because they did not seem to improve on the small sample performance of the present estimators in their simulations.

3.4 An Instrumental Variables Approach

Stock and Watson (2018) assume that the proxies satisfy the relevance, the exogeneity, and the lead-lag exogeneity conditions. For a scalar proxy z_t and mean-adjusted y_t, they note that, using z_t as an instrument, the standard IV estimator of the coefficient in the linear model

$$y_{t+h} = \theta_h y_{1t} + u_{t+h}^{(h)}, \quad (3.17)$$

is

$$\hat{\theta}_h(IV) = \left(\sum_{t=1}^{T-h} z_t y_{1t} \right)^{-1} \sum_{t=1}^{T-h} z_t y_{t+h}. \quad (3.18)$$

Clearly, the regressor in (3.17) is correlated with the error term and, hence, simple OLS regression is inconsistent. In contrast, the IV estimator, $\hat{\theta}_h(IV)$, converges in probability to θ_h because the instrument is uncorrelated with the error term, $u_{t+h}^{(h)}$, which contains leads and lags of u_t and y_t and, hence, of w_t. It can be shown that

$$\frac{1}{T-h}\sum_{t=1}^{T-h} z_t y_{1t} \xrightarrow{p} c \quad \text{and} \quad \frac{1}{T-h}\sum_{t=1}^{T-h} y_{t+h} z_t \xrightarrow{p} \theta_h c,$$

where $\xrightarrow{p}$ signifies convergence in probability. Hence, as θ_0 is assumed to be standardized such that the first component is one, $\hat{\theta}_h(IV) \xrightarrow{p} \theta_h$, and the estimator is asymptotically normal under general conditions. We denote the corresponding estimator for the $(K \times (H+1))$ matrix Θ by $\widehat{\Theta}_{IV}$.

Stock and Watson (2018) note that adding control variables to the basic model (3.17) may be necessary if, in their framework, the proxy does not satisfy the relevance, exogeneity, and lead-lag exogeneity conditions without the controls. Controls can also reduce the variance of the IV estimator. Stock and Watson (2018) mention that lagged y_t and leads of z_t are possible control variables that can improve the efficiency of IV estimation. Therefore, in our simulations comparing different estimators in Section 4, we use as control variables $(1, y'_{t-1}, \dots, y'_{t-p})'$ when $h = 0$ and $(1, y'_{t-1}, \dots, y'_{t-p})'$ or $(1, y'_{t-1}, \dots, y'_{t-p}, z_{t+1}, \dots, z_{t+h})'$ for $h > 0$. The corresponding estimators of the impulse responses are denoted by $\widehat{\Theta}_{IV}^{y}$ and $\widehat{\Theta}_{IV}^{yz}$, respectively, where y and yz stand for the respective controls. Note that adding the additional regressors may, of course, create degrees-of-freedom problems in the estimation if the gross sample size T is small, given that computing $\widehat{\Theta}_{IV}^{yz}$ requires p presample and up to H lead values. Moreover, there are up to $Kp + H + 2$ regressors. It may also be worth noting that the first column of $\widehat{\Theta}_{IV}^{y}$ and $\widehat{\Theta}_{IV}^{yz}$, i.e., the estimator of the impact effects, is identical to $\hat{\theta}_0$ in $\widehat{\Theta}_{VAR}$ if that estimator is based on OLS residuals (see Appendix A.2).

As a final note on the IV estimators we mention that, if there is a vector of proxies, z_t, rather than just a scalar, an IV estimator of θ_h corresponding to (3.18) may be chosen as

$$\hat{\theta}_h(IV) = \sum_{t=1}^{T-h} y_{t+h} z'_t Q_z \sum_{t=1}^{T-h} z_t y_{1t} \Bigg/ \sum_{t=1}^{T-h} y_{1t} z'_t Q_z \sum_{t=1}^{T-h} z_t y_{1t}, \tag{3.19}$$

where $Q_z = (\sum_{t=1}^{T} z_t z'_t)^{-1}$, as before, and an intercept should be added to the regression equation (3.17) if the y_t are not mean-adjusted. Adding control variables is also straightforward and, of course, the lead values of the proxy would become lead vectors so that the number of required degrees of freedom for estimation will increase even further (see also Stock and Watson (2018)).

3.5 Residual-Based LP-type Estimators

For completeness we now discuss two new estimators which are easy to compute and asymptotically valid. However, our simulations show that they have poor small sample properties relative to some of the other estimators. Considering again a scalar proxy and using the LP error $v_{t+H}^{(H)}$, a simple LP-type estimator is obtained by noting that

$$\mathbb{E}(v_{t+H}^{(H)} z_{t+H-h})/\mathbb{E}(v_{1,t+H}^{(H)} z_{t+H}) = \Phi_h \theta_0, \tag{3.20}$$

for $h = 0, \dots, H$. Here $v_{1,t+H}^{(H)}$ denotes the first component of $v_{t+H}^{(H)}$. Hence, we may estimate the LP form of the VAR model for $h = H$ and use the estimated residuals, $\hat{v}_{t+H}^{(H)}$, to obtain an estimator of the structural impulse responses as

$$\widehat{\Theta}_{LP}^{resid} = \sum_{t=p+1}^{T-H} \hat{v}_{t+H}^{(H)}(z_{t+H}, \dots, z_t) \bigg/ \sum_{t=p+1}^{T-H} \hat{v}_{1,t+H}^{(H)} z_{t+H}. \tag{3.21}$$

For this estimator, the quantities $\hat{v}_{t+H}^{(H)}$ are obtained by estimating a model with K equations only and not KH equations as in LP estimation. Note, however, that the estimator differs from $\widehat{\Theta}_{VAR}$ even for θ_0 if $H \geq 1$. While the estimator of θ_0 in $\widehat{\Theta}_{VAR}$ is based on estimated errors $\hat{u}_t$ of the original reduced-form VAR model, this is clearly not the case in (3.21), where estimated LP errors $\hat{v}_{t+H}^{(H)}$ are used.

It is easy to see that the residual-based estimator can be viewed as a differentiable function of a two-step GMM estimator in the sense of Newey and McFadden (1994) which has standard asymptotic properties under general assumptions. In Section S1.2 of the Online Supplement we show that

$$\frac{1}{T-H-p} \sum_{t=p+1}^{T-H} \hat{v}_{t+H}^{(H)} z_{t+h} = \frac{1}{T-H-p} \sum_{t=p+1}^{T-H} v_{t+H}^{(H)} z_{t+h} + o_p(T^{-1/2}), \tag{3.22}$$

under general conditions. This result implies that the first column of $\widehat{\Theta}_{LP}^{resid}$ has the same asymptotic distribution as $\hat{\theta}_0$, the first column of $\widehat{\Theta}_{VAR}$. In small samples, the two estimators may differ substantially, however, because $\hat{\theta}_{0,LP}^{resid}$ is estimated from serially dependent observations and the correlation between $v_{t+H}^{(H)}$ and z_t is smaller than between u_t and z_t as $v_{t+H}^{(H)}$ has a larger variance than u_t. This reduced correlation may undermine the small sample properties of our estimator. Moreover, $\hat{\theta}_{0,LP}^{resid}$ may even be based on fewer observations than $\hat{\theta}_0$, because the former estimator is based on $T - H - p$ observations only, while $\hat{\theta}_0$ is based on $T - p$ observations.

The estimator $\widehat{\Theta}_{LP}^{resid}$ may be inefficient at least for $H > p$ because, even if $\hat{v}_{t+H}^{(H)}$ is replaced by $v_{t+H}^{(H)}$, the estimator is just a sample mean of serially correlated observations which does not account for possible restrictions on Θ that are due to the VAR structure. However, $\widehat{\Theta}_{LP}^{resid}$ is consistent under general conditions and very easy to compute. Given the limited samples available for some empirical studies, the small sample performance of the estimator may be an issue, however. In Section 4 we also explore the finite sample properties of this estimator.

To improve the estimator $\widehat{\Theta}_{LP}^{resid}$ in small samples, one may want to consider an estimator

$$\widehat{\Theta}_{LP}^{ss} = \left(\sum_{t=p+1}^{T} \hat{u}_t z_t, \sum_{t=p+1}^{T-1} \hat{v}_{t+1}^{(1)} z_t, \dots, \sum_{t=p+1}^{T-H} \hat{v}_{t+H}^{(H)} z_t \right) \Big/ \sum_{t=p+1}^{T} \hat{u}_{1t} z_t. \quad (3.23)$$

It requires estimating all the models $y_{t+h} = \nu_h + A^{(h+1)} Y_{t-1} + v_{t+h}^{(h)}$ for $h = 0, 1, \dots, H$ and, thus, the advantage of computational savings relative to the standard LP estimator is lost. On the positive side, small sample efficiency gains are conceivable. It follows from the result in equation (3.22) that the asymptotic properties of $\widehat{\Theta}_{LP}^{ss}$ are the same as for $\widehat{\Theta}_{LP}^{resid}$. Note, however, that the impact effects ($h = 0$) of $\widehat{\Theta}_{LP}^{ss}$ are estimated exactly as in $\widehat{\Theta}_{VAR}$. In other words, the first column of $\widehat{\Theta}_{LP}^{ss}$ is the same as for $\widehat{\Theta}_{VAR}$ and it does not only have the same asymptotic properties.

If there is an N-dimensional vector of proxies z_t, the expression (3.20) generalizes to

$$\mathbb{E}(v_{t+H}^{(H)} z_{t+H-h}') Q \mathbb{E}(z_{t+H} v_{1,t+H}^{(H)}) / \mathbb{E}(v_{1,t+H}^{(H)} z_{t+H}') Q \mathbb{E}(z_{t+H} v_{1,t+H}^{(H)}) = \Phi_h \theta_0,$$

for $h = 0, \dots, H$. Here Q is again an arbitrary positive definite $(N \times N)$ matrix. The estimator $\widehat{\Theta}_{LP}^{resid}$ of the structural impulse responses becomes

$$\widehat{\Theta}_{LP}^{resid} = \sum_{t=p+1}^{T-H} \hat{v}_{t+H}^{(H)} (z_{t+H}', \dots, z_t') \times I_{H+1} \otimes \left(Q_z \sum_{t=p+1}^{T-H} z_{t+H} \hat{v}_{1,t+H}^{(H)} \Big/ \sum_{t=p+1}^{T-H} \hat{v}_{1,t+H}^{(H)} z_{t+H}' Q_z \sum_{t=p+1}^{T-H} z_{t+H} \hat{v}_{1,t+H}^{(H)} \right)$$

with $Q_z = (\sum_{t=1}^{T} z_t z_t')^{-1}$, as before. Furthermore, an estimator $\widehat{\Theta}_{LP}^{ss}$ may be computed using the more general expression

$$\widehat{\Theta}_{LP}^{ss} = \left(\sum_{t=p+1}^{T} \hat{u}_t z_t', \sum_{t=p+1}^{T-1} \hat{v}_{t+1}^{(1)} z_t', \dots, \sum_{t=p+1}^{T-H} \hat{v}_{t+H}^{(H)} z_t' \right) \times I_{H+1} \otimes \left(Q_z \sum_{t=p+1}^{T} z_t \hat{u}_{1t} \Big/ \sum_{t=p+1}^{T} \hat{u}_{1t} z_t' Q_z \sum_{t=p+1}^{T} z_t \hat{u}_{1t} \right).$$

Table 1: Equality of Estimators Based on OLS Estimation

	$\widehat{\Theta}_{VAR}$	$\widehat{\Theta}_{LP}$	$\widehat{\Theta}_{LP}^{GLS}$	$\widehat{\Theta}_{LP}^{aug}$	$\widehat{\Theta}_{BB}$	$\widehat{\Theta}_{BB}^{GLS}$	$\widehat{\Theta}_{IV}$	$\widehat{\Theta}_{IV}^{y}$	$\widehat{\Theta}_{IV}^{yz}$	$\widehat{\Theta}_{LP}^{resid}$	$\widehat{\Theta}_{LP}^{ss}$
$\widehat{\Theta}_{VAR}$	all h										
$\widehat{\Theta}_{LP}$	$h=0,1$	all h									
$\widehat{\Theta}_{LP}^{GLS}$	$h=0,1$	$h=0,1$	all h								
$\widehat{\Theta}_{LP}^{aug}$	$h=0$	$h=0$	$h=0$	all h							
$\widehat{\Theta}_{BB}$	$h=0$	$h=0$	$h=0$	all h	all h						
$\widehat{\Theta}_{BB}^{GLS}$	$h=0$	$h=0$	$h=0$	$h=0,1$	$h=0,1$	all h					
$\widehat{\Theta}_{IV}$	–	–	–	–	–	–	all h				
$\widehat{\Theta}_{IV}^{y}$	$h=0$	$h=0$	$h=0$	$h=0$	$h=0$	$h=0$	–	all h			
$\widehat{\Theta}_{IV}^{yz}$	$h=0$	$h=0$	$h=0$	$h=0$	$h=0$	$h=0$	–	$h=0$	all h		
$\widehat{\Theta}_{LP}^{resid}$	–	–	–	–	–	–	–	–	–	all h	
$\widehat{\Theta}_{LP}^{ss}$	$h=0$	$h=0$	$h=0$	$h=0$	$h=0$	$h=0$	–	$h=0$	$h=0$	–	all h

Note: h denotes the propagation horizon.

3.6 Summary of Numerical Relations Between Estimators

For assessing the small sample properties of the estimators, it is useful to keep in mind that, based on plain OLS estimation, $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{LP}$, $\widehat{\Theta}_{LP}^{GLS}$, $\widehat{\Theta}_{LP}^{aug}$, $\widehat{\Theta}_{BB}$, $\widehat{\Theta}_{BB}^{GLS}$, $\widehat{\Theta}_{IV}^{y}$, $\widehat{\Theta}_{IV}^{yz}$, and $\widehat{\Theta}_{LP}^{ss}$ all have the same first column and, hence, yield identical estimates of the impact effects. Moreover, $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{LP}$, and $\widehat{\Theta}_{LP}^{GLS}$ as well as $\widehat{\Theta}_{BB}$ and $\widehat{\Theta}_{BB}^{GLS}$ have identical first two columns by construction. Also, if bias-corrected OLS estimation is used for the reduced-form VAR, $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{LP}$, $\widehat{\Theta}_{LP}^{GLS}$, $\widehat{\Theta}_{LP}^{aug}$, and $\widehat{\Theta}_{LP}^{ss}$ share the same first column and for $\widehat{\Theta}_{BB}$ and $\widehat{\Theta}_{BB}^{GLS}$ the first two columns are identical. All nine estimators $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{LP}$, $\widehat{\Theta}_{LP}^{GLS}$, $\widehat{\Theta}_{LP}^{aug}$, $\widehat{\Theta}_{BB}$, $\widehat{\Theta}_{BB}^{GLS}$, $\widehat{\Theta}_{IV}^{y}$, $\widehat{\Theta}_{IV}^{yz}$, and $\widehat{\Theta}_{LP}^{ss}$ should provide very similar estimates for the impact effects also if bias-corrected OLS is applied, at least for larger sample sizes for which the estimated bias tends to be small. Only the estimators $\widehat{\Theta}_{IV}$ and $\widehat{\Theta}_{LP}^{resid}$ estimate the impact effects clearly differently for both plain OLS and bias-corrected OLS estimation of the reduced-form VAR parameters.

For propagation horizons $h > 0$, the estimators $\widehat{\Theta}_{LP}^{aug}$ and $\widehat{\Theta}_{BB}$ are identical for plain OLS estimation. For VAR(p) processes with little persistence, also $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ may be quite similar, in particular, if the lag order p is already large. In that case, adding an extra lag may not make much difference. These relations may be useful to remember in the next section where we explore the performance of the different estimators in small samples. All exact identities of the estimators based on plain OLS estimation are summarized in Table 1 for easy reference.

3.7 Other Proposals

Given the large number of parameters in the estimation equations underlying some of the estimators, Bayesian and other shrinkage estimators have also been used in the present context. In their study, Li et al. (2021) explicitly consider also a Bayesian approach, a penalized LP approach which shrinks the impulse responses to smooth functions, as proposed by Barnichon and Brownlees (2019), and a model averaging approach which addresses the uncertainty in the lag lengths to be considered in practice. Such modifications can be used with most of the estimators considered in our study. They raise issues such as Bayesian prior selection and selecting the degree of shrinkage etc. which are not the focus of our study. Therefore we compare the estimators in raw form as presented in the foregoing sections and we leave such modifications to future research.

4 Monte Carlo Comparison

We conjecture that the sample size T, the dimension, the lag order and the persistence of the VAR process as well as the correlation between the shock of interest and the proxy, i.e., the strength of the proxy, are features that have an impact on the small sample properties of the estimators for the impulse responses. Therefore we choose data generating processes (DGPs) accordingly.

We consider the RMSEs of the impulse response estimators as our main performance criterion for estimator comparison. As confidence intervals of impulse responses are often examined in empirical analysis, the coverage and length of bootstrap confidence intervals for the impulse responses are also important performance criteria. However, this raises the issue which bootstrap method to use. Different bootstraps have been considered in related contexts. For example, Stock and Watson (2018) and Breitung and Brüggemann (2019) use a parametric bootstrap (see Stock and Watson (2018, Appendix A.2)) and Montiel Olea and Plagborg-Møller (2021) recommend a wild bootstrap to construct equal-tailed percentile-t intervals for the lag augmented LP method. We decided to use a moving-block bootstrap (MBB) to construct percentile confidence intervals (see Section S2.2 of the Online Supplement for details). Jentsch and Lunsford (2019) show that the MBB yields asymptotically valid confidence intervals for structural impulse responses in proxy VAR analysis under general conditions. They also show that other bootstraps such as the wild bootstrap, that have been used in structural VAR analysis, do not yield confidence intervals with asymptotically correct coverage. Unfortunately, there is also evidence that the MBB may not be very accurate in small samples (e.g., Bruns and Lütkepohl (2020)). As we are primarily interested in the relative performance of the different estimators, we prefer the asymptotically valid MBB and assume that it imposes a similar small sample handicap on all estimators. It is well possible, however, that some of the estimators perform better with alternative bootstraps in terms of coverage and interval length. Therefore we give priority to the RMSE as performance criterion. If an estimator results in small RMSEs it may be possible to construct a suitable bootstrap superior to the MBB. Constructing better bootstraps is not the aim of this study, however.

To improve the small sample coverage rates of the MBB confidence intervals, we also use bias-corrected OLS estimation in addition to plain OLS estimation for the reduced-form VAR models (see Section S2.1 of the Online Supplement) and we primarily report the results for bias-corrected OLS estimation, if not otherwise stated. Bias-corrected OLS estimation was shown to improve small sample inference for impulse responses based on the standard VAR approach (see Kilian (1998)) and we confirmed that in unreported simulations. The corresponding residuals of bias-corrected OLS are used for generating the bootstrap samples for all other estimators as well. Moreover, the estimators $\widehat{\Theta}_{LP}$, $\widehat{\Theta}_{LP}^{aug}$, $\widehat{\Theta}_{LP}^{ss}$, $\widehat{\Theta}_{BB}$, $\widehat{\Theta}_{LP}^{GLS}$ and $\widehat{\Theta}_{BB}^{GLS}$ are also based on the residuals $\hat{u}_t$ of bias-corrected reduced-form OLS estimators wherever these residuals $\hat{u}_t$ enter the estimator.

Breitung and Brüggemann (2019) and Li et al. (2021) also perform Monte Carlo experiments to compare some of the estimators considered in the following. Breitung and Brüggemann (2019) use a DGP similar to DGP1 below and their performance criteria are the bias and standard deviation of the estimators as well as coverage and length of bootstrap confidence intervals based on their different bootstrap. As we will see later, their results are in line with our results for those estimators and simulation designs considered in their study. Li et al. (2021) consider a very large range of DGPs which are not finite-order VARs but are approximated by finite-order VARs in their study. They are specifically interested in the bias-variance trade-off of the estimators and include some shrinkage estimators in their comparison. Our results are roughly in line with their results for the overlapping estimators. Our focus is more limited, however, given that we consider finite-order VARs as DGPs and investigate the properties of the estimators in an idealized setting.

4.1 Monte Carlo Setup

4.1.1 DGP1

Our first DGP is a bivariate VAR(1), $y_t = A_1 y_{t-1} + u_t$, where y_t is a 2-dimensional vector of endogenous variables, A_1 is a matrix of autoregressive slope coefficients, and u_t is the white-noise reduced-form error term. The VAR slope coefficients are chosen similar to Kilian and Kim (2011), Breitung and Brüggemann (2019), Lütkepohl et al. (2015a) and other studies comparing impulse response estimators for VAR processes, where such a DGP has been considered. More precisely, we choose

$$A_1 = \begin{bmatrix} a_{11} & 0 \\ 0.5 & 0.5 \end{bmatrix},$$

with $a_{11} = 0.1, 0.5, 0.9, 0.95$. The process is stable and its persistence depends on a_{11}. If a_{11} is close to one, the persistence is high and it is low for a_{11} close to zero.

The structural shocks are standard normal, $w_t \sim \mathcal{N}(0, I_2)$, and $u_t = Bw_t$ with

$$B = \begin{bmatrix} 1 & 0 \\ 0.5 & 3 \end{bmatrix}.$$

In line with the related literature (e.g., Caldara and Herbst (2019), Lütkepohl and Schlaak (2021), Breitung and Brüggemann (2019)), a scalar proxy z_t is generated as

$$z_t = \phi w_{1t} + \eta_t, \tag{4.1}$$

where ϕ and the error η_t determine the strength of the correlation between z_t and w_{1t} and, hence, the strength of the instrument. The error term η_t is generated independently of w_t as $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$. Thus, the proxy not only ensures that the relevance and exogeneity conditions (2.7) and (2.8) hold but it also satisfies the lead-lag exogeneity condition (2.10).

Note that the strength of the relation between the instrument and the shock w_{1t} determines how well the impact effects of the shock can be estimated and these estimates are of central importance for estimating the impulse responses. Therefore different scenarios are considered. The variance of z_t is $\mathrm{Var}(z_t) = \phi^2 \mathrm{Var}(w_{1t}) + \sigma_\eta^2$. Hence, the correlation between w_{1t} and z_t is $\mathrm{Corr}(w_{1t}, z_t) = \phi\sqrt{\mathrm{Var}(w_{1t})} \Big/ \sqrt{\phi^2 \mathrm{Var}(w_{1t}) + \sigma_\eta^2}$. We consider the two different cases presented in Table 2. For correlation 0.9, z_t is a strong proxy while a correlation of 0.5 gives a proxy with intermediate strength.

Table 2: Specifications Used for the Proxy for DGP1

Case	ϕ	σ_η^2	$\text{Corr}(w_{1t}, z_t)$
1	1	0.2346	0.9 $\forall t$
2	1	3	0.5 $\forall t$

As the residual-based estimation also depends on the propagation horizon H, we consider an intermediate value of $H = 20$. As sample sizes we use $T = 100$, 200, and 500. The former value represents the order of magnitude used in macroeconometric studies based on quarterly data, whereas $T = 500$ is hoped to reflect the properties of the estimators for larger samples and $T = 200$ represents an intermediate sample size. The number of bootstrap replications is $N = 2000$ and the number of Monte Carlo replications is 1000 for all reported simulation results.

4.1.2 DGP2

Our second DGP is linked to an empirical model from the proxy VAR literature. More precisely, DGP2 is based on a model by Mertens and Ravn (2013), which employs seven variables at quarterly frequency from 1950Q1 - 2006Q4, giving $T = 228$ observations. We fit a VAR(1) process including a constant to their data and, after bias-adjustment, obtain the following set of parameters:

$$A_1 = \begin{bmatrix} 0.88 & 0.01 & 0.03 & 0.00 & 0.00 & -0.02 & -0.00 \\ -0.11 & 0.83 & -0.02 & 0.02 & -0.02 & -0.03 & 0.00 \\ 0.14 & -0.08 & 0.85 & 0.02 & -0.01 & 0.11 & 0.01 \\ -1.47 & -0.32 & -0.90 & 0.86 & -0.06 & 1.05 & 0.05 \\ -0.27 & 0.06 & 0.48 & 0.04 & 0.93 & -0.53 & 0.00 \\ -0.09 & -0.08 & 0.04 & 0.01 & -0.01 & 0.92 & 0.00 \\ -0.12 & -0.19 & -0.44 & -0.07 & 0.03 & 0.49 & 1.01 \end{bmatrix}$$

and

$$\Sigma_u = \begin{bmatrix} .021 & .004 & .007 & .003 & .013 & .005 & -.010 \\ .004 & .286 & .006 & .002 & -.024 & .010 & -.053 \\ .007 & .006 & .084 & .172 & .022 & .072 & .006 \\ .003 & .002 & .172 & 2.347 & -.025 & .328 & -.007 \\ .013 & -.024 & .022 & -.025 & .789 & .040 & .077 \\ .005 & .010 & .072 & .328 & .040 & .102 & .011 \\ -.010 & -.053 & .006 & -.007 & .077 & .011 & .078 \end{bmatrix} \times 10^{-3}.$$

The largest eigenvalue of A_1 has modulus 0.99995, implying a stable but very persistent process. The constant is estimated as

$$\nu = (0.09, -0.60, -0.46, 0.08, -1.02, -0.41, 0.23)'.$$

The VAR(1) with these parameters is used to generate y_t based on Gaussian u_t, $u_t \sim \mathcal{N}(0, \Sigma_u)$. We generate $2T$ observations starting from

$$y_0 = (0.17, 0.30, -15.15, -17.32, -14.66, -15.14, -15.86)',$$

the unconditional mean of y_t, and discard the first T observations to alleviate the effect of the starting value.

A proxy is generated so as to have similar properties as the proxy for shocks to personal income taxes in Mertens and Ravn (2013). More precisely, we estimate the b vector of impact effects of the first shock giving

$$b = (1.00, 2.07, 0.09, -9.67, 0.57, -1.11, 0.66)'$$

and generate the first shock using equation (3.11). Then we estimate the parameters ϕ and σ_η^2 of the proxy model (4.1) using the full sample from 1950Q1 - 2006Q4. This yields estimates $\phi = 464.18$ and $\sigma_\eta^2 = 0.32$. The original proxy by Mertens and Ravn (2013) has a correlation with the identified shock of 0.19, i.e., the proxy is rather weak. Only 7% of its values are non-zero. Instead, we employ a proxy with nonzero values for all sample periods and a correlation of 0.90 with the shock of interest, implying a strong proxy.

To mimic the situation in the Mertens/Ravn study where the proxy has many zero values, we follow Jentsch and Lunsford (2019) and also generate a proxy as

$$z_t = D_t(\phi w_{1t} + \eta_t), \tag{4.2}$$

where D_t is a series of independent, identically distributed Bernoulli 0-1 random variables with parameter d, $0 < d \leq 1$, which signifies the probability of a nonzero value. The random process D_t is assumed to be stochastically independent of w_{1t} and the error term η_t. The latter term is again generated to have mean zero and variance σ_η^2, i.e., $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, and it is distributed independently of w_{1t}. In this case, the correlation between w_{1t} and z_t is

$$\text{Corr}(w_{1t}, z_t) = \phi\sqrt{d}\sqrt{\text{Var}(w_{1t})} \Big/ \sqrt{\phi^2\text{Var}(w_{1t}) + \sigma_\eta^2}$$

and we choose the same values for ϕ and σ_η^2 as before ($\phi = 464.18$ and $\sigma_\eta^2 = 0.32$) and $d = 0.3$, which leads to a correlation of 0.5, implying a proxy with intermediate strength.

Note that the generation mechanism for DGP2 differs from that of DGP1, where the structural shocks are generated directly and the reduced-form data as well as the proxy are computed from the generated structural shocks and the generated η_t series. In contrast, we generate the reduced-form errors for DGP2, construct the first structural shock from the structural parameters b and the error covariance matrix Σ_u as in (3.11) and then generate z_t as in equation (4.1) or (4.2), depending on the strength of the considered proxy.

As for DGP1, we use a maximal propagation horizon of $H = 20$ but consider sample sizes $T = 200$ and 500 only. A sample size of $T = 100$ leaves insufficient degrees of freedom for some of the estimators for the higher-dimensional DGP2. The number of bootstrap replications for this DGP is again $N = 2000$ and the number of Monte Carlo replications is 1000.

4.2 Monte Carlo Results

4.2.1 Based on DGP1

A first assessment of the various projection estimators has shown that $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}^{aug}_{LP}$ dominate the IV estimators $\widehat{\Theta}_{IV}$, $\widehat{\Theta}^{y}_{IV}$, and $\widehat{\Theta}^{yz}_{IV}$ while $\widehat{\Theta}^{GLS}_{LP}$ and $\widehat{\Theta}^{GLS}_{BB}$ dominate $\widehat{\Theta}^{resid}_{LP}$, $\widehat{\Theta}^{ss}_{LP}$, and $\widehat{\Theta}_{BB}$ in terms of our performance criteria. Therefore we first compare the estimators $\widehat{\Theta}_{LP}$, $\widehat{\Theta}^{aug}_{LP}$, $\widehat{\Theta}^{GLS}_{LP}$, and $\widehat{\Theta}^{GLS}_{BB}$ with the benchmark estimator $\widehat{\Theta}_{VAR}$ in Figure 1. Results for the other estimators are collected in the Online Supplement and will be discussed subsequently.

In Figure 1, RMSEs, pointwise coverage rates of nominal 90% confidence intervals, and average interval lengths for the responses of variable 2 to the first structural shock are presented for selected simulation designs for DGP1.[3] The selected results provide an overview of the overall results for DGP1. We present results for simulation designs which vary the sample size T, the proxy-shock correlation and the persistence (a_{11}) of the DGP.[4]

It can be seen in the figure that all five estimators yield the same RMSEs, coverages, and interval lengths for the impact effects and are very similar for small propagation horizons. Recall that $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{LP}$, $\widehat{\Theta}^{GLS}_{LP}$, and $\widehat{\Theta}^{aug}_{LP}$ yield identical impact effects by construction and, as discussed in Section 3.3, the

[3]The corresponding results for variable 1 are presented in Figure S.2 in the Online Supplement. They confirm the conclusions drawn for variable 2. Note that, given the importance of the impact effects, the results for variable 2 are presented in Figure 1 because for that variable the impact effects of the shock are estimated whereas they are set to one for variable 1.

[4]Note that all average interval lengths are strictly positive. For some of the estimators they are relatively small, however, so that they cannot be distinguished from zero in Figure 1 and similar figures.

impact effects of $\widehat{\Theta}_{BB}^{GLS}$ would be the same as well if plain OLS estimation was used and they are very similar for bias-corrected OLS.

A striking result is that $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{VAR}$ yield very similar RMSEs and confidence intervals also for $h > 0$, which shows that $\widehat{\Theta}_{BB}^{GLS}$ is not only very efficient asymptotically, as shown by Breitung and Brüggemann (2019), but also in small samples. This result is fully in line with simulations reported by Breitung and Brüggemann (2019). The $\widehat{\Theta}_{LP}^{GLS}$ estimator outperforms $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ but is still less efficient in terms of RMSE than $\widehat{\Theta}_{BB}^{GLS}$, in particular for processes with medium persistence (see Figure 1(a), (b), (e), (f)).

Given previous simulation results by Kilian and Kim (2011), it is, of course, not surprising that $\widehat{\Theta}_{VAR}$ dominates $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ for our simulation designs. As $\widehat{\Theta}_{BB}^{GLS}$ performs about as well as $\widehat{\Theta}_{VAR}$, it is clearly the preferred projection estimator for DGP1. We stress, however, that the good coverage rates of confidence intervals associated with $\widehat{\Theta}_{BB}^{GLS}$ rely to some extent on the use of bias-corrected OLS estimators for the reduced-form VAR. In Figure S.4 in the Online Supplement, we show the corresponding results obtained when $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ are based on plain OLS estimation. Clearly, some coverage rates for these estimators for the more persistent processes ($a_{11} = .95$) are then far below the nominal 90% and much worse than with bias-correction (see Figure S.4(c), (d)). We note that Breitung and Brüggemann (2019) report better coverage rates for $\widehat{\Theta}_{BB}^{GLS}$ based on their alternative bootstrap method. Thus, using other bootstrap methods rather than bias-corrected VAR estimates may also improve the interval coverage associated with the GLS estimators.

For all simulation designs, the coverage rates of the confidence intervals associated with all estimators in Figure 1 are reasonably close to or larger than 90%. Most coverage rates are above 80% and in many cases the coverage rates are close to or at 100%. In other words, the estimators yield conservative intervals. Only for more persistent processes ($a_{11} = 0.95$), the coverage rates for $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ are below 80% for some propagation horizons (Figure 1(c), (d)). For some of the simulation designs the coverage rates of all estimators are actually rather similar, e.g., for designs with medium persistence ($a_{11} = 0.5$) (see Figure 1(a), (b), (e), (f)). Generally, the interval lengths tend to increase for processes with larger persistence, and if the proxy has lower correlation with the first shock. Note, however, that the interval lengths and RMSEs for $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ for longer propagation horizons are not much affected by these features. As one would expect, for all five estimators, interval lengths and RMSEs decline with increasing sample size.

To explore the impact of estimating models with larger lag orders, we show the effect of increasing the lag length to $p = 12$ in Figure S.1 of the

Online Supplement. A higher lag order leads to an increased RMSE for $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{BB}^{GLS}$, and $\widehat{\Theta}_{LP}^{GLS}$ at shorter horizons, while $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ are less affected. The ranking of the estimators is maintained, however. The coverage rates remain acceptable for all estimators. In terms of average length, the impact effect is estimated less precisely for all estimators for small sample size $T = 100$. The average length for $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ starts to approach zero after horizon $h = p = 12$, while $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ do not share this feature, owing to their horizon-specific estimation approach. Given the limited impact of changing the lag order on the ranking of the estimators, we have not explored the possibility of choosing the lag order by some model selection criterion.

Some results for the estimators $\widehat{\Theta}_{IV}$, $\widehat{\Theta}_{IV}^{y}$, $\widehat{\Theta}_{IV}^{yz}$, $\widehat{\Theta}_{LP}^{resid}$, and $\widehat{\Theta}_{LP}^{ss}$ are presented in Figure S.3 in the Online Supplement, where these estimators are compared with $\widehat{\Theta}_{VAR}$. Recall that $\widehat{\Theta}_{VAR}$, $\widehat{\Theta}_{IV}^{y}$, $\widehat{\Theta}_{IV}^{yz}$, and $\widehat{\Theta}_{LP}^{ss}$ are identical by construction for $h = 0$. For $h > 0$, it is obvious in Figure S.3 that $\widehat{\Theta}_{VAR}$ uniformly dominates the other five estimators in terms of RMSE. In other words, it yields very similar or smaller RMSEs for all simulation designs presented in the figure. Compared to the results in Figure 1, $\widehat{\Theta}_{IV}$, $\widehat{\Theta}_{IV}^{y}$, and $\widehat{\Theta}_{IV}^{yz}$ have higher RMSEs than $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$. In addition, $\widehat{\Theta}_{LP}^{resid}$, and $\widehat{\Theta}_{LP}^{ss}$ are dominated in terms of RMSE by $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$. As in Figure 1, the RMSEs of the estimators in Figure S.3 tend to increase with larger persistence (larger a_{11}) and smaller correlation between shock and proxy. Moreover, the RMSEs decrease for increasing sample size. In terms of coverage, all estimators in Figure S.3 are reasonably close to 90%. The average length for $\widehat{\Theta}_{IV}^{y}$ and $\widehat{\Theta}_{LP}^{resid}$ is much higher than for the other 4 estimators, especially when the proxy has a lower correlation with the shock.

The overall takeaway from the simulations of the bivariate DGP1 is that smaller samples, weaker instruments, and larger lag orders tend to increase RMSEs and the lengths of confidence intervals. The LP estimators $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ dominate the IV estimators and, among the projection estimators, the lag-augmented GLS estimator, $\widehat{\Theta}_{BB}^{GLS}$, dominates all other estimators including the residual-based LP-type estimators and the LP GLS estimator, $\widehat{\Theta}_{LP}^{GLS}$, clearly in terms of RMSE. Thus, using $\widehat{\Theta}_{BB}^{GLS}$ would be the best choice for DGPs such as DGP1.

4.2.2 Results Based on DGP2

In Figure 2, results for the seven-dimensional, very persistent DGP2 are presented. The estimators are grouped in the same way as in Figure 1 and we again consider Monte Carlo designs with different sample sizes and proxy strengths. However, now we present responses of two different variables to the first shock. There is some variation in the performance of the estimators across the seven variables. Therefore we present impulse responses of variables two and four, for which the performance is different. The results for variable four are in fact a bit special. In Figure S.9b in the Online Supplement, the original time series from Mertens and Ravn (2013) are plotted which are the basis for our DGP2. In that figure, variable 4 (Corporate income tax base) is seen to have rather distinct dynamics which may be reflected in our simulation results. The impulse responses of the remaining variables are also presented in Figures S.5a - S.5c in the Online Supplement. They display a similar overall picture and are mostly more similar to the results for variable two.

Before we discuss the figures in more detail, it may be worth mentioning that some crucial features are the same as for the bivariate DGP1. The coverage rates improve and the average lengths of the confidence intervals and the RMSEs tend to decline for all estimators with increasing sample size. Also using a stronger proxy tends to improve the estimation precision as measured by the RMSE. Note that the figures are scaled so as to bring out clearly the differences between and similarities of the alternative estimators. Therefore some RMSEs and average interval lengths in some of the figures in the Online Supplement had to be truncated at the upper limit of the vertical axis.

In Figure 2 the four estimators $\widehat{\Theta}_{LP}$, $\widehat{\Theta}_{LP}^{aug}$, $\widehat{\Theta}_{BB}^{GLS}$, and $\widehat{\Theta}_{LP}^{GLS}$ are compared to the standard VAR estimator $\widehat{\Theta}_{VAR}$. In contrast to what we found for DGP1, now $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ are both very similar to $\widehat{\Theta}_{VAR}$ in terms of RMSE and interval coverage and length. These three estimators clearly dominate $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ in terms of associated RMSE and interval length. Note, however, that the coverage rates associated with $\widehat{\Theta}_{BB}^{GLS}$, $\widehat{\Theta}_{LP}^{GLS}$, and $\widehat{\Theta}_{VAR}$ for some intermediate propagation horizons leave room for improvement (see Figure 2(e), (f), (g) and (h)).

Compared to the alternative estimators shown in Figure S.7, the results of DGP1 are broadly confirmed: For both variables $\widehat{\Theta}_{IV}$, $\widehat{\Theta}_{IV}^{y}$, and $\widehat{\Theta}_{IV}^{yz}$ are dominated in terms of RMSEs by $\widehat{\Theta}_{LP}$ and $\widehat{\Theta}_{LP}^{aug}$ at all horizons. In addition, $\widehat{\Theta}_{LP}^{resid}$, and $\widehat{\Theta}_{LP}^{ss}$ are dominated by $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$. When increasing the lag length to $p = 4$ (see Figure S.6 in the Online Supplement), the RMSE is affected only marginally, except for small samples ($T = 200$) and weak proxies. For this case, the RMSEs for all estimators increase for both variables (see Figure S.6(b) and (f)).

To show that the strong performance of $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ depends again to some extent on the use of bias-corrected OLS estimation of the VAR reduced form, we compare the estimators $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ based on plain OLS estimates in Figure S.8 in the Online Supplement to the VAR estimator with bias-correction. As DGP2 is very persistent, the coverage rates associated with $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ are very poor for most designs, while the impact of avoiding bias-correction on interval lengths and RMSEs is rather limited. Our recommendation is therefore to use $\widehat{\Theta}_{BB}^{GLS}$ and $\widehat{\Theta}_{LP}^{GLS}$ with bias-correction if a MBB is used for constructing confidence intervals. Again, one could argue that the features of the MBB confidence intervals are less important than the RMSE criterion because the former may reflect the properties of the MBB rather than the properties of the estimators.

It is perhaps worth mentioning, however, that the maximal lag order and propagation horizon $\widehat{\Theta}_{BB}^{GLS}$ can handle even for a gross sample size of $T = 200$ is, of course, a bit more limited than that of some of the other estimators because it needs p presample values and up to H lead values and it involves a rather substantial number of regressors. Thus, its net sample size quickly exhausts the degrees of freedom needed for estimation when the lag order or the propagation horizon increases for a model with many variables. Therefore, it is not surprising that, for larger propagation horizons, $\widehat{\Theta}_{LP}^{GLS}$ is occasionally marginally better in terms of RMSE than $\widehat{\Theta}_{BB}^{GLS}$ when the lag order is increased to $p = 4$ (see Figure S.6(e) and (f) in the Online Supplement).

In summary, the overall conclusion from the simulation results for DGP1 and DGP2 is that $\widehat{\Theta}_{BB}^{GLS}$ is the best projection estimator in terms of RMSE and $\widehat{\Theta}_{LP}^{GLS}$ comes in second. Apparently, for higher-dimensional processes the superior performance of $\widehat{\Theta}_{BB}^{GLS}$ that we observed for the bivariate DGP1 may decline relative to $\widehat{\Theta}_{LP}^{GLS}$.

5 Conclusions

This study compares a range of projection estimators for impulse responses of proxy VAR models. Using LP estimators in this context has become increasingly popular lately because these estimators are easy to apply and have a reputation of being robust to some model deficiencies. On the other hand, there is some evidence from simulation studies showing that the standard LP estimators may be quite inefficient if the true DGP is a finite-order VAR process. Such results have motivated researchers to look for modifications and alternatives to classical LP estimators. We review a number of alternative approaches and then compare them algebraically and in a simulation study. We present conditions for some estimators to be identical in small samples. In our simulation study, we use the RMSE as well as coverage rates and interval lengths of bootstrap confidence intervals as performance criteria.

We find that generally the estimators behave as expected in that the RMSEs and confidence intervals improve for increasing sample size and when stronger proxies (proxies with higher correlation with the shock of interest) are used. Moreover, estimation precision tends to decline when more heavily parametrized models with larger lag orders are considered. Furthermore, processes with higher persistence may lead to less precisely estimated impulse responses.

Ranking our estimators, we find that overall a lag-augmented GLS approach proposed by Breitung and Brüggemann (2019) and a GLS approach of Lusompa (2021) lead to the most precise projection estimators in small samples. All other estimators are typically less precise than the standard VAR estimators if a finite-order VAR process is the true DGP. In contrast, for moderately large samples the lag-augmented GLS approach performs about as well as the standard VAR approach, if it is used with bias-corrected OLS estimates of the reduced-form VAR model. As the Breitung and Brüggemann (2019) GLS approach involves many regressors, it is not a suitable choice in small samples if the number of variables in the model and/or the lag order is large and the desired propagation horizon for the impulse responses is also large because it quickly exhausts the degrees of freedom for estimation in that case. The coverage rates of MBB confidence intervals can be improved by using bias-corrected OLS estimation of the reduced-form VAR. Generally, it would be desirable to design bootstrap procedures that work well asymptotically and in small samples with the GLS estimators because bootstrap inference is quite common in structural VAR analysis. The fact, that Breitung and Brüggemann (2019) obtained better coverage rates for their lag-augmented GLS estimator by using an alternative bootstrap method in their simulation study suggests that there is scope for future research in that direction.

Our results suggest the following recommendations for empirical work if a finite-order vector autoregression is likely to be a good approximation to the true DGP and the proxy is a reasonably strong instrument. Given the simulation setup, a conventional proxy VAR approach is a benchmark which outperforms the alternative estimators. For these alternative estimators, our results imply: (i) If the VAR lag order and impulse response propagation horizon are such that the sample size is sufficiently large for estimation, then GLS LP approaches are worth considering because they yield more precise impulse response estimates than traditional proxy VAR LP estimators. (ii) Bias-correction improves bootstrap inference for the two LP GLS estimators if the MBB is used to construct confidence intervals for impulse responses.

It may also be worth noting that LP estimators based on a proxy VAR approach performed well in a large-scale simulation study by Li et al. (2021) which is based on a very different Monte Carlo design where finite-order VAR processes and LP estimators approximate more general DGPs. Thus, our results may be relevant for more general settings as well. Li et al. (2021) also consider shrinkage methods such as Bayesian methods and penalized estimation which shrinks to smooth impulse response functions to cope with the uncertainty induced by the large number of control variables in some of the LP equations. They find that, in their simulation scenario, shrinkage can indeed improve the bias-variance trade-off and, thus, it may reduce the RMSE. Of course, shrinkage can also be applied in conjunction with the estimators considered in the present study and is, hence, a topic that may be worth exploring in future research.

A Appendix

A.1 Equivalence of Lag-augmented and Breitung-Brüggemann Projection Estimators

As shown in Section 3.3, the equivalence of the lag-augmented LP estimator and the Breitung-Brüggemann estimator follows from the representation

$$y_{t+h} = \nu_{h-1} + \Theta_h w_t + A_*^{(h)} Y_{t-1} + v_{t+h}^{(h-1)} \tag{A.1}$$

which can be obtained by first replacing in (3.8) t by $t+1$ such that

$$\begin{aligned} y_{t+h+1} &= \nu_h + A^{(h+1)} Y_t + A_{p+1}^{(h+1)} y_{t-p} + v_{t+h+1}^{(h)} \\ &= \nu_h + \Phi_{h+1} y_t + A_{\#}^{(h+1)} Y_{t-1} + v_{t+h+1}^{(h)} \\ &= \nu_h + \Phi_{h+1} u_t + A_*^{(h+1)} Y_{t-1} + v_{t+h+1}^{(h)}. \end{aligned}$$

Using $\Phi_{h+1} u_t = \Theta_{h+1} w_t$ and then replacing h by $h-1$ gives the representation (A.1).

The expression for computing the first shock when $\theta_0 = b$ is given,

$$w_{1t} = b'\Sigma_u^{-1} u_t / b'\Sigma_u^{-1} b,$$

is obtained by noting that $\Sigma_u = B\Sigma_w B'$ implies $b'\Sigma_u^{-1} u_t = b'(B\Sigma_w B')^{-1} u_t = b'B'^{-1}\Sigma_w^{-1} B^{-1} u_t = w_{1t}/\sigma_{w_1}^2$, where $b'B'^{-1} = (1, 0, \dots, 0)$ has been used and $\sigma_{w_1}^2$ denotes the variance of w_{1t}. Moreover, $\Sigma_u^{-1} = B'^{-1}\Sigma_w^{-1} B^{-1}$ implies $b'\Sigma_u^{-1} b = 1/\sigma_{w_1}^2$. Putting things together, we get the above relation (see also Stock and Watson (2018, Footnote 6, p. 933)).

A.2 An Equivalence Result for the IV Estimator

Let $Y = (y_1, \dots, y_T)$, $\mathbf{y}_1 = (y_{11}, \dots, y_{1T})$, $Y_{-1} = (Z_0, \dots, Z_{T-1})$, where $Z_{t-1} = (1, y'_{t-1}, \dots, y'_{t-p})'$, $U = (u_1, \dots, u_T)$, $A = (\nu, A_1, \dots, A_p)$ and recall that OLS estimation of the model

$$Y = AY_{-1} + U$$

results in OLS errors $\hat{U} = Y - \hat{A}Y_{-1} = Y(I_T - Y'_{-1}(Y_{-1}Y'_{-1})^{-1}Y_{-1})$. Now consider

$$Y = \theta_0 \mathbf{y}_1 + AY_{-1} + \tilde{U}.$$

Estimating the model by OLS conditionally on θ_0 gives an estimator $\hat{A} = (Y - \theta_0\mathbf{y}_1)Y'_{-1}(Y_{-1}Y'_{-1})^{-1}$. Replacing A in the model equation by this estimator and rearranging terms gives

$$Y(I_T - Y'_{-1}(Y_{-1}Y'_{-1})^{-1}Y_{-1}) = \theta_0\mathbf{y}_1(I_T - Y'_{-1}(Y_{-1}Y'_{-1})^{-1}Y_{-1}) + \hat{\tilde{U}}$$

or

$$\hat{U} = \theta_0\hat{\mathbf{u}}_1 + \hat{\tilde{U}},$$

where $\hat{\mathbf{u}}_1$ is the first row of $\hat{U}$. In other words, instead of the model with controls, we can equivalently consider the model

$$\hat{u}_t = \theta_0\hat{u}_{1t} + error_t$$

so that an IV estimator with z_t as an instrument for $\hat{u}_{1t}$ is

$$\sum_{t=p+1}^{T} \hat{u}_t z_t \Bigg/ \sum_{t=p+1}^{T} \hat{u}_{1t} z_t,$$

the same as in (3.1).

References

Angelini, G. and Fanelli, L. (2019). Exogenous uncertainty and the identification of structural vector autoregressions with external instruments, *Journal of Applied Econometrics* **34**: 951–971.

Angrist, J. D., Jordà, O. and Kuersteiner, G. M. (2018). Semiparametric estimates of monetary policy effects: String theory revisited, *Journal of Business and Economic Statistics* **36**: 371–387.

Barnichon, R. and Brownlees, C. (2019). Impulse response estimation by smooth local projections, *Review of Economics and Statistics* **101**: 522–530.

Breitung, J. and Brüggemann, R. (2019). Projection estimators for structural impulse responses, *Technical report*, University of Konstanz.

Brugnolini, L. (2017). About local projection impulse response function reliability, *Technical report*, Central Bank of Malta.

Bruns, M. and Lütkepohl, H. (2020). An alternative bootstrap for proxy vector autoregressions, *Discussion Paper 1913*, DIW, Berlin.

Caldara, D. and Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy svars, *American Economic Journal: Macroeconomics* **11**: 157–192.

Choi, C.-Y. and Chudik, A. (2019). Estimating impulse response functions when the shock series is observed, *Economics Letters* **180**: 71–75.

Dolado, J. J. and Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems, *Econometric Reviews* **15**: 369–386.

Dufour, J.-M., Pelletier, D. and Renault, E. (2006). Short run and long run causality in time series: inference, *Journal of Econometrics* **132**: 337–362.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal: Macroeconomics* **7**: 44–76.

Jentsch, C. and Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Comment, *American Economic Review* **109**: 2655–2678.

Jordà, O. (2005). Estimation and inference of impulse responses by local projection, *American Economic Review* **95**: 161–182.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* **80**: 218–230.

Kilian, L. and Kim, Y. (2011). How reliable are local projection estimators of impulse responses?, *Review of Economics and Statistics* **93**: 1460–1466.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press, Cambridge.

Li, D., Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections vs. VARs: Lessons from thousands of DGPs, *Technical report*, Princeton University.

Lusompa, A. (2021). Local projections, autocorrelation, and efficiency, *Federal Reserve Bank of Kansas City Working Paper* (21-01).

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. and Schlaak, T. (2021). Heteroskedastic proxy vector autoregressions, *Journal of Business and Economic Statistics* (forthcoming).

Lütkepohl, H., Staszewska-Bystrova, A. and Winker, P. (2015a). Comparison of methods for constructing joint confidence bands for impulse response functions, *International Journal for Forecasting* **31**: 782–798.

Lütkepohl, H., Staszewska-Bystrova, A. and Winker, P. (2015b). Confidence bands for impulse responses: Bonferroni versus Wald, *Oxford Bulletin of Economics and Statistics* **77**: 800–821.

Meier, A. (2005). How big is the bias in estimated impulse responses? A horse race between var and local projection methods, *Manuscript*, European University Institute, Florence.

Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**: 1212–1247.

Montiel Olea, J. L. and Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think, *Econometrica* **89**: 1789–1823.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, Amsterdam, chapter 36, pp. 2111–2245.

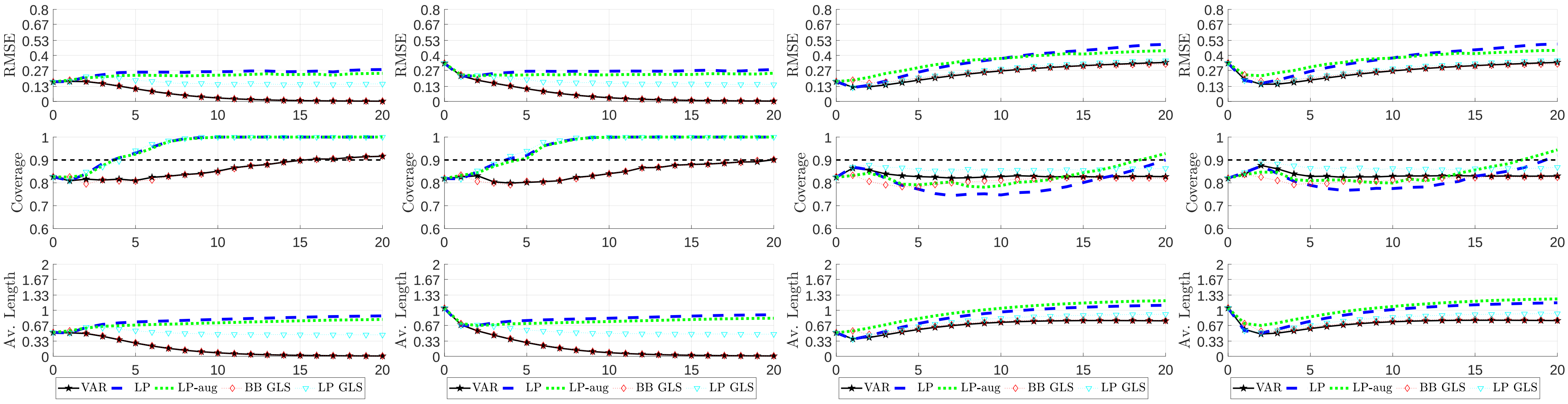
Plagborg-Møller, M. and Wolf, C. K. (2017). Instrumental variable identification of dynamic variance decompositions, *Technical report*, Princeton University.

Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections and VARs estimate the same impulse responses, *Econometrica* **89**: 955–980.

Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* pp. 81–135.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments, *Economic Journal* **128**: 917–948.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes, *Journal of Econometrics* **66**: 225–250.

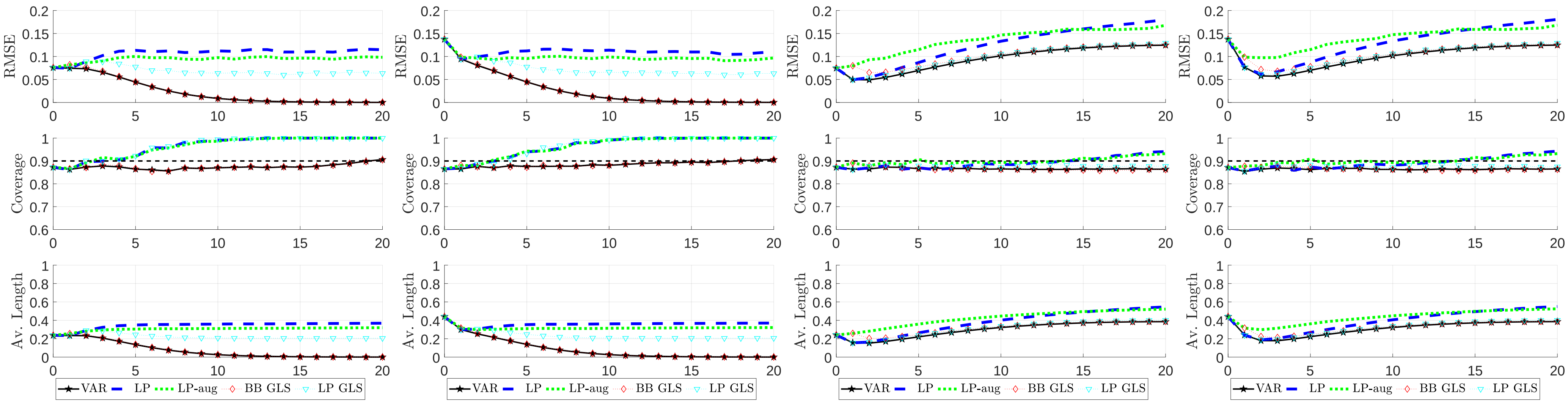


(a) $T = 100$, $a_{11} = 0.5$, Corr $= 0.9$, $p = 1$

(b) $T = 100$, $a_{11} = 0.5$, Corr $= 0.5$, $p = 1$

(c) $T = 100$, $a_{11} = 0.95$, Corr $= 0.9$, $p = 1$

(d) $T = 100$, $a_{11} = 0.95$, Corr $= 0.5$, $p = 1$



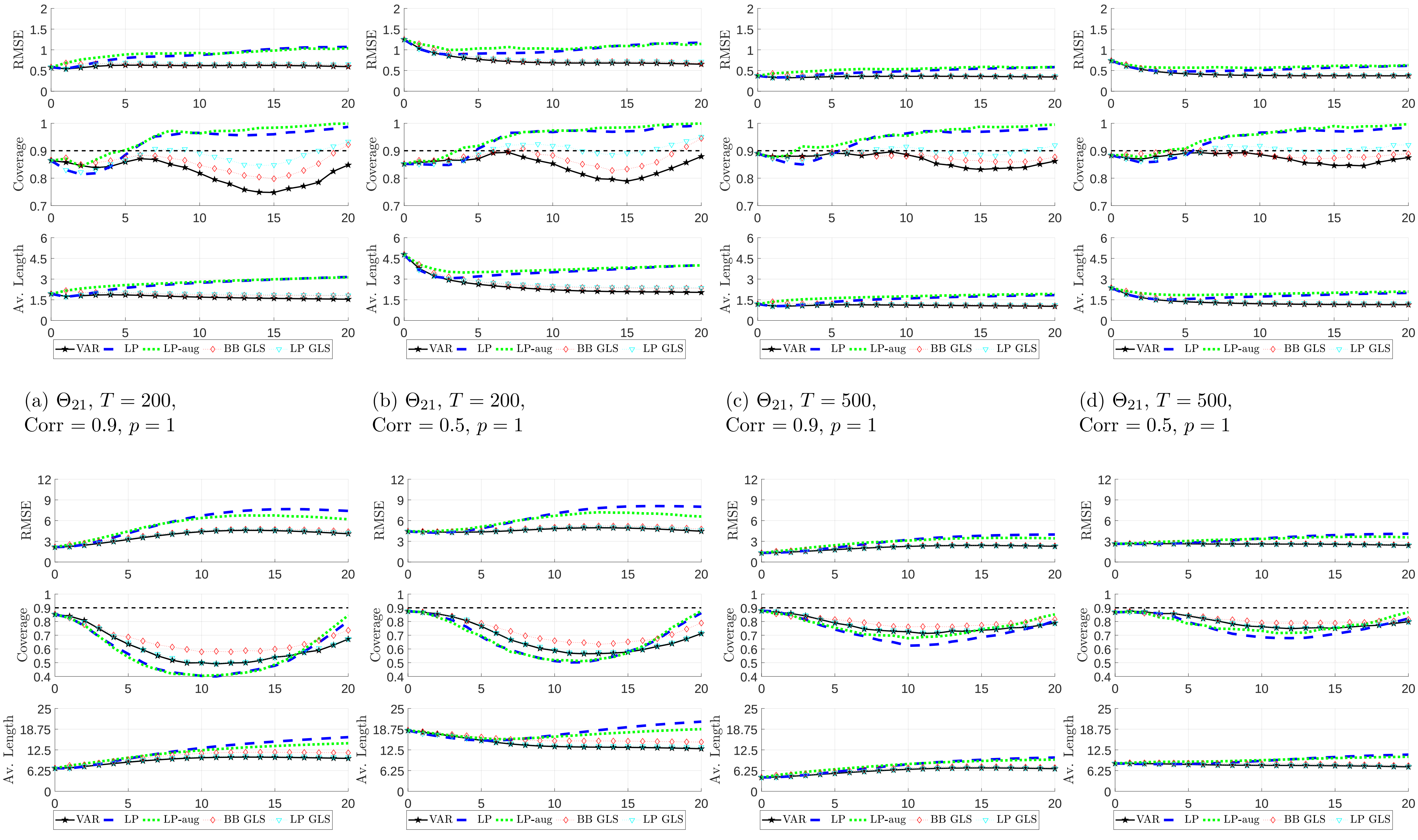
(e) $T = 500$, $a_{11} = 0.5$, Corr $= 0.9$, $p = 1$

(f) $T = 500$, $a_{11} = 0.5$, Corr $= 0.5$, $p = 1$

(g) $T = 500$, $a_{11} = 0.95$, Corr $= 0.9$, $p = 1$

(h) $T = 500$, $a_{11} = 0.95$, Corr $= 0.5$, $p = 1$

Figure 1: **DGP1** RMSEs as well as coverage and average lengths of pointwise MBB 90% confidence intervals based on alternative estimators for the responses of variable 2 to the first structural shock (with bias-correction). Sample sizes $T = 100$ (top row) and $T = 500$ (bottom row). Medium persistence ($a_{11} = 0.5$) and high persistence ($a_{11} = 0.95$). Strong proxy (Corr $= 0.9$) and weak proxy (Corr $= 0.5$). Lag length used for estimation $p = 1$.



(a) Θ_{21}, $T = 200$, Corr $= 0.9$, $p = 1$

(b) Θ_{21}, $T = 200$, Corr $= 0.5$, $p = 1$

(c) Θ_{21}, $T = 500$, Corr $= 0.9$, $p = 1$

(d) Θ_{21}, $T = 500$, Corr $= 0.5$, $p = 1$

(e) Θ_{41}, $T = 200$, Corr $= 0.9$, $p = 1$

(f) Θ_{41}, $T = 200$, Corr $= 0.5$, $p = 1$

(g) Θ_{41}, $T = 500$, Corr $= 0.9$, $p = 1$

(h) Θ_{41}, $T = 500$, Corr $= 0.5$, $p = 1$

Figure 2: **DGP2** RMSEs as well as coverage and average lengths of pointwise MBB 90% confidence intervals based on alternative estimators for the responses of variable 2 (top row) and variable 4 (bottom row) to the first structural shock (with bias-correction). Sample sizes $T = 200$ and $T = 500$. Strong proxy (Corr = 0.9) and weak proxy (Corr = 0.5). Lag length used for estimation $p = 1$.