Lütkepohl, Helmut; Schlaak, Thore

Working Paper

Heteroskedastic proxy vector autoregressions

DIW Discussion Papers, No. 1876

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Lütkepohl, Helmut; Schlaak, Thore (2020) : Heteroskedastic proxy vector autoregressions, DIW Discussion Papers, No. 1876, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/222858

DIW BERLIN

1876

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2020

Heteroskedastic Proxy Vector Autoregressions

Helmut Lütkepohl and Thore Schlaak

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

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: tschlaak@diw.de

This version: June 10, 2020

Abstract. In proxy vector autoregressive models, the structural shocks of interest are identified by an instrument. Although heteroskedasticity is occasionally allowed for, 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

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) and Alessi and Kerssenfischer (2019). Other authors using proxy VARs consider variables in their models that may induce heteroskedasticity and they assume that it does not affect the impulse responses (e.g., Gertler and Karadi (2015) and 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 important 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.

We reconsider studies by Cesa-Bianchi, Thwaites and Vicondoa (2020) and Dias and Duarte (2019) which use heteroskedastic proxy VAR models and assume time-invariant impulse responses. We apply our test to confront the data with the time-invariance assumption. The study by Cesa-Bianchi et al. (2020) considers 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. Our second example is based on Dias and Duarte (2019). It investigates the impact of monetary policy shocks on housing rents in the United States of America (US). We find that the model errors display clear changes in volatility during the sample period. However, our test does not find evidence against the implicit assumption of time-invariant impulse responses.

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 two example models 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, $w_t = (w_{1t}, \ldots, w_{Kt})'$, 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 an instrumental variable (proxy) z_t satisfying

$$\mathbb{E}(w_{1t} z_t) = c \neq 0 \quad \text{(relevance)}, \tag{1}$$
$$\mathbb{E}(w_{kt} z_t) = 0, \quad k = 2, \ldots, K, \quad \text{(exogeneity)}. \tag{2}$$

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

Now suppose that there is heteroskedasticity such that

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

where $\mathcal{T}_m = \{T_{m-1} + 1, \ldots, T_m\}$ $(m = 1, \ldots, M)$ are M given volatility regimes associated with consecutive time periods. The volatility changes occur at time periods T_m for $m = 1, \ldots, M-1$, 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, 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

$$\Sigma_u(m) = B\Lambda_m B', \quad m = 1, \ldots, M, \tag{4}$$

where $\Lambda_m = \text{diag}(\lambda_{1,m}, \ldots, \lambda_{K,m})$ $(m = 1, \ldots, M)$ 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 instrument and the first structural shock,

$$\mathbb{E}(w_{1t}z_t) = c_m \neq 0, \quad t \in \mathcal{T}_m, \quad m = 1, \ldots, M \quad \text{(relevance)}. \tag{5}$$

Hence, $\mathbb{E}(z_t u_t) = c_m b(m)$ 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 $m = 1, \ldots, M$. 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 may be 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.

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, $m = 1, \ldots, M$. 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 instrument identifies the first shock, as in the previous section and, for simplicity, we denote by $b(m)$ the vector of covariances $\mathbb{E}(z_t u_t)$ for $t \in \mathcal{T}_m$. Moreover, the first element of $b(m) = (b_1(m), \ldots, b_K(m))'$ is assumed to be nonzero, i.e., $b_1(m) \neq 0$, and we define the $(K-1)$-dimensional vector $\beta(m) = (b_2(m), \ldots, b_K(m))'/b_1(m)$. For $m, k \in \{1, \ldots, M\}$, $m \neq k$, we wish to test the pair of hypotheses

$$\mathbb{H}_0 : \beta(m) = \beta(k) \quad \text{versus} \quad \mathbb{H}_1 : \beta(m) \neq \beta(k). \tag{6}$$

In the following, a suitable test statistic is presented.

Let

$$\hat{b}(m) = \frac{1}{\tau_m T} \sum_{t \in \mathcal{T}_m} z_t u_t, \tag{7}$$

be an estimator of the first column $b(m)$ of $B(m)$ in volatility regime $m \in \{1, \ldots, M\}$. We assume that $\tau_m = (T_m - T_{m-1})/T$ is a fixed fraction of the sample size so that $T_m - T_{m-1} \to \infty$ with T. Moreover, u_t and z_t are such that $\hat{b}(m)$ is consistent and asymptotically normal,

$$\sqrt{T}\left(\hat{b}(m) - b(m)\right) \xrightarrow{d} \mathcal{N}(0, \tau_m^{-1}\Sigma_b(m)), \tag{8}$$

where $\xrightarrow{d}$ denotes convergence in distribution. Under general conditions, such a result follows from a central limit theorem, i.e., if $z_t u_t$ is independently, identically distributed (iid) for $\mathcal{T}_m$ or a martingale difference sequence. We also assume that $z_t u_t$ is such that

$$\widehat{\Sigma}_b(m) = \frac{1}{\tau_m T} \sum_{t \in \mathcal{T}_m} (z_t u_t - \hat{b}(m))(z_t u_t - \hat{b}(m))'$$

is a consistent estimator of $\Sigma_b(m)$.

Slutsky's theorem implies that for $\hat{\beta}(m) = (\hat{b}_2(m), \ldots, \hat{b}_K(m))'/\hat{b}_1(m)$,

$$\sqrt{T}\left(\hat{\beta}(m) - \beta(m)\right) \xrightarrow{d} \mathcal{N}\left(0, V(m)\right), \tag{9}$$

where

$$V(m) = \frac{1}{\tau_m} \frac{\partial \beta(m)}{\partial b(m)'} \Sigma_b(m) \frac{\partial \beta(m)'}{\partial b(m)}$$

is the $((K-1) \times (K-1))$ asymptotic covariance matrix with $\frac{\partial \beta(m)}{\partial b(m)'}$ denoting the $((K-1) \times K)$ matrix of partial derivatives of $\beta(m)$ with respect to the elements of $b(m)$, that is,

$$\frac{\partial \beta(m)}{\partial b(m)'} = \begin{bmatrix} -\frac{b_2(m)}{b_1(m)^2} & \frac{1}{b_1(m)} & 0 & \ldots & 0 \\ -\frac{b_3(m)}{b_1(m)^2} & 0 & \frac{1}{b_1(m)} & & 0 \\ \vdots & \vdots & & \ddots & 0 \\ -\frac{b_K(m)}{b_1(m)^2} & 0 & \ldots & & \frac{1}{b_1(m)} \end{bmatrix}$$

and $\frac{\partial \beta(m)'}{\partial b(m)} = \left(\frac{\partial \beta(m)}{\partial b(m)'}\right)'$. It may be worth noting that this matrix has rank $K-1$ and, hence, the covariance matrix of the asymptotic distribution in (9) is nonsingular. Consequently, by asymptotic independence of $\hat{\beta}(m)$ and $\hat{\beta}(k)$ for $m \neq k$, we get that, under the null hypothesis $\mathbb{H}_0 : \beta(m) = \beta(k)$, the test statistic

$$\eta = T\left(\hat{\beta}(m) - \hat{\beta}(k)\right)' \left(\widehat{V}(m) + \widehat{V}(k)\right)^{-1} \left(\hat{\beta}(m) - \hat{\beta}(k)\right) \xrightarrow{d} \chi^2(K-1). \tag{10}$$

Thus, we can use this statistic for testing the pair of hypotheses (6). In the test statistic in expression (10), the estimators of the covariance matrices are obtained as

$$\widehat{V}(m) = \frac{1}{\tau_m} \widehat{\frac{\partial \beta(m)}{\partial b(m)'}} \widehat{\Sigma}_b(m) \widehat{\frac{\partial \beta(m)'}{\partial b(m)}},$$

where all elements of $b(m)$ in $\frac{\partial \beta(m)}{\partial b(m)'}$ are replaced by their estimators from $\hat{b}(m)$.

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 (9) 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 $b(m)$ 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 (10) to hold is that the number of observations in each volatility regime goes to infinity such that the asymptotic result in (8) is valid.

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, the dependence on the strength of the instrument, the magnitude of the model in terms of the number of variables and the VAR lag order.

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

$$y_t = A_1 y_{t-1} + u_t \tag{11}$$

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 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. 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

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

Our setup implies that we draw the heteroskedastic structural innovation w_t from

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

and we compute the heteroskedastic reduced-form residuals as $u_t = B(m)w_t$. Note that the time-varying instantaneous impact effects matrix and the volatility change of w_t contribute to the heteroskedasticity of u_t. Using (11) and the previous specifications, we generate the data recursively with starting values $y_0 = 0$.

Finally, as in Caldara and Herbst (2019), the proxy z_t is generated by assuming a linear relation with the instrumented structural innovation w_{1t},

$$z_t = \phi w_{1t} + v_t, \tag{12}$$

where ϕ and the error v_t determine the strength of the correlation between z_t and w_{1t}. 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

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

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 (12) 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 (12). 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.

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.[1] 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.

Data are simulated for sample sizes $T = 300, 600, 1200$. 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.

[1] The underlying matrices $\Sigma_b(m)$ were computed by simulation.

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

4.1.2 DGP 2

To investigate the performance of our test for larger VAR models, 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 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 Appendix A.1.

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 Appendix A.1.

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.

The instrument, z_t, is generated as in equation (12) 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.

In Appendix A.1 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

$$\mathbb{E}(w_t w_t') = (1 - G(\gamma, \psi, t))\Lambda_1 + G(\gamma, \psi, t)\Lambda_2, \tag{13}$$

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 middle of the sample which 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. 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 $\Sigma_u(2) = B(2)\Lambda_2 B(2)'$. The number of replications of each simulation experiment is again 5000.

4.2 Simulation Results

4.2.1 DGP 1

The simulation results for the first DGP are summarized in Table 4 for the case when a VAR(1) is fitted. Since we also want to investigate the effect of the magnitude of the VAR model, we have in addition fitted VAR(12) processes to the data that are actually generated by a VAR(1) and report the results in Table 5. 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%.

Table 2: Relative Frequencies of F-Test for Weak Instrument Below 10 in Simulations of VAR(1)

	Case 1			Case 2			Case 3		
Sample size	$m=1$	$m=2$	$m=3$	$m=1$	$m=2$	$m=3$	$m=1$	$m=2$	$m=3$
$T=300$	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

Table 3: Relative Frequencies of F-Test for Weak Instrument Below 10 in Simulations of VAR(12)

	Case 1			Case 2			Case 3		
Sample size	$m=1$	$m=2$	$m=3$	$m=1$	$m=2$	$m=3$	$m=1$	$m=2$	$m=3$
$T=300$	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

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. In applied work, a threshold of 10 is usually

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	Data generated under $\mathbb{H}_0$ (size)			Data generated under $\mathbb{H}_1$ (power)		
	size	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
1	$T = 300$	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	$T = 300$	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	$T = 300$	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

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 an instrument would have been classified as weak in our simulations. This feature 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 $T = 300$, for the second regime the proxy would have been classified as weak instrument in more than 73% and 85% of the replications, if VAR(1) and VAR(12) processes are fitted, respectively. For larger sample sizes and a more strongly correlated proxy and shock, a weak instrument problem would not be diagnosed very often in any of the cases and regimes. However, it is, of course, of interest whether our test for time-varying impact effects also has power when the instrument is weak and thus, the impact effects cannot be estimated very precisely. This is the reason for including Case 2 in our simulations.

Now looking at the relative rejection frequencies of our test 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 close to the nominal size of 5%. This

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	Data generated under $\mathbb{H}_0$ (size)			Data generated under $\mathbb{H}_1$ (power)		
	size	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
1	$T = 300$	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	$T = 300$	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	$T = 300$	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

result holds even for the smaller sample sizes and if a large VAR(12) model is fitted. It holds for all three 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 ($T = 300$). Even then the 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 magnitude 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 $T = 300$ for testing $\beta(2) = \beta(3)$ is only 0.209 and 0.152 for the VAR(1) and VAR(12), respectively. Clearly, if the instrument 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.

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 magnitude of the model and the strength of the instrument. In contrast, the power declines with increasing magnitude 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 trimmed our second DGP towards an empirical model. The results are discussed next.

4.2.2 DGP 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 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 generation mechanism of the volatility change is a smooth

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)

	Assumed change point $T_1 = \frac{1}{2}T$		Assumed change point $T_1 = \frac{2}{5}T$	
Sample size	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)

	Assumed change point $T_1 = \frac{1}{2}T$		Assumed change point $T_1 = \frac{2}{5}T$	
Sample size	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

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)

	Data generated under $\mathbb{H}_0$ (size)			Data generated under $\mathbb{H}_1$ (power)		
Sample size	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$	$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
$T = 300$	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

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.

We have also performed simulations with three volatility regimes ($M = 3$) and show the results in Table 8. 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 too small. 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 instrument 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 $\mathbb{H}_0 : \beta(2) = \beta(3)$ and $T = 1200$, we get a relative rejection frequency of 0.286. 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 examples in Section 5 are considered.

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, although for large models the test can be conservative. However, these features affect the power

of the test in small samples. More precisely, the power declines for a model of larger magnitude, 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.

5 Empirical Analysis

In the following, two empirical examples from the proxy VAR literature are considered, where the authors use instrumental variables to identify shocks of interest and which also have heteroskedastic errors. None of the studies allows the heteroskedasticity to affect the impulse responses of the structural shocks. Therefore we apply our test to investigate whether the implicit assumption of time-invariant impulse responses is in line with the data.

5.1 Monetary Policy Transmission in the United Kingdom

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 instrumental variable 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 are monthly for the sample period 1992m1-2015m1 which implies a sample size of 277 observations, including presample values for the estimation, hence, $T = 275$. 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.[2]

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

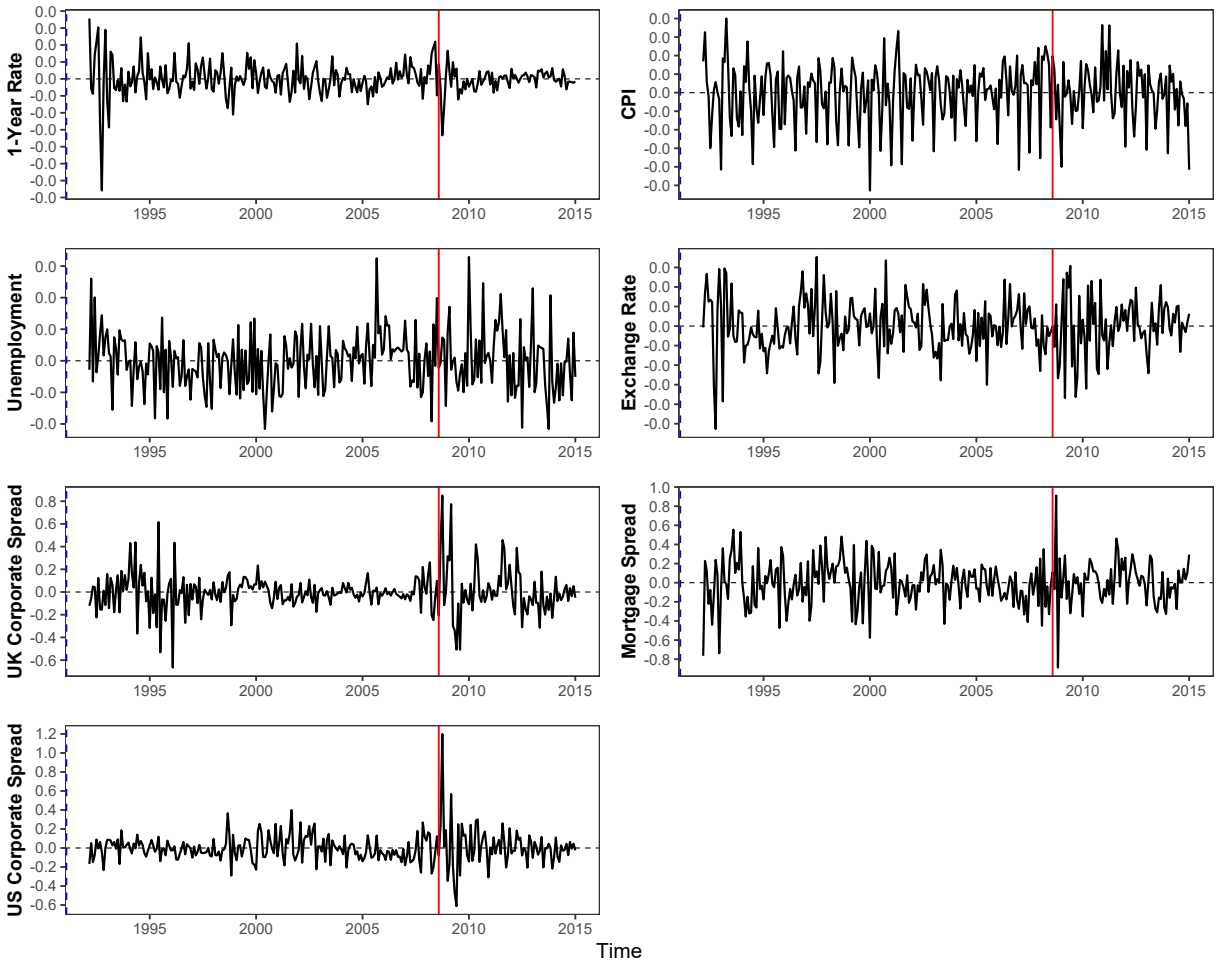

Figure 1: OLS residuals of UK monetary policy VAR(2) model. The solid red line indicates the variance break in August 2008 suggested by the likelihood based statistical criterion in (14).

Looking at the OLS residuals of the reduced-form VAR(2) model in Figure 1, it is obvious that the volatility of at least some of them has changed during the sample period. 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.[3] Moreover, to back

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

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

$$T_1 \log\det\widehat{\Sigma}_1 + (T - T_1)\log\det\widehat{\Sigma}_2, \tag{14}$$

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 $T_1 \in \{112,\ldots,T-48\}$ as we, first, exclude 63 observations at the beginning due to missing values of the instrument (see Section 3.3 in Cesa-Bianchi et al. (2020)) and, second, we eliminate another 48 observations at both the beginning and the end of the sample to have a sufficient amount of observations for 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). Hence, we apply our test for time-varying impact effects of the monetary policy shocks using a variance change in August 2008 (2008m8). The resulting p-value of our test is 0.010. 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 indicates that the assumption of time-invariant impact effects is strongly rejected by the data.

We checked whether the F-statistics resulting from a regression of the OLS residuals of the 1-year gilt on the instrument in both regimes exceeds the value of 10, which is a well-established condition for a sufficiently strong instrument in practice, as mentioned earlier. The F-statistics of 14.835 and 15.181, respectively, indicate that the instrument is informative for the underlying structural innovations in both periods. Therefore we have estimated separate impact effects for the two volatility regimes and present the corresponding impulse responses in Figure 2. The confidence bands are obtained by wild bootstraps to account for the possibility that there is more heteroskedasticity than what is captured by our model.[4]

[4]We generate the bootstrapped residuals $u_t^\star = \varphi_t\hat{u}_t$ and the bootstrapped instrument $z_t^\star = \varphi_t z_t$, where φ_t is a random variable following a Rademacher distribution, i.e., φ_t is either 1 or –1 with probability 0.5. The bootstrapped residuals are then used to generate bootstrapped data $y_t^\star$ recursively. Each bootstrap sample is based on identical presample values from the original data set as initial values, i.e., $y_{-p+1}^\star = y_{-p+1},\ldots,y_0^\star = y_0$. The bootstrap procedure is repeated 5,000 times. Instead, Cesa-Bianchi et al. (2020) use a moving-block-bootstrap to capture heteroskedasticity. It was shown by Brüggemann, Jentsch and Trenkler (2016) and Jentsch and Lunsford (2019) that the latter is asymptotically valid in heteroskedastic structural VAR models while the wild bootstrap does not provide asymptotically correct confidence intervals. However, in small samples such as ours with less than 100 observations for some subsamples, the wild bootstrap may have advantages in a related setting (see Brüggemann et al. (2016) and Lütkepohl and Schlaak (2019)).

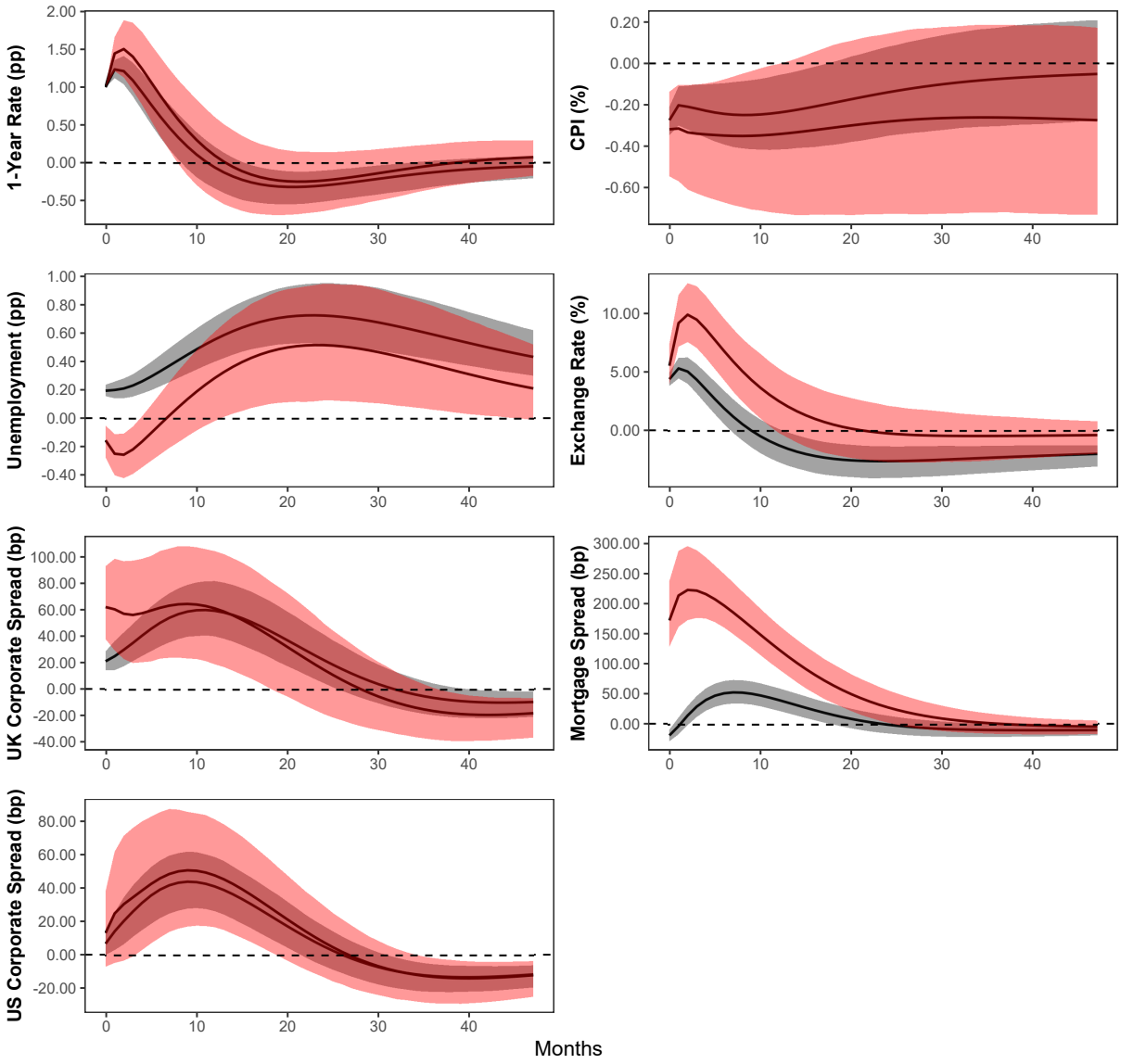

Figure 2: Responses to a monetary policy shock with pointwise 68% confidence bands based on a wild bootstrap. Grey areas indicate confidence intervals for the sample prior to the Great Recession from 1992m3-2008m8 and red areas represent confidence intervals for the the period during and after the Great Recession from 2008m9-2015m1.

In Figure 2 it can be seen that some impulse responses for short propagation horizons have non-overlapping 68% confidence intervals before and after the volatility change. For example, the responses of the mortgage spread is strikingly larger in the second volatility state for a propagation period of roughly two years. This finding is similar, though less pronounced, for other financial variables like the effective nominal exchange rate and the UK corporate bond spread. Interestingly, also the response of the unemployment rate to a contractionary monetary policy shock in the respective volatility regimes varies quite substantially at short propagation horizons of roughly ten months. In other words, according to our results, 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 sizable and significant 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.

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. The size of the respective impact effects in the original paper lies in between the impact effects in our Figure 2. Although Cesa-Bianchi et al. (2020) use a different bootstrap method to construct confidence intervals around their impulse responses, it is obvious that the estimated impact effects are quite different from ours.[5] Even though our wild bootstrap may understate the true uncertainty in the impulse responses and even though our confidence intervals overlap with those of Cesa-Bianchi et al. (2020), our test presents evidence that at least some of the impact effects are different before and after the Great Recession. In contrast, the longer term effects of the monetary policy shocks are similar in our Figure 2 and Figure 2 of Cesa-Bianchi et al. (2020). Thus, the data support primarily differences in the short-term effects of a monetary policy shock.

Despite the similarity of the long-term effects, 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 heteroskedas-

[5]As a robustness check, we have also computed impulse response functions with confidence bands relying on the moving-block-bootstrap used by Cesa-Bianchi et al. (2020) and show them in Figure A.1 in Appendix A.2. While our main results for the responses of the mortgage spread and the unemployment rate still hold, others get somewhat more ambiguous as expected from the small number of observations especially in the second volatility regime.

Table 9: Tests for Time-Varying Impact Effects of UK Model

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

tic. 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 described in (14). 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 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 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 9. 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 9. 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.

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 we have seen from our estimations, ignoring the possibility may lead to distorted impulse responses.

5.2 US Monetary Policy and Housing Rents

Dias and Duarte (2019) investigate the impact of monetary policy shocks

on housing rents in the US based on a proxy VAR. They use monthly data from January 1983 to December 2017. Thus, the full sample size amounts to 420 observations including presample values. The variables of their baseline model are

$$y_t = (r_t, ip_t, cpi_t, ebp_t, hr_t)',$$

where r_t is the 1-year treasury rate, ip_t is industrial production, cpi_t is the consumer price index, ebp_t is the excess bond premium, and hr_t represents housing rents. All variables apart from the interest rate (r_t) and the excess bond premium (ebp_t) are in logs. Dias and Duarte fit a VAR(12) model with an intercept as their baseline model and they use an instrument from Gertler and Karadi (2015) to identify the monetary policy shocks. The instrument consists of the changes in the 3-months-ahead monthly federal funds futures around the dates of monetary policy announcements. Dias and Duarte use a moving-block-bootstrap method for their impulse responses that accounts for possible heteroskedasticity. Thereby they acknowledge the possibility of heteroskedasticity in the model errors.

We use the baseline setup and data from Dias and Duarte (2019) to investigate the time-invariance assumption underlying their study.[6] Given the sample period and looking at the VAR residuals in Figure 3, it is plausible to allow for a change in volatility around the time of the outbreak of the financial crisis in 2008/09. We have investigated possible regime switches at the onset of the recession in November 2007 and one year later, in November 2008, around the time when short term interest rates reached the effective zero lower bound and unconventional monetary policy measures were implemented by the Federal Reserve. We also performed a search over possible volatility change points between January 1996 - December 2011 based on the Gaussian likelihood criterion function given in expression (14). It is minimized in September 2001. Thus, there is some evidence that there is heteroskedasticity in the model errors with a change in volatility during the recession following the burst of the Dotcom bubble and coinciding with the 9/11 terrorist attacks. All possible volatility breaks are marked with red lines in Figure 3.[7] The resulting p-values of our tests for time-varying impact effects are displayed in the upper panel of Table 10.

Considering any of the volatility change points, our test for time-varying impact effects gives a p-values by far exceeding any conventional significance level. Thus, our test does not find evidence of time-varying impact effects

[6]The data set can be found in the data archive of the *Journal of Applied Econometrics*.

[7]For all three possible break dates, the LM test for heteroskedasticity from Lütkepohl (2005) has p-values very close to zero and strongly supports heteroskedasticity.

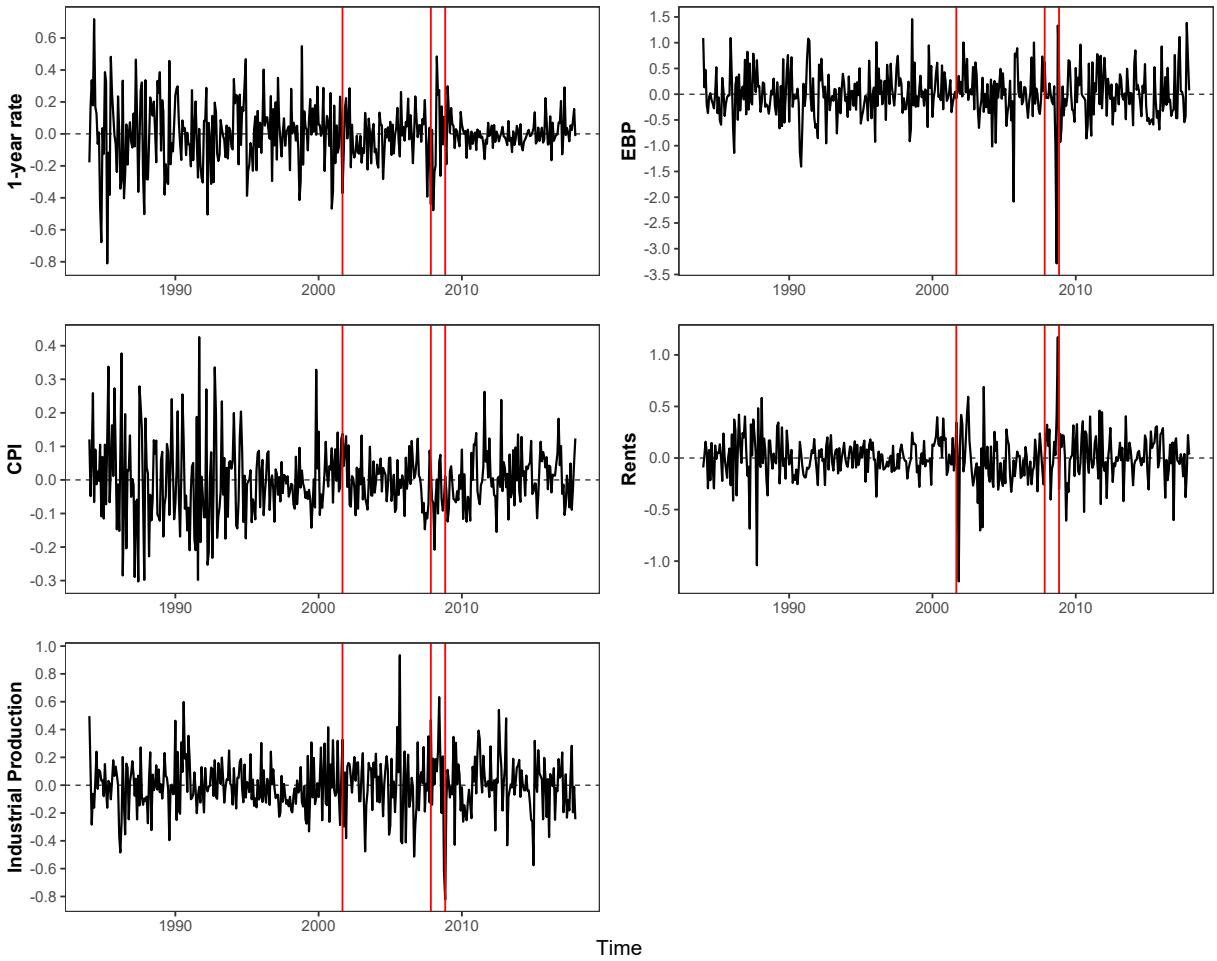

Figure 3: OLS residuals of US housing market VAR(12) model. The solid red lines indicate the three possible variance breaks in September 2001, November 2007, and November 2008 that are investigated.

in this model and sample period. Of course, the model is quite large and, hence, based on our simulation results, one would expect the power of the test to be low. However, in this situation the data give no indication for concern about the time-invariance of the impact effects.

Of course, one may wonder whether our volatility model is suitable for this data. After all, we are assuming only one volatility change point. Thus, we have also considered refining the volatility regimes by combining the breaks in the early 2000s and during the Great Recession. The results of these tests are displayed in the lower panel of Table 10 and do not provide evidence against time-invariant impact effects either. We also performed other robustness checks by considering another volatility model based on a Markov-switching in variances model and did not find evidence against time-invariant impact effects of the monetary policy shocks. Thus, in this case our tests provide the

Table 10: Results (p-values) of Tests for Time-Varying Impact Effects of US Model

Regimes	Change point(s)	Test for time-vary impact effects		
		$\beta(1) = \beta(2)$	$\beta(1) = \beta(3)$	$\beta(2) = \beta(3)$
$M = 2$	2001m9	0.953	–	–
	2007m11	0.888	–	–
	2008m11	0.997	–	–
$M = 3$	2001m9, 2007m11	0.999	0.902	0.995
	2001m9, 2008m11	0.947	0.998	0.982

reassuring result that the implicit assumption of time-varying impact effects of the structural shock of interest appears to be in line with the data.

6 Conclusions

In heteroskedastic proxy structural VAR analyses 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 it is not known precisely when a change in variance occurred. Misspecifying the volatility change point does not invalidate the test but may lead to reduced power. Even if the mechanism that drives the heteroskedasticity is not in line with the model assumed for our test, the test may have some power.

We have applied the test to two example models from the structural VAR literature. In one of the example models, time-invariance of the impact effects is rejected and in the other example model it is not rejected. The examples illustrate 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.

Appendix

A.1 Parametrization of DGP 2

DGP 2 used in our simulations is based on a data set for the crude oil market from Känzig (2019). He considers a 6-dimensional baseline VAR(13) model with a constant term for the variables

$$y_t = (rp_t, prod_t, inv_t, ip_t^{World}, ip_t^{US}, cpi_t^{US})',$$

where rp_t is the real price of oil, $prod_t$ is world oil production, inv_t represents world oil inventories, ip_t^{World} is world industrial production, ip_t^{US} is US industrial production, and cpi_t^{US} is the U.S. consumer price index. The data are monthly for the sample period 1974m1-2017m12 which implies a sample size of 528 observations.[8]

We fit a VAR(1) model to the data and get OLS estimates

$$\nu = (53.578, 132.134, -51.649, -17.081, -4.098, -15.050)',$$

$$A_1 = \begin{bmatrix} 0.980 & -0.066 & 0.028 & 0.142 & 0.013 & -0.125 \\ -0.003 & 0.869 & -0.018 & 0.065 & 0.000 & -0.002 \\ 0.005 & 0.059 & 0.991 & -0.058 & 0.013 & 0.023 \\ 0.000 & 0.017 & 0.002 & 0.983 & 0.006 & 0.003 \\ -0.002 & 0.008 & -0.004 & -0.008 & 0.993 & 0.012 \\ 0.002 & 0.015 & 0.006 & -0.005 & 0.008 & 0.987 \end{bmatrix}$$

which are used as parameters for DGP 2. The eigenvalues of A_1 have moduli

$$0.8681, 0.9734, 0.9859, 0.9859, 0.9984, 0.9915.$$

Thus, the process is close to a unit root process but is formally still stable and very persistent.

We consider two ($M = 2$) or three ($M = 3$) variance regimes. Under $\mathbb{H}_0$, we use the Cholesky decomposition of the estimated reduced-form error covariance matrix,

$$B = \begin{bmatrix} 7.660 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.012 & 1.480 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.120 & -0.018 & 1.033 & 0.000 & 0.000 & 0.000 \\ 0.109 & 0.056 & 0.033 & 0.593 & 0.000 & 0.000 \\ 0.052 & 0.139 & 0.011 & 0.418 & 0.561 & 0.000 \\ 0.113 & 0.017 & -0.002 & 0.005 & -0.006 & 0.233 \end{bmatrix}.$$

[8]The data set was kindly made available to us by Diego Känzig.

Under $\mathbb{H}_1$, the first column of $B(m)$ is replaced by the estimates that result from assuming a volatility break in April 1991 (i.e., after the end of the gulf war) for $M = 2$. For $M = 3$ an additional change in the volatility pattern is assumed to occur in December 2007 (i.e., at the onset of the Great Recession). The resulting estimates are rescaled with the standard deviation obtained under $\mathbb{H}_0$. All other columns of $B(m)$ remain as under $\mathbb{H}_0$. The precise impact effects matrices under $\mathbb{H}_1$ are as follows:

$$B(1) = \begin{bmatrix} 7.660 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.529 & 1.480 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.121 & -0.018 & 1.033 & 0.000 & 0.000 & 0.000 \\ -0.184 & 0.056 & 0.033 & 0.593 & 0.000 & 0.000 \\ -0.189 & 0.139 & 0.011 & 0.418 & 0.561 & 0.000 \\ 0.034 & 0.017 & -0.002 & 0.005 & -0.006 & 0.233 \end{bmatrix},$$

$$B(2) = \begin{bmatrix} 7.660 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.105 & 1.480 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.423 & -0.018 & 1.033 & 0.000 & 0.000 & 0.000 \\ -0.069 & 0.056 & 0.033 & 0.593 & 0.000 & 0.000 \\ 0.037 & 0.139 & 0.011 & 0.418 & 0.561 & 0.000 \\ 0.060 & 0.017 & -0.002 & 0.005 & -0.006 & 0.233 \end{bmatrix},$$

$$B(3) = \begin{bmatrix} 7.660 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ -0.143 & 1.480 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.469 & -0.018 & 1.033 & 0.000 & 0.000 & 0.000 \\ 0.177 & 0.056 & 0.033 & 0.593 & 0.000 & 0.000 \\ -0.162 & 0.139 & 0.011 & 0.418 & 0.561 & 0.000 \\ 0.208 & 0.017 & -0.002 & 0.005 & -0.006 & 0.233 \end{bmatrix}.$$

Thus,

$$\beta(1) = \begin{pmatrix} 0.0691 \\ -0.0158 \\ -0.0240 \\ -0.0247 \\ 0.0044 \end{pmatrix}, \quad \beta(2) = \begin{pmatrix} -0.0137 \\ 0.0552 \\ -0.0090 \\ 0.0048 \\ 0.0078 \end{pmatrix}, \quad \beta(3) = \begin{pmatrix} -0.0187 \\ 0.0612 \\ 0.0231 \\ -0.0211 \\ 0.0272 \end{pmatrix}.$$

Moreover, the asymptotic covariance matrices of the estimated $\beta(m)$ are

$$V(1) = \begin{bmatrix} 0.33 & -0.01 & 0.01 & 0.03 & 0.00 \\ -0.01 & 0.17 & 0.00 & 0.00 & 0.00 \\ 0.01 & 0.00 & 0.06 & 0.04 & -0.00 \\ 0.03 & 0.00 & 0.04 & 0.08 & -0.00 \\ 0.00 & 0.00 & -0.00 & -0.00 & 0.01 \end{bmatrix},$$

$$V(2) = \begin{bmatrix} 1.15 & -0.02 & 0.03 & 0.10 & -0.00 \\ -0.02 & 1.07 & -0.00 & 0.01 & 0.01 \\ 0.03 & -0.00 & 0.72 & 0.50 & -0.00 \\ 0.10 & 0.01 & 0.50 & 1.64 & -0.01 \\ -0.00 & 0.01 & -0.00 & -0.01 & 0.34 \end{bmatrix},$$

$$V(3) = \begin{bmatrix} 0.44 & -0.01 & 0.01 & 0.03 & 0.00 \\ -0.01 & 0.42 & 0.01 & -0.01 & 0.00 \\ 0.01 & 0.01 & 0.21 & 0.15 & -0.00 \\ 0.03 & -0.01 & 0.15 & 0.35 & -0.01 \\ 0.00 & 0.00 & -0.00 & -0.01 & 0.07 \end{bmatrix}.$$

A.2 Additional Impulse Responses

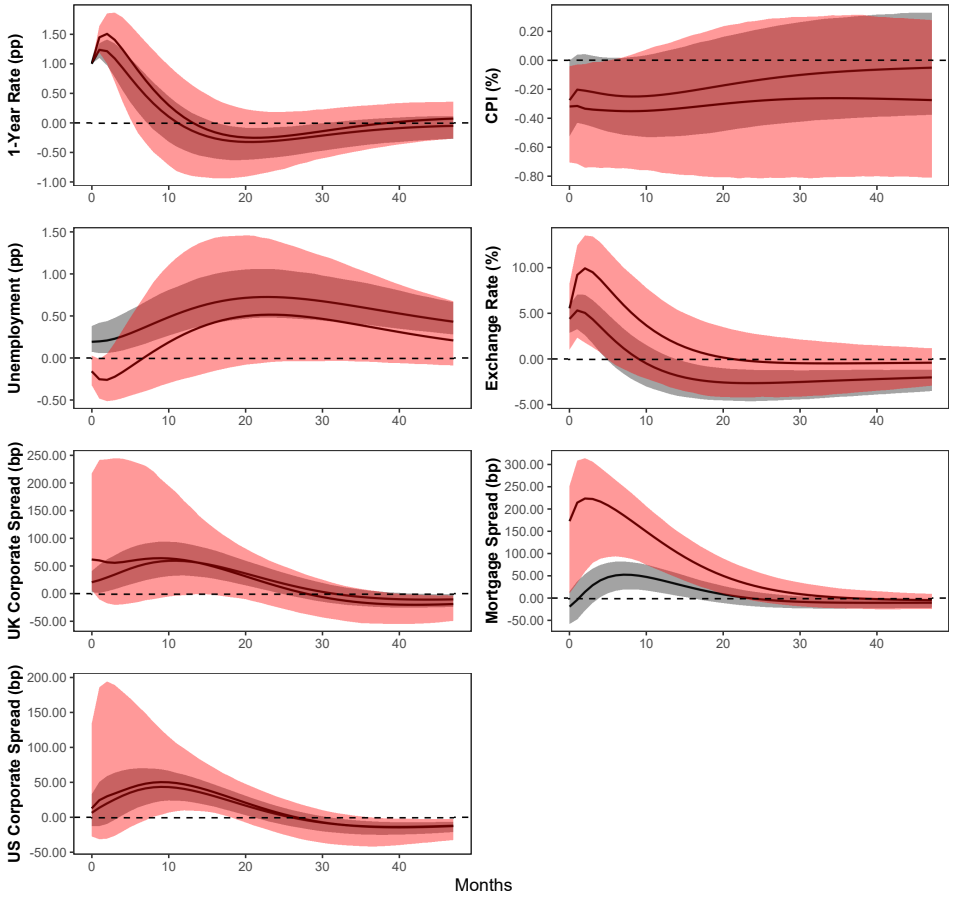

Figure A.1: Responses to a monetary policy shock with pointwise 68% confidence bands based on the moving-block-bootstrap methodology described in Jentsch and Lunsford (2019) using a block length of $l = 4$. Grey areas indicate confidence intervals for the sample prior to the Great Recession from 1992m3-2008m8 and red areas represent confidence intervals for the period during and after the Great Recession from 2008m9-2015m1.

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.

Angelini, G., Bacchiocchi, E., Caggiano, G. and Fanelli, L. (2019). Uncertainty across volatility regimes, *Journal of Applied Econometrics* **34**: 437–455.

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 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 Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**: 1212–1247.

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.

Piffer, M. and Podstawski, M. (2018). Identifying uncertainty shocks using the price of gold, *The Economic Journal* **128**: 3266–3284.

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.

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.