Projection estimators for structural impulse responses*

Jörg Breitung[a]
University of Cologne

Ralf Brüggemann[b]
University of Konstanz

preliminary version: November 26, 2019

comments welcome

Abstract

In this paper we provide a general framework for linear projection estimators for impulse responses in structural vector autoregressions (SVAR). An important advantage of our projection estimator is that for a large class of SVAR systems (that includes the recursive (Cholesky) identification scheme) standard OLS inference is valid without adjustment for generated regressors, autocorrelated errors or nonstationary variables. We also provide a framework for SVAR models that can be estimated by instrumental (proxy) variables. We show that this class of models (that includes also identification by long-run restrictions) result in a set of quadratic moment conditions that can be used to obtain the asymptotic distribution of this estimator, whereas standard inference based on instrumental variable (IV) projections is invalid. Furthermore, we propose a generalized least squares (GLS) version of the projections that performs similarly to the conventional (iterated) method of estimating impulse responses by inverting the estimated SVAR representation into the MA(∞) representation. Monte Carlo experiments indicate that the proposed OLS projections perform similarly to Jordà's (2005) projection estimator but enables us to apply standard inference on the estimated impulse responses. The GLS versions of the projections provide estimates with much smaller standard errors and confidence intervals whenever the horizon h of the impulse responses gets large.

JEL classification: C32, C51

Keywords: structural vector autoregressive models, impulse responses, local projections

*Financial support of the German Science Foundation (DFG grant number: BR 2941/3-1) is gratefully acknowledged.

[a]Corresponding author: Jörg Breitung: University of Cologne, Institute of Econometrics, Albertus Magnus Platz, 50923 Köln, Germany, email: breitung@statistik.uni-koeln.de.

[b]Ralf Brüggemann: University of Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: ralf.brueggemann@uni-konstanz.de

1 Introduction

The analysis of dynamic effects in vector autoregressive models (VAR) by means of impulse responses has become a standard tool in empirical macroeconomics. Following Sims (1980) the dynamic effect to shocks are typically measured by the moving average (MA) coefficients derived from the finite order VAR representation of the time series. Lütkepohl (1989) provide asymptotic methods for statistical inference. In recent years it has become popular to estimate the effects of structural shocks by "local projections" (e.g. Jordà (2005), Jordà, Schularick and Taylor (2015, 2019), Ramey and Zubairy (2018)). This method is based on a *direct* representation of the time series vector shifted h periods ahead, whereas the traditional method traces out the impulse responses *"iteratively"* from the first up to the h-th period. In contrast to the iterated method, where the MA parameters are nonlinear functions of the underlying VAR(p) model, the projection method results in a representation that is typically linear in the parameters. Since the errors of the direct representation are autocorrelated up to the horizon $h-1$, the projection estimator is inefficient, in general. On the other hand, the iterated method provides maximum likelihood (ML) estimators that are asymptotically efficient, whenever the likelihood function is correctly specified. As pointed out by Plagborg-Moller and Wolf (2019) "in population, they estimate the same thing, as long as we control flexibly for lagged data". An important difference is that the iterated estimator extrapolates longer-term impulse responses from the first p autocorrelations such that the impulse responses converge to zero as the horizon h gets large, whenever all roots are outside the unit circle. On the other hand, the estimated impulse responses of the projection method are functions of the autocorrelations from lag h up to lag $h+p$ which prevents the impulse responses to converge to zero as h tends to infinity. Hence, the estimates from the two methods may be quite different even for moderate horizons h.

In practice the identification of shocks and the estimation of the impulse responses are often performed in two separate steps. For instance, Romer and Romer (2004, 2010) first construct "narrative shocks" from a careful analysis of a rich information set, whereas in a second step the dynamic response of some variables of interest to the shocks is measured by an autoregressive distributed lag (ADL) regression. More formally the analysis can be characterized by two different steps:

$$\varepsilon_{j,t} = f(x_{j,t}, \beta_j) \tag{1}$$

$$y_{i,t+h} = \theta_{ij}^h \varepsilon_{j,t} + z_{j,t}' \pi_{ij}^h + e_{ij,t}^h \tag{2}$$

where $\varepsilon_{j,t}$ denotes the j-th structural shock ($j \in \{1, \ldots, k\}$), $x_{j,t}$ is a vector of time series used to identify the shocks, $f(\cdot)$ is a (possibly nonlinear) function, β_j is a parameter vector, $y_{i,t}$ denotes the target variable and $z_{j,t}$ is a vector of additional control variables, included to improve the efficiency of the estimator for θ_{ij}^h. The first equation (1) is used to identify the shocks based on a set of economic time series such that (i) the shock is unpredictable with respect to some information set $\mathcal{I}_t$ that includes the past of $x_{j,t}$, y_t $z_{j,t}$ and (ii) the shocks are "orthogonal" to all other shocks. Equation (2) is called the *projection step*, where the parameter θ_{ij}^h measures the effect of the j-th shock on the i-th variable h-steps ahead. In many applications it is convenient to estimate the parameter of interest θ_{ij}^h in two steps. In the first step the shocks result from by estimating β_j in (1). In the second step we replace the unobserved shock $\varepsilon_{j,t}$ by the estimated analog $\widehat{\varepsilon}_{j,t} = f(x_{j,t}, \widehat{\beta}_j)$. Alternatively it is possible to estimate all parameters jointly by estimating the equation

$$y_{i,t+h} = \theta_{ij}^h f(x_{j,t}, \beta_j) + z_{j,t}' \pi_{ij}^h + e_{ij,t}^h \ . \tag{3}$$

As a simple example, consider the bivariate VAR(1) model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

where the shocks are identified by the structural model $\varepsilon_t = \Gamma u_t$ and Γ is a lower triangular matrix with ones on the leading diagonal. Accordingly the shocks can be written as functions of $x_{1,t} = (y_{1,t}, y_{t-1}')'$ and $x_{2,t} = (y_t', y_{t-1}')'$:

$$\begin{aligned} \varepsilon_{1t} = u_{1,t} &= y_{1,t} - a_{11} y_{1,t-1} - a_{12} y_{2,t-1} \\ \varepsilon_{2t} = \gamma u_{1,t} + u_{2,t} &= y_{2,t} + \gamma y_{1,t} - (a_{21} + \gamma a_{11}) y_{1,t-1} - (a_{22} + \gamma a_{21}) y_{2,t-1} \ , \end{aligned}$$

where γ is the $(2,1)$-element of Γ. The first shock can be estimated by running a regression of $y_{1,t}$ on $y_{1,t-1}$ and $y_{2,t-1}$, whereas the second shock results from a regression of $y_{2,t}$ on $y_{1,t}$, $y_{1,t-1}$, $y_{2,t-1}$. In a second step, the estimated shock is inserted in the projection equation yielding

$$y_{i,t+h} = \theta_{ij}^h \widehat{\varepsilon}_{j,t} + z_{j,t}' \pi_{ij}^h + \tilde{e}_{ij,t}^h \ .$$

In our example it is not necessary to include any control variable as the candidate variables $y_{1,t-1}$ and $y_{2,t-1}$ are orthogonal to the estimated shocks. We may nevertheless include the vectors of control variables $z_{1,t} = y_{t-1}$ and $z_{2,t} = (y_{1,t}, y_{t-1}')'$ as otherwise the usual regression based inference on $\widehat{\theta}_{ij}^h$ is invalid (see the next section for more details). Alternatively the impulse response may be estimated in a single step using (3) which results in estimating the linear equation

$$y_{i,t+h} = \theta_{ij}^h y_{j,t} + \tilde{z}_{j,t}' \tilde{\pi}_{ij}^h + e_{ij,t}^h \tag{4}$$

with $\tilde{z}_{1,t} = y_{t-1}$ and $\tilde{z}_{2,t} = (y_{1,t}, y_{t-1}')'$.

These projection estimators are similar but not identical to the estimator suggested by Jordà (2005) which is based on the transformed "h-step ahead" representation of the VAR. For our simple VAR(1) example this approach results in

$$\begin{aligned} y_{t+h} &= A^h y_t + v_t^h, & (5) \\ &= \Theta_h y_t^* + v_t^h, & (6) \end{aligned}$$

where A is the $k \times k$ matrix of VAR coefficients, $y_t^* = \Gamma y_t$ and $\Theta_h = A^h \Gamma^{-1}$ is the impulse response matrix with respect to the vector of structural shocks $\varepsilon_t = \Gamma u_t$. In practice the matrix Γ is unknown and must be replaced by a consistent estimator. Hence, Jordà's (2005) estimator of the impulse response of the i-th variable to the j-th shock is obtained from the regression

$$y_{i,t+h} = \sum_{j=1}^{k} \theta_{ij}^h \widehat{y}_{j,t}^* + \widetilde{v}_{i,t}^h \ , \tag{7}$$

where $\widehat{y}_{j,t}^*$ is the j-th element of the vector $\widehat{y}_t^* = \widehat{\Gamma} y_t$ and $\widehat{\Gamma}$ is a consistent estimate of Γ. It should be noted that using an estimated matrix for the transformation of y_t involves a generated-regressor problem that introduces nuisance parameters in limiting distributions of the estimators. Accordingly, the usual heteroskedasticity and autocorrelation consistent (HAC) inference on the OLS estimator $\widehat{\theta}_{ij}^h$ is invalid and its limiting distribution needs to be adjusted by the estimation error $\widehat{\Gamma} - \Gamma$ (cf. Kilian and Kim (2011)). In contrast, no HAC or other adjustments are required for valid inference when using regression (4) with suitable control variables (see Section 2).

The remainder of this paper is structured as follows: Section 2 discusses OLS projection methods and their asymptotic properties and Section 3 considers IV projections. Section 4 suggests refinements leading to GLS projections. Section 5 presents some Monte Carlo evidence on the the relative performance of different projection estimators and Section 6 concludes.

2 OLS projections

Let us consider the VAR(p) model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \ , \tag{8}$$

where y_t is a k-dimensional vector of time series and u_t is a $k \times 1$ vector of white noise innovations with $\mathbb{E}(u_t) = 0$ and $\mathbb{E}(u_t u_t') = \Sigma$ (positive definite). The inclusion of further deterministic regressors like constants, trends or dummy variables is unproblematic and is therefore suppressed. The vector of structural shocks is related to the vector of innovations (forecast errors) by

$$\varepsilon_t = \Gamma u_t. \tag{9}$$

If the diagonal elements of Γ (or Γ^{-1}) are unity, $(k-1)k/2$ additional restrictions are necessary to identify the matrix Γ. A leading example is the assumption that Γ is a lower triangular matrix such that it can be determined from a Cholesky decomposition of the covariance matrix $\Sigma = \mathbb{E}(u_t u_t')$.

The impulse response of the i-th variable y_{it} with respect to the j-th shock h steps ahead is given by

$$\frac{\partial \mathbb{E}(y_{i,t+h}|\varepsilon_{j,t})}{\partial \varepsilon_{j,t}} = \theta_{ij}^h + \frac{\partial \mathbb{E}(z_{j,t}'|\varepsilon_{j,t})}{\partial \varepsilon_{j,t}} \pi_{ij}^h \tag{10}$$

and, thus, θ_{ij}^h is an unbiased measure of the impulse response whenever $z_{j,t}$ and $\varepsilon_{j,t}$ are independent. Since the shocks are uncorrelated with $y_{t-1}, \dots, y_{t-p}$, the lags of all variables are natural candidates for the vector of control variables. In a recursive (triangular) scheme we may also include the "previous" contemporaneous variable $y_{1,t}, \dots, y_{j-1,t}$ as these variables are uncorrelated with $\varepsilon_{j,t}$ as well.

If the shock $\varepsilon_{j,t}$ was observed, we may just run an OLS regression of the projection equation (2) to estimate the coefficient θ_{ij}^h consistently. In the more realistic case that the shock is estimated such that $\widehat{\varepsilon}_{j,t} = f(x_{j,t}, \widehat{\beta}_j)$, then the OLS estimator of θ_{ij}^h remains consistent provided that $\widehat{\beta}_j \xrightarrow{p} \beta_j$. In general, however, statistical inference is affected by the estimation error $\widehat{\beta}_j - \beta_j$. For important special cases the following proposition shows that standard regression inference is valid no matter of the estimation error in $\widehat{\varepsilon}_{j,t}$ and possible autocorrelation of $e_{ij,t}^h$.

Proposition 1. *Let y_t be generated by the VAR(p) model* (8) *and the estimated structural shock is obtained as $\widehat{\varepsilon}_{j,t} = f(x_{j,t}, \widehat{\beta}_j)$, where $\widehat{\beta}_j$ is a $\sqrt{T}$-consistent estimator for β_j. Assume that there exists some matrix C_j such that $g_{j,t}(\beta_j) = C_j z_{j,t}$ with*

$$g_{j,t}(\beta_j) = \frac{\partial f(x_{j,t}, \beta_j)}{\partial \beta_j}$$

and $\mathbb{E}(z_{j,t}\varepsilon_{j,t}) = 0$. Then (i) the OLS estimator $\widehat{\theta}_{ij}^h$ of θ_{ij}^h in the regression

$$y_{i,t+h} = \theta_{ij}^h \widehat{\varepsilon}_{j,t} + z_{j,t}' \pi_{ij}^h + \tilde{e}_{ij,t}^h \tag{11}$$

is a consistent and asymptotically normally distributed estimator for θ_{ij}^h. (ii) The ordinary OLS estimator of the standard error of $\widehat{\theta}_{ij}^h$ is consistent no matter of the serial correlation of $\tilde{e}_{ij,t}^h$.

A proof of the results is given in the appendix. This proposition shows that augmenting the regression by suitable control variables escapes the error-in-variables problem involved by working with estimated shocks. For illustration, let us first consider a VAR, where the popular triangular (recursive) identification scheme is applied. Accordingly, the j-th structural results as

$$\begin{aligned}\varepsilon_{j,t} &= \gamma_j'(y_t - A_1 y_{t-1} - \cdots - A_p y_{t-p}) \\ &= \beta_0' y_t + \beta_1' y_{t-1} + \cdots + \beta_p' y_{t-p}\end{aligned}$$

where γ_j' is the j-th row of Γ and β_ℓ' is the j-th row of $-\Gamma A_\ell$ for $\ell = 1, \dots, p$ and $\beta_0 = \gamma_j$. Denote by $\beta_{0,r}$ the r-th element of β_0. For a triangular identification scheme we set $\beta_{0,r} = 0$ for $r > j$. Furthermore, the unit effect normalization implies that the diagonal elements of Γ are equal to unity which implies $\beta_{0,j} = 1$. Accordingly, the derivative with respect to the remaining parameters $\tilde{\beta} = (\beta_{0,1}, \dots, \beta_{0,j-1}, \beta_1', \dots, \beta_p')'$ is given by $g_{j,t}(\beta) = \partial f(x_{j,t}, \beta_j)/\partial \tilde{\beta}_j = z_{j,t}$, where $z_{j,t} = (y_{1,t}, \dots, y_{j-1,t}, y_{t-1}', \dots, y_{t-p}')'$. Thus, the matrix C is the identity matrix and Proposition

1 applies. It should also be noted that for the triangular identification the regression (3) is equivalent to

$$y_{i,t+h} = \theta_{ij}^h \widehat{\varepsilon}_{j,t} + z_{j,t}' \pi_{ij}^h + e_{ij,t}^h \tag{12}$$

$$= \theta_{ij}^h y_{j,t} + z_{j,t}' \tilde{\pi}_{ij}^h + e_{ij,t}^h \tag{13}$$

as $\widehat{\varepsilon}_{j,t}$ results from a regression of $y_{j,t}$ on $z_{j,t}$. Accordingly, the sequential two-step estimator of (1) – (2) is equivalent to the joint estimation based on (3).

Let us compare our projection estimator for the triangular identification scheme with Jorda's (2005) projection estimator. First, since $\varepsilon_{j,t}$ is serially uncorrelated, the product $\varepsilon_{j,t} e_{ij,t}^h$ is serially uncorrelated as well and no HAC standard errors are required. Second, the regression (13) does not involve any errors-in-variable problem that has to be taken into account when computing standard errors. In contrast, for SVAR models the standard errors of Jordà's (2005) projection estimator needs to be adjusted by a (nonlinear) term resulting from the estimation error $(\widehat{\Gamma} - \Gamma)$ (cf. Kilian and Kim (2011)). Another advantage of our projection estimator is that inference is valid no matter whether the time series are stationary or not. This is due to the fact that the coefficient of interest θ_{ij}^h is attached to the stationary variable $\varepsilon_{j,t}$ in (11), whereas in Jordà's (2005) version of the projection estimator, the coefficient of interest may be attached to a nonstationary variable.

3 IV projections

3.1 Proxy VARs

There are different motivations for employing instrumental variables (IV) when estimating impulse response functions. One possibility is to use external shocks as instruments for identifying the internal shocks within a VAR system. Assume for example that there exists some external instrument (or proxy) w_t that is correlated with the shock $\varepsilon_{1,t}$ but exhibit some measurement error uncorrelated with all other shocks in the system. Accordingly, the relationship between the external instrument and the structural shock $\varepsilon_{1,t}$ is represented by

$$w_{1,t} = \psi \varepsilon_{1,t} + \eta_t \ , \tag{14}$$

where η_t is a measurement error with $\mathbb{E}(\eta_t \varepsilon_t) = 0$. We can use w_t as an instrument for estimating the parameters in the identification step. As a simple example consider the bivariate VAR system (in residuals) with structural equations:

$$u_{1,t} = \alpha_1 u_{2,t} + \varepsilon_{1,t} \tag{15}$$

$$u_{2,t} = \alpha_2 u_{1,t} + \varepsilon_{2,t} \ . \tag{16}$$

We are able to estimate the parameter α_2 consistently by using w_t as an instrument for $u_{1,t}$ in equation (16). The residual of this equation serves as an estimate of the second shock that can in turn be used as an instrument for $u_{2,t}$ in (15).

For estimating the impulse response with respect to the first shock, θ_{i1}^h, the usual projection equation is given by (e.g. Stock and Watson (2018))

$$y_{i,t+h} = \theta_{i1}^h y_{1,t} + \text{ lags } + e_{i1,t}^h \tag{17}$$

$$\text{or} \quad y_{i,t+h} = \theta_{i1}^h u_{1,t} + \text{ lags } + e_{i1,t}^h \tag{18}$$

where "lags" represent some linear combination of the vector $(y_{t-1}', \ldots, y_{t-p}')$. If the unit effect condition is neglected, we can estimate θ_{i1}^h by an IV estimator using the proxy variable w_t as instrument for the variable $y_{1,t}$ (resp. $u_{1,t}$). By solving the system (15) and (16) for the first structural shock we obtain

$$u_{1,t} = a_1\varepsilon_{1,t} + a_2\varepsilon_{2,t} \ ,$$

where $a_1 = 1/(1-\alpha_1\alpha_2)$ and $a_2 = \alpha_1/(1-\alpha_1\alpha_2)$. Accordingly, replacing $\varepsilon_{1,t}$ by $y_{1,t}$ or $u_{1,t}$ in the projection equations (17) and (18) implies that the (unit effect) impulse response is multiplied by the factor $1/a_1$. This is of no concern for performing standard significants tests on θ_{21}^h, but notice that the factor $1/a_1$ changes the sign of the response whenever $a_1 a_2 > 1$.

In order to impose the unit-effect normalization we replace (15) by

$$u_{1,t} = \varepsilon_{1,t}^* + a_2\varepsilon_{2,t} \ , \tag{19}$$

where $\varepsilon_{1,t}^* = a_1\varepsilon_{1,t}$. If $\varepsilon_{2,t}$ were observed, we would estimate the shock $\varepsilon_{1,t}^*$ as the residual from an OLS regression of $u_{1,t}$ on $\varepsilon_{2,t}$. Since $\varepsilon_{2,t}$ is not observed we may replace it by the residual of the second equation (16), estimated by using the proxy variable w_t as instrument for $u_{2,t}$. In a second step we estimate a_2 from inserting the residual in (19) and running the OLS regression

$$u_{1,t} = a_2\widehat{\varepsilon}_{2,t} + \widetilde{\varepsilon}_{1,t}^* \ , \tag{20}$$

The corresponding set of moment equations is given by

$$\mathbb{E}[w_t(u_{2,t} - \alpha_2 u_{1,t})] \ = \ 0 \tag{21}$$

$$\mathbb{E}(u_{2,t} - \alpha_2 u_{1,t})[u_{1,t} - a_2(u_{2,t} - \alpha_2 u_{1,t})] \ = \ 0 \tag{22}$$

$$\mathbb{E}\begin{pmatrix} [u_{1,t} - a_2(u_{2,t} - \alpha_2 u_{1,t})] \\ \text{lags} \end{pmatrix} \left(y_{i,t+h} - \theta_{i1}^h[u_{1,t} - a_2(u_{2,t} - \alpha_2 u_{1,t})] - \text{ lags }\right) \ = \ 0. \tag{23}$$

Again these equation can be solved sequentially. We first estimate the second shock from the IV regression (16). Second, we replace the unknown shock $\varepsilon_{2,t}$ by the residuals of (16) and run an OLS regression of $u_{1,t}$ on the residuals $\widehat{\varepsilon}_{2,t}$. Finally the parameter θ_{i1}^h is estimated by OLS after replacing the shock $\varepsilon_{1,t}^*$ by the residual (20) in the projection equation (2). This example shows that in many cases the two-step estimator is identical to a Method of Moments estimator of the system (21) – (23). The merit of characterizing the estimator by the corresponding moment equations is that from these moment conditions the limiting distribution can easily be derived (see below).

3.2 AB model

(Internal) instrumental variables are also useful for estimating a so-called "AB-model" (e.g. Amisano and Giannini (1997) and Breitung, Brüggemann and Lütkepohl (2004)). Let us assume again a simple example with the following structure:

$$u_{1,t} = a_1 u_{2,t} + \varepsilon_{1,1} \tag{24}$$

$$u_{2,t} = a_2 u_{1,t} + \varepsilon_{2,t} \tag{25}$$

$$u_{3,t} = b_1 \varepsilon_{1,t} + \varepsilon_{3,t} \tag{26}$$

The structural equations can be represented by the equation

$$Au_t = B\varepsilon_t$$

where

$$A = \begin{pmatrix} 1 & -a_1 & 0 \\ -a_2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and } B \;= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b_1 & 0 & 1 \end{pmatrix}$$

and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Since $u_{3,t}$ is uncorrelated with $\varepsilon_{2,t}$ we can use $u_{3,t}$ as an instrument[1] in equation (25). The residual of this equation serves as an instrument for $u_{2,1}$ in (24). The residual of the latter equation is used as regressor in (26) to obtain a consistent OLS estimator for b_1. Accordingly we can estimate all parameters in the AB-model consistently using instrumental variable methods. If the model is just identified, the IV method is asymptotically efficient (cf. Theil (1971)). It is interesting to note that this model is equivalent to the first example by letting $u_{3,t} = w_t$. Hence the distinction between internal and external instruments is not very stringent (see also Plagborg-Moller and Wolf (2019)).

The 2-step projection estimator results from estimating the projection equation

$$y_{i,t+h} = \theta_{ij}^h \widehat{\varepsilon}_{j,t} + \text{ lags } + \tilde{e}_{ij,t}^h$$

where the residual $\widehat{\varepsilon}_{j,t}$ is obtained by IV estimation of (24) – (26). If the unit-effect normalization is imposed, the corresponding moment conditions are equivalent to the moment conditions presented in section 3.1.

3.3 Long-run restrictions

The final example for the benefit of IV methods is the estimation of structural models identified by long-run restrictions as proposed by Blanchard and Quah (1989). Let $y_t = (y_{1t}, y_{2t})'$ denote

[1] In practice the vector of VAR innovations u_t is not observed but can be replaced by residuals without affecting the asymptotic properties. This is due to the fact that the estimation error in the VAR residuals does not affect the asymptotic distribution of the estimated structural parameters.

a bivariate vector of stationary time series with VAR(p) representation as given in (8). The identifying restriction is that the cumulated impulse response of the first variable with respect to the second shock tends to zero, that is,

$$\sum_{h=0}^{\infty} \theta_{12}^{h} \;=\; 0.$$

This implies that the long-run impact matrix $\bar{\Theta} = \sum_{h=0}^{\infty} \Theta_h$ is a lower triangular matrix which can be obtained from a Cholesky decomposition of the matrix

$$\bar{\Theta}\,\bar{\Theta}' = A(1)^{-1}\Sigma A(1)'^{-1}, \tag{27}$$

where $A(1) = I_k - \sum_{i=1}^{p} A_i$. The structural shocks result from $\varepsilon_t = \Gamma u_t$. We follow Shapiro and Watson (1988) and Fry and Pagan (2005) and employ a simple instrumental variable (IV) estimator for the columns of Γ that is equivalent to the estimator suggested by Blanchard and Quah (1989).

First, we re-write the model in the error correction format

$$\Delta y_t \;=\; -A(1)y_{t-1} + \;\Delta\text{lags}\; + u_t\;.$$

where "Δlags" represent a linear combination of the lagged differences $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. Next we multiply the system by the matrix Γ yielding

$$\Gamma\Delta y_t \;=\; -\Gamma A(1)y_{t-1} + \;\Delta\text{lags}\; + \varepsilon_t\;.$$

The identifying assumption is that the matrix $\bar{\Theta} = A(1)^{-1}\Gamma^{-1}$ is lower triangular and so is its inverse $\bar{\Theta}^{-1} = \Gamma A(1)$. By normalizing the diagonal elements of the matrix Γ to unity, the two equations of the system result as

$$\Delta y_{1,t} = -\gamma_{12}\Delta y_{2,t} - \overline{\theta}^{11} y_{1,t-1} + \;\Delta\text{lags}\; + \varepsilon_{1,t} \tag{28}$$

$$\Delta y_{2,t} = -\gamma_{21}\Delta y_{1,t} - \overline{\theta}^{21} y_{1,t-1} - \overline{\theta}^{22} y_{2,t-1} + \;\Delta\text{lags}\; + \varepsilon_{2,t} \tag{29}$$

where $\overline{\theta}^{ij}$ denotes the (i,j) element of the matrix $\bar{\Theta}^{-1}$. Note in the system (28) – (29) the error $\varepsilon_{1,t}$ is the permanent shock and $\varepsilon_{2,t}$ is the transitory shock. Obviously the size of the shocks are normalized in terms of a unit effect on y_{1t} and y_{2t}, respectively. By imposing the restriction that both shocks are orthogonal, the system is just identified and can be consistently estimated by the IV method. The first equation (28) can be estimated by using $y_{2,t-1}$ as an instrument for Δy_{2t}. For the second equation (29) the (estimated analog of the) residual $\varepsilon_{1,t}$ can be used as an instrument for $\Delta y_{1,t}$. This gives rise to the following moment equations for the impulse response with respect to the first shock:

$$\mathbb{E}\begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \Delta\text{lags} \end{pmatrix} \varepsilon_{1,t} \;= 0 \quad \text{and} \quad \mathbb{E}\begin{pmatrix} \varepsilon_{1,t} \\ \text{lags} \end{pmatrix} \left(y_{i,t+h} - \theta_{i1}^{h}\varepsilon_{1,t} - \;\text{lags}\;\right) \;=\; 0$$

where "lags" represent the lagged levels $y_{t-1}, \dots, y_{t-p}$ and for compactness we do not make explicit the dependence of $\varepsilon_{1,t}$ on the structural parameters (denoted as β_1 in (1)). For the impulse responses of the second shock the set of moment conditions is:

$$\mathbb{E}\begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \Delta\text{lags} \end{pmatrix} \varepsilon_{1,t} = 0 \quad , \quad \mathbb{E}\begin{pmatrix} \varepsilon_{1,t} \\ y_{1,t-1} \\ \Delta\text{lags} \end{pmatrix} \varepsilon_{2,t} = 0$$

$$\text{and } \mathbb{E}\begin{pmatrix} \varepsilon_{2,t} \\ \text{lags} \end{pmatrix} \left(y_{i,t+h} - \theta_{i2}^h \varepsilon_{2,t} - \text{ lags } \right) = 0.$$

Since the system is just identified, the estimators are equivalent to the ML estimator as well as the estimator based on the Cholesky decomposition (27).

3.4 Asymptotic properties of IV projections

In all applications discussed above, the projection estimator can be characterized by solving a sets of moment equations derived from the structural equation (1) and the projection step (2). Let $\theta = (\beta_j', \theta_{ij}^h, \pi_{ij}^{h\prime})'$ denote the vector of all parameters, where the notation suppresses the dependence on i, j, and h. The vector of moments can typically represented as

$$\sum_t m_t(\theta) = \begin{pmatrix} \sum\limits_t m_{1,t}(\beta_j) \\ \sum\limits_t m_{2,t}(\beta_j, \theta_{ij}^h, \pi_{ij}^{h\prime}) \end{pmatrix},$$

where the first set of moments depends linearly on the parameters of the identification step β_j. The projection step typically involves a set of nonlinear moment conditions, where the shock (that involves β_j) enters in terms of instruments. An exception is the linear OLS projection for the Cholesky VAR considered in Section 2. Note that in spite of the nonlinear nature of the second moment condition, the projection estimator can be estimated by OLS, where the unknown shock $\varepsilon_{j,t}$ is replaced by the residual $\widehat{\varepsilon}_{j,t} = f(x_{j,t}, \widehat{\beta}_j)$. Of course this involves a "generated regressor" problem but this can easily be coped with by considering the full system of moment conditions.

As all examples considered above result in a set of just identified moment conditions we focus on the simple method of moment (MM) estimators.[2] The asymptotic distribution of the MM estimator is given by

$$\sqrt{T}(\widehat{\theta}_{MM} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma_\theta) \tag{30}$$

where the asymptotic covariance matrix is given by

$$\Sigma_\theta = D(\theta)^{-1} V^m(\theta) D'^{-1}(\theta) \tag{31}$$

[2]The MM estimator is simpler than the GMM estimator as it does not require a weight matrix.

with

$$
\begin{aligned}
D(\theta) = \underset{T\to\infty}{\text{plim}}\, D_T(\theta) &= \underset{T\to\infty}{\text{plim}}\, \frac{1}{T}\sum_{t=p+1}^{T} \frac{\partial m_t(\theta)}{\partial \theta'} \\
V^m(\theta) = \underset{T\to\infty}{\text{plim}}\, V_T^m(\theta) &= \underset{T\to\infty}{\text{plim}}\, \frac{1}{T}\sum_{t=p+1}^{T} m_t(\theta) m_t(\theta)'
\end{aligned}
$$

(e.g. Newey and McFadden (1994)). To simplify the notation, we ignore the dependence of m_t, D_T and V_T^m on i, j, and h. Note further that in the examples considered above the vector of moments $m_t(\theta)$ are serially uncorrelated. Whenever (some of) the moments are autocorrelated, the asymptotic covariance matrix $V^m(\theta)$ is replaced by the long-run covariance matrix. In practice $\widehat{\Sigma}_\theta = D_T(\widehat{\theta})^{-1} V_T^m(\widehat{\theta}) D_T'^{-1}(\widehat{\theta})$ can be used as a consistent estimator of Σ_θ.

4 GLS projections

It is well known that the iterative VAR estimator for the impulse responses can be much more efficient than the (direct) projection estimator, in particular for larger values of h (cf. Kilian and Kim (2011)). In this section we therefore consider GLS type projection estimators for impulse responses. To simplify the exposition we focus on the simple univariate AR(1) model $y_t = \alpha y_{t-1} + u_t$ which is used to estimate the impulse response $\theta_2 = \alpha^2$ with respect to the shock u_t, $h = 2$ steps ahead. The direct representation is given by

$$
y_{t+2} = \theta_2 y_t + u_{t+2} + \alpha u_{t+1} \quad t = 1, 2, \ldots, T-2. \tag{32}
$$

If u_{t+1} was observed, we may include it as an additional regressor. In this case the regression only contains an uncorrelated error term u_{t+2} and the parameter θ_2 can be estimated efficiently by OLS. Since u_{t+1} is not observed, it is natural to replace it by the OLS residual from a regression of y_{t+1} on y_t, yielding $\widehat{u}_{t+1} = y_{t+1} - \widehat{\alpha} y_t$, for $t = 1, \ldots, T-1$. It is important to notice that $\widehat{u}_{t+1}$ is by construction orthogonal to y_t and, therefore, including $\widehat{u}_{t+1}$ as an additional regressor results in the same estimator as running the regression without the additional regressor $\widehat{u}_{t+1}$. This reasoning is however not fully correct, as OLS estimation implies $\sum_{t=1}^{T-1} \widehat{u}_{t+1} y_t = 0$, whereas the regression (32) involves the observations $t = 1, \ldots, T-2$ only. It is not difficult to see, however, that the difference between the feasible GLS estimator and the OLS estimator without $\widehat{u}_{t+1}$ as an additional regressor is $O_p(T^{-1})$, which is due to the different sample sizes of the regressions. Since the estimation error is $O_p(T^{-1/2})$ it follows that both estimators are asymptotically equivalent and this GLS estimator does not improve asymptotic efficiency of the projection estimator.[3]

Another possibility to construct a GLS version of the projection estimator is to consider a regression equation where we replace the dependent variable y_{t+2} by

$$
y_{t+2}^* = y_{t+2} - u_{t+2} = \theta_2 y_t + v_{t+1} \tag{33}
$$

[3]For $h \geq 3$ the additional regressors $\widehat{u}_{t+2}, \ldots, \widehat{u}_{t+h-1}$ may result in an efficiency gain, see Lusompa (2019).

with $v_{t+1} = \alpha u_{t+1}$. Note that the error in the regression of y^*_{t+2} on y_t is white noise rendering the OLS estimator asymptotically efficient. Obviously the error u_{t+2} that enters the dependent variable y^*_{t+2} is not observed and has to be replaced by the OLS residual $\widehat{u}_{t+2} = y_{t+2} - \widehat{\alpha} y_{t+1}$ yielding

$$\widehat{y}^*_{t+2} = y_{t+2} - (y_{t+2} - \widehat{\alpha} y_{t+1}) \;=\; \widehat{\alpha} y_{t+1}.$$

The OLS regression of $\widehat{y}^*_{t+2}$ on y_t yields

$$\begin{aligned}\widehat{\theta}_2 &= \frac{\sum_{t=1}^{T-2} \widehat{\alpha} y_{t+1} y_t}{\sum_{t=1}^{T-2} y_t^2} \\ &= \widehat{\alpha} \left(\frac{\sum_{t=1}^{T-1} y_{t+1} y_t}{\sum_{t=1}^{T-1} y_t^2} + O_p(T^{-1}) \right) \\ &= \widehat{\alpha}^2 + O_p(T^{-1}).\end{aligned}$$

The resulting estimator differs from the iterative estimator $\widehat{\alpha}^2$ by an asymptotically negligible term that is due to the fact that the direct regression (33) is based on $T-2$ observations instead of $T-1$ observations for the iterated estimator $\widehat{\alpha}^2$. Accordingly, the iterative estimator can be seen as a GLS version of the direct estimator for the impulse response $\theta_2 = \alpha^2$. Since $\widehat{\alpha}^2$ is the ML estimator for θ_2 it turns out that this GLS estimator is asymptotically efficient if the innovations are normally distributed.

This approach can be adapted to estimate the impulse responses from the projection step (2). Denote the vector of innovations by $u_t = y_t - \mathbb{E}(y_t | y_{t-1}, \ldots, y_{t-p})$. The GLS-type projection estimator is obtained from the regression

$$(y_{i,t+h} - u_{i,t+h}) = \theta^h_{ij} \varepsilon_{j,t} + z'_{j,t} \pi^h_{ij} + \gamma'_1 u_{t+h-1} + \cdots + \gamma'_{p-2} u_{t+2} + v^h_{ij,t} \tag{34}$$

where in practice the unknown innovations are replaced by their sample analogs (VAR residuals) and the parameters $\gamma_1, \ldots, \gamma_{p-2}$ are additional coefficient vectors. The error term $v^h_{ij,t}$ is a linear combination of all shocks in period t that are not included as regressors and u_{t+1}. Accordingly, the error term has an MA(1) representation. To remove this autocorrelation, all other shocks at time t can be included in the vector of control variables, which makes the GLS approach more efficient but inference is more complicated as the set of moment conditions becomes larger.

Replacing the innovations $u_{t-\ell}$ by $\widehat{u}_{t-\ell}$ in (34) involves an additional estimation error that affects statistical inference. In order to derive the asymptotic distribution of the resulting estimator we characterize the estimator by the following (just-identified) set of moments:

$$\begin{aligned}\mathbb{E}\left(m_{1t}(\theta_1)\right) = \mathbb{E}\left(\frac{\partial f(x_{j,t}, \beta_j)}{\partial \beta_j} \varepsilon_{j,t} \right) &= 0 \\ \mathbb{E}\left(m_{2t}(\theta_2)\right) = \mathbb{E}\left(Y^+_{t-1} u'_t \right) &= 0 \\ \mathbb{E}\left(m_{3t}(\theta_3)\right) = \mathbb{E}\left(Z_t v_{ij,t+h} \right) &= 0,\end{aligned}$$

where $Y^+_{t-1} = (y'_{t-1}, \dots, y'_{t-p})'$ and $Z_t = (\varepsilon_{j,t}, z'_{j,t}, u_{i,t+h}, u'_{t+h-1}, \dots, u'_{t+2})'$. Compared to the set of moment conditions considered in Section 3.4 we add the moment conditions $m_{2t}(\theta_2)$ for estimating the innovations of the VAR system that enter $v_{ij,t+h}$.

Note also that the three sets of moment conditions are recursive in the sense that the parameters of the previous moments may enter the subsequent moments but not vice versa. Hence we can solve the moment conditions by first solving $\sum_t m_{1,t}(\widehat{\theta}_1) = 0$ and $\sum_t m_{2,t}(\widehat{\theta}_2) = 0$ and inserting the resulting expressions $\widehat{\varepsilon}_{j,t}$ and $\widehat{u}_t$ in $m_{3t}(\widehat{\theta}_3|\widehat{\theta}_1, \widehat{\theta}_2) = 0$. As in the previous section the asymptotic covariance matrix results from adapting (31) accordingly. It should also be noted that in general the moments m_{3t} are autocorrelated. This is due to the fact that due to the inclusion of $u_{t+1}, \dots, u_{t+h}$ in Z_t the moment condition is no longer a martingale difference sequence.

It should also be noticed that the set of moment conditions may contain redundant moments, that is, the covariance matrix of $m_t(\theta) = (m_{1t}(\theta_1)', m_{2t}(\theta_1)', m_{3t}(\theta_1)')'$ may be singular. For example, in the recursive (triangular) identification scheme the first shock is identical to the first innovation and, thus, the first element of $m_{1t}(\theta)$ is identical to the first element of $m_{2t}(\theta)$. In these cases the redundant moment conditions need to be eliminated or a generalized inverse needs to be applied.

5 Monte Carlo Evidence

In this section we explore the finite sample properties of alternative impulse response estimators discussed in the previous sections by Monte Carlo simulations. For this purpose, we generate data from different data generating processes (DGPs) and report the bias, the standard deviation, the empirical coverage and the average length of confidence intervals for the structural impulse responses θ^h_{ij}.

5.1 Simulations for OLS Projections

We start by presenting simulation results for OLS projections using a recursive (Cholesky) identification scheme. We compare the conventional iterated estimator based on inverting the VAR representation ("iterated"), the projection estimator ("Jorda") based on Jordà (2005), the "2step" method based on (13), and the "2step-GLS" method given in (34).

The first DGP is a simple bivariate VAR(1)

$$y_t = Ay_{t-1} + u_t, \tag{35}$$

where $u_t = \Gamma^{-1}\varepsilon_t$. The contemporaneous impact matrix is set to

$$\Gamma^{-1} = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix},$$

where the structural shocks ε_t are generated as $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, I_k)$. We control the persistence of the process by specifying the eigenvalues of the VAR. For this purpose, we use $A = V\Lambda^* V^{-1}$, where V is the matrix of eigenvectors obtained from a spectral decomposition of some random matrix $X = ZZ'$, and Z is a matrix of i.i.d. standard normal random numbers. Within this setup, specifying different values for the diagonal elements of Λ^*, λ_1^* and λ_2^*, can be used to control the persistence of the DGP. All simulations results below are based on $M = 1000$ Monte Carlo replications.

Table 1 and Figure 1 report typical simulation results for this DGP type using different sample sizes and persistence properties. Results are obtained from estimating a VAR with lag order $p = 1$ and for different response horizons h. Panel A in Table 1 shows results for the VAR with low persistence using $\lambda_1^* = 0.2$ and $\lambda_2^* = 0.4$. We report bias and standard errors of the response coefficient θ_{21}^h, which corresponds to the response of the second variable to the first structural shock and has been in the center of interest in many other simulation studies.

A number of interesting results emerge. First, the table documents the well-known finding that in small samples ($T = 100$) and at larger horizons h, Jordà's (2005) projection estimator tends to be less efficient than the iterated estimator. Second, in terms of bias and standard errors the two-step estimator and Jordà's projection estimators perform quite similarly. The main advantage of the two-step estimator is, however, that the standard errors and confidence intervals can be computed without any correction generated regressors and serial correlation. Third, it appears that the GLS refinement improves the properties of the projection estimators substantially. Even in small samples, the 2step-GLS estimator behaves similarly to the iterated estimator with bias and standard deviation converging to zero as h gets large.

Figure 1 shows the corresponding empirical coverage rates and average lengths of intervals for θ_{21}^h obtained from the different estimators. Panel A and C on the left of Figure 1 show the empirical coverage of nominal 95% confidence intervals. The coverage rates across all methods are quite comparable.[4] What is striking is the substantial difference in the average lengths (see Panels B and D). The OLS projection methods lead to much wider confidence intervals at larger horizons h. In fact, while the widths of intervals based on the iterated method or the GLS projection method tend to zero with increasing h, intervals based on Jordà's and the 2-step estimator remain wide at large horizons. This illustrates the inefficiency of the OLS projection methods and the effectiveness of the suggested GLS refinement.

Panel B of Table 1 shows the results for the more persistent VAR(1) DGPs, where $\lambda_1^* = 0.8$ and $\lambda_2^* = 0.4$ has been used. The change in persistence leads to population responses that converge to zero more slowly as h increases. In turn, this leads to somewhat larger biases and standard deviations for larger h. Apart from this, the relative ranking of methods is not changing compared to Panel A: Jordà's and the OLS 2-step estimator still lead to more

[4] One exception for larger horizons h is that the iterated and the GLS projection method lead to coverage very close to 100%. This has also been observed in other simulation studies and is driven by population responses that are very close to zero.

biased estimates and higher standard deviations for larger h, while the GLS estimator (2S-GLS) behaves similarly to the iterated estimator (even for $T = 100$). Panels A and C of Figure 2 show the coverage rates. Similar to earlier results from the literature for more persistent VAR processes, in small samples and for larger h, all methods yield intervals that have coverage somewhat below the desired 95% level. The differences in coverage between the methods is, however, again fairly small. Moreover, we still observe the same striking difference in the widths of intervals (see Panels B and D of Figure 2).

We have also conducted a number of additional Monte Carlo experiments based on different DGPs, which included the vector autoregressive DGPs of Kilian and Kim (2011), vector moving average (VMA) processes and higher-dimensional, (four variables) VARs. We have also repeated the experiments using information criteria for lag selection instead of using the true lag order. The main conclusions from these additional simulations is the same as for the results reported above. We have therefore decided not to include the results here but they are available on request.

We complement the results for stylized DGPs above by also using a DGP obtained from an empirical 6-variable monetary VAR system based on the data in Jordà (2005). To fix the parameters, we fit a VAR with intercepts to the vector

$$y_t = (\mathrm{EM}_t, \mathrm{P}_t, \mathrm{PCOM}_t, \mathrm{FF}_t, \mathrm{NBRX}_t, \Delta\mathrm{M2}_t)',$$

where the variables are the log of non-agricultural employment, the log of personal consumption expenditure deflator (1996=100), the annual growth rate of an index of sensitive materials prices, the federal funds rate and the ratio of non-borrowed reserves to total reserves and the annual growth rate of M2. We use $p = 2$ as suggested by the Hannan-Quinn and the Schwarz information criterion. The results in Table 2 and Figure 3 are based on estimated VARs with a lag length selected by the Akaike information criterion (AIC) using a maximum lag length of $p_{\max} = 6$. We use $T = 500$, which corresponds approximately to the number of monthly observations originally used in Jordà (2005).

We report results for selected response coefficients, which are typically of interest in economic analysis. In particular, we report results for the responses of employment (θ^h_{14}), of consumer prices (θ^h_{24}) and of the growth rate of the monetary aggregate M2 (θ^h_{64}) to a monetary policy shock. This shock ε_{4t} is identified as an orthogonalized innovation of the federal funds rate equation (the fourth equation in the VAR) and consequently, a Cholesky model can be used.

Panel A of Table 2 reports results for the response of employment (θ^h_{14}). Similar to the results in lower dimensional VARs, we find that for larger h Jordà's and the OLS 2-step estimator lead to larger bias and standard deviations compared to the iterated and the GLS projection estimator. Panels B and C show results for the response of prices and money growth. Here the biases of different methods are in the same order magnitude and in some cases the iterated and the GLS estimators have even slightly larger bias. The OLS projection methods still have somewhat larger standard deviations, however, the differences among the methods seem to be smaller than in the bivariate DGPs considered above. Figure 3 shows the coverage and lengths of the corresponding 95% confidence intervals (see Panels A, C and D). As for the other DGPs, we only find small differences in the empirical coverage rates but larger differences in interval lengths at larger horizons. Again, as seen from Panels B, D, and F, the OLS projection estimators yield somewhat wider intervals (especially at longer horizons). We also find that the other methods produce intervals that are still fairly wide at the considered horizons and consequently, the difference in intervals lengths are not as pronounced as in the bivariate DGPs. This can be explained by the properties of the DGP. The empirical DGP used here is very persistent with a number of roots of the characteristic polynomial fairly close to the unit circle. This implies population impulse responses that are different from zero even at fairly large horizons h and consequently longer intervals. The fact that population responses are different from zero even for large h also explains the smaller differences in the biases of the different methods. Finally, we note that despite the large dimension of the system (and consequently a larger number of future residuals to be included in the refinement), the refined methods and the iterative estimator perform very similarly.

5.2 Results for IV Projections

To investigate the properties of IV projection estimators, we focus on a model where we identify the first structural shock ε_{1t} by a single external instrument w_t. Thus, we are interested in the first column of the contemporaneous impact matrix Γ^{-1}. For the simulations, we set

$$\Gamma^{-1} = \begin{pmatrix} 1 & 0 \\ \gamma^{21} & \gamma^{22} \end{pmatrix}$$

and let $\gamma^{21} = 0.3$ and $\gamma^{22} = \sqrt{1 - \gamma^{21}}$. The latter choice ensures that both reduced from errors have unit variances. To mimic an external instrument setup, we also need to simulate data for the instrument w_t that is related to the structural shocks of interest. We do so by making use of the augmented system

$$\begin{pmatrix} u_t \\ w_t \end{pmatrix} = \begin{pmatrix} \Gamma^{-1} & 0_{n\times k} \\ \Phi & \Sigma_\eta^{1/2} \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}, \qquad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim iid\, \mathcal{N}(0, I_{n+k}). \tag{36}$$

The standard relation between reduced form errors and structural shocks $u_t = \Gamma^{-1}\varepsilon_t$ is augmented by equations that relate the structural shocks to the external instruments. This setup has a measurement interpretation, where the η_t are the measurement errors (see e.g. Braun and Brüggemann (2017)). If there is only one instrument w_t and one structural shock ε_{1t}, we find that

$$w_t = \phi_1 \varepsilon_{1t} + \sigma_\eta \eta_t,$$

where ϕ_1, σ_η and η_t are scalar-valued. In this setup, it is easy to see that the correlation between w_t and ε_{1t} is given by $\rho_{w,\varepsilon_1} = \phi_1/\sqrt{\phi_1^2 + \sigma_\eta^2}$. For a given correlation ρ_{w,ε_1}, we may specify the standard deviation of the measurement error as $\sigma_\eta = \sqrt{\phi_1^2(1/\rho_{w,\varepsilon_1}^2 - 1)}$. In our baseline simulation, we use $\phi_1 = 0.5$ and choose σ_η to obtain a correlation of $\rho_{w,\varepsilon_1} = 0.5$. Using equation (36), we generate data on w_t and u_t. Data for y_t are then obtained from a VAR(1) in equation (35) with

$$A = \begin{pmatrix} a_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix}.$$

This reduced form VAR(1) DGP has been used in several studies on impulse response properties (among others in Kilian and Kim (2011)) and the parameter a_{11} can be used to control the persistence. For the results in Table 3 and Figure 4, we have set $a_{11} = 0.5$.[5] As above, we focus on the properties of estimators of θ_{21}^h, the response of the second variable to the first structural shock in the system. Table 3 and Figure 4 report results for $T = 100$ and $T = 500$ and the following five alternative response estimators: IV-SVAR, IV-LP and IV-LP$^+$ are the estimators described in Stock and Watson (2018). In IV-SVAR, only the (first column of the) contemporaneous impact matrix is obtained by the IV method and is then combined with iterated reduced form impulse response estimates. IV-LP estimates the dynamic responses to structural shocks by directly by regressing $y_{j,t+h}$ on y_{1t} using w_t as an instrument for y_{1t} (see equation (7) in Stock and Watson (2018)). IV-LP$^+$ adds additional control variables to the regression used in IV-LP (see equation (10) in Stock and Watson (2018)). 2S-IV denotes the 2-step IV projection estimator introduced in Section 3 and 2S-IV-GLS is the corresponding GLS version by adding future reduced form residuals to the projection equation. Intervals for impulse responses are obtained by using HAC standard errors for IV-LP and IV-LP$^+$ and a parametric bootstrap (with $B = 499$ bootstrap replications) for IV-SVAR, 2S-IV and 2S-IV-GLS following Appendix A.2 of Stock and Watson (2018).

For $T = 100$, we observe from Table 3 that IV-LP has a comparably large bias at very low and high horizons. Adding control variables as in IV-LP$^+$ decreases the bias somewhat at shorter horizons. Nevertheless, we observe the familiar pattern that for large h the projection based methods lead to larger biases than IV-SVAR, which is based on an iterated estimator for the impulse responses. Similar to the OLS projections considered above, we again find that the 2-step GLS projection estimator (2S-IV-GLS) performs very much like the IV-SVAR method. The differences between the two groups is also evident from the standard deviations: All projection methods without GLS refinement lead to substantially higher standard deviations compared to IV-SVAR and 2S-IV-GLS at all horizons. Interestingly, within the group projection estimators without GLS refinement, we find that the 2-step IV method often performs somewhat better in terms of bias and standard deviation. Increasing the sample size to $T = 500$, decreases biases and standard deviations for all methods but the relative differences between them still persist.

[5] Note that in this simulation setup for $T = 100$ the average first-stage F-statistics from regressing y_{1t} on w_t is 24.33 and about 95% of the F-statistics exceed 10. For $T = 500$, we find an average F-statistic of 115.9 with a minimal value off 47.91.

Panel A of Figure 4 shows the interval coverage. For $T = 100$ and at larger horizons h, the IV-SVAR method and 2S-IV-GLS show a somewhat lower coverage compared to the other projection methods. Panel B shows that this is mostly due to the much wider confidence intervals (especially for IV-LP and IV-LP$^+$). Remarkably, the lengths of intervals produced by 2S-IV is much smaller than those from IV-LP and IV-LP$^+$. It seems that within the group of projection estimators without GLS refinement, 2S-IV with bootstrap intervals has the best trade-off of coverage and interval length. For $T = 500$, all methods yield comparable coverage very close to the nominal level. However, the projection methods (without GLS refinement) still lead to much wider intervals. For this larger sample size, the IV-SVAR and and 2S-IV-GLS seem to have the most favorable properties (good coverage and at the same time short intervals).

In additional simulations, we vary the persistence a_{11} and the correlation ρ_{w,ε_1}. Note that the choice of both parameters influences the first-stage F-statistics. In small samples with very high persistence we find some erratic behavior of the LP-IV estimator, potentially driven by the fact that the regressor y_{1t} in the projection step is quite persistent, while the instrument is not persistent at all.[6] The other methods (including LP-IV$^+$) are more robust in this respect. Whenever we choose a combination of a_{11} and ρ_{w,ε_1} that implies fairly low first-stage F-statistics, we are essentially in weak instrument situations, which affects all considered methods adversely (larger bias and wider intervals that often show coverage larger than the nominal level in small samples). We find, however, that the IV-LP and IV-LP$^+$ estimators are affected most strongly. These estimators again show sometimes erratic behavior with huge bias and extremely wide intervals if ρ_{w,ε_1} is small. In practical work, these erratic cases would not be considered further given the low first-stage F-statistics. A more detailed study of differences between the methods in weak instrument situations is beyond the scope of this paper and thus left for future research. Whenever a parameter combination implies fairly large first-stage F-statistics, the conclusions are very similar to those from results reported in Table 3 and Figure 4 are therefore not shown here.

6 Conclusion

In this paper we propose a two-step projection estimator for impulse responses in a structural VAR framework. In the first step the structural shock is typically obtained from its relationship to the VAR residuals. In the second step the impulse response is estimated from a regression of the future variable of interest $y_{i,t+h}$ on the estimated shock $\widehat{\varepsilon}_{j,t}$ and a vector of control variables $z_{j,t}$. The control variables are included for improving the efficiency but they may also be used to eliminate the estimation error from the first estimation step. We show that if the control variables are uncorrelated with the shock and include the vector of derivatives $\partial\varepsilon_{j,t}/\partial\beta_j$,

[6]We find a similar behavior in simulations that use a DGP based on the empirical example in Stock and Watson (2018).

then the estimation error from the identification step is negligible in the projection step. Furthermore, standard OLS inference applies even if the projection residuals are autocorrelated. Another advantage is that inference is valid no matter whether the variables are stationary or nonstationary (integrated).

In many cases the shocks cannot be represented by observed variables. In such cases, an instrumental variable approach can be employed resulting in an IV projection method. We show how this method can be adapted for estimating popular SVAR models like the proxy VAR, the AB-model and long-run restrictions. The asymptotic distribution of these estimators can easily be derived from the method-of-moment (MM) representation of the two-step approach. Finally, we point out that the OLS and IV projection methods are inefficient as the projection residuals are correlated up to $h-1$ lags. In order to improve the efficiency we propose a GLS projection that removes the serial correlation from the projection equation. We show that GLS projections are closely related (but not identical) to the iterative method of estimating impulse responses from the MA representation of a finite order VAR.

In our Monte Carlo simulations we find that the original OLS projection method suggested by Jordà (2005) performs similarly to our two-step projection estimator. An important advantage of our projection estimator is that no corrections for serial correlation or generated regressors is required. Applying the GLS refinement we observe substantial efficiency gains relative to OLS (or IV) projections.

References

Amisano, G. and Giannini, C. (1997). *Topics in Structural VAR Econometrics*, 2nd ed. Springer.

Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* **79**(4): 655–673.

Braun, R. and Brüggemann, R. (2017). Identification of SVAR models by combining sign restrictions with external instruments, *Working Paper 2017-07*, Department of Economics, University of Konstanz.

Breitung, J., Brüggemann, R. and Lütkepohl, H. (2004). Structural vector autoregressive modeling and impulse responses, *in* H. Lütkepohl and M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 159–196.

Fry, R. and Pagan, A. (2005). Some issues in using VARs for macroeconometric research, *Report*, The Australian National University.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections, *American Economic Review* **95**(1): 161–182.

Jordà, O., Schularick, M. and Taylor, A. M. (2015). Betting the house, *Journal of International Economics* **96**: S2–S18.

Jordà, O., Schularick, M. and Taylor, A. M. (2019). The effects of quasi-random monetary experiments, *Journal of Monetary Economics* p. forthcoming.

Kilian, L. and Kim, Y. J. (2011). How reliable are local projection estimators of impulse responses?, *Review of Economics and Statistics* **93**(4): 1460–1466.

Lusompa, A. (2019). Local projections, autocorrelation, and efficiency, *Working paper*, University of California, Irvine.

Lütkepohl, H. (1989). A note on the asymptotic-distribution of impulse-response functions of estimated VAR models with orthogonal residuals, *Journal of Econometrics* **42**(3): 371–376.

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

Plagborg-Moller, M. and Wolf, C. K. (2019). Local projections and VARs estimate the same impulse responses, *Working paper*, Princeton University.

Ramey, V. A. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from U.S. historical data, *Journal of Political Economy* **forthcoming**.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications, *American Economic Review* **94**(4): 1055–1084.

Romer, C. D. and Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks, *American Economic Review* **100**(3): 763–801.

Shapiro, M. D. and Watson, M. D. (1988). Sources of business cycle fluctuations, *NBER Macroeconomic Annual 1998* pp. 111–148.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**: 1–48.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments, *Economic Journal* **128**(610): 917–948.

Theil, H. (1971). *Principles of Econometrics*, Wiley.

Appendix

Proof of Proposition 1

A first order Taylor expansion around the true value β_j yields

$$\widehat{\varepsilon}_{ij} = f(x_{j,t}, \beta_j) + g(x_{j,t}, \beta_j)'(\widehat{\beta}_j - \beta_j) + O_p(T^{-1}).$$

Using $g(x_{j,t}, \beta_j) = Cz_{j,t}$ and the Frisch-Waugh theorem, the OLS estimator of θ_{ij}^h in (11) is equivalent to the OLS estimator of the regression

$$\widetilde{y}_{i,t+h} = \theta_{ij}^h \widetilde{\varepsilon}_{j,t} + \widehat{e}_{ij,t}^h$$

where $\widetilde{y}_{i,t+h}$ and $\widetilde{\varepsilon}_{j,t}$ are residuals from the regressions of $y_{i,t+h}$ and $\varepsilon_{j,t}$ on $z_{j,t}$, respectively. Since $\varepsilon_{j,t}$ and $z_{j,t}$ are uncorrelated, the limiting distribution of the OLS estimator

$$\widehat{\theta}_{ij}^h = \theta_{ij}^h + \frac{\sum_{t=p+1}^{T-h} \widetilde{\varepsilon}_{j,t} \widehat{e}_{ij,t}^h}{\sum_{t=p+1}^{T-h} \widetilde{\varepsilon}_{j,t}^2}$$

is given by

$$\sqrt{T}(\widehat{\theta}_{ij}^h - \theta_{ij}^h) = \frac{1}{\sigma_j^2 \sqrt{T}} \sum_{t=p+1}^{T-h} \widetilde{\xi}_t + o_p(1)$$

where $\sigma_j^2 = \mathbb{E}(\widetilde{\varepsilon}_{j,t}^2)$ and $\widetilde{\xi}_t = \widetilde{\varepsilon}_{j,t} \widehat{e}_{ij,t}^h$. Since $z_{j,t}$ is uncorrelated with $\varepsilon_{j,t}$ and $e_{ij,t}^h$ it follows that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T-h} \widetilde{\xi}_t = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T-h} \xi_t + o_p(1) \ ,$$

where $\xi_t = \varepsilon_{j,t} e_{ij,t}^h$. Furthermore, since $\varepsilon_{j,t}$ is a serially uncorrelated shock, and ξ_t is also uncorrelated and, therefore,

$$\begin{aligned} \mathbb{E}\left(\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T-h} \xi_t \right)^2 &= \mathbb{E}\left(\frac{1}{T} \sum_{t=p+1}^{T-h} \xi_t^2 \right) \\ &= \sigma_j^2 \sigma_{h,ij}^2, \end{aligned}$$

where $\sigma_{h,ij}^2 = \mathbb{E}[(e_{ij,t}^h)^2]$. It follows that the limiting distribution of $\widehat{\theta}_{ij}^h$ is given by

$$\sqrt{T}(\widehat{\theta}_{ij}^h - \theta_{ij}^h) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{h,ij}^2 / \sigma_j^2\right)$$

Note that the asymptotic variance is consistently estimated by the usual OLS variance estimator.

Table 1: Bias and Standard Deviation of Impulse Response Estimators for θ_{21}^{h}

Panel A: DGP: VAR(1) with low persistence, $\lambda_1^* = 0.2$, $\lambda_2^* = 0.4$

		Bias				Standard Deviation			
	h	Jorda	2step	2step-GLS	iterated	Jorda	2step	2step-GLS	iterated
$T = 100$	0	-0.003	-0.003	-0.003	-0.003	0.106	0.106	0.106	0.106
	1	-0.013	-0.008	-0.008	-0.013	0.138	0.144	0.144	0.138
	2	-0.017	-0.013	-0.002	-0.004	0.131	0.129	0.077	0.075
	3	-0.015	-0.010	0.000	-0.001	0.128	0.126	0.038	0.037
	4	-0.015	-0.014	0.000	0.000	0.130	0.128	0.018	0.017
	8	-0.007	-0.006	0.000	0.000	0.129	0.126	0.001	0.00
	12	-0.005	-0.002	0.000	0.000	0.131	0.130	0.000	0.000
	16	-0.012	-0.009	0.000	0.000	0.134	0.133	0.000	0.000
$T = 500$	0	-0.001	-0.001	-0.001	-0.001	0.046	0.046	0.046	0.046
	1	-0.004	-0.004	-0.004	-0.004	0.095	0.096	0.096	0.095
	2	-0.002	0.000	-0.001	-0.001	0.072	0.072	0.056	0.055
	3	-0.004	-0.003	0.000	0.000	0.059	0.057	0.026	0.026
	4	-0.002	-0.002	0.000	0.000	0.055	0.054	0.012	0.012
	8	0.000	0.000	0.000	0.000	0.055	0.053	0.000	0.000
	12	-0.001	-0.001	0.000	0.000	0.054	0.051	0.000	0.000
	16	-0.002	-0.001	0.000	0.000	0.055	0.053	0.000	0.000

Panel B: DGP: VAR(1) with high persistence, $\lambda_1^* = 0.8$, $\lambda_2^* = 0.4$

		Bias				Standard Deviation			
	h	Jorda	2step	2step-GLS	iterated	Jorda	2step	2step-GLS	iterated
$T = 100$	0	0.005	0.005	0.005	0.005	0.102	0.102	0.102	0.102
	1	-0.014	-0.011	-0.011	-0.014	0.198	0.209	0.209	0.198
	2	-0.018	-0.011	-0.013	-0.015	0.235	0.240	0.219	0.214
	3	-0.019	-0.010	-0.012	-0.013	0.232	0.235	0.197	0.192
	4	-0.020	-0.012	-0.010	-0.012	0.216	0.215	0.167	0.161
	8	-0.024	-0.012	-0.005	-0.006	0.183	0.173	0.077	0.074
	12	-0.019	-0.010	-0.002	-0.002	0.188	0.165	0.038	0.036
	16	-0.029	-0.023	0.000	0.000	0.190	0.174	0.020	0.019
$T = 500$	0	0.000	0.000	0.000	0.000	0.045	0.045	0.045	0.045
	1	-0.004	-0.004	-0.004	-0.004	0.164	0.167	0.167	0.164
	2	-0.004	-0.004	-0.004	-0.004	0.197	0.199	0.194	0.193
	3	-0.003	-0.003	-0.003	-0.003	0.187	0.188	0.179	0.178
	4	-0.003	-0.002	-0.002	-0.002	0.163	0.164	0.153	0.152
	8	-0.005	-0.002	-0.001	-0.001	0.102	0.097	0.068	0.068
	12	-0.004	-0.001	0.000	0.000	0.081	0.072	0.030	0.030
	16	-0.006	-0.002	0.000	0.000	0.080	0.071	0.014	0.014

Note: Results show bias and standard deviation of different structural impulse response estimators identified by a recursive ordering (Cholesky). Panel A: DGP is 2-variable VAR(1) with low persistence. Panel B: DGP is 2-variable VAR(1) with high persistence. Estimated models use $p = 1$. Results are based on $M = 1000$ Monte Carlo replications.

Table 2: Bias and Standard Deviation of Impulse Response Estimators of θ_{ij}^h

Panel A: θ_{14}^h (response of employment)

		Bias				Standard Deviation			
	h	Jorda	2step	2step-GLS	iterated	Jorda	2step	2step-GLS	iterated
$T = 500$	1	0.000	0.000	0.000	0.000	0.015	0.016	0.016	0.015
	2	-0.001	-0.001	-0.002	-0.002	0.027	0.027	0.025	0.023
	3	-0.001	-0.001	-0.002	-0.004	0.037	0.035	0.030	0.028
	4	-0.001	-0.001	-0.003	-0.004	0.045	0.042	0.033	0.031
	8	0.008	0.008	-0.001	0.000	0.072	0.065	0.050	0.048
	12	0.025	0.028	0.010	0.014	0.093	0.085	0.071	0.068
	16	0.050	0.053	0.029	0.035	0.117	0.107	0.093	0.089

Panel B: θ_{24}^h (response of consumer prices)

		Bias				Standard Deviation			
	h	Jorda	2step	2step-GLS	iterated	Jorda	2step	2step-GLS	iterated
$T = 500$	1	0.000	0.000	0.000	0.000	0.012	0.014	0.014	0.012
	2	-0.002	-0.001	-0.002	-0.002	0.025	0.024	0.024	0.022
	3	-0.004	-0.004	-0.005	-0.005	0.036	0.034	0.031	0.029
	4	-0.006	-0.006	-0.008	-0.008	0.046	0.042	0.037	0.034
	8	-0.021	-0.022	-0.026	-0.027	0.082	0.076	0.061	0.000
	12	-0.040	-0.044	-0.050	-0.052	0.117	0.107	0.091	0.086
	16	-0.057	-0.062	-0.076	-0.077	0.157	0.142	0.121	0.115

Panel C: θ_{64}^h (response of M2 growth rate)

		Bias				Standard Deviation			
	h	Jorda	2step	2step-GLS	iterated	Jorda	2step	2step-GLS	iterated
$T = 500$	0	0.000	0.000	0.000	0.000	0.026	0.026	0.026	0.026
	1	0.002	0.001	0.001	0.002	0.046	0.047	0.047	0.046
	2	0.008	0.0	0.00	0.007	0.067	0.067	0.066	0.064
	3	0.015	0.010	0.007	0.012	0.087	0.084	0.080	0.076
	4	0.021	0.015	0.010	0.018	0.105	0.099	0.089	0.085
	8	0.031	0.023	0.027	0.036	0.156	0.143	0.115	0.110
	12	0.034	0.022	0.033	0.039	0.188	0.174	0.133	0.129
	16	0.028	0.015	0.023	0.026	0.207	0.191	0.142	0.138

Note: Results show bias and standard deviation of different structural impulse response estimators identified by a recursive ordering (Cholesky). DGP: Jorda 6-variable VAR(2). Estimated models use lag length selected by AIC using $p_{\max} = 6$. Results are based on $M = 1000$ Monte Carlo replications.

Table 3: Bias and Standard Deviation of Impulse Response Estimators for θ_{21}^{h}

		Bias					Standard Deviation				
	h	IV-LP	IV-LP^{+}	2S-IV	2S-IV-GLS	IV-SVAR	IV-LP	IV-LP^{+}	2S-IV	2S-IV-GLS	IV-SVAR
$T = 100$	0	-0.052	-0.004	-0.002	-0.002	-0.004	0.304	0.224	0.204	0.204	0.212
	1	-0.040	-0.012	-0.010	-0.010	-0.006	0.281	0.279	0.146	0.146	0.141
	2	-0.039	-0.032	-0.023	-0.030	-0.032	0.299	0.313	0.147	0.133	0.124
	3	-0.037	-0.038	-0.035	-0.032	-0.041	0.313	0.342	0.157	0.124	0.115
	4	-0.038	-0.041	-0.041	-0.029	-0.037	0.331	0.364	0.159	0.107	0.100
	5	-0.037	-0.034	-0.038	-0.022	-0.028	0.347	0.379	0.162	0.087	0.082
	6	-0.053	-0.048	-0.044	-0.015	-0.018	0.321	0.353	0.157	0.068	0.064
$T = 500$	0	-0.011	0.000	0.000	0.000	0.000	0.127	0.085	0.084	0.084	0.085
	1	-0.005	0.002	-0.001	-0.001	0.000	0.113	0.101	0.064	0.064	0.060
	2	-0.007	-0.004	-0.004	-0.005	-0.005	0.123	0.120	0.063	0.057	0.054
	3	-0.006	-0.004	-0.004	-0.005	-0.007	0.132	0.131	0.064	0.052	0.050
	4	-0.005	-0.004	-0.003	-0.005	-0.007	0.136	0.135	0.065	0.045	0.044
	5	-0.009	-0.008	-0.006	-0.004	-0.005	0.139	0.140	0.067	0.038	0.037
	6	-0.009	-0.008	-0.009	-0.003	-0.004	0.138	0.140	0.067	0.030	0.030

Note: Results show bias and standard deviation of different structural impulse response estimators using the IV approach. DGP: 2-variable VAR(1) with medium persistence. Population correlation of instrument and structural shock is $\rho = .5$. Estimated models use lag length $p = 1$. Results are based on $M = 1000$ Monte Carlo replications.

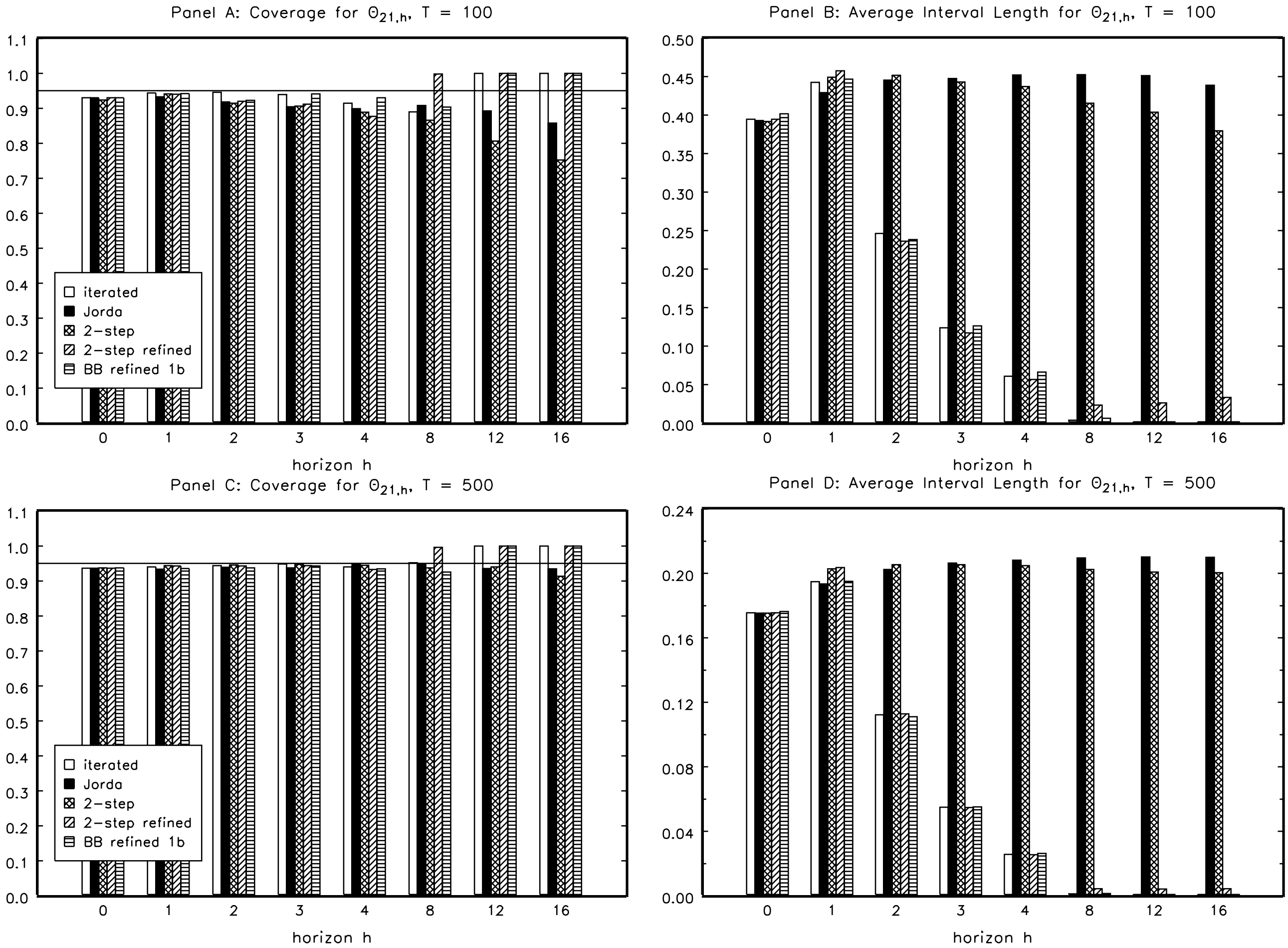

Figure 1: Coverage and average length of confidence intervals impulse response estimators of θ_{21}^{h} identified by a recursive ordering (Cholesky). DGP: 2-variable VAR(1) with low persistence. Estimated models use $p = 1$. Results are based on $M = 1000$ Monte Carlo replications.

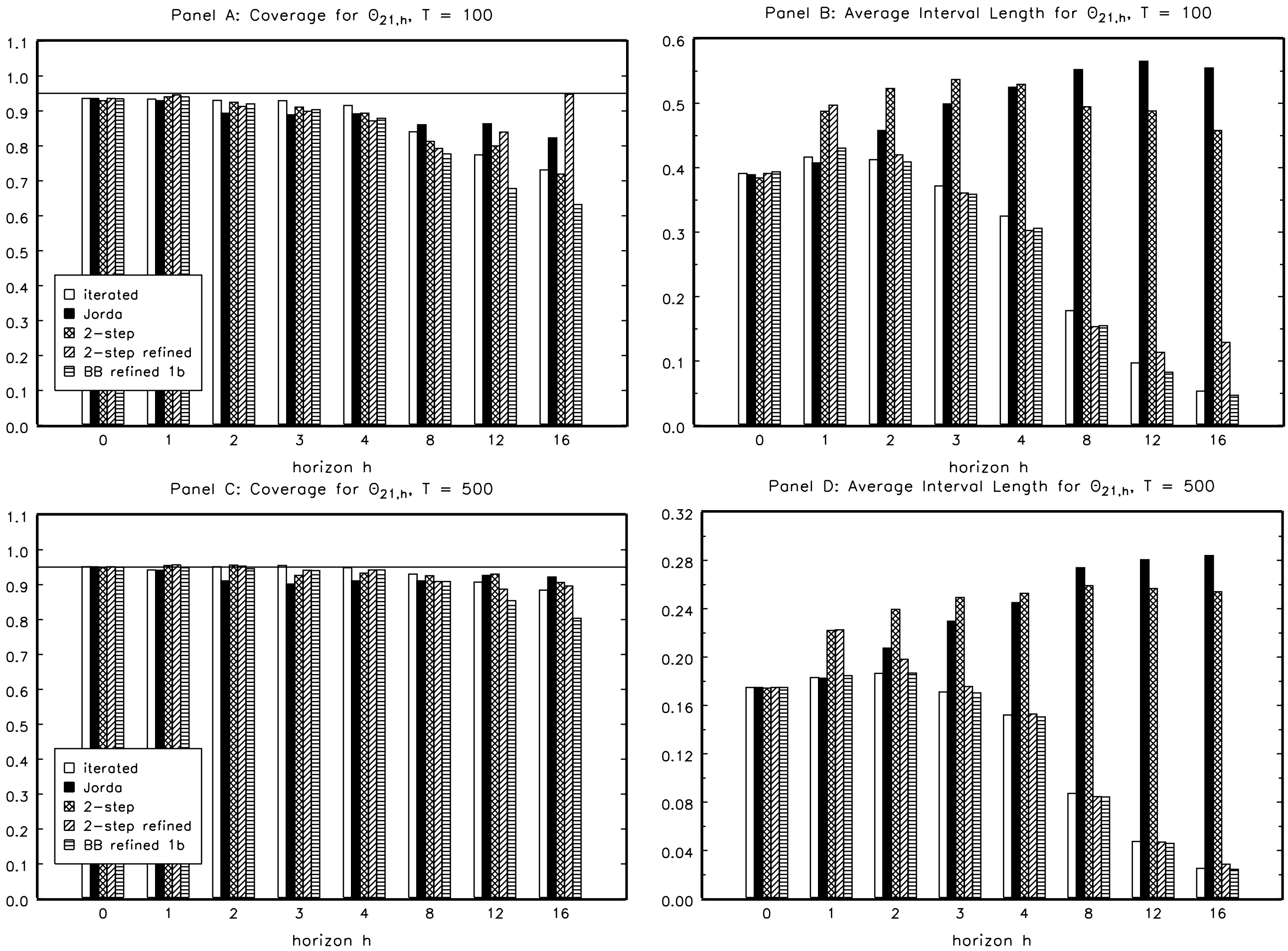

Figure 2: Coverage and average length of confidence intervals impulse response estimators of θ_{21}^{h} identified by a recursive ordering (Cholesky). DGP: 2-variable VAR(1) with high persistence. Estimated models use $p = 1$. Results are based on $M = 1000$ Monte Carlo replications.

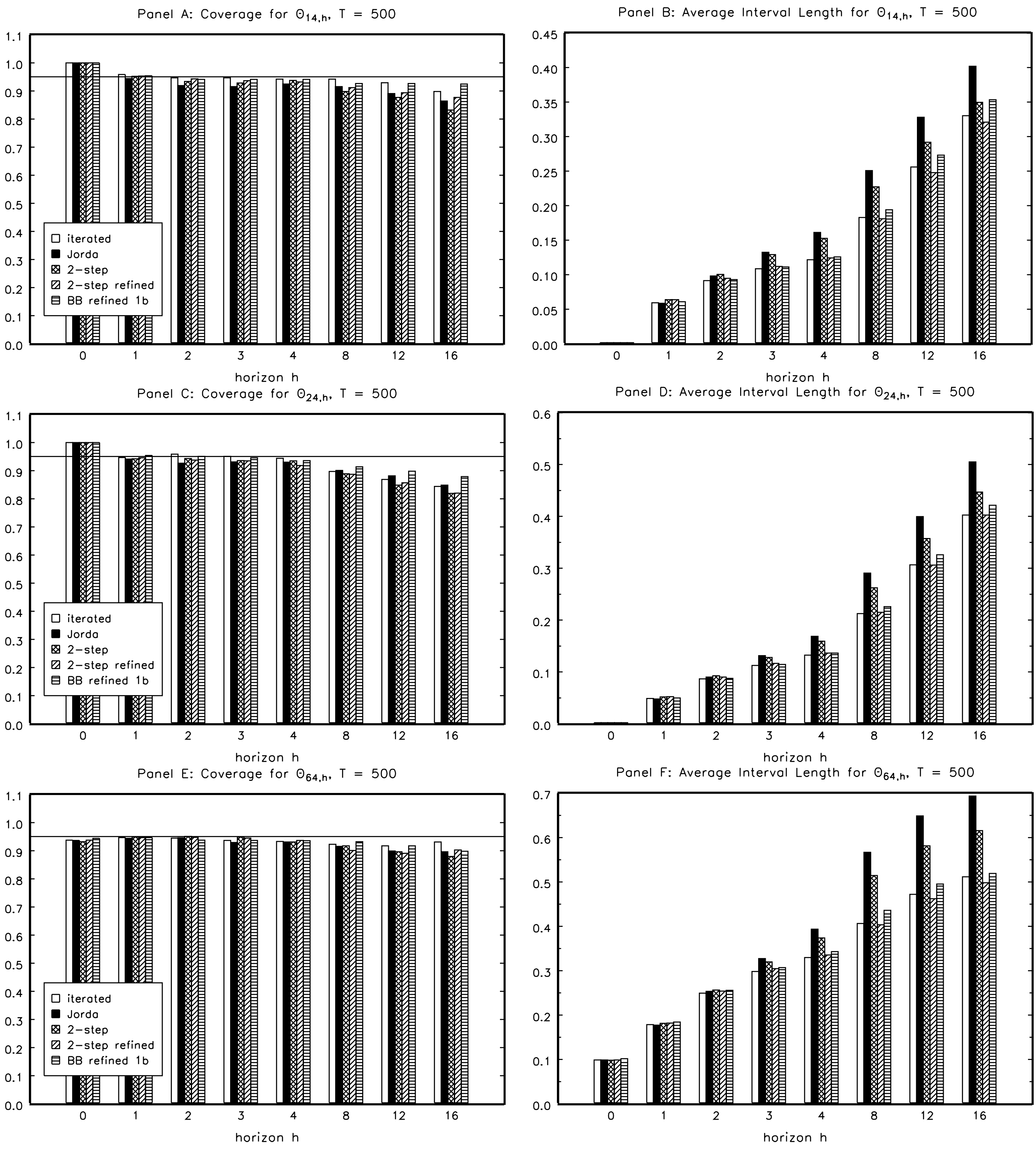

Figure 3: Coverage and average length of confidence intervals impulse response estimators of θ_{ij}^{h} identified by a recursive ordering (Cholesky). DGP: 6-variable VAR(2) with high persistence. Estimated models use lag length specified with $p_{\max} = 6$. Results are based on $M = 1000$ Monte Carlo replications.

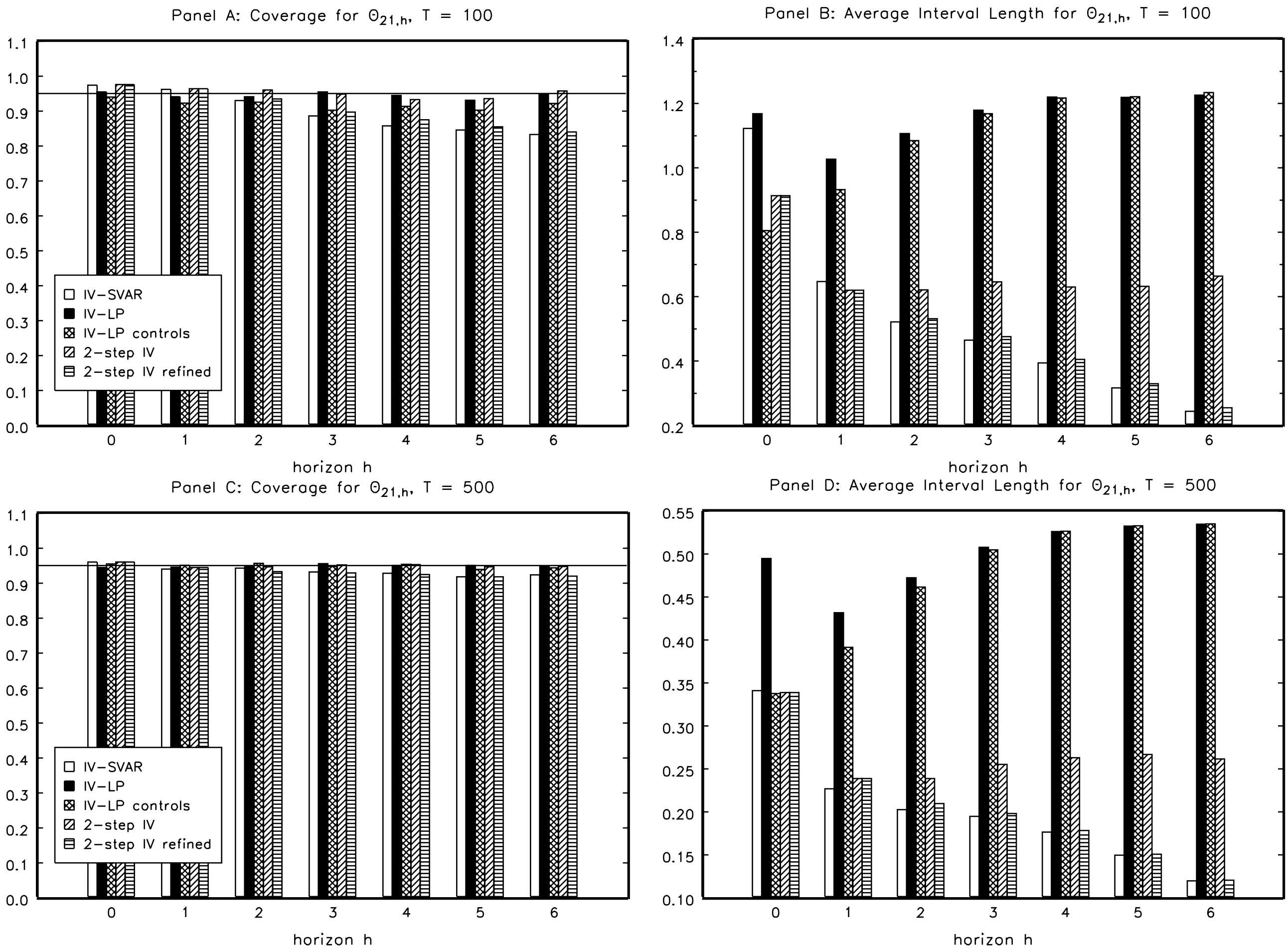

Figure 4: Coverage and average length of confidence intervals impulse response estimators of θ_{21}^h identified by IV. DGP: 2-variable VAR(1) with medium persistence. Population correlation of instrument and structural shock is $\rho = .5$. Estimated models use lag length $p = 1$. Results are based on $M = 1000$ Monte Carlo replications.