ON THE MAXIMUM AND MINIMUM RESPONSE TO AN IMPULSE IN SVARS

Bulat Gafarov[1] and José Luis Montiel Olea[2]

This paper proposes an intuitive, computationally simple, 'adjusted' delta method confidence interval for set-identified coefficients of the impulse-response function in a Structural Vector Autoregression. We establish the uniform asymptotic validity of our inference procedure in models that impose zero and sign restrictions only on the *contemporaneous* responses to one structural shock.

To illustrate our inference approach, we use a monetary Structural Vector Autoregression estimated with monthly U.S. data. We set-identify the dynamic responses of different variables to an *unconventional* monetary shock that decreases the two-year government bond rate upon impact, but has no effect over the nominal federal funds rate. We impose two additional sign restrictions on the contemporaneous responses of inflation and output. Based on the estimated parameters for this model, we present Monte-Carlo evidence that supports the validity of our adjusted delta method.

The construction of our confidence interval does not require random sampling from the space of rotation matrices or unit vectors. Instead, we treat the bounds of the identified set for the coefficients of impulse-responses as the *maximum and minimum value* of a mathematical program and we provide formulas for these values and their derivative. We also discuss the extension of our delta method approach to models with noncontemporaneous restrictions on one structural shock and models with contemporaneous sign restrictions on multiple shocks.

Keywords: Sign-restricted SVARs, Set-Identified Models, Delta Method, Hadamard Directional Differentiability.

1. INTRODUCTION

A Structural Vector Autoregression (SVAR) [Sims (1980, 1986)] is a time series model that brings restrictions into a reduced-form, linear forecasting multivariate regression. The restrictions are used to identify the dynamic response of forecasted variables to exogenous structural impulses in the economy. Depending on the restrictions imposed, the parameters of the reduced-form model can be associated to a unique collection of structural dynamic responses (point identification) or to many of them (set identification).

[1] Penn State University, Department of Economics and National Research University Higher School of Economics. E-mail: bzg134@psu.edu.

[2] I owe special thanks to Ennio Stacchetti for multiple references and useful insights concerning Lagrange Multipliers, Mathematical Programming and Theorems of the Alternative. I am also very grateful to Jonas Arias for sharing his code to construct Bayesian Credible Sets with sign/zero restrictions and for helpful conversations about SVARs. New York University, Department of Economics. E-mail: montiel.olea@nyu.edu.

[3] We are extremely grateful to Lutz Kilian for very helpful and detailed comments on an earlier draft of this paper. We would like to thank Saroj Bhattarai, Matias Cattaneo, Tim Cogley, Emmanuel Guerre, Patrik Guggenberger, Boyan Jovanovic, Peter Karadi, Sung Jae Jun, Adam McCloskey, Francesca Molinari, Joris Pinkse, Mikkel Plagborg-Møller, Jörg Stoye and seminar participants at the Harvard Econometrics Lunch, the NYU Econometrics Lunch, the NYU Macroeconomics Lunch, the econometrics workshop at Brown, Cornell, ITAM, Michigan, PSU, Rutgers, Vanderbilt, Wisconsin-Madison, and the 2014 NSF-Time Series Conference. The usual disclaimer applies. This draft: January 23th, 2015. Firs draft: November 19th, 2014.

It is now customary for empirical macroeconomic studies to impose sign and/or zero restrictions on structural dynamic responses in order to achieve set identification, as in the pioneering work of Faust (1998), Canova and Nicoló (2002), and Uhlig (2005). Most of these studies use numerical Bayesian methods to construct posterior *credible sets* for the coefficients of the impulse-response function.

This paper proposes an intuitive, computationally simple, 'adjusted' delta method procedure to conduct frequentist inference about the coefficients of impulse-response functions in SVARs. Our objective is to provide researchers with a simple frequentist tool to complement the information contained in a Bayesian credible set. We establish the uniform asymptotic validity of our inference procedure in models that impose zero and/or sign restrictions only on the *contemporaneous* responses to one structural shock, but we show that our approach is defined more generally.

The construction of our frequentist confidence interval does not require random sampling from the space of rotation matrices or unit vectors. Instead, we treat the bounds of the identified set for the coefficients of impulse-responses as the *maximum and minimum value* of a linear mathematical program with one quadratic constraint. For models with zero and sign restrictions on only one shock, we use Karush-Kuhn-Tucker conditions to provide formulas for these values and also for their derivatives. We 'adjust' the delta method formula to take into account possible violations of differentiability. The properties of the maximum and the minimum response are a key intermediate step to conduct frequentist inference about the coefficients of the impulse-response function.

The recent paper of Moon, Schorfheide, and Granziera (2013) [MSG] also studies frequentist inference procedures for SVARs with sign and/or zero restrictions. MSG consider both projection and Bonferroni inference using the moment-inequality framework in Andrews and Soares (2010). MSG establish the uniform validity of their inference procedures in a more general class of models than ours. We present Monte-Carlo evidence suggesting that our delta method approach has a reasonable performance beyond the 'one shock-contemporaneous restrictions' model. We also present a simpler version of MSG's projection method to compare it with our delta method confidence interval. The examples in this paper suggest that the adjusted delta method confidence intervals can be significantly tighter than those based on projection.

We see two advantages of our approach. The first one is computational tractability: we do not perform numerical inversion of a hypothesis test to construct our delta method confidence interval. The second advantage is a large-sample consideration: the delta method confidence interval has exact asymptotic coverage whenever the *maximum/minimum responses* to the shock of interest are continuously differentiable functions of the reduced-form parameters and the true structural parameter coincides with one of the bounds.[1]

[1]Let $\theta \in \Theta$ be a parameter and let $\overline{v}(\cdot)$ be some transformation of it. We say that a nominal $100(1-\alpha)\%$ confidence set CS_T for $\overline{v}(\theta)$ has *correct asymptotic coverage* for a sequence $\{\theta_T\} \subseteq \Theta$ if

The remainder of the paper is organized as follows. Section 2 presents the basic model and an overview of the main results. Section 3 presents the characterization of the maximum/minimum responses as value functions of a linear mathematical program with an additional quadratic equality constraint. Section 4 establishes the continuity and differentiability properties of the maximum/minimum response and relates them to the consistency and asymptotic normality of their plug-in estimators. Section 5 establishes the validity of our adjusted delta method confidence set for arbitrary drifting sequences of parameter values. This section also presents our own version of projection-based inference for SVARs (with no test inversion and no sampling of rotation matrices) and also an 'Asymptotic-Distribution bootstrap' (henceforth, AD bootstrap) implementation of our delta method confidence set. In section 6 we estimate a monetary SVAR with monthly U.S. data to set-identify the dynamic responses to an *unconventional*, expansionary monetary shock. We note that projection method confidence interval can be significantly larger than the adjusted delta method confidence interval. Section 7 presents the conclusions concerning the 'one shock-contemporaneous restrictions' model. In Section 8 we present applications of our delta method approach to models with a) noncontemporaneous restrictions on one shock and b) models with contemporaneous restrictions on multiple shocks. All the proofs are collected in the Appendix.

Generic Notation: If A is a matrix, A_{ij} denotes the ij-th element of A, $\text{vec}(A)$ denotes the vectorization of A, and $\text{vech}(A)$ denotes half-vectorization (applicable only if A is symmetric). The Kronecker product between matrices A and B is denoted by $A \otimes B$. The vector $e_i \in \mathbb{R}^n$ denotes the i-th column of the identity matrix of dimension n. If H is a matrix of dimension $n \times n$, $H_i \equiv He_i$ denotes its i-th column.

2. MODEL AND OVERVIEW OF THE MAIN RESULTS

This paper studies the n-dimensional Structural Vector Autoregression with p lags and unknown, invertible, $n \times n$ matrix H:

$$Y_t = A_1 Y_{t-1} + \dots A_p Y_{t-p} + \eta_t, \quad \eta_t \equiv H\varepsilon_t, \;\; \mathbb{E}[\varepsilon_t \varepsilon_t'] \equiv D = \text{diag}(\sigma_1^2, \dots \sigma_n^2). \tag{2.1}$$

The stochastic process in equation (2.1) is assumed to be covariance stationary. Consequently, equation (2.1) admits the structural moving average representation:

$$Y_t = \sum_{k=0}^{\infty} C_k(A) H \varepsilon_{t-k}, \tag{2.2}$$

$\liminf_{T\to\infty} \mathbb{P}(\overline{v}(\theta_T) \in \text{CS}_T) \geq 1 - \alpha$. We say that CS_T has *exact asymptotic coverage* for the sequence $\{\theta_T\}$ if $\lim_{T\to\infty} \mathbb{P}(\overline{v}(\theta_T) \in \text{CS}_T) = 1 - \alpha$.

where $C_0 = \mathbb{I}_n$ and

$$C_k(A) = \sum_{m=1}^{k} C_{k-m}(A)\, A_m, \quad k \in \mathbb{N}, \tag{2.3}$$

with $A_m = 0$ if $m > p$; see Lütkepohl (1990), p. 116. We will refer to $C_k(A)$ as the k-th *reduced-form* MA coefficient and, whenever convenient, we omit dependence on the reduced-form parameter $A = (A_1, A_2, \ldots A_p)$.

A fundamental result for the SVAR(p) model—first stated by Faust (1998), Canova and Nicoló (2002) and Uhlig (2005)—is that the structural parameter $HD^{1/2}$ can be set identified by sign/zero restrictions on the entries of H and the relation:

$$\Sigma \equiv \mathbb{E}[\eta_t \eta_t'] = HDH' = (HD^{1/2})(HD^{1/2})'. \tag{2.4}$$

In the majority of applications, the parameters of interest in (2.1) are the coefficients of the *structural impulse-response function* to some specific shock, $\varepsilon_{j,t}$.[2] For a given shock j under study, let $\text{IRF}_{k,ij}$ denote the response of the i-th component of Y_{t+k} to an impulse of magnitude σ_j in the shock $\varepsilon_{j,t}$—ceteris paribus. The k-th period ahead dynamic response of the the i-th component of Y_t to a j-th structural impulse is given by:

$$\text{IRF}_{k,ij}(A, H, D) \equiv \frac{\partial Y_{i,t+k}}{\partial \epsilon_{j,t}} \sigma_j = e_i' C_k(A) H e_j \sigma_j = e_i' C_k(A) H_j \sigma_j. \tag{2.5}$$

It is well-understood that under some conditions, for example Lemma 3.1 in the recent work of Giacomini and Kitagawa (2014), Lemma B.1 in Moon et al. (2013) or Lemma C-D in this paper, the following holds:

$$\text{IRF}_{k,ij}(A, H, D) \in [\underline{v}_{k,ij}(A, \Sigma)\ ,\ \overline{v}_{k,ij}(A, \Sigma)]. \tag{2.6}$$

That is, the identification region for the parameter of interest is an interval whose lower and upper bounds can, in principle, be estimated from the data as in Imbens and Manski (2004) or Stoye (2009).

The main objective of this paper is to provide researchers with a simple frequentist tool to complement the information contained in Bayesian credible sets, which are the most common statistics used to summarize uncertainty in set-identifed SVARs.[3]

[2] We focus on constructing frequentist confidence intervals for each of the (k, i, j) coefficients of the impulse-response function. For Bayesian approaches on how to conduct *joint* inference for all the set-identified coefficients see Inoue and Kilian (2013).

[3] We also want to show that frequentist inference for set-identified SVARs need not be computationally demanding nor conceptually awkward. This concern has been raised recently by Baumeister and Hamilton (2014) when explaining the pervasiveness of Bayesian methods in the literature of SVARs with sign restrictions.

MAIN CONTRIBUTION: Let $\widehat{\theta}_T$ denote the Ordinary Least Squares (OLS) estimators of the reduced-form parameters of the VAR model and let $z_{1-\alpha}$ be the $1-\alpha$ quantile of a standard normal random variable. We propose an asymptotically valid 100(1-α)%-confidence interval for $\text{IRF}_{k,ij}$ of the form:

$$\Big[\underline{v}_{k,ij}(\widehat{\theta}_T) - z_{1-\alpha}\, \widehat{\underline{\sigma}}_T/\sqrt{T},\ \overline{v}_{k,ij}(\widehat{\theta}_T) + z_{1-\alpha}\, \widehat{\overline{\sigma}}_T/\sqrt{T}\Big]. \tag{2.7}$$

We provide formulas for $\widehat{\overline{\sigma}}_T$ and $\widehat{\underline{\sigma}}_T$ based on the delta method and also an algorithm to evaluate $\underline{v}_{k,ij}(\widehat{\theta}_T)$ and $\overline{v}_{k,ij}(\widehat{\theta}_T)$ without sampling from rotation matrices or unit vectors.

For fixed values of θ, the functions $\underline{v}_{k,ij}(\theta)$ and $\overline{v}_{k,ij}(\theta)$ correspond to the minimum and maximum value that the structural parameter $\text{IRF}_{k,ij}$ can take over the identified set. We characterize the points for which the maximum and minimum response are differentiable functions of the reduced-form parameters. We show that differentiability is only compromised when the maximum (minimum) response is close to zero, or whenever there are nonnested sets of active/binding constraints associated to the same maximum (minimum) response.

If the reduced-form parameters suggest that the bounds of the identified set are not differentiable, we *adjust* the delta method by *switching* to a projection confidence interval. Our suggested projection-based confidence interval starts with an asymptotically valid $100(1-\alpha)\%$ confidence region for $\widehat{\theta}_T$, and then uses our algorithm to evaluate $\underline{v}_{k,ij}(\theta)$ and $\overline{v}_{k,ij}(\theta)$ over a grid in the confidence region. The values between the largest $\overline{v}_{k,ij}(\theta)$ in the grid and the smallest $\underline{v}_{k,ij}(\theta)$ in the grid correspond to a conservative, asymptotically valid $100(1-\alpha)\%$ confidence set for the structural parameter $\text{IRF}_{k,ij}$.

METHODOLOGICAL CONTRIBUTIONS OF THIS PAPER: In models with sign/zero restrictions on only one structural shock, the bounds of the identified set can be characterized as the *value function* of a linear mathematical program with an additional quadratic equality constraint. This observation has been made before by Faust (1998), Uhlig (2005) and the more recent paper of Giacomini and Kitagawa (2014). Our first methodological contribution is to show that the function $\overline{v}_{k,ij}(\cdot)$ can be evaluated using a simple algorithm that exploits closed-form solutions for the Karush-Kuhn-Tucker points of the linear program with the additional quadratic constraint.

Our second methodological contribution is to use the algorithm and the characterization of the Karush-Kuhn-Tucker points to establish the continuity and everywhere Hadamard directional differentiability of the value function, $\overline{v}_{k,ij}(\cdot)$. A key observation is that $\overline{v}_{k,ij}(\cdot)$ is only *fully* differentiable at reduced-form parameters (A, Σ) for which the mathematical program defining $\overline{v}_{k,ij}(\cdot)$ has a unique maximizer. It is only at points of full differentiability that we can guarantee that the asymptotic distribution of the plug-in estimator is *regular*—independent of drifts in the reduced-form parameters—and given by a delta method formula.

Our third methodological contribution is to establish the uniform validity of an *adjusted* version of the delta method. Our adjustment is simply to switch from the standard delta method confidence interval to a projection-based one (as described above) whenever the reduced-form parameters suggest multiplicity of solutions. Our switching rule is nothing else that a pre-test for multiplicity of solutions with a fixed threshold.

For those readers that are not interested in our methodological contributions, we suggest going directly to Section 6.

Relation to Recent Frequentist Literature: From an econometric perspective, our inference problem falls into the general framework of Bugni, Canay, and Shi (2014) and Kaido, Molinari, and Stoye (2014). In their papers, there is a set-identified parameter (e.g., $HD^{1/2} \in \mathbb{R}^n$) but the object of interest is a function of it (e.g., $\text{IRF}_{k,ij} = e_i' C_k(A) H_j \sigma_j \equiv f(HD^{1/2})$). Both papers work under the assumption that $f(\cdot)$ is known, but in our context $f(\cdot)$ needs to be estimated ($f(HD^{1/2}) = e_i' C_k(\widehat{A}_T) H_j \sigma_j$). Exploring this empirically relevant extension of the work of Bugni et al. and Kaido et. al. is outside the scope of our paper. If such an extension were available, however, we could propose an adjusted delta-method that switches to either of their procedures instead of switching to the more conservative projection.

3. MAXIMUM RESPONSE WITH CONTEMPORANEOUS ZERO AND SIGN RESTRICTIONS ON ONE SHOCK

In this section we consider the problem of finding the maximum response to an impulse in the j-th structural shock with m_z 'zero' restrictions and m_s 'sign' restrictions imposed on the response of variables in Y_t. The focus on the maximum and the minimum is an intermediate step to conduct frequentist inference about the coefficients of the impulse-response function.

This section makes two key, easily verifiable assumptions on the sign and zero restrictions allowed in the model. First, we assume that both zero and sign restrictions are imposed only on the response *upon impact* of the single j-th shock under study. Second, we require the number of zero restrictions in the model to be strictly less than $n-1$; implying that the zero restrictions do not point-identify the structural response to the j-th shock. Since most of the empirical applications of set-identified SVARs start with contemporaneous zero/sign restrictions, we provide a complete description of what frequentist inference can deliver in this case.

In terms of results, the two main assumptions in this section suffice to establish three fundamental properties of the identified set for the structural responses at any horizon: a) nonemptiness, b) path connectedness, and c) strict set-identification whenever the maximum or the minimum response are different from zero. In other words, the identified set for the coefficients of the structural impulse-response function is shown to be the interval:

$$\text{IRF}_{k,ij} \in \Big[\underline{v}_{k,ij}(A,\Sigma)\ ,\ \overline{v}_{k,ij}(A,\Sigma) \Big],$$

that becomes a singleton only when both the maximum and the minimum response to the j-th impulse are zero. More important, the two main assumptions in this section allow for simple and transparent arguments to establish the (directional) differentiability of the bounds of the identified set; i.e., the maximum and minimum response.[4]

We now formalize our assumptions and discuss their practical and theoretical implications.

Assumption 1 The restrictions of interest are imposed only on the j-th shock and have the form:

$$e'_{z_i} H_j = 0, \quad \text{for all } i = 1 \dots m_z, \quad \text{(zero restrictions)} \tag{3.1}$$

and

$$e'_{s_i} H_j \geq 0, \quad \text{for all } i = 1 \dots m_s, \quad \text{(sign restrictions)} \tag{3.2}$$

where $e'_{z_i} e_{s_j} = 0$ for all $(i, j) \in \{1, \dots m_z\} \times \{1, \dots m_s\}$.[5]

Assumption 1 has two main theoretical implications. First, we show that if $m_z \leq n-1$ the identified set for the column $H_j \sigma_j$ is never empty (see Lemma B in Appendix A.2). As a consequence, there is no concern that a collection of contemporaneous zero/sign restrictions could be 'incompatible' in neither the population nor the sample.[6] Second, the gradients of all the restrictions imposed are *linearly independent* in the population and in the sample. The linear independence property will play an important role in the characterization of the maximum and minimum response in terms of *Karush-Kuhn-Tucker* conditions and to establish the (directional) differentiability of the maximum response.

The second assumption in this section requires the number of zero restrictions to be strictly smaller than $n - 1$:

Assumption 2 $m_z < n - 1$.

Assumption 1 and 2 imply that the identified set for the column $H_j \sigma_j$ is path connected. Hence, the identified set for any coefficient of the impulse-response function is an interval

[4] We are aware that the nonemptiness and convexity of the identified set can be established under more general assumptions, for example Lemma B.1 in Moon et al. (2013) or Lemma 3.2 in Giacomini and Kitagawa (2014). Both results require a system of linear equalities and inequalities to have an *interior* solution. While this conditions is simple to understand, there does not seem to be a statistical test available for practitioners to verify this assumption. Therefore, we decided to work with the much simpler framework of *contemporaneous* zero and sign restrictions.

[5] The vector e_{s_i} has been defined as the s_i-th column $\mathbb{I}_n$. However, all the results in the paper allow e_{s_i} in (3.2) to be the s_i-th column of either $\mathbb{I}_n$ or -$\mathbb{I}_n$.

[6] The issue of nonemptiness of the identified set in models with sign restrictions on multiple horizons and/or multiple shocks is a more delicate matter.

(see Lemma C and D in Appendix A.3). Furthermore, Assumption 1 and 2 imply that the coefficients of the impulse-response function at any horizon are, in a sense, truly underidentified: the identified set becomes a singleton only if the maximum and minimum response are both zero (see Lemma E and F in Appendix A.4).

Comment: This section does not consider SVARs models in which there are restrictions imposed on multiple structural shocks. For example, the popular SVAR in Mountford and Uhlig (2009) which incorporates restrictions on both a business cycle shock and a monetary shock. Section 8.2 relaxes Assumption 1 to accommodate models with contemporaneous sign restrictions on multiple shocks. We study the 2-dimensional SVAR in Baumeister and Hamilton (2014) and the 3-dimensional SVAR in Kilian and Murphy (2012). Also, Section 8.1 considers models with sign/zero restrictions imposed on noncontemporaneous responses to one structural shock.

Maximum Response to a Shock: We now focus on the upper bound of the identified set. All the results for the lower bound are completely analogous. As we said before, the study of the maximum and the minimum is an intermediate step to conduct frequentist inference about the coefficients of the impulse-response function

Let D denote some diagonal matrix of dimension n controlling the magnitude of the structural innovations. Since we have assumed that

$$\eta_t = H\varepsilon_t, \quad \text{where } H \in \mathbb{R}^{n\times n} \text{ is of full rank,}$$

and we have imposed the orthogonality assumption $\mathbb{E}[\varepsilon_t\varepsilon_t'] = D$, it follows that:

$$\Sigma \equiv \mathbb{E}[\eta_t\eta_t'] = \mathbb{E}[H\varepsilon_t\varepsilon_t'H'] = HDH'.$$

Definition (Maximum Response at horizon k) The maximum response to a σ_j-impulse in the j-th structural shock is defined as the value function, $\overline{v}_{k,ij}(A,\Sigma)$, of the following mathematical program:

$$\overline{v}_{k,ij}(A,\Sigma) \equiv \max_{H_j\sigma_j} e_i'C_k(A)H_j\sigma_j \quad \text{subject to} \;\; HDH' = \Sigma, \;\; \text{and (3.1)} - \text{(3.2)}. \tag{3.3}$$

Thus, the parameter of interest $\overline{v}_{k,ij}(A,\Sigma)$ is simply given by the largest value of the $(t+k)$-structural impulse response to a shock $\epsilon_{j,t}$. The maximizer, however, needs to satisfy the orthogonality assumption and the zero and sign restrictions (3.1)-(3.2).

Two papers have studied mathematical programs that are closely related to (3.3) in the context of SVARs. The classical work of Faust (1998) conducts inference on the maximum value of forecast-error variance decomposition subject to sign restrictions (see p.14, 15 of his paper for a concrete comparison). Aside from the obvious difference in objective functions, the main distinction between our papers is that Faust focuses solely on Bayesian inference on the analogous of $\overline{v}_{k,ij}(A,\Sigma)$ while we focus on frequentist inference. As explained in his paper, Bayesian inference is not a challenge: it suffices to specify a prior for the reduced-form parameters and evaluate $\overline{v}_{k,ij}(\cdot)$ at any draw of the posterior for (A,Σ). Frequentist inference is slightly more complicated, as a minimum amount of regularity on $\overline{v}_{k,ij}(\cdot)$ is required to establish large-sample properties like consistency and asymptotic normality. This paper verifies such conditions for $\overline{v}_{k,ij}(\cdot)$; continuity and differentiability and not just measurability.

In a more recent paper, Giacomini and Kitagawa (2014) consider *robust* Bayes inference for set-identified SVARs. They show that robust posterior bounds can be obtained by solving the mathematical program in (3.3); see Step 3 in p.23 of their paper.[7] The implementation of their procedure is essentially the same as Faust's: they report posterior mean bounds for $\overline{v}_{k,ij}(A,\Sigma)$. A key contribution of their work is to establish *posterior consistency* for their mean bounds (see their section 5.4, p. 25). There are two main differences between our papers. First, we focus on the *large sample* frequentist properties of confidence sets for the parameter $\text{IRF}_{k,ij}$, whereas they focus on the robust Bayes interpretation of their procedure without providing asymptotic coverage statements. Second, we exploit the structure of the SVARs with contemporaneous sign restrictions to show that there is a simple, closed-form solution for (3.3). Giacomini and Kitagawa use the "auglag" function available in an R package "alabama", to solve the nonlinear optimization with equality and inequality constraints using the augmented Lagrangian multiplier method.[8] The similarities and/or differences between inference based on the robust Bayesian procedures in Giacomini and Kitagawa and our frequentist approach when the sample size grows large is an issue we plan to explore in future research.

Another recent paper that has studied the value of stochastic mathematical programs in a different context is Freyberger and Horowitz (2014). They show that the maximum and minimum value of certain linear program can be used to conduct frequentist inference about the value of an unidentified linear functional $L(g)$, where g satisfies the relation $Y = g(X) + U$, in a nonparametric instrumental variable model with instrument W. They propose a point-wise valid bootstrap procedure to conduct inference about the values of $L(g)$. To best of our knowledge, they do not study the uniform validity of their procedures.

[7] They do not allow for multiple shocks but they can accommodate restrictions on noncontemporaneous impulse response functions

[8] The output of their optimization routine could be used to construct our delta method formula. We decided to exploit the characterization of the Karush-Kuhn-Tucker points instead to avoid using external optimization routines.

3.1. *Karush-Kuhn-Tucker characterization of the maximum response*

In this section we show that in SVAR models that satisfy Assumption 1 and 2 there is a simple, closed-form characterization of the maximum response of any variable i to an impulse in $\varepsilon_{j,t}$. More precisely, we show that given any collection $z \in \mathbb{R}^{m\times n}$ of 'active' constraints the maximum response is determined in closed-form (and up to sign) by the Karush-Kuhn-Tucker conditions of the program (3.3). Furthermore, we show that the candidates for the maximum responses are differentiable at every (A, Σ) such that $\overline{v}_{k,ij}(A, \Sigma) \neq 0$.

Lemma 1 (Closed-form Solution for the maximum response) *Let z be a matrix of dimension $n \times m$, $m \leq (n-1)$ collecting the gradients of the active constraints at a solution of the mathematical program* (3.3)*. Let $\overline{v}_{k,ij}(A, \Sigma; z) \neq 0$ denote its value function. Under Assumption 1 and 2:*

a) $\overline{v}_{k,ij}(A, \Sigma; z)$ is given by either plus or minus the norm of the residual of the projection of $\Sigma^{1/2}C_k' e_i$ into the space spanned by the columns of $\Sigma^{1/2}z$; that is:

$$\overline{v}_{k,ij}(A, \Sigma; z) = \pm\Big(e_i' C_k \Sigma^{1/2} M_{\Sigma^{1/2}z} \Sigma^{1/2} C_k' e_i\Big)^{1/2}, \tag{3.4}$$

where

$$M_{\Sigma^{1/2}z} \equiv \mathbb{I}_n - \Sigma^{1/2} z (z'\Sigma z)^{-1} z' \Sigma^{1/2}.$$

b) In addition, the maximizer $(H_j^* \sigma_j) \in \mathbb{R}^n$ associated with the value $\overline{v}_{k,ij}(A, \Sigma; z)$ is unique and given by:

$$x^*(A, \Sigma; z) \equiv (H_j \sigma_j)^* = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}z}\Big)\Sigma^{1/2} C_k' e_i \Big/ \overline{v}_{k,ij}(A, \Sigma; z).$$

Consequently, the sign of $\overline{v}_{k,ij}(A, \Sigma; z)$ depends on which of the two solutions $x^*(A, \Sigma; z)$ satisfy the sign restrictions that are not in z.

c) Furthermore, $\overline{v}_{k,ij}(\cdot)$ is continuously differentiable at every (A, Σ) such that $\overline{v}_{k,ij}(A, \Sigma; z) \neq 0$. The derivative is given by:

$$\dot{\overline{v}}_k(A, \Sigma; z) = \begin{bmatrix} \frac{\partial \overline{v}_k(A,\Sigma)}{\partial vec(A)} \\ \\ \frac{\partial \overline{v}_k(A,\Sigma)}{\partial vec(\Sigma)} \end{bmatrix} = \begin{bmatrix} G_k'(A)(x^*(A, \Sigma; z) \otimes e_i) \\ \\ \frac{1}{2}\overline{v}_k(A, \Sigma; Z)\Big(\Sigma^{-1}x^*(A, \Sigma; z) \otimes \Sigma^{-1}x^*(A, \Sigma; z)\Big) \end{bmatrix},$$

where

$$G_k(A) \equiv \partial vec(C_k(A))/\partial vec(A)'$$

is the $n^2 \times n^2p$ matrix defined in Lütkepohl (1990), p. 118:

$$G_k(A) = \sum_{m=0}^{k-1} J(\mathcal{A}')^{k-1-m} \otimes C_m(A), \tag{3.5}$$

with $J \equiv [\mathbb{I}_n, \mathbf{0}_{n\times n(p-1)}]$ and

$$\mathcal{A} \equiv \begin{pmatrix} A_1 & \dots & A_{p-1} & A_p \\ & \mathbb{I}_{n(p-1)} & & \mathbf{0}_{n(p-1)\times n} \end{pmatrix}.$$

Proof: See Appendix B.1. *Q.E.D.*

Remark 1 The starting point in the proof of Lemma 1 is to show that the program in equation (3.3) is equivalent to *a linear program with sign/zero restrictions and one additional quadratic equality constraint*.[9] See Figure 1 below and its description.

Remark 2 The main argument in the proof of Lemma 1 uses the necessary Karush-Kuhn-Tucker conditions of the optimization problem to characterize the maximizers given a set z of active constraints. An alternative, graphical way to think about the solution to the problem of interest is as follows. Suppose there are only zero restrictions in (3.3). Note that $z'(H_j\sigma_j) = \mathbf{0}_{m\times 1}$ implies that re-parameterized choice variable $\widetilde{x} \equiv \Sigma^{-1/2}H_j\sigma_j$ must lie on the orthogonal space of $\Sigma^{1/2}z$. That is, the selected value of $\widetilde{x}$ should be of the form:

$$\widetilde{x} = M_{\Sigma^{1/2}z}y, \quad M_{\Sigma^{1/2}z} \equiv \left(\mathbb{I}_n - \Sigma^{1/2}z(z'\Sigma z)^{-1}z'\Sigma^{1/2}\right), \; y \in \mathbb{R}^n$$

The quadratic equality constraint also restricts the choice variable $\widetilde{x}$ to satisfy $\widetilde{x}'\widetilde{x} = 1$. See Figure 2 below. Consequently, the problem can be re-written as

$$\max_{y\in\mathbb{R}^n} e_i'C_k\Sigma^{1/2}M_{\Sigma^{1/2}z}y \quad \text{s.t.} \quad y'M_{\Sigma^{1/2}z}y = 1.$$

An application of the Cauchy-Schwartz inequality shows that the positive value in (3.4) gives the maximum response in (3.3).[10] For problems with additional sign and zero restrictions, one needs to check the feasibility of the solution $x^*(A,\Sigma,z)$ to uniquely determine the sign of $\overline{v}_{k,ij}(A,\Sigma)$ in (3.4).

[9]This comes from the fact that $HDH' = \Sigma$ can be replaced by a constraint that involves only the j-th column $HD^{1/2}$; namely, $(H_j\sigma_j)'\Sigma^{-1}(H_j\sigma_j) = 1$. This equivalence will not hold, in general, if there are restrictions on multiple shocks.

[10]Using the fact that $M_{\Sigma^{1/2}z}$ is idempotent and using the assumption that

$$\left(e_i'C_k\Sigma^{1/2}M_{\Sigma^{1/2}z}\Sigma^{1/2}C_k'e_i\right)^{1/2} \neq 0,$$

Figure 1: The mathematical program defining $\overline{v}_{k,ij}(A,\Sigma)$ $(n=3)$ with one zero restriction.

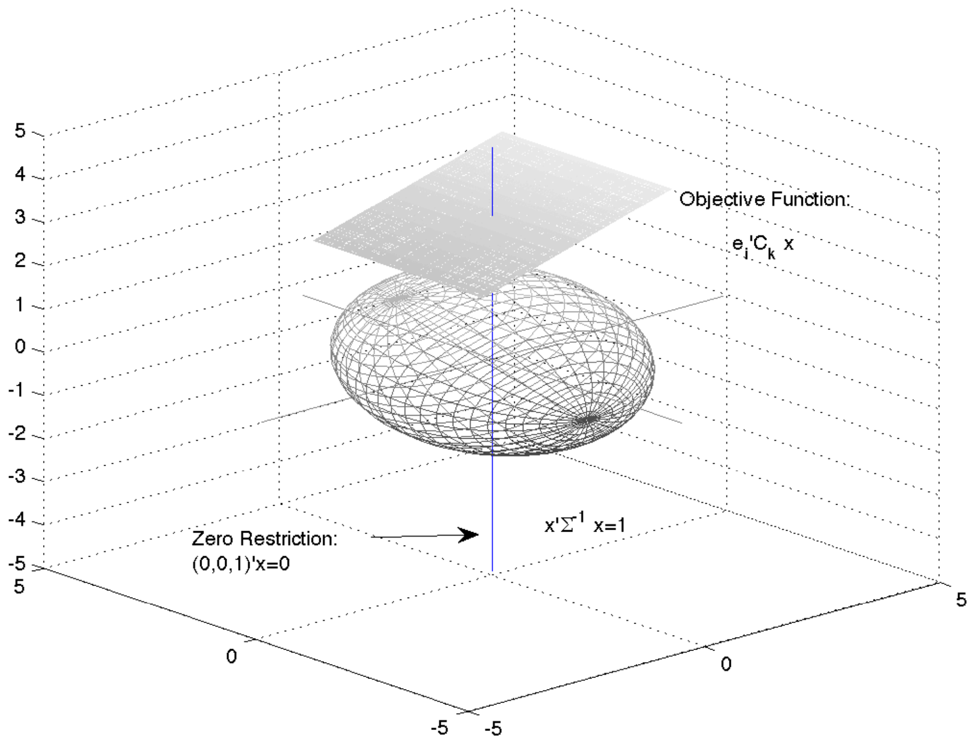


Figure 1 provides a graphical representation of the mathematical program (3.3), where $HDH' = \Sigma$ has been replaced by the 'ellipsoid' constraint $x'\Sigma^{-1}x = 1$, $x \equiv (H_j\sigma_j) \in \mathbb{R}^3$. The objective function corresponds to the hyperplane with normal vector $C_k'e_i \in \mathbb{R}^3$. In this example, there is only one zero restriction (BLUE, SOLID) which requires the contemporaneous impact of the j-th shock on the third variable to be zero. Note that without the zero restriction the maximizer and minimizer will be given by the point at which the hyperplane is tangent to the ellipsoid.

REMARK 3 The derivative of the value function is obtained by relying only upon matrix calculus. However, the *envelope theorem* sheds light on the derivative formula provided by

---

the problem of interest becomes:

$$\max_{y\in\mathbb{R}^n} \left(e_i'C_k\Sigma^{1/2}M_{\Sigma^{1/2}z}\Sigma^{1/2}C_k'e_i\right)^{1/2}\left[\frac{e_i'C_k\Sigma^{1/2}M_{\Sigma^{1/2}z}}{\left(e_i'C_k\Sigma^{1/2}M_{\Sigma^{1/2}z}\Sigma^{1/2}C_k'e_i\right)^{1/2}}\right]M_{\Sigma^{1/2}z}y,$$

s.t. $y'M_{\Sigma^{1/2}z}y = 1$. By the Cauchy-Schwartz inequality this program is bounded above by $(e_i'C_k\Sigma^{1/2}M_{\Sigma^{1/2}z}\Sigma^{1/2}C_k'e_i)^{1/2}$. This value can be achieved by $x^*(A,\Sigma;z)$ in Lemma 1.

Figure 2: Solving for $\overline{v}_{k,ij}(A,\Sigma)$ ($n=3$, $\Sigma=\mathbb{I}_3$) with one zero restriction.

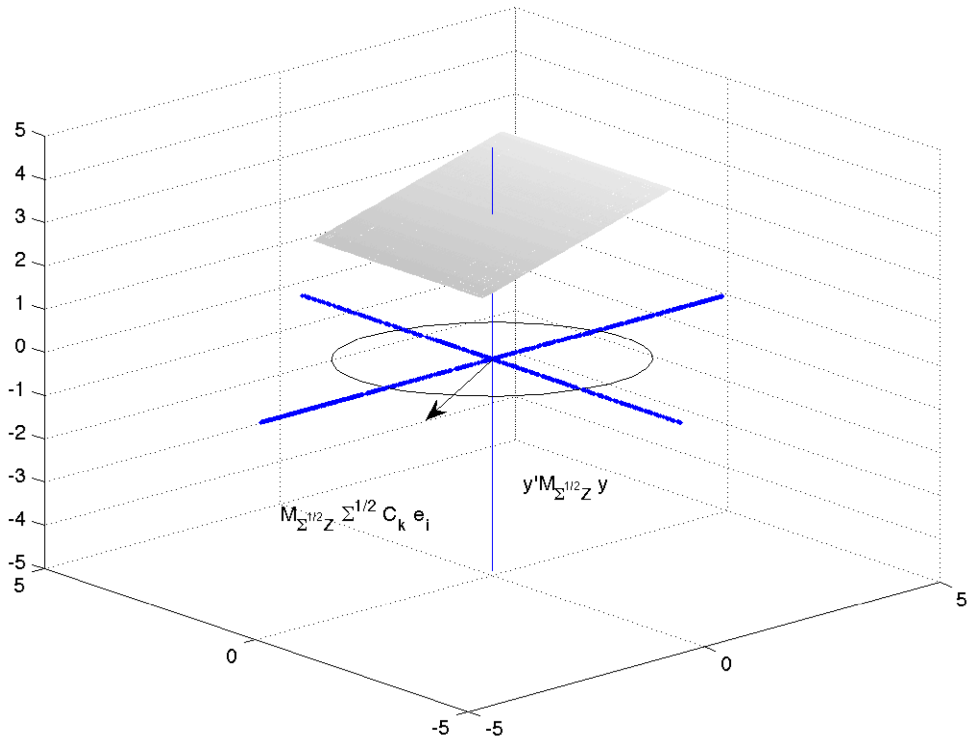


Figure 2 provides a graphical representation of the solution to the mathematical program (3.3) when $\Sigma=\mathbb{I}_3$ and there is only one zero restriction. The solution to the program must lie in the orthogonal complement of Z (BLUE, THIN, SOLID). In this picture the orthogonal complement corresponds to the space spanned by the BLUE, THICK, SOLID lines. This implies that the rotated solution, denoted $\widetilde{x}\equiv\Sigma^{-1/2}x$, must be of the form $M_{\Sigma^{1/2}Z}y$ for some $y\in\mathbb{R}^3$. Hence, the only relevant part of $x'\Sigma^{-1}x=1$ becomes the projected version of it: $y'M_{\Sigma^{1/2}}y=1$, represented by the BLACK, SOLID ellipsoid. One can find the value of this problem by projecting the gradient of the objective function on the orthogonal complement of $\Sigma^{1/2}z$ (arrow) and selecting a direction in the ellipsoid proportional to it. The value function $\overline{v}_{k,ij}(A,\Sigma)$ will be given by the norm of the arrow.

Lemma 1. Let $x\equiv(H_j\sigma_j)$ and let the *Lagrangian function* be given by:

$$\mathcal{L}(x;A,\Sigma,Z)=(x'\otimes e_i')\text{vec}(C_k(A))-w_1\Big((x'\otimes x')\text{vec}(\Sigma^{-1})-1\Big)-w_2'(S'x)-w_3'(Z'x),$$

where w_1 is the Lagrange multiplier corresponding to the quadratic equality restriction, $w_2\in\mathbb{R}^{m_s}$, $w_3\in\mathbb{R}^{m_z}$ are the vector of Lagrange multipliers corresponding to the m_s sign restrictions and the m_z zero restrictions. The envelope theorem provides conditions under which the value function is differentiable with derivative given by the gradient of the

Lagrangian evaluated at the optimal solutions. The envelope theorem suggests that:

$$\dot{\overline{v}}_k(A,\Sigma;Z) = \begin{bmatrix} \frac{\partial \mathcal{L}(x^*,A,\Sigma)}{\partial \text{vec}(A)} \\ \\ \frac{\partial \mathcal{L}(x^*,A,\Sigma)}{\partial \text{vec}(\Sigma)} \end{bmatrix} = \begin{bmatrix} (\partial \text{vec}(C_k(A))/\partial \text{vec(A)})'(x^* \otimes e_i) \\ \\ w_1^*(x^*)\Big(\Sigma^{-1}x^* \otimes \Sigma^{-1}x^*\Big) \end{bmatrix}.$$

Thus, the last step in the proof of Lemma 1 is to show that the envelope theorem formula applies to our problem whenever $\overline{v}_{k,ij}(A,\Sigma;z) \neq 0$.

3.2. *A population algorithm to compute the maximum response with zero and sign restrictions on one shock*

We have showed that $\text{IRF}_{k,ij} \in [\underline{v}_{k,ij}(A,\Sigma)\ ,\ \overline{v}_{k,ij}(A,\Sigma)]$ and we have provided a closed-form expression (up to a sign) for maximum response $\overline{v}_{k,ij}(A,\Sigma)$, given an active collection z of contemporaneous restrictions. We now answer the following question: how does one compute the maximum response $\overline{v}_{k,ij}(A,\Sigma)$ for known population values of (A,Σ)?

We use the result in Lemma 1 to state the solution of the mathematical program (3.3) that includes both contemporaneous zero and sign restrictions. The main result in this section is that a problem with zero and sign restrictions can be solved by 'activating' different combinations of inequality constraints. In other words, the problem in (3.3) can be solved by finding the largest value among the Karush-Kuhn-Tucker points that satisfy a *primal* feasibility constraint.

Additional Notation: In order to state the solution, we introduce additional notation. Define first:

$$Z_0 \equiv [e_{z_1}, e_{z_2}, \ldots e_{z_{m_z}}],$$

as the $\mathbb{R}^n$ matrix that collects all of the m_z zero restrictions. Define also:

$$Z_1 \equiv \Big\{ Z \in \mathbb{R}^{n \times (m_z+1)} \ \Big| \ Z = [Z_0, e_{s_i}], \quad i \in \{1, \ldots m_s\} \Big\},$$

as the collection of all matrices that *activate* one of the m_s inequality restrictions; that is, Z_1 corresponds to the collection of matrices that impose one of the sign-restrictions indexed by s_i as a zero-restriction. Denote by z_1 one of the typical elements in Z_1. More generally, consider the collection:

$$Z_l \equiv \Big\{ Z \in \mathbb{R}^{n \times (m_z+l)} \ \Big| \ Z = [z_0, e_{s_{m_1}}, \ldots e_{s_{m_l}}],\ \{m_i\}_{i=1}^{l} \text{is a subsequence of } \{1, \ldots m_s\} \Big\}.$$

The matrix $z_l \in Z_l$ activates l of the m_s sign-restrictions in the SVAR model.[11] Note that the collection Z_l has $m_s!/(l!(m_s-l)!)$ elements and Z_{m_s} has a unique element in which all the sign restrictions of the model are active. We define the *primal* feasibility of a vector $x \in \mathbb{R}^n$ as the indicator function

$$\mathbf{1}_{m_s}(x) = \mathbf{1}\Big\{ z'_{m_s} x \geq \mathbf{0}_{m_s \times 1} \Big\}, \tag{3.6}$$

where, following convention, $\geq$ is taken component-wise whenever the binary relation is applied to vectors. Hence, $x \in \mathbb{R}^n$ is a feasible point for the mathematical program (3.3) if and only if $\mathbf{1}_{m_s}(x) = 1$.[12]

Proposition 1 (Algorithm to compute the maximum response) *Consider the mathematical program*

$$\overline{v}_{k,ij}(A,\Sigma) \equiv \max_{H_j \sigma_j} e'_i C_k(A) H_j, \sigma_j \quad \textit{subject to} \;\; HDH' = \Sigma,$$

with contemporaneous zero restrictions:

$$e'_{z_i} H_j = 0, \quad \textit{for all } i = 1 \ldots m_z$$

and contemporaneous sign restrictions:

$$e'_{s_i} H_j \geq 0, \quad \textit{for all } i = 1 \ldots m_s,$$

Let

$$\mathcal{Z} \equiv \bigcup_{l=0}^{m_s} Z_l,$$

and for $z \in \mathcal{Z}$, with $\overline{v}_{k,ij}(A,\Sigma;z) \neq 0$ define:

$$f(A,\Sigma;z) \equiv \overline{v}_k(A,\Sigma;z) - (1 - \mathbf{1}_{m_s}(x^*(A,\Sigma;z)))c,$$

where $\overline{v}_k(A,\Sigma;z)$ and $x^*(A,\Sigma;z)$ are defined as in Lemma 1 and c is a positive constant such that $-c < \overline{v}_k(A,\Sigma;z)$ for all $z \in \mathcal{Z}$. If $\overline{v}_{k,ij}(A,\Sigma;z) = 0$, simply set $f(A,\Sigma;z) = 0$. Then:

$$\overline{v}_{k,ij}(A,\Sigma) = \max_{z \in \mathcal{Z}} f(A,\Sigma;z).$$

[11] We restrict $\{m_i\}_{i=1}^{l}$ to be a subsequence of $\{1, 2 \ldots m_s\}$—and not only a subset—to avoid unnecessary repetition in the elements of Z_l. For example, we do not allow elements that activate the same inequality restrictions in different order.

[12] We use the term primal feasibility in contrast with *dual* feasibility, which obtains whenever the Lagrange multipliers for the sign restrictions that are not active are all positive (or negative).

This is, the value function $\overline{v}_{k,ij}(A,\Sigma)$ is obtained by computing the Karush-Kuhn-Tucker points in Lemma 1 for each z, penalizing the value $\overline{v}_k(A,\Sigma;z)$ if unfeasible, and maximizing over all the possible values of z.

PROOF: The intuition behind the proof is as follows. Note that any combination of active sign restrictions z for which $x^*(A,\Sigma;z)$ is both well-defined and feasible must be, by definition, no larger than $\overline{v}_{k,ij}(A,\Sigma)$. Thus, we only have to show that $\max_{z\in\mathcal{Z}} f(A,\Sigma;z) \geq \overline{v}_{k,ij}(A,\Sigma)$. But Lemma 1 showed that the value of the program (3.3) should be of the form $f(A,\Sigma,z)$ for some $z \in \mathcal{Z}$. The proof is formalized in Appendix B.2. *Q.E.D.*

ALGORITHM: The proposition above shows that in order to solve the mathematical problem in (3.3) it is sufficient to apply the following algorithm:

1. *Activate* different combinations of the m_s sign restrictions. Collect the original m_z zero restrictions and the sign restrictions that were activated in the matrix z. Note that the total number of matrices z will be given by:

$$\sum_{k=0}^{m_s} \binom{m_s}{k} = 2^{m_s}.$$

2. Use Lemma 1 to compute the value function $\overline{v}_{k,ij}(A,\Sigma;z)$ for each of the elements $z \in \mathcal{Z}$. Such function corresponds to the value of the problem in which all the restrictions in z are equal to zero and the rest of the sign restrictions are ignored.

3. If $\overline{v}_{k,ij}(A,\Sigma;z) \neq 0$, verify if $x^*(A,\Sigma;z)$ in Lemma 1 satisfies the sign restrictions that were not included in z to determine the sign of $\overline{v}_{k,ij}(A,\Sigma;z)$. That is, verify the *primal feasibility* of the solution $x^*(A,\Sigma;z)$. If the primal feasibility condition is satisfied set

$$f(A,\Sigma;z) = \overline{v}_{k,ij}(A,\Sigma;z),$$

with $\overline{v}_{k,ij}(A,\Sigma;z)$ is defined as in Lemma 1 with a sign that corresponds to that of the feasible solution. If the primal feasibility condition is violated penalize $\overline{v}_{k,ij}(A,\Sigma;z)$ to guarantee that it is never a solution by setting:

$$f(A,\Sigma;z) \equiv \overline{v}_k(A,\Sigma;z) - (1 - \mathbf{1}_{m_s}(x^*(A,\Sigma;z)))c,$$

If $\overline{v}_{k,ij}(A,\Sigma;z) = 0$, set $f(A,\Sigma;z) = 0$.

4. Select the maximum value of $f(A,\Sigma;z)$ over $z \in \mathcal{Z}$; that is, consider the different combinations of active restrictions and select the maximum value $\overline{v}_{k,ij}(A,\Sigma,z)$ over

them.

4. PLUG-IN ESTIMATORS FOR THE MAXIMUM RESPONSE AND ITS LARGE-SAMPLE PROPERTIES

The previous section showed that under Assumption 1 and 2 the coefficients of the impulse response function are interval identified; that is:

$$\text{IRF}_{k,ij} \in \Big[\underline{v}_{k,ij}(A,\Sigma)\ ,\ \overline{v}_{k,ij}(A,\Sigma)\Big]$$

where the bounds of the identified set correspond the minimum and maximum response in the program (3.3). More important, the previous section presented a simple algorithm to solve for the maximum and minimum response in the population.

We now consider estimation and inference for the parameter $\overline{v}_{k,ij}(A,\Sigma)$. The obvious proposal is the plug-in estimator based on the mathematical program:

$$\overline{v}_{k,ij}(\widehat{A},\widehat{\Sigma}) \equiv \max_{H_j\sigma_j} e_i' C_k(\widehat{A}) H_j \sigma_j \quad \text{subject to} \quad HDH' = \widehat{\Sigma}, \quad \text{and } (3.1)-(3.2), \tag{4.1}$$

where $(\widehat{A},\widehat{\Sigma})$ are estimators of the reduced-form parameters (A,Σ). This section establishes the consistency of the plug-in estimator $\overline{v}_{k,ij}(\widehat{A},\widehat{\Sigma})$ and characterizes its asymptotic distribution.

Conceptually, the large sample properties of the plug-in estimator follow from a simple observation: the parameter of interest, $\overline{v}_{k,ij}(A,\Sigma)$, is the value of an optimization problem and, as such, $\overline{v}_{k,ij}(\cdot)$ is a reasonably well-behaved transformation. More precisely, we show that the restrictions imposed by Assumptions 1 and 2 on the structure of the mathematical program described in (3.3) are enough to guarantee the continuity and *directional* differentiability of the transformation $\overline{v}_{k,ij}(\cdot)$. In fact, we show that if (A,Σ) are such that (3.3) has a unique maximizer, the value function will be continuously differentiable. The readers that are not interested in the derivation of the asymptotic distribution of the maximum response can go directly to Section 5.

As one would expect, the consistency of the plug-in estimator is a direct consequence of the continuity of the value function. The characterization of the asymptotic distribution of the plug-in estimator requires slightly more work and—for the sake of clarity—we distinguish between two cases. On the one hand, the asymptotic distribution of the plug-in estimator at points of full differentiability can be characterized by combining the *delta method* with a conventional envelope theorem formula for the derivative of the value function.[13] On the

[13]For references on the delta method see Hogg, Mckean, and Allen (2006), Van der Vaart (2000), Van der Vaart and Wellner (1996), Ferguson (1996). For different versions of the envelope theorem for value functions see Theorem 4.2 in Fiacco and Ishizuka (1990), p. 223 or Theorem 4.24 in Bonnans and Shapiro (2000) p. 280

other hand, the characterization of the asymptotic distribution at points in the parameter space for which the value function is only directionally differentiable relies on the extensions of both the delta method and the envelope theorem used in the seminal work of Shapiro (1991).[14] The asymptotic distributions presented in this section consider both fixed and drifting sequences of parameter values.

4.1. *Continuity and Differentiability of the Value Function*

The large sample properties of the plug-in estimator $\overline{v}_{k,ij}(\widehat{A},\widehat{\Sigma})$ are obtained as corollaries of the following lemma:

Lemma 2 (Continuity and Differentiability of the Value Function) *Let $\overline{v}_{k,ij}(A,\Sigma)$ denote the value function of the program* (3.3) *and let $\mathcal{Z}(A,\Sigma)$ denote the collection of maximizers of $\overline{v}_{k,ij}(A,\Sigma)$. Suppose Assumption 1 and 2 hold and $\overline{v}_{k,ij}(A,\Sigma)\neq 0$. Suppose also that the smallest eigenvalue of Σ is bounded away from zero all over a parameter space Θ. Then:*

1. ***(Continuity)*** $(A_T,\Sigma_T)\to(A,\Sigma) \implies \overline{v}_{k,ij}(A_T,\Sigma_T)\to\overline{v}_{k,ij}(A,\Sigma)$.

2. ***(Everywhere Hadamard Directional Differentiability)*** *Let*

$$\theta=(vec(A)',vech(\Sigma)')'\in\Theta$$

denote the vectorized reduced-form parameters. For a sequence of 'directions' $h_T\in\mathbb{R}^r$, $r=n^2p+n(n+1)/2$, such that $h_T\to h$, consider the parameter $\theta_T\equiv\theta+(h_T/r_T)\in\Theta$, where $\{r_T\}\subseteq\mathbb{R}_+$ and $(1/r_T)\to 0^+$. Then

$$r_T\Big(\overline{v}_{k,ij}(\theta_T)-\overline{v}_{k,ij}(\theta)\Big)\to\max_{z\in\mathcal{Z}(A,\Sigma)}\Big[\dot{\overline{v}}_{k,ij}(A,\Sigma;z)'Vh\Big],$$

with $\dot{\overline{v}}_k(A,\Sigma;z)$ defined as in Lemma 1. The auxiliary block diagonal matrix V is given by:

$$V\equiv\begin{pmatrix}\mathbb{I}_{n^2p} & \mathbf{0}_{n^2p\times n(n+1)/2}\\ \mathbf{0}_{n^2\times n^2p} & D\end{pmatrix}$$

where D is the unique $(n^2\times n(n+1)/2\ )$ matrix such that $vec(\Sigma)=Dvech(\Sigma)$. Consequently, the value function $\overline{v}_{k,ij}$ has Hadamard directional derivative (in direction h at parameter θ) given by:

$$\dot{\overline{v}}_k(h,\theta)\equiv\max_{z\in\mathcal{Z}(A,\Sigma)}\Big[\dot{\overline{v}}_{k,ij}(A,\Sigma;z)'Vh\Big].^{15}$$

[14] A delta method for directionally differentiable functions is also considered in Dümbgen (1993). For a recent comprehensive exposition on the delta method for Hadamard directionally differentiable functions see Fang and Santos (2014).

[15] The map $\overline{v}:\mathbb{R}^r\to\mathbb{R}$ is said to be Hadamard directionally differentiable at $\theta\in\Theta\subseteq\mathbb{R}^r$, tangentially to

3. *(**Uniform Continuous Differentiability at points of unique solutions**) If θ is such that $\mathcal{Z}(\theta)$ is a singleton, then for any sequence $\theta_T \to \theta$:*

$$\sqrt{T}\Big(\overline{v}_{k,ij}(\theta_T + h_T/\sqrt{T}) - \overline{v}_{k,ij}(\theta_T)\Big) \to \dot{\overline{v}}_{k,ij}(\theta; z(\theta))'Vh.$$

PROOF: See Appendix B.3 for details. *Q.E.D.*

CONSISTENCY: The main implication of the continuity of the value function is, of course, the consistency of the plug-in estimator $\overline{v}_{k,ij}(\widehat{A},\widehat{\Sigma})$. The following corollary is a direct application of Lemma 2 and the continuous mapping theorem.

COROLLARY 1 (Consistency) *Let $(\widehat{A},\widehat{\Sigma})$ denote a sequence of estimators of the reduced-form parameters (A,Σ). If Assumptions 1, 2 are satisfied, $\overline{v}_{k,ij}(\widehat{A},\widehat{\Sigma}) \neq 0$ and $(\widehat{A},\widehat{\Sigma}) \xrightarrow{p} (A,\Sigma)$ then*

$$\overline{v}_{k,ij}(\widehat{A},\widehat{\Sigma}) \xrightarrow{p} \overline{v}_{k,ij}(A,\Sigma).$$

The continuity of the value function of the mathematical program (3.3) should be reminiscent of Berge (1963)'s maximum theorem. Indeed, one possibility to establish the continuity of $\overline{v}_{k,ij}(\cdot)$ is to verify the continuity of the choice correspondence. Instead, we used the closed-form solution of $\overline{v}_{k,ij}(\cdot)$ given in proposition 1 to verify continuity directly.

ASYMPTOTIC DISTRIBUTION: We use part 2 and 3 of Lemma 2 to characterize the asymptotic distribution of the plug-in estimator. We start by making a (uniform over Θ) Gaussian weak convergence assumption for the estimators of the reduced-form parameters:

ASSUMPTION 3 Fix a reduced-form parameter $\theta = (\text{vec}(A)', \text{vech}(\Sigma)')' \in \Theta \subseteq \mathbb{R}^r$, $r = n^2p + n(n+1)/2$, and let $\{\theta_T\} \subseteq \Theta$ be any converging sequence with limit θ. We assume that for any such sequence:

$$\sqrt{T}(\widehat{\theta}_T - \theta_T) \xrightarrow{d} \mathbb{G}_\theta,$$

where $\mathbb{G}_\theta \sim \mathcal{N}_{n^2p+n^2}(0, \Omega(\theta))$ and $\Omega(\theta)$ has full rank.[16]

The following corollary combines the weak convergence assumption above with the differentiability results in Lemma 2.

$\mathbb{R}^r$, if there is a continuous (not necessarily linear) map $\dot{\overline{v}}(\cdot,\theta) : \mathbb{R}^r \to \mathbb{R}$ such that:

$$\lim_{T\to\infty} \left| \frac{\overline{v}(\theta + t_T h_T) - \overline{v}(\theta)}{t_T} - \dot{\overline{v}}(h;\theta) \right| = 0$$

for all sequences $\{h_T\} \subseteq \mathbb{R}^r$ and $\{t_T\} \subseteq \mathbb{R}_+$ such that $t_T \to 0^+$, $h_T \to h \in \mathbb{R}^r$ and $\theta + t_T h_T \in \Theta$ for all T. The function $\overline{v}(\cdot)$ is Fully Differentiable at θ if and only if the mapping $\overline{v}(\cdot;\theta)$ is linear. See Fang and Santos (2014) for a recent elegant exposition on directionally differentiable functions. See also Shapiro (1990).

[16]Our high-level assumption is analogous to the weak convergence result derived in Hamilton (1994) p. 301 for stationary VARs. The main difference is that we implicitly assume that the weak convergence holds *uniformly* over compact sets in the parameter space Θ and not only point-wise.

Corollary 2 (Asymptotic Distribution) *Under Assumption 1, 2 and 3:*

1. *(Asymptotic distribution under $\sqrt{T}$-drifting sequences at points of multiplicity of solutions) For any θ_T of the form $\theta_T = \theta + c/\sqrt{T}$, such that $\overline{v}_{k,ij}(\theta) \neq 0$:*

$$\sqrt{T}(\overline{v}_{k,ij}(\widehat{\theta}_T) - \overline{v}_k(\theta_T)) \xrightarrow{d} \max_{z \in \mathcal{Z}(A,\Sigma)} \left[\dot{\overline{v}}_{k,ij}(A,\Sigma;z)' V(\mathbb{G}_\theta + c)\right] - \max_{z \in \mathcal{Z}(A,\Sigma)} \left[\dot{\overline{v}}_{k,ij}(A,\Sigma;z)' V c\right]$$

2. *(Asymptotic distribution under arbitrary sequences at points with unique maximizer) Let $\mathcal{Z}(A,\Sigma)$ denote the collection of maximizers of $\overline{v}_{k,ij}(A,\Sigma)$. If θ is such that $\mathcal{Z}(\theta)$ is a singleton, then for any sequence $\theta_T \to \theta$:*

$$\sqrt{T}(\overline{v}_{k,ij}(\widehat{\theta}_T) - \overline{v}_{k,ij}(\theta_T)) \xrightarrow{d} \dot{\overline{v}}_{k,ij}(A,\Sigma,z(\theta))' V \mathbb{G}_\theta,$$

where $z(\theta)$ is the set of active restrictions at the unique solution θ, $\dot{\overline{v}}_{k,ij}(A,\Sigma;z)$ is defined in Lemma 1, and V is defined in Lemma 2.

Proof: See Appendix B.4 *Q.E.D.*

The first part of this corollary shows that the lack of full differentiability of $\overline{v}_{k,ij}(\cdot)$ at points where the maximizer is not unique implies the lack of regularity (in the sense of Van der Vaart (2000) p. 115) of our plug-in estimator. This lack of regularity means that points in the parameter space for which there is multiplicity of solutions will require special attention in order to guarantee *uniform* inference about $\overline{v}_{k,ij}(\theta)$.[17]

In contrast, the second part of the corollary above shows that the plug-in estimator for the maximum response is regular whenever there is a unique solution to the program in (3.3). Not surprisingly, the asymptotic distribution uses the derivative of the value function we have presented in Lemma 1 and Proposition 1.

5. INFERENCE ABOUT THE COEFFICIENTS OF THE IMPULSE-RESPONSE FUNCTION WITH ZERO AND SIGN RESTRICTIONS ON ONE SHOCK

Section 3 showed that

$$\text{IRF}_{k,ij} \in \left[\underline{v}_{k,ij}(\theta)\ ,\ \overline{v}_{k,ij}(\theta)\right],$$

and we provided a simple algorithm to evaluate the maximum response $\overline{v}_{k,ij}(\theta)$ based on the Karush-Kuhn-Tucker conditions of the population program (3.3). The algorithm (coded in Matlab®) does not require sampling from rotation matrices.

[17] For references on the desirability of uniform inference and uniformity issues in different contexts see Andrews and Guggenberger (2010), Andrews, Cheng, and Guggenberger (2011), Andrews and Cheng (2013), Andrews and Cheng (2012), Romano, Shaikh et al. (2012), McCloskey (2013).

Section 4 showed that under Assumptions 1, 2 and 3 the plug-in estimator $\overline{v}_{k,ij}(\widehat{\theta}_T)$ is asymptotically normal for drifting sequences that converge to a point θ with a unique maximizer. More precisely:

$$\sqrt{T}(\overline{v}_{k,ij}(\widehat{\theta}_T) - \overline{v}_{k,ij}(\theta_T)) \xrightarrow{d} \dot{\overline{v}}_k(A, \Sigma, z(\theta))' V \mathbb{G}_\theta,$$

where $z(\theta)$ is the set of active restrictions at the unique solution, $\dot{\overline{v}}_k(\theta; z)$ is defined in Lemma 1, V is defined in Lemma 2, and $\mathbb{G}_\theta \sim \mathcal{N}_{n^2p+n^2}(0, \Omega(\theta))$.

In this section we use the large-sample properties of the plug-in estimator to conduct inference about the coefficients of the impulse-response function. Our objective is to construct functions $\underline{c}_T(\widehat{\theta}_T, \widehat{\Omega}, \alpha)$ and $\overline{c}_T(\widehat{\theta}_T, \widehat{\Omega}, \alpha)$ such that for all data generating processes that satisfy Assumption 3:

$$\liminf_{T\to\infty} \inf_{\text{IRF}_{k,ij},\theta} \mathbb{P}\Big\{\text{IRF}_{\text{k,ij}} \in \Big[\underline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; c)\ ,\ \overline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; c)\Big]\Big\} \geq 1 - \alpha, \tag{5.1}$$

where $\widehat{\Omega}_T$ is a consistent estimator for the covariance of $\mathbb{G}_\theta$. Since we have characterized the asymptotic distribution of $\overline{v}_{k,ij}(\widehat{\theta}_T)$ using the delta-method, a natural upper confidence bound is the plug-in estimator, $\overline{v}_{k,ij}(\widehat{\theta}_T)$, plus the 1-$\alpha$ quantile of $\dot{\overline{v}}_k(A, \Sigma, z(\theta))' V \mathbb{G}_\theta$. Unfortunately, this *standard* delta-method approach only satisfies the coverage requirement in (5.1) for sequences that converge to a point θ for which $\mathcal{Z}(\theta)$ is a singleton. To account for this issue, we propose an *adjusted* delta method that switches to a conservative *projection*-based confidence bound depending on whether the values of $\widehat{\theta}_T$ suggest multiplicity of solutions or not. To formalize the discussion we consider the following definition for an upper confidence bound $\overline{c}_T(\cdot)$. The lower bound is defined analogously.

Definition (Projection, Delta Method, and 'Asymptotic Distribution' Bootstrap bounds) We conduct inference based on the following bounds for our confidence interval:

1. The *projection*-method upper bound:

$$\overline{c}^p_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) \equiv \max_\theta \overline{v}_{k,ij}(\theta) \quad \text{s.t.} \quad T(\theta - \widehat{\theta}_T)' \widehat{\Omega}_T^{-1} (\theta - \widehat{\theta}_T) \leq \chi^2_r(\alpha). \tag{5.2}$$

where $\chi^2_r(\alpha)$ denotes the $1 - \alpha$ quantile of a central chi-squared random variable with r degrees of freedom.[18]

[18] Note that for a fixed variable i, the value $\overline{v}_k$ depends on θ only through $\theta^* \equiv \Big[(\mathbb{I}_n \otimes e_i')\text{vec}(C_k(A)), \text{vech}(\Sigma)\Big]$. Hence, we could define a θ^*-projection confidence bound:

$$\overline{c}^p_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) \equiv \max_{\theta^*} \overline{v}_{k,ij}(\theta^*)$$

2. The *delta method* upper bound is defined as:

$$\overline{c}^{d}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) \equiv \overline{v}_{k,ij}(\widehat{\theta}_T) + z_{1-\alpha}\widehat{\sigma}_T/\sqrt{T}. \tag{5.3}$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a $\mathcal{N}(0,1)$ random variable and

$$\widehat{\sigma}_T \equiv [\dot{\overline{v}}_{k,ij}(\widehat{\theta}_T, z(\widehat{\theta}_T))'V]\widehat{\Omega}_T[V'\dot{\overline{v}}_{k,ij}(\widehat{\theta}_T, z(\widehat{\theta}_T))].$$

3. The *'Asymptotic Distribution' bootstrap* upper bound based on sampling from the sample analogue of the limiting distribution derived in Lemma 2:

$$\overline{c}^{gb}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) \equiv \overline{v}_{k,ij}(\widehat{\theta}_T) + b_T(\widehat{\theta}_T, \widehat{\Omega}_T)/\sqrt{T}, \tag{5.4}$$

where $b_T(\widehat{\Omega}_T, \widehat{\Omega}_T)$ is the $1-\alpha$ quantile of the random variable

$$-\sqrt{T}(\overline{v}_{k,ij}(\theta^*_T) - \overline{v}_{k,ij}(\widehat{\theta}_T)),$$

and $\theta^*_T = \widehat{\theta}_T + (\widehat{\Omega}_T/T)^{-1/2}Z_r,\ Z_r \sim \mathcal{N}_r(\mathbf{0}, \mathbb{I}_r),\ Z_r \perp \mathbb{G}_\theta$.[19]

Remark 4 We make four observations concerning the definition above. First, a confidence interval based on the projection bounds satisfies the coverage requirement in (5.1), but is conservative relative to the delta-method and the AD bootstrap. Second, the confidence interval based on the delta-method bounds is guaranteed to satisfy the coverage requirement in (5.1) for any given sequence that converges to a point of full continuous differentiability and non-zero variance ($\widehat{\sigma}_T \to \sigma(\theta) > 0$). Third, under sequences $\theta_T \to \theta$ for which the set of maximizers is a singleton and the limiting variance is larger than zero, the AD bootstrap bound is equivalent to the delta-method bound. *Finally, we note that the projection-method confidence interval implemented in this paper is a computationally more tractable version of the projection method proposed in Moon et al. (2013), as we do not need to invert any hypothesis test.* We also present a comparison between our procedures in Appendix C to argue that the delta method is more efficient than MSG projection.

subject to

$$T(\theta^* - \widehat{\theta}^*_T)'\widehat{\Omega}^{-1}_T(\theta^*)(\theta^* - \widehat{\theta}^*_T) \leq \chi^2_{r^*}(\alpha),$$

where

$$\widehat{\Omega}_T(\theta^*) = \begin{bmatrix} (\mathbb{I}_n \otimes e_i')G_k(\widehat{A}) & \mathbf{0}_{n\times n(n+1)/2} \\ \mathbf{0}_{n(n+1)/2\times n^2p} & \mathbb{I}_{n(n+1)/2} \end{bmatrix} \widehat{\Omega}_T \begin{bmatrix} (\mathbb{I}_n \otimes e_i')G_k(\widehat{A}) & \mathbf{0}_{n\times n(n+1)/2} \\ \mathbf{0}_{n(n+1)/2\times n^2p} & \mathbb{I}_{n(n+1)/2} \end{bmatrix}',$$

and $r^* = n + (n(n+1)/2) < r$. This procedure is computationally more intensive than the standard projection.

[19]Our delta method approach relies on Lütkepohl (1990)'s asymptotic approximation for the distribution of reduced-form MA coefficients. Kilian (1998) has proposed a *bias-corrected bootstrap* to improve the accuracy of the confidence sets for the MA coefficients. Our inference procedure uses a 'Asymptotic Distribution' bootstrap (as defined in 3 above) to keep a close connection with our delta method results. However, it might be possible to adapt Kilian's 'bootstrap-after-bootstrap' idea to our set-up.

In order to combine the efficiency of the delta method/AD bootstrap with the robustness of the projection we suggest an approach analogous to the one described in Andrews and Cheng (2012). For tuning parameters $\varepsilon_1 > 0, \varepsilon_2 > 0$, and $s \in \{d, gb\}$ define:

$$\overline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; s) \equiv \begin{cases} \overline{c}^s_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) & \text{if} \quad \overline{v}_{k,ij}(\widehat{\theta}_T) \notin [-\varepsilon_1, \varepsilon_1] \\ & \text{and} \quad \overline{v}_{k,ij}(\widehat{\theta}_T) \geq f(\widehat{\theta}_T; z) + \varepsilon_2 \\ & \qquad \forall z \in \mathcal{Z} \ \text{ s.t. } \ z(\widehat{\theta}_T) \nsubseteq z \\ & \\ \overline{c}^p_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) & \text{in} \qquad \text{other case,} \end{cases} \tag{5.5}$$

We show that the upper bound in (5.5) satisfies the coverage requirement in (5.1), while preserving the delta method/AD bootstrap efficiency for a large class of sequences $\{\theta_T\}$. The lower bound is defined analogously

$$\underline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; s) \equiv \begin{cases} \underline{c}^s_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) & \text{if} \quad \underline{v}_{k,ij}(\widehat{\theta}_T) \notin [-\varepsilon_1, \varepsilon_1] \\ & \text{and} \quad \underline{v}_{k,ij}(\widehat{\theta}_T) \leq f(\widehat{\theta}_T; z) - \varepsilon_2 \\ & \qquad \forall z \in \mathcal{Z} \ \text{ s.t. } \ z(\widehat{\theta}_T) \nsubseteq z \\ & \\ \underline{c}^p_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) & \text{in} \qquad \text{other case,} \end{cases} \tag{5.6}$$

We now establish the uniform validity of the adjusted delta method:

Proposition 2 *Let $\overline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; s)$ be defined as in (5.5) and let $\underline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; s)$ be defined as in (5.6). Suppose Assumptions 1, 2 and 3 hold. Then, for any sequence $(IRF^T_{k,ij}, \theta_T)$ such that $\theta_T \to \theta$:*

$$\liminf_{T\to\infty} \mathbb{P}_{IRF^T_{k,ij},\theta_T} \Big\{ \underline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; s) \ \leq IRF^T_{k,ij} \leq \ \overline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha; s) \Big\} \geq 1 - \alpha$$

Furthermore, the coverage requirement is satisfied with equality for any sequence $\theta_T \to \theta$ for which the bounds are fully differentiable and the coefficients of the impulse-response converge (at $\sqrt{T}$-rate) to either the maximum or the minimum response. For example, $\overline{v}_{k,ij}(\theta) > \varepsilon_1$, $\overline{v}_{k,ij}(\theta) > f(\theta; z) + \varepsilon_2$ for all z such that $z(\theta) \nsubseteq z$ and $\sqrt{T}(\overline{v}_{k,ij}(\theta_T) - IRF^T_{k,ij}) \to 0$.

Proof: See Appendix B.5 *Q.E.D.*

The procedure described in (5.5) requires tuning parameters $(\varepsilon_1, \varepsilon_2)$. In our empirical application we select three configurations for the parameters $\varepsilon_1 = \varepsilon_2$ and report the Monte-Carlo rejection probability using a Gaussian VAR model with true parameter $\widehat{\theta}_T$ and T observations. We note that for all configurations the Monte-Carlo coverage is at most 15% below the desired 95% nominal coverage.

6. SIMPLE ILLUSTRATIVE EXAMPLE: UNCONVENTIONAL MONETARY SHOCKS

In conventional descriptions of monetary policy and its transmission mechanism, the short-term nominal interest rate is usually assumed to be the central bank's policy instrument. Following any adjustment by the monetary authority, the market participants—such as households, businesses and investors—use available information to form expectations about the future level of longer-term real interest rates relevant for their consumption and investment decisions.

The recent *Great Recession* has forced the Federal Reserve (and other central banks around the globe) to consider alternative mechanisms to affect market beliefs about the future of real interest rates. Two examples of such unconventional policies are the Federal Open Market Committee's *forward guidance* announcements and the Federal Reserve's *large-scale asset purchases program*. Broadly speaking, through forward guidance "the Federal Open Market Committee provides an indication to households, businesses, and investors about the stance of monetary policy expected to prevail in the future".[20] In a similar fashion, the asset purchase program of the Federal Reserve intends to "put downward pressure on yields of a wide range of longer-term securities, support mortgage markets, and promote a stronger economic recovery.[21]

Based on the description above, we suggest two identifying features of an expansionary *unconventional monetary policy shock* (UMP). First, a sign restriction (i.e., a UMP shock decreases long-term interest rates upon impact) and, second, a zero restriction (i.e., a UMP shock does not affect short-term rates upon impact). We also require the UMP expansionary shock to have a positive contemporaneous effect over prices and economic activity. These sign restrictions can be justified by the DSGE model calibrated in the work of Bhattarai, Eggertsson, and Gafarov (2014). We conduct frequentist inference about the responses of different macroeconomic variables to a UMP shock using our adjusted delta method. The main objective of this section is to compare our adjusted delta method with the AD bootstrap, the projection method and Bayesian Credible sets. Our application suggests that the projection confidence set can be much wider than the adjusted delta method confidence interval. Also, we note that the adjusted delta method and the AD bootstrap give confidence intervals that are very close to each other.

Our suggested set-identification strategy is not new. Baumeister and Benati (2013) study an analogous 'spread' monetary shock that leaves the short-term nominal rate unchanged, but affects the spread between the ten-year Treasury-bond yield and the policy rate. They consider a Bayesian SVAR with time varying parameters and stochastic volatility [as in Cogley and Sargent (2005), Primiceri (2005)] combined with demand and supply structural shocks that satisfy zero/sign restrictions [as in Rubio-Ramirez, Waggoner, and Zha (2010)].

[20]`Link:` http://www.federalreserve.gov/faqs/money_19277.htm
[21]`Link:`http://www.federalreserve.gov/faqs/what-are-the-federal-reserves-large-scale-asset-purchases.htm

Their main result is that the long-term yield spread exerts a powerful effect on both output growth and inflation. All their inference is Bayesian. We use the results in our section 5 to conduct frequentist inference on a simpler SVAR model that does not consider time varying parameters, stochastic volatility, and restrictions on other nonmonetary shocks. To justify the assumption of time invariant parameters, we only use data before September 2008. One advantage of our frequentist approach is that we do not need to specify any prior beliefs about the parameters of the model.

TABLE I

RESPONSE TO AN EXPANSIONARY UNCONVENTIONAL MONETARY POLICY SHOCK: RESTRICTIONS

Series	Acronym	UMP
Consumer Price Index	CPI	+
Industrial Production	IP	+
2-year Treasury Bond rate	2yTB	−
Fed Funds Rate	FF	0

DESCRIPTION: Restrictions on contemporaneous responses to a UMP shock. ‘0’ stands for a zero restriction, ‘−’ stands for a negative sign restriction and ‘+’ for positive sign restriction.

MONETARY SVAR: We consider a simple 4-variable model that includes the Consumer Price Index (CPI_t), the Industrial Production Index (IP_t), the Federal Funds rate (FF_t) and the 2-year Treasury Bond rate ($2yTB_t$).[22] The time span of the monthly series is July 1979 to August 2008 ($T = 342$). We take a logarithm transformation of CPI_t, IP_t and then work with first differences for all variables. We set the number of lags equal to 11 in the reduced form VAR model using the Bayesian Information Criterion. To keep our exposition as simple as possible, we ignore potential co-integration issues between short-term and long-term interest rates. We then argue that our simple SVAR does a remarkable job tracing the out-of-sample dynamics of the IP, CPI, 2yTB, and FF.

INFERENCE ON IMPULSE RESPONSE FUNCTIONS: Figure 3 reports maximum and minimum cumulative responses (bounds of the light gray area) corresponding to a σ-UMP shock under the contemporaneous sign/zero restrictions in Table 1.[23] Thus, the shaded light gray area is an estimate of the identified set—the collection of all the impulse response functions that are compatible with the sign/zero restrictions in Table 1. The point estimators $\overline{v}_{k,ij}(\widehat{A}, \widehat{\Sigma})$ are computed using our Matlab® code for the algorithm described at the end of Proposition

[22] All these variables are sourced from the data set of Gertler and Karadi (2014). We thank Peter Karadi for making their data set available to us.

[23] We scale the identified set to make the minimum response in the 2-year Treasury bond rate coincide with the decline in this variable from the en of July 2010 (.62) to the end of August 2010 (.52). This period corresponds to the last part of QE1 and the announcement of QE2.

1. *The estimation of the maximum and minimum response does not require any sampling from unit vectors in* $\mathbb{R}^4$.

a) Unadjusted delta method: For each horizon k, the solid blue line represents the delta-method confidence set for the (k,i,j)-th coefficient of the cumulative impulse-response function where no adjustment has been made to guarantee uniform inference. The unadjusted 95% delta-method confidence sets are simply given by:

$$\left[\underline{v}_{k,ij}(\widehat{\theta}_T) - 1.64\,\widehat{\underline{\sigma}}_T/\sqrt{T},\ \overline{v}_{k,ij}(\widehat{\theta}_T) + 1.64\,\widehat{\overline{\sigma}}_T/\sqrt{T}\right],$$

where the estimates of the asymptotic variances of the maximum and minimum response correspond to:

$$\widehat{\overline{\sigma}}_T \equiv [\dot{\overline{v}}_k(\widehat{\theta}_T, z(\widehat{\theta}_T))'V]\widehat{\Omega}_T[V'\dot{\overline{v}}_k(\widehat{\theta}_T, z(\widehat{\theta}_T))], \quad \widehat{\underline{\sigma}}_T \equiv [\dot{\underline{v}}_k(\widehat{\theta}_T, z(\widehat{\theta}_T))'V]\widehat{\Omega}_T[V'\dot{\underline{v}}_k(\widehat{\theta}_T, z(\widehat{\theta}_T))].$$

with gradients $\dot{\overline{v}}_{k,ij}$ and $\dot{\underline{v}}_{k,ij}$ based on Lemma 1.

b) Unadjusted AD bootstrap: Our delta-method uses a linear approximation for the non-linear functions $\overline{v}_{k,ij}(\cdot)$ and $\underline{v}_{k,ij}(\cdot)$. The AD bootstrap proposed in (5.3) is one informal way to assess the accuracy of the delta-method approximation while keeping the assumption that $\sqrt{T}(\widehat{\theta}_T - \theta)$ is close to a $\mathcal{N}_{n^2p+n^2}(\mathbf{0}, \Omega)$. As described in section 5, to implement our AD bootstrap we sample θ_T^* from the multivariate normal distribution $\mathcal{N}_{n^2p+n^2}(\widehat{\theta}_T, \widehat{\Omega}_T/\sqrt{T})$ and we evaluate $\overline{v}_{k,ij}(\theta_T^*)$ and $\underline{v}_{k,ij}(\theta_T^*)$ at each of the sampled values. The maximum and minimum response are computed using our Matlab® code. We consider 1,000 draws to implement the AD bootstrap. Note that neither the validity nor the implementation of the AD bootstrap relies on the Gaussianity of VAR innovations.

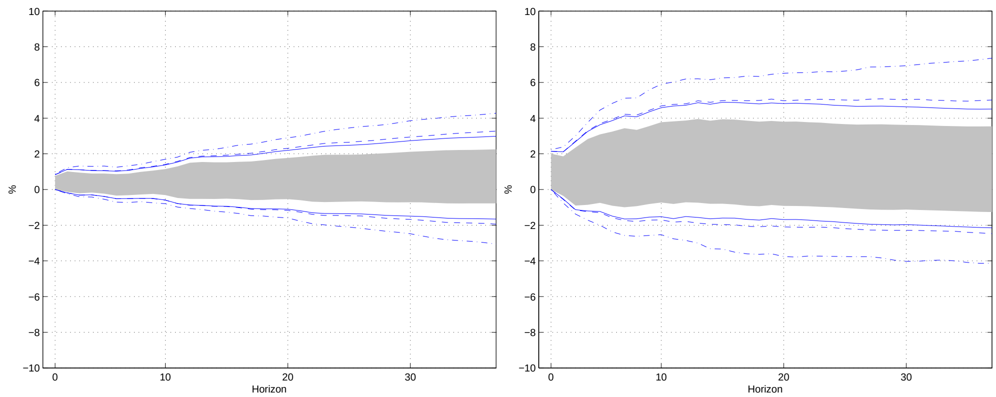
The dashed blue lines in Figure 3 represent the AD bootstrap confidence sets. For points in the parameter space where the value function is fully differentiable, the delta method and the AD bootstrap are asymptotically equivalent. We attribute differences between the delta-method and the AD bootstrap to one of two reasons: 1) the value function is only directional differentiable or 2) the value function is fully differentiable but our delta-method linear approximation is not very accurate for the sample size at hand. We note that the delta method and the AD bootstrap lie close to each other in Figure 3.

c) Projection Method: The wider, dotted-dashed, blue lines in Figure 3 are the projection confidence bands. The projection method is uniformly valid (without requiring any further adjustments), but it will give conservative confidence sets. For this particular application, the projection confidence sets seem very large relative to the unadjusted delta method, even when the value function is very different from zero and there are no close competing maximizers/minimizers.

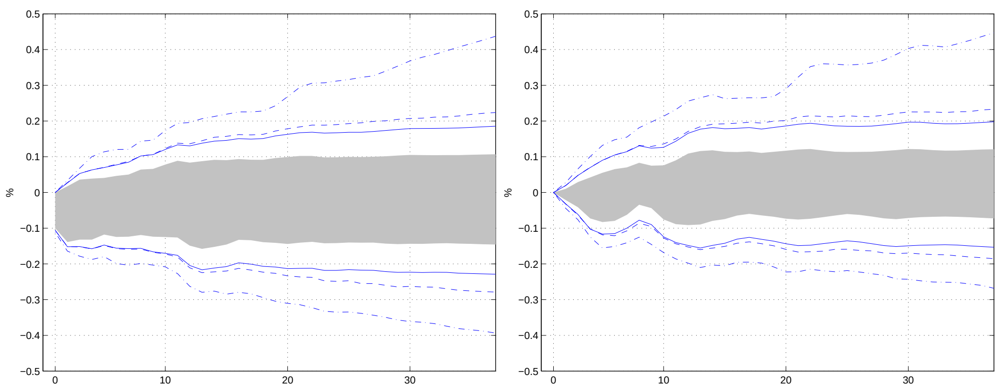
Figure 3: Frequentist Inference for the Coefficients of Cumulative Impulse-Responses.

Delta Method, AD bootstrap, and Projection 95% Confidence Sets

(Contemporaneous Restrictions on only One Shock)



(a) Consumer Price Index (b) Industrial Production Index



(c) 2-year Treasury Bond Rate (d) Federal Funds Rate

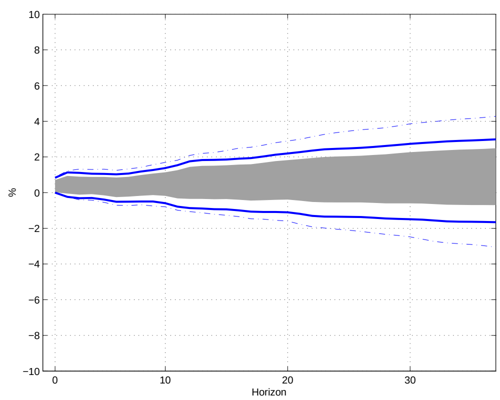
(Shaded, Light Gray Area) For each horizon k and each variable i the gray area represents the interval $\left[\underline{v}_{k,ij}(\widehat{\theta}) \, , \, \overline{v}_{k,ij}(\widehat{\theta})\right]$ for cumulative responses, where the j-th structural shock corresponds to the Unconventional Monetary Policy shock described in Table 1. (Solid, Blue Line) 95% Frequentist Confidence Set Based on the Delta-Method Formula with no adjustment. (Dashed, Blue Line) 95% Frequentist Confidence Set Based on the AD bootstrap with no adjustment. (Dotted-Dashed, Blue Line) 95% Projection Confidence Set.

Comparison with Bayesian Procedures: Figure 4 reports our 95% adjusted-delta method confidence set and a 95% Bayesian Credible Set based on the recent paper of

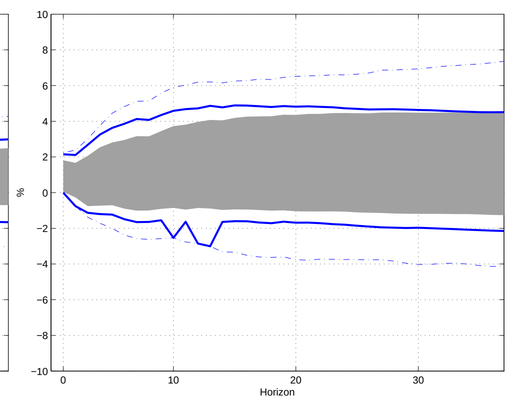
Arias, Rubio-Ramirez, and Waggoner (2014).[24] As anticipated—given the results in the paper of Moon and Schorfheide (2012)—frequentist confidence sets are larger than their Bayesian counterparts.

Figure 4: 95% Adjusted Delta Method Confidence Set and 95% Bayesian Credible Set for the Coefficients of Cumulative Impulse-Response Functions.

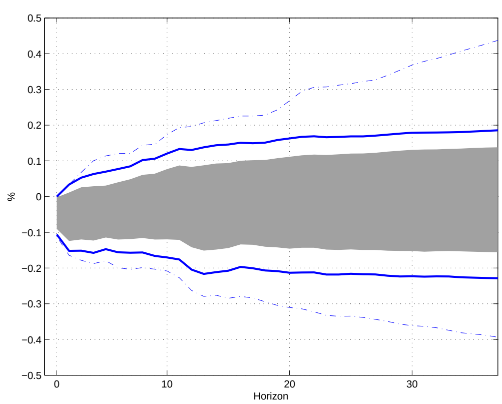
(Contemporaneous Restrictions on only One Shock)



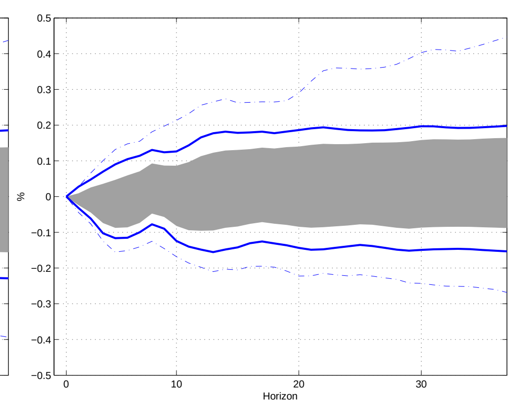
(a) Consumer Price Index



(b) Industrial Production Index



(c) 2-year Treasury Bond Rate



(d) Federal Funds Rate

(Shaded, Gray Area) 95% Bayesian Credible Set using Algorithm 1 described in Section 3.2 of Arias et al. (2014). (Blue, Solid Line) 95% adjusted delta method confidence set with $\varepsilon_1 = \varepsilon_2$ and the following tuning parameter for each of the time series: $[.15, .74, .02, .02]$. (Blue, Dotted Line) 95% projection confidence set based on the procedure described on Section 5.

[24] We are very grateful to Jonas Arias for sharing his extremely clean and clear Matlab code to replicate Panel (a) of Figure 4 in Arias et al. (2014). The Bayesian Credible set reported in Figure 4 is obtained after a slight modification of Arias et al. (2014) original code.

As explained in Section 5 our adjusted delta method requires two user-selected inputs $(\varepsilon_1, \varepsilon_2) > 0$ for each time series. These tuning parameters guarantee the uniform validity of our procedure. In practice, the tuning parameters provide a switching rule between the unadjusted delta method and the projection confidence bounds. For example, the tuning parameters used in the construction of Figure 4-b) are such that at horizon $k = 10$ there is a switch from the unadjusted delta method (blue solid line in Figure 3-b) to the projection method (dashed-dotted blue line in Figure 3-b). The adjusted delta method (blue, solid line in Figure 4-b) traces this switching.

From a theoretical perspective, any choice of $(\varepsilon_1, \varepsilon_2) > 0$ implies that the adjusted delta-method confidence set has the right coverage. From a practical perspective, the choice of (ϵ_1, ϵ_2) will affect the finite-sample coverage frequency of the adjusted-delta method. The tuning parameters used in the construction of Figure 4 are $[.15, .74, .02, .02]$ for each of the time series (and we set $\epsilon_1 = \epsilon_2$ for all of the time series). We note that under these tuning parameters, the adjusted delta method coincides with its unadjusted version for most of the horizons.

We conclude this section with a brief economic interpretation of our delta method confidence interval. As we mentioned before, the identification strategy in this paper was motivated by two mechanisms used by the Federal Reserve to affect market beliefs during the Great Recession: forward guidance announcements and the large-scale asset purchase program. We will focus on one particular episode of the Great Recession illustrating the role of forward guidance. In August 2010 the Federal Open Market Committee announced: "*The Committee will keep constant the Federal Reserve's holdings of securities at their current level by reinvesting principal payments from agency debt and agency mortgage-backed securities in longer-term Treasury securities.*" This announcement was an important prelude for the second part of the Quantitative Easing program (QE2) (see p. 244 in Krishnamurthy and Vissing-Jorgensen (2011) for a detailed discussion). In addition, this announcement generated a drop in the intraday yield for two- and ten- year treasury bond. In fact, from the end of July 2010 to the end of August 2010 the 2yrTB rate fell by 10 basis points.

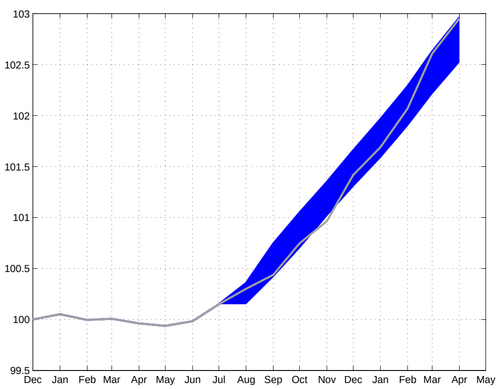
Figure 5 uses our delta method confidence interval to construct confidence bands for the evolution of the levels of the 4 variables in the monetary SVAR. We fix all the variables at their level on July 2010 and we trace their evolution according to the cumulative responses in Figure 3. Our thought exercise is the following. We have already used the data before September 2008 to conduct inference on the cumulative responses to a σ-unconventional monetary policy shock and linear trends for CPI and IP. We use these cumulative responses to get a rough idea of the evolution of the levels of the variables following the forward guidance announcement of the federal reserve in August 2010. For CPI and IP we ignore uncertainty coming from the estimation of the trend. Our exercise assumes that between

August 2010 and April 2011, there was not other shock affecting the variables in our SVAR.

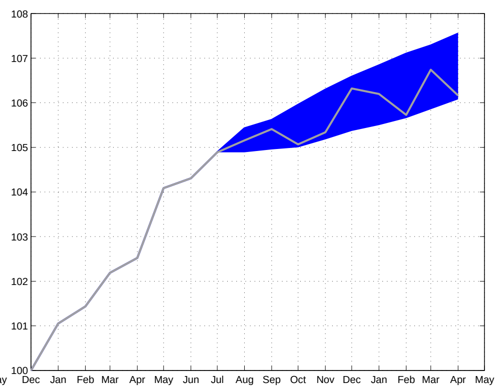
We note that the observed dynamics for CPI and IP after the August 2010 announcement seem to fall within the bounds motivated by our adjusted delta method confidence interval. We find this somewhat surprising, as we have only used the data before September 2008 to estimate our SVAR model. We conclude that the effects of the QE2 announcement seem to be in line with the σ-UMP monetary shock that we set-identified based on the pre-crisis data.

Figure 5: Evolution of CPI, IP, 2yTB, FF after the August 2010 announcement

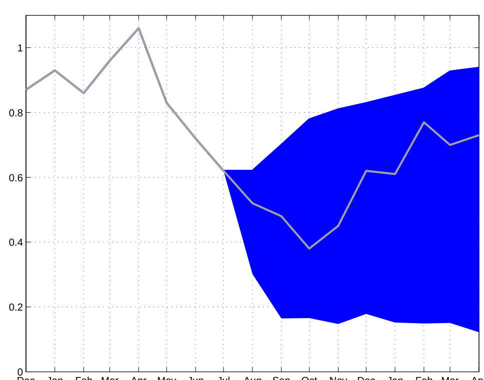
(Contemporaneous Restrictions on only One Shock; Out of sample analysis)



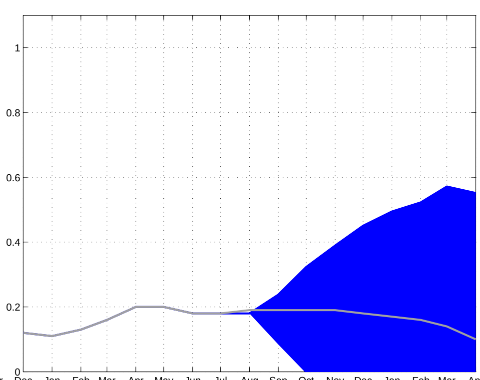
(a) Consumer Price Index



(b) Industrial Production Index



(c) 2-year Treasury Bond Rate



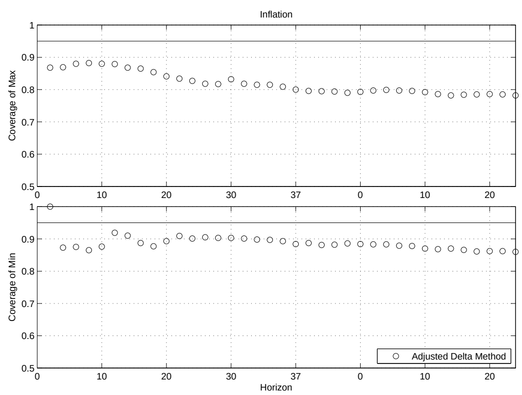
(d) Federal Funds Rate

(Shaded, Blue Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 95% adjusted delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Gray, Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to April 2011. Both the CPI index and the IP index were normalized to have a starting value of 100.

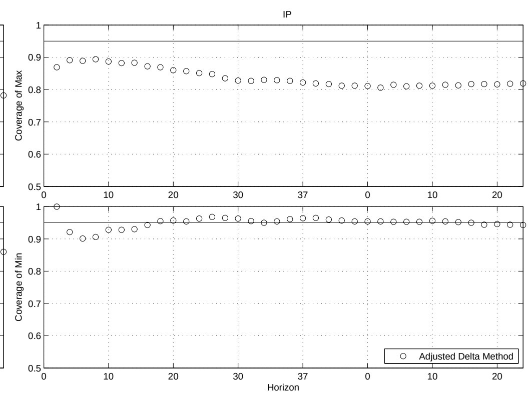
6.1. ***Monte-Carlo Exercise***

Figure 6: Monte-Carlo Coverage Frequency of 95% Adjusted Delta-method Confidence Sets

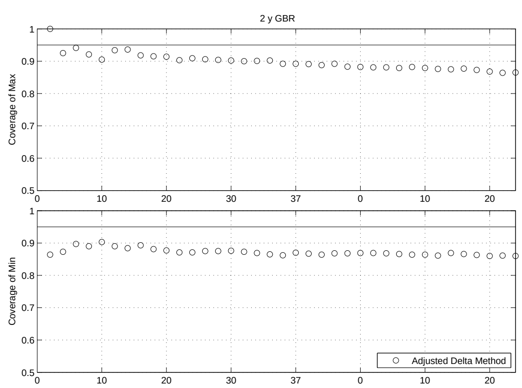
(Contemporaneous Restrictions on only One Shock)



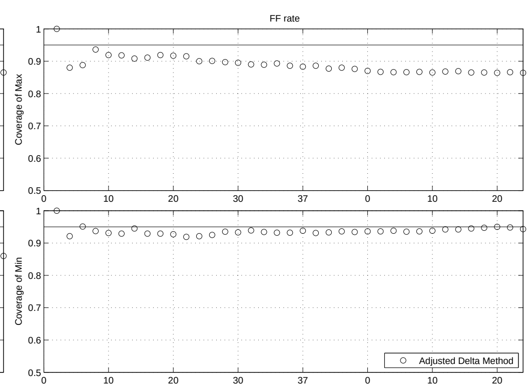
(a) Consumer Price Index



(b) Industrial Production



(c) 2-year Treasury Bond Rate



(d) Federal Funds Rate

(Upper Panel) Monte-Carlo coverage frequency of the maximum response. That is, we provide a MC estimate of $\mathbb{P}_\theta\left[\underline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha) \leq \overline{v}_{k,ij}(\theta) \leq \overline{c}_T(\widehat{\theta}_T, \widehat{\Omega}_T, \alpha)\right]$. (Lower Panel) The lower panel reports the Monte-Carlo coverage frequency of the minimum response.

As a robustness check, we suggest practitioners to report the coverage frequency for $\overline{v}_{k,ij}(\cdot)$ and $\underline{v}_{k,ij}(\cdot)$ in a particular Monte-Carlo (MC) exercise using different values of the tuning parameters. Our suggested MC exercise takes the estimated reduced-form parameters as the true parameters and generates data using serially uncorrelated, reduced-form Gaussian errors with covariance matrix $\widehat{\Sigma}$. In this example, we consider three values for the tuning parameter $\varepsilon = \varepsilon_1 = \varepsilon_2$: $[0, 0, 0, 0]$, $[.15, .74, .02, .02]$, and $[.3, 1, .03, .04]$.

We report the coverage frequency associated to the maximum and minimum responses. That is, we count the number of Monte-Carlo simulations for which the population's maximum (minimum) response is contained in the 95% adjusted delta method confidence set. We consider a sample size of $T = 350$ (close to the 342 observations in our example). The number of Monte-Carlo simulations is 1,000. The results of the study for $T = 350$ and $[.15, .74, .02, .02]$ are provided in Figure 6. The rest of the figures are in Appendix E.1.

In our MC exercise, the coverage frequency seems to be at least 80% for all confidence sets, even for those based on the unadjusted delta method ($\varepsilon = [0, 0, 0, 0]$) (see Figure 12).

7. CONCLUSION

This paper proposed an intuitive, computationally simple, adjusted delta method confidence interval for set-identified coefficients of the impulse-response function in a Structural Vector Autoregression. We established the asymptotic validity of our inference procedure in models that impose zero and sign restrictions only on the *contemporaneous* responses to one structural shock, but our approach is defined more generally (see Section 8). We also presented Monte-Carlo evidence supporting the validity of our adjusted delta method.

We showed that the construction of our frequentist confidence interval in the basic model does not require random sampling from the space of rotation matrices or unit vectors. Instead, we treated the bounds of the identified set for the coefficients of impulse-responses as the *maximum and minimum value* of a linear mathematical program with one quadratic constraint. We used Karush-Kuhn-Tucker conditions to provide formulas for these values and also for their derivatives. We adjusted the delta method formula to take into account possible violations of differentiability.

Preview of Extensions: In Section 8 we extend the delta method formula to accommodate models with contemporaneous and noncontemporaneous restrictions on one structural shock. Once again, the construction of our frequentist confidence interval exploits the characterization of the Karush-Kuhn-Tucker points in Lemma 1. We provide an empirical application to a 4-dimensional SVAR suggesting that the delta method confidence bounds can be more informative than the projection bounds. In addition, Monte-Carlo evidence suggests that the adjusted delta has reasonable coverage even in this case.

Section 8 also extends our delta method approach to accommodate models with contemporaneous restrictions on multiple shocks. First, we introduce a *relaxed*, adjusted delta method procedure for these class of models. Our *relaxed* procedure solves for the value function of an auxiliary linear program with only one quadratic constraint. The auxiliary value function bounds from above the maximum response of interest $\overline{v}_{k,ij}(\cdot)$. The bound is exact whenever there are only contemporaneous restrictions on two structural shocks. The relaxed, adjusted delta method allows us to exploit the characterization results in Lemma 1 and to use our Matlab® algorithm to solve for the bounds and construct the delta method confidence set.

We present an application to a 2-dimensional SVAR in Baumeister and Hamilton (2014) and also to the 3-dimensional SVAR in Kilian and Murphy (2012). We also present Monte-Carlo evidence suggesting that our adjusted delta method procedure has reasonable coverage.

Finally, Section 8 presents a general delta method procedure for models with contemporaneous restrictions on multiple shocks. Unfortunately, with more than two shocks, we do not have tractable closed-form expression for the Karush-Kuhn-Tucker points of the program defining $\overline{v}_{k,ij}(\cdot)$. However, we use Matlab®'s `fmincon` to solve for the maximum/minimum response and Lagrange multipliers. Once again, we present a Monte-Carlo exercise to verify the coverage of our proposal.

REFERENCES

Andrews, D. and X. Cheng (2013): "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure," *Econometric Theory*, 173(1), 36–56.

Andrews, D., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Cowles Foundation*, Discussion Paper No. 1813.

Andrews, D. W. and X. Cheng (2012): "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, D. W. and P. Guggenberger (2010): "Asymptotic size and a problem with subsampling and with the m out of n bootstrap," *Econometric Theory*, 26, 426.

Andrews, D. W. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Arias, J., J. F. Rubio-Ramirez, and D. F. Waggoner (2014): "Inference Based on SVAR Identified with Sign and Zero Restrictions: Theory and Applications," *Working paper, Duke University.*

Baumeister, C. and L. Benati (2013): "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound," *International Journal of Central Banking*, 9, 165–212.

Baumeister, C. and J. Hamilton (2014): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Working paper, Bank of Canada.*

Berge, C. (1963): *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*, Dover publications, 1st english ed.

Bhattarai, S., G. B. Eggertsson, and B. Gafarov (2014): "Time Consistency and the Duration of Government Debt: A Signalling Theory of Quantitative Easing," *Working paper, Penn State University.*

Billingsley, P. (1995): *Probability and Measure*, John Wiley & Sons, New York, 3rd ed.

Bonnans, J. F. and A. Shapiro (2000): *Perturbation analysis of optimization problems*, Springer.

Bugni, F., I. Canay, and X. Shi (2014): "Inference for functions of partially identified parameters in moment inequality models," *Working paper, Duke University.*

Canova, F. and G. D. Nicoló (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

Cogley, T. and T. J. Sargent (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic dynamics*, 8, 262–302.

Dümbgen, L. (1993): "On nondifferentiable functions and the bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Fang, Z. and A. Santos (2014): "Inference on Directionally Differentiable Functions," *Working paper, University of California at San Diego*.

Faust, J. (1998): "The Robustness of Identified VAR Conclusions about Money," in *Carnegie-Rochester Conference Series on Public Policy*, Elsevier, vol. 49, 207–244.

Ferguson, T. S. (1996): *A course in large sample theory*, vol. 49, Chapman & Hall London.

Fiacco, A. V. and Y. Ishizuka (1990): "Sensitivity and stability analysis for nonlinear programming," *Annals of Operations Research*, 27, 215–235.

Freyberger, J. and J. Horowitz (2014): "Identification and shape restrictions in nonparametric instrumental variables estimation," *Working paper, University of Wisconsin*.

Gertler, M. and P. Karadi (2014): "Monetary Policy Surprises, Credit Costs and Economic Activity," *Working paper, New York University*.

Giacomini, R. and T. Kitagawa (2014): "Robust-Bayes Inference for Non-Identified SVARs," *Working Paper, University College London*.

Hamilton, J. (1994): *Time series analysis*, vol. 2, Cambridge Univ Press.

Hogg, R. V., J. W. Mckean, and C. D. Allen (2006): *Introduction To Mathematical Statistics*, Pearson Education India.

Imbens, G. W. and C. F. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Inoue, A. and L. Kilian (2013): "Inference on impulse response functions in Structural VAR models," *Journal of Econometrics*, 177, 1–13.

Kaido, H., F. Molinari, and J. Stoye (2014): "Inference for Projections of Identified Sets," *Working Paper, Boston University*.

Kasy, M. and J. L. Montiel Olea (2014): "Uniformity and the delta method," *Working Paper, Harvard University*.

Kilian, L. (1998): "Small-sample confidence intervals for impulse response functions," *Review of economics and statistics*, 80, 218–230.

Kilian, L. and D. P. Murphy (2012): "Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models," *Journal of the European Economic Association*, 10, 1166–1188.

Krishnamurthy, A. and A. Vissing-Jorgensen (2011): "The effects of quantitative easing on interest rates: channels and implications for policy," *National Bureau of Economic Research*.

Lütkepohl, H. (1990): "Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models," *The Review of Economics and Statistics*, 116–125.

Mangasarian, O. L. (1993): *Nonlinear programming*, vol. 10, siam.

McCloskey, A. (2013): "Bonferroni-based size-correction for nonstandard testing problems," *Working paper, Brown University.*

Moon, H. R. and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80, 755–782.

Moon, H. R., F. Schorfheide, and E. Granziera (2013): "Inference for VARs Identified with Sign Restrictions," *University of Southern California.*

Mountford, A. and H. Uhlig (2009): "What are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics*, 24, 960–992.

Primiceri, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72, 821–852.

Romano, J. P., A. M. Shaikh, et al. (2012): "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 40, 2798–2822.

Rubio-Ramirez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural vector autoregressions: Theory of identification and algorithms for inference," *The Review of Economic Studies*, 77, 665–696.

Shapiro, A. (1990): "On concepts of directional differentiability," *Journal of optimization theory and applications*, 66, 477–487.

——— (1991): "Asymptotic analysis of stochastic programs," *Annals of Operations Research*, 30, 169–186.

Sims, C. A. (1980): "Macroeconomics and reality," *Econometrica*, 48, 1–48.

——— (1986): "Are forecasting models usable for policy analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2–16.

Stoye, J. (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, 77, 1299–1315.

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381–419.

Van der Vaart, A. (2000): *Asymptotic statistics*, 3, Cambridge Univ Pr.

Van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes.*, Springer, New York.

Extensions

8. EXTENSIONS

The results in this paper can be extended in at least three directions:[25]

1. *Forecast-error variance decomposition with contemporaneous zero/sign restrictions:* Frequentist inference about the maximum share of the forecast-error variance attributed to a single shock is one possible extension to the results in this paper. Conceptually, we need to replace the linear objective function in the mathematical program (3.3) by a quadratic objective function, with Lemma 1 adjusted accordingly.

2. *Zero/Sign restrictions on the responses to one shock at future horizons:* Extending our delta method approach to cases where zero/sign restrictions are imposed on future responses is another possibility. Establishing uniformity results for this case requires extra work (especially if we want to state assumptions that practitioners can easily verify). However, the extension of the delta method formula is straightforward. Section 8.1 presents the delta method formula for this case and presents an application to a 4-dimensional monetary SVAR. We do not present theoretical results, but we include a simple Monte-Carlo exercise to examine the finite-sample properties of adjusted delta method inference when there is sampling uncertainty in the sign restrictions.

3. *Sign Restrictions on the contemporaneous responses to multiple shocks:* In models with sign restrictions on only one shock, the Karush-Kuhn Tucker conditions that characterize the maximum and minimum response are very tractable. This means that the bounds of the identified set for the coefficients of the impulse-response function can easily be evaluated. This is no longer the case in models with sign restrictions on multiple shocks.
   To deal with this issue, we follow two approaches. First, we introduce a *relaxed*, adjusted delta-method to conduct frequentist inference in SVARs with contemporaneous sign restrictions on multiple structural shocks. Our procedure solves for the value function of an auxiliary linear program with only one quadratic constraint. The auxiliary value function *bounds* from above the maximum response of interest $\overline{v}_{k,ij}(\cdot)$. The bound is exact whenever there are only contemporaneous restrictions on two structural shocks. As an application, we use the 2-dimensional SVAR in Baumeister and Hamilton (2014).
   Second, we use Matlab®'s `fmincon` to solve for the maximum/minimum response and Lagrange multipliers and plug them into the delta method formula. As an application, we use the 3-dimensional SVAR model in Kilian and Murphy (2012).

[25] We are very grateful to Lutz Kilian for motivating us to write this section.

8.1. ***Zero/Sign restrictions on the responses to one shock at future horizons***

Zero/sign restrictions on the future response to a j-th shock have the form:

$$e'_{z_i} C_{k_{z_i}}(A) H_j = 0, \quad \text{for all } i = 1 \ldots m_z, \quad \text{(zero restrictions)}, \tag{8.1}$$

and

$$e'_{s_i} C_{k_{s_i}}(A) H_j \geq 0, \quad \text{for all } i = 1 \ldots m_s, \quad \text{(sign restrictions)}. \tag{8.2}$$

Once again, we assume that $m_z < n-1$. To guarantee that the algorithm in Proposition 1 can be extended to compute the maximum response under (8.1) and (8.2), we consider the following assumption:

Assumption 1-Ext: The reduced-form parameter A is such that any matrix of dimension $n \times n$ formed by all the gradients of the zero restrictions

$$[C_{k_{z_1}}(A)' e_{z_1}, C_{k_{z_2}}(A)' e_{z_2}, \ldots C_{k_{z_{m_z}}}(A)' e_{z_{m_z}}] \in \mathbb{R}^{n \times m_z}$$

and $n - m_z$ gradients of the m_s sign restrictions has full rank n. We refer to this assumption as n-Full Column Rank Assumption.

This assumption will be satisfied in the sample with probability one (as long as the researcher does not repeat zero or sign restrictions). Assumption 1-Ext and a unique solution of the mathematical program subject to (8.1)-(8.2) are enough to guarantee the differentiability of the value function.

The Lagrangian of the mathematical program with zero/sign restrictions, $\mathcal{L}(x; A, \Sigma)$, is given by:

$$\underbrace{(x' \otimes e'_i)\text{vec}(C_k)}_{\text{Objective Function}} - \underbrace{w_1\Big(x'\Sigma^{-1}x - 1\Big)}_{\text{ellipsoid constraint}} - \underbrace{\sum_{i=1}^{m_s} w_{2,i}(x' \otimes e'_{s_i})\text{vec}(C_{k_{s_i}})}_{\text{sign restrictions}} - \underbrace{\sum_{i=1}^{m_Z} w_{3,i}(x' \otimes e'_{z_i})\text{vec}(C_{k_{z_i}})}_{\text{zero restrictions}},$$

Suppose that there is a unique solution x^* and let $\{l_1, l_2, \ldots l_L\}$ denote the set indices of the L active constraints at x^*. The envelope theorem suggests that:

$$\dot{v}_k(A, \Sigma; z) = \begin{bmatrix} \frac{\partial \mathcal{L}(x^*, A, \Sigma)}{\partial \text{vec}(A)} \\ \\ \frac{\partial \mathcal{L}(x^*, A, \Sigma)}{\partial \text{vec}(\Sigma)} \end{bmatrix} = \begin{bmatrix} G_k(A)(x^* \otimes e_i) - \sum_{i=1}^{L} w_{l_i} G_{k_{l_i}}(A)(x^* \otimes e_{l_i}) \\ \\ w_1^* \Big(\Sigma^{-1}x^* \otimes \Sigma^{-1}x^*\Big) \end{bmatrix},$$

where $w_1^* = (1/2)\overline{v}_{k,ij}(A,\Sigma)$ and the vector of Lagrange multipliers for the restrictions that are active are given by

$$w = (w_{l_1}, w_{l_2}, \ldots w_{l_L})' = (z'\Sigma z)^{-1} z'\Sigma C_k' e_i.$$

The formula above extends part c) of Lemma 1. Since we have allowed for zero and sign restrictions to depend on the slope parameter, there is a new complication to guarantee the (uniform) differentiability of the value function: 'local' violations of the n-full rank column assumption. To avoid such violations one could require all the eigenvalues of all matrices of dimension $n \times n$ in Assumption 1-Ext to be bounded away from zero (for all A in the parameter space). At the moment, we will simply assume these violations away.

We now present our adjusted delta method confidence bounds for a 4-dimensional Monetary SVAR example based on Moon et al. (2013).

Moon et al. (2013) fit a 2-lag SVAR to per capita real GDP (in deviations from linear trend), inflation, the federal funds rate, and real money balances. They use quarterly U.S. data from 1964:I to 2005:I ($T = 165$). All the data is obtained from the file `SignAndZeroFreqBonfIDset.zip` in the following link link. Figure 7 uses the following sign/zero restrictions:

TABLE II

Response to a Contractionary Monetary Policy Shock: Restrictions

Series	Contemporaneous	Noncontemporaneous
Inflation	0	NA
Federal Funds Rate	+	+ (k=1)
Per Capita Real GDP	0	NA
Real Money Balances	−	NA

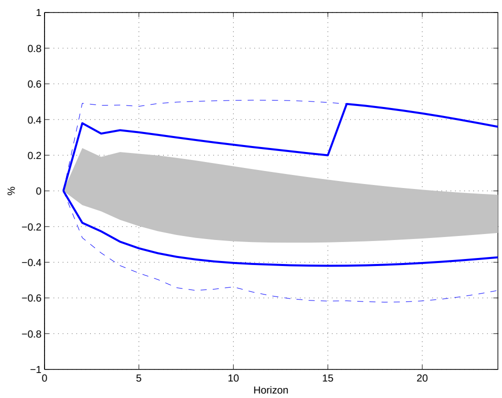
Description: Restrictions on responses to a contractionary monetary policy shock. '0' stands for a zero restriction, '−' stands for a negative sign restriction and '+' for positive sign restriction. 'NA' means there are no sign/zero restrictions imposed.

Figure 7 shows that the projection method confidence set can be less informative than the adjusted delta method. For example, the adjusted delta method confidence set for real GDP suggests that economic activity will fall below its trend during all of the second year after the contractionary shock. The projection confidence set, however, contains positive and negative values.

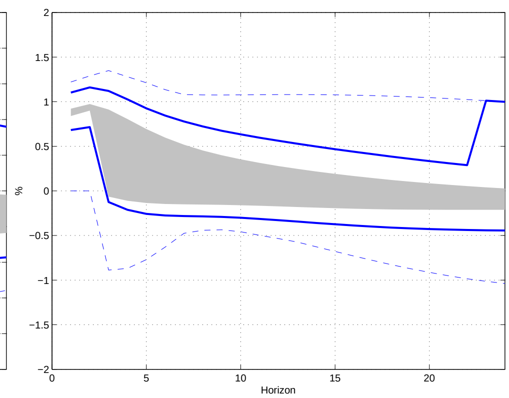
Figure 8 below reports the MC coverage frequency for inflation, federal Funds, per capita real GDP, and real money balances. For the first three time series, the MC coverage frequency is above 80% for all horizons. For the last time series, the MC coverage frequency is lower for the minimum response—but still above 70%. Figure 14 the appendix we also report the

Figure 7: 95% Adjusted Delta Method and 95% Projection Confidence Set for the Coefficients of Impulse-Response Functions.

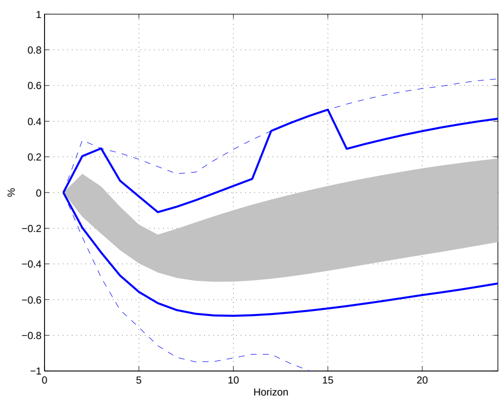
(Contemporaneous and Noncontemporaneous Restrictions on only One Shock)



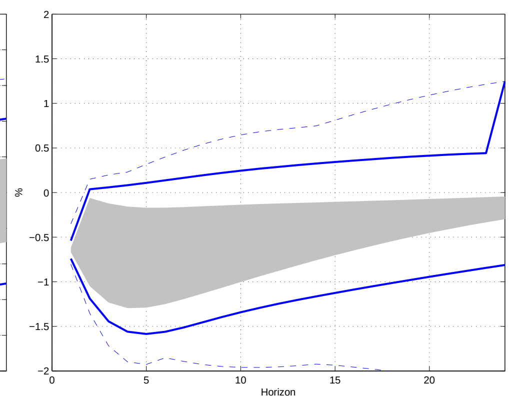
(a) Inflation



(b) Federal Funds Rate



(c) Per Capita Real GDP



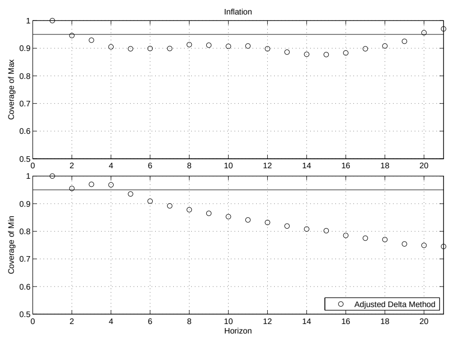
(d) Real Money Balances

(Shaded, Gray Area) For each horizon k and each variable i the gray area represents the interval $\left[\underline{v}_{k,ij}(\widehat{\theta})\ ,\ \overline{v}_{k,ij}(\widehat{\theta})\right]$, where the j-th structural shock corresponds to the contractionary monetary policy shock described in Table II. (Blue, Solid Line) 95% adjusted delta method with $\epsilon_1 = \epsilon_2$ and the following tuning parameter for each of the time series: [.05, .05, .05, .05]. (Blue, Dashed Line) 95% projection method confidence set.

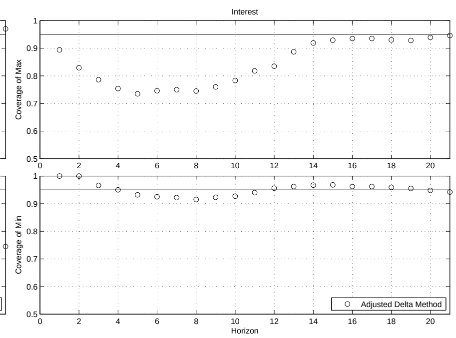
coverage probability of the adjusted delta-method for an unfeasible size of $T = 1,650$ to verify convergence of the adjusted delta method to the 95% nominal coverage.

Figure 8: Monte-Carlo Coverage Frequency of 95% Adjusted Delta-method Confidence Sets

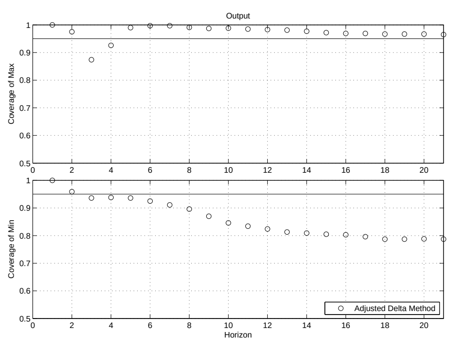
(Contemporaneous and Noncontemporaneous Restrictions on only One Shock)



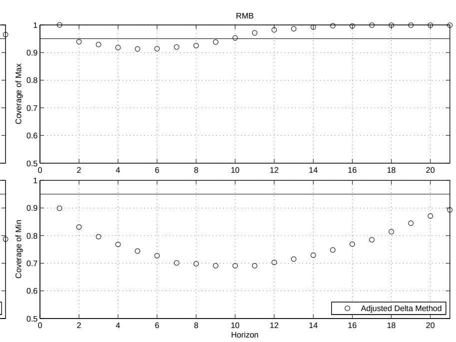
(a) Inflation



(b) Federal Funds Rate



(c) Per Capita Real GDP



(d) Real Money Balances

(Upper Panel) Monte-Carlo coverage frequency of the maximum response. (Lower Panel) The lower panel reports the Monte-Carlo coverage frequency of the minimum response.

8.2. *Sign restrictions on the contemporaneous responses to multiple shocks*

The results in Section 5 exploited three key properties of models that impose zero/sign restrictions on the responses to only one structural shock:

a) The value function $\overline{v}_{k,ij}(\cdot)$ is defined by a linear program with one additional quadratic constraint and, consequently, the Karush-Kuhn Tucker points admit simple, closed-form solutions.

b) It is easy to characterize the points in the parameter space for which the value function is not fully differentiable. As we have showed, full differentiability is only compromised if the value function is close to zero or if there are multiple maximizers/minimizers.

c) Given a collection of contemporaneous zero/sign restrictions, the identified set for the coefficients of the impulse response function is never empty and it is always an interval.

Unfortunately, these three properties are lost in models with contemporaneous restrictions on more than two shocks.

To deal with this issue, we introduces a *relaxed*, adjusted delta-method to conduct frequentist inference in SVARs with contemporaneous sign restrictions on multiple structural shocks. Our procedure solves for the value function of an auxiliary linear program with only one quadratic constraint. The auxiliary value function bounds from above the maximum response of interest $\overline{v}_{k,ij}(\cdot)$. The bound is exact whenever there are only contemporaneous restrictions on two structural shocks. We now present our main results.

Restrictions on only two shocks: Let $S_1, S_2 \in \mathbb{R}^{n\times n}$ collect the vectors associated with a full set of sign restrictions on the first two columns of $HD^{1/2}$. For example, if $n = 3$ and $He_1 \in \mathbb{R}^3$ is restricted to satisfy the sign restrictions $[+, +, +]$, then $S_1 = \mathbb{I}_3$. Likewise, if $He_2 \in \mathbb{R}^3$ is restricted to satisfy the sign restrictions $[+, -, +]$ then

$$S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrices of contemporaneous sign restrictions are diagonal and, hence, symmetric.

Lemma 3 *Let $S_1, S_2 \in \mathbb{R}^{n\times n}$ denote matrices of contemporaneous sign restrictions for the first and second structural shock. Consider the mathematical program:*

$$\overline{v}_{k,i1}(A,\Sigma) \equiv \max_{H\in\mathbb{R}^{n\times n}} e_i' C_k(A) H e_1 \sigma_1 \quad \text{s.t. } HDH' = \Sigma, \text{ and } S_1 He_1 \geq \boldsymbol{0}_{n\times 1}, \;\; S_2 He_2 \geq \boldsymbol{0}_{n\times 1},$$

and for some diagonal matrix S_k (with 1 and -1 diagonal elements) consider the auxiliary program:

$$\overline{v}_{k,i1}(A,\Sigma; S_k) \equiv \max_x e_i' C_k(A) x \quad \text{subject to } x'\Sigma^{-1} x = 1, \text{ and } S_1 x \geq \boldsymbol{0}_{n\times 1}, S_k \Sigma^{-1} x \geq \boldsymbol{0}_{n\times 1}.$$

Then

$$\overline{v}_{k,i1}(A,\Sigma) = \max_{S_k\in\mathcal{S}} \overline{v}_{k,i1}(A,\Sigma; S_k)$$

where $\mathcal{S}$ is the set of all matrices such that $S_2 S_k$ is neither positive nor negative definite.

Proof: See Appendix D.1. *Q.E.D.*

The proof of Lemma 3 relies on Tucker's Theorem of the Alternative, p. 34 in Mangasarian (1993). The objective function defining $\overline{v}_{k,i1}(\cdot)$ depends only on the first column of H, denoted $h_1 \in \mathbb{R}^n$. Not every h_1 satisfying $h_1'\Sigma^{-1}h_1 = 1$, and $S_1 h_1 \geq \mathbf{0}_{n\times 1}$ is feasible, though. To *support* a candidate h_1 we need to find $h_2 \neq 0 \in \mathbb{R}^n$ such that $h_1'\Sigma^{-1}h_2 = 0$ and $S_2 h_2 \geq \mathbf{0}$. Lemma 3 uses Tucker's Theorem of the alternative to show that only certain candidates can be supported; namely, those of the form $h_1 = \Sigma S_k \alpha, \alpha \geq \mathbf{0}, \alpha \neq 0$ for some S_k such that $S_2 S_k$ is neither positive nor negative definite.

Application to the 2-Dimensional SVAR of Baumeister and Hamilton using Tucker's Theorem of the Alternative: We apply Lemma 3 to conduct frequentist inference on the demand-supply SVAR model with two variables in Baumeister and Hamilton (2014). We report confidence bands for the coefficients of the cumulative impulse-response function based on the adjusted delta-method.

Baumeister and Hamilton (2014) fit an 8-lag VAR to U.S. data on growth rates of real labor compensation and total employment over $t = 1970 : \text{Q1} - 2014 : \text{Q2}$. They impose contemporaneous sign restrictions to set-identify an expansionary demand shock and a contractionary supply shock. We model the sign restrictions on the demand shock (first shock) by:

$$S_1 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Likewise, the sign restrictions that set-identify the supply shock (second shock) are given by:

$$S_2 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

For the sign restrictions considered in this model, the identified set for the coefficients of the impulse-response is an interval.[26] Consider first the maximum and minimum responses to a demand shock. The set $\mathcal{S}$ in Lemma 3 is given by $\mathcal{S} = \left\{\mathbb{I}_2, -\mathbb{I}_2\right\}$. Thus, to compute the $\overline{v}_{k,i1}(\cdot)$ it suffices to consider two programs:

$$\max_{h_1 \in \mathbb{R}^2} e_i' C_k h_1 \text{ s.t. } h_1'\Sigma^{-1}h_1 = 1, \;\; S_1 h_1 \geq \mathbf{0}_{2\times 1}, \text{ and } \Sigma^{-1}h_1 \geq \mathbf{0}_{2\times 1}, \tag{8.3}$$

and

$$\max_{h_1 \in \mathbb{R}^2} e_i' C_k h_1 \text{ s.t. } h_1'\Sigma^{-1}h_1 = 1, \;\; S_1 h_1 \geq \mathbf{0}_{2\times 1}, \text{ and } -\Sigma^{-1}h_1 \geq \mathbf{0}_{2\times 1}. \tag{8.4}$$

[26]This need not hold in general. See appendix TBA for an example of sign restrictions that generates an identified set that is not an interval. We also provide an example of sign restrictions for which the identified set is empty.

It is not difficult to see that the choice set in (8.4) is always empty. Therefore, $\overline{v}_{k,i1}(A,\Sigma)$ is given by the value function of the program (8.3), which can be solved using the algorithm in our Proposition 1.

The envelope theorem suggests that:

$$\dot{\overline{v}}_k(A,\Sigma;z) = \begin{bmatrix} \frac{\partial \mathcal{L}(x^*,A,\Sigma)}{\partial \mathrm{vec}(A)} \\ \\ \frac{\partial \mathcal{L}(x^*,A,\Sigma)}{\partial \mathrm{vec}(\Sigma)} \end{bmatrix} = \begin{bmatrix} G_k(A)(x^* \otimes e_i) \\ \\ w_1^* \Big(\Sigma^{-1}x^* \otimes \Sigma^{-1}x^* \Big) + \sum_{l=1}^{2} w_{2,l}^*(x'\Sigma^{-1} \otimes e_l'\Sigma^{-1}) \end{bmatrix}, \tag{8.5}$$

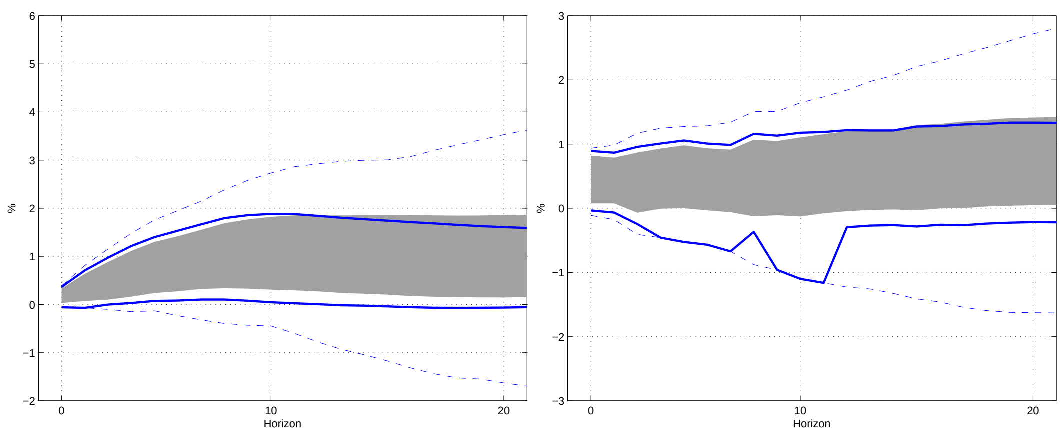
where $w_1^* = (1/2)\overline{v}_{k,ij}(A,\Sigma)$ and the vector of Lagrange multipliers for the restrictions that are active are given by:

$$w = (z'\Sigma z)^{-1} z'\Sigma C_k' e_i.$$

Figure 9: 95% Frequentist Confidence Bounds vs. 95% Bayesian Credible Set for

the Coefficients of Cumulative Impulse-Response Functions.

(Contemporaneous Restrictions on Two Shocks)



(a) Employment (b) Wage

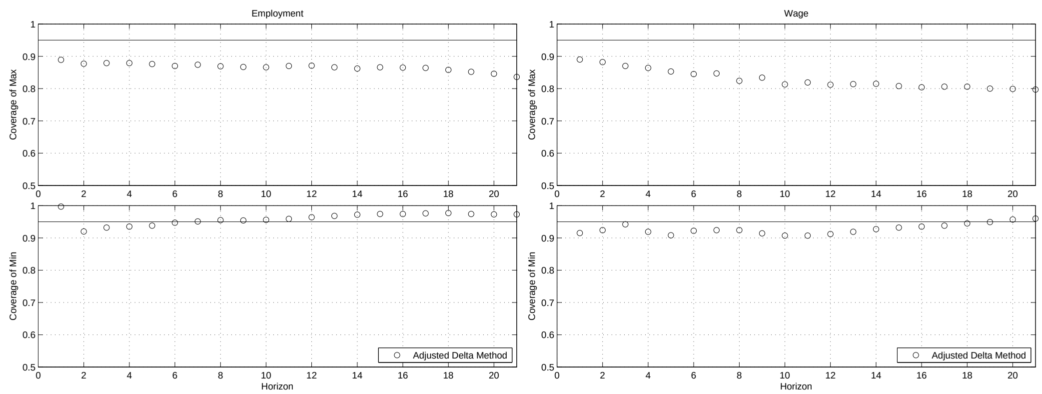
(Shaded, Gray Area) 95% Bayesian Credible Set using Algorithm 1 described in Section 3.2 of Arias et al. (2014) applied to the auxiliary program (8.4). (Blue, Solid Line) 95% adjusted delta method with $\varepsilon_1 = \varepsilon_2$ and the following tuning parameter for each of the time series: $[.05, .1]$. (Blue, Dashed Line) 95% projection method confidence set.

Figure 9 above reports the adjusted delta-method confidence bands for the cumulative responses to an expansionary demand shock. The delta-method formula in (8.5) now takes into account the sampling uncertainty in the constraint $\Sigma^{-1}h_1 \geq \mathbf{0}_{n\times 1}$. Just as before, we adjust the delta method formula to deal with multiplicity of solutions in the program (8.3).

Figure 10 below presents the Monte-Carlo coverage of the adjusted delta-method confidence sets.

Figure 10: Monte-Carlo Coverage Frequency of 95% Adjusted Delta-method Confidence Sets

(Contemporaneous Restrictions on Two Shocks)



(a) Employment (b) Wage

(Upper Panel) Monte-Carlo coverage frequency of the maximum response. (Lower Panel) The lower panel reports the Monte-Carlo coverage frequency of the minimum response.

Even though the Monte-Carlo coverage frequency in Figure 10 is above 80% for the two time series, the program in (8.3) poses the same challenge for the uniform validity of the adjusted delta method as the violations of the n-full column rank assumption discussed in Section 5.1.

Appendix D.3 shows that the value function $\overline{v}_{k,i1}$ will loose its differentiability if $\sigma_{12} \equiv e_1'\Sigma e_2 = 0$, as two of the constraints become linearly dependent. In the 2-dimensional example presented in this section if σ_{12} is bounded away from zero (all over the parameter space) the adjusted delta method of section 5 remains uniformly valid.

For more than two shocks, we can use Tucker's Theorem of the Alternative to modify Lemma 3 in the following way:

Lemma 4 *Let $S_1, S_2, \ldots S_n \in \mathbb{R}^{n\times n}$ denote matrices of contemporaneous sign restrictions for the n structural shocks of the model. Consider the mathematical program:*

$$\overline{v}_{k,ij}(A,\Sigma) \equiv \max_{H\in\mathbb{R}^{n\times n}} e_i' C_k(A) H e_j \sigma_j \quad s.t. \;\; HDH' = \Sigma, \;\; and \; S_l H e_l \geq \boldsymbol{0}_{n\times 1} \quad \forall l = 1, \ldots n.$$

For each $l \neq j$, let $S_{k,l}$ be a diagonal matrix with 1 and -1 diagonal elements. Let $\mathbf{S}_j =$

$(S_{k,1}, \ldots S_{k,j-1}, S_{k,j+1}, S_{k,n})$. *Consider the auxiliary program:*

$$\overline{v}_{k,ij}(A,\Sigma;\mathbf{S}_j) \equiv \max_x e_i' C_k(A)x \quad s.t \;\; x'\Sigma^{-1}x = 1, S_j x \geq \boldsymbol{0}_{n\times 1}, \;\; and \; S_{k,l}\Sigma^{-1}x \geq \boldsymbol{0}_{n\times 1}, l \neq j.$$

Then

$$\overline{v}_{k,ij}(A,\Sigma) \leq \max_{\mathbf{S}_j \in \mathcal{S}} \overline{v}_{k,ij}(A,\Sigma;\mathbf{S}_j)$$

where $\mathcal{S}$ is the set of all collections $(S_{k,1}, \ldots S_{k,j-1}, S_{k,j+1}, S_{k,n})$ such that the matrix product $(S_{k,l}S_l)$ is neither positive definite nor negative definite for all $l \neq j$.

Proof: See Appendix 4. *Q.E.D.*

Application the Delta Method to the 3-Dimensional SVAR of Kilian and Murphy using fmincon: Unfortunately, with more than two shocks we do not have tractable closed-form expression for the Karush-Kuhn-Tucker points of the program defining $\overline{v}_{k,ij}(\cdot)$. We now use Matlab®'s `fmincon` to solve for the maximum/minimum response and also for the Lagrange multipliers. Our application is the 3-dimensional SVAR in Kilian and Murphy (2012) with contemporaneous sign restrictions.

Kilian and Murphy (2012) fit a 24-lag SVAR to monthly data on the percent change in global crude oil production, an index of real economic activity representing the global business cycle, and the log of the real price of oil from $t = 1973 : M1 - 2008 : M9$.

There are three structural shocks in their model: a contractionary oil supply shock, an expansionary aggregate demand shock and an oil-market specific demand shock. The shocks are set-identified using the following contemporaneous sign restrictions:

$$S_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

For illustration, we focus on the dynamic responses to a contractionary supply shock. Let $D = \mathbb{I}_3$. We have shown that the constraint $HDH' = \Sigma$ is equivalent to the constraint $H'\Sigma^{-1}H = \mathbb{I}_n$. Let h_i denote the i-th column of the matrix H. The Lagrangian $\mathcal{L}(x; A, \Sigma, S)$ defining the maximum response in this program is given by:

$$\underbrace{(h_1' \otimes e_i')\text{vec}(C_k)}_{\text{Objective Function}} - \underbrace{\sum_{i=1}^{3} w_i\Big(h_i'\Sigma^{-1}h_i - 1\Big)}_{\text{ellipsoid constraint for } h_i} - \underbrace{\sum_{i=2}^{3} \kappa_{i-1}\Big(h_1'\Sigma^{-1}h_i\Big)}_{\text{orthogonality constraints for } h_1}$$
$$- \underbrace{\kappa_3\Big(h_2'\Sigma^{-1}h_3\Big)}_{\text{orthogonality between } h_2 \text{ and } h_3} - \underbrace{\sum_{i=1}^{3} \lambda_i' S_i h_i}_{\text{sign restrictions for } h_i}$$

The envelope theorem suggests that $\dot{\overline{v}}_k(A,\Sigma;z)$ is given by:

$$\begin{bmatrix} G_k(A)(h_1^*\otimes e_i) \\ \sum_{i=1}^3 w_i^*\Big(\Sigma^{-1}h_i^*\otimes\Sigma^{-1}h_i^*\Big) - \sum_{i=2}^3 \kappa_{i-1}^*\Big(\Sigma^{-1}h_i^*\otimes\Sigma^{-1}h_1^*\Big) - \kappa_3^*\Big(\Sigma^{-1}h_3^*\otimes\Sigma^{-1}h_2^*\Big)\end{bmatrix},$$

As we said before, the Karush-Kuhn-Tucker conditions for these problems are not as tractable as before. We can still establish some properties. For example, we can show that $w_2^* = w_3^* = 0$, which means that relaxing the ellipsoid constraints for h_2 and h_3 has no effect over the maximum response $e_i'C_kh_1^*$.

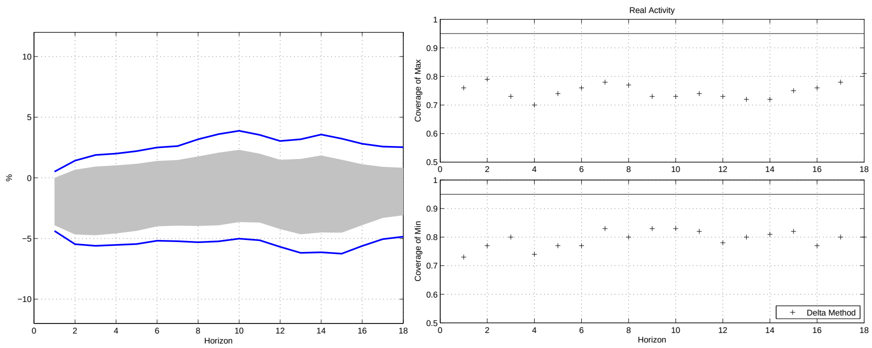
Comment on the Numerical Computation of Bounds: We use Matlab®'s `fmincon` to solve for h_i^* and for the Lagrange multipliers (ω_i^*, κ_i^*) for $i = 1, 2, 3$. First, we use the 'interior point solver' option to compute the maximizers. Then, we solve again the optimization problem—using the previous solution as an initial point—using `fmincon`'s 'active set solver'. We do this to get better precision in the estimation of the Lagrange multipliers. In our experience, such combination of solvers provides more robust and precise solutions for both the maximizers and the Lagrange multipliers.

We also use analytic gradients and Hessians for the constraints and the objective function to speed up the computation. We order the optimization problems according to the horizons of impulse-response function. Our initial guess for the interior point solver is the Cholesky decomposition of Σ. Then, we recursively use the obtained solutions as our initial guess for the consecutive horizons.

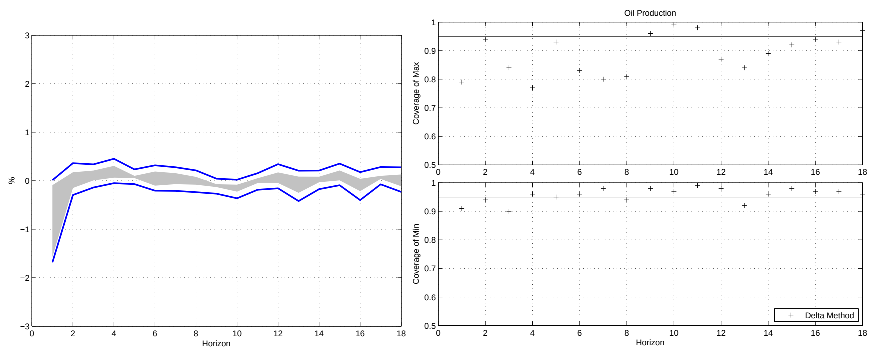
There is another interesting possibility to set the initial conditions of our optimization problem. Suppose that the researcher has sampled from rotation matrices—given the fixed OLS values of $(\widehat{A},\widehat{\Sigma})$—to approximate Bayesian credible sets. The researcher could use (for each horizon k) the rotation matrix Q that gives the maximum or the minimum response to construct a candidates $H^* = [h_1^*, h_2^*, h_3^*]$. We think about this procedure as a way of using `fmincon` after a first round of grid-search optimization. Since the problem of sampling efficiently from Q (uniformly) is a well-researched topic (see Arias et al. (2014), Rubio-Ramirez et al. (2010)), we could use available Matlab® codes to provide initial conditions for our `fmincon` routine. We plan to explore this possibility in future research.

Monte-Carlo: Figure 11 reports maximum and minimum cumulative responses (bounds of the light gray area) corresponding to a σ-contractionary oil supply shock under the contemporaneous restrictions S_1, S_2, S_3. Once again, we note that we are not sampling from rotation matrices to obtain these bounds. The figure also presents the Monte-Carlo coverage frequencies of our procedure for a sample size of $T = 417$.

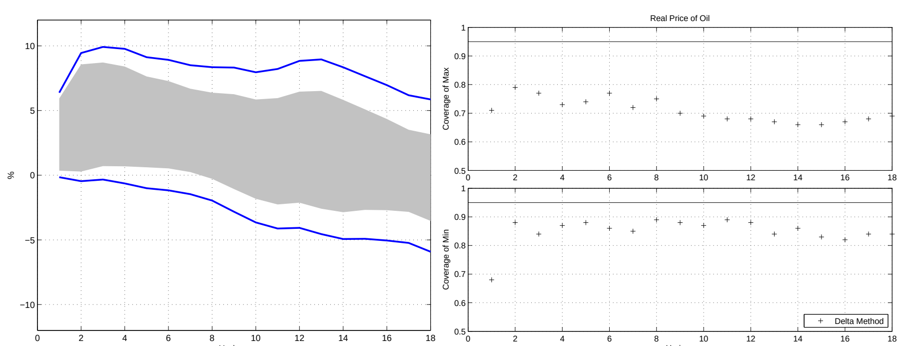
Figure 11: Unadjusted 95% Frequentist Confidence Bounds



(a) Global Crude Oil Production



(b) Global Business Cycle



(c) Real Price of Oil

Left Panel: (Shaded, Light Gray Area) For each horizon k and each variable i the gray area represents the interval $\left[\underline{v}_{k,ij}(\widehat{\theta})\ ,\ \overline{v}_{k,ij}(\widehat{\theta})\right]$. (Blue, Solid Line) Unadjusted 95% delta method based on the output obtained from Matlab®'s `fmincon`. **Right Panel:** Monte-Carlo coverage frequencies for the maximum and minimum response.