

Lütkepohl, Helmut; Schlaak, Thore

Conference Paper

Heteroskedastic Proxy Vector Autoregressions

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Lütkepohl, Helmut; Schlaak, Thore (2021) : Heteroskedastic Proxy Vector Autoregressions, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/242399>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Heteroskedastic Proxy Vector Autoregressions

Helmut Lütkepohl

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

and

Thore Schlaak

DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany
email: thore.schlaak@gmail.com

This version: December 15, 2020¹

Abstract. In proxy vector autoregressive models, the structural shocks of interest are identified by an instrument. Although heteroskedasticity is occasionally allowed for in inference, it is typically taken for granted that the impact effects of the structural shocks are time-invariant despite the change in their variances. We develop a test for this implicit assumption and present evidence that the assumption of time-invariant impact effects may be violated in previously used empirical models.

Key Words: Structural vector autoregression, proxy VAR, identification through heteroskedasticity

JEL classification: C32

¹The authors thank Lutz Kilian, Ambrogio Cesa-Bianchi and three anonymous reviewers for useful comments on an earlier version of the paper. The paper was presented in the Vienna Joint Economics Seminar in November 2020. The authors are grateful for the comments of the participants.

1 Introduction

In structural vector autoregressive (VAR) analysis, shocks of interest are sometimes identified by using an instrument variable or proxy (e.g., Stock and Watson (2012), Mertens and Ravn (2013), Piffer and Podstawski (2018), Kilian and Lütkepohl (2017, Chapter 15)). In these so-called proxy VARs it is commonly assumed that the impact effects of the structural shocks are time-invariant even if the reduced-form VAR is heteroskedastic. The assumption of time-invariant impact effects is often implicitly made although heteroskedasticity is allowed for in inference by using heteroskedasticity-robust bootstrap methods for impulse responses, for example. Such an approach is used, for instance, in Mertens and Ravn (2013), Piffer and Podstawski (2018), Dias and Duarte (2019), Gertler and Karadi (2015) and Alessi and Kerssenfischer (2019). Other authors using proxy VARs consider heteroskedastic variables in their models without accounting for it and they assume that heteroskedasticity does not affect the impulse responses (e.g., Caldara and Herbst (2019)).

The objective of this study is to clarify the assumption of time-invariant impact effects of the structural shocks in heteroskedastic proxy VAR models and draw attention to its implications. It is argued that time-invariance of the impulse responses cannot be taken for granted when the errors are heteroskedastic. We develop an asymptotically valid test which can be used to investigate the time-invariance of the impact effects of the shocks. The usefulness of the test in small samples is demonstrated by a Monte Carlo simulation experiment.

As an empirical example, we reconsider a study by Cesa-Bianchi, Thwaites and Vicondoa (2020) which uses a proxy VAR model and assumes time-invariant impulse responses in the presence of volatility changes of the data. The authors investigate the transmission of monetary policy shocks in the United Kingdom (UK). The VAR model errors are clearly heteroskedastic and our test rejects the assumption of time-invariant impulse responses. Some impulse responses are clearly distinct in the different volatility regimes. This outcome shows that in empirical studies, it is advisable to test the assumption of time-invariant impulse responses and do not take

it for granted.

The remainder of the paper is structured as follows. The precise model setup and the implicit assumption of time-invariance in heteroskedastic proxy VAR models are discussed in Section 2. The test for time-invariant impact effects of the structural shocks is developed in Section 3 and its small sample properties are explored in Section 4. The test is applied to the example model in Section 5 and conclusions are presented in the final section.

2 Proxy VAR Models

2.1 Model Setup

Consider a K -dimensional reduced-form VAR model

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

where u_t is a zero mean white noise process with nonsingular covariance matrix Σ_u , i.e., $u_t \sim (0, \Sigma_u)$.

The vector of structural errors, $w_t = (w_{1t}, \dots, w_{Kt})'$, is assumed to have instantaneously uncorrelated components, i.e., its covariance matrix is diagonal. The structural errors are obtained from the reduced-form errors, u_t , by a linear transformation, $u_t = Bw_t$, where B is the matrix of impact effects of the structural shocks. Suppose there is a set of N instrumental variables (proxies) $z_t = (z_{1t}, \dots, z_{Nt})'$ satisfying

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

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

Then the first column of a constant multiple of B can be estimated by using z_t as an instrument vector. More precisely, denoting the first column of B by b , $\mathbb{E}(u_t z_t') = bc'$. In other words, the covariance between the reduced-form errors u_t and the proxy

vector z_t is the first column of B times the row vector c' . In practice, often only one proxy variable is available to identify a specific shock of interest. In that case z_t is a scalar random variable and c is also a scalar.

Now suppose that there is heteroskedasticity such that

$$\mathbb{E}(u_t u_t') = \Sigma_t = \Sigma_u(m) \quad \text{for } t \in \mathcal{T}_m, \quad m = 1, \dots, M, \quad (3)$$

where $\mathcal{T}_m = \{T_{m-1} + 1, \dots, T_m\}$ ($m = 1, \dots, M$) are M given volatility regimes associated with consecutive time periods. The volatility changes occur at time periods T_m for $m = 1, \dots, M - 1$, with $T_0 = 0$ and $T_M = T$, the overall sample size. The change points T_m are assumed to be known to the investigator or they may be predetermined by some statistical procedure. In any case, in our basic model setup, the volatility changes are not driven by the variables of the model but are assumed to occur exogenously.

To justify the relevance condition in (1) for a heteroskedastic model, the impact effects of the first shock must be assumed to be invariant to the heteroskedasticity. Such an assumption may be justified if there exists a decomposition of the regime dependent covariance matrices such that

$$\Sigma_u(m) = B \Lambda_m B', \quad m = 1, \dots, M, \quad (4)$$

where $\Lambda_m = \text{diag}(\lambda_{1,m}, \dots, \lambda_{K,m})$ ($m = 1, \dots, M$) are diagonal matrices with strictly positive diagonal elements. Clearly, choosing $w_t = B^{-1} u_t$ provides a time-invariant transformation which results in uncorrelated components of w_t across the whole sample. Such covariance matrices are typically assumed in the literature on identification through heteroskedasticity (see Kilian and Lütkepohl (2017, Chapter 14)).

For $M = 2$, the covariance matrices $\Sigma_u(1)$ and $\Sigma_u(2)$ can always be decomposed as in (4) (e.g., Lütkepohl (1996, p. 86)). Although a time-invariant B matrix for the decomposition (4) exists in this case, that B matrix may not reflect the impact effects of the structural shocks of interest. Of course, there are many other decompositions

of $\Sigma_u(1)$ and $\Sigma_u(2)$ which imply that the impact effects of the shocks depend on the volatility regime. Consider, for example, Cholesky decompositions $\Sigma_u(m) = B(m)B(m)'$, $m = 1, 2$, where $B(m)$ is a lower-triangular matrix with positive diagonal elements. If these transformation matrices are used to get the structural errors as $w_t = B(m)^{-1}u_t$, the w_t have identity covariance matrix throughout the sample period and the heteroskedasticity in the reduced-form errors, u_t , is solely due to the time-varying impact effects, $B(m)$.

If there are three or more volatility regimes ($M \geq 3$), then a decomposition of the associated covariance matrices as in (4) imposes a restriction on the covariance matrices which may not be compatible with the data or the specification of the structural shocks (see Angelini, Bacchiocchi, Caggiano and Fanelli (2019) for an example). In that case, the matrix of impact effects of the shocks, $B(m)$, depends on the covariance regime. If the impact effects of the shocks vary over time, the relevance condition in (1) may have to be modified to allow for a time-varying covariance between the instruments and the first structural shock,

$$\mathbb{E}(w_{1t}z_t) = c'_m \neq 0, \quad t \in \mathcal{T}_m, \quad m = 1, \dots, M \quad (\text{relevance}). \quad (5)$$

Hence, $\mathbb{E}(u_t z'_t) = b(m)c'_m$ and the first column, $b(m)$, of $B(m)$ may also depend on the volatility state m .

For an impulse response analysis based on proxy VARs, the size of the shock of interest is typically fixed such that its impact effect is the same across volatility regimes. Estimating the impact effects of the first shock via the instrument in this case remains valid if $b(m)$ is time-invariant, i.e., $b(m) = b$ for $m = 1, \dots, M$. In other words, if only the variance of the first shock changes with the volatility regime while the impact effects remain time-invariant, the proxy VAR approach can be used in the usual manner. In general, it is worth checking the assumption of time-invariant impact effects if there is heteroskedasticity and we will propose a suitable test in the next section.

Although we have considered a model with exogenous volatility changes in the foregoing discussion, a similar analysis is relevant for models with other types of heteroskedasticity or conditional heteroskedasticity. For example, the volatility changes may be driven by a Markov switching process as in Lanne, Lütkepohl and Maciejowska (2010) and Herwartz and Lütkepohl (2014). In these models it is also assumed that there are M distinct volatility states which may or may not be associated with different impact effects of the shocks. Note, however, that the VAR slope coefficients are still assumed to be time-invariant. In other words, we are not proposing a Markov-switching model as considered, e.g., by Sims and Zha (2006). Our test is designed for the case where only the error covariance parameters are allowed to change.

3 Testing for Time-varying Impact Effects

In this section, a test is developed for time-varying impact effects if there is heteroskedasticity of the type specified in (3) with known volatility change points T_m , $m = 1, \dots, M$. For notational simplicity, we develop the test in terms of the unobserved reduced-form errors u_t and then discuss the modifications implied by replacing them with estimated residuals \hat{u}_t .

We continue to assume – without loss of generality – that the instruments identify the first shock, as in the previous section and, for simplicity, we denote by $C(m)$ the $(K \times N)$ matrix of covariances $\mathbb{E}(u_t z_t')$ for $t \in \mathcal{T}_m$. We use the notation $c_{ij}(m)$, $c_{\bullet j}(m)$ and $c_{i\bullet}(m)$ for the ij^{th} element, the j^{th} column and the i^{th} row of $C(m)$, respectively. Moreover, the first row of $C(m)$, $c_{1\bullet}(m) = (c_{11}(m), \dots, c_{1N}(m))$ is assumed to be a nonzero vector, i.e., $c_{1\bullet}(m) \neq 0$, and the size of the shock is normalized such that it has an instantaneous unit impact on the first variable. Thereby the effect of a time-varying covariance between the proxy and the first shock cancels if the impact effects remain time-invariant.

We define the $(K - 1)$ -dimensional vector

$$\beta(m) = \begin{pmatrix} c_{2\bullet}(m) \\ \vdots \\ c_{K\bullet}(m) \end{pmatrix} c_{1\bullet}(m)' / c_{1\bullet}(m)c_{1\bullet}(m)'. \quad (6)$$

Thus, the vector of impact effects of the first shock is $(1, \beta(m)')'$, $m = 1, \dots, M$. For $m, j \in \{1, \dots, M\}$, $m \neq j$, we wish to test the pair of hypotheses

$$\mathbb{H}_0 : \beta(m) = \beta(j) \quad \text{versus} \quad \mathbb{H}_1 : \beta(m) \neq \beta(j). \quad (7)$$

In the following, a suitable test statistic is presented.

Let

$$\widehat{C}(m) = \frac{1}{\tau_m T} \sum_{t \in \mathcal{T}_m} u_t z_t', \quad (8)$$

be an estimator of the covariance matrix $\mathbb{E}(u_t z_t')$ in volatility regime $m \in \{1, \dots, M\}$. We assume that $\tau_m = (T_m - T_{m-1})/T$ is a fixed fraction of the sample size so that $T_m - T_{m-1} \rightarrow \infty$ with T . Moreover, u_t and z_t are such that $\widehat{C}(m)$ is consistent and asymptotically normal, that is,

$$\sqrt{T} \text{vec} \left(\widehat{C}(m) - C(m) \right) \xrightarrow{d} \mathcal{N}(0, \tau_m^{-1} \Sigma_C(m)), \quad (9)$$

where vec denotes the column stacking operator and \xrightarrow{d} signifies convergence in distribution. Under general conditions, such a result follows from a central limit theorem, i.e., if $u_t z_t'$ is independently, identically distributed (iid) for $t \in \mathcal{T}_m$ or a martingale difference sequence. We also assume that $u_t z_t'$ is such that

$$\widehat{\Sigma}_C(m) = \frac{1}{\tau_m T} \sum_{t \in \mathcal{T}_m} \text{vec}(u_t z_t' - \widehat{C}(m)) [\text{vec}(u_t z_t' - \widehat{C}(m))]'$$

is a consistent estimator of $\Sigma_C(m)$. Note that we assume that the covariance between

the proxy and the shock does not depend on the sample size. In other words, we are not considering weak instrument asymptotics where the correlation between the instrument and the instrumented variable goes to zero when the sample size goes to infinity. An extension of our test to the weak instrument case may be an interesting topic for future research.

Let $\hat{\beta}(m)$ be the estimator of $\beta(m)$ obtained by replacing all unknown elements of $\beta(m)$ in expression (6) by the corresponding elements of $\hat{C}(m)$. Then Slutsky's theorem implies that

$$\sqrt{T} \left(\hat{\beta}(m) - \beta(m) \right) \xrightarrow{d} \mathcal{N} \left(0, V(m) \right), \quad (10)$$

where

$$V(m) = \frac{1}{\tau_m} \frac{\partial \beta(m)}{\partial \text{vec} C(m)'} \Sigma_C(m) \frac{\partial \beta(m)'}{\partial \text{vec} C(m)}$$

is the $((K-1) \times (K-1))$ asymptotic covariance matrix with $\frac{\partial \beta(m)}{\partial \text{vec} C(m)'}$ denoting the $((K-1) \times KN)$ matrix of partial derivatives of $\beta(m)$ with respect to the elements of $\text{vec} C(m)$. A closed form expression of this matrix is provided in Section S.1 of the Online Supplement. Having a closed form expression simplifies the computation of quantities based on the estimated covariance matrix, as we will see shortly. Alternatively, the covariance matrix $V(m)$ can be estimated by suitable bootstrap procedures which makes it computationally more involved.

It may be worth noting that the covariance matrix $V(m)$ has rank $K-1$ and, hence, the asymptotic distribution in (10) is nonsingular (see Section S.1 of the Online Supplement). Consequently, by asymptotic independence of $\hat{\beta}(m)$ and $\hat{\beta}(j)$ for $m \neq j$, we get that, under the null hypothesis $\mathbb{H}_0 : \beta(m) = \beta(j)$, the test statistic

$$\eta = T \left(\hat{\beta}(m) - \hat{\beta}(j) \right)' \left(\hat{V}(m) + \hat{V}(j) \right)^{-1} \left(\hat{\beta}(m) - \hat{\beta}(j) \right) \xrightarrow{d} \chi^2(K-1). \quad (11)$$

Thus, we can use this statistic for testing the pair of hypotheses (7). In the test

statistic in expression (11), the estimators of the covariance matrices are obtained as

$$\widehat{V}(m) = \frac{1}{\tau_m} \frac{\widehat{\partial\beta(m)}}{\partial\text{vec}C(m)'} \widehat{\Sigma}_C(m) \frac{\widehat{\partial\beta(m)'}}{\partial\text{vec}C(m)},$$

where all elements of $C(m)$ in $\frac{\partial\beta(m)}{\partial\text{vec}C(m)'}$ are replaced by their estimators from $\widehat{C}(m)$.

As mentioned earlier, in practice estimated errors \hat{u}_t have to be used instead of the true reduced-form errors u_t . Let $\hat{\eta}$ be the test statistic corresponding to η in (10) when the u_t are replaced by the ordinary least squares (OLS) residuals \hat{u}_t . The test statistic $\hat{\eta}$ has the same asymptotic distribution as η under mild conditions because the estimator of $C(m)$ can be viewed as a two-step GMM estimator as considered in Newey and McFadden (1994). For this result to hold, the assumption of consecutive observations in each subsample \mathcal{T}_m is not essential. The important condition for the asymptotic distribution in (11) to hold is that the number of observations in each volatility regime goes to infinity such that the asymptotic result in (9) is valid.

The test can, of course, be adjusted easily to test for time-varying individual elements of β . For example, to test the null hypothesis $\mathbb{H}_0 : \beta_i(m) = \beta_i(j)$, where $\beta_i(m)$ signifies the i^{th} element of $\beta(m)$, we denote the i^{th} column of I_{K-1} by ι_i and consider the test statistic

$$\hat{\eta}_i = T \left(\hat{\beta}(m) - \hat{\beta}(j) \right)' \iota_i \left(\iota_i' \widehat{V}(m) \iota_i + \iota_i' \widehat{V}(j) \iota_i \right)^{-1} \iota_i' \left(\hat{\beta}(m) - \hat{\beta}(j) \right) \quad (12)$$

with an asymptotic $\chi^2(1)$ distribution. Such a test is useful to investigate whether the impact effect of the shock on a single variable or a subset of the variables is time-varying.

In practice, the choice of the number of volatility regimes and the volatility change points is an important issue. In some cases, the number of regimes may be known from the subject matter and the change points may be known as dates where specific events have occurred that changed the volatility. Clearly, in practice the regimes have to be large enough so that a reliable estimation of the covariance between the proxy and the reduced form errors is possible. In the simulations in Section 4 we will

see that the test may even work if not all volatility regimes are taken into account, although its power may suffer if a change in volatility is ignored. In other words, even if short periods with higher or lower error variance are ignored, the test may find time-varying impact effects in other parts of the sample.

In practice, one may also consider standard statistical procedures to investigate heteroskedasticity, as discussed for the empirical example in Section 5. Of course, pretesting bias can result if statistical procedures are used to determine the change points. To assess the implications of misspecifying the change points, we will look at that situation in the simulations in Section 4.2.2. More generally, one may wonder about the applicability of the test if the actual generating mechanism of the heteroskedasticity is different from the model presented in equation (3). In the simulations in Section 4.2.2, we will also investigate the issue of applying the test under the assumption of the setup in equation (3) when, in fact, the true volatility generating mechanism is a different one.

Alternatively, one may consider a different volatility model. For example, one may consider a model where the volatility changes are generated by a Markov switching process as in Lanne et al. (2010) and Herwartz and Lütkepohl (2014). Such models have the advantage that the volatility change points are obtained from the data because they are endogenously generated by the model. If the volatility regimes are such that the result in (10) and hence the result in (11) holds, a test statistic such as $\hat{\eta}$ can be used as explained before. We emphasize, however, that our test is designed for models with a finite set of volatility regimes and, hence, our test cannot be used for stochastic volatility or GARCH models as considered by Primiceri (2005), Normandin and Phaneuf (2004) or Lütkepohl and Milunovich (2016).

We have presented our test for the case where just one structural shock is identified by the proxies. In the proxy VAR literature the case where several shocks are identified by proxies has also been considered (see, e.g., Mertens and Montiel-Olea (2018), Känzig (2019), Arias, Rubio-Ramírez and Waggoner (2020)). In that case, further assumptions are typically needed to separately identify the individual shocks.

If such assumptions are available, the impact effects can be estimated and tested separately for each identified shock with our test. Of course, one could also consider the impact effects of several shocks jointly and combine them in one test based on the same considerations used for constructing our test. Although such a generalization is straightforward in principle, we have refrained from presenting it because it would complicate the notation and it depends on the specific identifying assumptions used in addition to the proxies.

4 Monte Carlo Simulation

We have performed a simulation study to investigate the small sample properties of the test. In particular, we are interested in the rejection frequencies under the null and alternative hypotheses in finite samples and their dependence on the sample size, the size of the variance change, the strength of the instrument, the scale of the model in terms of the number of variables and the VAR lag order. We also investigate the behaviour of our test under different types of misspecifications that practitioners may be confronted with.

4.1 Setup of Monte Carlo Study

4.1.1 DGP 1

Our first data generating process (DGP) is a three-dimensional VAR(1) with reduced form

$$y_t = A_1 y_{t-1} + u_t \tag{13}$$

and a matrix of autoregressive slope coefficients

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

This matrix is based on an economic model discussed by An and Schorfheide (2007). It has eigenvalues 0.95, 0.90, and 0.51 and, hence, the process is stable with some persistence due to two roots close to the unit circle. Such features are occasionally met in practice.

Furthermore, we consider three volatility regimes ($M = 3$) with impact effects matrix $B(m) = I_3$ for $m = 1, 2, 3$ under the null hypothesis (\mathbb{H}_0) and

$$B(1) = I_3, \quad B(2) = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 4 \\ 4 & 6 & 6 \end{bmatrix}, \quad B(3) = \begin{bmatrix} 4 & 2 & 1 \\ -2 & 2 & 8 \\ 2 & 1 & 10 \end{bmatrix}$$

for the three volatility states under the alternative hypothesis (\mathbb{H}_1). This choice implies that

$$\beta(1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \beta(2) = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \quad \text{and} \quad \beta(3) = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix},$$

such that the three vectors to be tested are clearly distinct. The corresponding impulse responses are depicted in Figure S.1 of the Online Supplement. To study the properties of the test, we also have to consider the asymptotic variances $V(1)$, $V(2)$ and $V(3)$ of their estimators. They also depend on characteristics of the DGP and the proxy variable which will be specified subsequently.

Following standard conventions, we normalize the variances of the structural innovations in the first state to unity ($\Lambda_1 = I_3$). In line with comparable studies (Lütkepohl and Schlaak, 2018), we set the structural innovations' relative variances for the states $m = 2, 3$ such that a noticeable volatility pattern is generated. To achieve this, we choose

$$\Lambda_2 = \text{diag}(4, 9, 12) \quad \text{and} \quad \Lambda_3 = \text{diag}(1, 4, 9).$$

In our simulations, we draw the heteroskedastic structural innovation w_t from

$$w_t \sim \mathcal{N}(0, \Lambda_m), \quad \text{for } m = 1, 2, 3,$$

and we compute the heteroskedastic reduced-form residuals as $u_t = B(m)w_t$. Note that, under \mathbb{H}_1 , the time-varying instantaneous impact effects matrix and the volatility change of w_t contribute to the heteroskedasticity of u_t while, under \mathbb{H}_0 , the heteroskedasticity of u_t is driven exclusively by the variance changes of w_t . Using (13) and the previous specifications, we generate the data recursively with starting values $y_0 = 0$.

In most of our simulations, we consider a scalar proxy z_t because this is the most common case in practice. As in Caldara and Herbst (2019), the proxy is generated by assuming a linear relation with the instrumented structural innovation w_{1t} ,

$$z_t = \phi w_{1t} + v_t, \tag{14}$$

where ϕ and the error v_t determine the strength of the correlation between z_t and w_{1t} and, hence, the strength of the instrument. In principle, ϕ may vary for the different volatility regimes. We keep it constant at $\phi = 1$, however, and instead vary the strength of the proxy by changing the variance of the error v_t which is a noise term that mimics measurement errors of the instrument. It is generated independently of w_{1t} as

$$v_t \sim \mathcal{N}(0, \sigma_v^2(m)),$$

where the variance may change with the volatility regime. This setup implies a time-varying volatility of the instrument which can be observed in many time series of instruments that are used in the literature (see, for example, Romer and Romer, 2010; Miranda-Agrippino and Ricco, 2017).

Note that the strength of the relation between the instrument and the shock

w_{1t} determines how well the impact effects of the shock can be estimated and these estimates are the basis of our test statistic. Therefore it is important to note that the correlation between the instrument and the structural error w_{1t} can change with the variance of w_{1t} . The variance of z_t is $\text{Var}(z_t) = \phi^2 \text{Var}(w_{1t}) + \sigma_v^2(m)$. Hence, the correlation between w_{1t} and z_t is

$$\text{Corr}(w_{1t}, z_t) = \phi \sqrt{\text{Var}(w_{1t})} / \sqrt{\phi^2 \text{Var}(w_{1t}) + \sigma_v^2(m)}.$$

We consider the three different cases presented in Table 1. In Cases 1 and 2, the correlation between w_{1t} and z_t remains time-invariant across the sample, whereas in Case 3 the correlation and, hence, the strength of the instrument changes with the volatility regime, while the parameters of the relation (14) remain time-invariant. Instead of varying the measurement error variance $\sigma_v^2(m)$, we could have kept this variance constant throughout the sample period and instead vary the parameter ϕ in the relation (14). That would not affect the empirical rejection frequencies of our test as long as the correlation between the proxy and the error term remains unchanged. The correlations in Cases 2 and 3 in Table 1 are roughly in line with the correlations between the proxies and shocks used in a range of studies from the related literature (see Table S.1 in the Online Supplement for this article). The larger correlation of 0.9 in Case 1 in Table 1 was chosen to explore the potential power gains for our tests due to having a proxy which is strongly correlated with the shock.

Table 1: Specifications Used for DGP 1

Case	$\sigma_v^2(m)$	$\text{Corr}(w_{1t}, z_t)$	$V(1)$	$V(2)$	$V(3)$
1	$\sigma_v^2(1) = 0.2346$ $\sigma_v^2(2) = 0.9383$ $\sigma_v^2(3) = 0.2346$	0.9 $\forall t$	$\begin{bmatrix} 3.73 & -0.10 \\ -0.10 & 3.71 \end{bmatrix}$	$\begin{bmatrix} 48.86 & 87.69 \\ 87.69 & 334.90 \end{bmatrix}$	$\begin{bmatrix} 150.71 & 159.37 \\ 159.37 & 178.28 \end{bmatrix}$
2	$\sigma_v^2(1) = 3$ $\sigma_v^2(2) = 12$ $\sigma_v^2(3) = 3$	0.5 $\forall t$	$\begin{bmatrix} 12.09 & -0.33 \\ -0.33 & 11.81 \end{bmatrix}$	$\begin{bmatrix} 161.94 & 291.98 \\ 291.98 & 1098.91 \end{bmatrix}$	$\begin{bmatrix} 490.49 & 517.39 \\ 517.39 & 577.28 \end{bmatrix}$
3	$\sigma_v^2(m) = \sigma_v^2 = 1$	0.7071, $t \in \mathcal{T}_1$ 0.8944, $t \in \mathcal{T}_2$ 0.7071, $t \in \mathcal{T}_3$	$\begin{bmatrix} 6.05 & -0.17 \\ -0.17 & 5.96 \end{bmatrix}$	$\begin{bmatrix} 49.48 & 88.81 \\ 88.81 & 339.15 \end{bmatrix}$	$\begin{bmatrix} 244.26 & 257.86 \\ 257.86 & 287.96 \end{bmatrix}$

Having specified the characteristics of our DGP, we can also work out the asymptotic covariance matrices, $V(m)$, of the estimated $\beta(m)$ and we present them also in Table 1.² It is important to note that asymptotic covariances differ considerably for the different volatility regimes. The estimator $\hat{\beta}(1)$ has clearly the smallest asymptotic variances, while $\beta(2)$ and $\beta(3)$ are estimated less precisely in all three cases. This feature is expected to have an impact on the small sample power of the test as well. Note also that the $V(m)$ depend on the strength of the instrument. Generally, a stronger instrument implies more precise estimators of the $\beta(m)$, as one would expect.

We also consider the situation where two instruments are available for DGP 1. Specifically we consider instruments

$$z_t = \phi w_{1t} + v_t, \tag{15}$$

where z_t is a bivariate vector, $\phi = (\phi_1, \phi_2)' = (1, 1)'$ and $v_t \sim \mathcal{N}(0, \Sigma_v(m))$ is bivariate normal with

$$\Sigma_v(m) = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Thus, the two proxies are generated with properties as in Case 3 of Table 1, but they are correlated via the measurement error term v_t .

Data are simulated for sample sizes $T = 150, 300, 600, 1200$. The two breakpoints of the volatility states are set at $T_1 = \frac{1}{3}T$ and $T_2 = \frac{2}{3}T$. We use 5000 replications of each simulation design.

4.1.2 DGP 2

To investigate the performance of our test for VAR models with more variables, as they are often encountered in practice, we consider a specification based on a data set for the crude oil market used in a proxy VAR study by Känzig (2019). Changes in

²The underlying matrices $\Sigma_b(m)$ were computed by simulation.

volatility of oil market data have been diagnosed and studied in a number of previous papers (e.g., Lütkepohl and Netšunajev (2014), Lütkepohl, Meitz, Netšunajev and Saikkonen (2020)). Exact details of the model setup and its parameters are given in Section S.4 of the Online Supplement.

In our simulations we are fitting 6-dimensional VAR(13) models to the generated series. Thus, the model involves a large number of parameters. We consider two ($M = 2$) or three ($M = 3$) variance regimes. The B matrix under \mathbb{H}_0 and the $B(1)$, $B(2)$, $B(3)$ matrices used under \mathbb{H}_1 are also given in Section S.4 of the Online Supplement. The corresponding population impulse responses of the first structural shock are depicted in Figure S.2 of the Online Supplement.

The variances of the structural shocks are specified as follows:

$$\Lambda_1 = I_6, \quad \Lambda_2 = \text{diag}(0.5, 1, 2, 4, 8, 12), \quad \Lambda_3 = \text{diag}(2, 4, 8, 12, 16, 24),$$

if three volatility regimes are used. When two volatility regimes are considered, only Λ_1 and Λ_2 are used. The structural errors have Gaussian distributions such that $w_t \sim \mathcal{N}(0, \Lambda_m)$ and the reduced-form errors are computed as $u_t = B(m)w_t$. These errors are used to generate y_t samples for $T = 300, 600$, and 1200 . Volatility change points are at $T_1 = \frac{1}{2}T$ for $M = 2$ and at $T_1 = \frac{1}{3}T$ and $T_2 = \frac{2}{3}T$ for $M = 3$. In the simulations, we assume knowledge of the correct change points but also investigate the properties of the test when a change point is misspecified or one of the change points is ignored.

The instrument, z_t , is generated as in equation (14) with time-invariant $\phi = 0.5$ and $\sigma_v^2(m) = \sigma_v^2 = 0.5$ which implies the following correlations between shock and proxy:

$$\text{Corr}(w_{1t}, z_t) = \begin{cases} 0.5774 & \text{for } t \in \mathcal{T}_1, \\ 0.4472 & \text{for } t \in \mathcal{T}_2, \\ 0.7071 & \text{for } t \in \mathcal{T}_3. \end{cases}$$

Thus, for this DGP, the correlation is moderate at best.³

In Section S.4 of the Online Supplement we also show the $\beta(m)$ vectors and the corresponding asymptotic covariance matrices of their estimators. Clearly, $\beta(1)$ is estimated more precisely than $\beta(2)$ and $\beta(3)$. In fact, the estimator of $\beta(2)$ has the largest asymptotic variances. Again, this is expected to affect the power of our test for time-varying impact effects.

To investigate how important it is to specify the volatility model correctly, we not only consider the implications of misspecifying the volatility change points, but we also perform the test by assuming a heteroskedasticity model as in (3) when the change in variance is in fact generated by some other mechanism. More precisely, we have also generated a change in volatility by a smooth transition (ST) mechanism such that

$$\mathbb{E}(w_t w_t') = (1 - G(\gamma, \psi, t))\Lambda_1 + G(\gamma, \psi, t)\Lambda_2, \quad (16)$$

where a standard logistic transition function, $G(\gamma, \psi, t) = (1 + \exp[-\exp(\gamma)(t - \psi)])^{-1}$, is used. It depends on parameters γ and ψ and time t is chosen as transition variable. Using this model to generate the volatility changes implies that the volatility change from Λ_1 to Λ_2 occurs smoothly over a number of periods, depending on the parameters γ and ψ . The latter parameter determines the location of the variance change while γ determines the speed of the transition from variance regime Λ_1 to Λ_2 . Large values of γ imply a fast transition.

A transition from one volatility regime to another one over a number of sample periods may be more plausible in some situations in empirical studies than an instantaneous volatility adjustment in one period. Therefore it is of interest to explore the properties of our test if a one time volatility change is assumed although the actual volatility change is spread out over a longer period. We have used $\psi = 0.5T$ and $\gamma = -\sqrt{T}/10$ in our simulations which ensures a variance change roughly in the

³The correlation in Känzig's (2019) study is even slightly lower. For our simulations we have chosen correlations similar to Cesa-Bianchi et al. (2020) which is considered as an example in Section 5 (see Table S.2 of the Online Supplement for details).

Table 2: Relative Frequencies of F -Statistic for Weak Instrument Below 10 in Simulations of VAR(1) for DGP 1 and one Proxy

Sample size	Case 1			Case 2			Case 3		
	$m = 1$	$m = 2$	$m = 3$	$m = 1$	$m = 2$	$m = 3$	$m = 1$	$m = 2$	$m = 3$
$T = 150$	0.000	0.394	0.092	0.112	0.888	0.797	0.001	0.404	0.417
$T = 300$	0.000	0.053	0.002	0.011	0.731	0.478	0.000	0.057	0.061
$T = 600$	0.000	0.000	0.000	0.000	0.325	0.078	0.000	0.000	0.000
$T = 1200$	0.000	0.000	0.000	0.000	0.023	0.001	0.000	0.000	0.000

middle of the sample. The choice of γ implies that the variance change is largely completed over about 20% of the respective sample size. For example, for $T = 600$, the transition from variance regime Λ_1 to Λ_2 is basically completed between observations 240 and 360 of the sample. Under \mathbb{H}_1 , we also let the impact effects of the shocks change smoothly with the transition function $G(\gamma, \psi, t)$ from $B(1)$ to $B(2)$ such that the covariance structure of the reduced-form residuals smoothly changes from $\Sigma_u(1) = B(1)B(1)'$ to $\Sigma_u(2) = B(2)\Lambda_2B(2)'$. The number of replications of each simulation experiment is again 5000.

4.2 Simulation Results

4.2.1 DGP 1

Before looking at the results for our test for time-varying impact effects, it may be worth taking a look at the results in Tables 2 and 3, where we record the relative frequencies of standard F -tests for weak instruments (e.g., Kilian and Lütkepohl (2017, Section 15.2.1)) to produce values below 10 for the \mathbb{H}_1 processes in our simulations when a single proxy is used. In applied work, a threshold of 10 for the F -statistic is usually used to classify an instrument as sufficiently strong (see also Stock, Wright and Yogo (2002)). In other words, the numbers in Tables 2 and 3 indicate how often a proxy would have been classified as a weak instrument in our simulations with a scalar proxy. We also want to investigate the effect of the scale of the VAR model. Therefore, we have fitted VAR(1) and VAR(12) models to the data that are actually generated by a VAR(1).

Table 3: Relative Frequencies of F -Statistic for Weak Instrument Below 10 in Simulations of VAR(12) for DGP 1 and one Proxy

Sample size	Case 1			Case 2			Case 3		
	$m = 1$	$m = 2$	$m = 3$	$m = 1$	$m = 2$	$m = 3$	$m = 1$	$m = 2$	$m = 3$
$T = 150$	0.021	0.801	0.542	0.663	0.977	0.957	0.245	0.806	0.822
$T = 300$	0.000	0.170	0.012	0.107	0.859	0.673	0.000	0.176	0.176
$T = 600$	0.000	0.001	0.000	0.000	0.427	0.131	0.000	0.001	0.001
$T = 1200$	0.000	0.000	0.000	0.000	0.033	0.001	0.000	0.000	0.000

The strength of the instrument is clearly linked to the correlation between the shock and the proxy and it is seen in the tables that for Case 2 and small and moderate sample sizes, the instrument would be classified as weak in a large number of cases. For example, for $T = 150$, for the second volatility regime the proxy would have been classified as weak instrument in 88.8% of the replications when a VAR(1) is fitted and in 97.7%, if a VAR(12) process is fitted. For larger sample sizes and a more strongly correlated proxy and shock, the weak instrument classifications decline. However, it is, of course, of interest whether our test for time-varying impact effects also has power for weaker instruments for which the impact effects cannot be estimated very precisely. This is the reason for including Case 2 in our simulations.

The simulation results of the tests for time-varying impact effects for the first DGP are summarized in Table 4 for the case when a VAR(1) is fitted and in Table 5 for the VAR(12) case. In both cases there are three variance regimes and we have generated data both under \mathbb{H}_0 and \mathbb{H}_1 to explore the finite sample test size and power. The nominal significance level in both tables is 5%. Looking at the relative rejection frequencies of our test for the three cases of scalar proxies specified in Table 1 and the bivariate proxy in Tables 4 and 5, the following observations can be made.

1. The empirical test size and power tend to improve with larger sample sizes as they should be. In other words, the relative rejection frequencies tend to move closer to the nominal significance level of 5% under \mathbb{H}_0 and they increase with increasing sample size under \mathbb{H}_1 .
2. The empirical size in all cases is reasonably close to the nominal size of 5%. This

Table 4: Relative Rejection Frequencies for 3-dimensional VAR(1) Model of Pairwise Tests of $\mathbb{H}_0 : \beta(m) = \beta(k)$ for $m, k = 1, 2, 3$ and $m \neq k$ (Nominal Significance Level 5%, Volatility Change Points $T_1 = \frac{1}{3}T$ and $T_2 = \frac{2}{3}T$, 5000 MC Repetitions)

Case	Sample size	Data generated under \mathbb{H}_0 (size)			Data generated under \mathbb{H}_1 (power)		
		$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
1	$T = 150$	0.071	0.085	0.072	0.674	0.630	0.418
	$T = 300$	0.068	0.065	0.065	0.866	0.972	0.756
	$T = 600$	0.057	0.056	0.056	0.979	1.000	0.962
	$T = 1200$	0.046	0.056	0.049	1.000	1.000	1.000
2	$T = 150$	0.033	0.038	0.036	0.397	0.124	0.091
	$T = 300$	0.042	0.048	0.046	0.580	0.395	0.209
	$T = 600$	0.054	0.050	0.046	0.747	0.819	0.499
	$T = 1200$	0.047	0.048	0.051	0.915	0.995	0.840
3	$T = 150$	0.064	0.067	0.067	0.644	0.352	0.256
	$T = 300$	0.064	0.062	0.063	0.857	0.802	0.575
	$T = 600$	0.059	0.057	0.054	0.977	0.996	0.905
	$T = 1200$	0.050	0.054	0.051	1.000	1.000	0.998
2 proxies	$T = 150$	0.066	0.064	0.064	0.679	0.446	0.309
	$T = 300$	0.056	0.061	0.059	0.872	0.862	0.619
	$T = 600$	0.061	0.054	0.052	0.981	0.998	0.931
	$T = 1200$	0.051	0.054	0.052	1.000	1.000	0.998

Table 5: Relative Rejection Frequencies for 3-dimensional VAR(12) Model of Pairwise Tests of $\mathbb{H}_0 : \beta(m) = \beta(k)$ for $m, k = 1, 2, 3$ and $m \neq k$ (Nominal Significance Level 5%, Volatility Change Points $T_1 = \frac{1}{3}T$ and $T_2 = \frac{2}{3}T$, 5000 MC Repetitions)

Case	Sample size	Data generated under \mathbb{H}_0 (size)			Data generated under \mathbb{H}_1 (power)		
		$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
1	$T = 150$	0.077	0.070	0.063	0.450	0.244	0.241
	$T = 300$	0.073	0.066	0.066	0.805	0.882	0.651
	$T = 600$	0.059	0.059	0.057	0.967	1.000	0.941
	$T = 1200$	0.050	0.059	0.053	1.000	1.000	1.000
2	$T = 150$	0.028	0.023	0.020	0.131	0.033	0.039
	$T = 300$	0.045	0.036	0.034	0.455	0.233	0.152
	$T = 600$	0.052	0.047	0.044	0.703	0.731	0.447
	$T = 1200$	0.046	0.048	0.051	0.906	0.991	0.821
3	$T = 150$	0.063	0.053	0.045	0.368	0.116	0.143
	$T = 300$	0.070	0.055	0.059	0.786	0.616	0.458
	$T = 600$	0.057	0.056	0.053	0.964	0.982	0.867
	$T = 1200$	0.052	0.054	0.053	1.000	1.000	0.998
2 proxies	$T = 150$	0.066	0.052	0.048	0.415	0.146	0.160
	$T = 300$	0.058	0.058	0.053	0.801	0.701	0.509
	$T = 600$	0.060	0.054	0.051	0.970	0.994	0.905
	$T = 1200$	0.053	0.052	0.051	1.000	1.000	0.997

result holds even for the smaller sample sizes and if a large VAR(12) model is fitted. It holds for all four types of proxy variables. For Case 1, where the correlation between the proxy and the first structural shock is 0.9 and, hence, it is rather large across the sample, the rejection frequencies are slightly larger than 5% when the sample size is small ($T = 150$). Even then the largest relative rejection frequency is only 0.085 and most relative rejection frequencies under \mathbb{H}_0 are well below 0.075 in all situations reported in the tables.

3. Despite the similarity of the rejection frequencies under \mathbb{H}_0 , the empirical power is lower for the large VAR(12) model than for the smaller VAR(1) (compare the corresponding entries of Tables 4 and 5). Thus, the dimension of the parameter space of the model affects the power but not the test size in small samples.
4. A comparison of Cases 1 and 2 shows that the smaller correlation in Case 2 undermines the power of the test in small samples but does not have much of an effect on the rejection frequencies under \mathbb{H}_0 , for the sample sizes considered in our simulations. The power can be rather low for small samples. For example, the power for Case 2 for $T = 150$ for testing $\beta(2) = \beta(3)$ is only 0.091 and 0.039 for the VAR(1) and VAR(12), respectively. Clearly, if the proxy is not strong enough to estimate the impact effects with reasonable precision, the test is not able to discriminate well between the estimated impact effects of different volatility regimes.
5. In Case 3, the correlation between the proxy and the first shock varies over time. For all three volatility regimes it is between the correlation of Cases 1 and 2. This feature is reflected in the power results in Tables 4 and 5 which are typically in between the corresponding results for Cases 1 and 2.
6. The power increases if more identifying information in the form of two proxies is available. In that case, in particular for small samples, the power is larger than for the corresponding tests for Case 3, where only one proxy of the same type is available.

Overall our results based on DGP 1 suggest that the empirical size of the test and its power improve with increasing sample size. The empirical test size is not much affected by the scale of the model and the strength of the instrument. In contrast, the power declines with increasing scale of the model and it increases with the strength of the instrument.

Of course, one may argue that the DGPs encountered in applied work have rather different properties than our DGP 1. Therefore we have based our second DGP on an empirical model. The results are discussed next.

4.2.2 DGP 2

As explained in Section 4.1.2, our second DGP is similar to a model from Känzig (2019) based on actual data. The VAR model used is 6-dimensional and has lag order $p = 13$. Thus, the fitted model is large. Results for the case with one variance change are reported in Table 6. Given the results for DGP 1, it may be good to remember that the correlation between the proxy and the first shock is 0.5774 and 0.4472 in the first and second volatility regimes, respectively. Thus, the correlation is rather low.

For such a large model the relative rejection frequency of our test under \mathbb{H}_0 is slightly below the nominal 5%, if the sample size is small ($T = 300$). Even for a moderate sample size of $T = 600$, the relative rejection frequency under \mathbb{H}_0 is only 0.037 if the true volatility change point is used for the test. Thus, for large models the test tends to be conservative in small and even moderate samples.

Interestingly, if the volatility change point is misspecified, as it may easily happen in practice, the relative rejection frequencies under \mathbb{H}_0 on the right-hand side of Table 6 are slightly larger. However, the power declines if the volatility change point is misspecified. Given the simulation results for DGP 1, the low power for small samples is likely to be partly due to the low correlation between the proxy and the structural shock.

The performance of the test for a misspecified volatility change point is further

Table 6: Relative Rejection Frequencies of 6-dimensional VAR(13) Model of Tests of $\mathbb{H}_0 : \beta(1) = \beta(2)$ (Nominal Significance Level 5%, True Volatility Change Point at $T_1 = \frac{1}{2}T$, 5000 MC Repetitions)

Sample size	Assumed change point $T_1 = \frac{1}{2}T$		Assumed change point $T_1 = \frac{2}{5}T$	
	Under \mathbb{H}_0	Under \mathbb{H}_1	Under \mathbb{H}_0	Under \mathbb{H}_1
$T = 300$	0.018	0.194	0.026	0.132
$T = 600$	0.037	0.605	0.040	0.416
$T = 1200$	0.040	0.934	0.043	0.819

Table 7: Relative Rejection Frequencies of 6-dimensional VAR(13) Model of Tests of $\mathbb{H}_0 : \beta(1) = \beta(2)$ (Nominal Significance Level 5%, Volatility Change Generated by Smooth Transition, 5000 MC Repetitions)

Sample size	Assumed change point $T_1 = \frac{1}{2}T$		Assumed change point $T_1 = \frac{2}{5}T$	
	Under \mathbb{H}_0	Under \mathbb{H}_1	Under \mathbb{H}_0	Under \mathbb{H}_1
$T = 300$	0.022	0.184	0.026	0.161
$T = 600$	0.037	0.564	0.046	0.469
$T = 1200$	0.048	0.903	0.051	0.859

investigated in Table 7. In that table the relative rejection frequencies of our test are given for two different volatility change points although the true underlying generating mechanism of the volatility change is a smooth transition as explained in Section 4.1.2. Again the rejection frequencies under \mathbb{H}_0 are not much affected compared to those reported in Table 6 for a correctly specified change point ($T_1 = \frac{1}{2}T$). Even the reduction in power is quite moderate if the incorrect volatility model with one change point at a fixed fraction of the sample is assumed. Thus, our test is quite robust to slight misspecifications of the volatility change point and the type of heteroskedasticity.

Table 8: Relative Rejection Frequencies for 6-dimensional VAR(13) Model of Pairwise Tests of $\mathbb{H}_0 : \beta(m) = \beta(k)$ for $m, k = 1, 2, 3$ and $m \neq k$ (Nominal Significance Level 5%, Volatility Change Points $T_1 = \frac{1}{3}T$ and $T_2 = \frac{2}{3}T$, 5000 MC Repetitions)

Sample size	Data generated under \mathbb{H}_0 (size)			Data generated under \mathbb{H}_1 (power)		
	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
$T = 300$	0.013	0.058	0.016	0.096	0.503	0.037
$T = 600$	0.027	0.060	0.028	0.307	0.912	0.124
$T = 1200$	0.039	0.052	0.039	0.685	1.000	0.286

Table 9: Relative Rejection Frequencies of 6-dimensional VAR(13) Model with two Volatility Change Points When the Second Change Point is Ignored, $\mathbb{H}_0 : \beta(1) = \beta(2)$ with Volatility Change Point Assumed at $T_1 = \frac{1}{3}T$ (Nominal Significance Level 5%, 5000 MC Repetitions)

Sample Size	Under \mathbb{H}_0	Under \mathbb{H}_1
$T = 300$	0.047	0.462
$T = 600$	0.054	0.900
$T = 1200$	0.053	1.000

We have also performed simulations with three volatility regimes ($M = 3$) and show the results in Tables 8 and 9. In Table 8 the three volatility regimes and the change points are assumed to be known. In that case, for the null hypotheses $\mathbb{H}_0 : \beta(1) = \beta(2)$ and $\mathbb{H}_0 : \beta(2) = \beta(3)$, the relative rejection frequencies under \mathbb{H}_0 are slightly smaller than 5%. Although they move towards 0.05 for increasing sample size, they remain below 5% even for $T = 1200$. Note that both null hypotheses involve the impact effects vector for the second volatility regime where the correlation between the proxy and the structural shock is as low as 0.4472. Hence, it is also not surprising that the corresponding empirical power is small. In fact, for testing the null hypothesis $\beta(2) = \beta(3)$ and $T = 1200$, we get a relative rejection frequency of 0.286 if \mathbb{H}_1 is true. Such a low power is likely to be due to the low rejection frequency under \mathbb{H}_0 and the quite low correlation between the proxy and the shock. In any case, it may be worth remembering the potentially low power of the test when the example in Section 5 is considered.

In Table 9 we investigate the case where one of the volatility changes is ignored. Such a situation may arise in practice when there is uncertainty about the number of volatility regimes. The results in Table 9 are obtained by considering the first volatility change point at $T_1 = \frac{1}{3}T$ only and ignoring the second volatility change at $T_2 = \frac{2}{3}T$. In other words, the second volatility regime is incorrectly assumed to be $\mathcal{T}_2 = \{\frac{1}{3}T+1, \dots, T\}$ and the corresponding estimator $\hat{\beta}(2)$ mixes the second and third volatility regimes in the data. As can be seen in Table 9, this model misspecification does not undermine the empirical test size which is still close to 5% for all three

sample sizes. Also the empirical power of the test does not suffer much from the model misspecification. It is almost as large as for testing $\mathbb{H}_0 : \beta(1) = \beta(3)$ under \mathbb{H}_1 in Table 8.

4.2.3 Summary

The simulation results can be summarized briefly as follows. Larger samples improve the empirical size and power of the test. The dimension and lag order of the VAR process and precise knowledge of the type of heteroskedasticity and the timing of the variance change do not affect the small sample rejection frequencies substantially if the null hypothesis is true. However, these features affect the power of the test in small samples. More precisely, the power declines for a model of larger dimension, i.e., with more variables and/or larger lag order. The power can be improved by a more precise specification of the model for the second moments and the timing of the volatility change. Moreover, the power improves when the proxy has a stronger correlation with the structural shock.

5 Empirical Analysis

In the following, an empirical example from Cesa-Bianchi et al. (2020) is considered, where the authors use an instrumental variable to identify a monetary policy shock for the United Kingdom (UK). Their model has heteroskedastic errors. The study does not allow the heteroskedasticity to affect the impulse responses of the structural shock. Therefore we apply our test to investigate whether the implicit assumption of time-invariant impulse responses is in line with the data.

Cesa-Bianchi et al. (2020) construct a time series of intra-day price variation of the 3-month Sterling future contracts around policy decisions of the Monetary Policy Committee of the Bank of England. This series is used as proxy for monetary policy shocks in order to investigate their impact on macroeconomic and financial variables in the United Kingdom. In their baseline specification, Cesa-Bianchi et al. (2020)

consider a 7-dimensional VAR(2) model with a constant term for the variables

$$y_t = (r_t, cpi_t, ur_t, e_t, cs_t^{uk}, ms_t, cs_t^{us})',$$

where r_t is the nominal yield on the 1-year gilt, cpi_t is the consumer price index, ur_t describes the unemployment rate, e_t represents the nominal effective exchange rate, cs_t^{uk} and ms_t are UK corporate and mortgage bond spreads, respectively, whereas cs_t^{us} is a US corporate bond spread. All variables enter the model in levels, the only exception to this rule is cpi_t which is in natural logarithms.

The data for y_t are monthly for the sample period 1992m1-2015m1 which implies a sample size of 277 observations, including presample values for the estimation, hence, $T = 275$. For the proxy, data is only available for 1997m6-2015m1. We set the missing values for 1992m1-1997m5 to zero, as suggested by Mertens and Ravn (2019). This amounts to ignoring the sample period 1992m1-1997m5 in estimating the impact effects of the monetary policy shock (see equations (6) and (8)). We use the data from Cesa-Bianchi et al. (2020) and refer to that paper for further details on the precise variable specifications and data sources.⁴

In Figure 1, the OLS residuals of the reduced-form VAR(2) model are shown for the period 1997m6-2015m1, where the proxy is available and, hence, the impact effects of the monetary policy shock can be estimated. Obviously, the volatility of at least some of them has changed during the period 1997m6-2015m1. Especially as of the onset of the Great Recession the residual series of all financial variables display a markedly distinct volatility pattern. To identify the exact timing of the volatility break point, we assume that the bankruptcy filing of Lehman Brothers bank in the US and the turmoil on financial markets thereafter turned mild recessionary tendencies in the UK into a severe long-lasting economic crisis. Therefore, we choose August 2008 as preliminary volatility break and argue that September 2008 constitutes the beginning of a new volatility regime.⁵

⁴The data set is available as supplementary material to the paper on the web page of the *European Economic Review*.

⁵An LM test for heteroskedasticity as described in Lütkepohl (2005, pp. 600-601) yields a test

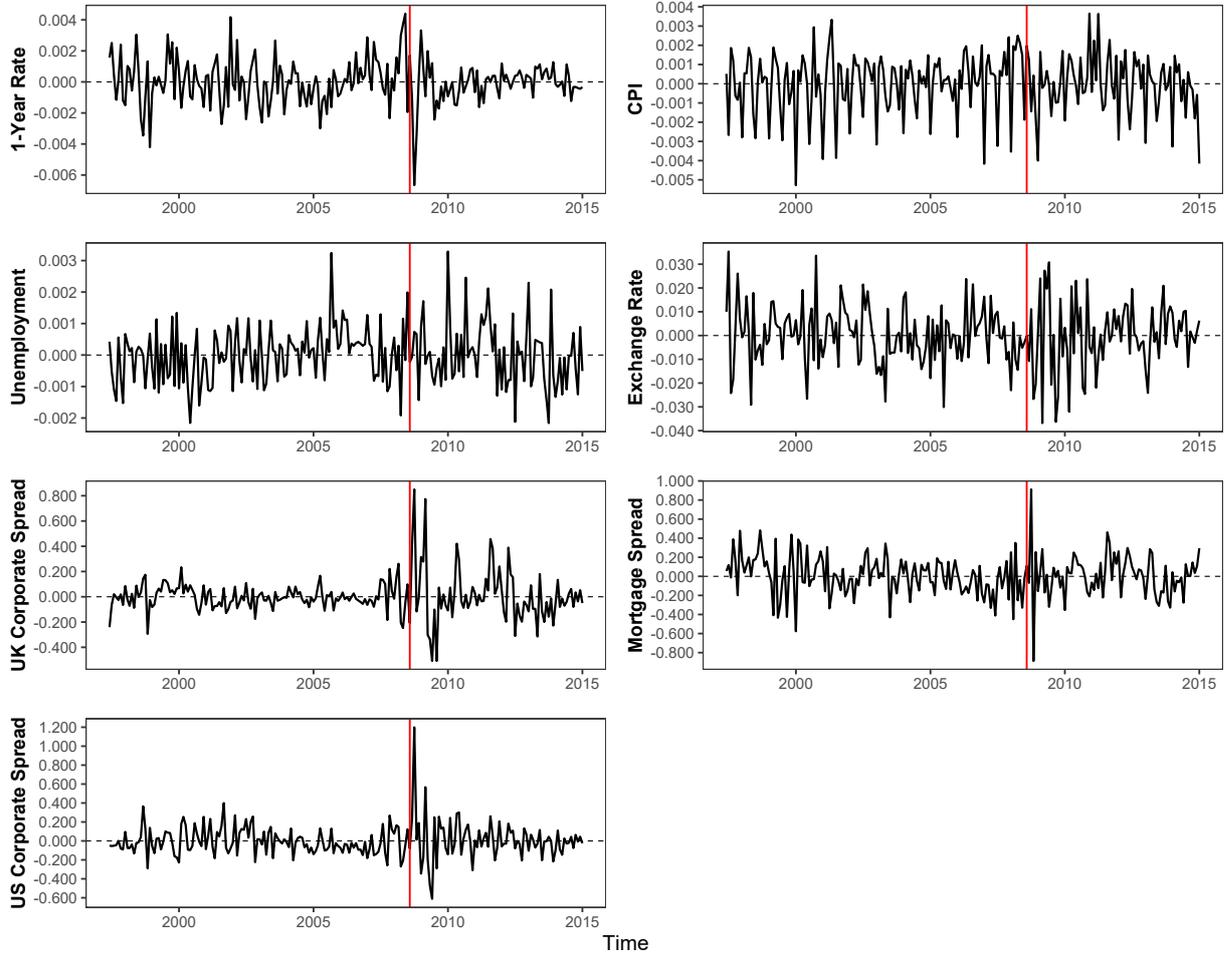


Figure 1: OLS residuals for 1997m6-2015m1 of the UK monetary policy VAR(2) model. The solid red line indicates the variance break in August 2008 suggested by the likelihood based statistical criterion in (17).

To back up our economic arguments, for the precise timing of the volatility change, we consider the Gaussian likelihood based criterion function

$$T_1 \log \det \widehat{\Sigma}_1 + (T - T_1) \log \det \widehat{\Sigma}_2, \quad (17)$$

where $\widehat{\Sigma}_1 = T_1^{-1} \sum_{t=1}^{T_1} \hat{u}_t \hat{u}_t'$ and $\widehat{\Sigma}_2 = (T - T_1)^{-1} \sum_{t=T_1+1}^T \hat{u}_t \hat{u}_t'$ are computed from OLS residuals \hat{u}_t . We minimize this objective function over $T_1 \in \{112, \dots, T - 48\}$ as we, first, exclude 63 observations at the beginning due to the missing values of the proxy and, second, we eliminate another 48 observations at both the beginning and the end statistic of 138.33 which clearly rejects the null hypothesis at the 1% significance level and, hence, provides strong evidence for a change in variance.

of the sample to have a sufficient amount of observations for a reasonably reliable estimation of the regime-specific covariance matrices and the underlying structural parameters. The procedure confirms the volatility change point in August 2008 (i.e., $T_1 = 198$, see solid red line in Figure 1).

Table 10: p -values of Tests for Time-Varying Impact Effects of UK Model for Different Variance Change Periods

Change period	Joint test	Tests for individual impact effects					
		cpi_t	ur_t	e_t	cs_t^{uk}	ms_t	cs_t^{us}
June 2008	0.009	0.461	0.044	0.393	0.242	0.010	0.675
July 2008	0.008	0.461	0.033	0.408	0.244	0.010	0.694
August 2008	0.010	0.886	0.028	0.653	0.363	0.012	0.867
September 2008	0.001	0.975	0.026	0.715	0.219	0.011	0.684
October 2008	0.001	0.975	0.026	0.715	0.219	0.011	0.684
November 2008	0.123	0.305	0.556	0.253	0.200	0.488	0.499
December 2008	0.148	0.534	0.934	0.487	0.196	0.274	0.611

Note: The p -values for September and October are identical because the proxy value for October is zero.

We have applied our test for time-varying impact effects of the monetary policy shocks for a number of covariance change periods around August 2008 and, in Table 10, we show the resulting p -values of our joint tests and tests of the last six variables individually as specified in equation (12). The results for August 2008 are highlighted in boldface in Table 10. Because we found in the simulations in Section 4 that the test may have low power for large models and relatively small samples sizes, the p -value 0.010 of the joint test for August 2008 indicates that the assumption of time-invariant impact effects is strongly rejected by the data.⁶ It is also apparent, however, that this is primarily due to time-varying impact effects of the unemployment rate, ur_t , and the mortgage spread, ms_t . Testing the impact effects of these two variables individually results in p -values well below 5% while all the other individual tests associated with a volatility change in August 2008 have rather large p -values.

In Table 10 we also show p -values for joint and individual tests for time-varying impact effects obtained by placing the volatility change in other months around the

⁶We have also computed correlations between the proxy and the estimated monetary policy shock and show them in Table S.2 of the Online Supplement. They are in the lower range of what was used in the simulations. Thus, the proxy is not a particularly strong instrument.

Lehman collapse. As our simulation results indicate that missing the actual volatility change period slightly does not invalidate the tests but reduces their power, one would expect the small p -values for a change in August 2008 to increase if the volatility change point is moved away from August 2008. It can indeed be seen in Table 10 that all p -values obtained for an assumed volatility change in November or December 2008 are greater than 10% and, hence, do not support a change in the impact effects of the variables.

We note in passing that the heteroskedasticity robust F -statistics for the strength of the proxy are well above 10 for all the subperiods corresponding to the volatility change points June to October 2008. They are presented in Table S.3 of the Online Supplement and indicate that the instrument is informative for the underlying structural innovations in both subperiods for each of the assumed volatility change periods. In contrast, the robust F -statistics for the subperiods 2008m12-2015m1 and 2009m1-2015m1, corresponding to volatility change periods November and December 2008, are smaller than 10 in Table S.3. Thus, the large p -values for November and December 2008 may just be a weak instrument problem. They may also be due to shifting part of the more volatile period following the Lehman collapse to the pre-crisis period and thereby reducing the power of the tests.

On the other hand, the p -values of the joint tests and the individual tests for ur_t and ms_t are still all clearly smaller than 5% if volatility changes in June, July, September or October are assumed. Overall, these p -values are well in line with our simulation results for the case of only slightly misspecified change periods and support the conclusion drawn for a volatility change in August 2008. In other words, the data support a change in the impact effects of the unemployment rate and the mortgage spread in that period.

In order to further investigate the implications of changes in the impact effects in some of the variables in August 2008, we have estimated impulse responses under the assumption of time-varying impact effects and show them in Figure 2. The confidence bands are obtained by a moving-block bootstrap as proposed by Brüggemann, Jentsch

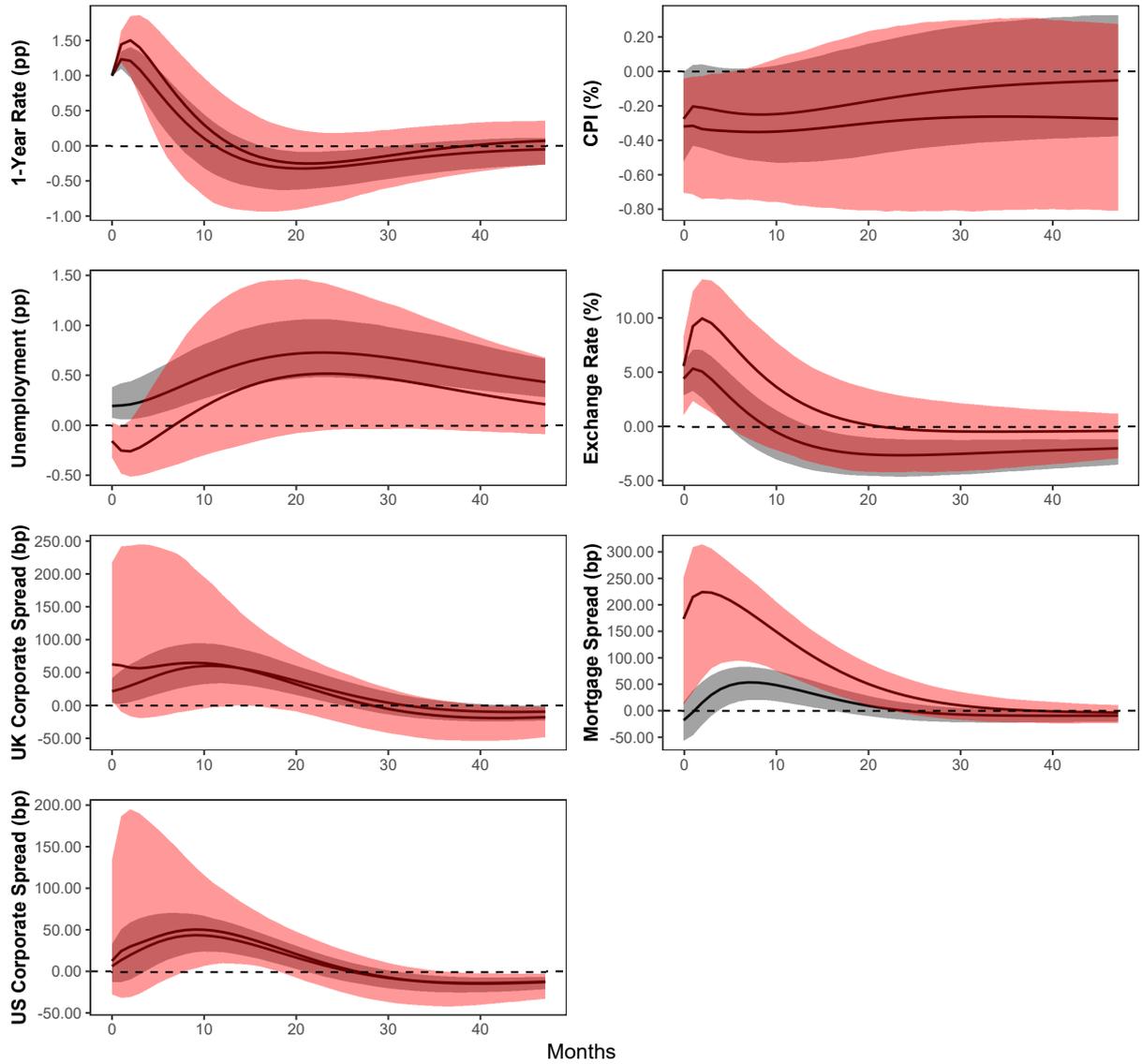


Figure 2: Responses to a monetary policy shock with pointwise 68% Efron percentile confidence bands centered at the point estimator, based on the moving-block bootstrap methodology described in Jentsch and Lunsford (2019) using a block length of $l = 4$ and 5000 bootstrap replications. Grey areas indicate confidence intervals for the sample prior to the Great Recession from 1992m3-2008m8 and red areas represent confidence intervals for the period during and after the Great Recession from 2008m9-2015m1.

and Trenkler (2016) and Jentsch and Lunsford (2019) who show the asymptotic validity of the bootstrap. Such a bootstrap can account for the possibility that there is more heteroskedasticity than what is captured by our model and was also used by Cesa-Bianchi et al. (2020). Unfortunately, the confidence bands in Figure 2 may be of limited value as the bootstrap is known to be unreliable in small samples (see Brüggemann et al. (2016) and Lütkepohl and Schlaak (2019)). Hence, the bootstrap confidence intervals in Figure 2 should be regarded with caution.⁷

Taking that into account, what can be seen in Figure 2 is that the bootstrap confidence intervals line up well with the test results for the individual impact effects. For the unemployment rate and the mortgage spread some impulse responses have non-overlapping 68% confidence intervals for short propagation horizons for the periods before and after the volatility change. For example, the responses of the mortgage spread are markedly larger in the second volatility state for a propagation period of one year although the statistical significance of such differences may not be clear from the figure. However, the test results in Table 10 provide evidence for significant effects for the unemployment rate. Hence, our test results indicate that monetary policy shocks in the UK have had different implications for the real economy in the two subsamples. Prior to the Great Recession an increase of interest rates led to a sizeable rise in unemployment. However, in the period during and after the Great Recession, contractionary monetary policy shocks were coupled with an initial decline in unemployment that lasted for several months. For all other variables, apart from the mortgage spread and the unemployment rate, the confidence intervals overlap for all propagation horizons which reflects the large p -values of our tests for these variables. Note that, although our test compares the impact effects only, these are the only parameters that are allowed to change in our setup. Thus, the differences in the

⁷We have also computed confidence intervals for the impulse responses using a wild bootstrap and present them in Figure S.4 of the Online Supplement. Although they provide somewhat smaller intervals for some of the impulse responses they are also problematic as Jentsch and Lunsford (2019) show that they are invalid asymptotically. Apart from that they convey a similar picture as Figure 2. The same is true for Hall confidence intervals that could be used instead of the Efron intervals in Figure 2. We show them in Figure S.3 of the Online Supplement. We also emphasize that we show 68% confidence intervals in Figure 2 and the other figures. Using a larger confidence level, such as 90%, the intervals would, of course, be even wider than those in the figures.

impulse responses at larger propagation horizons are determined by the differences in the impact effects suggested by our tests.

The impact effects of a monetary policy shock in Figure 2 are also different from the corresponding quantities presented in Figure 2 of Cesa-Bianchi et al. (2020) for their baseline model, where, for example, the initial response of unemployment and the mortgage spread are not significantly different from zero. We have plotted the impulse responses of Cesa-Bianchi et al. (2020) obtained by assuming time-invariant structural impulse responses in Figure S.5 in the Online Appendix. It can be seen that the respective impact effects in the original paper lie in between the impact effects in our Figure 2. It is obvious that the estimated impact effects are quite different from ours. Of course, given that, based on the test results in Table 10, some of our estimated impact effects are significantly different, at least some of them must also be different from those of Cesa-Bianchi et al. (2020) even if that may not be apparent from Figure S.5.

Despite the similarity of some of the impulse responses, this example shows that it is important to allow for the possibility of time-varying impact effects if shocks are identified by an instrument and the model errors are heteroskedastic. Not accounting for this possibility can lead to biased impulse responses.

There are, of course, other events that may have caused changes in the volatility of the series. We determined a second volatility change point by keeping the break in August 2008 fixed and performing a search over the remaining observations using a similar procedure as above, based on the criterion given in expression (17). Specifically, we ensure a minimum of at least 48 observations within one volatility regime. The criterion is minimized in August 2001, coinciding with the start of a long lasting economic upswing following the Dotcom bubble slowdown of the early 2000s. Assuming three volatility regimes, the first regime runs from March 1992 to August 2001 and may be interpreted as a medium volatility regime, whereas the second regime from September 2001 to August 2008 is a period of markedly low volatility in particular in the UK corporate bond spread series. The third regime coincides with the period

Table 11: Tests for Time-Varying Impact Effects of UK Model

\mathbb{H}_0	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
p -value	0.938	0.071	0.002

during and after the Great Recession from September 2008 to January 2015 and is coupled with high volatility notably in the two corporate bond spreads (see Figure 1).

We have applied our test for time-varying impact effects and present the p -values in Table 11. While the first null hypothesis, $\mathbb{H}_0 : \beta(1) = \beta(2)$, gives a p -value of 0.938 and, hence, cannot be rejected, the p -values are much smaller for the other two null hypotheses in Table 11. The null hypothesis $\mathbb{H}_0 : \beta(2) = \beta(3)$ results in a p -value of 0.002 and is clearly rejected at common significance levels. For $\mathbb{H}_0 : \beta(1) = \beta(3)$, with a p -value of 0.071, the situation is not quite so clear. Of course, not rejecting the null hypothesis at a 5% level of significance may just be a problem of low power of our test, as diagnosed in the small sample simulations reported in Section 4. Note also that the heteroskedasticity robust F -statistics for instrument strength for the periods 1997m1-2001m8 and 2001m9-2008m8 are 11.675 and 6.810, respectively (see Table S.3 in the Online Supplement). Thus, the proxy may be too weak to generate good power for the test for time-varying impact effects at least during the second volatility regime. In addition, the rather small sample periods associated with the first and second volatility regimes may cause low power of our tests. In any case, the test results in Table 11 do not provide much support for a change in the impact effects due to a possible variance change in August 2001.

Whatever the reason for our test results in this specific case may be, it appears to be a good idea to check for time-varying impact effects in heteroskedastic proxy VARs. As can be seen from our analysis of the volatility change in August 2008, ignoring the possibility may lead to distorted impulse responses.

6 Conclusions

In structural analyses with proxy VARs in the presence of heteroskedasticity of the data, it is typically assumed that the impact effects of the shocks of interest and, hence, the impulse responses at all propagation horizons are time-invariant and are not affected by the change in variance. We show that this assumption cannot be taken for granted in empirical studies in general. If the assumption is violated it can lead to biased impulse response estimates.

In this study, we have developed a test for time-varying impact effects of the shocks which is very easy to conduct. It can be applied even if the timing of the change in variance is not known precisely. Misspecifying the volatility change point does not invalidate the test but may lead to reduced power. Even if the mechanism that drives the heteroskedasticity of the data is not in line with the model assumed for our test, the test may have some power.

We have applied the test to an empirical example model from the structural VAR literature and reject time-invariance of the impact effects. The empirical example illustrates that the issue is relevant for applied work. Therefore we propose to test routinely for time-varying impact effects of a structural shock identified by a proxy variable when the data underlying a proxy VAR analysis are heteroskedastic. Our test is easy to perform and, hence, it opens up an easy way to improve the credibility of empirical results based on heteroskedastic proxy VARs.

References

- Alessi, L. and Kerssenfischer, M. (2019). The response of asset prices to monetary policy shocks: Stronger than thought, *Journal of Applied Econometrics* **34**: 661–672.
- An, S. and Schorfheide, F. (2007). Bayesian analysis of DSGE models, *Econometric Reviews* **26**: 113–172.
- Angelini, G., Bacchiocchi, E., Caggiano, G. and Fanelli, L. (2019). Uncertainty across volatility regimes, *Journal of Applied Econometrics* **34**: 437–455.
- Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2020). Inference in Bayesian proxy-SVARs, manuscript.

- Brüggemann, R., Jentsch, C. and Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **191**: 69–85.
- Caldara, D. and Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy svars, *American Economic Journal: Macroeconomics* **11**: 157–192.
- Cesa-Bianchi, A., Thwaites, G. and Vicondoa, A. (2020). Monetary policy transmission in the United Kingdom: A high frequency identification approach, *European Economic Review* **123**: 103375.
- Dias, D. A. and Duarte, J. B. (2019). Monetary policy, housing rents, and inflation dynamics, *Journal of Applied Econometrics* **34**: 673–687.
- Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal: Macroeconomics* **7**: 44–76.
- Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**: 104–116.
- Jentsch, C. and Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Comment, *American Economic Review* **109**: 2655–2678.
- Känzig, D. R. (2019). The macroeconomic effects of oil supply news: Evidence from OPEC announcements, *Technical report*, London Business School.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press, Cambridge.
- Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics and Control* **34**: 121–131.
- Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley & Sons, Chichester.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Lütkepohl, H., Meitz, M., Netšunajev, A. and Saikkonen, P. (2020). Testing identification via heteroskedasticity in structural vector autoregressive models, *Econometrics Journal* **23**.
- Lütkepohl, H. and Milunovich, G. (2016). Testing for identification in SVAR-GARCH models, *Journal of Economic Dynamics and Control* **73**: 241–258.
- Lütkepohl, H. and Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**: 479–496.

- Lütkepohl, H. and Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis, *Oxford Bulletin of Economics and Statistics* **80**: 715–735.
- Lütkepohl, H. and Schlaak, T. (2019). Bootstrapping impulse responses of structural vector autoregressive models identified through GARCH, *Journal of Economic Dynamics and Control* **101**: 41–61.
- Mertens, K. and Montiel-Olea, J. L. (2018). Marginal tax rates and income: New time series evidence, *Quarterly Journal Of Economics* **133**: 1803–1884.
- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**: 1212–1247.
- Mertens, K. and Ravn, M. O. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Reply, *American Economic Review* **109**(7): 2679–2691.
- Miranda-Agrippino, S. and Ricco, G. (2017). The Transmission of Monetary Policy Shocks, *Economic Research Papers 269310*, University of Warwick - Department of Economics.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing, in R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, Amsterdam, chapter 36, pp. 2111–2245.
- Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* **51**: 1217–1243.
- Piffer, M. and Podstawski, M. (2018). Identifying uncertainty shocks using the price of gold, *The Economic Journal* **128**: 3266–3284.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**: 821–852.
- 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.
- Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**: 54–81.
- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* pp. 81–135.
- Stock, J. H., 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.