School of Economics Working Paper 2024-05

University of East Anglia

SCHOOL OF ECONOMICS

Avoiding Unintentionally Correlated Shocks in Proxy Vector Autoregressive Analysis

Martin Bruns *
Helmut Lütkepohl **
James McNeil ***

* University of East Anglia
** DIW Berlin and Freie Universität Berlin
*** Dalhousie University

School of Economics
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/economics

Avoiding Unintentionally Correlated Shocks in Proxy Vector Autoregressive Analysis

Martin Bruns

University of East Anglia, School of Economics,

Norwich Research Park, NR4 7TJ, Norwich, United Kingdom

Email: martin.j.bruns@gmail.com

Helmut Lütkepohl

DIW Berlin and Freie Universität Berlin, Mohrenstr. 58, 10117 Berlin, Germany

Email: luetkepohlh@gmail.com

James McNeil

Department of Economics, Dalhousie University, Halifax, Nova Scotia, Canada

Email: mcneilj@dal.ca

This version: July 12, 2024[1]

[1]The research presented in this paper was partly circulated earlier in working papers entitled *Unintentionally Correlated Shocks in Proxy Vector Autoregressive Analysis* and *Over-identifying structural vector autoregressions with scarce external instruments*. We are grateful to Luca Fanelli, Allan W. Gregory, and Gregor W. Smith for helpful comments. Earlier versions of the research were presented at the 2022 annual conferences organized by the Royal Economics Society, the Canadian Economics Association, the International Association for Applied Econometrics, the Atlantic Canada Economics Association, the annual meeting of the Econometrics Committee of the Verein für Socialpolitik in Berlin, 21 - 23 March 2024, and at the Workshop in Empirical Macroeconomics at King's College London, 16 April 2024. We thank the participants for helpful comments. Martin Bruns thanks the British Academy for financial support, grant number TDA22\220046. James McNeil acknowledges the support of the Social Sciences and Humanities Research Council (SSHRC), Insight Development Grant number 430-2022-0007.

Abstract

The shocks in structural vector autoregressive (VAR) analysis are typically assumed to be instantaneously uncorrelated. This condition may easily be violated in proxy VAR models if more than one shock is identified by a proxy variable. Correlated shocks may be obtained even if the proxies are uncorrelated and satisfy the usual relevance and exogeneity conditions individually. Examples from the recent proxy VAR literature are presented. It is shown that assuming uncorrelated proxies that satisfy the usual relevance and exogeneity conditions individually actually over-identifies the shocks of interest and a Generalized Method of Moments (GMM) algorithm is proposed that ensures orthogonal shocks and provides efficient estimators of the structural parameters. It generalizes an earlier GMM proposal that works only if at least $K-1$ shocks are identified by proxies in a VAR with K variables.

Key Words: Structural vector autoregression, proxy VAR, external instruments, correlated shocks, Generalized Method of Moments

JEL classification: C32, C36, E52

1 Introduction

In recent years it has become increasingly popular to identify structural shocks in structural VAR analysis by external instruments or proxies. In a number of studies, more than one shock of interest is identified in this way. In monetary policy analysis, this approach is, for instance, used by Altavilla, Brugnolini, Gürkaynak, Motto and Ragusa (2019) to account explicitly for the different nature of monetary policy shocks due to the range of tools at the disposal of central banks. Other examples of studies using a set of proxies to identify more than one structural shock are Mertens and Ravn (2013) who use two proxies to identify two tax shocks, Lunsford (2015) who identifies a consumption and an investment TFP shock by two proxies, Piffer and Podstawski (2017) who identify an uncertainty and a news shock, Lakdawala (2019) who identifies two monetary policy shocks, a fed funds rate shock and a forward guidance shock, Jarociński and Karadi (2020) who identify a monetary policy and a central bank information shock, and Känzig (2021) who uses two proxies to identify two oil market related shocks. Further examples are Arias, Rubio-Ramírez and Waggoner (2021) and Giacomini, Kitagawa and Read (2022), where Bayesian methods are applied.

If a set of proxies identifies several shocks, in general only linear combinations of the shocks are identified but they are not identified individually. For disentangling the shocks of interest, further information is required. In some studies, it is assumed that the proxies satisfy the usual relevance and exogeneity conditions for valid proxies individually and then they are used one-by-one to identify the shocks individually. It was pointed out by Gregory, McNeil and Smith (2024), however, that in this case the shocks may not be instantaneously uncorrelated any more and, hence, violate a standard assumption of structural VAR analysis. This can happen, even if the proxies are mutually uncorrelated and individually satisfy the standard relevance and exogeneity conditions that the proxy VAR literature typically assumes for valid proxies.

Having uncorrelated structural shocks is important for the interpretation of impulse responses, for example. If structural shocks are instantaneously correlated, then they are not likely to occur in isolation and, hence, impulse responses to individual shocks may not reflect properly the dynamics of the system under consideration. In fact, for realistically assessing the responses of the variables, one would have to consider the situation where several shocks hit at the same time. Clearly, responses to several shocks hitting jointly can be quite different from the impulse responses of the individual shocks. Hence, the causal interpretation of isolated shocks may be questionable in that case (see also the related discussions in Ramey (2016), Stock and Watson (2018), and Gregory et al. (2024)). Thus, it may be problematic that some of the shocks considered in recent structural VAR analysis are not instantaneously uncorrelated.

In this study we will first state the problem and its sources more formally. We show that even if the proxies are mutually uncorrelated and individually satisfy the standard assumptions for valid proxies, the resulting shocks may be correlated. However, if each proxy is correlated with one shock only, that information can be used to identify the shocks. In fact, imposing that the shocks are instantaneously uncorrelated, the assumption of each proxy being correlated with one shock only even over-identifies the shocks identified by the proxies.

Gregory et al. (2024) present a GMM method that provides uncorrelated shocks if the condition of each proxy being correlated with a single shock only is satisfied. Their method requires that, in a K-dimensional VAR model with K variables, at least $K-1$ shocks are identified by proxies. Unfortunately, that condition is not satisfied in a range of proxy VAR studies (e.g., Lunsford (2015), Lakdawala (2019), Jarociński and Karadi (2020), Känzig (2021), Piffer and Podstawski (2017), Fanelli and Marsi (2022)). We propose a simple GMM method that also works more generally even if there are fewer proxies and shocks to be identified by them. The method permits to focus exclusively on the structural parameters of interest in the GMM objective function and, hence, will typically result in a computationally simple optimization problem. A device proposed by Crepon, Kramarz and Trognon (1997) is used to ensure that the method provides asymptotically efficient estimators under standard assumptions. It also ensures a valid J-test for model misspecification.

We will present examples of empirical studies, where using the proxies one-by-one in the usual way leads to correlated shocks and we will show that using our GMM approach for estimation can, for instance, make a difference for the impulse responses. Thereby we show that avoiding correlated shocks is not only a theoretical problem but is relevant for applied work.

The remainder of the paper is organized as follows. In the next section we present the model setup, formulate the problem of getting correlated shocks formally, and present the GMM procedure that can be used for solving the problem. In Section 4 we present the empirical examples and Section 5 concludes.

The following general notation will be used throughout. The operator $\mathrm{vec}(\cdot)$ is the usual column vectorization operator for a matrix. $\mathrm{vech}(\cdot)$ is the corresponding operator vectorizing a square matrix from the main diagonal downwards, and $\mathrm{vh}(\cdot)$ is the vectorization operator that collects only the elements below the main diagonal of a matrix in a vector. Moreover, we use the following special matrices (see also Lütkepohl (1996) for some of the definitions and properties of the matrices): $\mathbf{S}_m$ is a $(\frac{1}{2}m(m-1) \times m^2)$ selection matrix that selects the elements below the main diagonal of an $(m \times m)$ matrix M from $\mathrm{vec}(M)$, i.e., $\mathrm{vh}(M) = \mathbf{S}_m \mathrm{vec}(M)$. The $(\frac{1}{2}m(m+1) \times m^2)$ elimination matrix $\mathbf{L}_m$ is defined such that, for an $(m \times m)$ matrix M, $\mathrm{vech}(M) = \mathbf{L}_m \mathrm{vec}(M)$. Furthermore, $\mathbf{D}_m$ is the $(m^2 \times \frac{1}{2}m(m+1))$ duplication matrix defined such that, for a symmetric $(m \times m)$ matrix M, $\mathrm{vec}(M) = \mathbf{D}_m \mathrm{vech}(M)$. Finally, $\mathbf{K}_{mm}$ denotes the $(m^2 \times m^2)$ commutation matrix defined such that, for an $(m \times m)$ matrix M, $\mathrm{vec}(M') = \mathbf{K}_{mm} \mathrm{vec}(M)$.

2 Model Setup and Problem Discussion

2.1 The Model

Our basic model is a K-dimensional reduced-form VAR process,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t = (\nu, A_1, \ldots, A_p)Y_{t-1} + u_t, \tag{1}$$

where u_t is a zero mean white noise process with nonsingular covariance matrix Σ_u, i.e., $u_t \sim (0, \Sigma_u)$ and $Y_{t-1} = (1, y'_{t-1}, \ldots, y'_{t-p})'$ is a $(Kp+1)$-dimensional column vector.

The vector of structural shocks is denoted by $\mathbf{w}_t = (w_{1t}, \ldots, w_{Kt})'$. It is obtained from the reduced-form errors, u_t, by a linear transformation, $\mathbf{w}_t = B^{-1}u_t$. The $(K \times K)$ matrix $B = [b_{ij}]$ contains the impact effects of the structural shocks and $B\Sigma_{\mathbf{w}}B' = \Sigma_u$, where $\Sigma_{\mathbf{w}}$ is the covariance matrix of $\mathbf{w}_t$. As in much of the structural VAR literature, the shocks are assumed to be instantaneously uncorrelated and, hence, the transformation matrix B is such that the covariance matrix $\Sigma_{\mathbf{w}}$ is diagonal and the structural shocks $\mathbf{w}_t$ are instantaneously uncorrelated by construction.

As we are also considering partially identified models, we partition $\mathbf{w}_t$ in K_1- and $(K-K_1)$-dimensional subvectors $\mathbf{w}_{1t} = (w_{1t}, \ldots, w_{K_1t})'$ and $\mathbf{w}_{2t} = (w_{K_1+1,t}, \ldots, w_{Kt})'$ such that $\mathbf{w}'_t = (\mathbf{w}'_{1t}, \mathbf{w}'_{2t})$. The first K_1 shocks, $\mathbf{w}_{1t}$, are the structural shocks of interest. They have to be identified properly, while $\mathbf{w}_{2t}$ contains shocks which are not in the focus of the analysis and are, hence, not necessarily identified as proper economic shocks. The matrix of impact effects, B, is partitioned accordingly as $B = [B_1 : B_2]$, where B_1 is $(K \times K_1)$ and B_2 is $(K \times (K - K_1))$. In other words, B_i contains the impact effects of the shocks $\mathbf{w}_{it}$, $i = 1, 2$.

The impact effects, B, are the structural parameters of the model. The shocks $\mathbf{w}_{1t}$ are identified if the B_1 matrix is identified. Having the matrix B_1, we can compute the structural impulse responses to the $\mathbf{w}_{1t}$ shocks for propagation horizon h as

$$\Theta_{1h} = \Phi_h B_1,$$

where the Φ_h are reduced-form quantities obtained recursively from the VAR slope coefficients as $\Phi_h = \sum_{j=1}^{h} \Phi_{h-j} A_j$, with $\Phi_0 = I_K$, for $h = 0, 1, \dots$, and $A_j = 0$ for $j > p$ (see, e.g., Lütkepohl (2005, Sec. 2.1.2)).

Each column of B contains the impact effects of a single shock on all the K variables. Denoting by b_k the k-th column of B, the k-th shock can be obtained from the reduced-form residuals as

$$w_{kt} = b_k' \Sigma_u^{-1} u_t / b_k' \Sigma_u^{-1} b_k \tag{2}$$

(see, e.g., Stock and Watson (2018), Bruns and Lütkepohl (2022, Appendix A.1)).

2.2 Identification via Proxy Variables

Identification of the structural parameters and, hence, the structural shocks is assumed to be based on a set of N instrumental variables (proxies) $z_t = (z_{1t}, \dots, z_{Nt})'$ satisfying

$$\mathbb{E}(\mathbf{w}_{1t} z_t') = \Sigma_{\mathbf{w}_1 z} \neq 0, \quad \Sigma_{\mathbf{w}_1 z}\ (K_1 \times N), \quad \mathrm{rk}(\Sigma_{\mathbf{w}_1 z}) = K_1 \quad \text{(relevance)}, \tag{3}$$

$$\mathbb{E}(\mathbf{w}_{2t} z_t') = 0 \quad \text{(exogeneity)}. \tag{4}$$

These conditions imply that

$$\mathbb{E}(u_t z_t') = B\mathbb{E}(\mathbf{w}_t z_t') = B_1 \Sigma_{\mathbf{w}_1 z}. \tag{5}$$

Obviously, there must be at least as many proxies as there are identified shocks such that $N \geq K_1$, to satisfy the rank condition for $\Sigma_{\mathbf{w}_1 z}$ which ensures that the N proxies contain identifying information for all shocks in $\mathbf{w}_{1t}$. As we can estimate $B_1 \Sigma_{\mathbf{w}_1 z}$ by the usual covariance matrix estimator

$$\overline{\hat{u}z} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t z_t', \tag{6}$$

where the $\hat{u}_t$ are reduced-form least squares (LS) residuals, the proxies contain identifying information for the first K_1 structural shocks collectively but the shocks are not necessarily individually identified. However, if each proxy is correlated with just one shock such that $\Sigma_{\mathbf{w}_1 z}$ is a diagonal square matrix, the shocks will be identified individually because the right-hand side of (5) will consist of multiples of the impact effects of the shocks that will provide multiples of the shocks via the relation (2). In general, if $\Sigma_{\mathbf{w}_1 z}$ is not a diagonal matrix, the proxies identify only linear combinations of the shocks of interest.

For ease of exposition, we assume from now on that there are precisely as many proxies as there are structural shocks of interest, i.e., $N = K_1$ and $\Sigma_{\mathbf{w}_1 z}$ is a square matrix. In that case, if $\Sigma_{\mathbf{w}_1 z}$ is a diagonal matrix, each proxy satisfies the relevance and exogeneity conditions individually, i.e.,

$$\mathbb{E}(w_{kt} z_{kt}) = c_k \neq 0, \quad \text{(relevance)}, \tag{7}$$

$$\mathbb{E}(w_{jt} z_{kt}) = 0, j \in \{1, \dots, K\}, j \neq k, \quad \text{(exogeneity)}, \tag{8}$$

such that $\mathbb{E}(u_t z_{kt}) = B\mathbb{E}(\mathbf{w}_t z_{kt}) = c_k b_k$. Thus, z_{kt} identifies a multiple of the impact effects of the k-th shock. Note that, if we just have a multiple of b_k and use that in (2) in place of b_k, we get a multiple of the k-th shock. That is not a problem in practice because the analyst has to take a stand on the size of the shock she wants to look at anyway and she has to scale the shock accordingly.

If the impact effects of an individual shock are estimated as

$$\widehat{c_k b_k} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t z_{kt},$$

we will refer to this approach as the conventional proxy VAR approach to estimating the impact effects of the shocks of interest. Obviously, this estimator is identical to the one obtained by using the estimator in equation (6) and thereby estimating the impact effects of all K_1 shocks of interest jointly. The shocks obtained in this way may, however, be instantaneously correlated, as pointed out already by Gregory et al. (2024), because there is no mechanism that enforces uncorrelatedness. Effectively, with each of the proxies we are estimating a column of a B matrix, say $B^{(i)}$, that satisfies $B^{(i)}\Sigma_{\mathbf{w}}^{(i)}B^{(i)\prime} = \Sigma_u$ with a diagonal matrix $\Sigma_{\mathbf{w}}^{(i)}$. For two proxies z_{1t} and z_{2t} we do not necessarily estimate a column of the same B matrix, however, and, hence, it is possible that $B^{(1)} \neq B^{(2)}$.

If we were to specify two proxies z_t and three shocks $\mathbf{w}_t$ such that

$$\begin{bmatrix} z_t \\ \mathbf{w}_t \end{bmatrix} \sim \mathcal{N}\left(0, \left[\begin{array}{cc|ccc} 1 & 0 & * & 0 & 0 \\ 0 & 1 & 0 & * & 0 \\ \hline * & 0 & 1 & * & 0 \\ 0 & * & * & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]\right), \tag{9}$$

where * denotes a nonzero element chosen such that the covariance matrix is positive definite, then the proxies are mutually uncorrelated and satisfy individually the relevance and exogeneity conditions (7) and (8) such that each proxy is correlated with one shock only. Still, that can go together with correlated shocks, as the example in (9) shows. Thus, correlated shocks may in principle come up also in empirical work based on the conventional proxy VAR approach.

Clearly, it is attractive to estimate the structural shocks without having to draw on additional identifying information because identifying information based on exclusion restrictions on the impact or long-run effects of the shocks or on sign restrictions is often debatable and not shared by the entire profession. That may have motivated some researchers to construct their proxies such that each of them is correlated with exactly one shock only. In the next section it will be shown that such a property together with the requirement of uncorrelated shocks actually over-identifies the shocks of interest. Taking all the identifying information properly into account avoids the problem of getting correlated shocks.

2.3 Orthogonalizing the Shocks

Gregory et al. (2024) propose a generalized method of moments (GMM) approach that ensures uncorrelated (orthogonal) shocks. They use moment conditions obtained from the assumption that each proxy is correlated with one shock only which implies that all the other shocks are uncorrelated with the proxy, giving $K-1$ moment conditions for each proxy. In addition, they use that all the shocks are mutually uncorrelated. Thereby they obtain $\frac{1}{2}K(K-1)$ further moment conditions. Finally, they standardize the B matrix to have a unit diagonal. In other words, they assume that each shock has a unit impact effect for one of the variables. Thereby they have to estimate just $K(K-1)$ structural parameters in the B matrix. A drawback of the Gregory et al. (2024) approach is that it works only if at least $K-1$ shocks are identified by proxies. Otherwise they do not have enough moment conditions to identify all the parameters. Clearly, there are many examples of proxy VAR studies, where less than $K-1$ shocks are identified by proxies and, hence, their approach does not work any more. It is also a disadvantage of their approach that, for each shock, they have to take a stand on a specific variable having a nonzero instantaneous response. In the literature, there are many studies, where the response of the variables to the shocks is uncertain and is only the outcome of the analysis and not known before the analysis.

In the following, we present a GMM method that works more generally also if $K_1 < K-1$ because we use a different set of moment conditions. Our moment conditions are focussed on the first K_1 shocks of interest that are identified by proxies and do not involve moment conditions related to other shocks. We assume that $\Sigma_{\mathbf{w}_1 z}$ is diagonal and standardize the variances of the shocks such that $\Sigma_{\mathbf{w}_1 z} = I_{K_1}$. Thereby we get KK_1 moment conditions

$$\mathbb{E}(u_t z_t' - B_1) = 0 \tag{10}$$

from (5). Moreover, using $u_t = B\mathbf{w}_t$ and, hence, $\Sigma_u = B\Sigma_{\mathbf{w}}B'$, where $\Sigma_{\mathbf{w}}$ is the covariance matrix of the standardized $\mathbf{w}_t$ shocks, we have

$$\mathbb{E}(B_1'\Sigma_u^{-1}u_tu_t'\Sigma_u^{-1}B_1) = B_1'B'^{-1}\Sigma_{\mathbf{w}}^{-1}B^{-1}B_1 = [I_{K_1} : 0]\Sigma_{\mathbf{w}}^{-1}\begin{bmatrix} I_{K_1} \\ 0 \end{bmatrix}.$$

Uncorrelated shocks imply that $\Sigma_{\mathbf{w}}$ is a diagonal matrix. Hence, considering only the elements below the main diagonal of the left-hand side matrix, we get a further set of $\frac{1}{2}K_1(K_1-1)$ moment conditions

$$\mathbb{E}[\text{vh}(B_1'\Sigma_u^{-1}u_tu_t'\Sigma_u^{-1}B_1)] = 0. \tag{11}$$

Note that there are KK_1 free parameters in B_1, which are already identified by the KK_1 moment conditions (10) so that the additional moment conditions (11) over-identify B_1 if we have multiple proxies. In other words, assuming that each proxy is correlated with a single shock only does not only identify the shocks but even over-identifies them if $K_1 > 1$.

Note also that the moment conditions depend on the reduced-form VAR parameters $\alpha = \text{vec}(\nu, A_1, \dots, A_p)$ via $u_t(\alpha) = y_t - (Y_{t-1}' \otimes I_K)\alpha$ and $\sigma = \text{vech}(\Sigma_u)$. Thus, if we focus on estimating $\beta = \text{vec}(B_1)$, we still have to account for nuisance parameters $\gamma = (\alpha', \sigma')'$. The standard moment conditions for the reduced-form VAR parameters are

$$\mathbb{E}[(Y_{t-1} \otimes I_K)y_t - (Y_{t-1}Y_{t-1}' \otimes I_K)\alpha] = 0 \tag{12}$$

and

$$\mathbb{E}(\Sigma_u - u_tu_t') = 0. \tag{13}$$

The empirical moments corresponding to the moment conditions for the structural parameters (10) and (11) are

$$\bar{m}^{\beta}(\beta,\gamma) = \frac{1}{T}\sum_{t=1}^{T} m_t^{\beta}(\beta,\gamma) \qquad \big(KK_1 + \tfrac{1}{2}K_1(K_1-1)\big)\times 1$$

with

$$m_t^{\beta}(\beta,\gamma) = \begin{bmatrix} \operatorname{vec}(u_t(\alpha)z_t' - B_1) \\ \\ \operatorname{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1) \end{bmatrix} \tag{14}$$

and for the reduced-form parameters we get empirical moments

$$\bar{m}^{\gamma}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} m_t^{\gamma}(\gamma) \qquad \big(K(Kp+1) + \tfrac{1}{2}K(K+1)\big)\times 1$$

with

$$m_t^{\gamma}(\gamma) = \begin{bmatrix} (Y_{t-1}\otimes I_K)y_t - (Y_{t-1}Y_{t-1}'\otimes I_K)\alpha \\ \\ \operatorname{vech}(\Sigma_u - u_t(\alpha)u_t(\alpha)') \end{bmatrix} \tag{15}$$

Using a result by Crepon et al. (1997, Proposition 1), we can set up an efficient GMM procedure by specifying the GMM objective function as

$$J(\beta) = T\bar{m}^{\beta}(\beta,\hat{\gamma})'\Omega(\hat{\beta},\hat{\gamma})^{-1}\bar{m}^{\beta}(\beta,\hat{\gamma}), \tag{16}$$

where $\Omega(\beta,\gamma)$ is a suitable GMM weighting matrix, $\hat{\beta}$ is a consistent first-stage estimator of β and $\hat{\gamma}$ is a consistent estimator of γ that satisfies the condition

$$\bar{m}^{\gamma}(\hat{\gamma}) = 0. \tag{17}$$

In the present setup, we can use the least squares (LS) estimator $\hat{\gamma}^{LS}$ as estimator for γ because it satisfies condition (17).

An asymptotically efficient estimator for β, denoted by $\hat{\beta}^{GMM}$ in the following, is obtained if the weighting matrix $\Omega(\beta, \gamma)$ is chosen as

$$\Omega(\hat{\beta}, \hat{\gamma}^{LS}) = \frac{1}{T} \sum_{t=1}^{T} \omega_t(\hat{\beta}, \hat{\gamma}^{LS}) \omega_t(\hat{\beta}, \hat{\gamma}^{LS})', \tag{18}$$

where $\hat{\beta}$ is some consistent first-stage estimator of β and

$$\omega_t(\beta, \gamma) = m_t^{\beta}(\beta, \gamma) - \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial m_t^{\beta}(\beta, \gamma)}{\partial \gamma'} \right) \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial m_t^{\gamma}(\gamma)}{\partial \gamma'} \right)^{-1} m_t^{\gamma}(\gamma) \tag{19}$$

(see Crepon et al. (1997)). It is shown in Appendix A that the correction term simplifies to

$$\omega_t(\beta, \hat{\gamma}^{LS}) = m_t^{\beta}(\beta, \hat{\gamma}^{LS}) - \left[\begin{array}{c} \left((\frac{1}{T} \sum_{t=1}^{T} z_t Y_{t-1}')(\frac{1}{T} \sum_{t=1}^{T} Y_{t-1} Y_{t-1}')^{-1} Y_{t-1} \otimes I_K \right) \hat{u}_t \\ -2\mathbf{S}_{K_1} \mathrm{vec} \left(B_1' \widehat{\Sigma}_u^{-1} (\widehat{\Sigma}_u - \hat{u}_t \hat{u}_t') \widehat{\Sigma}_u^{-1} B_1 \right) \end{array} \right] \tag{20}$$

if γ is replaced by the LS estimator $\hat{\gamma}^{LS}$. Here $\hat{u}_t$ denotes again reduced-form LS residuals and $\widehat{\Sigma}_u = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$. For $\beta = \mathrm{vec}(B_1)$ we may choose a consistent first-stage estimator obtained, e.g., by minimizing the GMM objective function (16) with $\Omega(\beta, \gamma) = I_{KK_1 + \frac{1}{2}K(K+1)}$. The procedure can also be iterated, using the latest estimate of β and $\hat{\gamma}^{LS}$ in each iteration step.

Using this approach, we estimate the impact effects of the first K_1 structural shocks that can be obtained as $\mathbf{w}_{1t} = B_1' \Sigma_u^{-1} u_t$ from the reduced-form residuals. The GMM approach aims at estimating the structural shocks of interest in such a way that they are instantaneously uncorrelated. It should be clear that, due to the over-identifying moment conditions, the empirical correlation between the components of the estimated $\mathbf{w}_{1t}$ may be nonzero because $J(\hat{\beta}^{GMM}) > 0$. This feature can also be used to set up Hansen's J-test for misspecification using that

$$J(\hat{\beta}^{GMM}) \xrightarrow{d} \chi^2 \left(\tfrac{1}{2} K_1 (K_1 - 1) \right) \tag{21}$$

under standard GMM assumptions, if the model and the moment conditions are correctly specified. Note that the asymptotic χ^2-distribution is obtained because we have corrected for the nuisance parameters.

The diagonality of $\Sigma_{\mathbf{w}_1 z}$ is one assumption that may not hold in practice and, diagnosing model misspecification by a Hansen test may be a signal of incorrect moment conditions. The advantage of our GMM approach is that it provides estimates of the shocks and, hence, we can also estimate the $\Sigma_{\mathbf{w}_1 z}$ matrix and directly check whether it is diagonal.

We have standardized the shocks such that $\Sigma_{\mathbf{w}_1 z}$ is an identity matrix. For an empirical analysis we can, of course, rescale the shocks to have the desired size. For example, if a monetary policy shock is identified that moves an interest rate on impact, we can rescale the column of $\hat{B}_1$ corresponding to the shock such that the interest rate changes by, say, 25 basis points on impact.

Another approach for estimating the structural parameters B_1 is discussed by Angelini and Fanelli (2019).[2] These authors assume a parametric model for the proxies and augment the VAR model by the proxies. Then they set up a minimum distance procedure that minimizes the distance of the structural parameters from the reduced-form parameters. Their approach also works for proxy VAR models where less than $K-1$ shocks are identified by proxies. However, in addition to the B_1 matrix, the minimum distance method also estimates parameters of the model for the proxies. Given the way some proxies are constructed in the recent literature, their model is not a universally good approximation of the generating mechanism of the proxies. In particular, their model does not match the situation where observations for the proxy are only available at infrequent event dates and the proxy is set to zero on all other dates as, e.g., in Piffer and Podstawski (2017), Wright (2012), Boer and Lütkepohl (2021), Gertler and Karadi (2015) and many other studies. Clearly, our GMM approach has the advantage of focussing exclusively on the parameters of interest, B_1, and it does not assume a specific model for the generating mechanism

[2] Related research is also reported by Garchi Casal and Zimic (2023).

of the proxies and therefore also accommodates proxies with many zero values during the sample period.

3 Monte Carlo Study

An attractive feature of the orthogonality conditions introduced in the previous section is that they provide over-identifying moments, which are testable with the J-statistic. Unless the nuisance parameters are properly accounted for, however, this test will not follow the expected distribution. One might also expect the over-identified model to yield more precise estimates compared with the usual proxy SVAR procedure since it incorporates additional information to estimate the same number of parameters. That both of these points are observed in finite samples is demonstrated with the following small Monte Carlo experiment.

The data generating process is

$$y_t = A_1 y_{t-1} + B\mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, \Sigma_{\mathbf{w}}), \tag{22}$$

where y_t and $\mathbf{w}_t$ are (3×1) vectors and $\Sigma_{\mathbf{w}}$ is a diagonal matrix. We fit VAR(4) models with constant term and consider sample sizes, exclusive of pre-sample values, of $T = 100$ and 500 and the parameter matrices are

$$A_1 = \begin{bmatrix} 0.9 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}, \quad \Sigma_{\mathbf{w}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma^2_{w_3} \end{bmatrix}.$$

There are $K_1 = 2$ instruments available to identify the first two shocks. These two shocks are related to the vector of proxies, z_t, according to

$$z_t = \mathbf{w}_{1t} + v_t, \quad v_t \sim \mathcal{N}(0, \Sigma_v), \tag{23}$$

where v_t is independent of $\mathbf{w}_{1t}$ and Σ_v is a diagonal matrix. The matrix A_1 has a maximum eigenvalue of 0.9 and, thus, the VAR process is stable but quite persistent. We set $\sigma_{w_3}^2 = 0.01$ or 1 and simulate 5000 Monte Carlo replications. The low value of $\sigma_{w_3}^2 = 0.01$ reflects an environment in which the two identified shocks account for a majority of the variation in all three variables. This case may be of practical interest as macroeconomists often study the most important sources of economic variation. A value of $\sigma_{w_3}^2 = 1$ reflects a situation where all three shocks are equally important.

The strength of the proxies as instrumental variables is determined by the correlations between the two identified shocks and their respective proxy variables

$$\mathrm{Corr}(w_{it}, z_{it}) = \frac{\mathrm{Var}(w_{it})}{\sqrt{\mathrm{Var}(w_{it}) + \mathrm{Var}(v_{it})}}.$$

Since the two identified shocks have unit variance, we adjust the diagonal elements of Σ_v to achieve an intermediate correlation of 0.5 between the instruments and their associated shocks.

For each simulation we produce three estimates of the first two columns of B, i.e. for B_1. First, we apply the usual proxy SVAR procedure, which uses the two instrumental variables (proxies) but places no restriction on the correlation between the two identified shocks. Thus, it uses the moment conditions (10) such that the estimator for the impact effects is $\hat{B}_1 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t z_t'$. Second, we apply the GMM method outlined in the previous section, which incorporates the over-identifying moment condition based on the orthogonality of the estimated shocks, and adjusts the GMM weighting matrix using the Crepon et al. (1997) method. The moment conditions for this estimator are (10) and (11) and the GMM weighting matrix is (18). As first stage estimates for β and γ we use $\hat{\beta} = \mathrm{vec}(\hat{B}_1)$ and $\hat{\gamma}^{LS}$, respectively. The resulting estimator will be referred to as adjusted GMM estimator in the following. Third, we again use the moments (10) and (11) but ignore the nuisance parameters α and Σ_u and use as GMM weighting matrix

$$\frac{1}{T}\sum_{t=1}^{T} m_t^{\beta}(\hat{\beta}, \hat{\gamma}^{LS}) m_t^{\beta}(\hat{\beta}, \hat{\gamma}^{LS})'. \tag{24}$$

The corresponding estimator will be called unadjusted GMM estimator.

Figures 1 and 2 show kernel densities for the estimates of the six identified parameters of B_1, for each of the three estimators when we set $\sigma^2_{w_3} = 0.01$ and 1, respectively. The vertical black lines correspond with the true values of the parameters. The following observations emerge from the figures:

1. The two GMM estimators clearly dominate the conventional proxy VAR estimator in that their densities are generally more or at least not less concentrated than the densities of the conventional proxy VAR estimates and, hence, the GMM estimators have smaller variances.

2. For some of the parameters the densities of all three estimators are concentrated around values different from the true parameter value in small samples. In other words, they are biased. The bias is similar for all three estimators and declines quickly with growing sample size.

3. The adjusted and unadjusted GMM estimators have very similar small sample densities. In other words, for our limited simulation designs, the small sample properties of the estimators do not depend much on the adjustment of the GMM weighting matrix.

Table 1 shows rejection frequencies for the J-test calculated for the two over-identified GMM estimators at three popular significance levels. If the test statistic has the correct size, we would expect it to exceed the 10% critical value approximately 10% of the time, and likewise for any other critical value. We see that this is the case even for a relatively small sample size of $T = 100$ for the adjusted GMM but the test is severely undersized for the unadjusted GMM, indicating that a test of the over-identifying restrictions would under-reject. The rejection frequencies for the unadjusted GMM do not improve as the sample size increases and it is found for both values of $\sigma^2_{w_3}$. This illustrates that the adjustment for the GMM procedure outlined in the previous section is crucial for obtaining a reliable J-test.

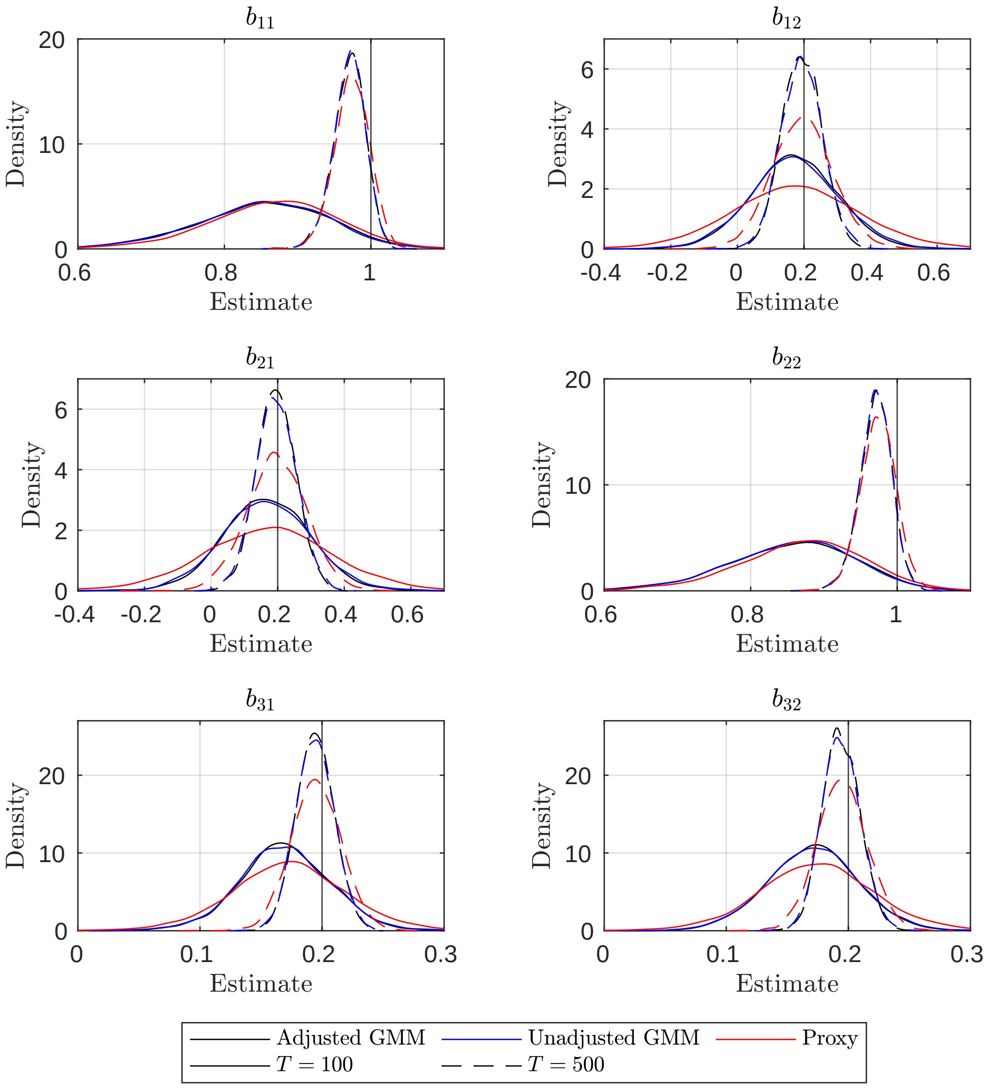

Figure 1: Monte Carlo Kernel Densities of Estimates of B_1 when $\sigma^2_{w_3} = 0.01$.

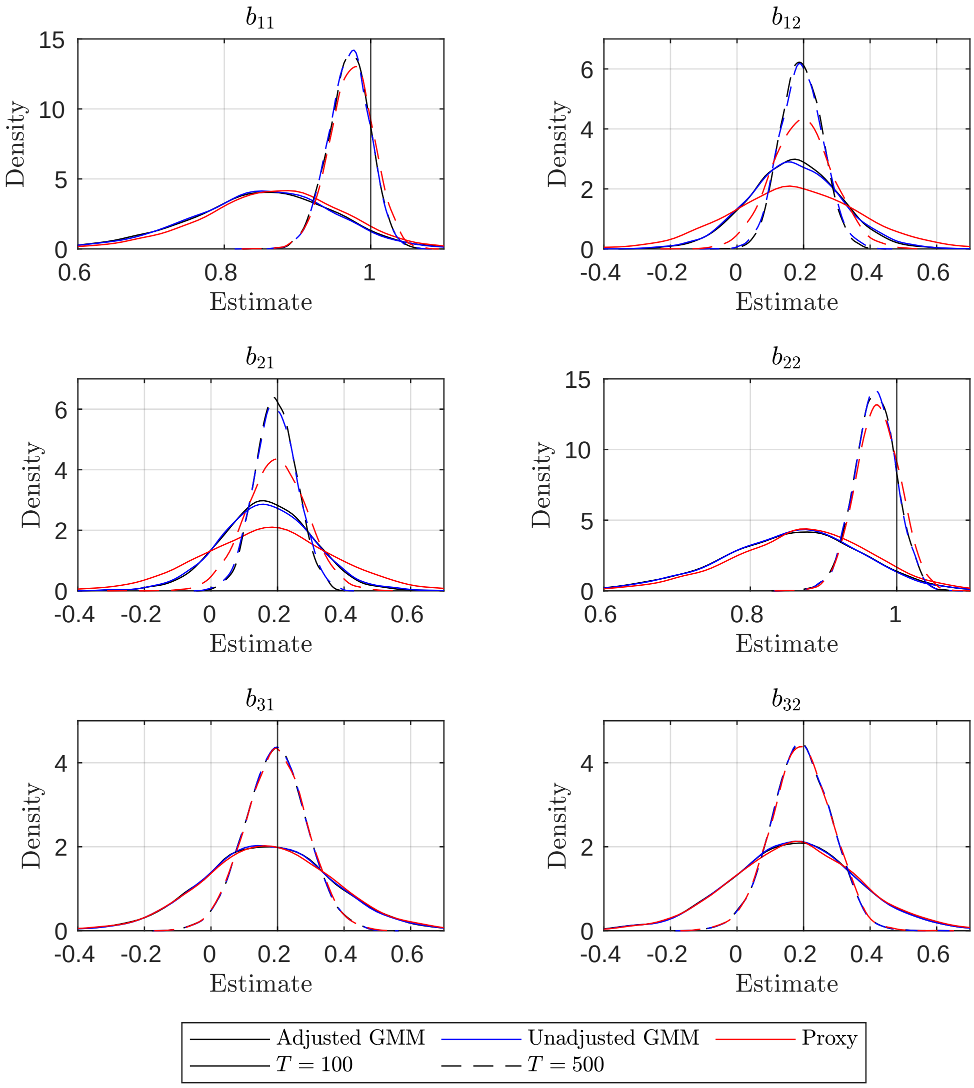

Figure 2: Monte Carlo Kernel Densities of Estimates of B_1 when $\sigma^2_{w_3} = 1$.

Table 1: Monte Carlo Rejection Frequencies of J-Test

		$T = 100$			$T = 500$		
		10%	5%	1%	10%	5%	1%
$\sigma^2_{w_3} = 0.01$	adjusted GMM	11.38%	5.72%	1.24%	11.22%	5.72%	1.32%
	unadjusted GMM	2.30%	0.68%	0.04%	1.84%	0.60%	0.02%
$\sigma^2_{w_3} = 1$	adjusted GMM	11.38%	5.80%	1.18%	11.22%	5.70%	1.22%
	unadjusted GMM	2.24%	0.72%	0.06%	1.8%	0.60%	0.02%

In summary, the simulation results suggest that the adjusted GMM procedure should be used in applied work because it generally dominates the conventional proxy VAR estimator in terms of small sample precision and it is crucial for taking full advantage of the inference possibilities that come with the GMM approach. In the following examples, we use the adjusted GMM procedure throughout.

4 Empirical Examples

There are a number of VAR studies where several shocks are identified by a set of proxies. If the proxies are instantaneously correlated, the authors typically use additional information to identify the shocks of interest individually. For example, Mertens and Ravn (2013) study the impact of tax shocks on the U.S. economy. They use two proxies to identify two shocks simultaneously and use an additional identifying zero restriction for the impact effects of the shocks to identify them individually. If the proxies individually satisfy the relevance and exogeneity conditions, it is tempting to use them one-by-one without additional, possibly controversial identifying restrictions. Clearly such an approach can be used if there are good reasons to assume that each proxy is only correlated with precisely one shock. As discussed in the previous section, using the conventional proxy VAR approach in this case may lead to correlated shocks. However, our GMM approach can be used to get around the problem of correlated shocks and to increase estimation efficiency. In the following we will present examples that illustrate the relevance of the issue for applied work.

Table 2: Empirical Correlations of Proxies for Altavilla el al. (2019) Example, Sample Period Jan 1, 2014 - Sep 13, 2018, with 95% Bootstrap Confidence Intervals

	z_t^{Target}	z_t^{Timing}	z_t^{FG}	z_t^{QE}
z_t^{Target}	1.000	0.094	0.253	0.147
		$(-0.255, 0.361)$	$(-0.368, 0.592)$	$(-0.389, 0.557)$
z_t^{Timing}		1.000	-0.558	0.428
			$(-0.830, -0.056)$	$(0.251, 0.601)$
z_t^{FG}			1.000	-0.015
				$(-0.324, 0.359)$
z_t^{QE}				1.000

4.1 European Monetary Policy Analysis

Altavilla et al. (2019) investigate the impact of monetary policy in the euro area (EA) by a proxy VAR model. They consider four shocks related to monetary policy to be identified by four proxies in a daily financial VAR model containing the following variables: the 2-year Overnight Index Swap rate (OIS2Y), the log EUR-USD exchange rate (EURUSD), the log Euro Stoxx 50 (STOXX), and the 2-year inflation linked swap (ILS2Y). They use their four proxies $z_t = (z_t^{Target}, z_t^{Timing}, z_t^{FG}, z_t^{QE})'$ one-by-one to identify four shocks, labelled Target, Timing, Forward Guidance (FG), and Quantitative Easing (QE) shocks. The Target shock is meant to capture conventional monetary policy action reflected in changes in the OIS2Y, the Timing and FG shocks are associated with central bank communication and capture short-term and medium term guidance, respectively. Finally, the QE shock is relevant for the time of unconventional monetary policy starting in 2014. It captures the longer term assessment of the economic situation by the central bank.

The overall sample considered by Altavilla et al. (2019) runs from January 2002 till September 2018. They conduct a subsample analysis for three subperiods. We only consider the last subperiod from January 2014 to September 2018 for which all the variables and all proxies are available. We perform our analysis based on VARs with 12 lags and an intercept term as in Altavilla et al. (2019).

The empirical correlations of the proxies for our sample period are presented in

Table 3: Empirical Correlations of Conventional Proxy VAR Shocks and Proxies for Altavilla el al. (2019) Example, Sample Period Jan 1, 2014 - Sep 13, 2018, with 95% Bootstrap Confidence Intervals

	$\hat{w}_t^{Target}(PVAR)$	$\hat{w}_t^{Timing}(PVAR)$	$\hat{w}_t^{FG}(PVAR)$	$\hat{w}_t^{QE}(PVAR)$
$\hat{w}_t^{Target}(PVAR)$	1.000	0.990	0.979	0.991
		$(0.985, 0.993)$	$(0.972, 0.985)$	$(0.987, 0.994)$
$\hat{w}_t^{Timing}(PVAR)$		1.000	0.959	0.992
			$(0.946, 0.970)$	$(0.989, 0.994)$
$\hat{w}_t^{FG}(PVAR)$			1.000	0.972
				$(0.962, 0.980)$
$\hat{w}_t^{QE}(PVAR)$				1.000

	$\hat{w}_t^{Target}(PVAR)$	$\hat{w}_t^{Timing}(PVAR)$	$\hat{w}_t^{FG}(PVAR)$	$\hat{w}_t^{QE}(PVAR)$
z_t^{Target}	0.379	0.375	0.370	0.375
	$(-0.040, 0.609)$	$(-0.038, 0.602)$	$(-0.044, 0.601)$	$(-0.042, 0.604)$
z_t^{Timing}	0.153	0.154	0.148	0.153
	$(-0.028, 0.337)$	$(-0.026, 0.339)$	$(-0.030, 0.333)$	$(-0.025, 0.336)$
z_t^{FG}	0.250	0.245	0.255	0.248
	$(0.043, 0.442)$	$(0.040, 0.435)$	$(0.049, 0.443)$	$(0.042, 0.438)$
z_t^{QE}	0.252	0.252	0.247	0.254
	$(0.046, 0.471)$	$(0.048, 0.469)$	$(0.047, 0.463)$	$(0.050, 0.470)$

Table 2 along with 95% bootstrap confidence intervals.[3] In the table it can be seen that the proxies are not uncorrelated. For example, z_t^{Timing} is significantly correlated with z_t^{FG} and z_t^{QE}. Although our earlier discussion suggests that correlation between the proxies is no problem as long as each of the proxies is correlated with a single shock only, many researcher would take correlated proxies as a warning signal that the shocks cannot be identified one-by-one using no further identifying information or additional assumptions.

We still follow Altavilla et al. (2019) and use the proxies one-by-one to identify the four shocks. The correlations between the estimated shocks and the proxies are presented in Table 3. Apparently, the resulting estimated shocks are highly correlated. The results in Table 3 also do not provide a case for the assumption of a diagonal $\Sigma_{\mathbf{w}_1 z}$ matrix because the FG and QE proxies are significantly correlated with all the

[3]The confidence intervals are generated by the bootstrap method presented in Appendix B.

Table 4: Empirical Correlations of GMM Shocks for Altavilla el al. (2019) Example, Sample Period Jan 1, 2014 - Sep 13, 2018, with 95% Bootstrap Confidence Intervals

	$\hat{w}_t^{Target}(GMM)$	$\hat{w}_t^{Timing}(GMM)$	$\hat{w}_t^{FG}(GMM)$	$\hat{w}_t^{QE}(GMM)$
$\hat{w}_t^{Target}(GMM)$	1.000	-0.031	-0.017	0.066
		$(-0.108, 0.042)$	$(-0.111, 0.076)$	$(-0.019, 0.156)$
$\hat{w}_t^{Timing}(GMM)$		1.000	0.003	0.003
			$(-0.111, 0.118)$	$(-0.096, 0.104)$
$\hat{w}_t^{FG}(GMM)$			1.000	0.026
				$(-0.095, 0.153)$
$\hat{w}_t^{QE}(GMM)$				1.000

	$\hat{w}_t^{Target}(GMM)$	$\hat{w}_t^{Timing}(GMM)$	$\hat{w}_t^{FG}(GMM)$	$\hat{w}_t^{QE}(GMM)$
z_t^{Target}	0.357	0.021	0.077	-0.062
	$(-0.047, 0.570)$	$(-0.034, 0.064)$	$(-0.042, 0.158)$	$(-0.119, 0.018)$
z_t^{Timing}	0.149	0.018	0.012	-0.020
	$(-0.015, 0.325)$	$(-0.021, 0.066)$	$(-0.023, 0.064)$	$(-0.108, 0.062)$
z_t^{FG}	0.220	0.050	0.080	-0.062
	$(0.032, 0.407)$	$(0.001, 0.093)$	$(0.018, 0.148)$	$(-0.117, 0.009)$
z_t^{QE}	0.242	0.037	0.047	-0.017
	$(0.045, 0.450)$	$(-0.020, 0.090)$	$(-0.034, 0.126)$	$(-0.094, 0.060)$

estimated shocks.

As, strictly speaking, the relevant $\Sigma_{\mathbf{w}_1 z}$ matrix to look at is the covariance matrix between the proxies and the orthogonal (uncorrelated) shocks, we have also determined shocks with our GMM procedure and present their correlations and correlations with the proxies in Table 4.[4] Not surprisingly, there is no significant correlation between the GMM shocks. Recall, however, that the GMM approach is based on the assumption of a diagonal $\Sigma_{\mathbf{w}_1 z}$ matrix which may not hold in this case. Therefore it is useful to take a look at the estimated $\Sigma_{\mathbf{w}_1 z}$ matrix for the GMM shocks also shown in Table 4. Clearly, that matrix does not look like a diagonal matrix. Three elements on the main diagonal are not even significantly different from zero while some off-diagonal elements are (e.g., the correlations between z_t^{FG} and $\hat{w}_t^{Target}(GMM)$ and $\hat{w}_t^{Timing}(GMM)$ as well as z_t^{QE} and $\hat{w}_t^{Target}(GMM)$). It is perhaps worth noting that

[4] The weighting matrix $\Omega(\beta, \hat{\gamma}^{LS})$ of equation (18) used in the GMM objective function (16) is chosen iteratively, using as stopping rule a relative change of the objective function of less than 5%.

the J-statistic takes a value of 36.77 which corresponds to a p-value of less than 1% of the relevant $\chi^2(6)$ distribution. Hence, the J-test clearly rejects a proper specification.

These results indicate that additional assumptions are necessary for identifying the four shocks properly as uncorrelated shocks based on the present proxies. Given the rather substantial correlation of the proxies, this outcome is perhaps not surprising.

4.2 U.S. Monetary Policy

Jarociński and Karadi (2020) investigate the impact of monetary policy in the U.S. and the EA. They consider two relevant shocks, a monetary policy shock which we denote by w_t^{mp} and a central bank information shock, denoted by w_t^{cbi} in the following. The first one captures conventional monetary policy action such as changes in interest rates, while w_t^{cbi} captures the impact of the assessment of the economic outlook conveyed by the central bank. Jarociński and Karadi (2020) construct different sets of proxies z_t^{mp} and z_t^{cbi} to identify the shocks. Furthermore, they use sign restrictions to properly identify their shocks of interest and Bayesian methods to perform their analysis. We will focus on one of their U.S. models, a specific pair of proxies and frequentist methods, thereby deviating from Jarociński and Karadi (2020), to illustrate some of the points we have made in Section 2.

Our model involves five U.S. variables, the one-year constant-maturity Treasury yield, log S&P 500, log real GDP, the log GDP deflator, and the excess bond premium (EBP) as a measure for the recession risk in the next 12 months. We use monthly data for the period 1984m2 - 2016m12 from Jarociński and Karadi (2020), where further details on their construction are provided. The model fitted is a VAR(12) with intercept term.

The two proxies are constructed as follows: A series of Federal Funds futures surprises at the time of FOMC announcements is constructed and aggregated to monthly frequency. That monthly series is split up in two proxies by taking into account S&P 500 changes. When the S&P 500 moves in opposite direction to the Fed Funds futures, the Fed Funds futures surprise is classified as a value of z_t^{mp}, while it

Table 5: Empirical Correlations of Proxies and Conventional Proxy VAR Shocks for Jarociński/Karadi (2020) Example with 95% Bootstrap Confidence Intervals

	$\hat{w}_t^{cbi}(PVAR)$	$\hat{w}_t^{mp}(PVAR)$
z_t^{cbi}	0.203 $(0.067, 0.299)$	0.089 $(-0.090, 0.242)$
z_t^{mp}	0.101 $(-0.018, 0.218)$	0.232 $(0.121, 0.335)$
$\hat{w}_t^{cbi}(PVAR)$	1	0.436 $(0.338, 0.534)$
$\hat{w}_t^{mp}(PVAR)$		1

Table 6: Empirical Correlations of Proxies and GMM Shocks for Jarociński/Karadi (2020) Example with 95% Bootstrap Confidence Intervals

	$\hat{w}_t^{cbi}(GMM)$	$\hat{w}_t^{mp}(GMM)$
z_t^{cbi}	0.172 $(-0.016, 0.338)$	0.068 $(-0, 115, 0.242)$
z_t^{mp}	0.023 $(-0.095, 0.138)$	0.230 $(0.123, 0.331)$
$\hat{w}_t^{cbi}(GMM)$	1	-0.025 $(-0.163, 0.124)$
$\hat{w}_t^{mp}(GMM)$		1

is classified as a value of z_t^{cbi} for all other periods. For all periods where no value is available, the proxies are set to zero. Thus, $z_t^{mp} z_t^{cbi} = 0$ by construction and, as the proxies have nonzero means, their correlation is nonzero but small by construction.

In Table 5, we show the empirical correlations between the shocks and the proxies when the shocks are estimated using the conventional proxy VAR approach. The correlation between z_t^{cbi} and $\hat{w}_t^{mp}(PVAR)$ and between z_t^{mp} and $\hat{w}_t^{cbi}(PVAR)$ is small and not significantly different from zero. Given that the estimated $\Sigma_{\mathbf{w}_1 z}$ matrix is thus nearly diagonal, one may conclude that identifying the two shocks one-by-one may be justified. However, the resulting shocks have significant correlation as high as 0.436 if the conventional proxy VAR approach is used for estimating them.

As it may be reasonable to assume in the present case that each of the proxies is only correlated with a single shock ($\Sigma_{\mathbf{w}_1 z}$ is diagonal), we can use our GMM procedure to obtain uncorrelated shocks without having to worry about further identifying information. In Figure 3, the shocks obtained by GMM with a 2-step weighting matrix $\Omega(\beta, \hat{\gamma}^{LS})$ in the objective function (16) are plotted in a scatter diagram against the shocks obtained with the conventional proxy VAR approach. Obviously, the shocks are similar but not identical, $\hat{w}_t^{cbi}(PVAR)$ and $\hat{w}_t^{cbi}(GMM)$ being not quite as similar as $\hat{w}_t^{mp}(PVAR)$ and $\hat{w}_t^{mp}(GMM)$. In any case, corresponding shocks appear to be highly positively correlated.

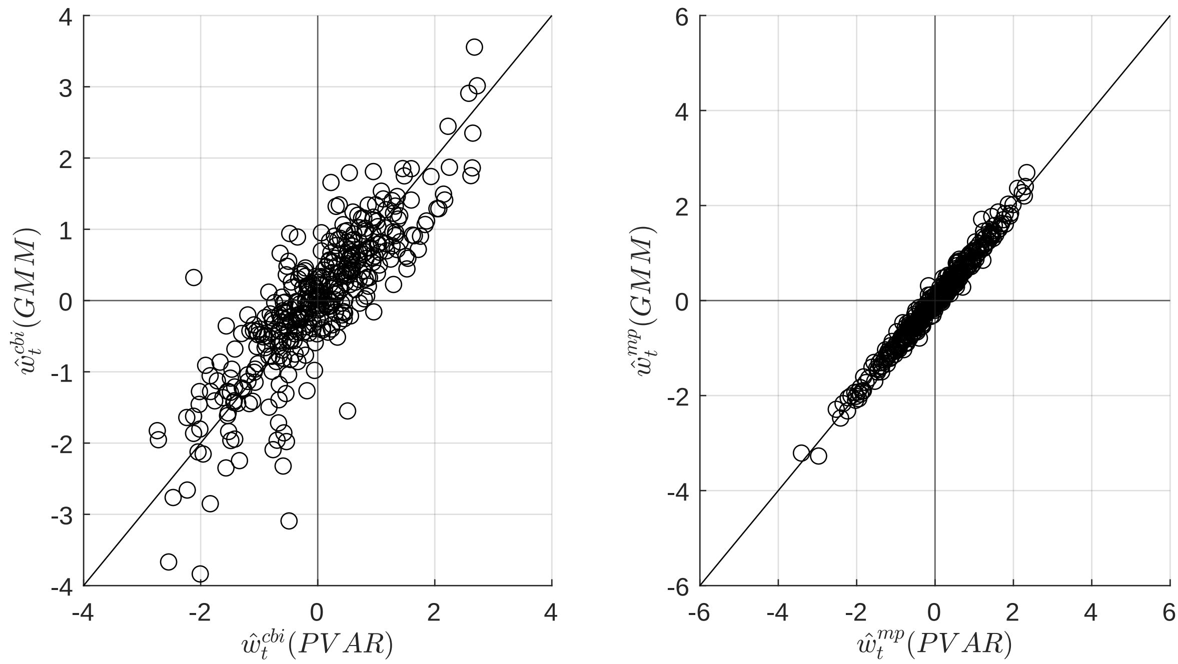

Figure 3: U.S. monetary policy example: Scatter plots of shocks obtained by the conventional proxy VAR and GMM approaches.

The correlations of the GMM shocks and their correlations with the proxies are presented in Table 6. In this case, the empirical correlation between the estimated $\hat{w}_t^{cbi}(GMM)$ and $\hat{w}_t^{mp}(GMM)$ is -0.025 and, hence, very small and not significantly different from zero. The fact that the estimated correlation matrix corresponding to $\Sigma_{\mathbf{w}_1 z}$ based on the conventional proxy VAR shocks in Table 5 is diagonal is, of course, no insurance for getting also a diagonal $\Sigma_{\mathbf{w}_1 z}$ for the GMM shocks. Thus, we also present the estimated correlation matrix corresponding to $\Sigma_{\mathbf{w}_1 z}$ for these shocks in

Table 6 and find that the assumption of a diagonal $\Sigma_{\mathbf{w}_1 z}$ matrix is supported by the very small and insignificant off-diagonal elements of the estimated correlation matrix. Thereby we also support the use of our GMM approach for estimation in this case. We have also performed the J-test and obtained a test value of $J = 1.3462$ and a p-value of 0.25 of the corresponding $\chi^2(1)$ distribution which indicates that our test provides no evidence against the moment conditions being correct.

Given that an argument against using correlated shocks is that the corresponding impulse responses may reflect a distorted picture of the actual responses of the variables as isolated shocks are not likely to occur in practice, we present the impulse responses obtained with the conventional proxy VAR approach and the GMM approach in Figures 4 and 5. The shocks underlying the impulse responses in the figures are scaled such that they increase the interest rate by 25 basis points on impact to make them comparable in size despite having potentially different variances.[5] The confidence intervals around the impulse responses in the figures are computed by a moving block bootstrap. That method was proposed by Jentsch and Lunsford (2019) and produces intervals that are robust to conditional heteroskedasticity. We use exactly the implementation described in Bruns and Lütkepohl (2023).

In the two figures, the conventional proxy VAR impulse responses with 68% confidence intervals are shown on the left-hand side of the figure and the GMM confidence intervals are presented on the right-hand side. The point estimates obtained with both approaches are shown on the left-hand as well as the right-hand side to facilitate the comparison. Of course, given the similarity of the shocks in Figure 3, one would also expect the impulse responses to be similar as well. This is actually reflected in Figures 4 and 5. However, it may still be worth pointing out some important differences.

The responses to $\hat{w}_t^{cbi}(PVAR)$ and $\hat{w}_t^{cbi}(GMM)$ shocks are compared in Figure 4. Overall the confidence intervals of the GMM approach are somewhat smaller than

[5] For the scaling, a variable has to be chosen that has impact effects well away from zero for shocks estimated with either approach to avoid undesirable effects on the bootstrap confidence intervals (see Lütkepohl (2013)).

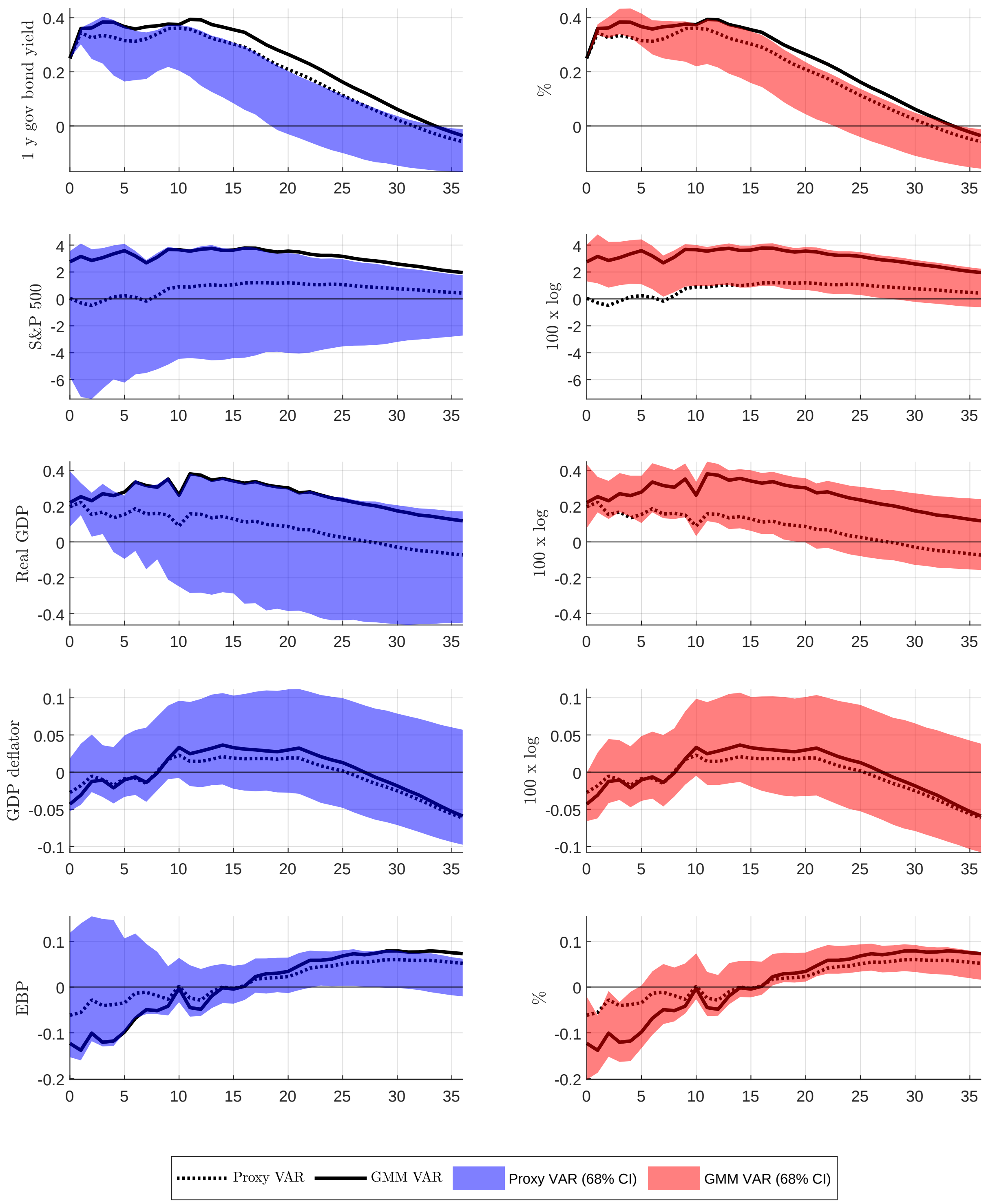

Figure 4: U.S. monetary policy example: Comparison of impulse responses of a w_t^{cbi} shock estimated by the conventional proxy VAR approach (dotted lines, blue confidence intervals) and with the GMM approach (solid lines, red confidence intervals). The impulse responses are normalized to yield a 25 basis points increase in the one-year government bond yield on impact. The confidence intervals around the impulse responses are based on 5000 bootstrap samples.

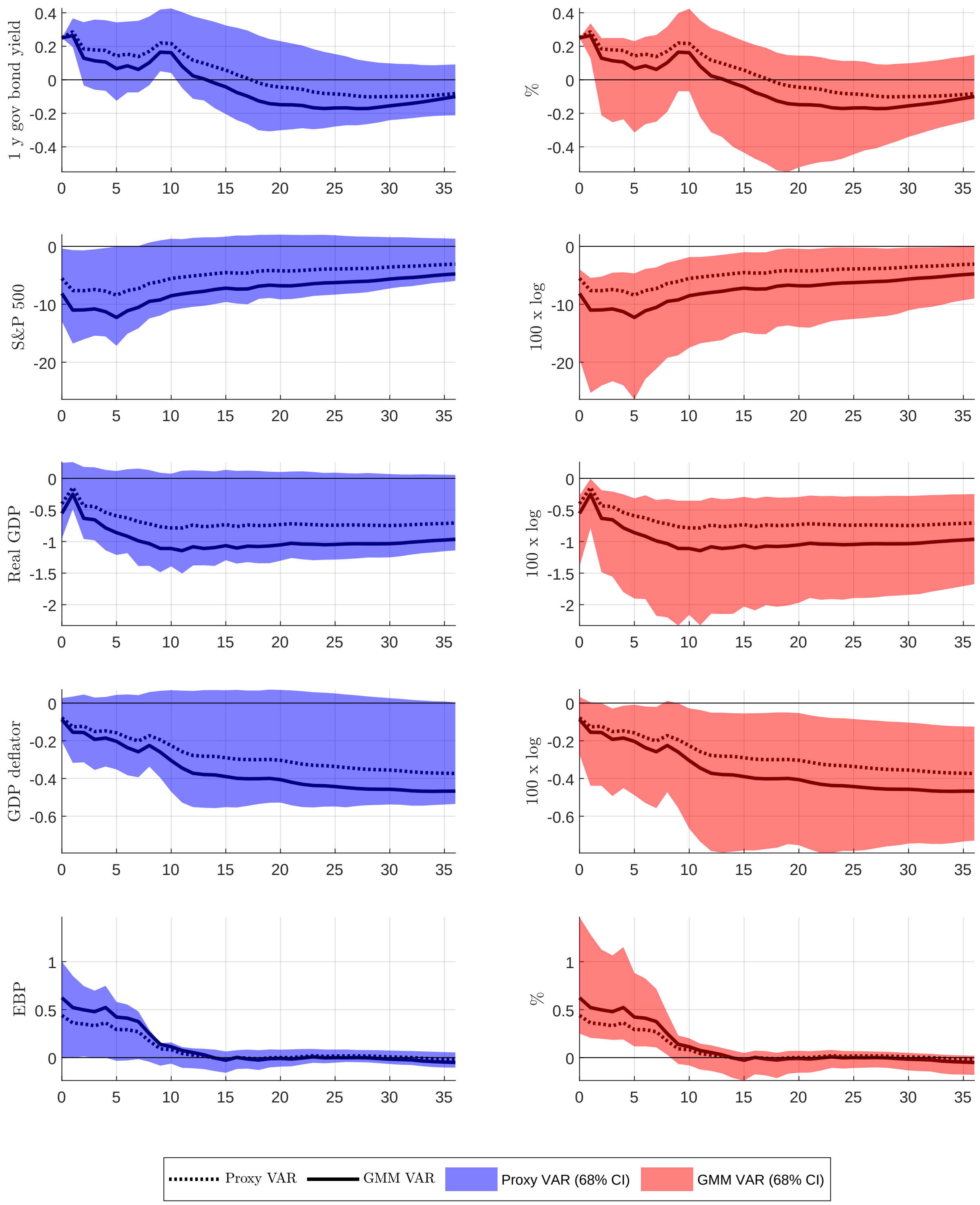

Figure 5: U.S. monetary policy example: Comparison of impulse responses of a w_t^{mp} shock estimated by the conventional proxy VAR approach (dotted lines, blue confidence intervals) and with the GMM approach (solid lines, red confidence intervals). The impulse responses are normalized to yield a 25 basis points increase in the one-year government bond yield on impact. The confidence intervals around the impulse responses are based on 5000 bootstrap samples.

the corresponding conventional proxy VAR intervals. Most point estimates of one approach are covered by the confidence intervals of the other approach. However, the responses of the stock index estimated by the conventional proxy VAR approach for the first 12 months after the shock do not fall inside the GMM confidence intervals and are much closer to zero than the responses estimated by GMM. In other words, for the orthogonalized central bank information shock the impact effect is estimated to be much stronger on the stock market than indicated by the correlated shock. It is in fact quite plausible that the information released by the central bank is closely monitored by the stock market participants and, hence, the stock market response estimated by the GMM approach may be the more realistic one. Another striking difference is the response of EBP to the GMM shock compared to the conventionally estimated shock. The confidence interval of the EBP response on impact to a GMM central bank information shock does not cover zero and, hence, one may conclude that the central bank can successfully reduce the risk of a recession by its communication despite increasing the interest rate. In contrast, relying on the conventional proxy VAR approach, zero is well inside the confidence intervals of propagation horizons of up to more than one year. Hence, in this case, one may underestimate the impact of the central bank communication shocks when using the correlated shocks.

The responses to $\hat{w}_t^{mp}(PVAR)$ and $\hat{w}_t^{mp}(GMM)$ shocks presented in Figure 5 are again rather similar. In this case, the confidence intervals produced by the GMM approach are partly wider than the corresponding conventional proxy VAR intervals. Even in this case, considering the point estimates, a 25 basis points interest rate shock is estimated by the GMM approach to have a stronger impact on the S&P 500, and the GDP deflator than the shock estimated by the conventional proxy VAR approach. The real GDP and EBP responses to a GMM shock are significant on impact while the conventional approach estimates insignificant impact effects. Clearly, it is not implausible that an interest rate hike lowers real GDP and increases the risk of a recession if the marginal effect of increasing the interest rate is clearly separated from the central bank communication that goes along with it. In other words, the orthogonal GMM shock provides the more plausible impulse responses. Thus, considering uncorrelated shocks makes a difference.

We emphasize again that Jarociński and Karadi (2020) use quite different identification and estimation methods. Therefore it is not surprising that their impulse responses differ from those obtained by our estimation approaches. We have deviated from their analysis to illustrate some of the theoretical points made in Section 2 of our paper.

5 Conclusions

This study shows that using proxies to identify structural shocks in a VAR analysis can lead to unintentionally correlated shocks. Such shocks are usually ruled out by assumption in structural VAR analysis because correlated shocks may lead to distorted impulse responses. When several proxies are used to identify a set of shocks and no further identifying information is available, in general, the proxies will identify only linear combinations of the impact effects and, hence, the shocks. To avoid having to worry about additional identifying information, it is desirable to use the proxies one-by-one if they satisfy the usual relevance and exogeneity conditions for valid proxies individually. If each proxy is correlated with exactly one shock only, that approach is in fact theoretically sound. Using that feature and also imposing uncorrelatedness of the shocks implies even over-identifying restrictions for the shocks of interest. We have proposed a simple, efficient GMM approach that takes full advantage of the over-identifying restrictions and ensures uncorrelated shocks if each proxy is correlated with exactly one shock only.

We present examples of structural VAR studies that use multiple proxies to identify more than one structural shock and where the structural shocks are not instantaneously uncorrelated if the proxies are used in the conventional way to identify the shocks. Enforcing uncorrelated shocks by using our GMM approach makes a difference for the impulse responses. Thereby we show that the problem of correlated shocks is relevant in practice.

Proxies have also been used to identify structural shocks in factor models (see, e.g., Stock and Watson (2012)). Obviously, correlated shocks may also be obtained in that setting if the proxies are used one-by-one. Therefore, extending our analysis in that direction would be of interest. We leave it for future research because identifying structural shocks in factor models involves additional considerations (see, e.g., Kilian and Lütkepohl (2017, Chapter 16)).

Appendix

A Derivation of the GMM Correction Term

To set up an efficient GMM procedure for the parameters of interest, β, we need the derivatives considered in the GMM weighting matrix in expressions (18) and (19). We will derive closed-form expressions for these derivatives first.

A.1 Derivatives

$$\frac{\partial m_t^\beta(\beta,\gamma)}{\partial\gamma'} = \begin{bmatrix} \dfrac{\partial \mathrm{vec}(u_t(\alpha)z_t' - B_1)}{\partial\gamma'} \\ \\ \dfrac{\partial \mathrm{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial\gamma'} \end{bmatrix}_{\left((KK_1+\frac{1}{2}K_1(K_1-1))\times(K(Kp+1)+\frac{1}{2}K(K+1))\right)}$$

where

$$\frac{\partial \mathrm{vec}(u_t(\alpha)z_t' - B_1)}{\partial\gamma'} = (z_t \otimes I_K)\frac{\partial u_t(\alpha)}{\partial\gamma'} = -(z_t \otimes I_K)(Y_{t-1}' \otimes I_K)\frac{\partial\alpha}{\partial\gamma'}$$

$$= -(z_t Y_{t-1}' \otimes I_K)[I_{K(Kp+1)} : 0_{K(Kp+1)\times\frac{1}{2}K(K+1)}]_{\left(K_1K\times(K(Kp+1)+\frac{1}{2}K(K+1))\right)}$$

and

$$\frac{\partial \mathrm{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial\alpha'} = \mathbf{S}_{K_1}\frac{\partial \mathrm{vec}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial\alpha'}$$

$$= \mathbf{S}_{K_1}(B_1'\Sigma_u^{-1} \otimes B_1'\Sigma_u^{-1})\frac{\partial \text{vec}(u_t(\alpha)u_t(\alpha)')}{\partial \alpha'} \quad \left(\frac{1}{2}K_1(K_1-1) \times K(Kp+1)\right)$$

$$= \mathbf{S}_{K_1}(B_1'\Sigma_u^{-1} \otimes B_1'\Sigma_u^{-1})(I_{K^2} + \mathbf{K}_{KK})(u_t(\alpha) \otimes I_K)\frac{\partial u_t(\alpha)}{\partial \alpha'}$$

$$= -\mathbf{S}_{K_1}(B_1'\Sigma_u^{-1} \otimes B_1'\Sigma_u^{-1})(I_{K^2} + \mathbf{K}_{KK})(u_t(\alpha)Y_{t-1}' \otimes I_K)$$

(see Rule (2)(b) from Lütkepohl (1996, Section 10.5.1)). Moreover,

$$\frac{\partial \text{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial \sigma'} = \mathbf{S}_{K_1}\frac{\partial \text{vec}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial \text{vec}(\Sigma_u)'}\frac{\partial \text{vec}(\Sigma_u)}{\partial \text{vech}(\Sigma_u)'}$$

$$= -\mathbf{S}_{K_1}\left(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1} \otimes B_1'\Sigma_u^{-1} + B_1'\Sigma_u^{-1} \otimes B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}\right)\mathbf{D}_K,$$

where Rule (5)(c) from Lütkepohl (1996, Section 10.6.1) has been used. Obviously,

$$\frac{\partial \text{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial \gamma'}$$

$$= \left[\frac{\partial \text{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial \alpha'} : \frac{\partial \text{vh}(B_1'\Sigma_u^{-1}u_t(\alpha)u_t(\alpha)'\Sigma_u^{-1}B_1)}{\partial \sigma'}\right].$$

Furthermore,

$$\frac{\partial m_t^\gamma(\gamma)}{\partial \gamma'} = \begin{bmatrix} -\dfrac{\partial (Y_{t-1}Y_{t-1}' \otimes I_K)\alpha}{\partial \gamma'} \\ \\ \dfrac{\partial \text{vech}(\Sigma_u - u_t(\alpha)u_t(\alpha)')}{\partial \gamma'} \end{bmatrix}_{\left((K(Kp+1)+\frac{1}{2}K(K+1))\times(K(Kp+1)+\frac{1}{2}K(K+1))\right)}$$

where

$$-\frac{\partial (Y_{t-1}Y_{t-1}' \otimes I_K)\alpha}{\partial \gamma'} = \left[-(Y_{t-1}Y_{t-1}' \otimes I_K) : 0_{K(Kp+1)\times\frac{1}{2}K(K+1)}\right]$$

and

$$\frac{\partial \text{vech}(\Sigma_u - u_t(\alpha)u_t(\alpha)')}{\partial \gamma'} = \left[-\frac{\partial \text{vech}[u_t(\alpha)u_t(\alpha)']}{\partial \alpha'} : \frac{\partial \text{vech}(\Sigma_u)}{\partial \sigma'}\right]$$

with

$$\frac{\partial \text{vech}[u_t(\alpha)u_t(\alpha)']}{\partial \alpha'} = \mathbf{L}_K \frac{\partial \text{vec}[u_t(\alpha)u_t(\alpha)']}{\partial \alpha'} = -\mathbf{L}_K(I_{K^2} + \mathbf{K}_{KK})(u_t(\alpha)Y_{t-1}' \otimes I_K)$$

(see Rule (2)(b) from Lütkepohl (1996, Section 10.5.1)) and

$$\frac{\partial \text{vech}(\Sigma_u)}{\partial \sigma'} = I_{\frac{1}{2}K(K+1)}.$$

A.2 Correction Term

The correction term for computing the GMM weighting matrix is of the form

$$-\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial m_t^{\beta}(\beta,\gamma)}{\partial \gamma'}\right)\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial m_t^{\gamma}(\gamma)}{\partial \gamma'}\right)^{-1} m_t^{\gamma}(\gamma).$$

Using the previously derived derivatives as well as the rules for the partitioned inverse and the inverse of a Kronecker product, we get

$$\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial m_t^{\gamma}(\gamma)}{\partial \gamma'}\right)^{-1}$$

$$= \begin{bmatrix} -(\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}Y_{t-1}')^{-1} \otimes I_K & 0_{K(Kp+1)\times\frac{1}{2}K(K+1)} \\ -\mathbf{L}_K(I_{K^2}+\mathbf{K}_{KK})\left((\frac{1}{T}\sum_{t=1}^{T} u_t(\alpha)Y_{t-1}')(\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}Y_{t-1}')^{-1} \otimes I_K\right) & I_{\frac{1}{2}K(K+1)} \end{bmatrix}.$$

Thus, evaluating the derivatives at the LS estimator $\hat{\gamma}^{LS}$ for γ, denoting the LS residuals by $\hat{u}_t$ and using $\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t Y_{t-1}' = 0$, gives

$$-\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial m_t^{\beta}(\beta,\hat{\gamma}^{LS})}{\partial \gamma'}\right)\left(\frac{1}{T}\sum_{t=1}^{T} \frac{\partial m_t^{\gamma}(\hat{\gamma}^{LS})}{\partial \gamma'}\right)^{-1}$$

$$= \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} z_t Y_{t-1}' \otimes I_K & 0_{KK_1\times\frac{1}{2}K(K+1)} \\ 0_{\frac{1}{2}K_1(K_1-1)\times K(Kp+1)} & 2\mathbf{S}_{K_1}(B_1'\widehat{\Sigma}_u^{-1} \otimes B_1'\widehat{\Sigma}_u^{-1})\mathbf{D}_K \end{bmatrix}$$

$$\times \begin{bmatrix} -(\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}Y_{t-1}')^{-1} \otimes I_K & 0_{K(Kp+1)\times\frac{1}{2}K(K+1)} \\ \\ 0_{\frac{1}{2}K(K+1)\times K(Kp+1)} & I_{\frac{1}{2}K(K+1)} \end{bmatrix}$$

$$= \begin{bmatrix} -(\frac{1}{T}\sum_{t=1}^{T} z_t Y_{t-1}')(\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}Y_{t-1}')^{-1} \otimes I_K & 0_{KK_1\times\frac{1}{2}K(K+1)} \\ \\ 0_{\frac{1}{2}K_1(K_1-1)\times K(Kp+1)} & 2\mathbf{S}_{K_1}(B_1'\widehat{\Sigma}_u^{-1} \otimes B_1'\widehat{\Sigma}_u^{-1})\mathbf{D}_K \end{bmatrix}$$

Right-multiplying the latter matrix by

$$m_t^{\gamma}(\hat{\gamma}^{LS}) = \begin{bmatrix} (Y_{t-1} \otimes I_K)y_t - (Y_{t-1}Y_{t-1}' \otimes I_K)\hat{\alpha}^{LS} \\ \\ \text{vech}(\widehat{\Sigma}_u - \hat{u}_t\hat{u}_t') \end{bmatrix} = \begin{bmatrix} (Y_{t-1} \otimes I_K)\hat{u}_t \\ \\ \text{vech}(\widehat{\Sigma}_u - \hat{u}_t\hat{u}_t') \end{bmatrix}$$

gives the correction term in expression (20).

B Construction of Bootstrap Intervals for Correlations

To construct the correlation confidence intervals shown in Tables 2 - 6, an i.i.d. bootstrap with 10,000 repetitions as in Lunsford (2015) proceeds as follows:

1. Draw T times with replacement from $\begin{bmatrix} z_t \\ w_t \end{bmatrix}$ to generate a bootstrap sample $\begin{bmatrix} z_{t,n}^* \\ w_{t,n}^* \end{bmatrix}$, $t = 1, \ldots, T$.

2. Compute $R_n = (\rho_{ij,n}) = \text{corr}(\begin{bmatrix} z_{t,n}^* \\ w_{t,n}^* \end{bmatrix})$ of dimension $(2K_1 \times 2K_1)$.

3. Repeat steps 1. - 2. 10,000 times and store $(R_1, \ldots, R_{10{,}000})$.

4. Compute the 2.5th and 97.5th percentile $CI_{\rho_{ij},2.5}$ and $CI_{\rho_{ij},97.5}$ for each $\rho_{ij}, i < j$.

References

Altavilla, C., Brugnolini, L., Gürkaynak, R. S., Motto, R. and Ragusa, G. (2019). Measuring euro area monetary policy, *Journal of Monetary Economics* **108**: 162–179.

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.

Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2021). Inference in Bayesian Proxy-SVARs, *Journal of Econometrics* **225**: 88 – 106.

Boer, L. and Lütkepohl, H. (2021). Qualitative versus quantitative external information for proxy vector autoregressive analysis, *Journal of Economic Dynamics and Control* **127**: 104118.

Bruns, M. and Lütkepohl, H. (2022). Comparison of local projection estimators for proxy vector autoregressions, *Journal of Economic Dynamics & Control* **134**: 104277.

Bruns, M. and Lütkepohl, H. (2023). An alternative bootstrap for proxy vector autoregressions, *Computational Economics* **62**: 1857–1882.

Crepon, B., Kramarz, F. and Trognon, A. (1997). Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error component models, *Journal of Econometrics* **82**: 135–156.

Fanelli, L. and Marsi, A. (2022). Sovereign spreads and unconventional monetary policy in the Euro area: A tale of three shocks, *European Economic Review* **150**: 104281.

Garchi Casal, S. and Zimic, S. (2023). A multivariate proxy identification of SVARs with plausibly exogenous instruments, mimeo.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal: Macroeconomics* **7**(1): 44–76.

Giacomini, R., Kitagawa, T. and Read, M. (2022). Robust Bayesian inference in proxy SVARs, *Journal of Econometrics* **228**: 107–126.

Gregory, A. W., McNeil, J. and Smith, G. W. (2024). US fiscal policy shocks: Proxy-SVAR overidentification via GMM, *Journal of Applied Econometrics* .

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises - the role of information shocks, *American Economic Journal: Macroeconomics* **12**: 1–43.

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.

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements, *American Economic Review* **111**(4): 1092–1125.

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

Lakdawala, A. (2019). Decomposing the effects of monetary policy using an external instruments SVAR, *Journal of Applied Econometrics* **34**: 934–950.

Lunsford, K. (2015). Identifying structural VARs with a proxy variable and a test for a weak proxy, *Technical report*, Federal Reserve Bank of Cleveland.

Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley & Sons, Chichester.

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

Lütkepohl, H. (2013). Reducing confidence bands for simulated impulse responses, *Statistical Papers* **54**: 1131–1145.

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.

Piffer, M. and Podstawski, M. (2017). Identifying uncertainty shocks using the price of gold, *The Economic Journal* **128**(616): 3266–3284.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation, *in* J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2, Elsevier, chapter 2, pp. 71–162.

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.

Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound?, *Economic Journal* **122**: F447–F466.