DISCUSSION PAPER SERIES

No. 9892

JOINT CONFIDENCE SETS FOR STRUCTURAL IMPULSE RESPONSES

Atsushi Inoue and Lutz Kilian

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

JOINT CONFIDENCE SETS FOR STRUCTURAL IMPULSE RESPONSES

Atsushi Inoue, Southern Methodist University
Lutz Kilian, University of Michigan and CEPR

ABSTRACT

Joint Confidence Sets for Structural Impulse Responses*

Many users of structural VAR models are primarily interested in learning about the shape of structural impulse response functions. This requires joint inference about sets of structural impulse responses, allowing for dependencies across time as well as across response functions. Such joint inference is complicated by the fact that the joint distribution of structural impulse response becomes degenerate when the number of structural impulse responses of interest exceeds the number of model parameters, as is often the case in applied work. This degeneracy may be overcome by transforming the estimator appropriately. We show that the joint Wald test is invariant to this transformation and converges to a nonstandard distribution, which can be approximated by the bootstrap, allowing the construction of asymptotically valid joint confidence sets for any subset of structural impulse responses, regardless of whether the joint distribution of the structural impulse responses is degenerate or not. We demonstrate by simulation the coverage accuracy of these sets in finite samples under realistic conditions. We make the case for representing these joint confidence sets in the form of "shotgun plots" rather than joint confidence bands for impulse response functions. Several empirical examples demonstrate that this approach not only conveys the same information as confidence bands about the statistical significance of response functions, but provides economically relevant additional information about the shape of response functions that is lost when reducing the joint confidence set to two-dimensional bands.

Atsushi Inoue
Department of Economics
Southern Methodist University
3300 Dyer Street
Dallas, TX 75275-0496
USA

Lutz Kilian
Department of Economics
University of Michigan
238 Lorch Hall
611 Tappan Street
Ann Arbor, MI 48109-1220
USA

Email: ainoue@smu.edu

Email: lkilian@umich.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158326

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=132598

1 Introduction

It is well known that impulse response estimates from structural vector autoregressive (VAR) models tend to be imprecisely estimated, given the short samples typical of applied work. This fact makes it important to assess the reliability of these estimates by constructing confidence sets. It has become standard in the literature to evaluate the statistical significance of the estimated structural impulse response functions using pointwise confidence intervals (see, e.g., Lütkepohl 1990, Kilian 1998). This practice is questionable because pointwise intervals treat structural impulse response estimators as mutually independent, when we know them to be dependent in general both across horizons and across impulse response functions. Confidence bands obtained by connecting pointwise confidence intervals tend to be too narrow and lack coverage accuracy, resulting in spurious findings of statistical significance. This problem has been recognized for a long time, but there is no consensus on how to overcome these distortions.

Analogous problems also arise in Bayesian inference. In related work, Sims and Zha (1999) and Inoue and Kilian (2013) have discussed possible solutions to this problem from a Bayesian point of view. The current paper, in contrast, takes a frequentist perspective. To the extent that the problem of joint impulse response confidence sets has been discussed in the frequentist VAR literature, it has often been reduced to a problem of conducting joint inference across a range of horizons for a given impulse response function. For example, Jordà (2009) proposes one solution to this problem and Lütkepohl et al. (2013) several alternatives. Simulation evidence on the finite-sample accuracy of these confidence bands is discussed in Kilian and Kim (2011) and Lütkepohl et al. (2013).

It is important to stress that these approaches, while representing an important step forward, are too restrictive for applied work. Many users of structural VAR models are interested in conducting inference on multiple impulse response functions at the same time. Sometimes the economic question of interest involves multiple impulse response functions. For example, a macroeconomist may be interested in whether an oil price shock creates

1

stagflation in the domestic economy, which by necessity involves studying the responses of inflation as well as real output. The same would be true if we studied the effect of a U.S. monetary policy shock because the loss function of the Federal Reserve depends on both real output and inflation. It is also common to be interested in testing the implications of economic theory for a range of different impulse response functions simultaneously. For example, Blanchard (1989) uses a macroeconomic VAR model to test the implication of standard Keynesian models that (1) positive demand innovations increase output and decrease unemployment persistently, and (2) that a favorable supply shock triggers an increase in unemployment without a decrease in output. This example involves inference about four impulse response functions simultaneously. There even are cases in which users of structural VAR models are interested in studying the responses of all model variables to all structural shocks simultaneously. A good example is recent structural VAR models of the global market for industrial commodities such as crude oil (e.g., Kilian 2009).

A proper solution to this problem requires taking account of the dependence of all structural impulse responses of interest, not just of the responses in a given impulse response function. This is the objective of the current paper. We propose a novel approach to constructing asymptotically valid joint confidence sets for any subset of the structural impulse responses of interest that is based on inverting a Wald test statistic. One difficulty in this context is that in many situations the asymptotic normality of the joint distribution of the structural impulse responses does not hold even in stationary vector autoregressions. Specifically, when the number of structural impulse responses exceeds the number of model parameters, the joint asymptotic distribution is degenerate, and the distribution of the Wald test statistic is nonstandard. This problem has also been noted in Lütkepohl and Poskitt (1991, p. 493), for example.

This degeneracy may be overcome by transforming the estimator appropriately. We show that the joint Wald test statistic is invariant to this transformation and converges to a nonstandard distribution, which can be approximated by the bootstrap, thus providing a

theoretical justification for the use of the bootstrap Wald test statistic in constructing joint confidence sets for structural impulse responses even in the absence of joint asymptotic normality. This result greatly extends the range of problems that can be addressed with this bootstrap method. A similar econometric problem has been described in a different context by Andrews (1987). Andrews provides a sufficient condition that allows standard inference based on asymptotic χ^2-critical values even when the covariance matrix used in constructing the test statistic is asymptotically singular. This condition does not apply in our setting, however, because the bootstrap covariance matrix is almost always positive definite. Its eigenvalues are positive in finite samples and equal to zero only in the limit. Thus, our approach is of independent interest.

We show that in idealized settings the joint Wald confidence region is expected to be smaller than alterrnative confidence sets such as the Bonferroni set. We also analyze the coverage accuracy of the these confidence sets in finite samples by simulation. Our simulation design focuses on data generating processes with many lags, large roots, and realistic sample sizes. We find that the bootstrap Wald confidence set to be reasonably accurate even in large-dimensional and highly persistent VAR models, while the Bonferroni approach is conservative. The latter result is not unexpected because the number of structural impulse responses of interest in these models is large.

A closely related approach has been proposed independently in Lütkepohl et al. (2014). One difference between our analysis is that we focus on the Wald test statistic for the vector of structural impulse responses, γ, whereas Lütkepohl et al. (2014) utilize a Wald test statistic for the parameters of the VAR model, θ. They then infer the confidence region for γ from the mapping $\gamma = \gamma(\theta)$. We contrast these two approaches and note that the Wald test statistic in Lütkepohl et al. (2014) has certain theoretical advantages in constructing joint confidence regions for impulse responses compared with our approach. Our simulation study, however, suggests that the differences in coverage accuracy tend to be small. In the few cases in which there is a larger difference in coverage accuracy, the

3

Wald test statistic for γ yields more accurate confidence regions.

The second difference between our analysis and Lütkepohl et al. (2014) is that we propose to plot the sets of impulse responses associated with each bootstrap draw that is contained in the joint Wald confidence set, resulting in plots for each impulse response function that resemble a shotgun trajectory chart ("shotgun plot"). In contrast, Lütke- pohl et al. (2014) connect the pointwise maxima of the impulse responses in the joint set to form the upper bound of a confidence band and similarly connect the pointwise minima of the impulse responses in the joint set to form the lower bound of a confidence band. This approach results in a loss of information compared with our shotgun plots because one is no longer able discern the evolution over time of the impulse response functions as- sociated with any one structural model estimate in the joint confidence set. It is precisely this evolution of the response function that users of structural VAR models typically are interested in (see, e.g., Cochrane 1994). For example, many macroeconomists have aban- doned frictionless neoclassical models and adopted models with nominal or real rigidities based on VAR evidence of sluggish or delayed responses of inflation and output (see, e.g., Woodford 2003). This is also true in other applications. Whereas macroeconomists may be interested in whether a response function for real output is hump-shaped or not, for example, users of structural VAR models in international economics may be interested in whether there is delayed overshooting in the response of the exchange rate to monetary policy shocks. It is difficult to answer such questions about the shape of a given impulse response function based on two-dimensional joint confidence bands in general because such bands are consistent with a multitude of different response patterns.

These difficulties are compounded when considering the analysis of more than one impulse response function at a time, as is common in applied work. For example, one may interested in whether an unanticipated increase in the real price of oil is stagflationary. A stagflationary responses means that the response of real output to this shock is negative at all horizons, while the response of the price level in the same model is positive at all

horizons (see Kilian 2008). A confidence band of the type proposed in Lütkepohl et al. (2014) cannot answer this question in general. To illustrate this point, consider two model realizations in the joint Wald confidence band, one of which implies a positive response of both real output and the price level, while the other model implies a negative response of both real output and the price level such that neither model implies a stagflationary response. The resulting confidence band will nevertheless be observationally equivalent to one in which some models have stagflationary responses in that the confidence bands cover both negative and positive responses of both real output and the price level.

We illustrate these points based on two empirical examples. Our first empirical example is a semi-structural model of U.S. monetary policy; the second example is a semi-structural model of the response of the U.S. economy to oil price shocks. We use these examples to illustrate that in some situations, the use of shotgun plots and of joint confidence bands will yield the same answer by construction. For example, if we are interested only in one impulse response function at a time, the lower bound of the joint confidence band will include zero at some horizon, if and only if some of the impulse response functions in the shotgun plot crosses zero. In other situations, the shotgun plots implied by joint Wald confidence sets may reveal features of the data that are obscured by more traditional point-wise confidence intervals or by two-dimensional joint confidence bands. For example, the shotgun plot provides strong evidence against the hypothesis of stagflationary responses to oil price shocks that cannot be detected based on joint confidence bands. Likewise, the shotgun plot provides evidence of hump-shaped responses of real output to monetary policy shocks which is not conveyed by joint confidence bands.

The remainder of the paper is organized as follows. In section 2, we define basic notation. The asymptotic results are developed in section 3. We also provide a simple AR(1) example to illustrate the difference between our problem and the related testing problem analyzed in Andrews (1987). Section 4 reviews the practical implementation of the Wald and Bonferroni approach to constructing joint confidence sets for structural

impulse responses. In section 5, we compare the power of alternative test statistics used in constructing joint confidence regions, and in section 6 we contrast the Wald test statistics proposed in Lütkepohl et al. (2014) and in the current paper. The simulation results are summarized in section 7 and the empirical examples are discussed in section 8. We conclude in section 9. Technical arguments are relegated to the appendix.

2 Model

Consider an n-variate structural VAR of order p:

$$A_0 y_t \; = \; a + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \tag{1}$$

with reduced-form representation:

$$y_t \; = \; c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t, \tag{2}$$

where a and c are $n \times 1$ vectors of intercepts, A_j and Φ_j are $n \times n$ slope parameter matrices for $j = 0, 1, ..., p$, with $\Phi_0 = I_n$, and the $n \times 1$ vectors of disturbance terms satisfy $\varepsilon_t \; \sim \; WN(0_{n \times 1}, I_n)$ and hence $u_t \; \sim \; WN(0_{n \times 1}, V)$. V is an $n \times n$ positive definite matrix. The structural impulse responses may be identified from model (2), for example, by imposing short-run exclusion restrictions directly on A_0, so that there is a unique nonsingular matrix A_0 that satisfies $A_0 A_0' = V$. Then the jth step ahead structural impulse response of (1) can be written as $\Psi_j A_0^{-1}$, where Ψ_j is the jth step ahead reduced-form impulse response matrix of (2).

Let γ denote a $q \times 1$ vector that is obtained from stacking the elements of structural impulse responses of interest. For example, γ could contain all impulse response functions stacked into one vector or it could contain any subset of the structural impulse responses of interest. Let $\widehat{\gamma}_T$ denote the estimator of γ. Note that $\widehat{\gamma}_T$ can be written as a nonlinear

function of sample averages of the data, i.e., $\hat{\gamma}_T = \gamma(\bar{X}_T)$, where $\bar{X}_T$ is a $k \times 1$ vector consisting of the first and second moments of $y_t, y_{t-1}, ..., y_{t-p}$.

Although we choose to present our results in the context of stationary VAR models identified based on short-run exclusion restrictions, our framework could be adapted to allow for the imposition of exclusion restrictions on long-run impulse responses in suitably transformed VAR models containing some I(1) variables. In contrast, our analysis does not allow for set-identified structural impulse responses.

3   Asymptotic Theory

Our objective is to find an asymptotic approximation to the distribution of the Wald statistic of the vector-valued structural impulse response estimator under the null hypothesis:

$$H_0 : \gamma(\mu) \ = \ \gamma_0,$$

where $\mu = E(\bar{X}_T)$. Our proposal is to invert the Wald statistic to form an asymptotically valid joint confidence set for this null hypothesis. We begin with some conditions required for characterizing the joint asymptotic distribution of the elements of $\gamma(\bar{X}_T)$.

Assumptions.

(a) $Z_T \equiv \Omega^{-\frac{1}{2}}\sqrt{T}(\bar{X}_T - \mu) \ \overset{d}{\to} \ Z \equiv N(0, I_k)$ and $Z_T^* \equiv \Omega^{-\frac{1}{2}}\sqrt{T}(\bar{X}_T^* - \bar{X}_T) \ \overset{d}{\to} \ Z^* \equiv$ $N(0, I_k)$ where Ω is a positive definite matrix and the convergence of Z_T^* is with respect to the bootstrap probability measure conditional on the data.

(b) There are conformable matrices $B_0,...,B_h$ and $\hat{B}_0,...,\hat{B}_h$ such that

$$
\begin{aligned}
T^{\frac{1}{2}}(\gamma(\bar{X}_T) - \gamma(\mu)) &= B_0 Z_T + T^{-\frac{1}{2}} B_1 (Z_T \otimes Z_T) + \dots \\
&\quad + T^{-\frac{h}{2}} B_h (Z_T \otimes \dots \otimes Z_T) + o_p(T^{-\frac{h}{2}}), \quad (3) \\
T^{\frac{1}{2}}(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T)) &= \hat{B}_0 Z_T^* + T^{-\frac{1}{2}} \hat{B}_1 (Z_T^* \otimes Z_T^*) + \dots \\
&\quad + T^{-\frac{h}{2}} \hat{B}_h (Z_T^* \otimes \dots \otimes Z_T^*) + o_p^*(T^{-\frac{h}{2}}), \quad (4)
\end{aligned}
$$

where $\hat{B}_j = B_j + o_p(1)$ for $j = 0, 1, ..., h$ and $o_p^*(T^{-\frac{h}{2}})$ is with respect to the bootstrap probability measure conditional on the data with probability one.

Remarks.

1. Assumption (a) holds when applying the residual-based bootstrap to stationary vector autoregressive processes. It also covers vector error correction models with known cointegrating vectors and VAR models in which integrated variables have been differenced to achieve stationarity. Conditional heteroskedasticity can be accommodated by the use of the wild bootstrap (see Gonçalves and Kilian 2004).

2. Assumption (b) follows from a Taylor series expansion of the left-hand side of equations (3) and (4). The delta method is based on the first-order term of the stochastic expansion on the right-hand side (see Lütkepohl, 1990, for the application of the delta method to structural impulse responses). The higher-order stochastic terms on the right-hand side have also been used to develop Edgeworth expansions of the distribution of estimators (see Hall, 1992). Assumption (b) holds, for example, for stationary vector autoregressive processes with positive definite error covariance matrices and short-run exclusion restrictions. For more primitive assumptions for the existence of higher-order asymptotic expansions of the distribution of estimators in stationary time series models see Bao and Ullah (2007) and Bao (2007).

Theorem 1 (Joint Asymptotic Distribution of Structural Impulse Responses).
Suppose that Assumptions (a) and (b) hold. Then there are $0 \leq r \leq h$ and $q \times q$ nonsingular

matrix S such that

$$\Upsilon_T S'(\gamma(\bar{X}_T) - \gamma(\mu)) \xrightarrow{d} \xi \equiv \begin{bmatrix} S_0' B_0 Z \\ S_1' B_1 (Z \otimes Z) \\ \vdots \\ S_r' B_r (Z \otimes \cdots \otimes Z) \end{bmatrix}, \tag{5}$$

$$\Upsilon_T S'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T)) \xrightarrow{d} \xi^* \equiv \begin{bmatrix} S_0' B_0 Z^* \\ S_1' B_1 (Z^* \otimes Z^*) \\ \vdots \\ S_r' B_r (Z^* \otimes \cdots \otimes Z^*) \end{bmatrix}, \tag{6}$$

where $S = [S_0 \ S_1 \ \cdots \ S_r]$, S_j is $q \times q_j$ with $q_0 = k$, $q_j \geq 0$ and $\sum_{j=0}^r q_j = q$, and

$$\Upsilon_T = \begin{bmatrix} T^{\frac{1}{2}} I_{q_0} & 0_{q_0 \times q_1} & \cdots & 0_{q_0 \times q_r} \\ 0_{q_1 \times q_0} & T I_{q_1} & \cdots & 0_{q_1 \times q_r} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{q_r \times q_0} & 0_{q_r \times q_1} & \cdots & T^{\frac{r+1}{2}} I_{q_r} \end{bmatrix}.$$

Remarks.

1. We allow for $k = q$ and $k < q$. When B_0 is of full rank such that $k = q$, the delta method can be applied to the entire vector $\hat{\gamma}_T$ so that . This is the case that Lütkepohl et al. (2014) appeal to. When B_0 is not of full rank, which is the case of primary interest in the current paper, we have $k < q$ and the delta method fails. The distribution of the impulse responses in the AR(1) model with zero slope analyzed in Benkwitz, Lütkepohl and Neumann (2000) is a special case of $k < q$.

2. Theorem 1 shows that the estimator of structural impulse responses can be rotated by some matrix S such that some elements may converge at a faster rate than others if $k < q$. Such coordinate rotations have been employed in various contexts in the literature (see, e.g., Phillips 1989; Sims, Stock and Watson 1990; Antoine and Renault 2012).

Assumptions.

(c)

$$\Upsilon_T S'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T))(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T))'S\Upsilon_T$$

is uniformly integrable.

(d) $J = E(\xi\xi')$ is nonsingular.

Remarks.

1. Assumption (c) guarantees that the bootstrap method can be used to estimate the limiting covariance matrix.

2. Assumption (d) requires that the scaled rotated vector of impulse responses in Theorem 1 has a nonsingular asymptotic covariance matrix.

Theorem 2 (Asymptotic Null Distribution of the Wald Statistic)

Suppose that Assumptions (a), (b), (c) and (d) hold. Then under the null hypothesis

$$W_T^{\gamma} \quad \xrightarrow{d} \quad \xi' J^{-1}\xi, \tag{7}$$

$$W_T^{\gamma*} \quad \xrightarrow{d} \quad \xi^{*\prime} J^{-1}\xi^*, \tag{8}$$

where the convergence of $W_T^{\gamma*}$ is with respect to the bootstrap probability measure conditional on the data with probability one.

Remarks.

Theorem 2 shows that the Wald statistic has a nonstandard asymptotic null distribution when $k < q$. This situation occurs when the number of structural impulse responses exceeds the number of slope parameters in the fitted VAR model. Theorem 2 shows that this null distribution may be approximated by bootstrap methods because Z and Z^* both are vectors of standard normal random variables, so the limiting distribution of $W_T^{\gamma*}$ is the same as that of W_T^{γ}. The bootstrap method is generally applicable in that it also recovers the correct asymptotic null distribution of the Wald statistic in the case of $k = q$. In the

latter case, $W_T^{\gamma} \xrightarrow{d} \chi_k^2$ and $W_T^{\gamma*} \xrightarrow{d} \chi_k^2$.

Theorem 2 has the following corollary:

Corollary (Asymptotic Validity of Joint Confidence Sets)

Under the stated assumptions and the null hypothesis,

$$\lim_{T \to} P^*(W_T^{\gamma} < c_{1-\alpha}^*) = 1 - \alpha, \tag{9}$$

where P^* is the bootstrap probability measure conditional on the data and $c_{1-\alpha}^*$ denotes the $100(1 - \alpha)\%$ bootstrap critical value for the Wald statistic obtained using the same bootstrap methods conventionally used for inference about t-statistics.

The effective coverage accuracy of the nominal $100(1 - \alpha)\%$ Wald confidence set can be evaluated by evaluating in repeated samples with what probability W_T^{γ} is contained in the confidence region implied by the $c_{1-\alpha}^*$ critical values.

The next section illustrates the results in Theorems 1 and 2 in the context of a stylized example in which B_0 is less than full rank such that the conventional asymptotic approximation fails. We also demonstrate how our approach differs from that in Andrews (1987) who considered a related problem in which singular covariance matrices are used in Wald test statistics. We show that his results do not apply in our setting because the bootstrap covariance matrix is nonsingular with probability one, even though its probability limit is singular.

3.1   An Illustrative Example

Suppose that

$$y_t = \rho y_{t-1} + u_t, \tag{10}$$

where $|\rho| < 1$ and $u_t \overset{iid}{\sim} (0, \sigma^2)$. Assumption (a) is satisfied for strictly stationary autoregressive processes with iid disturbance terms with finite variance (e.g., Freedman

1984). Consider jointly testing the null hypothesis that $\rho = \rho_0$ and $\rho^2 = \rho_0^2$, where ρ and ρ^2 correspond to the responses of y_{t+h} to a unit shock in u_t for $h = 1, 2$. Because

$$
\begin{aligned}
\widehat{\rho}_T - \rho &= \frac{\sum_{t=2}^{T} y_{t-1} u_t}{\sum_{t=2}^{T} y_{t-1}^2} = \frac{1}{E(y_{t-1}^2)} \frac{1}{T} \sum_{t=2}^{T} y_{t-1} u_t + o_p\left(\frac{1}{T}\right) \\
&= \frac{1-\rho^2}{\sigma^2} \frac{1}{T} \sum_{t=2}^{T} y_{t-1} u_t + o_p\left(\frac{1}{T}\right) \\
&= \left(\frac{1-\rho^2}{\sigma^2}\right)^{\frac{1}{2}} T^{-\frac{1}{2}} Z_T, + o_p\left(\frac{1}{T}\right),
\end{aligned}
\tag{11}
$$

and

$$
\widehat{\rho}_T^2 - \rho^2 = 2\rho(\widehat{\rho}_T - \rho) + (\widehat{\rho}_T - \rho)^2,
\tag{12}
$$

where $Z_T = ((1-\rho^2)/\sigma^2)^{\frac{1}{2}} T^{-\frac{1}{2}} \sum_{t=2}^{T} y_{t-1} u_t$, we can write

$$
\begin{aligned}
\begin{bmatrix} \widehat{\rho}_T - \rho \\ \widehat{\rho}_T^2 - \rho^2 \end{bmatrix} &= \begin{bmatrix} \frac{1}{E(y_{t-1}^2)} \\ \frac{2\rho}{E(y_{t-1}^2)} \end{bmatrix} \frac{1}{T} \sum_{t=2}^{T} y_{t-1} u_t + \begin{bmatrix} 0 \\ \frac{1}{[E(y_{t-1}^2)]^2} \end{bmatrix} \left(\frac{1}{T} \sum_{t=2}^{T} y_{t-1} u_t\right)^2 + \begin{bmatrix} o_p(T^{-1}) \\ o_p(T^{-1}) \end{bmatrix} \\
&= \begin{bmatrix} \frac{(1-\rho^2)^{\frac{1}{2}}}{\sigma} \\ \frac{2\rho(1-\rho^2)^{\frac{1}{2}}}{\sigma} \end{bmatrix} T^{-\frac{1}{2}} Z_T + \begin{bmatrix} 0 \\ \frac{(1-\rho^2)}{\sigma^2} \end{bmatrix} T^{-1} Z_T^2 + \begin{bmatrix} o_p(T^{-1}) \\ o_p(T^{-1}) \end{bmatrix}.
\end{aligned}
\tag{13}
$$

Thus Assumption (b) is satisfied. By the Schur decomposition theorem, we have

$$
S' B_0 B_0' S = \Lambda,
\tag{14}
$$

where

$$
S = \frac{1}{(4\rho^2 + 1)^{\frac{1}{2}}} \begin{bmatrix} 1 & 2\rho \\ 2\rho & -1 \end{bmatrix}, \quad B_0 = \begin{bmatrix} \frac{(1-\rho^2)^{\frac{1}{2}}}{\sigma} \\ \frac{2\rho(1-\rho^2)^{\frac{1}{2}}}{\sigma} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \frac{(1-\rho^2)(4\rho^2+1)}{\sigma^2} & 0 \\ 0 & 0 \end{bmatrix}.
$$

Thus it can be shown that

$$\Upsilon_T S' \begin{bmatrix} \widehat{\rho}_T - \rho \\ \widehat{\rho}_T^2 - \rho^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{[(1-\rho^2)(4\rho^2+1)]^{\frac{1}{2}}}{\sigma} \\ 0 \end{bmatrix} Z_T + \begin{bmatrix} \frac{2\rho(1-\rho^2)}{\sigma^2(4\rho^2+1)^{\frac{1}{2}}}T^{-\frac{1}{2}} \\ -\frac{1-\rho^2}{\sigma^2(4\rho^2+1)^{\frac{1}{2}}} \end{bmatrix} Z_T^2 + o_p(1), \qquad (15)$$

where

$$\Upsilon_T = \begin{bmatrix} T^{\frac{1}{2}} & 0 \\ 0 & T \end{bmatrix}.$$

Note that the first element of (15) converges in distribution to a normal random variable, whereas the second element converges in distribution to a chi-square random variable up to scale, which is exactly the implication of Theorem 1. The limiting covariance matrix of (15) takes the form of

$$\begin{bmatrix} \frac{(1-\rho^2)(4\rho^2+1)}{\sigma^2} & 0 \\ 0 & \frac{3(1-\rho^2)^2}{\sigma^4(4\rho^2+1)} \end{bmatrix}, \qquad (16)$$

which the bootstrap covariance matrix approximates under the assumption of uniform integrability. Thus, the asymptotic null distribution of the Wald statistic is

$$Z^2 + \frac{1}{3}Z^4, \qquad (17)$$

where Z is the asymptotic normal distribution of Z_T. The latter result is an implication of Theorem 2. While the Wald test statistic is pivotal, it is not chi-square, so Andrews' (1987) solution does not apply in our context.

4 Implementation of the Proposed Methods

Before proceeding to simulation evidence and empirical examples, it is useful to review the implementation of the Wald approach in practice. We also review the application of the Bonferroni approach as the leading alternative to the Wald approach. There are a number

of potential refinements of the original Bonferroni approach designed to reduce the width of the joint confidence bands. We ignore these refinements given evidence in Lütkepohl et al. (2014) that for processes with degrees of persistence similar to our simulation designs and empirical examples these refinements undermine coverage accuracy and hence cannot be recommended.

4.1 Joint Wald Confidence Sets

Upon stacking the estimates of the structural impulse responses of interest in a $q \times 1$ vector $\widehat{\gamma}_T$, we bootstrap the structural impulse responses and denote the bootstrap estimates by $\widehat{\gamma}_T^{*(j)}$ for $j = 1, 2, ..., B$, where B is the number of bootstraps. In practice, we rely on the standard residual-based bootstrap method for parametric models with iid errors. For a review of this and alternative bootstrap methods for vector autoregressions see Berkowitz and Kilian (2000). If we are testing a specific null hypothesis, $\gamma = \gamma_0$, we form the Wald test statistic by

$$W_T^{\gamma} \;=\; (\widehat{\gamma}_T - \gamma_0)' \widehat{\Sigma}_T^{*-1} (\widehat{\gamma}_T - \gamma_0), \tag{18}$$

where

$$\widehat{\Sigma}_T^* = (1/B) \sum_{j=1}^{B} (\widehat{\gamma}_T^* - \widehat{\gamma}_T)(\widehat{\gamma}_T^* - \widehat{\gamma}_T)'. \tag{19}$$

To test a given null hypothesis, the value of the Wald statistic would be compared to the $100(1 - \alpha)$ percentile of the empirical distribution of the bootstrap Wald test statistics

$$W_T^{\gamma*(j)} \;=\; (\widehat{\gamma}_T^{*(j)} - \widehat{\gamma}_T)' \widehat{\Sigma}_T^{**-1} (\widehat{\gamma}_T^{*(j)} - \widehat{\gamma}_T), \tag{20}$$

for $j = 1, 2, ..., B$. Generating the bootstrap critical values requires a nested bootstrap loop, because for each bootstrap realization of the Wald statistic the term $\widehat{\Sigma}_T^*$ in turn must be evaluated by bootstrap simulation.

In the absence of a specific null value, γ_0, joint confidence sets may be constructed

by comparing the value of each of the bootstrap Wald test statistics (20) against the $100(1 - \alpha)$ percentile of the bootstrap distribution. It the bootstrap Wald test statistic is less than this critical value, this bootstrap draw, $\widehat{\gamma}_T^{*(j)}$ is retained and becomes a member of the joint confidence set. This allows us to characterize the confidence region with increasing accuracy, as $B \to \infty$. In our empirical work, we rely on $2,000 \times 2,000$ bootstrap replications.

An important question is how to represent the members of that set of estimates. Our approach differs from that in Lütkepohl et al. (2014) who propose to construct a joint confidence band by constructing an envelope around the shotgun plot. The upper bound of this band is obtained by connecting the largest realization for each element of $\widehat{\gamma}_T^{*(j)}$ across j to form a line; the lower bound is obtained by connecting the smallest realization for each element of $\widehat{\gamma}_T^{*(j)}$ across j. We instead follow Inoue and Kilian (2013) in representing the impulse response draws by plotting all sets of impulse responses associated with the models contained in the joint Wald confidence set. As a result, each impulse response function viewed in isolation looks like a shotgun trajectory chart. Unlike conventional representations of confidence sets, this "shotgun plot" may be frayed around the edges.

The fact that we do not reduce the information to a two-dimensional confidence band is not a drawback of our method in that the shotgun plot conveys the same information as confidence bands would, but it also conveys additional information. In fact, when structural VAR users are interested in the shape of the impulse response functions of interest, for example, or in the relationship of different impulse response functions, only displaying a confidence band may involve an important loss of information, as we will illustrate in section 8.

4.2   Bonferroni Confidence Sets

A well-known alternative for constructing joint confidence regions for γ are Bonferroni bounds. In implementing the Bonferroni method we follow Lütkepohl et al. (2013, 2014)

in forming q t-test statistics for testing $\gamma_i = \gamma_{0,i}$ for $i = 1, 2, ..., q$:

$$t_i = \frac{\widehat{\gamma}_{T,i} - \gamma_{0,i}}{\sqrt{\widehat{\Sigma}^*_{T,ii}}}, \tag{21}$$

where $\widehat{\gamma}_{T,i}$, $\gamma_{0,i}$ and $\widehat{\Sigma}^*_{T,ii}$ denote the ith element of $\widehat{\gamma}_T$, the ith element of γ_0 and the (i,i)th element of $\widehat{\Sigma}^*_T$, respectively. Each of the q t test statistics is compared to the $100(1 - \alpha/q)$ percentile of the empirical distribution of the corresponding bootstrap t test statistics:

$$t_i^{*(j)} = \frac{\widehat{\gamma}^{*(j)}_{T,i} - \widehat{\gamma}_{T,i}}{\sqrt{\widehat{\Sigma}^{**}_{T,ii}}}, \tag{22}$$

for $i = 1, ..., q$ and $j = 1, ..., B$. If any of the q t tests rejects its null hypothesis, the null hypothesis, $\gamma = \gamma_0$ is rejected.

In constructing the joint set, the Bonferroni method only utilizes the marginal distributions of the structural impulse responses, which are asymptotically normal under our assumptions; hence, this method is unaffected by any degeneracy of the joint distribution. The Bonferroni bounds will remain asymptotically valid in the sense of providing a bound on the joint confidence set with at least $100(1 - \alpha)\%$ coverage.

5   Power Considerations

In related work, Lütkepohl et al. (2014) evaluate the power of the Wald test statistic $W_T^{\gamma(\theta)}$ based on the average width of the two-dimensional joint confidence bands implied by the joint confidence set $\left\{\gamma(\theta)|\theta \in W^\theta_{1-\alpha}\right\}$, where $W^\theta_{1-\alpha}$ denotes the $100(1 - \alpha)\%$ joint confidence region for the VAR model parameters θ. This approach results in conservative bands in that the coverage probability of the Wald confidence band is at least $1 - \alpha$, as is the coverage probability of the corresponding Bonferroni confidence band. Lütkepohl et al. (2014) demonstrate by example that it is possible for a Bonferroni confidence band for impulse responses to have lower average interval width than a Wald confidence

band, conditional on both bands having coverage accuracy of at least $1 - \alpha$ (although not necessarily the same coverage accuracy).

To further illustrate this point, they consider a stylized numerical example involving an M-dimensional Gaussian vector $\widehat{\theta}$ with covariance matrix I_M, for which one can determine a precise confidence box with exact confidence level $1 - \alpha$, obtained by choosing individual confidence intervals with confidence level $1 - \alpha_{ind}$ where $\alpha_{ind} = 1 - (1 - \alpha)^{1/M}$. This precise confidence box has width $2 \times z_{1 - \alpha_{ind}/2}$, where z refers to a quantile of the $N(0, 1)$ distribution, while the width of the Bonferroni box is $2 \times z_{1 - \alpha/(2M)}$ and the width of the $W_{1-\alpha}^{\theta, band}$ box is $2 \times \sqrt{\chi^2(M)_{1-\alpha}}$. Lütkepohl et al. (2014) illustrate by simulation that under these idealized conditions the Bonferroni box comes close to the width of the precise confidence box, while the Wald joint confidence box becomes increasingly larger, as the dimension of θ increases. On the basis of this result as well as other simulation evidence, they make the case that the Bonferroni method deserves serious consideration for applied work.

Focusing on the average width of the bands is appropriate, if one restricts the analysis to joint confidence bands. If one focuses on the performance of the M-dimensional confidence region implied by the Wald test statistic instead, as we do in our analysis, the relevant metric is the volume of the M-dimensional confidence region instead. This volume may be computed in the example of Lütkepohl et al. (2014) by raising the width of the precise confidence box and of the Bonferroni confidence box to the power of M. The corresponding volume of the Wald confidence region under the assumptions above is $\pi^{M/2} \chi^2(M)_{1-\alpha}^{M/2} / \Gamma(M/2 + 1)$. This result is based on the volume of a ball of radius r in the M-dimensional Euclidian space, which is $\pi^{M/2} r^M / \Gamma(M/2 + 1)$, where $r = \chi^2(M)_{1-\alpha}^{1/2}$ in our case.

Table 1 shows that the Wald confidence box, as measured by the volume of the implied joint confidence region, is much larger than the precise confidence box. Even for $M = 2$, there is a 22% increase in volume. For $M = 10$, the increase in volume reaches a factor

of 87. This result indicates a substantial loss in power resulting from the use of the projection method in constructing the joint confidence band. In contrast, the volume of the joint confidence region implied by the Bonferroni confidence box is only 7% higher than the volume of the precise confidence box for $M = 10$. Thus, based on its volume, the Bonferroni confidence box is clearly preferred. While this result is qualitatively similar to that in Lütkepohl et al. (2014), Table 1 also shows that the volume of the joint Wald confidence set is much smaller than that of the precise confidence boxes. Its volume is lower for every choice of M. For $M = 2$, the reduction in volume amounts to 5%. For $M = 10$, the reduction in volume increases to 80%.

Table 1 shows that it makes a difference whether one evaluates the volume in M dimensions or in two dimensions. On the basis of the volume of the joint confidence set, we conclude that the joint Wald confidence set is preferred to the Bonferroni bounds. This example is necessarily stylized, and there is no guarantee that these results will carry over to the more complicated setting of mutually correlated estimators of structural impulse responses. Nevertheless, the analysis suggests that the joint Wald confidence set will not only have accurate coverage $1 - \alpha$ asymptotically, but that it may have much smaller volume than the alternative joint confidence regions implied by the Bonferroni set or by joint Wald confidence bands.

6 Alternative Wald Confidence Regions

Extending this analysis to vectors of structural impulse responses further complicates the analysis. In section 3, we proposed a joint confidence region for γ obtained by inverting the Wald test statistic W_T^γ to obtain the set $\left\{\gamma | \gamma \in W_{1-\alpha}^\gamma\right\}$. An alternative approach would have been to construct a joint confidence region by inverting the Wald test statistic $W_T^{\gamma(\theta)}$, on the basis of which Lütkepohl et al. (2014) construct their confidence bands $W_{1-\alpha}^{\gamma(\theta),band}$, to obtain the joint confidence set $\left\{\gamma(\theta) | \theta \in W_{1-\alpha}^\theta\right\}$, where θ denotes the parameters of the vector autoregression and where we made use of the fact that inverting $W_T^{\gamma(\theta)}$ is equivalent

to inverting

$$W_T^\theta \;=\; (\widehat{\theta}_T - \theta_0)'\widehat{\Omega}_T^{*-1}(\widehat{\theta}_T - \theta_0),$$

where $\widehat{\Omega}_T^* = (1/B)\sum_{j=1}^{B}(\widehat{\theta}_T^* - \widehat{\theta}_T)(\widehat{\theta}_T^* - \widehat{\theta}_T)'$ and $\gamma(\theta_0) = \gamma_0$.

Although Lütkepohl et al. (2014) do not consider this approach, it has several merits in the context of our analysis. First, it avoids the singularity of the joint limiting distribution of the impulse responses, when M exceeds the number of model parameters. This point is of no consequence for the first-order asymptotic validity of these methods, given that both W_T^γ and $W_T^{\gamma(\theta)}$ must be bootstrapped in practice and given that the potential degeneracy of the joint distribution can be dealt with by bootstrapping W_T^γ, as we have shown. It matters, however, for two other reasons.

First, because γ has a higher dimension than θ, one would expect the Wald test statistic W_T^γ to be less powerful than $W_T^{\gamma(\theta)}$. This means that the implied confidence regions for $W_{1-\alpha}^{\gamma(\theta)}$ should have even smaller volume than $W_{1-\alpha}^\gamma$. Second, while bootstrapping either W_T^γ or $W_T^{\gamma(\theta)}$ results in first-order asymptotically valid confidence sets, the coverage accuracy of $W_{1-\alpha}^{\gamma(\theta)}$ is determined by the coverage accuracy of $W_{1-\alpha}^\theta$, which is asymptotically pivotal, whereas $W_{1-\alpha}^\gamma$ is not. This suggests that bootstrapping $W_T^{\gamma(\theta)}$ should yield asymptotic refinements and may improve the finite-sample coverage accuracy of the joint confidence set (see Hall 1992). How relevant such asymptotic refinements are in practice is a different matter (see Kilian 1999). We therefore include $\left\{\gamma(\theta)|\theta \in W_{1-\alpha}^\theta\right\}$ as an additional competitor in our simulation study below and compare its coverage accuracy to that of $\left\{\gamma|\gamma \in W_{1-\alpha}^\gamma\right\}$ as well as the Bonferroni set. We do not compare the volume of these confidence regions, because it is not clear how to construct measures of the volume of these confidence regions in general, but the example in section 5 suggests that both Wald confidence regions are likely to have lower volume than any of the alternatives.

7 Simulation Study

We assess the coverage accuracy of the proposed joint Wald confidence set based on two data generating processes that are representative for the types of models of interest in applied work, but are not so large as to make a simulation study computationally prohibitive. We also compare the coverage accuracy of the joint Wald confidence set to that of the Bonferroni set. The coverage accuracy of the Wald confidence set is evaluated by verifying in repeated trials how often W_T^γ and W_T^θ, respectively, are outside the rejection region constructed by bootstrap methods. The coverage accuracy of the Bonferroni set is evaluated by verifying in repeated trials whether the bootstrap Bonferroni confidence band for γ_k includes all of the population impulse responses of interest. Both methods are implemented based on a nested bootstrap loop with $1,000 \times 1,000$ replications for each Monte Carlo trial. The coverage results are based on 500 Monte Carlo trials, given the high computational cost of the exercise.

7.1 DGP 1: Monthly VAR Model

The first data generating process (DGP) is a recursively identified three-variable VAR(p) model of the global market for crude oil proposed by Kilian (2009). The model includes the growth rate of global crude oil production, a measure of global real economic activity (expressed as a business cycle index), and the real price of oil. Notwithstanding the model's recursive structure, all three shocks in this model can be given a structural economic interpretation. There is a flow supply shock, a flow demand shock and an oil-market specific demand shock, designed to capture, for example, precautionary demand shocks. As in Kilian (2009), we are interested in the responses of all three model variables to each of these shocks at horizons $h = 0, ..., 17$.

The DGP is based on estimates of this VAR model on the original data set used in Kilian (2009) with lag order $p \in \{6, 12, 24\}$, reflecting common choices of the lag order for monthly data. All models include an intercept. The estimation period is 1973.2-2007.12.

The sample size in the simulation study matches that in the original data to make the analysis as realistic as possible. For details of the data construction the reader is referred to Kilian (2009). The lag order is treated as known in the simulations, as an increasing number of empirical studies simply postulates a lag order rather than relying on lag order selection criteria. One reason is that lag order estimates tend to be biased downward in practice (see Kilian 2001). A case in point is Kilian (2009) who set $p = 24$. For expository purposes, the error term is modelled as iid based on the empirical error distribution.

Table 2 compares the effective coverage rates of nominal 68% and 95% joint confidence sets for all 162 structural impulse response coefficients of interest. The choice of the significance level is motivated by the focus on one- and two-standard pointwise error bands in applied macroeconomic research, which corresponds to 68% and 95% confidence intervals under normality. When the number of structural impulse responses of interest exceeds the number of model parameters, as is the case in this design when $p = 6$ and $p = 12$, the conventional asymptotic approximation for the W_T^γ-test statistic breaks down. Theorem 2 shows that our bootstrap asymptotic approximation, in contrast, remains equally valid regardless of the choice of p and the maximum horizon.

Table 2 confirms that this approximation works well in that the coverage rates of the $W_{1-\alpha}^\gamma$ confidence set are fairly accurate and stable across p. For the 68% joint confidence set, the $W_{1-\alpha}^\gamma$ coverage probabilities range from 66% to 71%; for the 95% joint confidence set from 94% to 96%. These coverage rates are closer to nominal coverage than for the Bonferroni approach. For example, the 68% Bonferroni bound yields effective coverage rates of between 79% and 84% and the 95% Bonferroni bound between 97% and 98%. As expected, given the magnitude of q, the Bonferroni bound is conservative. Bootstrapping $W_T^{\gamma(\theta)}$ instead of W_T^γ does not systematically improve coverage accuracy. In fact, in one case the $W_{1-\alpha}^{\gamma(\theta)}$ joint confidence region is noticeably less accurate than the $W_{1-\alpha}^\gamma$ region.

7.2   DGP 2: Quarterly VAR Model

The second DGP for the simulation study is a semi-structural recursively identified monetary policy VAR model of the type discussed in Christiano, Eichenbaum and Evans (1999). We focus on a prototypical model with only three variables, in which the Federal Reserve Board controls the interest rate by setting the policy innovation after observing the forecast errors for deflator inflation and real GDP growth. The interest rate is ordered last such that policy shocks do not affect inflation and growth within the current quarter. Only the monetary policy shock is identified. We are interested in the own-response of the interest rate to a monetary policy shock as well as the responses of inflation and economic growth at horizons of $h = 0, ..., 15$. The choice of a horizon of four years is typical in applied work on responses to monetary policy shocks based on quarterly data, given the persistence of the responses in question. The error term is again modelled as iid based on the empirical error distribution.

The DGP is based on estimates of this VAR model on U.S. data with lag order $p \in \{4, 6, 8\}$, reflecting common choices of the lag order for quarterly data. All models include an intercept. The estimation period is restricted to 1954.IV-2007.IV in order to exclude the recent period of unconventional monetary policy measures. The sample size in the simulation study matches that in the underlying data for maximum realism. All data are obtained from the FRED data base of the Federal Reserve Bank of St. Louis. Quarterly GNP is expressed in chained 2005 dollars and seasonally, as is the corresponding GNP deflator. The effective federal funds rate is averaged to quarterly frequency.

Table 3 shows that the effective coverage rates of the Wald confidence set for the 48 structural impulse response coefficients of interest are again reasonably accurate, consistent with our theoretical results. At the 68% significance level, coverage rates are between 63% and 66%; at the 95% significance level, between 92% and 93%. In contrast, the Bonferroni bound results in excessively high coverage rates ranging from 83% to 88% and from 96% to 98%, respectively. Again, bootstrapping $W_T^{\gamma(\theta)}$ instead of W_T^{γ} does not systematically

improve coverage accuracy. In the few cases in which there is a larger difference in coverage accuracy, Table 3 shows that the $W_{1-\alpha}^{\gamma(\theta)}$ confidence region is less accurate. On the basis of this evidence and the corresponding results in Table 2 we restrict attention to $W_{1-\alpha}^{\gamma}$ in the empirical analysis discussed below.

7.3   Comparison with Other Simulation Evidence

We conclude that inverting the Wald test statistic of Theorem 2 produces reasonably accurate joint confidence sets even in high-dimensional problems. Although our asymptotic approximation did not allow for roots that are local to unity, these examples suggest that our method provides a good approximation even when the dominant autoregressive root is quite close to unity. For example, in the monthly simulation design, the roots of the data generating processes range from 0.967 to 0.989, depending on p. In the quarterly simulation design the dominant root may be as high as 0.948.

Likewise, it is worth pointing out that our application of the bootstrap does not involve any bias adjustments for the slope parameters of the type proposed by Kilian (1998) for pointwise inference about impulse responses. Such adjustments do not appear necessary in our context. It is also of interest to note that the joint confidence sets remain accurate even at relatively long horizons, at which conventional pointwise asymptotic and bootstrap intervals already tend to become inaccurate (see Kilian and Chang 2000). Finally, it is useful to highlight some differences between our analysis and the simulation evidence reported in Lütkepohl et al. (2014). Their simulation design was a bivariate VAR(1) model with a wide range of different degrees of persistence and sample sizes. The relevant comparison is to their results for highly persistent VAR processes.

First, whereas Lütkepohl et al. (2014) show that Bonferroni sets are at best mildly conservative in terms of coverage, we found them to be considerably more conservative for our simulation designs. One would expect this difference to be largely explained by the fact that q in our simulation designs is much larger than in their study, which puts the

23

Bonferroni method at a disadvantage.

Second, Lütkepohl et al. (2014) find that the Wald confidence band $W_{1-\alpha}^{\gamma(\theta),band}$ has reasonably accurate coverage for the simulation designs in question when q is moderately large. We show the Wald confidence sets based on $W_{1-\alpha}^{\gamma}$ and $W_{1-\alpha}^{\gamma(\theta)}$ to be reasonably accurate for much larger q. There is one important difference, however. Because we are not concerned with the construction of joint confidence bands, we evaluate the coverage accuracy of the joint Wald confidence sets instead by assessing with what probability the joint Wald test statistics W_T^{γ} and W_T^{θ} lie outside their respective bootstrap critical region. Lütkepohl et al. (2014), in contrast, assess the coverage accuracy of the joint confidence band by recording the relative frequency with which the joint confidence band entirely includes the corresponding vector of true impulse responses. Hence, the results are not directly comparable.

Third, because forming two-dimensional confidence bands by the projection method involves a loss of power, the $W_{1-\alpha}^{\gamma(\theta),band}$ confidence band may be wider on average than the Bonferroni band. Lütkepohl et al. (2014) show by simulation that the Bonferroni bounds indeed in many situations can produce tighter confidence bands than $W_{1-\alpha}^{\gamma(\theta),band}$, while preserving coverage rates of at least $1 - \alpha$. These situations do not include the highly persistent VAR processes representative of much of applied work, however, as shown in their Table 4. There is little to choose between these two approaches in the latter case.

An interesting extension of Lütkepohl et al.'s work would be to compare the finite-sample coverage accuracy and average width of the confidence bands based on $W_{1-\alpha}^{\gamma,band}$ to $W_{1-\alpha}^{\gamma(\theta),band}$. We do not pursue this question because our analysis focuses on joint confidence regions rather than bands. The reason is that the construction of joint confidence bands based on the Wald confidence region, while facilitating comparisons with the Bonferroni bounds, negates some of the advantages of working with shotgun plots of the impulse responses based on the joint Wald confidence set. This point is illustrated in the next section. For expository purposes we focus on the shotgun plots associated with the $W_{1-\alpha}^{\gamma}$

confidence region.

8   Empirical Examples

8.1   A VAR Model of U.S. Monetary Policy

The first example is a VAR(4) model of U.S. monetary policy estimated on quarterly data for 1959.I-2007.IV. The variables include GNP deflator inflation, real GNP growth and the federal funds rate. The dominant autoregressive root is 0.95. The model is identified recursively with the federal funds rate ordered last. We are interested in the responses of the model variables to an unexpected monetary policy tightening, corresponding to an unexpected increase in the federal funds rate, so $q = 48$.

Figure 1 shows the three shotgun plots obtained by plotting all sets of structural impulse responses contained in the 68% joint confidence set based on W_T^γ. An easy way of testing whether any of these responses is zero is to search for impulse response functions that cross the zero line. We conclude that neither the response function of real GNP growth nor the response function of inflation is distinguishable from the zero line at the 68% significance level. Exactly the same result would be obtained by constructing the joint confidence band as in Lütkepohl et al. (2014) by connecting the uppermost realizations of the shotgun plot at each horizon to form the upper confidence band and by connecting the lowest realizations by horizon to form the lower band. In this sense, these two approaches convey the same information.

Such a joint confidence band, however, would tend to obscure information about the shape of a given response function. Consider, for example, the response of the level of U.S. real GNP to an unexpected monetary policy tightening, which is obtained by cumulating the responses of the growth rate. Standard business cycle theory implies that the increase in real GDP caused by a unexpected loosening of monetary policy should be hump-shaped. Figure 3 shows this response obtained by flipping the sign of the cumulated growth rate

25

response shown in Figure 2. We also added a visual representation of the joint confidence band obtained by constructing an envelope around the shotgun plot. It is unclear from inspecting the confidence band, what the shape of the response function is. Indeed, the confidence band is wide enough to accommodate any number of shapes of the response function, some consistent with economic theory and some not.

Most economists would be interested in the question of whether economic models implying a hump-shaped response of real output to a monetary policy shock are consistent with the evidence in Figure 2. If we are willing to define a hump shape, as most macroeconomists are prone to be based on our reading of the VAR literature, we can run an iterative search on the response functions contained in the 68% joint confidence set to determine whether any of the response functions in the set are inconsistent with a hump shape. Evidence that all response functions in the 68% confidence set are hump-shaped would lead us to reject the hypothesis that there is no hump in the response function. Evidence that none of the response functions in the set is hump-shaped, in contrast, would imply a rejection of the hypothesis of a hump-shaped response. Evidence that some response functions in the set are hump-shaped and some are not, would be imply that neither hypothesis can be rejected at the 68% confidence level.

For example, for expository purposes, we may define a *hump-shaped* response function for U.S. real GNP as one whose maximum occurs between horizons 1 and 15. Given that the response starts at zero by construction, effectively this definition rules out response functions that reach their maximum at horizon 15. Table 4 shows that 28.6% of the response functions in Figure 3 are not hump-shaped by this minimal definition of a hump shape, so we cannot be confident at the 68% level that the response function of U.S. real GNP is hump-shaped. The data in the shotgun plot are consistent with other interpretations of the shape of the response function. Similar results also hold at the 95% significance level.

This is more information than could have been gleaned from the joint confidence band.

In this particular example, one could have inferred that there must be at least one response function that is not hump-shaped from the fact that upper bound of the confidence band peaks at horizon 15. On the other hand, one could not have inferred the presence of hump-shaped response functions from the joint confidence band.

8.2 A VAR Model of the Stagflationary Effects of Oil Price Shocks

Our second empirical example focuses on a hypothesis that involves multiple response functions. Evidence that the economy remains below potential, while inflation continues to rise, is inconsistent with the standard accelerationist model of the macroeconomy and thus would seem to require a different explanation, presumably one based on domestic supply shocks that shift the Phillips curve. A popular argument in macroeconomics has been that oil price shocks in particular may act as such supply shocks for the U.S. economy. Thus, the question of whether oil price shocks are stagflationary has a long tradition in macroeconomics (see, e.g., Blanchard 2002; Barsky and Kilian 2004; Kilian 2008).

We address this question by postulating a VAR(4) model for the U.S. economy for the percent change in the real WTI price of crude oil, GDP deflator inflation and real GDP. The data are quarterly and the estimation period is 1987.I-2013.II. The starting date was chosen for illustrative purposes, given evidence for a possible structural break in the relationship between the real price of oil and U.S. real GDP in 1987. The dominant autoregressive root is 0.85. The model is identified recursively with the real price of oil ordered first. We focus on the effect of an unanticipated increase in the real price of oil on the real price of oil, on the change in inflation and on U.S. real GDP growth.

Figure 3 shows that the shotgun plots for the change in inflation and for real GDP growth cover the zero line at all horizons. This fact alone, however, tells us nothing about the question of whether these responses are stagflationary. To assess the latter question we need to look at these two response functions pairwise for each model in the joint confidence

27

set and verify whether the responses of $\Delta\pi_{t+h}$ and Δy_{t+h}, where π stands for GDP deflator inflation and Δy for real GDP growth, to an oil price shock are of opposite sign for all horizons h of interest. This may easily be accomplished with some extra code. Table 5 shows that, at the 68% significance level and looking jointly at horizons $h = 1, ..., 15$, not a single structural model estimate in the joint confidence set is consistent with the hypothesis of stagflationary responses to oil price shocks. Thus, we can rule out that hypothesis. One might have conjectured that increasing the confidence level would lead us to revise this statement, but Table 5 shows that the same result holds at the 95% significance level.

If we change the hypothesis of interest and focus on the first year following the oil price shocks only, however, in about 0.5% of the models within the joint confidence set is there evidence of a stagflationary response. In the remainder, there is not. Thus, the data would not allow us to reject either hypothesis of interest at the 68% and 95% significance levels.

It would have been tempting to conclude based on a joint confidence band that Figure 3 allows for stagflationary responses because both bands include positive as well as negative values, so hypothetically a stagflationary response would fit within this confidence band for $h = 1, ..., 15$. This conclusion, however, would have been wrong, considering the evidence in Table 5 based on the shotgun plot. This example reinforces our point that joint confidence bands obscure important information about the shapes of impulse response functions.

Figure 4 illustrates how the sets of response functions associated with the models in the joint confidence region that exhibit stagflationary responses, as defined in Table 2, may be highlighted within the shotgun plot. This allows us, for example, to infer that none of these stagflationary models are associated with large reductions in real GDP. The largest reduction in real GDP at horizon 15 that is contained in the joint confidence set is -0.6 percent, so even if we cannot rule out stagflationary effects completely at the 68% confidence level, we can be confident that such effects are not quantitatively important

for real output. This is one more example in which the shotgun plot reveals additional information not conveyed by confidence bands.

9 Conclusion

We considered the problem of constructing joint confidence sets for subsets of structural impulse responses that remain asymptotically valid even when the joint limiting distribution of the structural impulse responses becomes degenerate, which occurs when the number of responses considered exceeds the number of model parameters. We made the case that applied users should invert the joint Wald test statistic to obtain such a joint confidence set. We considered two alternative specifications of the Wald test statistic. Both are asymptotically valid and were shown to imply joint confidence sets with reasonably accurate coverage in finite samples. Our simulation evidence suggested that in the few cases, in which there is a noticeable difference in coverage accuracy, our preferred specification is more accurate in finite samples.

We proposed to represent the sets of structural responses associated with the estimates in the joint confidence set in the form of shotgun plots of the impulse response functions. We made the case that this approach preserves additional information about the shape of the impulse response functions that is lost when the results are presented in the form of confidence bands. In fact, the use of shotgun plots is essential for answering many of the key questions applied users want to answer based on structural VAR models. These questions relate not so much to whether a particular set of responses is significantly different from zero, although that question as well can be answered with the help of shotgun plots, but whether multiple sets of response functions follow the shape and pattern implied by economic theory. We demonstrated by example that the latter type of question cannot be answered based on joint confidence bands, whether constructed from Bonferroni bounds of by applying the projection method to Wald confidence sets. Appropriate answers can be obtained, however, based on the shotgun plots of the impulse response functions.

Our theoretical analysis involved a novel approach to approximating the distribution of test statistics with an asymptotically singular covariance matrix that is distinct from alternative approaches such as Andrews (1987) and is of independent interest for other applications. Although we focused on the construction of joint confidence sets about structural impulse responses, our results on bootstrapping Wald statistics for impulse response vectors with a degenerate joint asymptotic distribution can also be applied in a variety of other contexts. For example, they facilitate the design of statistical tests about various features of structural impulse response functions (e.g., Lütkepohl 1996). They also are relevant for the development of impulse response matching estimators of the structural parameters of dynamic stochastic general equilibrium models (see Guerron-Quintana, Inoue and Kilian 2014).

In this paper, we followed much of the VAR literature on impulse response analysis by focusing on Wald test statistics. An obvious extension would be to apply our approach to the LR test statistic. We defer this extension to future research, given the high coverage accuracy of the Wald confidence region in our simulation study.

10 Technical Appendix

Proof of Theorem 1.

It follows from the Schur decomposition theorem (Theorem 13 of Magnus and Neudecker, 1999, p.16) that there exists an orthonormal matrix S whose columns are eigenvectors of $B_0 B_0'$ and a diagonal matrix Λ whose diagonal elements are the eigenvalues of $B_0 B_0'$ such that

$$S' B_0 B_0' S = \Lambda \tag{23}$$

Stack the eigenvectors associated with the k largest eigenvalues of $B_0 B_0'$ in S_0 and let $q_0 = k$. Using a subset of the $q - q_0$ remaining eigenvectors that are not used in S_0, form S_1 such that $S_1' B_1$ contains no row vector of zeros. Let q_1 denote the number of columns in S_1. Using a subset of the $q - q_0 - \cdots - q_{j-1}$ remaining eigenvectors, that are not used in $S_0, S_1, \ldots S_{j-1}$, form S_j so that $S_j' B_j$ contains no row vector of zeros. Let q_j be the number of columns in S_j. Stop when $q_0 + q_1 + \cdots + q_r = q$. With some abuse of notation, let $S = [S_1 \ S_2 \ \cdots \ S_r]$. Then it follows from the definition of S_j and Assumption (b) that

$$
\begin{aligned}
\Upsilon_T S'(\gamma(\bar{X}_T) - \gamma(\mu)) &= S' B_0 Z_T + T^{-\frac{1}{2}} S' B_1 (Z_T \otimes Z_T) + o_p(T^{-\frac{1}{2}}) \\
&= \begin{bmatrix} S_0' B_0 Z_T \\ S_1' B_1 (Z_T \otimes Z_T) \\ \vdots \\ S_r' B_r (Z_T \otimes \cdots \otimes Z_T) \end{bmatrix} + o_p(1),
\end{aligned} \tag{24}
$$

It furthermore follows from (24) and Assumption (a) that

$$
\Upsilon_T S'(\gamma(\bar{X}_T) - \gamma(\mu)) \xrightarrow{d} \begin{bmatrix} S_0' B_0 Z \\ S_1' B_1 (Z \otimes Z) \\ \vdots \\ S_r' B_r (Z \otimes \cdots \otimes Z) \end{bmatrix}. \tag{25}
$$

Because of Assumptions (a) and (b) and because of the continuity of eigenvalues and eigenvectors as a function of matrices, we also know that

$$
\Upsilon_T S'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T)) \xrightarrow{d} \begin{bmatrix} S_0' B_0 Z^* \\ S_1' B_1 (Z^* \otimes Z^*) \\ \vdots \\ S_r' B_r (Z^* \otimes \cdots \otimes Z^*) \end{bmatrix}, \tag{26}
$$

where the convergence is with respect to the bootstrap probability measure conditional on the data with probability one, Therefore Theorem 1 follows from (25) and (26).

Proof of Theorem 2. It follows from (26) that the limiting covariance matrix of

$\Upsilon_T S'(\gamma(\bar{X}_T) - \gamma(\mu))$ is

$$J = E\left\{\begin{bmatrix} S_0'B_0Z \\ S_1'B_1(Z \otimes Z) \\ \vdots \\ S_r'B_r(Z \otimes \cdots \otimes Z) \end{bmatrix}\begin{bmatrix} S_0'B_0Z \\ S_1'B_1(Z \otimes Z) \\ \vdots \\ S_r'B_r(Z \otimes \cdots \otimes Z) \end{bmatrix}'\right\}. \tag{27}$$

Similarly, the limiting covariance matrix of $\Upsilon_T S'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T))$ is also (27) conditional on the data with probability one. It follows from Assumption (c) that

$$\Upsilon_T S'\widehat{\Sigma}_T^* S\Upsilon_T \xrightarrow{p} J. \tag{28}$$

Theorem 2 follows from Theorem 1, (27) and (28).

11 References

1. Andrews, D.W.K., (1987), "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348-358.

2. Antoine, B., and E. Renault (2012), "Efficient Minimum Distance Estimation with Multiple Rates of Convergence," *Journal of Econometrics*, 1709, 350-367.

3. Bao, Y., and A. Ullah, (2007), "The Second-Order Bias and Mean Squared Error of Estimators in Time Series Models," *Journal of Econometrics*, 140, 650-669.

4. Bao, Y. (2007), " The Approximate Moments of the Least Squares Estimator for the Stationary Autoregressive Model Under a General Error Distribution," *Econometric Theory*, 23, 1013-1021.

5. Barsky, R.B., and L. Kilian (2004), "Oil and the Macroeconomy since the 1970s," *Journal of Economic Perspectives*, 18, Fall, 115-134.

6. Benkwitz, A., Lütkepohl, H., and M.H. Neumann (2000), "Problem Related to Bootstrapping Impulse Responses of Autoregressive Processes," *Econometric Reviews*, 19, 69-103.

7. Berkowitz, J., and L. Kilian (2000), "Recent Developments in Bootstrapping Time Series," *Econometric Reviews*, 19, 1-54.

8. Blanchard, O.J. (1989), "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review*, 79, 1146-1164.

9. Blanchard, O.J. (2002), "Comments on 'Do We Really Know that Oil Caused the Great Stagflation? A Monetary Alternative' by Robert Barsky and Lutz Kilian," in: Bernanke, B.S., and K. Rogoff (eds.), *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press, 183-192.

10. Cochrane, J.H. (1994), "Shocks," *Carnegie-Rochester Conference Series on Public Policy*, 41, 295-364.

11. Christiano, L., M. Eichenbaum and C. Evans (1999), "Monetary policy shocks: what have we learned and to what end?", in J.B Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, 1, Amsterdam: Elsevier/North Holland, 65-148.

12. Freedman, D., (1984), "On Bootstrapping Two-Stage Least-Squares Estimators in Stationary Linear Models," *Annals of Statistics*, 12, 827-842.

13. Gonçalves, S., and L. Kilian (2004), "Bootstrapping Autoregressions in the Presence of Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89-120.

14. Guerron-Quintana, P., Inoue, A., and L. Kilian (2014), "Impulse Response Matching Estimators for DSGE Models," mimeo, University of Michigan.

15. Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer: New York.

16. Inoue, A., and L. Kilian (2013), "Inference on Impulse Response Functions in Structural VAR Models," *Journal of Econometrics*, 177, 1-13.

17. Jordà, Ò. (2009), "Simultaneous Confidence Regions for Impulse Responses," *Review of Economics and Statistics*, 91, 629-647.

18. Kilian, L. (1998), "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80, 218-230.

19. Kilian, L. (1999), "Finite-Sample Properties of Percentile and Percentile-*t* Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics*, 81, 652-660.

20. Kilian, L. (2001), "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order", *Journal of Forecasting*, 20, 161-179.

21. Kilian, L. (2008), "A Comparison of the Effects of Exogenous Oil Supply Shocks on Output and Inflation in the G7 Countries", *Journal of the European Economic Association*, 6, 78-121.

22. Kilian, L. (2013), "Structural Vector Autoregressions," in: *Handbook of Research Methods and Applications in Empirical Macroeconomics*, N. Hashimzade and M. Thornton (eds.), Cheltenham, UK: Edward Elgar, 515-554.

23. Kilian, L., and P.-L. Chang (2000), "How Accurate are Confidence Intervals for Impulse Responses in Large VAR Models?", *Economics Letters*, 69, 299-307.

24. Kilian, L., and Y.J. Kim (2011), "How Reliable Are Local Projection Estimators of Impulse Responses?" *Review of Economics and Statistics*, 93, 1460-1466.

25. Lütkepohl, H. (1990), "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics*, 72, 116-25.

26. Lütkepohl, H. (1996), "Testing for Nonzero Impulse Responses in Vector Autoregressive Processes," *Journal of Statistical Planning and Inference*, 50, 1-20.

27. Lütkepohl, H., Staszewska-Bystrova, A., and P. Winker (2013), "Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions," forthcoming: *International Journal of Forecasting*.

28. Lütkepohl, H., Staszewska-Bystrova, A., and P. Winker (2014), "Confidence Bands for Impulse Responses: Bonferroni versus Wald," manuscript, DIW Berlin.

29. Lütkepohl, H., and D.S. Poskitt (1991), "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," *Econometric Theory*, 7, 487-496.

30. Magnus, J.R., and H. Neudecker (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition, Wiley: Chichester: England.

31. Phillips, P.C.B. (1989), "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

32. Sims, C.A., J.H. Stock and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.

33. Sims, C.A., and T. Zha, (1999), "Error Bands for Impulse Responses," *Econometrica*, 67, 1113-1156.

34. Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton: Princeton University Press.

Table 1: Relative Volumes of Bonferroni and Wald Confidence Regions

M	Confidence level $1-\alpha = 0.9$		
	$B_{1-\alpha}^{\theta}$	$W_{1-\alpha}^{\theta,band}$	$W_{1-\alpha}^{\theta}$
2	1.01	1.22	0.95
3	1.02	1.65	0.87
4	1.03	2.46	0.76
5	1.03	3.94	0.65
6	1.04	6.67	0.54
7	1.05	11.89	0.44
8	1.05	22.17	0.35
9	1.06	43.02	0.28
10	1.07	86.54	0.22

NOTES: The volume of the $1-\alpha$ confidence region refers to an M-dimensional cube, where M is the length of the vector θ. All results are normalized relative to the volume of the joint confidence region associated with the precise confidence box that represents the tightest two-dimensional confidence box possible. $B_{1-\alpha}^{\theta}$ refers to the joint confidence region implied by Bonferroni confidence band. $W_{1-\alpha}^{\theta,band}$ refers to the joint confidence region implied by the confidence band proposed in Lütkepohl et al. (2014), and $W_{1-\alpha}^{\theta}$ relates to the joint confidence region implied by the Wald test statistic.

Table 2: Joint inference for all impulse responses at horizon $h = 0,...,17$ in a monthly oil market VAR model

	Coverage Rates (%)					
	$W_{1-\alpha}^{\gamma}$		$W_{1-\alpha}^{\gamma(\theta)}$		$B_{1-\alpha}^{\gamma}$	
Nominal	68	95	68	95	68	95
$p = 6$	70.8	93.8	63.0	93.8	84.0	97.0
$p = 12$	65.6	96.0	65.4	94.0	80.6	98.0
$p = 24$	67.2	94.2	71.6	95.4	78.8	98.0

NOTES: Coverage rates based on 500 trials, each of which involves $1,000 \times 1,000$ bootstrap replications. The model is motivated by the analysis in Kilian (2009). Based on the oil market data set in Kilian (2009), alternative DGPs are constructed for lag orders $p \in \{6, 12, 24\}$. $W_{1-\alpha}^{\gamma}$ and $W_{1-\alpha}^{\gamma(\theta)}$ stand for the two alternative joint Wald confidence regions discussed in section 6 and $B_{1-\alpha}^{\gamma}$ for the Bonferroni set.

Table 3: Joint inference for all responses to the monetary policy shock at horizon
$h = 0, ..., 15$ **in a quarterly monetary policy VAR model**

| | Coverage Rates (%) | | | | | |
| | $W_{1-\alpha}^{\gamma}$ | | $W_{1-\alpha}^{\gamma(\theta)}$ | | $B_{1-\alpha}^{\gamma}$ | |
Nominal	68	95	68	95	68	95
$p = 4$	63.4	93.2	61.0	92.0	87.6	98.2
$p = 6$	66.4	93.4	64.2	93.0	87.6	98.0
$p = 8$	66.4	92.2	62.6	89.2	83.2	96.4

NOTES: Coverage rates based on 500 trials, each of which involves $1,000 \times 1,000$ bootstrap replications. The DGP is a prototypical recursively identified monetary policy VAR model for U.S. deflator inflation, real GDP growth and the federal funds rate, as discussed in Kilian (2013). Based on data for 1954.III-2007.IV, alternative DGPs are constructed from the fitted models for lag orders $p \in \{4, 6, 8\}$. $W_{1-\alpha}^{\gamma}$ and $W_{1-\alpha}^{\gamma(\theta)}$ stand for the two alternative joint Wald confidence regions discussed in section 6 and $B_{1-\alpha}^{\gamma}$ for the Bonferroni set.

Table 4: Percentage of models in joint Wald confidence set consistent with a hump-shaped response function of real GNP to an unexpected loosening of monetary policy

| Joint Confidence Level | |
68%	95%
71.4%	72.4%

NOTES: Based on the model underlying Figure 2 and a minimal definition of a hump shaped response as a response reaching its maximum between horizon 1 and 15.

Table 5: Percentage of models in joint Wald confidence set inconsistent with a stagflationary response to an unexpected increase in the real price of oil

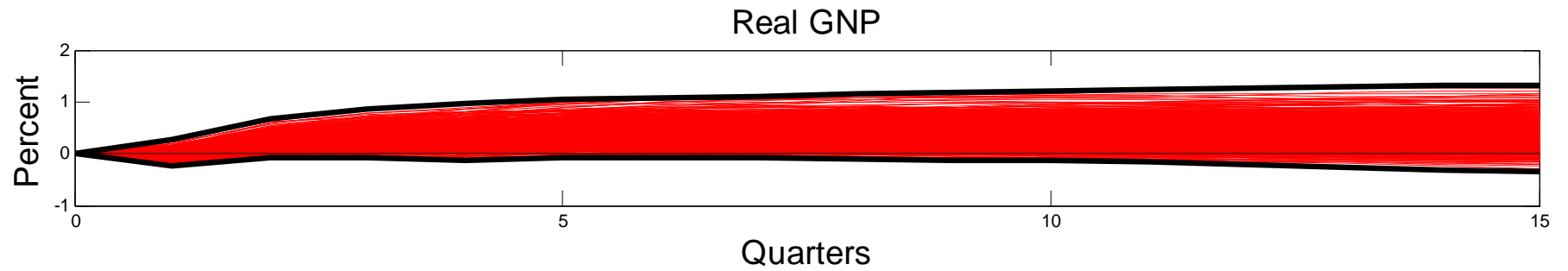| | Joint Confidence Level | |
Horizons (h)	68%	95%
1,…,15	100.0%	100.0%
1,…,4	99.5%	99.4%

NOTES: Based on the model underlying Figure 3. The stagflation hypothesis translates to the responses of $\Delta\pi_{t+h}$ and Δy_{t+h} to the oil price shock being of opposite sign for all h of interest and the inflation acceleration being positive at horizon 1.

Figure 1: Responses to an Unexpected Tightening of Monetary Policy
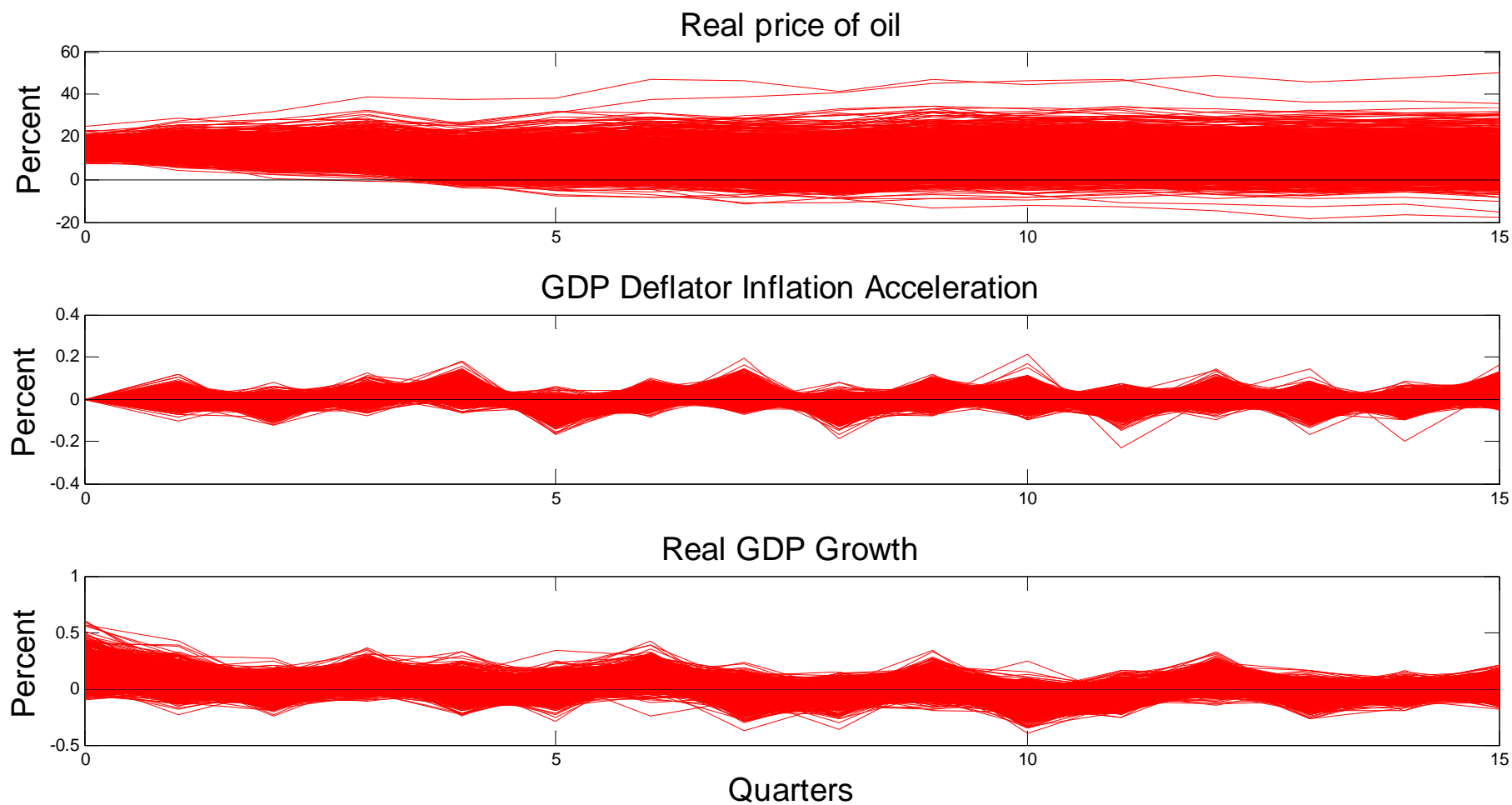**Shotgun Plot Implied by Joint 68% Wald Confidence Set**
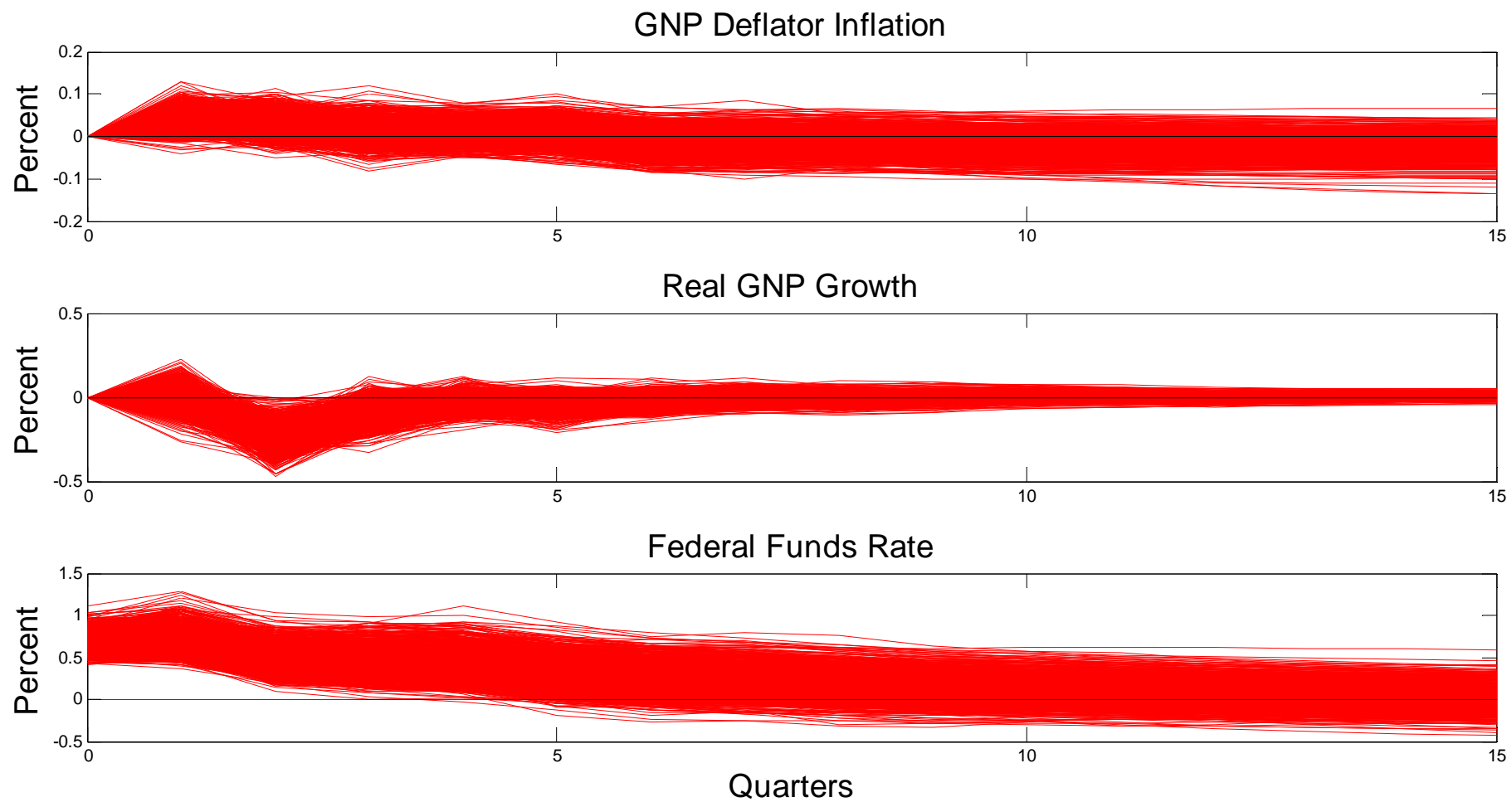
NOTES: Estimates based on a VAR(4) model with intercept including deflator inflation, real GNP growth and the federal funds rate obtained by fitting U.S. data for 1959.I-2007.IV All results based on 2,000×2,000 bootstrap replications.

Figure 2: Responses of U.S. Real GNP to an Unexpected Loosening of Monetary Policy
**Shotgun Plot Implied by Joint 68% Wald Confidence Set**

Real GNP

NOTES: Obtained from Figure 2 by cumulating the responses of the growth rate of real GNP and flipping the sign. The joint confidence band obtained by constructing an envelope around the shotgun plot is shown as bold lines.

Figure 3: Responses to an Unexpected Increase in the Real Price of Oil
Shotgun Plot Implied by Joint 68% Wald Confidence Set

NOTES: Estimates based on a VAR(4) model with intercept including the real price of oil, GDP deflator inflation, and real GDP growth obtained by fitting U.S. data for 1987.I-2013.II. All results based on 2,000×2,000 bootstrap replications.

Figure 4: Responses to an Unexpected Increase in the Real Price of Oil
Shotgun Plot Implied by Joint 68% Wald Confidence Set with Subset of Stagflationary Responses Highlighted
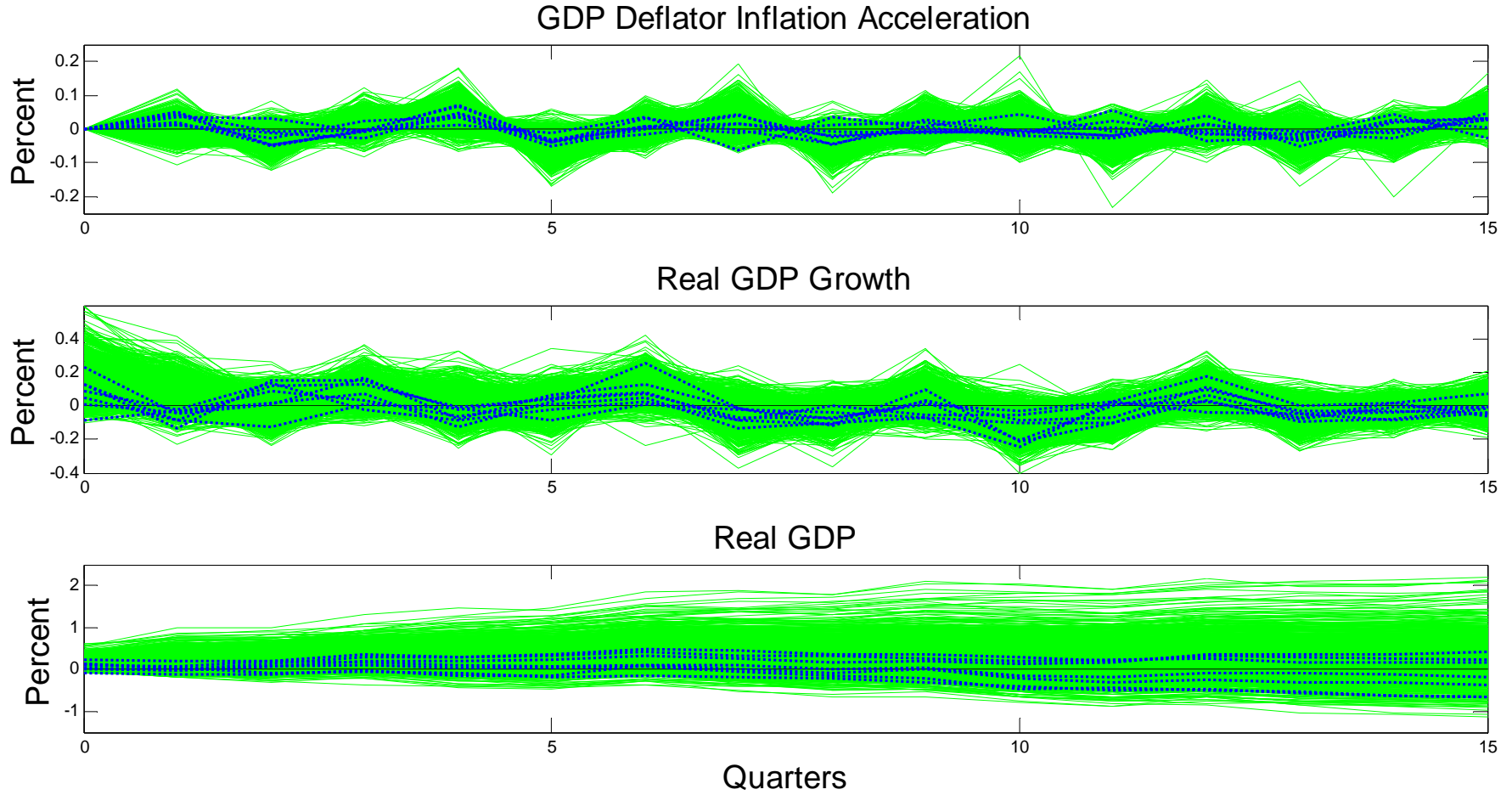
NOTES: The pairs of response functions associated with the 3% of models in the 68% joint confidence set that exhibit stagflationary responses at horizons $0, ..., 4$ and a positive inflation acceleration response at horizon 1 are highlighted as dark dotted lines. There are no models with stagflationary responses at all horizons.