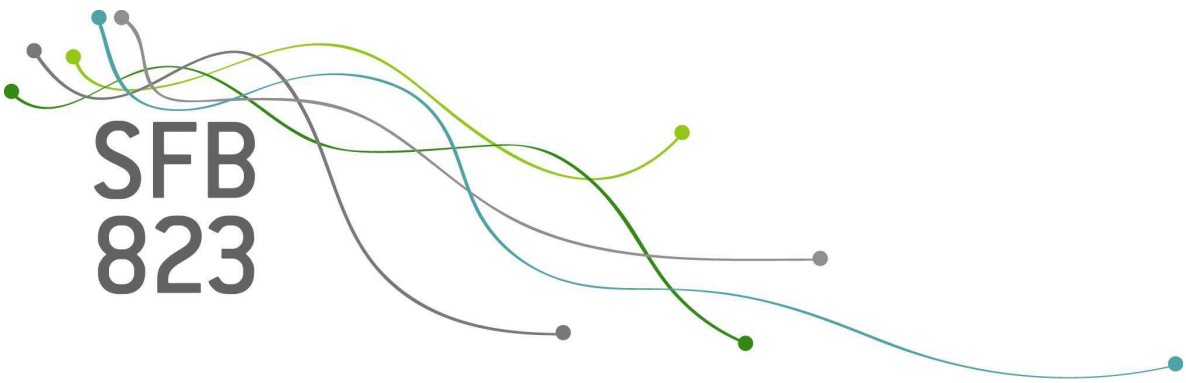
SFB 823

Efficiency gains in structural vector autoregressions by selecting informative higher-order moment conditions

Sascha Alexander Keweloh,
Stephan Hetzenecker

Nr. 26/2021

Discussion Paper



Efficiency Gains in Structural Vector Autoregressions by Selecting Informative Higher-Order Moment Conditions

Sascha Alexander Keweloh* Stephan Hetzenecker**

November 24, 2021

This study combines block-recursive restrictions with non-Gaussian and mean independent shocks to derive identifying and overidentifying higher-order moment conditions for structural vector autoregressions. We show that overidentifying higher-order moments can contain additional information and increase the efficiency of the estimation. In particular, we prove that in the non-Gaussian recursive SVAR higher-order moment conditions are relevant and therefore, the frequently applied estimator based on the Cholesky decomposition is inefficient. Even though incorporating information in valid higher-order moments is asymptotically efficient, including many redundant and potentially even invalid moment conditions renders standard SVAR GMM estimators unreliable in finite samples. We apply a LASSO-type GMM estimator to select the relevant and valid higher-order moment conditions, increasing finite sample precision. A Monte Carlo experiment and an application to quarterly U.S. data illustrate the improved performance of the proposed estimator.

JEL Codes: C32, E52, E44

Keywords: SVAR, efficiency, non-Gaussianity, block-recursive, LASSO, monetary policy

* TU Dortmund University / RGS Econ, E-Mail: sascha.keweloh@tu-dortmund.de, Postal Address: TU Dortmund, Department of Economics, D-44221 Dortmund. Financial support from the German Science Foundation, DFG - SFB 823, is gratefully acknowledged.

** University Duisburg-Essen / RGS Econ, E-Mail: stephan.hetzenecker@vwl.uni-due.de, Postal Address: University Duisburg-Essen, Universitätsstraße 12, D-45117 Essen. Financial support by the Ruhr Graduate School in Economics is gratefully acknowledged.

1 Introduction

Identification of structural vector autoregressions (SVARs) requires to assume an a priori structure of the model. Traditionally, identification is based on imposing structure on the interaction of the variables, ideally derived from macroeconomic theory (e.g., short-run restrictions Sims (1980) or long-run restrictions Blanchard and Quah (1993)). However, theoretical restrictions are rare and oftentimes debatable. More recently, data-driven approaches allow to identify the SVAR without imposing any restrictions on the interaction. Instead, identification is achieved by imposing structure on the stochastic properties of the shocks (e.g., time-varying volatility as discussed in Rigobon (2003), Lanne et al. (2010), Lütkepohl and Netšunajev (2017), and Lewis (2021) or non-Gaussian and independent shocks as discussed in Gouriéroux et al. (2017), Lanne et al. (2017), Lanne and Luoto (2021) Keweloh (2021b), and Guay (2021)).

Guay (2021) and Keweloh et al. (2021) combine the restriction and non-Gaussian identification approaches. Guay (2021) proposes a test to determine which part of the SVAR is identified based on non-Gaussianity and solely uses restrictions to identify the remaining part of the SVAR. Therefore, the identification approach relies as heavily on non-Gaussianity as possible and as little on restrictions as necessary. In contrast to that, Keweloh et al. (2021) put forward an approach which relies as heavily on restrictions as possible and as little on non-Gaussianity as necessary. In particular, their approach imposes a block-recursive order, meaning shocks in a given block only influence variables in the same block or blocks ordered below. Identification of the shocks within a certain block is based on higher-order moment conditions derived from mean independent shocks in the given block. However, the impact of the shocks in one block on variables in another block is identified based only on second-order moments and only requires uncorrelated shocks between the blocks. If each block contains exactly one shock such that the SVAR is recursive, the proposed estimator is equal to the recursive estimator obtained by applying the Cholesky decomposition. If there is only one block containing all shocks, the proposed estimator is equal to the unrestricted SVAR GMM estimators based on higher-order moment conditions proposed by Lanne and Luoto (2021) or Keweloh (2021b).

If identification is the only concern, relying on restrictions appears unnecessarily restrictive in a non-Gaussian SVAR with independent shocks. That said, Keweloh et al. (2021) demonstrate that purely data-driven estimators perform poorly in small samples and that restrictions are required to achieve decent small sample performance.

We extend the setting of Keweloh et al. (2021) who derive a conservative set of assumptions on the dependence and Gaussianity of the shocks, ensuring identification for a block-recursive structure. However, imposing more structure on the dependence of the shocks than required for identification also yields overidentifying moment conditions.

Our first contribution is to show that these overidentifying conditions can contain relevant conditions which increase the efficiency of the estimator. More precisely, we prove that in the recursive SVAR some higher-order moment conditions derived from mean independent structural shocks decrease the asymptotic variance if at least one shock has non-zero skewness or non-zero excess kurtosis.[1] Therefore, in the recursive SVAR with non-Gaussian and mean independent shocks the frequently used estimator obtained by applying the Cholesky decomposition is inefficient. Moreover, we show that some higher-order moment conditions are redundant, meaning they do not decrease the asymptotic variance of the estimator. Redundant moment conditions have no impact on the asymptotic variance of the estimator, however, including many of them can distort finite sample performance.

Our second contribution is to use the LASSO-type GMM estimator proposed by Cheng and Liao (2015a), referred to as the penalized GMM estimator (pGMM), to distinguish between redundant and relevant as well as valid and invalid overidentifying conditions in a data-driven way. That is, we apply the pGMM to select only the relevant and valid moment conditions. The pGMM estimator is asymptotically normal and as efficient as the asymptotically efficent block-recursive SVAR GMM estimator, including all higher-order moment conditions. Our Monte Carlo simulation illustrates the pGMM estimator's ability to successfully ignore redundant moments. The finite

[1]Note that this is not trivial. For example, in a linear regression model $y_t = \beta_1 x_t + \epsilon_t$ the GMM estimator with the moment condition $E[x_t \epsilon_t] = 0$ is identified and efficient under (conditional) homoscedastic errors. Therefore, including additional higher-order moment conditions like $E[x_t^2 \epsilon_t] = 0$ does not decrease the asymptotic variance of the GMM estimator even if the shocks or variables are non-Gaussian.

sample performance of the pGMM estimator is good in terms of bias and finite sample variance in small and large samples, which is not the case for the considered alternatives. For instance, the Cholesky estimator, which relies only on second-order moment conditions, performs well in small but not in large samples. The asymptotically efficent recursive SVAR GMM estimator including all higher-order moment conditions performs well in large but not in small samples. Thus, the pGMM solves finite sample problems induced by many (potentially redundant) moments.

The remainder of the paper is organized as follows: Section 2 reviews the SVAR and different identification schemes. Section 3 introduces the block-recursive SVAR, derives identifying and overidentifying moment conditions, analyzes which of the overidentifying moment are redundant or relevant in a recursive SVAR, and describes the pGMM estimator. In Section 4, we present a Monte Carlo experiment that investigates the performance of the pGMM estimator. Section 5 applies the pGMM estimator to analyze the impact of supply, demand, cost-push, stock market and monetary policy shocks on the U.S. economy. Section 6 concludes.

2 Overview SVAR

This section briefly explains the identification problem and common identification approaches of SVAR models. A detailed overview can be found in Kilian and Lütkepohl (2017). Consider the SVAR

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + B_0 \varepsilon_t, \tag{1}$$

with parameter matrices $A_1, ..., A_p \in \mathbb{R}^{n \times n}$, an invertible matrix B_0, an n-dimensional vector of time series $y_t = [y_{1,t}, ..., y_{n,t}]'$ and an n-dimensional vector of i.i.d. structural shocks $\varepsilon_t = [\varepsilon_{1,t}, ..., \varepsilon_{n,t}]'$ with mean zero and unit variance.

W.l.o.g., we simplify and omit the lag terms in Equation (1). That is, we focus on the simultaneous

interaction of the SVAR given by

$$u_t = B_0 \varepsilon_t, \tag{2}$$

with the reduced form shocks $u_t = y_t - A_1 y_{t-1} - ... - A_p y_{t-p}$, which can be estimated consistently by OLS. The reduced form shocks are equal to an unknown mixture B_0 of the unknown structural shocks ε_t. So far, neither the mixing matrix B_0 nor the structural shocks ε_t are identified. To see this, define the unmixed innovations $e(B)$ as the innovations obtained by unmixing the reduced form shocks with some matrix B

$$e_t(B) := B^{-1} u_t. \tag{3}$$

Note that for $B = B_0$, the unmixed innovations are equal to the structural shocks ε_t, i.e., $e_t(B_0) = \varepsilon_t$. However, the structural shocks ε_t and the mixing matrix B_0 are unknown and without imposing further structure, there is no criterion to verify, whether our mixing matrix B and our unmixed innovations $e_t(B)$ are equal to the true mixing matrix B_0 and the structural shocks ε_t.

To identify B_0 and the shocks ε_t, the researcher has to impose structure on the SVAR. The structure can be specified in two dimensions: We may

(i) impose more structure on the interaction of the shocks (see, Sims (1980) for short-run restrictions or Blanchard (1989) for long-run restrictions), or

(ii) impose more structure on the stochastic properties of the structural shock (see, Lanne et al. (2010) for time-varying volatility or Gouriéroux et al. (2017), Lanne et al. (2017), Lanne and Luoto (2021) Keweloh (2021b), or Guay (2021) for non-Gaussian shocks).

Imposing structure on the stochastic properties of the shocks can be used to derive conditions for the unmixed innovations, while imposing structure on the interaction narrows the space of possible mixing matrices used to unmix the reduced form shocks.

In applied work, the probably most frequently imposed structure are uncorrelated structural shocks (meaning $\varepsilon_{i,t}$ is restricted to be uncorrelated with $\varepsilon_{j,t}$ for $i \neq j$) and a recursive interaction (meaning restricting on B_0 such that $b_{ij} = 0$ for $i < j$ where b_{ij} denotes the element at row i and column j of B_0). Uncorrelated shocks with unit variance can be used to derive $(n+1)n/2$ moment conditions from

$$I = E\left[\varepsilon_t \varepsilon_t'\right] \stackrel{!}{=} E\left[e_t(B) e_t(B)'\right], \tag{4}$$

while a recursive order implies that $n(n-1)/2$ parameters of B_0 are known a priori, leaving only $(n+1)n/2$ unknown parameters in the mixing matrix B. It is then straightforward to show that, if the remaining $(n+1)n/2$ parameters of the restricted B matrix generate unmixed innovations $e(B)$, which satisfy the $(n+1)n/2$ moment conditions in Equation (4), the matrix B has to be equal to B_0 and hence, the unmixed innovations are equal to the structural shocks, meaning the SVAR is identified.[2]

However, economic theory rarely allows to derive the required $n(n-1)/2$ parameter restrictions to ensure identification. More recently, identification methods based on non-Gaussian and independent shocks have been put forward in the literature (see, Gouriéroux et al. (2017), Lanne et al. (2017), Lanne and Luoto (2021) Keweloh (2021b), or Guay (2021)). These identification schemes do not require to impose any restrictions on the impact of the shocks, in particular on the matrix B_0. Instead, the researcher has to impose structure on the stochastic properties of the shocks. If the structural shocks are not only uncorrelated but independent, this property can be used to derive additional moment conditions. For example, consider the coskewness matrices of the structural shocks $S_i = E\left[\varepsilon_t \varepsilon_t' \varepsilon_{i,t}\right]$ for $i = 1, \ldots, n$. Then, independent and mean zero shocks imply that all entries of S_i are zero except for the ith diagonal element, which contains the (unknown) skewness of the shock $\varepsilon_{i,t}$. Hence, the unmixing matrix B has to generate unmixed

[2]Note that this GMM approach is equivalent to the frequently used estimator based on the Cholesky decomposition.

innovations, which satisfy the third-order moment conditions derived from

$$S_i = E\left[\varepsilon_t \varepsilon_t' \varepsilon_{i,t}\right] \stackrel{!}{=} E\left[e_t(B) e_t(B)' e_{i,t}(B)\right], \tag{5}$$

for $i = 1, \ldots, n$ and similarly, B has to generate unmixed innovations, which satisfy the fourth-order moment conditions derived from

$$K_{ij} = E\left[\varepsilon_t \varepsilon_t' \varepsilon_{i,t} \varepsilon_{j,t}\right] \stackrel{!}{=} E\left[e_t(B) e_t(B)' e_{i,t}(B) e_{j,t}(B)\right], \tag{6}$$

for $i, j = 1, \ldots, n$. If at most one component of ε_t is unskewed (has an excess kurtosis), second- and third-order moment conditions (second- and fourth-order moment conditions) identify the SVAR without any further restrictions up to labeling of the shocks, see, e.g., Keweloh (2021b).

3 Block-recursive SVAR

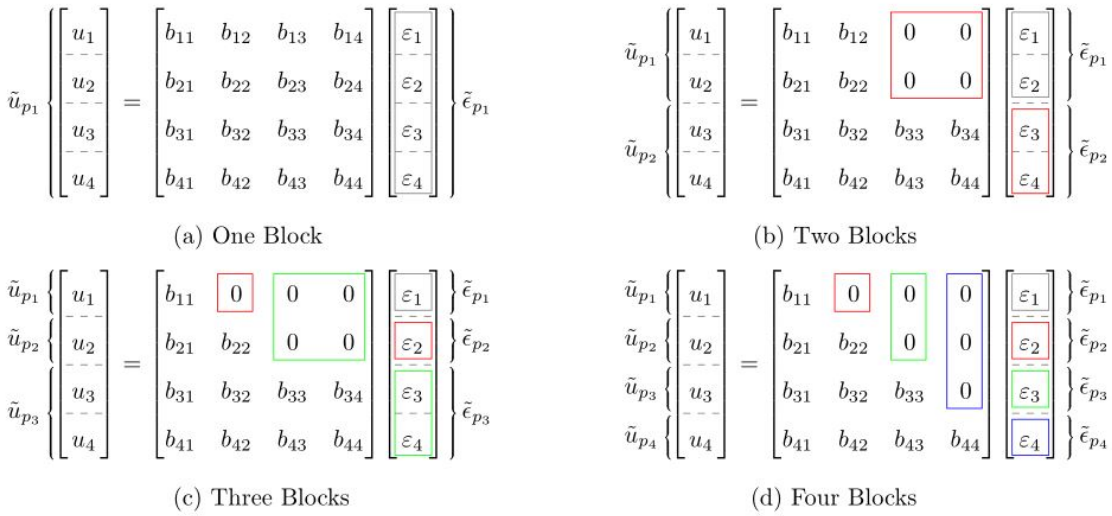
In this section, we propose a generalization between identification based on recursiveness restrictions and identification based on non-Gaussian shocks. The proposed GMM estimator allows the researcher to impose an arbitrary block-recursive structure and for a given block-recursive structure we derive a conservative set of assumptions on the independence and non-Gaussianity of the shocks to ensure identification. In particular, we prove that it is sufficient to assume (mean) independence of the structural shocks in a given block, while shocks of different blocks only need to be uncorrelated.

3.1 Imposing structure on the interaction of shocks

Traditionally, identification of the SVAR is based on a structure imposed on the interaction of the shocks, e.g. short-run or long-run restrictions. These restriction based approaches require a fixed number of restrictions on the interaction of the shocks to ensure identification. A frequently

imposed structure on the interaction is a recursive structure, meaning that each structural shock is restricted to have no simultaneous impact on variables ordered above the shock. The reasoning behind a recursive structure is oftentimes the prejudice that some variables, e.g., some macroeconomic variables like inflation, tend to move slowly, while other variables, e.g. financial variables like stock prices, react faster. However, in practice this intuitive reasoning oftentimes allow to order only some, but not all variables recursively. This motivates us to consider the block-recursive SVAR, meaning that the structural shocks are ordered in blocks of consecutive shocks and each structural shock can simultaneously affect all variables in the same block and in blocks ordered below but explicitly not variables in blocks ordered above.[3] Figure 1 shows different block-recursive structures in a SVAR with four variables. The examples show that a

Figure 1: Examples of Different Block-Recursive SVAR Models.



Note: The figure illustrate how the the block structure can be defined by the structural shocks and our definition of $\tilde{\varepsilon}_{p_i}$ and $\tilde{u}_{p_i}$, $i = 1, \ldots, m$.

block-recursive structure generalizes the non-recursive SVAR and the fully-recursive SVAR and includes both as extreme cases.

[3] Zha (1999) derives identifying restrictions for the block-recursive SVAR. The author restricts not only the simultaneous interaction, but also the lagged interaction. Our proposed block-recursive structure affects only the simultaneous interaction, while the lagged interaction remains unrestricted.

We now introduce the notation for the block-recursive SVAR. Suppose that the structural shocks can be ordered into $m \leq n$ blocks of consecutive shocks. Let the indices $p_1 = 1 < p_2 < \ldots < p_m \leq n$ denote the beginning of a new block and for p_i let $\tilde{\varepsilon}_{p_i,t}$ denote the vector of all structural form shocks in the ith block. Furthermore, $\tilde{u}_{p_i,t}$ denotes reduced form shocks in block i such that

$$\tilde{\varepsilon}_{p_i,t} := \left[\varepsilon_{p_i,t}, \varepsilon_{p_i+1,t}, \ldots, \varepsilon_{p_{i+1}-1,t}\right]', \tag{7}$$

$$\tilde{u}_{p_i,t} := \left[u_{p_i,t}, u_{p_i+1,t}, \ldots, u_{p_{i+1}-1,t}\right]', \tag{8}$$

for $i = 1, \ldots, m$ and $p_{m+1} := n + 1$ for ease of notation. Moreover, let l_i denote the number of shocks in block i for $i = 1, ..., m$. The vector of all structural shocks ε_t can then be decomposed into the m blocks $\varepsilon_t = [\tilde{\varepsilon}'_{p_1,t}, \ldots, \tilde{\varepsilon}'_{p_m,t}]'$ and the reduced form shocks can be decomposed analogously $u_t = [\tilde{u}'_{p_1,t}, \ldots, \tilde{u}'_{p_m,t}]'$. The SVAR is block-recursive with $m \leq n$ blocks with $p_1 = 1 < p_2 < \ldots < p_m \leq n$, if shocks in the ith block have no simultaneous impact on reduced form shocks in blocks j with $j < i$ such that for $i = 1, \ldots, m$

$$b_{ql} = 0, \text{ for } l \geq p_i \text{ and } q < p_i. \tag{9}$$

More generally, any block-recursive structure can be described by the following assumption.

Assumption 1. *Block-recursive interaction:*
For $m \leq n$ blocks with $p_1 = 1 < p_2 < \ldots < p_m \leq n$ and $q, l = 1, ..., n$ let
$B_0 \in \mathbb{B}_{brec} := \{B \in \mathbb{B} |\ b_{ql} = 0$ *if* $\exists p_i \in \{p_1, ..., p_m\}$ *with* $l \geq p_i$ *and* $q < p_i\}$.

3.2 Imposing structure on the stochastic properties of shocks

Imposing structure according to Assumption 1 on the interaction is not sufficient to ensure identification and further assumptions on the dependence and potential non-Gaussianity of the shocks are required. Almost all identification approaches at least assume mutually uncorrelated

structural shocks such that $E[\varepsilon_{i,t}\varepsilon_{j,t}] = E[\varepsilon_{i,t}]\,E[\varepsilon_{j,t}]$ for $i \neq j$.[4] Mutually uncorrelated shocks are justified by the idea that different structural shocks are orthogonal, e.g., a structural monetary policy shock should not depend on other structural shocks. In general, imposing uncorrelated structural shocks does not rule out that the structural shocks are dependent. In this case, the interpretation of the estimated SVAR via impulse response functions can be misleading. For example consider the two random variables $\varepsilon_1 \sim \mathcal{N}(0,1)$ and $\varepsilon_2 = \varepsilon_1^2 - 1$ such that both random variables are uncorrelated, but dependent. Policy analysis based on impulse response functions typically uses the ceteris paribus assumption that only a single shock varies, while the other shocks remain unchanged. In the example above, both shocks are uncorrelated, but nevertheless always move simultaneously. Therefore, uncorrelated structural shocks are not sufficient to guarantee that the ceteris paribus assumption holds.

A more rigorous implementation of the idea of orthogonal shocks is to assume mutually independent shocks such that $E[h(\varepsilon_{i,t})g(\varepsilon_{j,t})] = E[h(\varepsilon_{i,t})]\,E[g(\varepsilon_{j,t})]$ for $i \neq j$ with bounded, measurable functions $g(\cdot)$ and $h(\cdot)$. If shocks are independent, a structural shock cannot contain any information on any other structural shock. Therefore, independent structural shocks justify the ceteris paribus interpretation used in policy analysis based on impulse response functions. However, several authors argue that the assumption of independent structural shocks is too strong (cf. Kilian and Lütkepohl (2017, Chapter 14), Lanne and Luoto (2021), or Lanne et al. (2021)). In particular, independence implies that also the volatility processes of the shocks are independent, which may be too restrictive for some macroeconomic applications. For example suppose that $\tilde{\varepsilon}_{1,t}$ and $\tilde{\varepsilon}_{2,t}$ are drawn independently of each other and represent unscaled structural shocks. Additionally, in each period an additional volatility shock v_t is drawn independently of the other shocks and the structural shocks are given by $\varepsilon_{1,t} = \tilde{\varepsilon}_{1,t}v_t$ and $\varepsilon_{2,t} = \tilde{\varepsilon}_{2,t}v_t$. These structural shocks are uncorrelated but dependent since the variance of one shock contains information on the variance of the other shock.

A compromise between the two extreme cases of mutually uncorrelated and mutually indepen-

[4] Proxy-variable identification approaches are different and instead assume that structural shocks are uncorrelated with an external proxy variable.

dent shocks is the assumption of mutually mean independent shocks, such that $E\left[\varepsilon_{i,t} g(\varepsilon_{j,t})\right] = E\left[\varepsilon_{i,t}\right] E\left[g(\varepsilon_{j,t})\right]$ for $i \neq j$ with a bounded, measurable function $g(\cdot)$. If shocks are mutually mean independent, a structural shock cannot contain any information about the mean of other structural shock. Mutually mean independent shocks can justify the ceteris paribus assumption used in impulse response analysis and at the same time allow for depended volatility processes. In particular, the two shocks $\varepsilon_{1,t} = \tilde{\varepsilon}_{1,t} v_t$ and $\varepsilon_{2,t} = \tilde{\varepsilon}_{2,t} v_t$ defined above are mean independent since a given shock contains no information on the mean of the other shock.

Imposing structure on the dependence of the structural shocks allows to derive moment conditions, see, e.g. Lanne and Luoto (2021), Keweloh (2021b), or Guay (2021). In particular, uncorrelated structural shocks with mean zero and unit variance imply n variance and $n(n-1)$ covariance conditions

$$E[e(B)_{i,t}^2] = 1 \quad \text{and} \quad E[e(B)_{i,t} e(B)_{j,t}] = 0, \text{ for } i, j = 1, ..., n \text{ and } i \neq j. \tag{10}$$

Additionally, mean independent structural shocks implies asymmetric cokurtosis conditions

$$E[e(B)_{i,t}^3 e(B)_{j,t}] = 0, \text{ for } i, j = 1, ..., n \text{ and } i \neq j. \tag{11}$$

Moreover, mean independent structural shocks imply additional cokurtosis conditions

$$E[e(B)_{i,t}^2 e(B)_{j,t} e(B)_{k,t}] = 0, \text{ for } i, j, k = 1, ..., n \text{ and } i \neq j \neq k, \tag{12}$$

$$E[e(B)_{i,t} e(B)_{j,t} e(B)_{k,t} e(B)_{l,t}] = 0, \text{ for } i, j, k, l = 1, ..., n \text{ and } i \neq j \neq k \neq l. \tag{13}$$

Independent structural shocks would imply further symmetric cokurtosis conditions

$$E[e(B)_{i,t}^2 e(B)_{j,t}^2] = 1, \text{ for } i, j = 1, ..., n \text{ and } i \neq j. \tag{14}$$

Furthermore, all coskewness conditions

$$E[e(B)_{i,t}^2 e(B)_{j,t}] = 0, \text{ for } i,j = 1,...,n \text{ and } i \neq j, \tag{15}$$

$$E[e(B)_{i,t} e(B)_{j,t} e(B)_{k,t}] = 0, \text{ for } i,j,k = 1,...,n \text{ and } i \neq j \neq k, \tag{16}$$

can be derived from mutually mean independent structural shocks.

3.3 Identification and estimation

In this section, we show that identification in a block-recursive SVAR can be achieved by including all variance-covariance conditions and the asymmetric cokurtosis conditions $E[e(B)_{i,t}^3 e(B)_{j,t}] = 0$ of innovations $e(B)_{i,t}$ and $e(B)_{j,t}$ in the same block. Higher-order moment conditions of shocks in different blocks are not necessary for identification and can be left out of the estimation. Thereby, identification of the block-recursive SVAR model can be achieved by relying as little on higher-order moment conditions as possible, avoiding a many moment problem in finite samples.

Let $E[f_{\mathbf{2}}(B,u_t)] = 0$ contain all variance-covariance conditions from Equation (10) and let $E[f_{\mathbf{4_{p_i}}}(B,u_t)] = 0$ contains all asymmetric cokurtosis conditions from Equation (11) corresponding to shocks in block k, e.g., $E[e(B)_{i,t}^3 e(B)_{j,t}] = 0$ for $i,j = p_k,...,p_{k+1}-1$ and $i \neq j$. We define the conservative set of moment conditions as

$$E[f_{\mathbf{N}}(B,u_t)] := E\begin{bmatrix} f_{\mathbf{2}}(B,u_t) \\ f_{\mathbf{4_{p_1}}}(B,u_t) \\ \vdots \\ f_{\mathbf{4_{pm}}}(B,u_t) \end{bmatrix}. \tag{17}$$

Note that the set $E[f_{\mathbf{N}}(B,u_t)]$ does not contain asymmetric cokurtosis conditions of shocks in different blocks, e.g., the condition $E[e(B)_{i,t}^3 e(B)_{j,t}] = 0$ for shocks $e(B)_{i,t}$ and $e(B)_{j,t}$ in different blocks is not contained in $E[f_{\mathbf{N}}(B,u_t)]$. Let $k_{\mathbf{N}}$ denote the number of conditions in $f_{\mathbf{N}}(B,u_t)$. The conditions $E[f_{\mathbf{N}}(B,u_t)]$ can be justified by the following assumption.

Assumption 2. *Block-recursive mean independence:*

For $m \leq n$ blocks with $p_1 = 1 < p_2 < \ldots < p_m \leq n$,

(i) all shocks are mutually uncorrelated, i.e., $E[\varepsilon_{i,t}\varepsilon_{j,t}] = 0$ for $i \neq j$.

(ii) all shocks within the same block are mutually mean independent, i.e., $E[\varepsilon_{i,t}|\varepsilon_{-i,t}] = 0$ for $i \in \{p_k, p_k + 1, ..., p_{k+1} - 1\}$ and $-i = \{p_k, p_k + 1, ..., p_{k+1} - 1\} \backslash i$ for $k = 1, \ldots, m$.

The conservative set of moment conditions contains n variance conditions, $n(n-1)/2$ covariance conditions and $\sum_{k=1}^{m} l_k(l_k-1)/2$ asymmetric cokurtosis conditions, where $l_k := p_{k+1} - p_k$ denotes the number of shocks in block k. Therefore, each additional specified block refines the conservative set $f_{\mathbf{N}}(B, u_t)$ such that it contains less higher-order moment conditions. In the extreme case, when the SVAR is specified recursively, meaning each block contains only one variable, the conservative set of moment conditions contains no higher-order moment conditions. In the other extreme case of a single block containing all variables, the conservative set of moment conditions contains all $n(n-1)$ asymmetric cokurtosis conditions and it is similar to the conditions proposed in Lanne and Luoto (2021).[5] Under Assumption 1 and Assumption 2, Keweloh et al. (2021) show that the conservative set of moment conditions is sufficient to locally identify the block-recursive SVAR, i.e., $E[f_{\mathbf{N}}(X, u_t)] = 0$ is locally identified at $X = B_0$ with $X \in \mathbb{B}_{brec}$.

Define the block-recursive SVAR GMM estimator which minimizes the variance, covariance and the asymmetric cokurtosis conditions over the set of block-recursive matrices as

$$\hat{B}_{\mathbf{N}} := \underset{B \in \mathbb{B}_{brec}}{\arg\min}\, g_{\mathbf{N}}(B)' W_{\mathbf{N}} g_{\mathbf{N}}(B), \tag{18}$$

with a suitable weighting matrix $W_{\mathbf{N}}$ and $g_{\mathbf{N}}(B) := 1/T \sum_{t=1}^{T} f_{\mathbf{N}}(B, u_t)$. Consistency, asymp-

[5] Lanne and Luoto (2021) propose to select $n(n-1)/2$ asymmetric cokurtosis conditions, which is sufficient for local identification if none of the asymmetric conditions does include the third power of a Gaussian shock. They advocate to rely on a moment selection criterion to avoid including redundant conditions or conditions of Gaussian shocks. Additionally, Lanne and Luoto (2021) note that including all $n(n-1)$ asymmetric cokurtosis conditions ensures local identification even if conditions related to Gaussian shocks are included. We argue that the degree of overidentification remains reasonably small even if we include all asymmetric cokurtosis conditions and therefore, including redundant conditions can be expected to be rather harmless. For example, in a SVAR with four variables and no restrictions the conservative set has 22 conditions to identify 16 parameters. Thus, we suggest to use all asymmetric cokurtosis conditions in order to avoid the cumbersome process of selecting a subset of the conditions.

totic normality and asymptotically optimal weighting of the block-recursive SVAR GMM estimator follow from the identification result of Keweloh et al. (2021) and standard assumptions, see Hall (2005). That is,

$$\hat{B}_{\mathbf{N}} \xrightarrow{p} B_0 \tag{19}$$

$$\sqrt{T}\left(\left(vec\left(\hat{B}_{\mathbf{N}}\right) - vec\left(B_0\right)\right)\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\mathbf{N}}\right) \tag{20}$$

where the formula of the asymptotic variance, $V_{\mathbf{N}}$, is standard but lengthy and, therefore, deferred to Appendix A.1. Using the weighting matrix $W^*_{\mathbf{N}} := S^{-1}_{\mathbf{N}}$ with $S_{\mathbf{N}} := \lim_{T\to\infty} E[g_{\mathbf{N}}(B)g_{\mathbf{N}}(B)']$ leads to the estimator $\hat{B}^*_{\mathbf{N}}$ with lowest possible asymptotic variance.

3.4 Efficiency gains and moment selection

In the previous section, we showed that identification of a block-recursive SVAR can be achieved by the conservative set of moment conditions $E[f_{\mathbf{N}}(B, u_t)]$, which contains no higher-order moment conditions except for the asymmetric cokurtosis conditions of shocks in the same block. In this section, we show that some of the remaining coskewness and cokurtosis conditions implied by mean independent shocks can contain additional information and decrease the asymptotic variance of the GMM estimator. Recognizing that the number of coskewness and cokurtosis conditions implied by mean independent shocks increases quickly with the dimension of the SVAR, which leads to many moment problems, we propose to included these additional higher-order moment conditions via a LASSO estimator. In particular, we use the pGMM estimator proposed by Cheng and Liao (2015a) which is able to detect and select only the valid and relevant higher-order moment conditions in a data-driven way.

3.4.1 Overidentified block-recursive SVAR GMM estimator

For a given block-recursive SVAR, define the overidentifying higher-order moment conditions as

$$f_{\mathbf{D}}(B, u_t) = \begin{bmatrix} f_{\mathbf{3}}(B, u_t) \\ f_{\mathbf{4}\setminus\mathbf{N}}(B, u_t) \end{bmatrix}, \tag{21}$$

where $E[f_{\mathbf{4}\setminus\mathbf{N}}(B, u_t)] = 0$ denotes the asymmetric cokurtosis conditions from Equation (11) – (12) not contained in $\mathbf{N}$ and $E[f_{\mathbf{3}}(B, u_t)] = 0$ all coskweness conditions, i.e., $E[e(B)_{i,t}e(B)_{j,t}e(B)_{k,t}] = 0$ for $i, j, k = 1, ..., n$ and $j \neq k$. Let $k_{\mathbf{D}}$ denote the number of conditions in $f_{\mathbf{D}}(B, u_t)$ and $\widetilde{D} := \{1, \ldots, k_{\mathbf{D}}\}$ the indices for all overidentifying moment conditions. Moreover, $f_{\mathbf{D}_j}(B, u_t)$ for $j \in \widetilde{D}$ corresponds to one specific moment of $f_{\mathbf{D}}(B, u_t)$.

The overidentified block-recursive SVAR GMM estimator is defined as

$$\hat{B}_{\mathbf{N+D}} := \underset{B \in \mathbb{B}_{brec}}{\arg\min} \begin{bmatrix} g_{\mathbf{N}}(B) \\ g_{\mathbf{D}}(B) \end{bmatrix}' W_{\mathbf{N+D}} \begin{bmatrix} g_{\mathbf{N}}(B) \\ g_{\mathbf{D}}(B) \end{bmatrix}, \tag{22}$$

with a suitable weighting matrix $W_{\mathbf{N+D}}$ and $g_{\mathbf{D}}(B) := 1/T \sum_{t=1}^{T} f_{\mathbf{D}}(B, u_t)$. In contrast to the block-recursive SVAR GMM estimator in Equation (18), consistency and asymptotic normality of the overidentified block-recursive SVAR GMM estimator requires not only the conservative set of moment conditions $E[f_{\mathbf{D}}(B, u_t)]$ but also the set of overidentfying moment conditions $E[f_{\mathbf{D}}(B, u_t)]$ to be valid, which holds under Assumption 1 and Assumption 2. That is, analogously to Equation (19) and Equation (19), $\hat{B}_{\mathbf{N+D}}$ is a consistent estimator of B_0 and asymptotically normal with asymptotic variance $V_{\mathbf{N+D}}$, which depends on both the set of conservative and overidentfying moment conditions (details are given in Appendix A.1). Using the weighting matrix $W^*_{\mathbf{N+D}} := S^{-1}_{\mathbf{N+D}}$ with $S_{\mathbf{N+D}} := \lim_{T\to\infty} E[g_{\mathbf{N+D}}(B_0)g_{\mathbf{N+D}}(B_0)']$, where $g_{\mathbf{N+D}}(B_0) := [g_{\mathbf{N}}(B_0)', g_{\mathbf{D}}(B_0)']'$, leads to the estimator $\hat{B}^*_{\mathbf{N+D}}$ with lowest possible asymptotic variance.

It is well known that adding additional valid moment conditions can never increase the asymptotic variance of the GMM estimator, see, e.g., Breusch et al. (1999). Therefore, if the structural shocks

are mean independent such that the overidentifying conditions hold, the asymptotic variance of $\hat{B}^*_{\mathbf{N+D}}$ is equal to or smaller than the asymptotic variance of $\hat{B}^*_{\mathbf{N}}$. If including an additional moment condition decreases the asymptotic variance of the estimator, the moment condition is called relevant, otherwise the moment condition is called redundant, see Breusch et al. (1999). A moment condition is called partially relevant for a subset of parameters if it decreases the asymptotic variance of a subset of parameters. If this is not the case, the moment conditions is called partially redundant, see Breusch et al. (1999).

In the following proposition, we show that overidentifying moment conditions higher-order moment conditions in $f_{\mathbf{D}}(B, u_t)$ can decrease the asymptotic variance of the estimator. To this end, we consider the special case of a recursive SVAR which is identified solely by second-order moment conditions and all coskewness and cokurtosis moment conditions are overidentifying.

Proposition 1. *Efficiency gains in the recursive SVAR.*
In a recursive SVAR with mean independent structural shocks, it holds that:

1. *If at least one structural shock has non-zero skewness, the set $f_{\mathbf{3}}(B, u_t)$ which contains all coskewness conditions implied by mean independent shocks is relevant w.r.t. $f_{\mathbf{2}}(B, u_t)$.*

2. *If at least one structural shock has non-zero excess kurtosis or zero skewness, the set $f_{\mathbf{4}}(B, u_t)$ which contains all cokurtosis conditions implied by mean independent shocks is relevant w.r.t. $f_{\mathbf{2}}(B, u_t)$.*

Proof. The proof is based on the redundancy conditions from Breusch et al. (1999) and available on request. □

The Proposition implies that the recursive SVAR GMM estimator based solely on second-order moments in Equation (18) has a higher asymptotic variance than the overidentified recursive SVAR GMM estimator based on second- and higher-order moments in Equation (22) if at least one structural shock has a non-zero skewness or excess kurtosis. Hence, in a recursive SVAR with mean independent structural shocks the frequently used estimator obtained by applying the

Cholesky decomposition is inefficient if at least one structural shock has non-zero skewness or excess kurtosis.[6]

The proof of Proposition 1 involves more detailed statements on the relevance or redundancy of specific coskewness and cokurtosis conditions in a recursive SVAR. In particular, we proof that certain coskewness and cokurtosis conditions are always redundant, while the (partial) redundancy of other conditions depends on conditions on the true B_0 matrix and skewness and kurtosis of the structural shocks. In practice, these conditions cannot be verified since the B_0 matrix and the skewness or kurtosis of the structural shocks is unknown a priori. Furthermore, Proposition 1 only covers recursive SVARs, i.e., for more general block-recursive SVARs it is unclear, which moment conditions are relevant and which are not.

3.4.2 Data-driven moment selection

We apply the pGMM estimator of Cheng and Liao (2015a) to detect and include only the relevant and valid overidentifying moment conditions, without relying on a priori knowledge of the skewness and excess kurtosis. The overidentified SVAR GMM estimator in Equation (22) is asymptotically efficient but the number of overidentifying moment conditions increases quickly with the dimension of the SVAR, which leads to a many moments problem. This distorts the finite sample performance of the estimator (see, e.g., Cheng and Liao (2015b), Hall (2005), and Hall (2015)).

By including valid and relevant moment conditions in the estimation, we exploit the asymptotic efficiency gains of relevant moments in finite samples. By leaving out the remaining moment conditions, which do not lower the asymptotic variance but increase finite sample variance of the estimation, we address the many moments problem arising in large SVARs. In general, the overidentifying higher-order moment conditions $f_{\mathbf{D}}(B, u_t)$ can be separated into three sets: $f_{\mathbf{A}}(B, u_t)$ contains valid and relevant moment conditions, $f_{\mathbf{R}}(B, u_t)$ contains valid but redundant

[6] The proposition also holds if we replaced the assumption of mean independence of the structural shocks by independence.

conditions, and $f_{\mathbf{I}}(B, u_t)$ contains invalid moment conditions. The goal is to select the moments $f_{\mathbf{A}}(B, u_t)$ and to leave out the moments $f_{\mathbf{R}}(B, u_t)$ and $f_{\mathbf{I}}(B, u_t)$. However, in practice the researcher does not know whether a given moment condition is invalid, redundant, or valid and relevant. Therefore, we propose to detect and select the relevant and valid overidentifying moment conditions in a data-driven way. Based on Cheng and Liao (2015a), define the LASSO-type SVAR GMM estimator

$$\{\widehat{B}^{pGMM}_{\mathbf{N+D}}, \widehat{\beta}\} := \underset{\{B,\beta\}\in\Lambda}{\arg\min} \begin{bmatrix} g_{\mathbf{N}}(B) \\ g_{\mathbf{D}}(B) - \beta \end{bmatrix}' W_{\mathbf{N+D}} \begin{bmatrix} g_{\mathbf{N}}(B) \\ g_{\mathbf{D}}(B) - \beta \end{bmatrix} + \lambda \sum_{j\in\widetilde{D}} \omega_j \, |\beta_j|, \tag{23}$$

where $\lambda \geq 0$ is a tuning parameter specified by the researcher, $\beta \in \mathbb{R}^{k_{\mathbf{D}}}$ is the vector of slackness parameters, $\Lambda := \{\mathbb{B}_{brec}, \mathbb{R}^{1\times k_{\mathbf{D}}}\}$ is the parameter space of $\{B, \beta\}$, and $\omega \in \mathbb{R}^{k_{\mathbf{D}}}$ is a vector of weights used in the penalty term. The vector of slackness parameters β allows the moment conditions generated by $\mathbf{D}$ to deviate from zero without increasing the first part of the loss function and therefore, to decrease their impact on the estimation. However, each element of β gets penalized in the second part of the loss function and consequently, giving slack to overidentifying moments adds a cost, i.e., increases the loss function. The vector of weights ω and the tuning parameter λ govern the cost of giving slack to moment conditions. In particular, a smaller λ makes it cheaper to give slack to all overidentifying moments and a smaller ω_j makes it less costly to give slack to a specific overidentifying moment j.

The pGMM estimator in Equation (23) has two special cases. First, if $\lambda = 0$, adding slack to the overidentifying moments is not penalized. Therefore, the solution of the pGMM estimator is $\widehat{B}^{pGMM}_{\mathbf{N+D}} = \hat{B}_{\mathbf{N}}$ and $\widehat{\beta} = g_{\mathbf{D}}\left(\hat{B}_{\mathbf{N}}\right)$, where $\hat{B}_{\mathbf{N}}$ is the solution of the the conservative block-recursive SVAR GMM estimator in Equation (18) using only the moments generated by $\mathbf{N}$ and the weighting matrix $W_{\mathbf{N}}$, corresponding to the leading $k_{\mathbf{N}} \times k_{\mathbf{N}}$-block of the weighting matrix $W_{\mathbf{N+D}}$. Second, if $\lambda = \infty$, deviations of β from zero become infinitely costly for overidentifying moments with $\omega_j > 0$. Assuming $\omega > 0$, the pGMM estimator cannot give slack to any overidentifying moment condition. Thus, $\widehat{B}^{pGMM}_{\mathbf{N+D}} = \hat{B}_{\mathbf{N+D}}$ and $\widehat{\beta} = 0$ minimize the loss function of the pGMM

estimator, where $\hat{B}_{\mathbf{N+D}}$ is the solution of the the overidentified block-recursive SVAR GMM estimator in Equation (22)) using the weighting matrix $W_{\mathbf{N+D}}$. Other choices of λ than $\lambda = 0$ or $\lambda = \infty$ lead to solutions which lie between these extreme cases. In practice, we recommend using cross-validation to find the optimal value of λ.

The penalty term uses weights $\omega_j \geq 0,\ \forall j \in \tilde{D}$, to shrink the elements of β differently. That is, a higher ω_j, leads to more shrinkage for β_j and consequently, makes it more likely that β_j becomes zero, meaning that the corresponding moment $f_{\mathbf{D}_j}(B, u_t)$ gets selected. Furthermore, $\omega_j = 0$ implies that even if the tuning parameter λ is large, there is no cost for giving slack to the moment condition $f_{\mathbf{D}_j}(B, u_t)$, implying that those moments do not influence the estimated $\widehat{B}^{pGMM}_{\mathbf{N+D}}$. Since we aim to select only the relevant and valid moment conditions $f_{\mathbf{A}}(B, u_t)$, and not the invalid or redundant moment conditions $f_{\mathbf{R}}(B, u_t)$ and $f_{\mathbf{I}}(B, u_t)$, we would specify $\omega_j > 0$ for all valid and relevant conditions, and $\omega_j = 0$ for all invalid or redundant conditions. To achieve this without prior knowledge on $f_{\mathbf{A}}(B, u_t)$, $f_{\mathbf{R}}(B, u_t)$, and $f_{\mathbf{I}}(B, u_t)$, Cheng and Liao (2015a) construct ω_j allowing information-based adaptive adjustment for each moment in $f_{\mathbf{D}}(B, u_t)$. More precisely, they use

$$\omega_j = \frac{\mu_j^{r_1}}{|\beta_j^{*r_2}|},\ j \in \tilde{D}, \tag{24}$$

where μ_j is a measure for the empirical relevance of the moment condition $f_{\mathbf{D}_j}(B, u_t)$, relative to the moment conditions of conservative set $f_{\mathbf{N}}(B, u_t)$, and β_j^* is a preliminary consistent estimator of $E[f_{\mathbf{D}_j}(B_0, u_t)]$ and $r_1 \geq r_2 \geq 0$ are constants specified by the researcher. The use of $1/|\beta_j^{*r_2}|$ resembles an adpative LASSO penalty (cf. Zou (2006)) and implies that moments with small β_j^* are subject to more shrinkage. Since β_j^* is a consistent estimator and the true value of β_j^* for a valid moment is zero, the adaptive penalty ensures that valid moments get selected. However, using only the adpative penalty, we would unintendedly incentivize the estimator to select also redundant moments since, by definition, these are also valid. To avoid selecting redundant

moments, Cheng and Liao (2015a) suggest to multiply the adaptive penalty with

$$\mu_j = \rho_{\max}\left(\widehat{V}_{\mathbf{N}} - \widehat{V}_{\mathbf{N}+\mathbf{D}_j}\right),\ j \in \tilde{D}, \tag{25}$$

where $\rho_{\max}(A)$ is the maximum eigenvalue of an arbitrary square matrix A and $\widehat{V}_{\mathbf{N}}$ and $\widehat{V}_{\mathbf{N}+\mathbf{D}_j}$ are consistent estimators of the efficient asymptotic variance-covariance matrices $V^*_{\mathbf{N}}$ and $V^*_{\mathbf{N}+\mathbf{D}_j}$, defined in Appendix A.1. If the maximum eigenvalue of $V^*_{\mathbf{N}} - V^*_{\mathbf{N}+\mathbf{D}_j}$ is positive, then adding moment condition $f_{\mathbf{D}_j}(B, u_t)$ to the conditions $f_{\mathbf{N}}(B, u_t)$ decreases the asymptotic variance of the estimator and hence, moment condition $f_{\mathbf{D}_j}(B, u_t)$ is relevant. Therefore, μ_j estimates the empirical relevance of the moment generated by $\mathbf{D}_j$.[7]

Cheng and Liao (2015a) show that, under conditions, the pGMM estimator consistently selects the valid and relevant moments, i.e., $\lim_{T\to\infty} P(\hat{\beta}_j = 0) = 1$ if the moment condition $f_{\mathbf{D}_j}(B, u_t)$ is in $f_{\mathbf{A}}(B, u_t)$, and does not select the invalid or redundant moments, i.e., $\lim_{T\to\infty} P(\hat{\beta}_j = 0) = 0$ if the moment condition $f_{\mathbf{D}_j}(B, u_t)$ is in $f_{\mathbf{R}}(B, u_t)$ or $f_{\mathbf{I}}(B, u_t)$. They also derive that, under conditions, the pGMM estimator is a consistent estimator of B_0 and asymptotically normal with asymptotic variance $V_{\mathbf{N}+\mathbf{A}}$.[8] Even though the pGMM estimator uses the moment conditions $f_{\mathbf{N}}(B, u_t)$ and $f_{\mathbf{D}}(B, u_t)$ for the estimation, its asymptotic variance does only depend on the moments conditions $f_{\mathbf{D}}(B, u_t)$ and $f_{\mathbf{A}}(B, u_t)$. That is, the pGMM estimator successfully ignores the redundant and invalid moments and decreases the asymptotic variance by incorporating the information contained in the relevant and valid moments. The weighting matrix $W^*_{\mathbf{N}+\mathbf{D}} := S^{-1}_{\mathbf{N}+\mathbf{D}}$ leads to the estimator with the lowest possible asymptotic variance (Hall, 2005), corresponding to the asymptotic variance of the oracle estimator. The oracle estimator uses only moment

[7] Cheng and Liao (2015a) show that $V^*_{\mathbf{N}} - V^*_{\mathbf{N}+\mathbf{D}_j}$ is positive semidefinite for every $j \in \widetilde{D}$, implying that the maximum eigenvalue of $V^*_{\mathbf{N}} - V^*_{\mathbf{N}+\mathbf{D}_j}$ is nonnegative. Furthermore, note that both $\widehat{V}_{\mathbf{N}} \equiv \widehat{V}_{\mathbf{N}}\left(\hat{B}_{\mathbf{N}}\right)$ and $\widehat{V}_{\mathbf{N}+\mathbf{D}_j} \equiv \widehat{V}_{\mathbf{N}+\mathbf{D}_j}\left(\hat{B}_{\mathbf{N}}\right)$ are evaluated at $\hat{B}_{\mathbf{N}}$, which is obtained from Equation (18). Thereby, we do not rely on $\hat{B}_{\mathbf{N}+\mathbf{D}_j}$ to estimate $V^*_{\mathbf{N}+\mathbf{D}_j}$ since the moment associated with $\mathbf{D}_j$ may be invalid and hence, $\widehat{V}_{\mathbf{N}+\mathbf{D}_j}\left(\hat{B}_{\mathbf{N}+\mathbf{D}_j}\right)$ inconsistent for $V^*_{\mathbf{N}+\mathbf{D}_j}$.

[8] This result is not explicitly stated in Cheng and Liao (2015a) but follows from their Remark 3.5 using the Cramér-Wold device, an arbitrary weighting matrix W and replacing the variance of the sample GMM estimator with the asymptotic variance. We proof the result in Appendix A.3, assuming that Assumption 1 and Assumption 2 hold.

conditions $f_{\mathbf{N}}(B, u_t)$ and $f_{\mathbf{A}}(B, u_t)$ and is infeasible in practice without prior knowledge on the index sets $\mathbf{D}$ and $\mathbf{A}$. However, the pGMM estimator is as efficient as the oracle estimator asymptotically.

4 Finite sample performance

In this section, we conduct a Monte Carlo study to illustrate that the pGMM estimator selects relevant moments in a data-driven way and thereby, improves the finite sample performance compared to the SVAR GMM estimator based on all third and fourth moments. Asymptotically, the pGMM estimator is as efficient as the SVAR GMM estimator based on second, third and fourth moments and both are more efficient than the SVAR GMM estimator based only on second moments, which is equivalent to a Cholesky decomposition.

We simulate a recursive SVAR using $n = 4$ variables and

$$B_0 = \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 0.50 & 1.00 & 0 & 0 \\ 0.50 & 0.50 & 1.00 & 0 \\ 0.50 & 0.50 & 0.50 & 1.00 \end{pmatrix}. \tag{26}$$

For the estimation of B_0, we impose a recursive order for all considered estimators, i.e., we use zero restrictions for all elements where B_0 is zero. We consider five different sample sizes $T = \{100, 200, 500, 1\,000, 5\,000\}$ to analyze the influence of the sample size on the performance of the estimators. We independently and identically draw each structural shock ϵ_{ti}, $t = 1, \ldots, T$, $i = 1, \ldots n$, from the two-component mixture

$$\epsilon_{ti} \sim 0.79\, \mathcal{N}(-0.2, 0.7^2) + 0.21\, \mathcal{N}(0.75, 1.5^2),$$

implying that the shocks have skewness 0.90 and kurtosis 5.41.

With the imposed recursive order, the set $f_{\mathbf{N}}(B, u_t)$ contains 10 and the set $f_{\mathbf{D}}(B, u_t)$ moment 47 conditions. All moment conditions in $f_{\mathbf{D}}(B, u_t)$ are valid. More precisely, 17 of its moment conditions are in $f_{\mathbf{R}}(B, u_t)$, i.e., redundant, and 30 are in $f_{\mathbf{A}}(B, u_t)$, i.e., valid and relevant. We compare the finite sample performance[9] of the following recursive estimators:

- Cholesky estimator: We apply the conservative estimator in Equation (18), which uses only the set of conservative moment conditions $f_{\mathbf{N}}(B, u_t)$. The estimator is just identified and equal to the frequently used estimator obtained by applying the Cholesky decomposition.
- CUE estimator: We apply the overidentified estimator given in Equation (22), which uses all moments generated by **N** and **D**. We use the continuously updating estimator instead of the GMM estimator in Equation (22).
- CUE-Oracle estimator: We apply the overidentified estimator given in Equation (22), using only the identifying moment conditions $f_{\mathbf{N}}(B, u_t)$ and the relevant overidentifying moment conditions $f_{\mathbf{A}}(B, u_t)$. We use the continuously updating estimator instead of the SVAR GMM estimator in Equation (22).
- CUE-LASSO estimator: We apply the LASSO estimator in Equation (23). We use a continuously updating version of the LASSO estimator instead of the pGMM estimator in Equation (23).

We use continuously updating estimators proposed by Hansen et al. (1996) instead of GMM estimators since these are known to reduce finite sample bias compared to two-step GMM estimators. Additionally, we specify W in a parametric way as proposed in Keweloh (2021a), which also increases finite sample performance since it exploits the assumed structure on the shocks when estimating W.

The construction of the weights for the pGMM estimator as in Equation (24) requires an initial consistent estimate $\hat{B}$ to estimate β^* and the asymptotic variance in Equation (25). To this

[9]The estimators are implemented in python Van Rossum and Drake (2009), e.g., the pGMM estimation the solvers of Defferrard et al. (2017).

end, we apply the conservative estimator, which is Cholesky estimator in this case. Moreover, we again use mutually independent shocks to estimate the asymptotic variance, as proposed by Keweloh (2021a). Additionally, we use $r_1 = 2$ and $r_2 = 1$ in Equation (24).[10]

We choose the optimal λ for the pGMM estimator with 5-fold cross-validation from a sequence of 10 potential values. The maximum value of the sequence of λ's depends on the sample size, ensuring that it is large enough to selects all momtents j for which $\omega_j / \sum_{k \in \widetilde{D}} \omega_k > 10^{-4}$.[11] We also include $\lambda = 0$ in the range of possible values to allow our estimator to simplify to the recursive SVAR. The selection of the optimal tuning parameter is based on the median of the GMM loss of each left-out fold.

Table 1 summarizes the results of $M = 1000$ Monte Carlo simulations for $T = \{100, 200, 5\,000\}$. The table shows the average of each estimated element $\bar{b}_{ij} = 1/M \sum_{m=1}^{M} \hat{b}_{ij}^m$ and the estimated population variance (scaled by the square root of the sample size), $\hat{\sigma}_{i,j}^2 = 1/M \sum_{m=1}^{M} T \left(\hat{b}_{ij}^m - b_{ij} \right)^2$, where b_{ij} denotes the element of B_0 in row i and column j and $\hat{b}_{ij}^m$ its estimated value in Monte Carlo run m. Moreover, we calculate the sum over the empirical biases, $Bias := \sum_{i=1}^{n} \sum_{j=1}^{i} \left(\bar{b}_{ij} - b_{ij} \right)$ and the sum over the scaled variances, $Var := \sum_{i=1}^{n} \sum_{j=1}^{i} \hat{\sigma}_{i,j}^2$. Additionally, we report the number of moments used by the recursive SVAR GMM estimator, the CUE estimator, and the CUE-oracle estimator and the average number of moments selected by the pGMM estimator across Monte Carlo runs. Furthermore, Table 1 shows the median of the chosen λ's for the pGMM estimator across Monte Carlo runs. In Appendix B, we display results including all sample sizes and the Post-pGMM estimator which uses the moments selected by pGMM in a second stage estimation.

The Cholesky estimator already performs good in the smallest sample size in terms of bias and scaled variance. However, its scaled variance does not decrease further as T increases. In fact, for the largest sample size the Cholesky estimator has the largest scaled variance among all consid-

[10] Cheng and Liao (2015a) use $r_1 = 3$ and $r_2 = 2$ in their simulation study, which performs qualitatively similar in our setting.

[11] We specify the maximum value of the sequence of λ's in a data-driven way using the subgradient of Equation (23) with respect to β. We give more details on how to construct the maximum value of the sequence of λ's in the cross-validation in Appendix A.2

Table 1: Summary Statistics of 1000 Monte Carlo Runs

		Cholesky	CUE	CUE-Oracle	CUE-LASSO
$T = 100$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.16) & & & \\ 0.50 & 0.99 & . & . \\ (1.26) & (1.13) & & \\ 0.50 & 0.49 & 0.99 & . \\ (1.51) & (1.28) & (1.08) & \\ 0.50 & 0.50 & 0.49 & 0.98 \\ (1.72) & (1.47) & (1.26) & (1.09) \end{bmatrix}$	$\begin{bmatrix} 0.98 & . & . & . \\ (1.08) & & & \\ 0.48 & 0.97 & . & . \\ (1.46) & (1.13) & & \\ 0.49 & 0.48 & 0.97 & . \\ (1.66) & (1.52) & (1.12) & \\ 0.49 & 0.48 & 0.49 & 0.97 \\ (2.01) & (1.87) & (1.44) & (1.12) \end{bmatrix}$	$\begin{bmatrix} 0.98 & . & . & . \\ (1.07) & & & \\ 0.49 & 0.97 & . & . \\ (1.28) & (1.13) & & \\ 0.49 & 0.48 & 0.97 & . \\ (1.51) & (1.35) & (1.12) & \\ 0.49 & 0.49 & 0.49 & 0.97 \\ (1.75) & (1.56) & (1.25) & (1.13) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.16) & & & \\ 0.50 & 0.99 & . & . \\ (1.20) & (1.10) & & \\ 0.51 & 0.49 & 0.99 & . \\ (1.44) & (1.27) & (1.12) & \\ 0.50 & 0.50 & 0.50 & 0.99 \\ (1.68) & (1.49) & (1.26) & (1.10) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	23.99
	Bias	−0.0526	−0.1998	−0.2005	−0.0331
	Var	12.97	14.42	13.15	12.83
	λ	.	.	.	103.91
		Cholesky	CUE	CUE-Oracle	CUE-LASSO
$T = 200$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.10) & & & \\ 0.50 & 1.00 & . & . \\ (1.22) & (1.15) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.42) & (1.22) & (1.13) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.76) & (1.54) & (1.22) & (1.15) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (1.02) & & & \\ 0.49 & 0.99 & . & . \\ (1.19) & (1.08) & & \\ 0.49 & 0.49 & 0.99 & . \\ (1.47) & (1.26) & (1.04) & \\ 0.49 & 0.49 & 0.49 & 0.99 \\ (1.82) & (1.58) & (1.21) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (1.02) & & & \\ 0.50 & 0.99 & . & . \\ (1.10) & (1.09) & & \\ 0.49 & 0.49 & 0.99 & . \\ (1.34) & (1.11) & (1.05) & \\ 0.49 & 0.49 & 0.49 & 0.99 \\ (1.60) & (1.39) & (1.05) & (1.09) \end{bmatrix}$	$\begin{bmatrix} 1.01 & . & . & . \\ (1.09) & & & \\ 0.50 & 1.00 & . & . \\ (1.08) & (1.12) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.34) & (1.13) & (1.07) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.59) & (1.40) & (1.08) & (1.09) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	26.37
	Bias	−0.0222	−0.0846	−0.0848	0.0030
	Var	12.93	12.77	11.84	11.99
	λ	.	.	.	114.89
		Cholesky	CUE	CUE-Oracle	CUE-LASSO
$T = 5000$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.21) & & & \\ 0.50 & 1.00 & . & . \\ (1.34) & (1.25) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.62) & (1.37) & (1.22) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.95) & (1.63) & (1.24) & (1.22) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.10) & & & \\ 0.50 & 1.00 & . & . \\ (1.09) & (1.15) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.29) & (1.11) & (1.11) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.57) & (1.33) & (1.08) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.10) & & & \\ 0.50 & 1.00 & . & . \\ (1.08) & (1.15) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.27) & (1.10) & (1.11) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.55) & (1.32) & (1.06) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 1.01 & . & . & . \\ (1.14) & & & \\ 0.50 & 1.00 & . & . \\ (1.09) & (1.19) & & \\ 0.50 & 0.50 & 1.01 & . \\ (1.30) & (1.12) & (1.15) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.58) & (1.35) & (1.08) & (1.13) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	28.95
	Bias	0.0297	0.0284	0.0285	0.0319
	Var	14.05	11.91	11.83	12.12
	λ	.	.	.	22.23

The table reports the average summary statistics over all Monte Carlo replicates for the recursive SVAR estimator (Cholesky), the continuously updating estimator (CUE), the continuously updating oracle estimator (CUE-oracle), and the continuously updating pGMM estimator (CUE-LASSO).

ered estimators. Due to many moments, the CUE estimator performs worst in terms of bias and scaled variance among the considered estimators in small samples. Yet, its performance increases with the sample size and it eventually outperforms the Cholesky estimator. The scaled variance of the CUE-Oracle estimator is already comparable to the Cholesky estimator in small samples. Unlike the Cholesky estimator, its performance further increases with the sample size and it performs similar to the CUE estimator in the largest sample size. In general, the CUE-Oracle

estimator is infeasible since the redundant moments are unknown a priori.[12] In contrast to that, the CUE-LASSO estimator is feasible and uses a data-driven approach to select the relevant and valid moments. The CUE-LASSO estimator performs good across all sample sizes in terms of bias and scaled variance. For $T = 100$, its scaled variance is notably smaller than the one of the CUE and the CUE-Oracle estimator and comparable to the one of the Cholesky estimator. Already for $T = 200$, its scaled variance is below the one of the Cholesky estimator. In the largest sample, the CUE-LASSO estimator performs as good as the CUE and CUE-Oracle estimator.[13] The simulation shows that the CUE-LASSO estimator can, without prior specification, distinguish informative from non-informative overidentifying moments, which solves the many moments problem of the CUE estimator and allows to exploit information in overidentifying higher-order moments already in small samples.

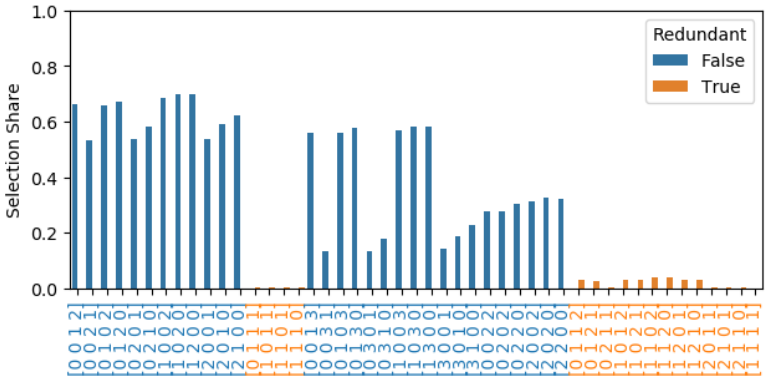
Table 1 indicates that the average number of selected moments increases only slightly as T increases. Even for $T = 5000$, the CUE-LASSO estimator only selects 29 out of 40 valid and relevant moments. That said, the remaining 11 moments would only decrease the scaled variance from 11.40 to 11.22, indicating that the moments being not selected would not lower the scaled variance much. Figure 2 illustrates that CUE-LASSO estimator only selects relevant moments and manages to leave redundant moments out, especially as T increases. Moreover, the share of selections of each moment across all Monte Carlo runs rises with the sample size for the majority of relevant moments.[14]

[12] Even if we knew the non-Gaussianity of the shocks, we would not be able to derive the oracle estimator if the block-recursive structure was not just purely recursive. In this case, we still lack the information on which moments are redundant and which are relevant.

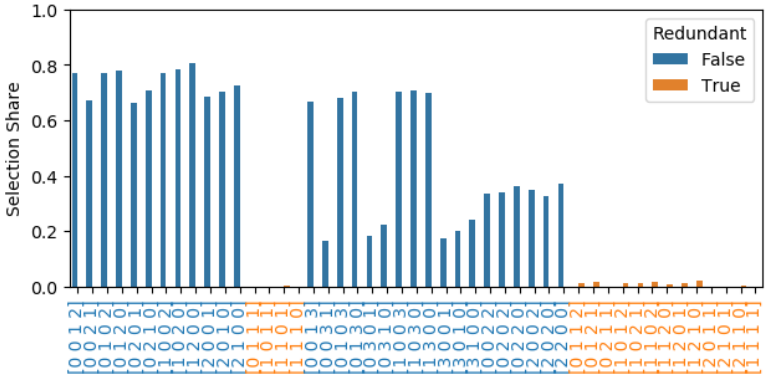
[13] The Post-pGMM estimator reported in Appendix B performs similar to the CUE-LASSO estimator.

[14] In Figure B.6, we plot the average weight of each moment across Monte Carlo runs. By comparing Figure 2 and Figure B.6, we argue that there is a clear correlation between the average weight and the number of selections of each moment. More precisely, all redundant moments have an average weight which is very close to zero and hence, they are not selected by the CUE-LASSO estimator.

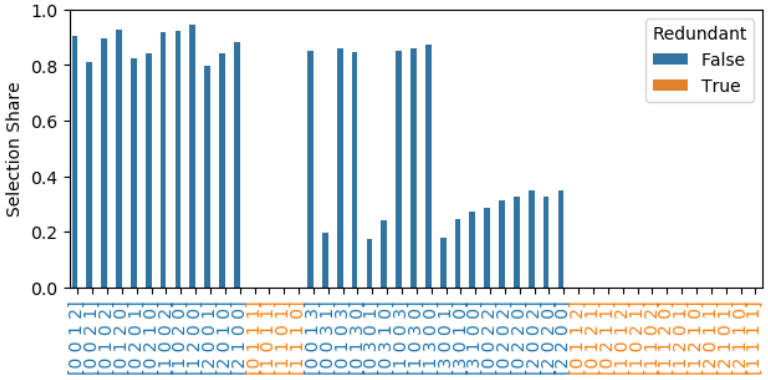
Figure 2: Share of Selections of Moments across Monte Carlo Runs



(a) $T = 100$



(b) $T = 200$

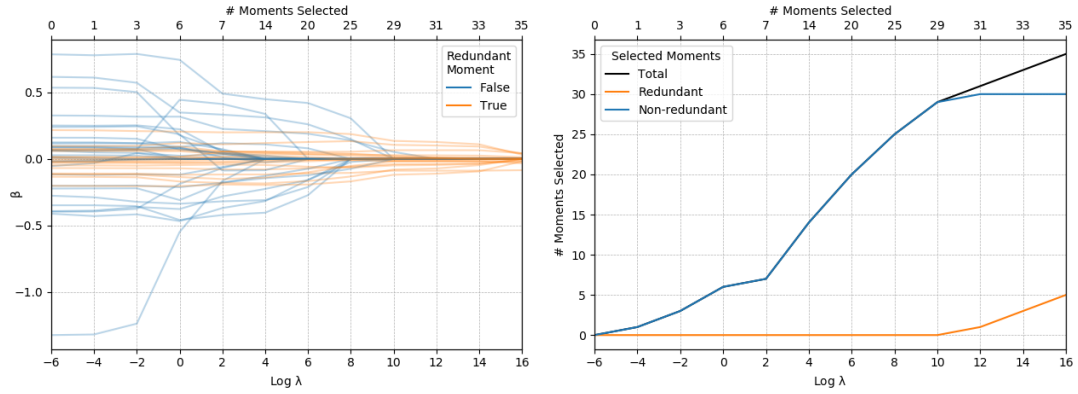


(c) $T = 5000$

Note: The figure shows how often each moment gets selected across $M = 1000$ Monte Carlos simulations. Redundant moments (orange) and relevant moments (blue) are displayed on the x-axis.

Figure 3 highlights the influence of λ on β and hence, on the number of selected moments in $\widetilde{D}$ for one Monte Carlo run[15] For instance, for $log(\lambda) = -6$ no moments in $\widetilde{D}$ are selected and the solution of the pGMM estimator corresponds to the one of the Cholesky estimator. Further, the number of selected moments increases as λ increases, i.e., the penalty shrinks the elements of β to zero. Furthermore, the relevant moments get selected first when λ increases and we do not select any redundant moment until λ becomes very large.

Figure 3: Illustration of Influence of λ on β.



(a) Trace Plot

(b) Selected Moments

Note: Panel (a) of the figure shows the values of β in dependence on $log(\lambda)$ for one Monte Carlo run for $T = 100$ and the corresponding number of selected moments in $\widetilde{D}$. Panel (b) of the figure splits the number of selected moments into the number of selected redundant and the number of selected relevant moments for each $log(\lambda)$.

As a robustness-check, we conduct a Monte Carlo study where we draw each structural shock independently and identically from a t-distribution with $\nu = 9$ degrees of freedom[16]. The results are reported in Appendix B and qualitatively similar to the ones presented in this section.

[15] For the purpose of illustration, we use a wider range of of λ values for this plot.

[16] We normalize the structural shocks to unit variance by multiplying each shock with $1/\sqrt{\nu)/(\nu-2)}$.

5 Empirical illustration

To illustrate the block-recursive estimation, we consider the following SVAR in five variables with quarterly U.S. Data from $1983Q1$ to $2019Q1$ of the form

$$\begin{bmatrix} y_t \\ tfp_t \\ \pi_t \\ s_t \\ i_t \end{bmatrix} = \alpha + \gamma t + \sum_{i=1}^{p} A_i \begin{bmatrix} y_{t-i} \\ tfp_{t-i} \\ \pi_{t-i} \\ s_{t-i} \\ i_{t-i} \end{bmatrix} + \begin{bmatrix} u_t^y \\ u_t^{tfp} \\ u_t^{\pi} \\ u_t^{s} \\ u_t^{i} \end{bmatrix}, \tag{27}$$

with $p = 2$ lags as suggested by the AIC and BIC criteria. The variable y denotes the log of real GDP, tfp the utilization-adjusted total factor productivity from Fernald (2009), π the inflation rate, i the federal funds rate, and s the real stock returns[17].

We assume the following block-recursive structure with two blocks

$$\begin{bmatrix} u_t^y \\ u_t^{tfp} \\ u_t^{\pi} \\ u_t^{s} \\ u_t^{i} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & 0 & 0 \\ b_{21} & b_{22} & b_{23} & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{54} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \\ \varepsilon_{4,t} \\ \varepsilon_{5,t} \end{bmatrix}, \tag{28}$$

such that the macroeconomic variables GDP, tfp, and inflation can simultaneously only react to the first three structural shocks while the federal funds rate and stock returns can simultaneously respond to two additional structural shocks.

The simultaneous relationship is estimated with the CUE-LASSO estimator with the parametric

[17] The inflation rate is defined as the quarter to quarterly growth rate in the quarterly chain-type GDP price index retrieved from the FRED. Real GDP is retrieved from the FRED. The nominal interest rate is defined as the Federal Funds Rate (FFR), where the effective FFR (retrieved from FRED) is replaced by the shadow FFR provided by Wu and Xia (2016) for the Zero Lower Bound observations during the Great Recession. Stock returns are defined as the quarterly log-difference in real stock prices, where real stock prices are given by the S&P 500 index (retrieved from macrotrends.net) divided by the chain-type GDP price index.

weighting matrix as proposed in Keweloh (2021a). With the imposed block-recursive structure, we can divide the third- and fourth-order moment conditions into 39 identifying conditions $f_{\mathbf{N}}(B, u_t)$ and 71 overidentifying conditions $f_{\mathbf{D}}(B, u_t)$. We use the same specifications to construct the weights as in the Monte Carlo simulation, i.e., we use $r_1 = 2$ and $r_2 = 1$ in Equation (24). The weights are displayed in Figure B.8. For the cross-validation, we consider a range of 28 values for λ, including $\lambda = 0$ to allow our estimator to simplify to a recursive SVAR. The maximum value of λ is chosen such that all moments generated by $\mathbf{D}$ for which $\omega_j / \sum_{k \in \widetilde{D}} \omega_k > 10^{-7}$ get selected.[18] With the chosen $\lambda = 2112$, which is the 17th value of the considered sequence, 26 out of the 71 overidentifying conditions $f_{\mathbf{D}}(B, u_t)$ are selected. Figure B.8 illustrates that there is a correlation between the selected moments and their corresponding weights which we already observed in our Monte Carlo simulation.

To ensure identification, at most one structural shock in a given block may be Gaussian. Based on the Jarque-Bera test for normality, we reject the null hypothesis of Gaussian reduced form shocks for one out of three shocks in the first block and for two out of two shocks in the second block. For our estimated structural shocks, we reject the null hypothesis of Gaussian structural shocks for two out of three shocks in the first block and two out of two shocks in the second block.

Table 2: Non-Gaussianity of shocks

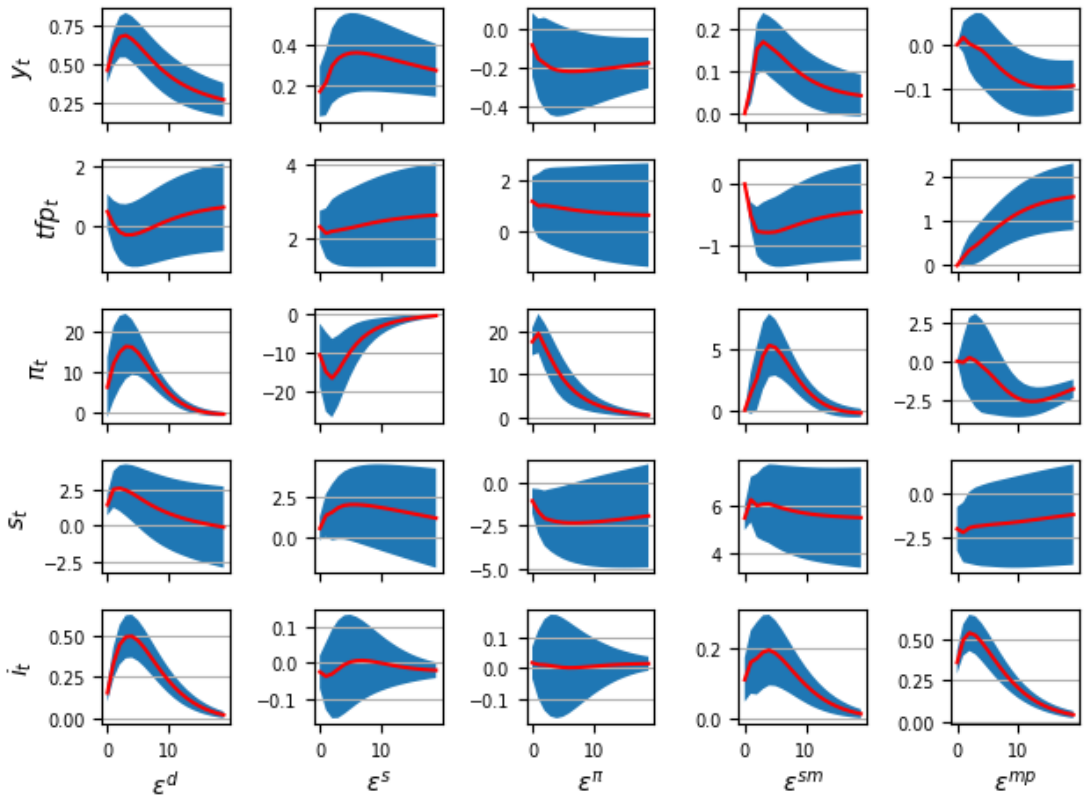
	u^y	u^{tfp}	u^π	u^s	u^i	ε^1	ε^2	ε^3	ε^4	ε^5
Skewness	−0.59	0.15	−0.03	−0.7	−1.16	−0.51	0.48	−0.25	−0.61	−0.51
Kurtosis	4.72	3.5	2.72	4.65	11.23	5.17	4.07	2.71	4.15	15.12
JB-Test	0	0.36	0.79	0	0	0	0	0.38	0	0

Note: Skewness, kurtosis and the p-value of the Jarque-Bera test for normality.

Figure 4 shows the corresponding impulse response functions (IRFs) to one standard deviation shocks.

[18] As a robustness-check, we repeated the cross-validation with a maximum value of λ for which all moments generated by $\mathbf{D}$ for which $\omega_j / \sum_{k \in \widetilde{D}} \omega_k > 10^{-4}$ get selected. However, in this case we choose the maximum value of the sequence of λ's and therefore, we extend the range of maximum values for the main results. That being said, with the extended λ sequence the number of selected moments only rises from 25 to 26 moments.

Figure 4: Impulse Responses.



Note: Impulse responses to normalized shocks for the block-recursive SVAR. The rows tfp_t and s_t show cumulative responses. Confidence bands are 68% bands based on standard errors and 1000 replications.

The shocks in the first block are labeled as follows. The first shock leads to an increase of GDP and inflation and is labeled as a demand shock, ε^d. The second shock increases GDP, tfp, and decreases inflation and is, therefore, labeled as a supply shock, ε^s. The third shock increases inflation and decreases GDP in the long-run and is labeled as a cost-push shock, ε^π. We find that stock returns increase in response to the demand and supply shock and decrease in response to the cost-push shock. Moreover, interest rates increase in response to the demand shock. However, we find no significant response of interest rates to the supply and demand shock. In the second block, one shock leads to an increase of interest rates and decrease of GDP and inflation in the

long-run and, therefore, is labeled as the monetary policy shock, ε^{mp}. The remaining shock is labeled as a stock market shock, ε^{sm}. We find that stock returns decrease after the monetary policy shock, while interest rates increase in response to the stock market shock.

6 Conclusion

Imposing a block-recursive structure on the interactions of the shocks in the SVAR allows to derive identfying and overidentfying moment conditions. We exploit that including these overidentfying moment conditions can increase the efficiency of the block-recursive SVAR GMM estimator. For the special case of a recursive SVAR, we prove the Cholesky estimator, which relies only on second-order moment conditions, to be inefficient if the shocks are mean independent and at least one shock has non-zero skewness or non-zero excess kurtosis.

The asymptotically efficient block-recursive SVAR GMM estimator includes all valid higher-order moment conditions. However, some of the valid higher-order moment conditions may be redundant. While redundant moment conditions do neither increase nor decrease the asymptotic variance of the estimator, they inflate the finite sample variance of the estimator. Therefore, we apply the pGMM estimator of Cheng and Liao (2015b) to discriminate between relevant and the remaining moment conditions. A Monte Carlo experiment illustrates that the pGMM estimator only selects relevant moment conditions and improves the finite sample performance compared to the Cholesky estimator and the overidentified block-recursive SVAR GMM estimator. In the empirical illustration, we employ a block-recursive structure with two blocks and analyze the impact of various shocks on the U.S. economy. Our results suggest that the pGMM estimator, selecting only 26 out of the 71 overidentifying conditions, increases finite sample precision.

References

Blanchard, O. J. (1989). A traditional interpretation of macroeconomic fluctuations. *The American Economic Review*, 1146–1164.

Blanchard, O. J. and D. Quah (1993). The dynamic effects of aggregate demand and supply disturbances: Reply. *The American Economic Review 83*(3), 653–658.

Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999). Redundancy of moment conditions. *Journal of econometrics 91*(1), 89–111.

Cheng, X. and Z. Liao (2015a). Select the valid and relevant moments: An information-based lasso for gmm with many moments. *Journal of Econometrics 186*(2), 443–464.

Cheng, X. and Z. Liao (2015b). Select the valid and relevant moments: An information-based lasso for gmm with many moments. *Journal of Econometrics 186*(2), 443–464.

Defferrard, M., R. Pena, and N. Perraudin (2017, Dec). PyUNLocBoX: Optimization by Proximal Splitting.

Fernald, J. (2009). Utilization-adjusted quarterly-tfp series for the us business sector. In *Federal Reserve Bank of San Francisco*.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2017). Statistical Inference for Independent Component Analysis: Application to Structural VAR Models. *Journal of Econometrics 196*(1), 111–126.

Guay, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics 225*(1), 27–46.

Hall, A. R. (2005). *Generalized method of moments*. Oxford university press.

Hall, A. R. (2015). Econometricians have their moments: Gmm at 32. *Economic Record 91*, 1–24.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics 14*(3), 262–280.

Keweloh, S., S. Hetzenecker, and A. Seepe (2021). Block-recursive non-gaussian structural vector autoregressions.

Keweloh, S. A. (2021a). A feasible approach to incorporate information in higher moments in structural vectorautoregressions. *Discussion papers SFB 823*.

Keweloh, S. A. (2021b). A generalized method of moments estimator for structural vector autoregressions based on higher moments. *Journal of Business & Economic Statistics 39*(3), 772–782.

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

Lanne, M., K. Liu, and J. Luoto (2021). Identifying structural vector autoregression via large economic shocks. *Available at SSRN 3910532*.

Lanne, M. and J. Luoto (2021). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics 39*(1), 69–81.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control 34*(2), 121–131.

Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and Estimation of Non-Gaussian Structural Vector Autoregressions. *Journal of Econometrics 196*(2), 288–304.

Lewis, D. J. (2021, 05). Identifying Shocks via Time-Varying Volatility. *The Review of Economic Studies*. rdab009.

Lütkepohl, H. and A. Netšunajev (2017). Structural vector autoregressions with smooth transition in variances. *Journal of Economic Dynamics and Control 84*, 43–57.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics 85*(4), 777–792.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Wainwright, M. J. (2006). Sharp thresholds for sparsity recovery in the noisy and high-dimensional setting using ℓ_1-constrained quadratic programming. *44th Annual Allerton Conference on Communication, Control, and Computing 2006 1*(5), 217–224.

Wu, J. C. and F. D. Xia (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking 48*(2-3), 253–291.

Zha, T. (1999). Block recursion and structural vector autoregressions. *Journal of Econometrics 90*(2), 291–316.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101*(476), 1418–1429.

Appendix

A Supplementary Proofs

A.1 Asymptotic variance of the (overidentified) block-recursive SVAR GMM estimator

For completeness, we state the formulas of the asymptotic variances of the estimators in Section 3 here. The asymptotic variance of the block-recursive SVAR GMM estimator defined in Equation (18) is given by

$$V_{\mathbf{N}} := M_{\mathbf{N}} S_{\mathbf{N}} M_{\mathbf{N}}' \tag{A.1}$$

where

$$M_{\mathbf{N}} := \left(G_{\mathbf{N}}' W_{\mathbf{N}} G_{\mathbf{N}}\right)^{-1} G_{\mathbf{N}}' W_{\mathbf{N}}, \qquad S_{\mathbf{N}} := \lim_{T\to\infty} E\left[T g_{\mathbf{N}}(B_0) g_{\mathbf{N}}(B_0)\right],$$
$$G_{\mathbf{N}} := E\left[\frac{\partial f_{\mathbf{N}}(B_0, u_t)}{\partial vec(B)'}\right].$$

Consequently, using the weighting matrix $W_{\mathbf{N}}^* := S_{\mathbf{N}}^{-1}$ leads to the estimator $\hat{B}^*$ with the asymptotic variance

$$V_{\mathbf{N}}^* := (G_{\mathbf{N}}' S_{\mathbf{N}}^{-1} G_{\mathbf{N}})^{-1}, \tag{A.2}$$

which is the lowest possible asymptotic variance, see Hall (2005).

Similarly, the asymptotic variance of the overidentified block-recursive SVAR GMM estimator defined in Equation (22) is given by

$$V_{\mathbf{N+D}} := M_{\mathbf{N+D}} S_{\mathbf{N+D}} M_{\mathbf{N+D}}', \tag{A.3}$$

where

$$M_{\mathbf{N+D}} := \left(G_{\mathbf{N+D}}' W_{\mathbf{N+D}} G\right)^{-1} G_{\mathbf{N+D}}' W_{\mathbf{N+D}}, \qquad S_{\mathbf{N+D}} := \lim_{T\to\infty} E\left[g_{\mathbf{N+D}}(B_0) g_{\mathbf{N+D}}(B_0)'\right],$$
$$G_{\mathbf{N+D}} := \begin{bmatrix} G_{\mathbf{N}} \\ G_{\mathbf{D}} \end{bmatrix}, \qquad g_{\mathbf{N+D}}(B_0) := \begin{bmatrix} g_{\mathbf{N}}(B_0) \\ g_{\mathbf{D}}(B_0) \end{bmatrix},$$
$$G_{\mathbf{D}} := E\left[\frac{\partial f_{\mathbf{D}}(B_0, u_t)}{\partial vec(B)'}\right].$$

Using the weighting matrix $W_{\mathbf{N+D}}^* := S_{\mathbf{N+D}}^{-1}$ leads to the estimator $\hat{B}_{\mathbf{N+D}}^*$ with the asymptotic variance

$$V_{\mathbf{N+D}}^* := (G_{\mathbf{N+D}}' S_{\mathbf{N+D}}^{-1} G_{\mathbf{N+D}})^{-1}, \tag{A.4}$$

which is the lowest possible asymptotic variance, see Hall (2005). To construct $V_{\mathbf{N+D}_j}$ and $V_{\mathbf{N+D}_j}^*$, $j \in \widetilde{D}$, we replace the moment conditions $f_{\mathbf{D}_j}(B, u_t)$ by moment condition $f_{\mathbf{D}_j}(B, u_t)$, $j \in \widetilde{D}$, in Equation (A.3) and (A.4).

A.2 Choice of maximum λ in the cross-validation

In the following, we illustrate how to choose the maximum value of λ in the cross-validation. Define the loss function of the pGMM estimator as

$$L^*(B,\beta) := L(B,\beta) + \lambda \sum_{i\in\widetilde{D}} \omega_i \, |\beta_i|, \tag{A.5}$$

where $L(B,\beta) := \begin{bmatrix} g_{\mathbf{N}}(B) \\ g_{\mathbf{D}}(B,\beta) \end{bmatrix}' W \begin{bmatrix} g_{\mathbf{N}}(B) \\ g_{\mathbf{D}}(B,\beta) \end{bmatrix}$.

Further, let $z \in \partial||\beta||_1$, where $z \in \mathbb{R}^{k_T-k_0}$, denote the subgradient for the ℓ_1-norm evaluated at β, i.e.,

$$\begin{aligned} z_i &= sign(\beta_i), \text{ if } \beta_i \neq 0, \\ z_i &\in [-1,1], \quad \text{if } \beta_i = 0, \end{aligned} \tag{A.6}$$

for $i = 1,\ldots,k_T-k_0$ (Wainwright (2006)). Then, the first order condition of the pGMM estimator with respect to β_i, $i = 1,\ldots,k_T-k_0$, evaluated at β and B is

$$\frac{\partial L^*(B,\beta)}{\partial \beta_i} = \frac{\partial L(B,\beta)}{\partial \beta_i} + \lambda \omega_i \, z_i = 0 \tag{A.7}$$

Note that $\omega_i \geq 0$. However, if $\omega_i = 0$, β_i is not penalized and therefore, we only consider $i \in \widetilde{P} := \{j \in \tilde{D} |\ \omega_j > 0\}$ for which, by definition, $\omega_i > 0$ when choosing the maximum value of λ in the cross-validation. By (A.6) and (A.7), $\beta = \mathbf{0} = (0,\ldots,0)'$ and $B = B_0$ minimize the loss function in (A.5) only if

$$\frac{1}{\omega_i} \frac{\partial L(B_0,\mathbf{0})}{\partial \beta_i} \in \lambda[-1,1],$$

for $i \in \widetilde{P}$. Thus,

$$\max_{i\in\widetilde{P}} \left| \frac{1}{\omega_i} \frac{\partial L(B_0,\mathbf{0})}{\partial \beta_i} \right| \leq \lambda.$$

This motivates us to use

$$\lambda_{\max} = \max_{i\in\widetilde{P}} \left| \frac{1}{\omega_i} \frac{\partial L(B_0,\mathbf{0})}{\partial \beta_i} \right|.$$

as the largest value in the cross-validation. Note that any $\lambda > \lambda_{\max}$ would not have an effect on β as $\lambda_{\max}$ already shrinks all elements of β to zero. In practice, we replace B_0 and ω_i by consistent estimators to obtain $\lambda_{\max}$. Furthermore, we consider a weight ω_j to be positive and hence, $j \in \widetilde{P}$, if $\omega_j / \sum_{k\in\widetilde{D}} \omega_k > 10^{-4}$.

A.3 Proof of Asymptotic Variance of pGMM Estimator

We show how to derive the asymptotic variance of the pGMM estimator, $V_{\mathbf{N}+\mathbf{A}}$, based on Remark 3.5 of Cheng and Liao (2015a). We first show Lemma 1 and then apply the result in Remark 3.5 of Cheng and Liao (2015a). Denote $\mathbf{U} := \mathbf{I} \cup \mathbf{R}$ as the moments in either $\mathbf{I}$ or $\mathbf{R}$ and the number of moments in $\mathbf{U}$ by k_U. Further, define the number of elements in $vec(B)$ as d_B. In the proof of Lemma 1, we use the index sets $1 \equiv \mathbf{N}+\mathbf{A}$, $2 \equiv (\mathbf{N}+\mathbf{A},\mathbf{U})$, $3 \equiv (\mathbf{U},\mathbf{N}+\mathbf{A})$, and $4 \equiv \mathbf{U}$ to keep notation uncluttered. Let $\iota^* = (\iota', \mathbf{0}'_{d_U})'$ where $\iota = (1,\ldots,1)'$ is a $d_B \times 1$ vector, i.e., $\iota^{*\prime} A \iota^{*\prime}$ gives the leading $d_B \times d_B$-upper west block of an arbitrary $(d_B+d_U)\times(d_B+d_U)$ matrix A.

Lemma 1.

$$\iota^{*\prime}\left(\Gamma' W \Gamma\right)^{-1}\left(\Gamma' W S_{\mathbf{N+D}} W \Gamma\right)\left(\Gamma' W \Gamma\right)^{-1}\iota^{*} = V_{\mathbf{N+A}},$$

where

$$\Gamma := \begin{bmatrix} G_{\mathbf{N+A}} & \mathbf{0}_{(k_T-k_U)\times k_U} \\ G_{\mathbf{U}} & -I_{k_U} \end{bmatrix}, \qquad V_{\mathbf{N+A}} := M_{\mathbf{N+A}} S_{\mathbf{N+A}} M'_{\mathbf{N+A}}$$

with

$$M_{\mathbf{N+A}} := \left(G'_{\mathbf{N+A}} W^{pi}_{\mathbf{N+A}} G_{\mathbf{N+A}}\right)^{-1} G'_{\mathbf{N+A}} W^{pi}_{\mathbf{N+A}}, \qquad S_{\mathbf{N+A}} := \lim_{T\to\infty} E\left[g_{\mathbf{N+A}}(B_0) g_{\mathbf{N+A}}(B_0)'\right],$$

$$G_{\mathbf{N+A}} := \begin{bmatrix} G_{\mathbf{N}} \\ G_{\mathbf{A}} \end{bmatrix}, \qquad W^{pi}_{\mathbf{N+A}} := \left(W_{\mathbf{N+A}} - W_{\mathbf{N+A},\mathbf{I\cup R}} W^{-1}_{\mathbf{I\cup R}} W_{\mathbf{I\cup R},\mathbf{N+A}}\right),$$

$$G_{\mathbf{A}} := E\left[\frac{\partial f_{\mathbf{A}}(B_0, u_t)}{\partial vec(B)'}\right], \qquad W_{\mathbf{N+D}} := \begin{bmatrix} W_{\mathbf{N+A}} & W_{\mathbf{N+A},\mathbf{I\cup R}}, \\ W_{\mathbf{I\cup R},\mathbf{N+A}} & W_{\mathbf{I\cup R}} \end{bmatrix},$$

$$W_{\mathbf{N+A}} \in \mathbb{R}^{(k_{\mathbf{N}}+k_{\mathbf{A}})\times(k_{\mathbf{N}}+k_{\mathbf{A}})}, \qquad W_{\mathbf{N+A},\mathbf{I\cup R}} \in \mathbb{R}^{(k_{\mathbf{N}}+k_{\mathbf{A}})\times(k_{\mathbf{D}}-k_{\mathbf{A}})},$$

$$W_{\mathbf{I\cup R},\mathbf{N+A}} = W'_{\mathbf{N+A},\mathbf{I\cup R}}, \qquad W_{\mathbf{I\cup R}} \in \mathbb{R}^{(k_{\mathbf{D}}-k_{\mathbf{A}})\times(k_{\mathbf{D}}-k_{\mathbf{A}})}.$$

Proof. Recall that $G_{\mathbf{N+A}}$ and $G_{\mathbf{U}}$ have dimension $(k_T - k_U) \times d_B$ and $k_U \times d_B$, respectively. We define

$$L := \begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} := \left(\Gamma' W \Gamma\right)^{-1}.$$

Additionally, let

$$N := \begin{bmatrix} N_1 & N_2 \\ N_3 & N_4 \end{bmatrix} := \left(\Gamma' W S_{\mathbf{N+D}} W \Gamma\right),$$

and denote the inverse of W by

$$W^{ipi} := \begin{bmatrix} W_1^{ipi} & W_2^{ipi} \\ W_3^{ipi} & W_4^{ipi} \end{bmatrix} := W^{-1} = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix}^{-1}.$$

Let $W_1^{pi} := \left(W_1 - W_2 W_4^{-1} W_3\right)$. Then, by the partitioned inverse, $W_1^{ipi} := \left(W_1^{pi}\right)^{-1}$. By similar arguments as leading to (2.18) in the Online Appendix of Cheng and Liao (2015b), we get that

$$L_1 = \left(G_1'\left(W_1 - W_2 W_4^{-1} W_3\right) G_1\right)^{-1} = \left(G_1' W_1^{pi} G_1\right)^{-1}$$

and, by using the partitioned inverse formula again, and similar arguments as leading to (2.10), (2.11) and (2.18) in the Online Appendix of Cheng and Liao (2015b), that

$$\begin{aligned} L_3 &= -W_4^{-1}\left(-G_1' W_2 - G_4' W_4\right)' \left(G_1' W_1^{pi} G_1\right)^{-1} \\ &= \left(W_4^{-1} W_3 G_1 + G_4\right) L_1 \\ &= X L_1, \end{aligned} \tag{A.8}$$

where we used that $W_4' = W_4$, $W_3 = W_2'$ and $X := \left(W_4^{-1} W_3 G_1 + G_4\right)$. Further, let

$$H := \begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \end{bmatrix} := W S_{\mathbf{N+D}} W,$$

where

$$\begin{aligned}
H_1 &:= W_1 S_1 W_1 + W_2 S_3 W_1 + W_1 S_2 W_3 + W_2 S_4 W_3 \\
H_2 &:= W_1 S_1 W_2 + W_2 S_3 W_2 + W_1 S_2 W_4 + W_2 S_4 W_4 \\
H_3 &:= W_3 S_1 W_1 + W_4 S_3 W_1 + W_3 S_2 W_3 + W_4 S_4 W_3 \\
H_4 &:= W_3 S_1 W_2 + W_4 S_3 W_2 + W_3 S_2 W_4 + W_4 S_4 W_4.
\end{aligned}$$

Note that $H_3 = H_2'$ since $W_3 = W_2'$, $W_1 = W_1'$, $W_4 = W_4'$, $S_3 = S_2'$, $S_1 = S_1'$ and $S_4 = S_4'$. Hence, similar to (2.11) in the Online Appendix of Cheng and Liao (2015b),

$$\begin{aligned}
N_1 &= G_1' H_1 G_1 + G_4' H_3 G_1 + G_1' H_2 G_4 + G_4' H_4 G_4 \\
&= G_1' H_1 G_1 + G_4' H_2' G_1 + G_1' H_2 G_4 + G_4' H_4 G_4 \\
N_2 &= -G_1' H_2 - G_4' H_4 \\
N_3 &= N_2' \\
N_4 &= H_4.
\end{aligned}$$

Then,

$$\begin{aligned}
\iota^{*\prime} \left(\Gamma' W \Gamma\right)^{-1} \left(\Gamma' W S_{\mathbf{N+D}} W \Gamma\right) \left(\Gamma' W \Gamma\right)^{-1} \iota^* &= \iota^{*\prime} L N L \iota^* \\
&= L_1 N_1 L_1 + L_2 N_3 L_1 + L_1 N_2 L_3 + L_2 N_4 L_3 \\
&= L_1 N_1 L_1 + L_3' N_3 L_1 + L_1 N_2 L_3 + L_3' N_4 L_3 \\
&\overset{(A.8)}{=} L_1 N_1 L_1 + L_1' X' N_2' L_1 + L_1 N_2 X L_1 + L_1' X' N_4 X L_1 \\
&= L_1 \left(N_1 + X' N_2' + N_2 X + X' N_4 X\right) L_1, \qquad \text{(A.9)}
\end{aligned}$$

where we used that $L_1' = L_1$, $L_3' = L_2$, and $N_3' = N_2$.

Next, define $Y := N_1 + X' N_2' + N_2 X + X' N_4 X$. Then, multiplying out gives

$$\begin{aligned}
Y =& G_1' H_1 G_1 + G_4' H_3 G_1 + G_1' H_2 G_4 + G_4' H_4 G_4 + \left(G_1' W_2 W_4^{-1} + G_4'\right)\left(-H_2' G_1 - H_4' G_4\right) \\
&+ \left(-G_1' H_2 - G_4' H_4\right)\left(W_4^{-1} W_2' G_1 + G_4\right) + \left(G_1' W_2 W_4^{-1} + G_4'\right) H_4 \left(W_4^{-1} W_2' G_1 + G_4\right) \\
=& G_1' W_2 W_4^{-1} H_4 W_4^{-1} W_2' G_1 + G_1' H_1 G_1 - G_1' W_2 W_4^{-1} H_2' G_1 - G_1' H_2 W_4^{-1} W_2' G_1 \\
=& G_1' \left(W_2 W_4^{-1} H_4 W_4^{-1} W_2' + H_1 - W_2 W_4^{-1} H_2' - H_2 W_4^{-1} W_2'\right) G_1 \\
=& G_1' \left(W_2 W_4^{-1} W_3 S_1 W_2 W_4^{-1} W_3 + W_1 S_1 W_1 - W_2 W_4^{-1} W_3 S_1 W_1 - W_1 S_1 W_2 W_4^{-1} W_3\right) G_1 \\
=& G_1' \left(W_1 - W_2 W_4^{-1} W_3\right) S_1 \left(W_1 - W_2 W_4^{-1} W_3\right) G_1 \\
=& G_1' W_1^{pi} S_1 W_1^{pi} G_1 \qquad \text{(A.10)}
\end{aligned}$$

Plugging (A.10) into (A.9), we obtain

$$\begin{aligned}
&\iota^{*\prime}\left(\Gamma^{\prime}W\Gamma\right)^{-1}\left(\Gamma^{\prime}WS_{\mathbf{N+D}}W\Gamma\right)\left(\Gamma^{\prime}W\Gamma\right)^{-1}\iota^{*} \\
&\quad = L_1\left(G_1^{\prime}W_1^{pi}S_1W_1^{pi}\right)G_1L_1 \\
&\quad = \left(G_1^{\prime}W_1^{pi}G_1\right)^{-1}\left(G_1^{\prime}W_1^{pi}S_1W_1^{pi}G_1\right)\left(G_1^{\prime}W_1^{pi}G_1\right)^{-1} \\
&\quad = \left(G_{\mathbf{N+A}}^{\prime}W_{\mathbf{N+A}}^{pi}G_{\mathbf{N+A}}\right)^{-1}\left(G_{\mathbf{N+A}}^{\prime}W_{\mathbf{N+A}}^{pi}S_{\mathbf{N+A}}W_{\mathbf{N+A}}^{pi}G_{\mathbf{N+A}}\right)\left(G_{\mathbf{N+A}}^{\prime}W_{\mathbf{N+A}}^{pi}G_{\mathbf{N+A}}\right)^{-1}
\end{aligned}$$

which was to show. □

Note that in the following proposition, we treat k_T as a fixed constant to keep our asymptotic results for the pGMM estimator in line with the asymptotic results for the block-recursive SVAR GMM estimator in Equation (22). Cheng and Liao (2015b) allow k_T to increase with the sample size. However, their results also hold when the number of moment conditions is fixed.

Proposition 2. *Assume that the Assumptions in Theorem 3.3 of Cheng and Liao (2015b) hold. Further, assume that* $E\left[\frac{\partial f_{\mathbf{A}}(B_0,u_t)}{\partial vec(B)^{\prime}}\right] = \frac{\partial E[f_{\mathbf{A}}(B_0,u_t)]}{\partial vec(B)^{\prime}}$ *and that the assumptions needed to derive the asymptotic distribution used to derive the asymptotic distribution of the block-recursive SVAR GMM estimator in Equation* (22) *hold. Then,*

$$\sqrt{T}\left(vec(\hat{B}_{\mathbf{N+D}}) - vec(B_0)\right) \xrightarrow{d} \mathcal{N}(0, V_{\mathbf{N+A}})$$

Proof. Define $\Sigma_{CL} := \left(\Gamma^{\prime}W\Gamma\right)^{-1}\left(\Gamma^{\prime}WS_{\mathbf{N+D}}W\Gamma\right)\left(\Gamma^{\prime}W\Gamma\right)^{-1}$ and $\gamma = \left(\nu^{\prime}, \mathbf{0}_{d_U}^{\prime}\right)^{\prime}$ where $\nu \in \mathbb{R}^{d_B}$ is an arbitrary vector. Then, by Remark 3.5 of Cheng and Liao (2015b) (and fixing a typo in this remark),

$$\left|\left|\Sigma_{CL}^{1/2}\,\gamma\right|\right|^{-1/2}\sqrt{T}\nu^{\prime}\left(vec(\hat{B}_{\mathbf{N+D}}) - vec(B_0)\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where $||a|| := a^{\prime}a$ is the squared ℓ_2-norm of an arbitrary vector a.

Note that Lemma 1 immediately implies $\left|\left|\Sigma_{CL}^{1/2}\gamma\right|\right| = \gamma^{\prime}\Sigma_{CL}\gamma = \nu^{\prime}V_{\mathbf{N+A}}(W)\nu$. Hence,

$$\left|\left|V_{\mathbf{N+A}}(W)^{1/2}\nu\right|\right|^{-1/2}\sqrt{T}\nu^{\prime}\left(vec(\hat{B}_{\mathbf{N+D}}) - vec(B_0)\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where $V_{\mathbf{N+A}}(W)$ is the asymptotic variance of $vec(\hat{B}_{\mathbf{N+D}})$ since it holds that

$$\nu^{*\prime}V_{\mathbf{N+A}}(W)\nu^{*} = \left|\left|V_{\mathbf{N+A}}(W)^{1/2}\nu\right|\right|^{-1}\nu^{\prime}V_{\mathbf{N+A}}(W)\nu = 1$$

where $\nu^{*} := \left|\left|V_{\mathbf{N+A}}(W)^{1/2}\nu\right|\right|^{-1/2}\nu$.

Consequently, using the Cramér-Wold device, we get

$$\sqrt{T}\left(vec(\hat{B}_{\mathbf{N+D}}) - vec(B_0)\right) \xrightarrow{d} \mathcal{N}(0, V_{\mathbf{N+A}}).$$

□

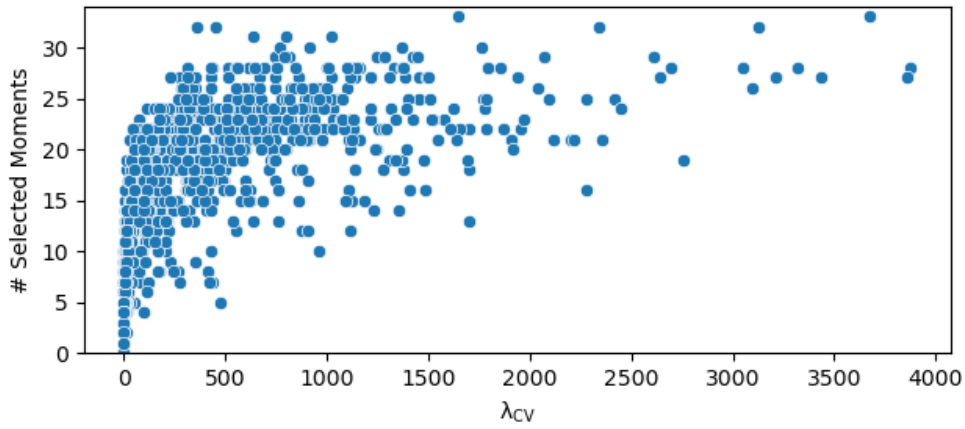
B Supplementary Tables

Table B.1: Finite sample performance including Post-LASSO

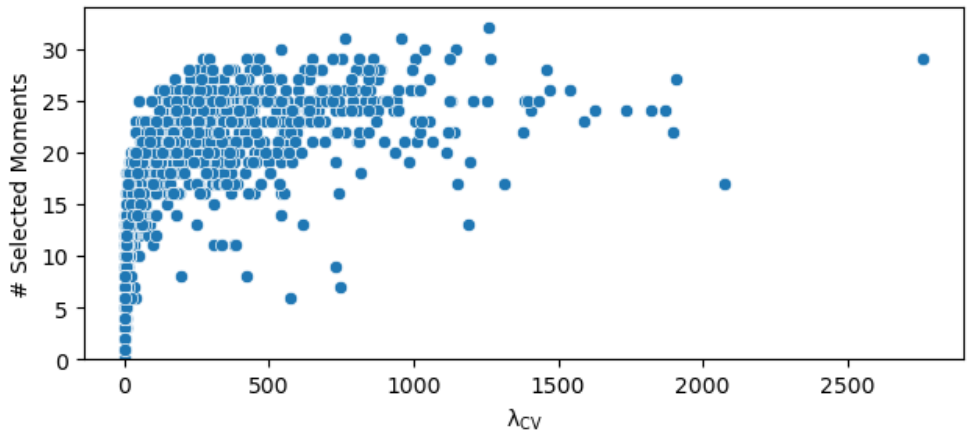
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 100$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.16) & & & \\ 0.50 & 0.99 & . & . \\ (1.26) & (1.13) & & \\ 0.50 & 0.49 & 0.99 & . \\ (1.51) & (1.28) & (1.08) & \\ 0.50 & 0.50 & 0.49 & 0.98 \\ (1.72) & (1.47) & (1.26) & (1.09) \end{bmatrix}$	$\begin{bmatrix} 0.98 & . & . & . \\ (1.08) & & & \\ 0.48 & 0.97 & . & . \\ (1.46) & (1.13) & & \\ 0.49 & 0.48 & 0.97 & . \\ (1.66) & (1.52) & (1.12) & \\ 0.49 & 0.48 & 0.49 & 0.97 \\ (2.01) & (1.87) & (1.44) & (1.12) \end{bmatrix}$	$\begin{bmatrix} 0.98 & . & . & . \\ (1.07) & & & \\ 0.49 & 0.97 & . & . \\ (1.28) & (1.13) & & \\ 0.49 & 0.48 & 0.97 & . \\ (1.51) & (1.35) & (1.12) & \\ 0.49 & 0.49 & 0.49 & 0.97 \\ (1.75) & (1.56) & (1.25) & (1.13) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.16) & & & \\ 0.50 & 0.99 & . & . \\ (1.20) & (1.10) & & \\ 0.51 & 0.49 & 0.99 & . \\ (1.44) & (1.27) & (1.12) & \\ 0.50 & 0.50 & 0.50 & 0.99 \\ (1.68) & (1.49) & (1.26) & (1.10) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (1.10) & & & \\ 0.50 & 0.98 & . & . \\ (1.16) & (1.11) & & \\ 0.50 & 0.49 & 0.98 & . \\ (1.39) & (1.23) & (1.10) & \\ 0.50 & 0.49 & 0.49 & 0.98 \\ (1.59) & (1.40) & (1.20) & (1.08) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	23.99	23.99
	Bias	−0.0526	−0.1998	−0.2005	−0.0331	−0.1114
	Var	12.97	14.42	13.15	12.83	12.36
	λ	.	.	.	103.91	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 200$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.10) & & & \\ 0.50 & 1.00 & . & . \\ (1.22) & (1.15) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.42) & (1.22) & (1.13) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.76) & (1.54) & (1.22) & (1.15) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (1.02) & & & \\ 0.49 & 0.99 & . & . \\ (1.19) & (1.08) & & \\ 0.49 & 0.49 & 0.99 & . \\ (1.47) & (1.26) & (1.04) & \\ 0.49 & 0.49 & 0.49 & 0.99 \\ (1.82) & (1.58) & (1.21) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (1.02) & & & \\ 0.50 & 0.99 & . & . \\ (1.10) & (1.09) & & \\ 0.49 & 0.49 & 0.99 & . \\ (1.34) & (1.11) & (1.05) & \\ 0.49 & 0.49 & 0.49 & 0.99 \\ (1.60) & (1.39) & (1.05) & (1.09) \end{bmatrix}$	$\begin{bmatrix} 1.01 & . & . & . \\ (1.09) & & & \\ 0.50 & 1.00 & . & . \\ (1.08) & (1.12) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.34) & (1.13) & (1.07) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.59) & (1.40) & (1.08) & (1.09) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.04) & & & \\ 0.50 & 0.99 & . & . \\ (1.04) & (1.11) & & \\ 0.50 & 0.50 & 0.99 & . \\ (1.28) & (1.10) & (1.07) & \\ 0.50 & 0.50 & 0.49 & 0.99 \\ (1.55) & (1.36) & (1.07) & (1.10) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	26.37	26.37
	Bias	−0.0222	−0.0846	−0.0848	0.0030	−0.0508
	Var	12.93	12.77	11.84	11.99	11.72
	λ	.	.	.	114.89	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 500$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.06) & & & \\ 0.50 & 1.00 & . & . \\ (1.25) & (1.05) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.54) & (1.23) & (1.10) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.89) & (1.47) & (1.24) & (1.17) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.99) & & & \\ 0.50 & 1.00 & . & . \\ (1.23) & (0.98) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.42) & (1.15) & (1.01) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.73) & (1.34) & (1.14) & (1.03) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.00) & & & \\ 0.50 & 1.00 & . & . \\ (1.10) & (0.98) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.29) & (1.06) & (1.01) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.55) & (1.27) & (1.06) & (1.03) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.04) & & & \\ 0.50 & 1.00 & . & . \\ (1.10) & (1.00) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.33) & (1.06) & (1.03) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.61) & (1.27) & (1.07) & (1.07) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.03) & & & \\ 0.50 & 1.00 & . & . \\ (1.10) & (1.01) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.32) & (1.07) & (1.04) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.59) & (1.26) & (1.07) & (1.09) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	28.35	28.35
	Bias	−0.0010	−0.0267	−0.0278	0.0083	−0.0153
	Var	13.00	12.02	11.34	11.60	11.59
	λ	.	.	.	90.16	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 1000$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.13) & & & \\ 0.50 & 1.00 & . & . \\ (1.32) & (1.09) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.53) & (1.20) & (1.19) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.85) & (1.69) & (1.29) & (1.21) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.00) & & & \\ 0.50 & 1.00 & . & . \\ (1.12) & (0.99) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.29) & (1.04) & (1.05) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.61) & (1.38) & (1.10) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.00) & & & \\ 0.50 & 1.00 & . & . \\ (1.08) & (0.99) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.23) & (0.99) & (1.04) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.49) & (1.30) & (1.04) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 1.01 & . & . & . \\ (1.06) & & & \\ 0.50 & 1.00 & . & . \\ (1.07) & (1.02) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.24) & (1.01) & (1.09) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.49) & (1.33) & (1.05) & (1.11) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.04) & & & \\ 0.50 & 1.00 & . & . \\ (1.08) & (1.02) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.26) & (1.00) & (1.10) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.51) & (1.34) & (1.06) & (1.13) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	29.10	29.10
	Bias	0.0271	0.0155	0.0165	0.0329	0.0212
	Var	13.51	11.64	11.24	11.47	11.53
	λ	.	.	.	59.32	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 5000$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.21) & & & \\ 0.50 & 1.00 & . & . \\ (1.34) & (1.25) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.62) & (1.37) & (1.22) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.95) & (1.63) & (1.24) & (1.22) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.10) & & & \\ 0.50 & 1.00 & . & . \\ (1.09) & (1.15) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.29) & (1.11) & (1.11) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.57) & (1.33) & (1.08) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.10) & & & \\ 0.50 & 1.00 & . & . \\ (1.08) & (1.15) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.27) & (1.10) & (1.11) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.55) & (1.32) & (1.06) & (1.08) \end{bmatrix}$	$\begin{bmatrix} 1.01 & . & . & . \\ (1.14) & & & \\ 0.50 & 1.00 & . & . \\ (1.09) & (1.19) & & \\ 0.50 & 0.50 & 1.01 & . \\ (1.30) & (1.12) & (1.15) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.58) & (1.35) & (1.08) & (1.13) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (1.13) & & & \\ 0.50 & 1.00 & . & . \\ (1.10) & (1.19) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.33) & (1.14) & (1.15) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.61) & (1.38) & (1.09) & (1.14) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	28.95	28.95
	Bias	0.0297	0.0284	0.0285	0.0319	0.0296
	Var	14.05	11.91	11.83	12.12	12.25
	λ	.	.	.	22.23	.

The table reports the average summary statistics over all Monte Carlo replicates for the recursive SVAR estimator (Cholesky), the continuously updating estimator (CUE), the continuously updating oracle estimator (CUE-oracle), the continuously updating pGMM estimator (CUE-LASSO), and the Post-continuously updating pGMM estimator (CUE-LASSO-Post).

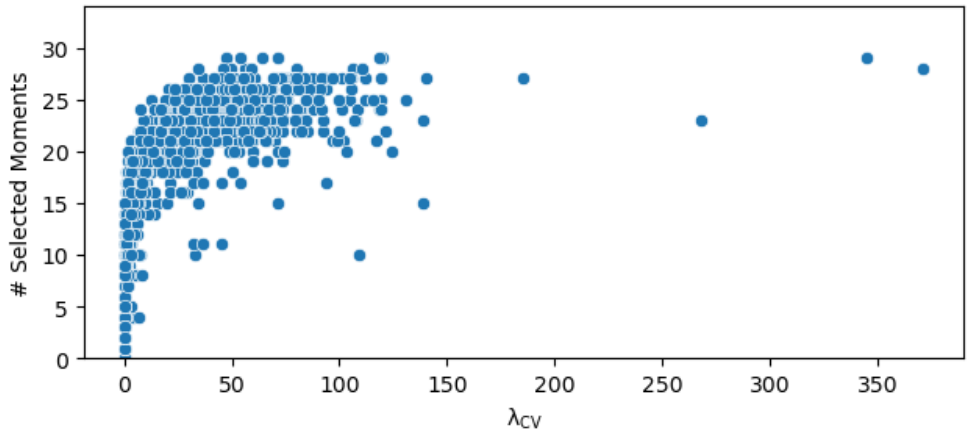
Figure B.5: Relationship of chosen λ_{CV} and Number of Selected Moments across Monte Carlo runs



(a) $T = 100$

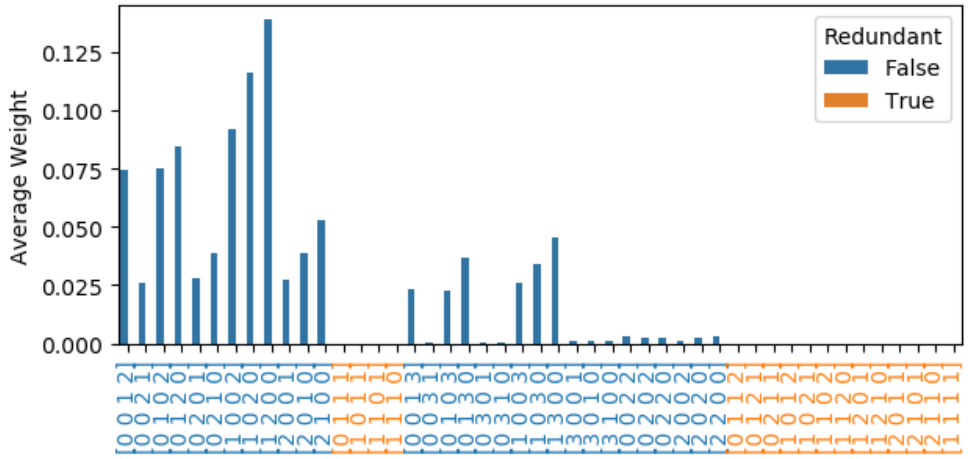


(b) $T = 200$

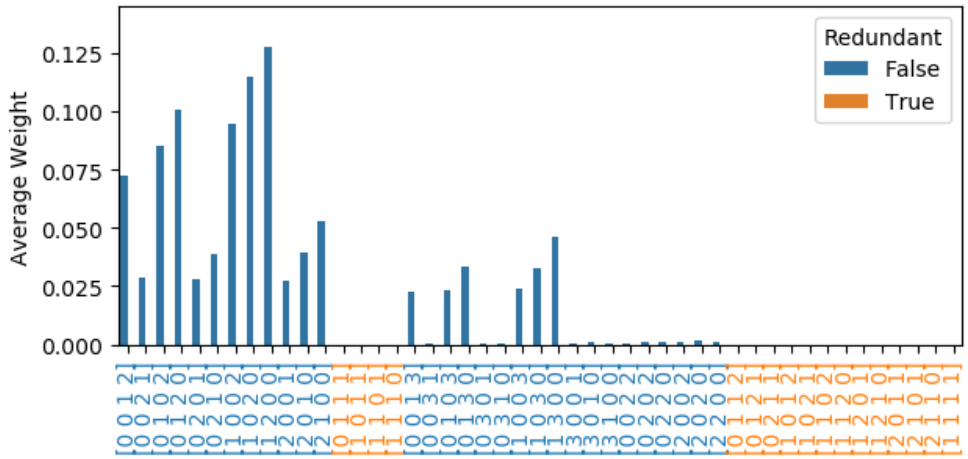


(c) $T = 5000$

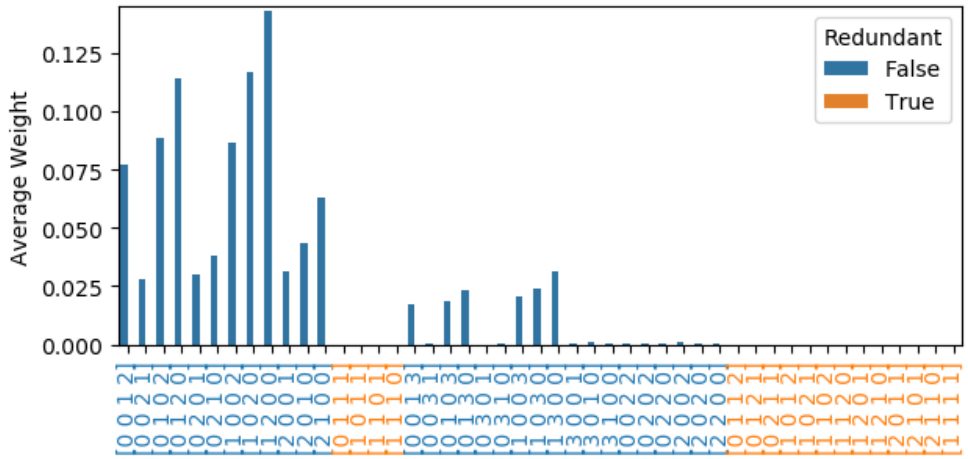
Figure B.6: Average Weight of Moments across Monte Carlo runs



(a) $T = 100$

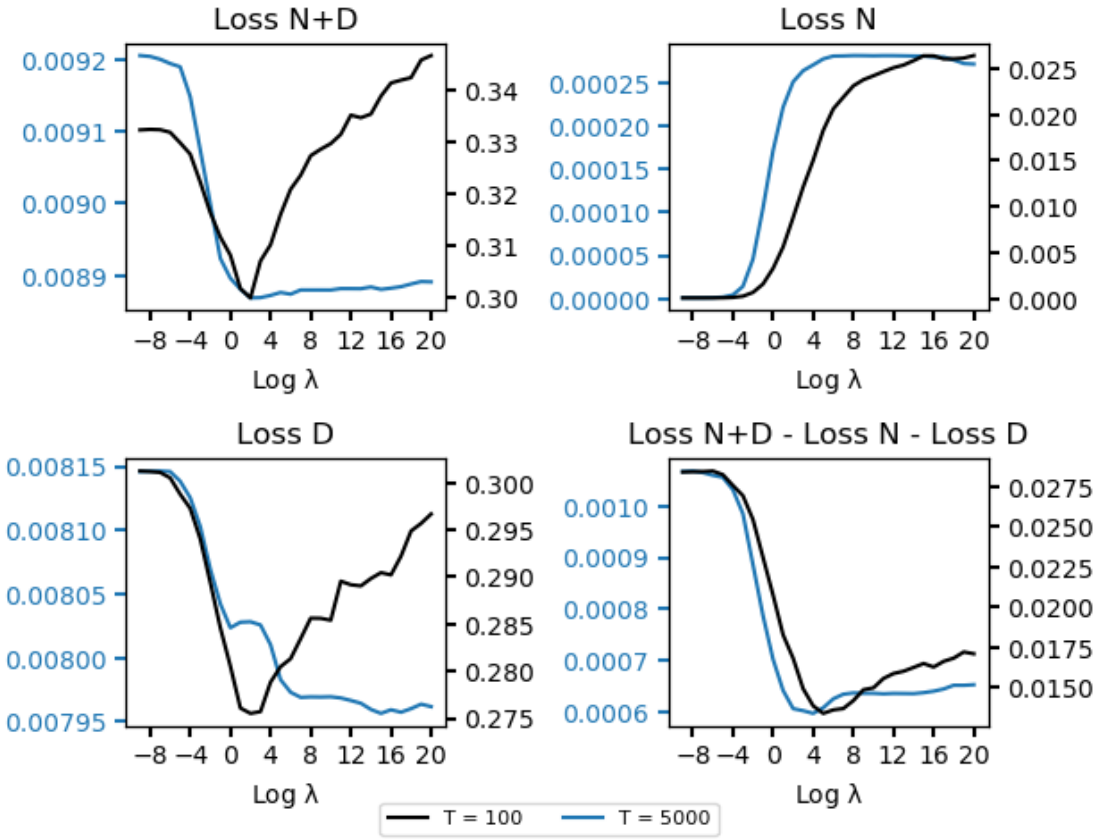


(b) $T = 200$



(c) $T = 5000$

Figure B.7: Relationship of λ and the Loss Function of the CUE-LASSO estimator without β



Note: The figure shows the median of the loss function of the overidentified SVAR GMM estimator in Equation (22) evaluated at the estimates of the CUE-LASSO estimator in dependence on $log(\lambda)$ across 1000 Monte Carlo runs for $T = 100$ (black line, right y-axis) and $T = 5000$ (blue line, left y-axis). For the evaluation, we use the efficient weighting matrix $W^*_{\mathbf{N+D}}$ and denote the blocks of $W^*_{\mathbf{N+D}}$ corresponding to $\widetilde{N}$ and to $\widetilde{D}$ by $W^*_{\mathbf{N}}$ and $W^*_{\mathbf{D}}$, respectively. To be clear, Loss N+D $:= \left[g_{\mathbf{N}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})', \quad g_{\mathbf{D}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})'\right] W^*_{\mathbf{N+D}} \left[g_{\mathbf{N}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})', \quad g_{\mathbf{D}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})'\right]'$, Loss N $:= g_{\mathbf{N}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})' W^*_{\mathbf{N}} g_{\mathbf{N}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})$, and Loss D $:= g_{\mathbf{D}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})' W^*_{\mathbf{D}} g_{\mathbf{D}}(\widehat{B}^{pGMM}_{\mathbf{N+D}})$.

Figure B.8: Weight of each Moment for the block-recursive SVAR in five variables with quarterly U.S. Data.

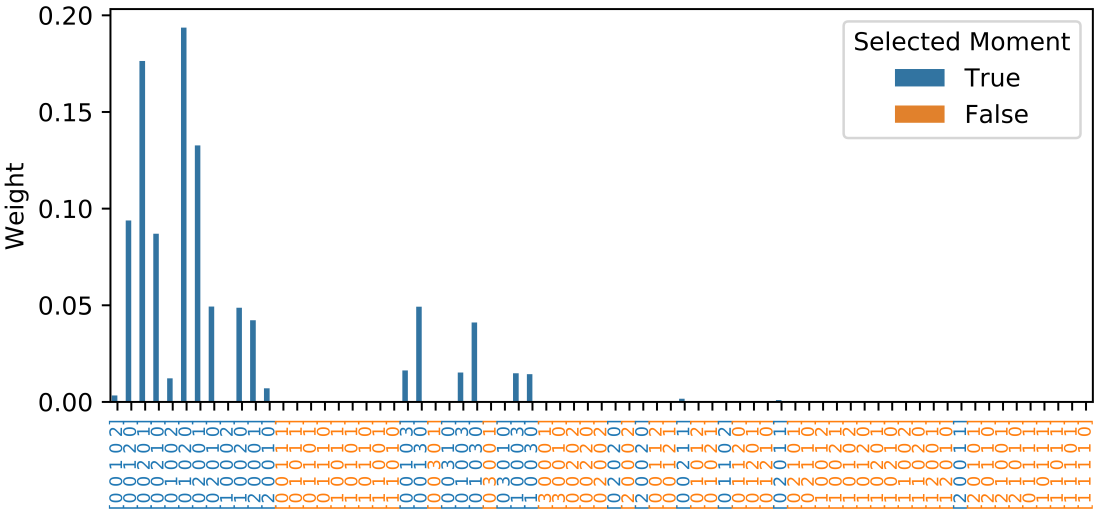


Table B.2: Finite sample performance including Post-LASSO using t-distributed Errors with $df = 9$

		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 100$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.74) & & & \\ 0.50 & 0.99 & . & . \\ (1.24) & (0.75) & & \\ 0.50 & 0.50 & 0.99 & . \\ (1.52) & (1.28) & (0.71) & \\ 0.50 & 0.49 & 0.49 & 0.99 \\ (1.76) & (1.42) & (1.30) & (0.86) \end{bmatrix}$	$\begin{bmatrix} 0.98 & . & . & . \\ (0.77) & & & \\ 0.48 & 0.97 & . & . \\ (1.68) & (0.85) & & \\ 0.48 & 0.48 & 0.97 & . \\ (1.98) & (1.70) & (0.81) & \\ 0.49 & 0.48 & 0.48 & 0.97 \\ (2.19) & (1.91) & (1.60) & (0.93) \end{bmatrix}$	$\begin{bmatrix} 0.98 & . & . & . \\ (0.77) & & & \\ 0.48 & 0.97 & . & . \\ (1.55) & (0.85) & & \\ 0.48 & 0.48 & 0.97 & . \\ (1.75) & (1.47) & (0.82) & \\ 0.49 & 0.48 & 0.48 & 0.97 \\ (2.04) & (1.68) & (1.52) & (0.92) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.80) & & & \\ 0.50 & 0.99 & . & . \\ (1.36) & (0.81) & & \\ 0.50 & 0.49 & 0.99 & . \\ (1.58) & (1.38) & (0.76) & \\ 0.50 & 0.50 & 0.49 & 0.99 \\ (1.87) & (1.59) & (1.41) & (0.89) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (0.76) & & & \\ 0.50 & 0.98 & . & . \\ (1.28) & (0.79) & & \\ 0.50 & 0.49 & 0.98 & . \\ (1.51) & (1.35) & (0.74) & \\ 0.50 & 0.49 & 0.49 & 0.98 \\ (1.80) & (1.53) & (1.34) & (0.88) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	21.10	21.10
	Bias	−0.0517	−0.2015	−0.2128	−0.0547	−0.1003
	Var	11.59	14.42	13.36	12.45	11.97
	λ	.	.	.	65.86	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 200$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.77) & & & \\ 0.50 & 1.00 & . & . \\ (1.20) & (0.85) & & \\ 0.51 & 0.50 & 0.99 & . \\ (1.54) & (1.24) & (0.81) & \\ 0.50 & 0.50 & 0.49 & 0.99 \\ (1.66) & (1.44) & (1.23) & (0.75) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (0.80) & & & \\ 0.50 & 0.99 & . & . \\ (1.45) & (0.86) & & \\ 0.49 & 0.49 & 0.98 & . \\ (1.69) & (1.43) & (0.88) & \\ 0.49 & 0.49 & 0.49 & 0.98 \\ (1.97) & (1.78) & (1.48) & (0.80) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (0.79) & & & \\ 0.50 & 0.99 & . & . \\ (1.32) & (0.86) & & \\ 0.49 & 0.50 & 0.98 & . \\ (1.59) & (1.33) & (0.88) & \\ 0.49 & 0.49 & 0.49 & 0.98 \\ (1.85) & (1.61) & (1.38) & (0.80) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.82) & & & \\ 0.50 & 1.00 & . & . \\ (1.27) & (0.89) & & \\ 0.50 & 0.50 & 0.99 & . \\ (1.54) & (1.29) & (0.84) & \\ 0.50 & 0.50 & 0.49 & 0.99 \\ (1.73) & (1.56) & (1.26) & (0.77) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (0.78) & & & \\ 0.50 & 0.99 & . & . \\ (1.21) & (0.86) & & \\ 0.50 & 0.50 & 0.99 & . \\ (1.49) & (1.25) & (0.83) & \\ 0.50 & 0.50 & 0.49 & 0.99 \\ (1.67) & (1.51) & (1.25) & (0.77) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	21.85	21.85
	Bias	−0.0185	−0.1028	−0.1039	−0.0209	−0.0462
	Var	11.48	13.15	12.41	11.96	11.61
	λ	.	.	.	99.00	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 500$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.78) & & & \\ 0.50 & 1.00 & . & . \\ (1.19) & (0.75) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.47) & (1.16) & (0.81) & \\ 0.49 & 0.50 & 0.50 & 1.00 \\ (1.81) & (1.46) & (1.19) & (0.81) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (0.79) & & & \\ 0.49 & 1.00 & . & . \\ (1.29) & (0.77) & & \\ 0.49 & 0.50 & 0.99 & . \\ (1.57) & (1.21) & (0.84) & \\ 0.49 & 0.50 & 0.50 & 0.99 \\ (1.95) & (1.46) & (1.33) & (0.83) \end{bmatrix}$	$\begin{bmatrix} 0.99 & . & . & . \\ (0.80) & & & \\ 0.49 & 1.00 & . & . \\ (1.21) & (0.77) & & \\ 0.49 & 0.50 & 0.99 & . \\ (1.47) & (1.14) & (0.85) & \\ 0.49 & 0.50 & 0.50 & 0.99 \\ (1.82) & (1.38) & (1.21) & (0.83) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.80) & & & \\ 0.50 & 1.00 & . & . \\ (1.20) & (0.79) & & \\ 0.50 & 0.50 & 0.99 & . \\ (1.44) & (1.13) & (0.83) & \\ 0.49 & 0.50 & 0.50 & 1.00 \\ (1.80) & (1.37) & (1.17) & (0.83) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.78) & & & \\ 0.50 & 1.00 & . & . \\ (1.18) & (0.76) & & \\ 0.50 & 0.50 & 0.99 & . \\ (1.41) & (1.12) & (0.83) & \\ 0.49 & 0.50 & 0.50 & 1.00 \\ (1.77) & (1.36) & (1.17) & (0.82) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	20.96	20.96
	Bias	−0.0177	−0.0505	−0.0511	−0.0188	−0.0273
	Var	11.42	12.05	11.47	11.36	11.20
	λ	.	.	.	102.57	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 1000$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.77) & & & \\ 0.50 & 1.00 & . & . \\ (1.16) & (0.80) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.43) & (1.21) & (0.83) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.69) & (1.51) & (1.15) & (0.81) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.77) & & & \\ 0.50 & 1.00 & . & . \\ (1.16) & (0.80) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.42) & (1.23) & (0.83) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.70) & (1.56) & (1.23) & (0.83) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.77) & & & \\ 0.50 & 1.00 & . & . \\ (1.14) & (0.80) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.37) & (1.17) & (0.83) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.62) & (1.49) & (1.15) & (0.84) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.77) & & & \\ 0.50 & 1.00 & . & . \\ (1.13) & (0.80) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.38) & (1.17) & (0.83) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.60) & (1.48) & (1.13) & (0.83) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.77) & & & \\ 0.50 & 1.00 & . & . \\ (1.13) & (0.80) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.37) & (1.18) & (0.83) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.59) & (1.49) & (1.13) & (0.82) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	19.39	19.39
	Bias	−0.0018	−0.0178	−0.0182	−0.0023	−0.0047
	Var	11.35	11.52	11.18	11.12	11.10
	λ	.	.	.	63.62	.
		Cholesky	CUE	CUE-Oracle	CUE-LASSO	CUE-LASSO-Post
$T = 5000$	$\hat{B}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.73) & & & \\ 0.50 & 1.00 & . & . \\ (1.16) & (0.81) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.35) & (1.21) & (0.76) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.75) & (1.51) & (1.23) & (0.84) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.73) & & & \\ 0.50 & 1.00 & . & . \\ (1.13) & (0.81) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.30) & (1.16) & (0.76) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.69) & (1.45) & (1.20) & (0.85) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.73) & & & \\ 0.50 & 1.00 & . & . \\ (1.12) & (0.81) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.29) & (1.15) & (0.76) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.69) & (1.44) & (1.18) & (0.85) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.74) & & & \\ 0.50 & 1.00 & . & . \\ (1.12) & (0.82) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.30) & (1.16) & (0.76) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.69) & (1.45) & (1.18) & (0.84) \end{bmatrix}$	$\begin{bmatrix} 1.00 & . & . & . \\ (0.73) & & & \\ 0.50 & 1.00 & . & . \\ (1.13) & (0.81) & & \\ 0.50 & 0.50 & 1.00 & . \\ (1.31) & (1.17) & (0.76) & \\ 0.50 & 0.50 & 0.50 & 1.00 \\ (1.70) & (1.47) & (1.19) & (0.84) \end{bmatrix}$
	#Mo	10.00	57.00	40.00	16.06	16.06
	Bias	−0.0016	−0.0052	−0.0050	−0.0016	−0.0019
	Var	11.37	11.08	11.02	11.05	11.09
	λ	.	.	.	1.81	.

The table reports the average summary statistics over all Monte Carlo replicates for the recursive SVAR estimator (Cholesky), the continuously updating estimator (CUE), the continuously updating oracle estimator (CUE-oracle), the continuously updating pGMM estimator (CUE-LASSO), and the Post-continuously updating pGMM estimator (CUE-LASSO-Post).