School of Economics Working Paper 2023-07

UEA University of East Anglia

SCHOOL OF ECONOMICS

Testing for Strong Exogeneity in Proxy-VARS

Martin Bruns *
Sascha A. Keweloh **

* University of East Anglia
** TU Dortmund University

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

Testing for strong exogeneity in Proxy-VARs

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

Sascha A. Keweloh
TU Dortmund University
email: sascha.keweloh@tu-dortmund.de

Abstract. Proxy variables have gained widespread prominence as indispensable tools for identifying structural VAR models. Analogous to instrumental variables, proxies need to be exogenous, i.e. uncorrelated with all non-target shocks. Assessing the exogeneity of proxies has traditionally relied on economic arguments rather than statistical tests. We argue that the economic rational underlying the construction of commonly used proxy variables aligns with a stronger form of exogeneity. Specifically, proxies are typically constructed as variables not containing any information on the expected value of non-target shocks. We show conditions under which this enhanced concept of proxy exogeneity is testable without additional identifying assumptions.

Key Words: Structural vector autoregression, proxy VAR, exogeneity test

JEL classification: C32

1 Introduction

Structural vector autoregressions (SVARs) are routinely combined with external instruments, so-called proxies, to achieve identification. Proxies are valid for identification under two conditions: They need to be relevant, i.e. contemporaneously correlated with the target shock; and they need to be exogenous, i.e. contemporaneously uncorrelated with the non-target shocks.

The traditional conception is that proxy exogeneity cannot be tested using only the proxy variable. Instead, the majority of applications using proxy variables rely on economic arguments to justify the exogeneity condition and not on a statistical test. More recently, several studies propose to test the proxy exogeneity assumption by identifying the model using other identifying assumptions, i.e. using heteroskedasticity in Schlaak, Rieth and Podstawski (2023), sign restrictions in Braun and Brüggemann (2022), independent and non-Gaussian shocks in Keweloh, Klein and Prüser (2023), or changes in unconditional volatility, which can be used to test for proxy exogeneity ex-post in Angelini, Fanelli and Neri (2023). However, these proxy exogeneity tests of course rely on the validity of the chosen identification approach.

In this study, we propose a proxy exogeneity test that exclusively leverages the information embedded in the proxy variable, not relying on the validity of another identification approach. Instead, we argue that the economic reasoning employed to construct proxy variables typically implies a stronger form of exogeneity. Specifically, proxy variables are typically constructed not merely to exhibit no correlation with non-target shocks but to contain no information at all on the expected value of the non-target shocks, which we denote as a strongly exogenous proxy. If we are willing to extend our notion of proxy exogeneity to this stronger exogeneity assumption, a proxy variable can contain information beyond its correlation with the reduced form shocks and this information can be used to detect endogenous proxy variables. Specifically, for a strongly exogenous proxy z_t, we can generate a synthetic proxy $\tilde{z}_t = z_t^2$ which is also exogenous. Therefore, we obtain two proxies and an overidentified system, so that a simple J-test can be used to test the strong exogeneity assumption.

Hence, the crux of our study lies in the expansion of the concept of an exogenous proxy beyond mere uncorrelation with non-target shocks to encompass the absence of any information on the expected value of non-target shocks. Consider, for example, the tax proxy in Mertens and Ravn (2014), carefully constructed as a series of tax shocks based on narrative documents, a monetary policy shock proxy in Gertler and Karadi (2015) constructed as federal funds futures movements around FOMC announcements, or an oil supply news proxy in Känzig (2021), constructed as changes in oil price futures in narrow windows OPEC announcements. In each case, the economic rational underlying the construction of the proxy is that the proxy is a function of the target shock but not affected by non-target shocks. However, this rationale not only justifies the proxy's lack of correlation with non-target shocks but also implies that the does not contain any information on non-target shocks, aligning with the concept of strong exogeneity. For instance, if the tax proxy is constructed as a series of tax shocks, the synthetic tax proxy $\tilde{z}_t = z_t^2$ is equal to a series of squared tax shocks. Consequently, the same economic rational that supports the exogeneity of the tax proxy also justifies exogeneity of the synthetic tax proxy. Moreover, for a linear relationship between proxy and shocks our extended assumption of strong exogeneity directly follows from the traditional assumption of proxy uncorrelatedness. Such linear relationships are often employed in the Bayesian literature, see e.g. Caldara and Herbst (2019), Arias, Rubio-Ramirez and Waggoner (2021), or Giacomini, Kitagawa and Read (2022), but also in frequentist settings, e.g. in Angelini and Fanelli (2019), Bruns and Lütkepohl (2022a), or Angelini, Caggiano, Castelnuovo and Fanelli (2023).

Having established the economic rational justifying the exogeneity of the synthetic proxy, the remaining question is under which circumstances does a strongly exogenous proxy contain sufficient information to detect endogeneity? Put differently, what are the conditions to ensure that the synthetic proxy contains additional information not contained the original proxy? A proxy variable is a function of potentially all structural shocks and additional exogenous variation. If this function is linear and all shocks are Gaussian, a strongly exogenous proxy contains no information beyond its correlation with the reduced form shocks, such that no synthetic proxy can contain any information to detect endogeneity. However, if the proxy does not adhere perfectly to a linear function or if not all shocks influencing the proxy adhere strictly to a Gaussian distribution, a strongly exogenous proxy can contain additional information which can be leveraged to detect endogeneity of the proxy.

First, consider a proxy z_t generated as a linear function of the skewed target shock and a noise term. In this scenario, the information of the proxy is not entirely contained in its correlation with the reduced form shocks, instead, the synthetic proxy $\tilde{z}_t = z_t^2$ will be relevant and can contain information to detect endogeneity. Intuitively, if tax shocks exhibit a left-skewed distribution, implying that large negative tax shocks are more likely than large positive ones, high values of the synthetic tax proxy $\tilde{z}_t = z_t^2$ correlate with negative values of the tax shock, implying that the synthetic tax proxy is correlated with the target shock and hence provides over-identifying restrictions. The same argument can be made if a non-target shock which is correlated with the proxy (a so-called "contaminating shock") exhibits skewness. Second, our approach is not limited to non-Gaussian shocks. Even with exclusively Gaussian shocks, the information of the proxy may not be entirely contained in the second moments if the proxy generating function is non-linear. Only in the special case of exactly Gaussian shocks and a perfectly linear proxy model a synthetic proxy cannot contain additional information such that our test has no power.

We relate to the literature identifying SVARs via independence and non-Gaussianity, see Matteson and Tsay (2017), Gouriéroux, Monfort and Renne (2017), Keweloh (2021), Guay (2021), or Mesters and Zwiernik (2022). The crucial difference is that all of these approaches do not only rely on non-Gaussian shocks, but also require to impose stronger assumptions on the (in)dependencies of the shocks. In contrast, our approach only requires uncorrelated structural shocks. Moreover, identification via non-Gaussianity allows for at most one normally distributed structural shock, while for our test a single (or for a non-linear proxy even no) non-Gaussian shock can be sufficient.

We investigate the finite sample properties of our proposed test in a Monte Carlo study using a stylised and a more realistic data generating process (DGP). We find that our test has a precise nominal level in settings usually encountered in macroeconomic datasets. Its power increases with sample size, with stronger proxies, higher skewness of the target and/ or contaminating shock, stronger forms of proxy non-linearity, and a higher correlation of the proxy with the contaminating shock.

We apply the proposed exogeneity test to three proxy variables frequently utilized in the fiscal proxy SVAR literature. These proxies include the narrative tax variable employed in Mertens and Ravn (2014) as a proxy for tax shocks, the total factor productivity measure from Fernald (2012) as an output shock proxy, and changes in military spending, which have been employed as a government expenditure shock proxy, as seen in Klein and Linnemann (2019), for instance. Our findings reveal a lack of strong exogeneity in the tax proxy, while providing no evidence against exogeneity for the output and spending proxies.

The remainder of this paper is organized as follows: Section 2 discusses the model assumption and develops the proxy exogeneity tests. Section 3 shows the Monte Carlo simulation. Section 4 applies the test to three proxy variables in a fiscal SVAR. Section 5 concludes.

2 Model Setup

Our model is a K-dimensional SVAR(p) process,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \quad \text{with} \quad u_t = B_0 \varepsilon_t \tag{1}$$

where y_t is a $K \times 1$ vector of endogenous variables, u_t is a zero mean white noise process with nonsingular covariance matrix Σ, i.e., $u_t \sim (0, \Sigma)$, ν is an intercept term, and $A_1, ..., A_p$ are the auto-regressive slope coefficients. The structural shocks are denoted by $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$. . They are obtained from the reduced-form errors, u_t, by a linear transformation, $\varepsilon_t = B_0^{-1} u_t$. The $(K \times K)$ matrix B_0 contains the impact effects of the structural shocks and we normalize the diagonal elements of B_0 to one. Without loss of generality let the first shock, ε_{1t}, be the target shock of interest, $\boldsymbol{\varepsilon}_{2t} = (\varepsilon_{2t}, ..., \varepsilon_{Kt})'$ be the vector of the $(K-1)$ non-target shocks, and let $\boldsymbol{u}_{2t} = (u_{2t}, ..., u_{Kt})'$ denote the vector of the last $(K-1)$ reduced form shocks. Moreover, let β_0 be equal to the last $(K-1)$ elements of the first column of B_0, such that β_0 denotes the simultaneous impact of the target shock on the last $(K-1)$ variables and the impact on the first variable is normalized to one.

The reduced form shocks, u_t, can be estimated via OLS. Identifying and estimating the simultaneous interaction in $u_t = B_0 \varepsilon_t$ requires additional assumptions. In the next section, we describe how a proxy variable can be used to identify and estimate the impact of the target shock and afterwards derive a test for the exogeneity of the proxy variable. For simplicity, we omit the lag structure and only focus on the simultaneous interaction $y_t = u_t = B_0 \varepsilon_t$. An extension to a fully specified VAR(p) model with $p > 0$ is straightforward and can be found in appendix A.

2.1 Proxy SVAR

Typically, the underlying rational of a proxy variable is to find or construct a variable z_t that contains information about the target shock, ε_{1t}, but no information about the non-target shocks, $\boldsymbol{\varepsilon}_{2t}$. These properties are technically imposed using second-order moments, i.e. a proxy is valid if it is correlated with the target shock and uncorrelated with all non-target shocks.

Assumption 1 (Valid proxy)
The proxy z_t for the shock ε_{1t} is relevant and exogenous

1. *Relevance:* $E[\varepsilon_{1t} z_t] \neq 0$

2. *Exogeneity:* $E[\boldsymbol{\varepsilon}_{2t} z_t] = \mathbf{0}$

Using exogeneity and relevance allows to identify the impact of the target shock with $\beta_0 = \frac{\mathbb{E}(u_t z_t)}{\mathbb{E}(u_{1t} z_t)}$. However, with only a single proxy variable, we can not statistically test the identifying uncorrelatedness assumption $E[\boldsymbol{\varepsilon}_{2t} z_t] = 0$. Therefore, it typically remains up to the researcher to find convincing economic arguments why the proxy should not contain information on the non-target shocks.

Assumption 2 (Strong exogeneity)
The proxy z_t for the shock ε_{1t} is strongly exogenous if $E[\boldsymbol{\varepsilon}_{2t} | z_t] = 0$.

Strong exogeneity is a stronger assumption than uncorrelatedness of the proxy with all non-target shocks and we propose a test for the strong exogeneity assumption. The relevant question thus is whether the strong exogeneity assumption is too strong. In other words, are there proxy variables that are uncorrelated with the non-target shocks but do not meet the criteria for strong exogeneity? While it is statistically possible do define such a variable, we are not aware of any reasonable economic examples where proxy variables are arguably uncorrelated with the non-target shocks yet still contain predictive power on the expected value of the non-target shocks.[1]

Furthermore, in the case of a linear proxy variable, uncorrelatedness of the proxy with the non-target shocks also implies strong exogeneity. Specifically, consider a linear proxy variable

$$z_t = \psi_1 \varepsilon_{1t} + \psi_2 \varepsilon_{2t} + v_t, \tag{2}$$

with a white noise measurement error v_t. Suppose that the target shock ε_{1t} and the measurement error v_t contain no information on the expected value of the contaminating shock ε_{2t}, i.e. $E[\varepsilon_{2t} | v_t] = E[\varepsilon_{2t} | \varepsilon_{1t}] = 0$, such that the endogenous component of z_t is linear and equal to $\psi_2 \varepsilon_{2t}$. In this case, the traditional uncorrelatedness assumption $E[\varepsilon_{2t} z_t] = 0$ implies $\psi_2 = 0$, which immediately implies strong exogeneity, i.e. $E[\varepsilon_{2t} | z_t] = 0$ holds.

2.2 A Strong exogeneity test

In contrast to the commonly used proxy exogeneity assumption, which imposes uncorrelatedness of the proxy with all non-target shocks in Assumption 1, strong exogeneity in Assumption 2 can easily be tested. The key idea is that a strongly exogenous proxy can be used to construct an additional synthetic proxy, $\tilde{z}_t$. If the synthetic proxy provides over-identifying restrictions, a simple J-test can be employed to test for strong exogeneity.

By definition, a strongly exogenous proxy contains no information on the expected value of all non-target shocks, $E[\boldsymbol{\varepsilon}_{2t} | z_t] = 0$. This is equivalent to $E[\boldsymbol{\varepsilon}_{2t} h(z_t)] = 0$

[1] Note that strong exogeneity does not imply independence of the proxy variable and the non-target shocks. For example, strong exogeneity still allows that the proxy variable and the non-target shock are driven by the same volatility process. Furthermore, strong exogeneity even allows that the non-target shocks can have predictive power for the expected value of the proxy variable, only the opposite is prohibited.

for any measurable function $h(\cdot)$ with $E[h(\cdot)] < \infty$. Under strong exogeneity, the variable

$$\tilde{z}_t := h(z_t) \tag{3}$$

is uncorrelated with the non-target shocks and thus satisfies the traditional proxy exogeneity assumption imposing uncorrelatedness of the proxy with all non-target shocks in Assumption 1. Consequently, we refer to $\tilde{z}_t$ as a synthetic proxy variable. While Equation 3 theoretically allows the generation of infinitely many synthetic proxy variables, for simplicity, we focus on a straightforward example and use the synthetic proxy $\tilde{z}_t := z_t^2$.

Assume for now that both z_t and $\tilde{z}_t$ are relevant. If the original proxy z_t is strongly exogenous, it follows that both proxies are exogenous, i.e. uncorrelated with the non-target shocks, and thus valid proxies according to Assumption 1. The original proxy z_t yields $(K-1)$ moment conditions

$$E[f_z(\beta, u_t)] = 0 \quad \text{with} \quad f_z(\beta, u_t) = \boldsymbol{u}_{2t} z_t - \beta u_{1t} z_t. \tag{4}$$

These moment conditions allow us to identify β and the synthetic proxy $\tilde{z}_t$ yields additional $(K-1)$ moment conditions:

$$E[f_{\tilde{z}}(\beta, u_t)] = 0 \quad \text{with} \quad f_{\tilde{z}}(\beta, u_t) = \boldsymbol{u}_{2t} \tilde{z}_t - \beta u_{1t} \tilde{z}_t, \tag{5}$$

which lead to a potentially overidentified system. If the proxy z_t is strongly exogenous, all moment conditions hold[2] and the GMM estimator

$$\hat{\beta}_T := \underset{\beta \in \mathbb{R}^{K-1}}{\arg\min} \ g_T(\beta)' W g_T(\beta) \quad \text{with} \quad g_T(\beta) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^{T} f_z(\beta, u_t) \\ \frac{1}{T} \sum_{t=1}^{T} f_{\tilde{z}}(\beta, u_t) \end{bmatrix} \tag{6}$$

with a suitable weighting matrix W is consistent, i.e. $\hat{\beta}_T \xrightarrow{p} \beta$. Moreover, since the model is overidentified, the data can provide evidence that the moment conditions do not hold and thus, provide evidence against the strong exogeneity assumption. Therefore, a simple J-test can be used to test the strong exogeneity assumption. The J-test statistic proposed by (Hansen, 1982) is given by

$$J_T = T g_T(\hat{\beta}_T)' S^{-1} g_T(\hat{\beta}_T), \tag{7}$$

with $S := \underset{T\to\infty}{lim} E\left[T g_T(\beta_0) g_T(\beta_0)'\right]$.[3] Under the null hypothesis of a correctly specified model with $E[f_z(\beta, u_t)] = E[f_{\tilde{z}}(\beta, u_t)] = 0$, the distribution of the test statistic is given by $J_T \xrightarrow{d} \chi^2_{r-q}$ where $r = 2(K-1)$ is equal to the number of moment conditions and $q = K-1$ is equal to the number of elements in β. Rejecting the J-test yields evidence that at least one of the moment conditions is not correct, thus providing evidence against strong exogeneity of the proxy variable.

[2] If the proxy is strongly exogenous, both moment conditions hold since

$$\begin{aligned} E[f_z(\beta, u_t)] &= E[\boldsymbol{u}_{2t} z_t - \beta u_{1t} z_t] \overset{*}{=} \beta_0 E[\varepsilon_{1t} z_t] - \beta E[\varepsilon_{1t} z_t] = 0 \\ E[f_{\tilde{z}}(\beta, u_t)] &= E[\boldsymbol{u}_{2t} \tilde{z}_t - \beta u_{1t} \tilde{z}_t] \overset{*}{=} \beta_0 E[\varepsilon_{1t} \tilde{z}_t] - \beta E[\varepsilon_{1t} \tilde{z}_t] = 0, \end{aligned}$$

where the equality highlighted by $*$ uses the exogeneity assumption.

[3] We follow standard practice and use a two-step GMM estimator with $W = I_n$ in the first step and in the second step $W = \hat{S}(\hat{\beta}_T)^{-1}$ with $\hat{\beta}_T$ from the first step and $\hat{S}(\hat{\beta}_T) = \frac{1}{T} \sum_{t=1}^{T} \left[f_t(\tilde{\beta}_T)) f_t(\tilde{\beta}_T))' \right]$. The J-test also uses $\hat{S}(\hat{\beta}_T) = \frac{1}{T} \sum_{t=1}^{T} \left[f_t(\tilde{\beta}_T)) f_t(\tilde{\beta}_T))' \right]$, which in general is a consistent estimator for S if the moment conditions are serially uncorrelated, see Hall (2005).

2.3 Power properties

The proposed test only leverages information contained in the proxy variable. This section derives the conditions under which the proxy contains sufficient information to provide evidence against the strong proxy exogeneity assumption. Specifically, we show how non-Gaussian shocks and non-linearity of the proxy can lead to informative synthetic proxy variables which allow to provide evidence against strong exogeneity.

The ability to provide evidence against strong exogeneity depends on whether the information of the proxy variable is entirely contained in the second moments $E(\varepsilon_{1t} z_t)$ and $E(\varepsilon_{2t} z_t)$. Consider the synthetic proxy moment conditions from equation (5) with

$$E[f_{\tilde{z}}(\beta, u_t)] = E[\boldsymbol{u}_{2t} z_t^2 - \beta u_{1t} z_t^2] = (\beta_0 - \beta) E[\varepsilon_{1t} z_t^2] + (B_{22,0} - \beta B_{12,0}) E[\varepsilon_{2t} z_t^2] = 0,$$

where $B_{12,0}$ is the upper-right $(1 \times K-1)$ matrix of B_0 and $B_{22,0}$ is the lower-right $(K-1 \times K-1)$ matrix of B_0. If the synthetic proxy contains no information on the target or non-target shocks, i.e. $E(\varepsilon_{1t} z_t^2) = E(\varepsilon_{2t} z_t^2) = 0$, this set of moment conditions is equal to zero for any finite β. Moreover, if the synthetic proxy contains no information beyond the information contained in the second moments $E(\varepsilon_{1t} z_t)$ and $E(\varepsilon_{2t} z_t)$, such that $E(\varepsilon_{1t} z_t) = E(\varepsilon_{1t} z_t^2)$ and $E(\varepsilon_{2t} z_t) = E(\varepsilon_{2t} z_t^2)$, the moment conditions are fulfilled by the same β which fulfills the traditional proxy moment conditions in equation (4).[4] Consequently, if the information of the proxy is entirely contained in the second moments and the synthetic proxy contains no additional information, the J-test is not able to detect proxy endogeneity and will reject at the nominal level.

Conversely, if the information of the proxy variable is not entirely contained in the second moments, meaning the synthetic proxy is correlated with the target or contaminating shock, the synthetic proxy moment conditions may not be fulfilled for a given β vector, leading to a rejection of the J-test. This correlation can result from non-Gaussian shocks or a non-linear proxy process.

First, consider a linear proxy process as in equation (2). In this case, the moments $E(\varepsilon_{1t} z_t^2)$ and $E(\varepsilon_{2t} z_t^2)$ of the synthetic proxy are equal to

$$E(\varepsilon_{1t} z_t^2) = \psi_1^2 E[\varepsilon_{1t}^3] \quad \text{and} \quad E(\varepsilon_{2t} z_t^2) = \psi_2^2 E[\varepsilon_{2t}^3].$$

For a skewed target shock $E[\varepsilon_{1t}^3] \neq 0$ or a skewed contaminating shock $E[\varepsilon_{2t}^3] \neq 0$, the synthetic proxy contains information which can be used for detecting proxy endogeneity.[5] Specifically, the synthetic proxy moment conditions in equation (5) may not be fulfilled by the β vector which fulfills the traditional proxy moment conditions in equation (4), such that the J-test rejects.

Second, consider an SVAR with exclusively Gaussian shocks and a proxy equation featuring a non-linearity, $z_t = g(\epsilon_{1t}, \epsilon_{2t}, v_t)$. For such a proxy equation, the covariances are given by:

$$E(\varepsilon_{1t} z_t^2) = E[\varepsilon_{1t} g(\epsilon_{1t}, \epsilon_{2t}, v_t)^2] \quad \text{and} \quad E(\varepsilon_{2t} z_t^2) = E[\varepsilon_{2t} g(\epsilon_{1t}, \epsilon_{2t}, v_t)^2)]$$

Depending on the type of non-linearity, either or both correlations can be non-zero even in the presence of fully Gaussian shocks. As a simple example, consider the

[4] This case, for example, occurs for a binary proxy variable.

[5] Note that if only the non-contaminating non-target shock, ε_{3t}, displays skewness, then the synthetic proxy does not contain information which can be used to test for strong proxy exogeneity.

process $z_t = \varepsilon_{1t} + \varepsilon_{1t}^2 + \eta_t$ where the proxy is affected by the volatility of the target shock and let ε_{1t} and η_t be independently drawn from a standard normal distribution. In this case, the moment $E(\varepsilon_{1t} z_t^2)$ is equal to $E(\varepsilon_{1t}(\varepsilon_{1t}+\varepsilon_{1t}^2+\eta_t)^2) = 2E(\varepsilon_{1t}^4) \neq 0$, such that synthetic proxy moment conditions can again provide evidence against strong exogeneity.

In summary, the proposed test only leverages information contained in the proxy variable. The proxy variable contains sufficient information to detect proxy endogeneity if the synthetic proxy is informative about the target or contaminating shock. Potential sources for such information are skewed shocks or non-linear proxy variables. Only for the special case of exactly Gaussian shocks and an exactly linear proxy model, synthetic proxy variables cannot provide additional information and the test has no power to detect deviations from strong proxy exogeneity.

3 Monte Carlo Simulation

To investigate the small sample performance of our proposed test statistic we set up a Monte Carlo experiment. We find that the size and power depend on sample size, degree of proxy endogeneity, proxy strength, and the degree of the synthethic proxy informativeness, which is affected by the skewness of the structural shocks and the non-linearity in the proxy equation. We use two data-generating processes (DGPs).

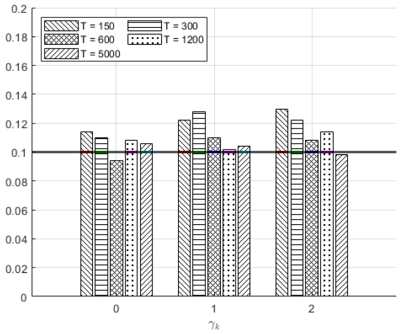
3.1 DGP1: Stylised SVAR Simulation

The first simulation uses a simplification of the DGP in Lütkepohl and Schlaak (2022) and Bruns and Lütkepohl (2022b), which these authors use for a related issue. The SVAR has three variables, one proxy to identify one shock, and follows a VAR(0) process:

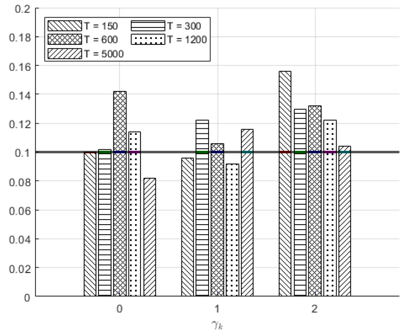
$$u_t = B_0 \varepsilon_t \quad \text{with} \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 4 \\ 4 & 6 & 6 \end{bmatrix}$$

Structural shocks are generated from the Pearson family of distributions, i.e. $\boldsymbol{\varepsilon}_{kt} \sim \mathcal{P}(\mu_k, \sigma_k^2, \gamma_k, \kappa_k)$, where μ_k is the mean, σ_k^2 the variance, γ_k the skewness, and κ_k the kurtosis. We assume $\mu_k = 0$, $\sigma_k^2 = 1$ and vary $\gamma_k = [0, 1, 2]$ for all shocks jointly. We keep κ_k constant across simulations at $\kappa_k = max(\gamma_k)^2 + 2 = 6$. Proxies are generated as in equation (2) with $v_t \sim N(0, 1)$. We choose (ψ_1, ψ_2) to simulate samples with three different levels of proxy strength with $corr(\varepsilon_{1t}, z_t) = (0.5, 0.7, 0.9)$ and three levels of proxy exogeneity with $corr(\varepsilon_{2t}, z_t) = (0, -0.2, -0.5)$. We generate $T = [150, 300, 600, 1200, 5000]$ observations (plus pre-sample values) and use 500 repetitions for each simulation.

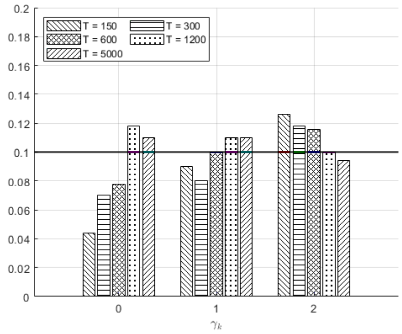
Figure 1 panels (a) - (c) show the rejection frequencies when data are generated under the Null of proxy exogeneity, i.e. $corr(\varepsilon_{2t}, z_t) = 0$, with a nominal level of 10%. While an increase in the sample size, T, leads to a better matching of the nominal level, we note that even for the smallest sample size, $T = 150$ the rejection frequency does not exceed 16% for any setup. Variations in proxy strength, $corr(\varepsilon_{1t}, z_t)$ do not seem to affect the nominal level much. The skewness has little effect on the nominal



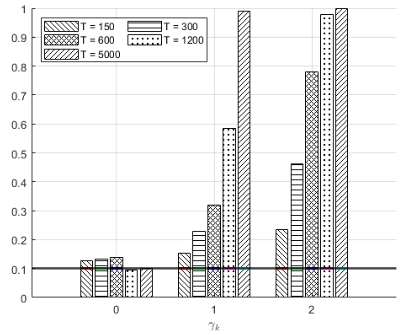
(a) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = 0$



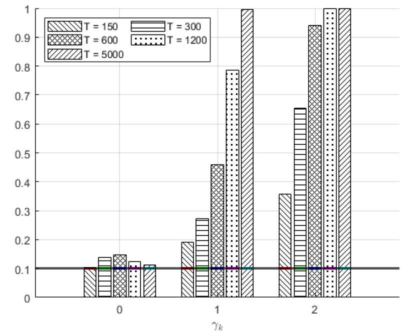
(b) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = 0$



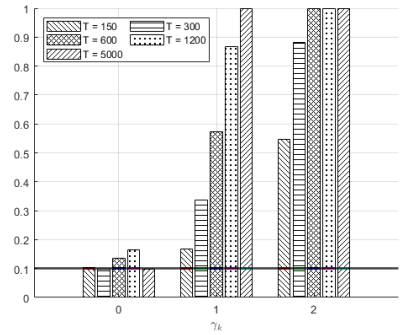
(c) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = 0$



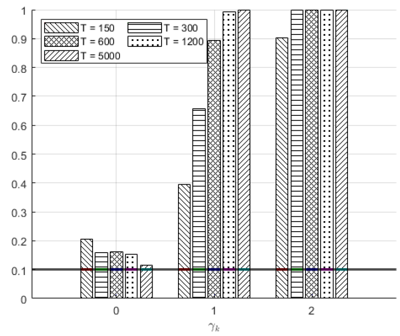
(d) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.2$



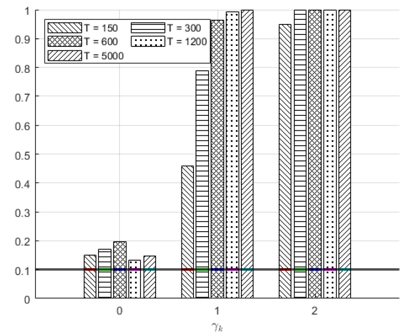
(e) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.2$



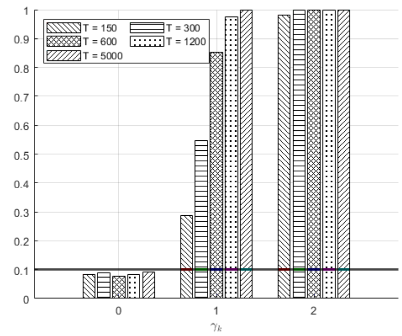
(f) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.2$



(g) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.5$



(h) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.5$



(i) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.5$

Figure 1: Relative rejection frequencies for DGP1. Nominal significance level 10%. $p = 0$.

level overall as well, except for shocks without skewness, $\gamma_k = 0$, and low sample sizes, $T = 150$, which leads to a nominal level of around 5%.

Figure 1 panels (d) - (i) show the rejection frequencies when data are generated under the alternative of proxy endogeneity, i.e. $corr(\varepsilon_{2t}, z_t) \neq 0$. When shocks have zero skewness ($\gamma_k = 0$) the test has no power and rejects close to the nominal level of 10%. Intuitively, in this case z_t^2 is not informative and cannot provide evidence against strong exogeneity. A larger skewness, i.e. increasing γ_k leads to more power. Stronger proxies, i.e. higher $corr(\varepsilon_{1t}, z_t)$ as well as larger deviations from the Null in the form of a higher proxy endogeneity, i.e. higher $corr(\varepsilon_{2t}, z_t)$, also increase the test's power.

In Appendix A, we introduce two approaches for incorporating lags into our testing methodology. First, we propose a one-step approach, where the J-test incorporates supplementary moment conditions corresponding to the lags of the SVAR. Second, we adopt a two-step approach, where the VAR is estimated in the first step and the J-test is conducted similar to the previous section using the reduced form shocks from the VAR. Notably, it's worth mentioning that a similar two-step procedure is typically employed for testing proxy strength using an F-test, see, Stock, Wright and Yogo (2002). While the two-step approach can lead to incorrect coverage, our simulations, both in the next section and within the appendix, show that the two-step approach yields conservative outcomes. Specifically, the rejection frequencies for the two step-approach are below the nominal level for all simulations including lags and exogenous proxy variables.

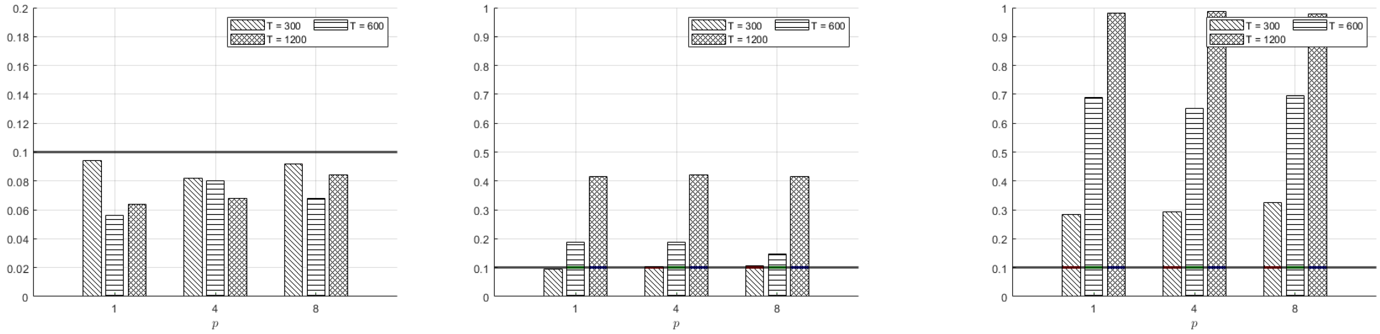
The simulation above is based on a very stylised VAR(0) model. In Appendix B we conduct a wide range of alternative simulations based on a VAR(1) including models estimated with $p = (1, 12)$ lags (figure A.1 and A.2). We also show that for our test to have power, it is enough for either the target (ε_{1t}) or the contaminating shock (ε_{2t}) to be skewed, while a skewed non-contaminating non-target shock (ε_{3t}) does not provide power (figure A.3). Lastly, we investigate a fully Gaussian set of shocks and three different types of non-linearities in the proxy equation and find that the test has power in these scenarios as well (figure A.4).

In the next section, we conduct a simulation based on a more realistic DGP with parameters from the fiscal SVAR estimated in Mertens and Ravn (2014).

3.2 DGP2: Fiscal SVAR Simulation

A second, more realistic, DGP is based on the variables by Mertens and Ravn (2014): total tax revenues, government spending, and output. The exact parameters used to generate data are given in the appendix. We estimate models with $p = [1, 4, 8]$ lags and intercept for $T = [300, 600, 1200]$ and 500 repetitions per simulation.

Figure 2 shows the results for the two-step testing procedure. In panel (a) data are generated under the Null of strong proxy exogeneity, i.e. $corr(\varepsilon_{2t}, z_t) = 0$. The empirical rejections frequencies are within 5%-points of the nominal level of 10% for all sample sizes and lag lengths shown. Panels (b) and (c) show that the power of the test increases with sample size and distance from the Null, i.e. higher correlation of the proxy with the non-target shock, but is only marginally affected by the lag length, p. In the Appendix we show that using a one-step testing procedure taking into account the moment conditions for the autoregressive coefficients leads to higher power at the cost of a less precise nominal level in smaller samples (figure A.5).



(a) $corr(\varepsilon_{2t}, z_t) = 0$ (b) $corr(\varepsilon_{2t}, z_t) = -0.1$ (c) $corr(\varepsilon_{2t}, z_t) = -0.2$

Figure 2: Relative rejection frequencies for DGP2 (2-step test). Nominal significance level 10%.

Overall, DGP2 shows that the test has good size and power properties in a realistic environment.

4 Application

This section applies the proposed exogeneity test to three proxy variables used in the fiscal proxy SVAR literature. Specifically, we test exogeneity of the tax proxy used in Mertens and Ravn (2014), the Fernald (2012) total factor productivity measure as an output shock proxy, and military spending changes as a government spending shock proxy (see e.g. Klein and Linnemann (2019)). We find evidence against strong exogeneity of the tax proxy, but no evidence against exogeneity of the output and spending proxies. However, we find that the synthetic output proxy is weak which may limit our ability to detect exogeneity violations of the output proxy.

We consider the fiscal SVAR as proposed by Mertens and Ravn (2014) for the US. The variables are federal tax revenues (τ_t), federal government consumption (g_t), and output (y_t), all in log real per capita terms for the sample 1950Q2 to 2006Q4, leading to $T = 228$ observations. The SVAR has four lags a constant, linear and quadratic trends, and a dummy for $1975Q2$ all contained in X_t with

$$\begin{bmatrix} \tau_t \\ g_t \\ y_t \end{bmatrix} = \gamma X_t + \sum_{i=1}^{4} A_i \begin{bmatrix} \tau_{t-i} \\ g_{t-i} \\ y_{t-i} \end{bmatrix} + \begin{bmatrix} u_{\tau,t} \\ u_{g,t} \\ u_{y,t} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} u_{\tau,t} \\ u_{g,t} \\ u_{y,t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{\tau,t} \\ \varepsilon_{g,t} \\ \varepsilon_{y,t} \end{bmatrix}, \tag{8}$$

tax shocks $\varepsilon_{\tau,t}$, government spending shocks $\varepsilon_{g,t}$, and output shocks $\varepsilon_{y,t}$.

We test the exogeneity of the following three proxies. First, the narrative tax proxy $z_{\tau,t}$ for tax shocks $\varepsilon_{\tau,t}$ constructed by Mertens and Ravn (2014) based on the Romer and Romer (2010) tax shocks identified by studying narrative records of tax policy decisions. Second, the Fernald (2012) TFP measure as a proxy $z_{y,t}$ for output shocks $\varepsilon_{y,t}$ used by Caldara and Kamps (2017) as a non-fiscal proxy. Third, military spending changes as a proxy $z_{g,t}$ for government spending shocks $\varepsilon_{g,t}$ used in Klein and Linnemann (2019).

Table 1 displays the skewness of the three proxy variables and the F-statistic for the original and synthetic proxies. The F-statistic indicates that the original tax proxy may be weak. However, the tax proxy exhibits a strong negative skewness (see also figure A.6 panel (b) in Appendix C), potentially driven by left skewed structural tax shocks, which leads to a F-statistic of the synthetic tax proxy close to 10. The spending proxy is right skwewed, potentially driven by right skewed spending shocks, and the F-statistic indicates relevant original and synthetic spending proxies. The F-statistic also indicates a relevant output proxy. However, the output proxy has a skewness close to zero which leads to an irrelevant synthetic output proxy. Based on the results presented in the previous section, we expect that exogeneity test may have low power for the tax and output proxy.

Table 1: Skewness and strength of proxy variables

	Skewness original proxy	**F-statistic original proxy**	**F-statistic synthetic proxy**
Tax Proxy	-4.47	4.22	9.6
Spending Proxy	3.22	131	25.7
Output Proxy	-0.02	56.5	0.01

Table 2 shows the results of the proxy exogeneity tests. First, despite the potential limitations stemming from the low relevance of the tax proxy, we find evidence that the tax proxy is endogenous. Specifically, we reject the null hypothesis that the proxy and synthetic proxy moment conditions (4) and (5) hold at the 10% level, which indicates that the tax proxy is not strongly exogenous. This outcome aligns with the findings of Lewis (2021) and Keweloh et al. (2023), both of whom provide evidence against exogeneity of the tax proxy. However, these studies rely on additional identifying assumptions, i.e. time-varying volatility or non-Gaussian and independent shocks. Our contribution to this body of literature is the provision of additional evidence challenging the exogeneity of the tax proxy, without relying on any supplementary identifying assumptions, such as heteroskedasticity or the assumption of independent and non-Gaussian shocks.

Second, we find no evidence against exogeneity of the spending and output proxy. Specifically, we can not reject the null hypothesis that the proxy and synthetic proxy moment conditions (4) and (5) for the spending and output proxy hold at the 10% level. However, we stress that the output proxy displays almost no skeweness and the synthetic output proxy appears to be not relevant. Therefore, the exogeneity test may have little power to detect exogeneity violations of the output proxy. In contrast, the spending proxy has a positive skewness and both, the original and the synthetic spending proxy, are found to be relevant, indicating that our exogeneity test may have power to detect exogeneity violations of the spending proxy.

Consequently, we present evidence against strong exogeneity of the tax proxy. Specifically, our findings suggest that the tax proxy carries information related on the expected value of the non-target shocks. It is essential to note that while it is theoretically conceivable that the tax proxy is uncorrelated with all non-target shocks, while the squared tax proxy, i.e. the synthetic tax proxy, is correlated with a non-target shock. In such a scenario, our test would indeed reject the strong exogeneity of the tax proxy, even when the tax proxy itself shows no correlation with non-target shocks. Nevertheless, the critical inquiry revolves not around the technical possibility of this scenario, but rather its economic plausibility. Mertens and Ravn

Table 2: Proxy exogeneity test

	J-statistic	**P-Value**
Tax Proxy	5.03	0.08
Spending Proxy	0.63	0.73
Output Proxy	1.56	0.46

Note: The table presents the results of the two-stage proxy exogeneity test, which involves estimating the VAR in the first step and subsequently conducting the exogeneity test in the second step using the original proxy variable z_t and its corresponding synthetic proxy $\tilde{z}_t = z_t^2$.

(2014) construct the tax proxy as a series of unanticipated tax shocks based on the Romer and Romer (2010) narrative tax shocks. If indeed the tax proxy represents a series of unforeseen tax shocks, then the squared tax proxy equates to a series of squared unforeseen tax shocks. Therefore, the same reasoning which is used to motivate exogeneity of the tax proxy can also be applied to argue for exogeneity of the synthetic tax proxy. Consequently, in line with the economic rationale underlying the proxy's construction, we expect both the original and synthetic proxy variables, to be exogenous. However, the data provide evidence against this hypothesis. Taking into account the economic rational of the proxy's construction, our results provide evidence against the validity of the proxy itself.

5 Conclusions

Our study addresses the issue of proxy exogeneity in structural vector autoregressions. Traditionally, asserting the exogeneity of a proxy has rested on economic justifications rather than statistical assessments. While recent statistical tests have emerged, they necessitate additional identifying assumptions beyond those inherent in the proxy. In contrast, we have introduced a novel approach that extends the concept of exogeneity to a stronger notion, assuming that the proxy variable is not just uncorrelated with non-target shocks but contains no information at all on the expected value of non-target shocks. This extension allows for the direct testing of the enhanced notion of proxy exogeneity, offering a more robust framework for evaluating the reliability of proxy-based SVAR models than just relying on some narrative justification for the exogeneity of the proxy. Importantly, the proposed framework also does not rely on the availability and validity of some additional identification assumptions but exploits only information contained in the proxy itself. We show that the proxy itself can contain the necessary information to provide evidence against exogeneity if the proxy is a non-linear function of the shocks or affected by non-Gaussian shocks. By applying our approach to widely-used proxy variables in the fiscal SVAR literature, we have demonstrated its effectiveness in uncovering deviations from strong exogeneity.

References

Angelini, G., Caggiano, G., Castelnuovo, E. and Fanelli, L. (2023). Are fiscal multipliers estimated with Proxy-SVARs robust?, *Oxford Bulletin of Economics and Statistics* **85**(1): 95–122.

Angelini, G. and Fanelli, L. (2019). Exogenous uncertainty and the identification of structural vector autoregressions with external instruments, *Journal of Applied Econometrics* **34**(6): 951–971.

Angelini, G., Fanelli, L. and Neri, L. (2023). Invalid proxies and volatility changes, *Technical report*, manuscript.

Arias, J. E., Rubio-Ramirez, J. F. and Waggoner, D. F. (2021). Inference in Bayesian Proxy-SVARs, *Journal of Econometrics* **225**(1): 88–106.

Braun, R. and Brüggemann, R. (2022). Identification of SVAR models by combining sign restrictions with external instruments, *Journal of Business & Economic Statistics* **41**(4): 1077–1089.

Bruns, M. and Lütkepohl, H. (2022a). An Alternative Bootstrap for Proxy Vector Autoregressions, *Computational Economics* .

Bruns, M. and Lütkepohl, H. (2022b). Identifying multiple shocks in heteroskedastic proxy vector autoregressions, *Technical report*, DIW Berlin.

Caldara, D. and Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from Bayesian Proxy SVARs, *American Economic Journal: Macroeconomics* **11**(1): 157–92.

Caldara, D. and Kamps, C. (2017). The analytics of svars: a unified framework to measure fiscal multipliers, *The Review of Economic Studies* **84**(3): 1015–1040.

Fernald, J. G. (2012). A quarterly, utilization-adjusted series on total factor productivity, *Working Paper Series 2012-19*, Federal Reserve Bank of San Francisco.

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

Giacomini, R., Kitagawa, T. and Read, M. (2022). Robust bayesian inference in proxy svars, *Journal of Econometrics* **228**(1): 107–126.

Gouriéroux, C., Monfort, A. and Renne, J.-P. (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.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the econometric society* pp. 1029–1054.

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

Keweloh, S. A. (2021). A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments, *Journal of Business & Economic Statistics* **39**(3): 772–782.

Keweloh, S. A., Klein, M. and Prüser, J. (2023). Estimating the effects of fiscal policy using a novel proxy shrinkage prior, *arXiv preprint arXiv:2302.13066* .

Klein, M. and Linnemann, L. (2019). Tax and spending shocks in the open economy: Are the deficits twins?, *European Economic Review* **120**: 103300.

Lewis, D. J. (2021). Identifying shocks via time-varying volatility, *The Review of Economic Studies* **88**(6): 3086–3124.

Lütkepohl, H. and Schlaak, T. (2022). Heteroscedastic proxy vector autoregressions, *Journal of Business & Economic Statistics* **40**(3): 1268–1281.

Matteson, D. S. and Tsay, R. S. (2017). Independent component analysis via distance covariance, *Journal of the American Statistical Association* **112**(518): 623–637.

Mertens, K. and Ravn, M. O. (2014). A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers, *Journal of Monetary Economics* **68**: S1–S19.

Mesters, G. and Zwiernik, P. (2022). Non-independent components analysis, *arXiv preprint arXiv:2206.13668* .

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

Schlaak, T., Rieth, M. and Podstawski, M. (2023). Monetary policy, external instruments, and heteroskedasticity, *Quantitative Economics* **14**(1): 161–200.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business & Economic Statistics* **20**(4): 518–529.

Appendix

A Extension to $p > 0$

For the realistic setting of a VAR(p) model with $p > 0$ one needs to decide how to account for the estimation uncertainty in the residuals, $\hat{u}_t$. Here, we suggest two approaches: As a baseline, the J- test presented in the paper can be directly applied conditioning on OLS-residuals, $\hat{u}^{OLS}$, without further changes. We refer to this as the "two-step" testing procedure since it involves first estimating the model via OLS to obtain residuals and then applying our test. Alternatively, the moment conditions in (6) can be augmented by the $K(Kp+3)$ moment conditions relating to the autoregressive slope coefficients, an intercept term as well as a linear and quadratic time trend as follows:

$$\hat{\theta}_T := \underset{\theta \in \mathbb{R}^{n-1+K(Kp+3)}}{\arg\min} \quad g_T(\theta)' W g_T(\theta) \quad \text{with} \quad g_T(\theta) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} x_t(y_t - \Pi x_t) \\ \frac{1}{T}\sum_{t=1}^{T} f_z(\beta, u_t) \\ \frac{1}{T}\sum_{t=1}^{T} f_{\tilde{z}}(\beta, u_t) \end{bmatrix}, \tag{A.9}$$

with $\theta = [vec(\Pi), \beta]'$, $\Pi = [A_1, ..., A_p]$, and $x_t = [1, t, t^2, y_{t-1}, ..., y_{t-p}]'$. We label this the "one-step" testing procedure since the VAR slope coefficients, Π, and the impact effects of the identified structural shock, β, are jointly rather than sequentially estimated. Compared to the "two-step" procedure it has the advantage that the weighting matrix, W, can be chosen to be consistent so that the implied J-statistic has an asymptotic χ^2-distribution. In finite samples, and depending on the model's dimensions, the "one-step" procedure is likely to be less precise. The two approached are numerically identical for $p = 0$.

B Additional Simulation Results

B.1 DGP1 with lags

To investigate the test's performance when including lags in the model, we generate data y_t from a VAR(1) following Lütkepohl and Schlaak (2022) and Bruns and Lütkepohl (2022b) and augment DGP1 by the following autoregressive parameters:

$$A_1 = \begin{bmatrix} 0.79 & 0.00 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0.00 & 0.62 \end{bmatrix}.$$

The largest Eigenvalue of A_1 is 0.95, implying a persistent but stable process. We generate data recursively without intercept starting from $y_1 = [0, 0, 0]'$.

B.2 DGP1 with linear proxy model and only one skewed shock

To investigate the test's performance when the proxy model is linear and not all, but only shock ε_{kt} exhibit skewness, we modify DGP1 to allow for only one skewed shock. The results are shown in figure A.3. If shocks ε_{1t} (target shock) or ε_{2t} (contaminating shock) are skewed, then the proxy still contains information beyond its first two moments, leading to power to detect a false Null. If only shock ε_{3t} is skewed, then the proxy does not contain such information since the third shock does not enter the proxy equation (2).

B.3 DGP1 with non-linear proxy model and Gaussain shocks

To investigate the test's performance when all shocks are exactly Gaussian, but the proxy equation is non-linear, we draw from $\varepsilon_{kt} \sim N(0,1), \forall k$ and modify (2) as

NL 1:

$$z_t = \begin{cases} \psi_1 \varepsilon_{1t} + v_t & \varepsilon_{1t} + \varepsilon_{2t} < abs(\Phi^{-1}(\psi_3)) \\ \psi_1 \varepsilon_{1t} + \psi_2 \varepsilon_{2t} + v_t & \varepsilon_{1t} + \varepsilon_{2t} > abs(\Phi^{-1}(\psi_3)) \end{cases}, \tag{A.10}$$

where $\Phi^{-1}(x)$ is the cdf of a normal distribution with standard deviation 2. In words, the proxy is contaminated by shock ε_{2t} if the sum $\varepsilon_{1t} + \varepsilon_{2t}$ exceeds a threshold. The non-linearity is stronger for higher values of ψ_3. We investigate $\psi_1 = 1$, $\psi_2 = [0, 0.25, 0.5]$, and $\psi_3 = [0.8, 0.7, 0.5]$. The results are shown in figure A.4 (top panel).

Another non-linearity in the proxy is introduced as

NL 2:

$$z_t = \psi_1 \varepsilon_{1t} + \psi_2 \varepsilon_{2t} + \psi_3 \varepsilon_{1t}^2 + v_t \tag{A.11}$$

we set $\psi_1 = 1$, vary $\psi_2 = [0, 0.1, 0, 25]$, and $\psi_3 = [0.01, 0.05, 0.1]$. In this case, the non-linearity arises from the squared term, ε_{1t}^2. Again, the non-linearity is stronger for higher values of ψ_3. The results are shown in figure A.4 (middle panel).

A third non-linearity in the proxy is motivated by an interest rate zero lower bound. The time series i_t is a truncated AR(1) process with $i_t = \psi_3 + 0.9 i_{t-1} + u_t$ and

$$w_t = \begin{cases} \psi_1 \varepsilon_{1t} & \psi_3 + 0.9 i_{t-1} + \psi_1 \varepsilon_{1t} > 0 \\ max(\psi_1 \varepsilon_{1t}, -(\psi_3 + 0.9 i_{t-1})\lambda_t) & \psi_3 + 0.9 i_{t-1} + \psi_1 \varepsilon_{1t} \leq 0 \end{cases} \tag{A.12}$$

and $\lambda_t \sim \mathcal{U}_{[0,1]}$ such that if a shock ε_{1t} would drive i_t below the zero lower bound, only a random fraction of the shock is realized ensuring that i_t remains above the zero lower bound. The proxy z_t is given by

NL 3:

$$z_t = w_t + \psi_2 \varepsilon_{2t} + v_t \tag{A.13}$$

with $\psi_1 = 1$, $\psi_2 = [0, 0.25, 0.5]$ determining the degree of endogeneity and proxy strength, and $\psi_3 = [1, 0.1, 0.]$ governing the degree of non-linearity where higher values lead to a larger distance to the zero lower bound and thus less non-linearity.

The results are shown in figure A.4 (bottom panel). The figure shows that our test leads to the correct nominal level if the proxy is exogenous, i.e. $\psi_2 = 0$, irrespective of the non-linearity of the proxy. For endogenous proxy variables with $\psi_2 \neq 0$, we find that the power increases with the degree of non-linearity, the sample size, and proxy endogeneity.

B.4 DGP2 Details

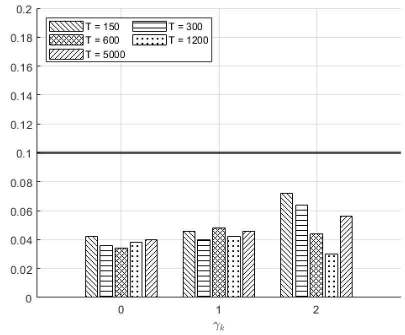
To obtain parameters for DGP2 we estimate a VAR(1) model with constant, but without trends or time dummies for the variables in (Mertens and Ravn, 2014) to obtain the following parameters:

$$A_1 = \begin{bmatrix} 0.930 & 0.053 & -0.055 \\ -0.063 & 0.770 & 0.250 \\ -0.014 & -0.043 & 1.045 \end{bmatrix}$$

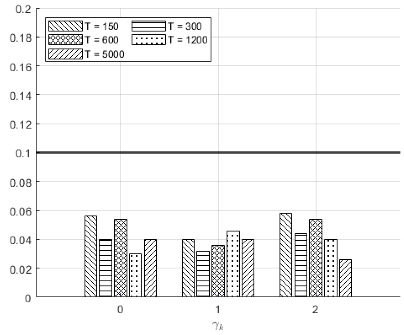
The maximum Eigenvalue of the associated companion form is 0.9975 indicating a very persistent but stable process. We then follow Mertens and Ravn (2014)'s identification approach and assume that all shocks are uncorrelated with unit variance and restrict the simultaneous impact of output shocks on government spending to zero to obtain $B = \begin{bmatrix} 0.004 & 0 & 0.028 \\ 0.013 & 0.026 & 0.000 \\ -0.005 & 0.008 & 0.002 \end{bmatrix}$.

Structural shocks are generated with $\gamma_k = -4.43 \; \forall \; k$ and $\kappa_k = \gamma_k^2 + 2 = 21.62 \; \forall \; k$, mimicking moments of the proxy in Mertens and Ravn (2014).

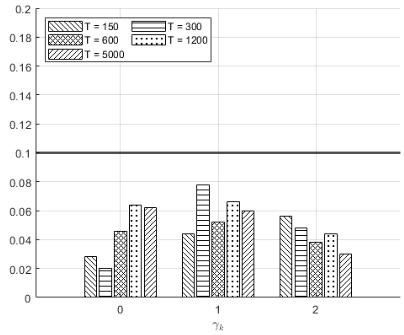
The proxy z_t is generated from equation (2) with $corr(\varepsilon_{1t}, z_t) = 0.26$ as in Mertens and Ravn (2014). We investigate $corr(\varepsilon_{2t}, z_t) = (0, -0.1, -0.2)$, which is informed by the finding in Keweloh et al. (2023) that the tax proxy in Mertens and Ravn (2014) has a correlation of -0.2 with the output shock, suggesting some degree of endogeneity.



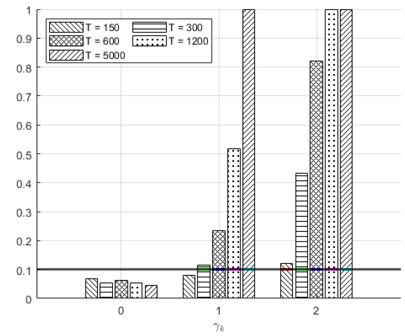
(a) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = 0$



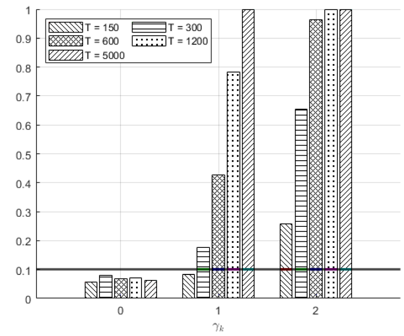
(b) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = 0$



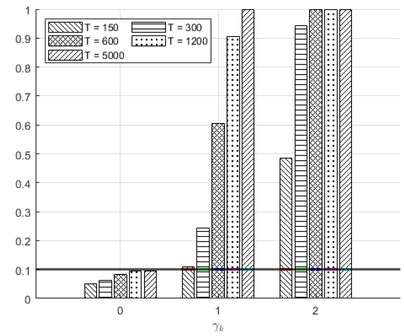
(c) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = 0$



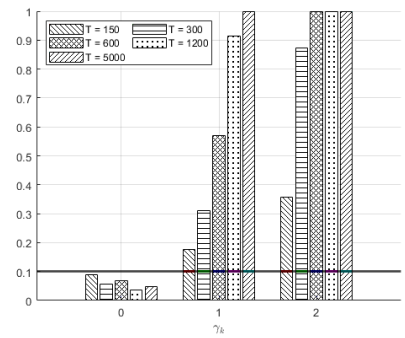
(d) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.2$



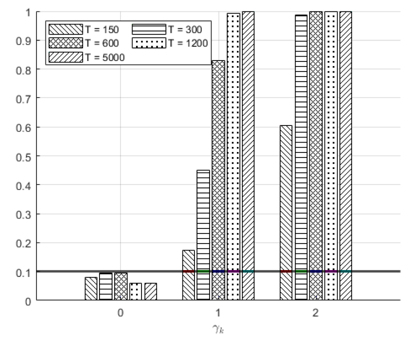
(e) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.2$



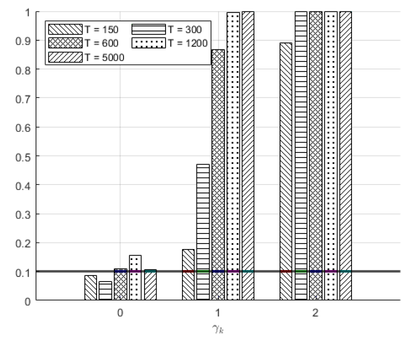
(f) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.2$



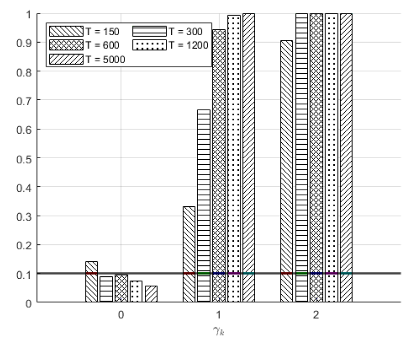
(g) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.3$



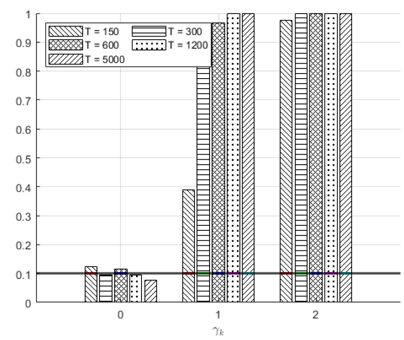
(h) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.3$



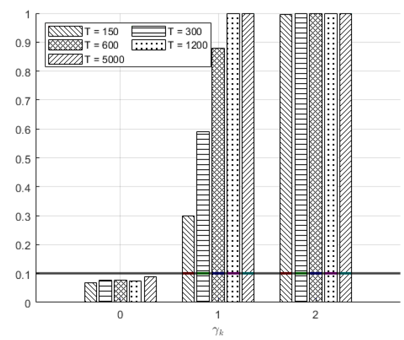
(i) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.3$



(j) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.5$

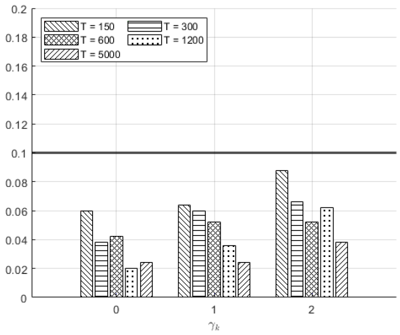


(k) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.5$

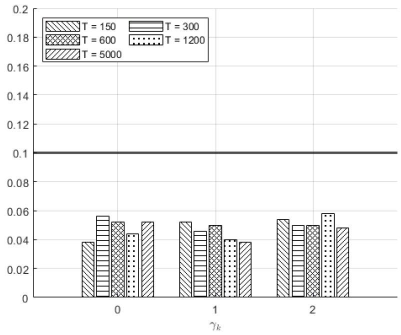


(l) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.5$

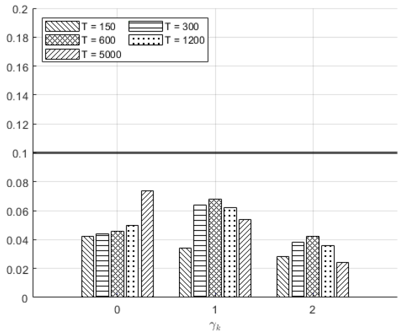
Figure A.1: Relative rejection frequencies for DGP1 (2-step test). Nominal significance level 10%. $p = 1$.



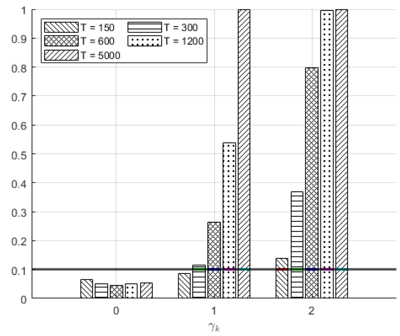
(a) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = 0$



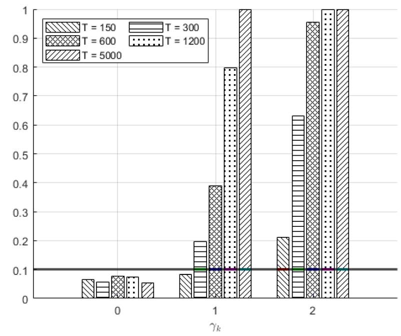
(b) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = 0$



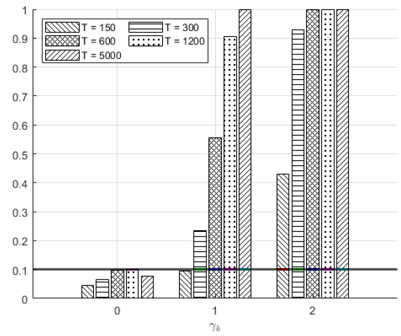
(c) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = 0$



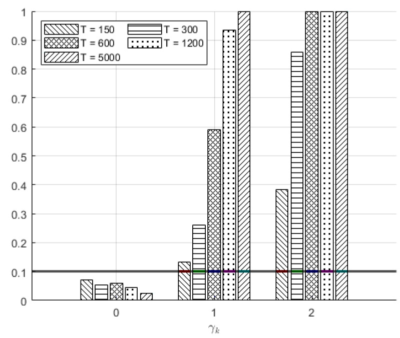
(d) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.2$



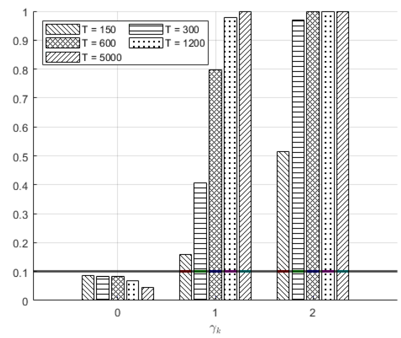
(e) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.2$



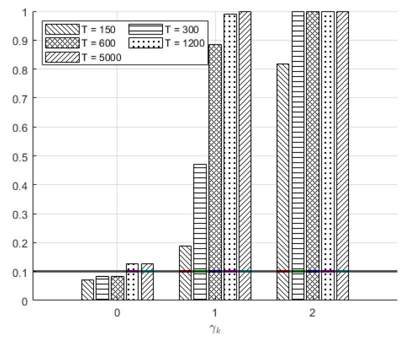
(f) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.2$



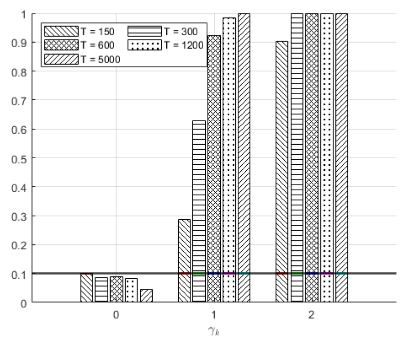
(g) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.3$



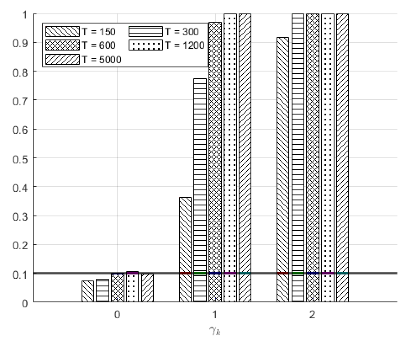
(h) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.3$



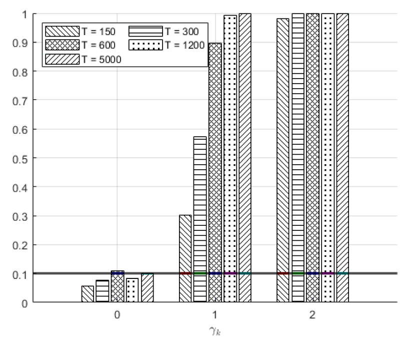
(i) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.3$



(j) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = -0.5$

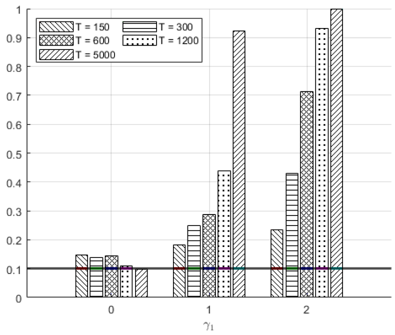


(k) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = -0.5$

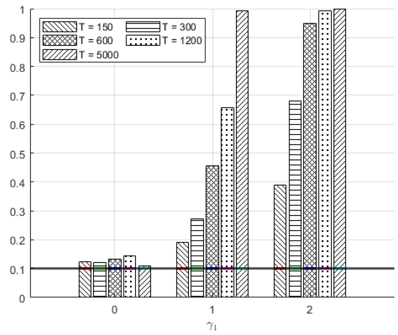


(l) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = -0.5$

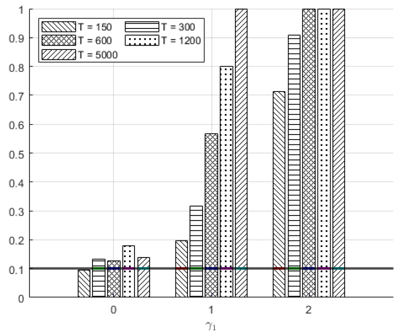
Figure A.2: Relative rejection frequencies for DGP1 (2-step test). Nominal significance level 10%. $p = 12$.



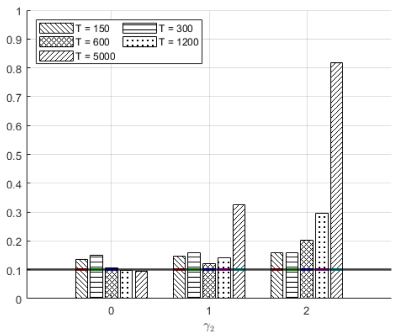
(a) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_2 = \gamma_3 = 0$



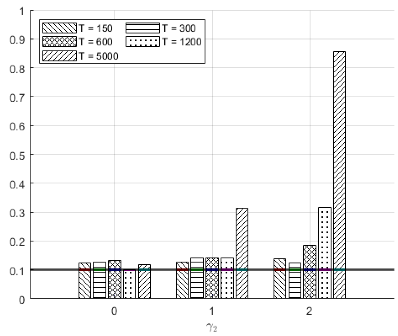
(b) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_2 = \gamma_3 = 0$



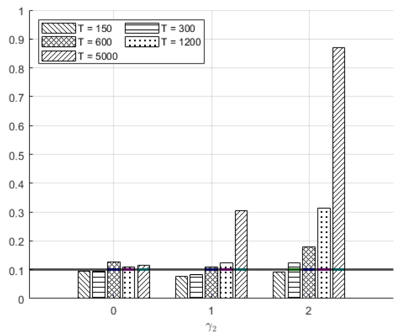
(c) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_2 = \gamma_3 = 0$



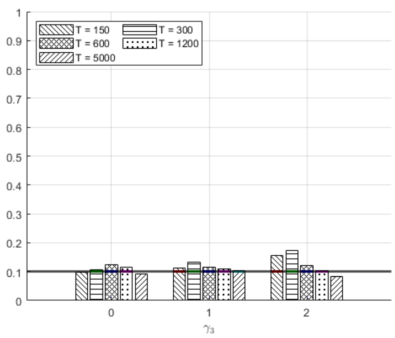
(d) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_1 = \gamma_3 = 0$



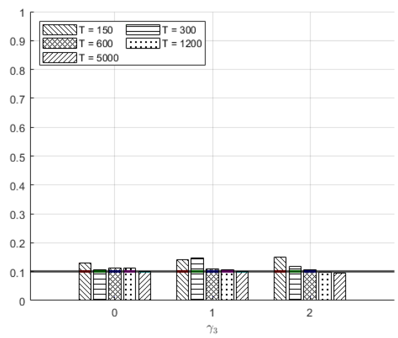
(e) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_1 = \gamma_3 = 0$



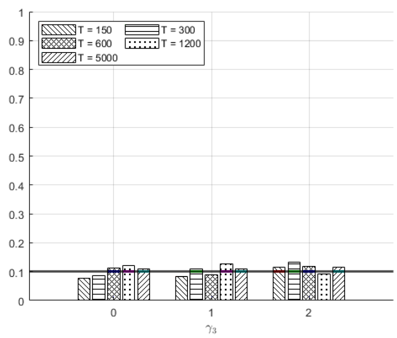
(f) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_1 = \gamma_3 = 0$



(g) $corr(z_t, \varepsilon_{1t}) = 0.5$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_2 = \gamma_3 = 0$

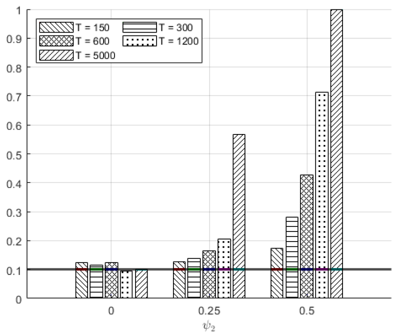


(h) $corr(z_t, \varepsilon_{1t}) = 0.7$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_2 = \gamma_3 = 0$

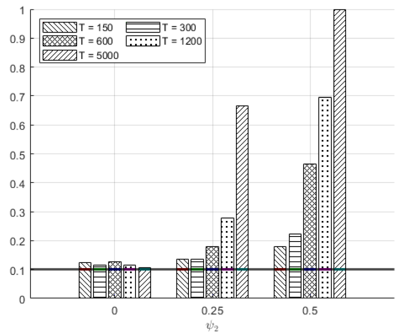


(i) $corr(z_t, \varepsilon_{1t}) = 0.9$, $corr(z_t, \varepsilon_{2t}) = 0$, $\gamma_2 = \gamma_3 = 0$

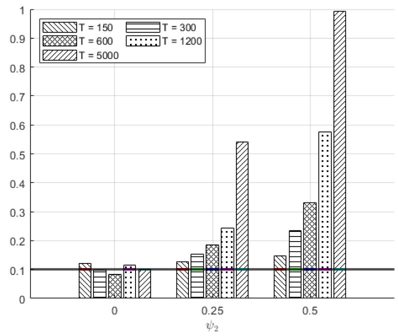
Figure A.3: Relative rejection frequencies for DGP1 when only one shock is skewed. Nominal significance level 10%. $p = 0$. Shock skewness is non-zero only for shock w_{1t} (panels (a) -(c)), only for shock w_{2t} (panels (d) - (f)) or only for shock w_{3t} (panels (g) - (i)).



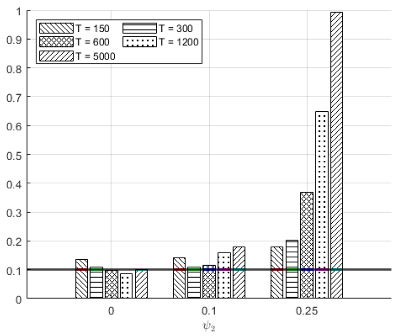
(a) $\psi_3 = 0.5$



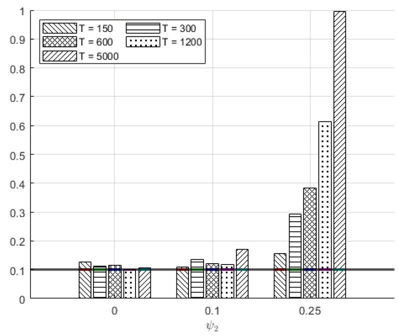
(b) $\psi_3 = 0.7$



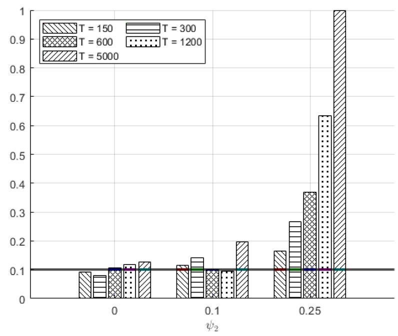
(c) $\psi_3 = 0.8$



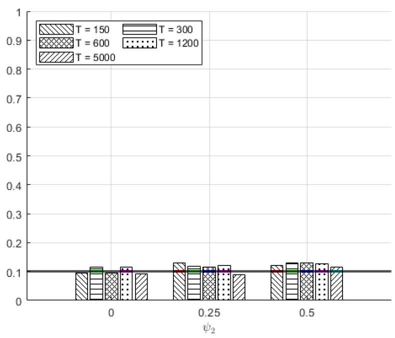
(d) $\psi_3 = 0.01$



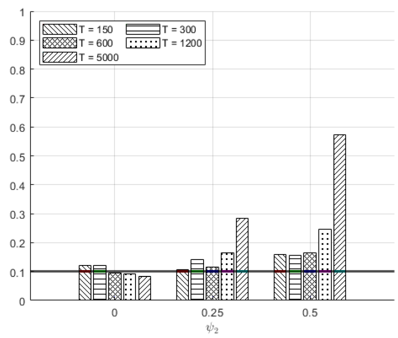
(e) $\psi_3 = 0.1$



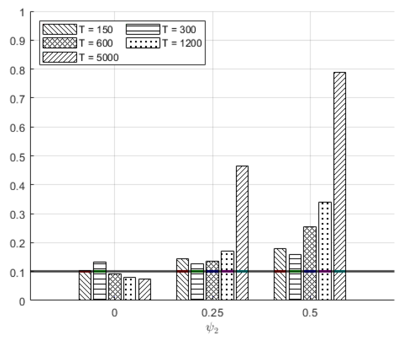
(f) $\psi_3 = 0.05$



(g) $\psi_3 = 1$

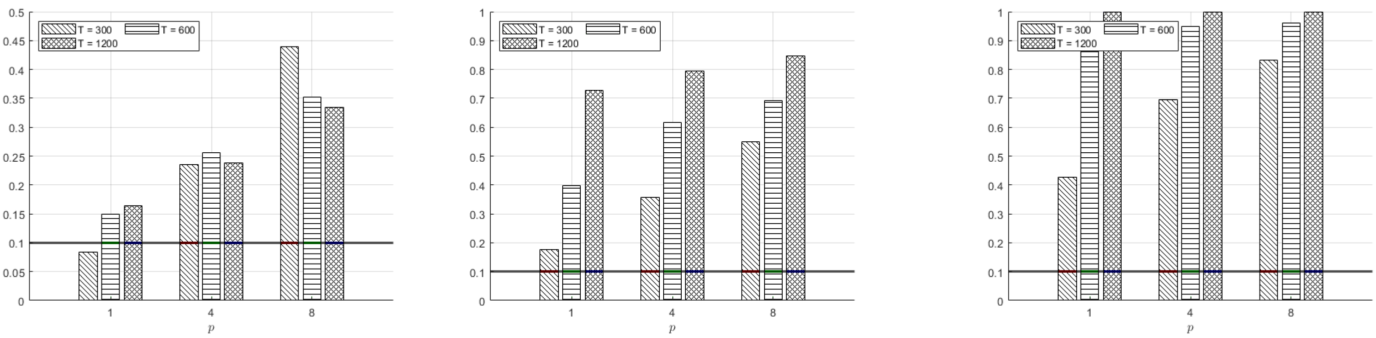


(h) $\psi_3 = 0.1$



(i) $\psi_3 = 0$

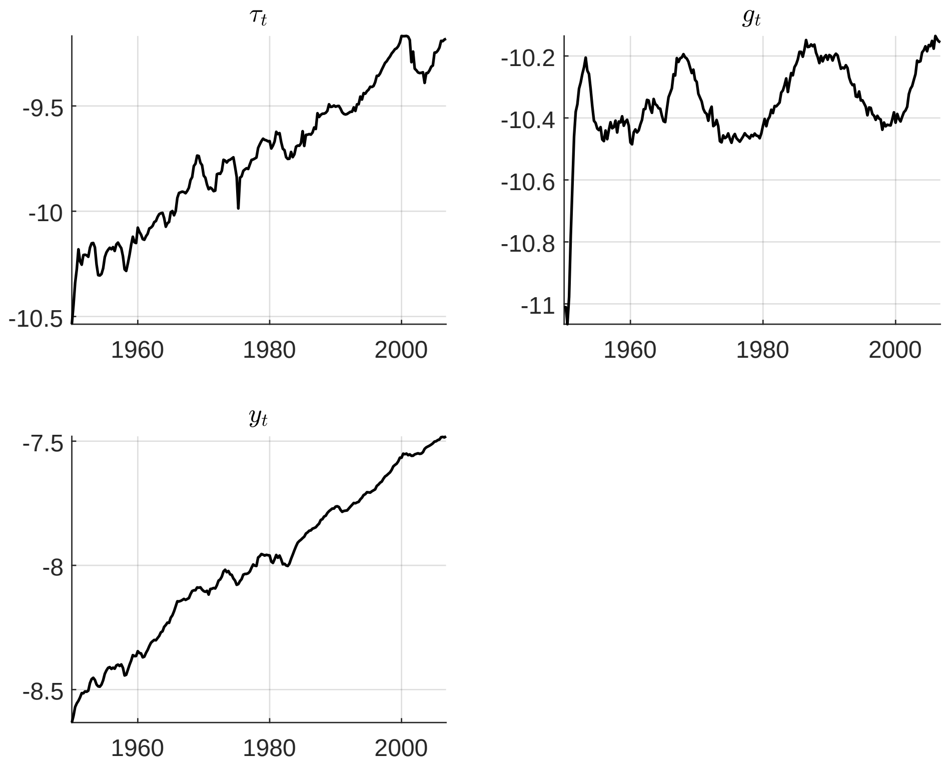
Figure A.4: Relative rejection frequencies for DGP1 when proxy equation is non-linear according to *NL 1* (top panel), *NL 2* (middle panel), and *NL 3* (bottom panel). Nominal significance level 10%. $p = 0$.



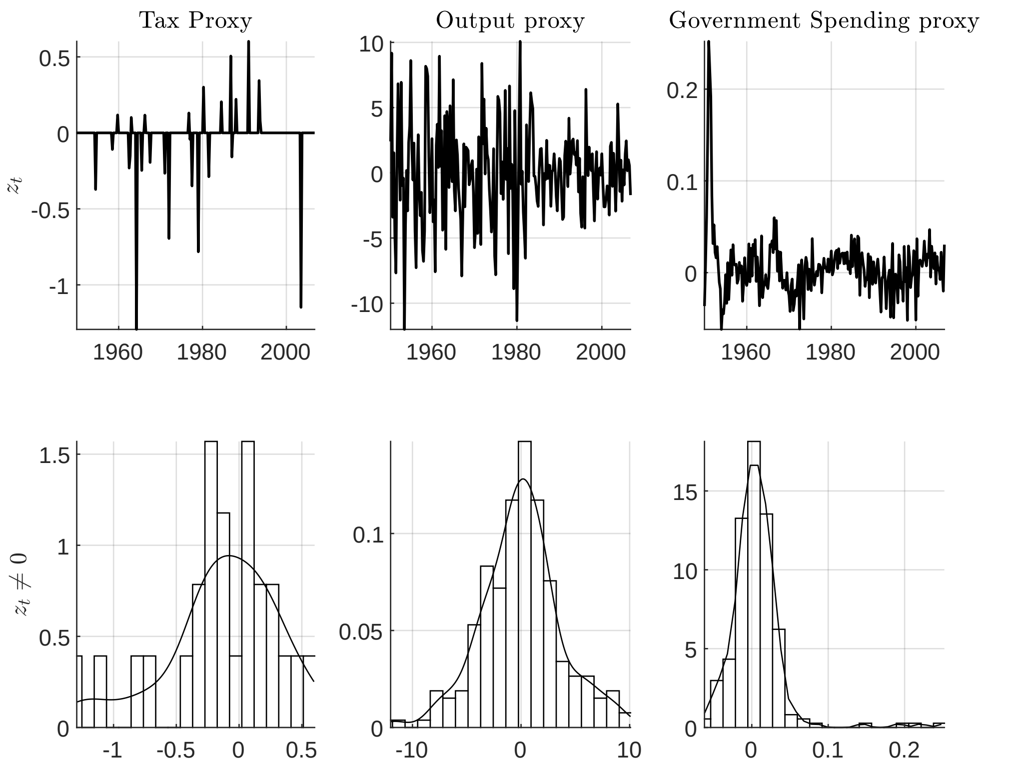
(a) $corr(\varepsilon_{2t}, z_t) = 0$ (b) $corr(\varepsilon_{2t}, z_t) = -0.1$ (c) $corr(\varepsilon_{2t}, z_t) = -0.2$

Figure A.5: Relative rejection frequencies for DGP2 (1-step test). Nominal significance level 10%.

C Application



(a) Mertens and Ravn (2014) variables



(b) Proxies by Mertens and Ravn (2014), Fernald (2012), and Klein and Linnemann (2019). Time series plot (top panels) together with histogram and kernel density estimate (bottom panel).

Figure A.6: Data.