DELTA-METHOD INFERENCE FOR A CLASS OF SET-IDENTIFIED SVARS[1]

Bulat Gafarov[2], Matthias Meier[3] and José Luis Montiel Olea[4]

We study vector autoregressions that impose equality and/or inequality restrictions to set-identify the dynamic responses to a single structural shock. We make three contributions. First, we present an algorithm to compute the largest and smallest value that an impulse-response coefficient can attain over its identified set. Second, we provide conditions under which these largest and smallest values are directionally differentiable functions of the model's reduced-form parameters. Third, we propose a delta-method approach to conduct inference about the structural impulse-response coefficients. We use our results to assess the effects of the announcement of the Quantitative Easing program in August 2010. (JEL-Classification: C1, C32, E47).

Keywords: Set-Identification, Sign Restrictions, SVAR, Directional Differentiability, Unconventional Monetary Policy.

---

[1] Recepient of the 2016 award for best paper in applied economics presented by young researchers at the 69th European Meeting of the Econometric Society.

[2] University of California, Davis, Department of Agricultural and Resource Economics. E-mail: bgafarov@ucdavis.edu.

[3] Mannheim University, Department of Economics. E-mail: m.meier@uni-mannheim.de.

[4] Corresponding author; Columbia University, Department of Economics. E-mail: jm4474@columbia.edu.
This draft: December 11, 2017. First draft: November 19th, 2014.

1. INTRODUCTION

An increasingly popular practice in empirical macroeconomics is to set-identify the parameters of a structural vector autoregression [SVAR] by means of exclusion and/or sign restrictions. Most studies working with this type of models have relied on Bayesian methods to construct posterior credible sets for the structural parameters of interest (for example, Inoue and Kilian (2013), Arias, Rubio-Ramirez, and Waggoner (2017), and Baumeister and Hamilton (2015)).

A practical concern with Bayesian analysis in set-identified SVARs is that posterior inference continues to be influenced by prior beliefs even if the sample size is infinite (Poirier (1998), Gustafson (2009), Moon and Schorfheide (2012)). This observation has motivated the study of alternative approaches to inference that dispense with the specification of a prior distribution over structural parameters that are only set-identified.

There are two existing proposals that characterize the estimation uncertainty of set-identified structural responses, without postulating a specific prior for the parameters of the structural model. On the one hand, Granziera, Moon, and Schorfheide (2017) [GMS17] have proposed a *frequentist* confidence interval for structural impulse-response coefficients based on a moment-inequality-minimum-distance framework. On the other hand, Giacomini and Kitagawa (2015) [GK15] have proposed a *robust Bayes* credible interval that achieves a given credibility level regardless of the prior specified over the model's set-identified structural parameters.

We contribute to the analysis of set-identified SVARs by proposing a novel delta-method interval for the coefficients of the impulse-response function [IRF]. We show that our delta-method interval is *point-wise consistent in level* and, under certain regularity conditions, has *asymptotic robust Bayesian credibility* of at least the nominal level. Thus, our inference approach can be interpreted both from a frequentist and a robust Bayes perspective. We also argue that the computational cost of our procedure compares favorably with GMS17 and GK15.

Broadly speaking, our approach is based on a closed-form characterization of the endpoints of the identified set and their directional derivatives. Our delta-method interval— which may be viewed as a generalization of the pioneering work of Lütkepohl (1990) on delta-method inference for point-identified VARs—takes the form of a plug-in estimator for the identified set plus/minus standard errors.

The main limitation of our approach is that the delta-method interval is only defined for SVAR models that impose equality and inequality restrictions on a single structural shock (e.g., a monetary policy shock). Admittedly, this is problematic, as some popular applications of set-identified SVARs feature restrictions on multiple structural innovations.[1]

[1]SVAR applications for the oil market set-identify both demand and supply shocks using sign restrictions and elasticity bounds [Kilian and Murphy (2012)]. The same is true for recent labor market applications [Baumeister and Hamilton (2015)]. Also Mountford and Uhlig (2009)—one of the most cited applications of

In spite of this observation, single-shock set-identified models have been applied in several empirical studies: for example, to study the effects of monetary policy on output [Uhlig (2005)], the impact of monetary policy on the housing market [Vargas-Silva (2008)], the effects of labor market shocks on worker flows [Fujita (2011)], the effects of exchange rates on aggregate prices [An and Wang (2012)], and the effect of optimism shocks on business cycles fluctuations [Beaudry, Nam, and Wang (2011)]. Thus, we think there is room for our results to have an impact on empirical work.

To illustrate the usefulness of our main results, we estimate a monetary structural vector autoregression using monthly U.S. data from July 1979 to December 2007 (a sample that deliberately ends a half-year before the financial crisis begins). The goal of our exercise is to use pre-crisis data to learn about the responses of macroeconomic variables to shocks that have effects similar to the 'unconventional' monetary policy interventions implemented after the crisis.

We set-identify an *unconventional* monetary policy [UMP] shock as an innovation that decreases the two-year government bond rate upon impact, but has no effect over the nominal federal funds rate.[2] We consider two additional sign restrictions on the contemporaneous responses of inflation and output. Namely, we assume that—upon impact—neither inflation nor output can respond negatively to a UMP shock. Since the model is only set-identified, our analysis effectively captures the effects of any historical economic shock that affected the economy in the same way as an UMP shock.

We apply our delta-method approach to construct a confidence interval for the dynamic responses of industrial production, inflation, the two-year government bond rate, and the nominal federal funds rate. We use our delta-method intervals to assess the effects of the announcement of the second part of the so-called Quantitative Easing program (QE2) in August 2010. Pre-crisis data turns out to be extremely useful to learn about the post-crisis response of macroeconomic aggregates to unconventional monetary policy.

The remainder of the paper is organized as follows. Section 2 presents an overview of the main methodological results in this paper. Section 3 introduces our empirical application, which is used as a running example throughout the paper. Section 4.1 presents our algorithm to evaluate the endpoints of the identified set. Section 4.2 establishes the differentiability properties of the endpoints. Section 4.3 presents our delta-method approach and establishes its asymptotic frequentist validity as well as its asymptotic robust Bayesian credibility. Section 5 presents the delta-method intervals for the dynamic responses to the QE2 program. Section 6 concludes. All of our proofs are collected in Appendix A. Additional figures and implementation details of different procedures are collected in Appendix B.

set-identified SVARs—use sign restrictions to identify a government revenue shock as well as a government spending shock, while controlling for a generic business cycle shock and a monetary policy shock.

[2] The paper focuses on the two-year rate as this variable changed considerably after the announcement of the second round of the Quantitative Easing program. See Krishnamurthy and Vissing-Jorgensen (2011)

Generic Notation: If A is a matrix, A_{ij} denotes the ij-th element of A, vec(A) denotes the vectorization of A, and 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 $e_i^m \in \mathbb{R}^m$ denotes the i-th column of the identity matrix—denoted $\mathbb{I}_m$—of dimension m. If B is a matrix of dimension $n \times n$, $B_i \equiv B e_i^n$ denotes its i-th column. If the dimension of e_i^n is obvious, we ignore the superscript n.

2. MODEL, SET-IDENTIFYING RESTRICTIONS, AND OVERVIEW OF MAIN THEORETICAL RESULTS

This section presents the baseline SVAR model, discusses the class of set-identifying restrictions that we consider, and provides an overview of our main methodological results.

2.1. *SVAR model and impulse-response coefficients*

We study the n-dimensional structural vector autoregression (SVAR) with p lags; i.i.d. structural shocks distributed according to F; and unknown $n \times n$ structural matrix B:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + B\varepsilon_t, \;\; \mathbb{E}_F[\varepsilon_t] = 0_{n\times 1}, \;\; \mathbb{E}_F[\varepsilon_t \varepsilon_t'] \equiv \mathbb{I}_n. \tag{2.1}$$

The object of interest is the k-th period ahead structural impulse response function of variable i to a particular shock j (e.g., a monetary policy shock):

$$\lambda_{k,i,j}(A, B) \equiv e_i' C_k(A) B_j, \tag{2.2}$$

where $B_j \equiv B e_j$ and e_i and e_j denote the i-th and j-th column of $\mathbb{I}_n$.[3] We refer to the parameter in (2.2) as the (k, i, j)-*coefficient of the structural impulse-response function.*

An auxiliary object in the estimation of (2.2) is the vector of *reduced-form VAR parameters*:

$$\mu \equiv (\text{vec}(A)', \text{vec}(\Sigma)')' \in \mathcal{M} \subseteq \mathbb{R}^d, \; A \equiv (A_1, A_2, \ldots, A_p), \quad \Sigma \equiv BB'. \tag{2.3}$$

The reduced-form parameter space is denoted as $\mathcal{M}$. The parameter A denotes the autoregressive coefficients of the VAR model, while Σ denotes the covariance matrix of residuals. These parameters can be estimated directly from the data by multivariate Least-Squares (LS). Our main high-level assumption will be the approximate normality of the distribution

[3] The transformation $C_k(A)$ that appears in equation (2.2) is defined recursively by the formula $C_0 \equiv \mathbb{I}_n$:

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

$A_m = 0$ if $m > p$; see Lütkepohl (1990), p. 116.

of the LS estimator of μ. This condition will be satisfied even in the presence of unit roots and possible cointegration of unknown form (see Sims, Stock, and Watson (1990), Toda and Yamamoto (1995), Dolado and Lütkepohl (1996), Inoue and Kilian (2002), and Proposition 7.1 in Lütkepohl (2007)). Our main assumption is less demanding than the asymptotic normality of the reduced-form impulse-responses in GMS17 (see Kilian (1998), Benkwitz, Neumann, and Lütekpohl (2000)).[4]

2.2. *Set-Identifying Restrictions*

A common practice in empirical macroeconomics is to use equality and inequality restrictions to *set-identify* the structural IRFs in (2.2). An example of an equality restriction in a monetary VAR is that prices do not react contemporaneously to monetary policy shocks. An example of an inequality restriction is that a contractionary monetary policy shock cannot increase prices.

Let $\mathcal{R}(\mu) \subseteq \mathbb{R}^n$ be the set of values of B_j that satisfy the inequality and equality restrictions. In our paper, the set $\mathcal{R}(\mu)$ takes the form

$$\mathcal{R}(\mu) \equiv \left\{ B_j \in \mathbb{R}^n \;\middle|\; Z(\mu)' B_j = \mathbf{0}_{m_z \times 1} \text{ and } S(\mu)' B_j \geq \mathbf{0}_{m_s \times 1} \right\}, \tag{2.4}$$

where $Z(\mu)$ is a matrix of dimension $n \times m_z$ and $S(\mu)$ is a matrix of dimension $n \times m_s$. The matrix $Z(\mu)$ collects the equality restrictions specified by the researcher (we assume there are m_z of them). The matrix $S(\mu)$ collects the inequality restrictions (we assume there are m_s of them).

The simple formulation in (2.4) allows the researcher to incorporate the following identifying restrictions:

a) Sign restrictions on the responses of variable i at horizon k to an impulse on the j-th shock:

$$e_i' C_k(A) B_j \geq \text{ or } = 0,$$

as in Uhlig (2005).

b) Long-run restrictions on the response of variable i to an impulse on the j-th shock:

$$e_i'(\mathbb{I}_n - A_1 - \ldots - A_p)^{-1} B_j \geq \text{ or } = 0,$$

as in Blanchard and Quah (1989).

c) Short-run restrictions on the coefficients of the j-th structural equation. For example,

[4] We would like to thank an anonymous referee for suggesting this clarification.

the contemporaneous coefficient of the i-th variable in the j-th structural equation:

$$e_i'(B')^{-1}e_j = e_i'\Sigma^{-1}B_j \geq \text{ or } = 0,$$

as in Rubio-Ramirez, Caldara, and Arias (2015).

d) Elasticity bounds as in Kilian and Murphy (2012); for example, for some $b \in \mathbb{R}$:

$$e_i'B_j/e_{i'}'B_j \geq b \iff (e_i - be_{i'})'B_j \geq 0,$$

provided $e_{i'}'B_j > 0$.

Sign-Normalization: In order to make sure that the impulse response of interest is with respect to a fixed-sign shock one should always impose a sign-normalization. Our framework allows at least two different ways of imposing such a normalization: i) restricting the sign of the direct effect of the j-th variable on the j-th equation, or ii) restricting the sign of an arbitrary IRF coefficient. The first type of sign normalization is covered in c) as the short-run restriction $e_j'B^{-1}e_j \geq 0$, while the second is covered in a) as a typical sign restriction on the IRFs.

2.3. *Overview of the main results*

The main results in this paper concern the 'endpoints' of the identified set for a given structural impulse-response coefficient, $\lambda_{k,i,j}$. These endpoints (which we sometimes refer to as the *maximum and minimum* response) are defined as follows:

Definition 1: Given a vector of reduced-form parameters μ we define the endpoints of the identified set for $\lambda_{k,i,j}$ as the functions:

$$\overline{v}_{k,i,j}(\mu) \equiv \sup_{B \in \mathbb{R}^{n\times n}} e_i'C_k(A)Be_j, \text{ s.t. } BB' = \Sigma \text{ and } Be_j \in \mathcal{R}(\mu), \tag{2.5}$$

and

$$\underline{v}_{k,i,j}(\mu) \equiv \inf_{B \in \mathbb{R}^{n\times n}} e_i'C_k(A)Be_j, \text{ s.t. } BB' = \Sigma \text{ and } Be_j \in \mathcal{R}(\mu). \tag{2.6}$$

The functions $\overline{v}_{k,i,j}(\mu), \underline{v}_{k,i,j}(\mu)$ correspond to the largest and smallest value of the structural parameter over its identified set.

Our delta-method approach is supported by the three results described in the abstract, which can be summarized as follows:

• THEOREM 1 *(Algorithm to evaluate the maximum and minimum response)*: We present an algorithm that allows a researcher to evaluate the endpoints of the identified set given a vector of reduced-form parameters. The algorithm—inspired by the earlier work of Faust (1998)—evaluates all different collections of 'active' constraints and selects those that generate the largest (or smallest) value function—after checking that the inequality constraints not included in the set of active constraints are satisfied.[5]

Our algorithm does not require sampling from the space of structural matrices B. Instead, we show that $\overline{v}_{k,i,j}(\mu)$ and $\underline{v}_{k,i,j}(\mu)$ are the *value functions* of a mathematical program whose *Karush-Kuhn-Tucker* points can be described analytically—up to a set of active inequality constraints. More concretely, Lemma 1 shows that the maximum response for $\lambda_{k,i,j}$ is equal to either plus or minus the function

$$v_{k,i,j}(\mu;r) \equiv \Big(e_i' C_k(A)\Sigma^{1/2} M_{\Sigma^{1/2}r} \Sigma^{1/2} C_k(A)' e_i\Big)^{1/2},$$

where

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

and r is a matrix collecting the gradient vectors of the constraints in $\mathcal{R}(\mu)$ that are active at a maximum. Evaluating the function above for different values of r and checking the feasibility of the corresponding solution yields the maximum response. The minimum response is obtained analogously.

• THEOREM 2 *(Directional Differentiability of the endpoints)*: We show that the functions $\overline{v}_{k,i,j}(\cdot)$ and $\underline{v}_{k,i,j}(\cdot)$ are *directionally* differentiable. More precisely, let $X^*(\mu)$ denote the set of maximizers of program (2.5). Consider a sequence of 'perturbations' of μ each of them in a 'direction' $h_N \in \mathbb{R}^d$. We show that for any sequence $h_N \in \mathbb{R}^d$ such that $h_N \to h \in \mathbb{R}^d$, and any sequence $t_N \to \infty$:

$$t_N\Big(\overline{v}_{k,i,j}(\mu + h_N/t_N) - \overline{v}_{k,i,j}(\mu)\Big) \to \max_{x \in X^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r(\mu;x))' h\Big],$$

where $r(\mu;x)$ collects the gradient of the constraints that are active at a point x and $\dot{v}_{k,i,j}(\cdot;r)'$ is a gradient of $v_{k,i,j}(\cdot;r)$. The proof of the result above builds on Lemma 2 which establishes the differentiability of the function $v_{k,i,j}$ for a fixed set of active constraints. We relate the expression of the directional derivative with the generalized versions of the envelope theorems in the work of Fiacco and Ishizuka (1990) and Bonnans and Shapiro (2000). We argue that directional differentiability of the value functions (as opposed to full differentiability) arises due to the possibility that different structural models lead to the maximum (or minimum) response.

[5] Given a point x, we refer to any collection of binding restrictions defining $\mathcal{R}(\mu)$ as *active* constraints at x. The term 'active constraints' or 'active set of constraints' is the common terminology used in numerical optimization; see p. 308 in Nocedal and Wright (2006).

• THEOREM 3 *(Large-sample properties)*: We establish the point-wise consistency in level and the asymptotic robust Bayes credibility of our delta-method interval. Our suggested interval takes the form

$$\mathrm{CS}_T(1-\alpha;\lambda_{k,i,j}) \equiv \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\;\; \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\Big],$$

where $\widehat{\mu}_T$ is the typical LS estimator for the VAR reduced-form parameters, $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of a standard normal, and $\widehat{\sigma}_{(k,i,j),T}$ is our formula for the standard errors based on the directional derivatives.

3. RUNNING EXAMPLE: UNCONVENTIONAL MONETARY POLICY SHOCKS

This section introduces our empirical application, which will be used as a running example to illustrate our assumptions and results.

We consider a simple 4-variable model that includes the Consumer Price Index (CPI_t), the Industrial Production Index (IP_t), the 2-year Treasury Bond rate ($2yTB_t$), and the Federal Funds rate (FF_t).[6] We take a logarithmic transformation of CPI_t, IP_t and then work with first differences for all variables. Thus, our vector of macro variables is:

$$Y_t \equiv \Big(\ln CPI_t - \ln CPI_{t-1},\quad \ln IP_t - \ln IP_{t-1},\quad 2yTB_t - 2yTB_{t-1},\quad FF_t - FF_{t-1}\Big)'.$$

We set the number of lags equal to $p = 12$ following Gertler and Karadi (2015). The time span of the monthly series is July 1979 to August 2008 ($T = 342$). To keep our exposition as simple as possible, we ignore potential co-integration issues between short-term and long-term interest rates. Without loss of generality, we assume that the column of B corresponding to an UMP shock is the first column; $B_1 \equiv Be_1$. Our equality/inequality restrictions are summarized in Table I. These sign restrictions can be justified by the DSGE model calibrated in the work of Bhattarai, Eggertsson, and Gafarov (2015).

TABLE I
SET-IDENTIFICATION OF AN UNCONVENTIONAL MONETARY POLICY SHOCK: RESTRICTIONS

Series	Acronym	UMP	Notation
Consumer Price Index	CPI	+	$e_1' B_1 \geq 0$
Industrial Production	IP	+	$e_2' B_1 \geq 0$
2-year Treasury Bond rate	2yTB	−	$e_3' B_1 \leq 0$
Fed Funds Rate	FF	0	$e_4' B_1 = 0$

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.

[6] All these variables are sourced from the data set of Gertler and Karadi (2015). We thank Peter Karadi for making their data set available to us.

Baumeister and Benati (2013) study a related identification scheme. They consider a Bayesian SVAR to study an analogous 'spread' monetary policy shock that leaves the short-term nominal rate unchanged, but affects the spread between the ten-year Treasury-bond yield and the policy rate.

OUTLINE FOR THE REST OF OUR PAPER: We have already presented an overview of our main results and described our running example. In the remaining part of the paper, we formalize Theorems 1, 2, 3 and use them to conduct inference about the responses to an *unconventional monetary policy shock*.

4. THEOREMS

4.1. *Theorem 1*

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

4.1.1. *Assumptions*

We make two assumptions on the sign and zero restrictions allowed in the model:

ASSUMPTION 1 The choice set in program (2.5) is not empty at μ.

This assumption simply requires that the identifying restrictions do not contradict each other.

Now, let $e_1^{m_s}, e_2^{m_s}, \ldots e_{m_s}^{m_s}$ denote the m_s different columns of the identity matrix $\mathbb{I}_{m_s}$. Let $e(k)$ denote an $m_s \times k$ matrix formed by collecting any of the $k \leq m_s$ columns of $\mathbb{I}_{m_s}$.

DEFINITION 2: We say that $Z(\mu)$ and $S(\mu)$ are linearly independent at μ if for any $k \in \mathbb{Z}$, $0 \leq k \leq m_s$ and any $e(k)$ the matrix

$$R(\mu; e(k)) \equiv [Z(\mu), S(\mu)e(k)] \in \mathbb{R}^{n \times (m_z + k)}$$

has full rank.

ASSUMPTION 2 $Z(\mu)$ and $S(\mu)$ are linearly independent at μ.

This assumption has two important implications. The first implication is that at most $n - 1$ constraints can be active at a solution of program (2.5) (in particular, it implies $m_z \leq n - 1$). The second implication is that it will allow us to characterize the maximum and minimum response in terms of *Karush-Kuhn-Tucker* conditions. We verify (and discuss) this assumption for the UMP example in Section 4.1.3.

4.1.2. ***Algorithm***

We now show that the value function $\overline{v}_{k,i,j}(\mu)$ in (2.5) can be obtained by applying a simple algorithm. Let r be the matrix that collects all the columns of $Z(\mu)$ and whatever columns of $S(\mu)$ that are active at a solution of program (2.5). Our first observation is that the value function $\overline{v}_{k,i,j}(\mu)$ equals plus or minus

$$v_{k,i,j}(\mu;r) = \Big(e_i' C_k(A)\Sigma^{1/2} M_{\Sigma^{1/2}r} \Sigma^{1/2} C_k(A)' e_i\Big)^{1/2},$$

and the corresponding maximizer equals either

$$x_+^*(\mu;r) \equiv \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu;r)$$

or

$$x_-^*(\mu;r) \equiv -\Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu;r),$$

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

This result is shown formally in Lemma 1 in Appendix A.1 (where we also provide intuition). The lemma implies that if we knew the program's binding constraints, the value function could be computed directly—up to its sign—as $v_{k,i,j}(\mu, r)$. Moreover, the sign of value function is positive if $x_+^*(\mu;r)$ satisfies the inequality restrictions that are not included in r, and negative otherwise.

Let R denote the set of all possible matrices r that could arise from collecting all of the m_z columns of the matrix $Z(\mu)$ and k out of the m_s columns of the matrix $S(\mu)$, where $0 \leq k \leq n - m_z - 1$. Take any c larger than

$$\bar{c} \equiv \max_{i,k} \Big(e_i' C_k(A)\Sigma C_k(A)' e_i\Big)^{1/2}.$$

The parameter c will be used to 'penalize' candidate solutions that do not satisfy the inequality restrictions in $S(\mu)$.[7] The penalization works as follows. Consider first the case in which $v_{k,i,j}(\mu;r) \neq 0$. Since $x_+^*(\mu;r)$ and $x_-^*(\mu;r)$ above are well defined, we can verify if these candidate solutions satisfy the sign restrictions that were not included in r (that is, we verify the *primal feasibility* of the solutions). If the primal feasibility condition is satisfied we store the candidate values; else we penalize them to guarantee that they are never a

[7] The constant $\bar{c}$ is the maximum value of the following programs:

$$\bar{c} \equiv \max_{i,k} \sup_{B\in\mathbb{R}^{n\times n}} e_i' C_k(A) B e_j, \text{ s.t. } BB' = \Sigma. \tag{4.1}$$

solution. More concisely, we define the auxiliary functions:

$$\begin{aligned} f^{+}_{max}(\mu;r) &\equiv v_{k,i,j}(\mu;r) - 2(1-\mathbf{1}_{m_s}(x^{*}_{+}(\mu;r)))c, \\ f^{-}_{max}(\mu;r) &\equiv -v_{k,i,j}(\mu;r) - 2(1-\mathbf{1}_{m_s}(x^{*}_{-}(\mu;r)))c, \end{aligned}$$

where $\mathbf{1}_{m_s}(x) \equiv \mathbf{1}\{S(\mu)'x \geq \mathbf{0}_{m_s\times 1}\}$ is 1 if and only if x satisfies all the inequality restrictions in $S(\mu)$. The functions $f^{+}_{\max}, f^{-}_{\max}$ allow us to keep track of the candidate values (and their feasibility) for each combination of active restrictions.

Consider now the penalization in the case in which $v_{k,i,j}(\mu;r) = 0$. This case is slightly different from the one considered in the previous paragraph, as the candidate solutions (x^{*}_{+} and x^{*}_{-}) are not always defined in this case. If there is a point $x^{*} \neq 0$ satisfying the equality restrictions in r and also the inequality restrictions that are not included in r, we set

$$f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = 0.$$

If no such point $x^{*} \neq 0$ exists, $v_{k,i,j}(\mu, r) = 0$ cannot be a solution and we set

$$f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = -2c.$$

The following theorem shows that we can compute the value function of the mathematical program (2.5) by selecting the maximum value of $\max\{f^{+}_{max}(\mu;r),\ f^{-}_{max}(\mu;r)\}$ over $r \in R$. That is, we can solve for $\overline{v}_{k,i,j}(\mu)$ by considering the different combinations of active restrictions and select the maximum value $\pm v_{k,i,j}(\mu, r)$ over them.

Theorem 1 *Suppose that Assumptions 1 and 2 hold, then:*

$$\overline{v}_{k,i,j}(\mu) = \max_{r\in R}\Big(\max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\}\Big).$$

The minimum value is obtained analogously.

Proof: The intuition behind the proof is as follows. Note that value achieved by any combination of active sign restrictions r for which $x^{*}_{+}(\mu;r)$ or $x^{*}_{-}(\mu;r)$ is well-defined and feasible must be, by definition, no larger than $\overline{v}_{k,i,j}(\mu)$. Thus, we only have to show that

$$\max_{r\in R}\Big(\max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\}\Big) \geq \overline{v}_{k,i,j}(\mu).$$

Since Lemma 1 showed that the value of the program (2.5) should be of the form $f^{+}_{\max}(\mu;r)$ or $f^{-}_{\max}(\mu;r)$ for some $r \in R$, the result must follow. The proof is formalized in Appendix A.2.

4.1.3. ***Using the algorithm in the UMP example***

We verify Assumptions 1 and 2 at the estimated LS values of μ, denoted $\widehat{\mu}_T$. The simplest way to verify Assumption 1 is to consider the different candidate solutions for the different combinations of active constraints and check whether one of these solutions is feasible. For Assumption 2, note that regardless of the number of k columns selected from S the resulting matrix $R(\mu, e(k))$ will always have full column rank. Thus, Assumption 2 is also verified.[8]

We now use our algorithm to evaluate the identified set and report $\overline{v}_{k,i,j}(\widehat{\mu}_T)$ and $\underline{v}_{k,i,j}(\widehat{\mu}_T)$ for the cumulative IRFs.[9] The bounds in Figure 1 correspond to a one standard deviation structural UMP shock.

We consider first the equality/inequality restrictions in Table I. Evaluating the endpoints of the identified set for the 4 variables in the VAR, over 36 horizons, takes around 0.1 seconds. We then include an additional inequality restriction on the response of output to an expansionary UMP shock. Namely, we assume that even one period after the shock, the cumulative effect on IP cannot be negative ($e_2'(C_0 + C_1(A))B_1 \geq 0$). Figure 1 shows that the upper bounds of the identified sets under the two identification schemes almost overlap. The figure suggests that the noncontemporaneous constraint has thus little additional identification power.

There are at least two other ways of evaluating the maximum and minimum response (although only our algorithm is guaranteed to provide a global solution in a finite number of steps). One approach is to simply use a numerical solver (such as Matlab's `fmincon`) to get the value of the non-linear, non-convex program in (2.5). The result in Theorem 1 allows us to avoid the specification of the standard tuning parameters for numerical optimization routines (such as initial conditions, algorithms for the solver, tolerance levels for the solutions, and number of iterations).

Another approach is to rely on a version of the Bayesian algorithm in Uhlig (2005). Given reduced-form parameters μ and D draws of a unit vector $q \in \mathbb{R}^n$, one could report the maximum and minimum value over $\{\lambda_{k,i,j}(\mu, q^d)\}_{d=1}^{D}$. Note that such algorithm is essentially a random grid search approach to solve the program (2.5). The grid search approach underestimates the identified set. In our application the bias is negligible for $D = 10,000$ draws (the algorithm, however, takes around 300 seconds to run).

---

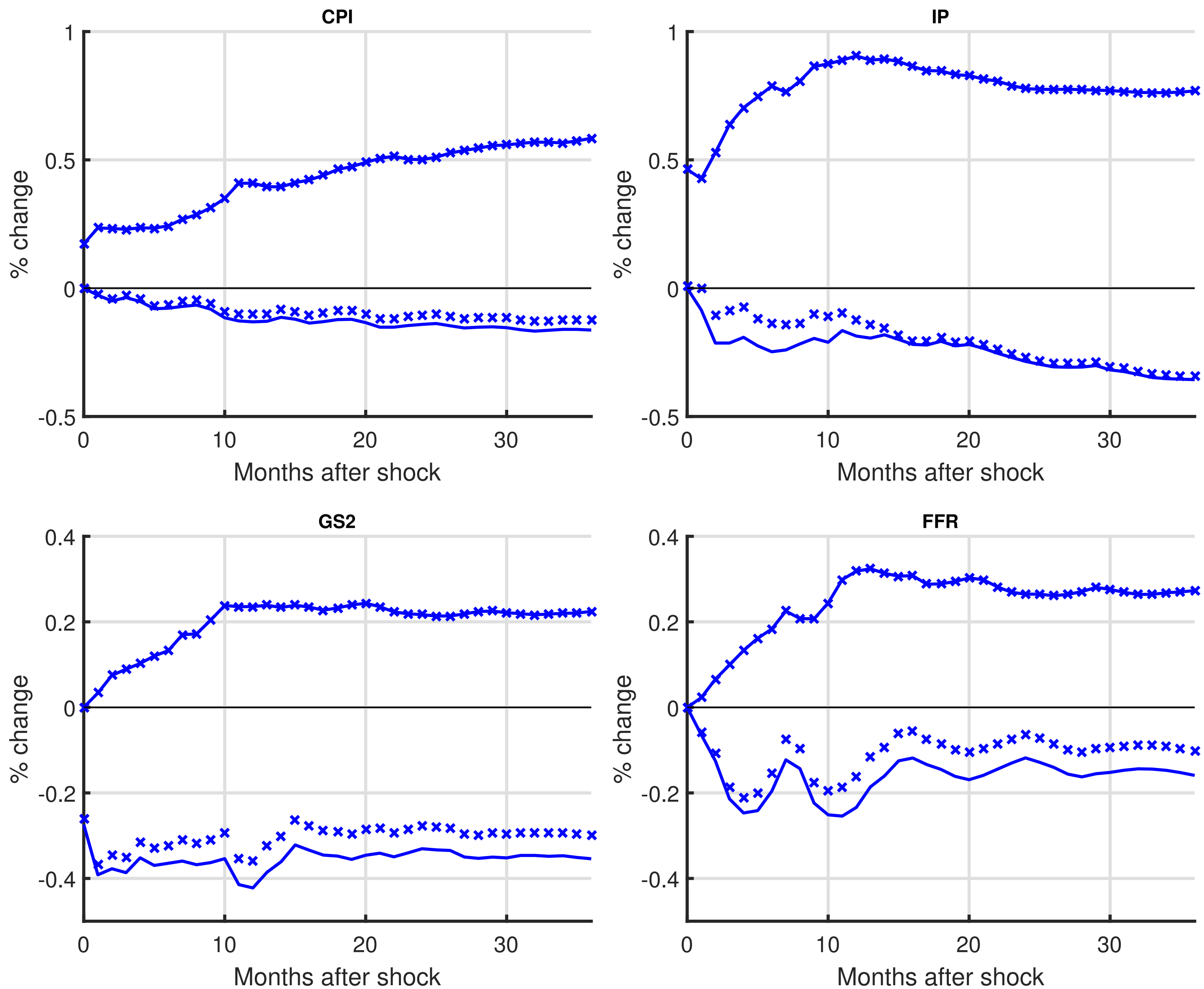
[8] Verifying Assumption 2 with more general restrictions requires additional work. For example, suppose that the researcher is interested in including the restriction:

$$e_2' C_1(A) B_1 \geq 0.$$

This restriction says that the UMP shock cannot decrease the growth rate in Industrial Production even one-period after the shock. Since $C_1(A) = A_1$, the vector $e_2' C_1(A)$ is equal to the second row of A_1, which we can denote as $(A_{1,(2,1)}, A_{1,(2,2)}, A_{1,(2,3)}, A_{1,(2,4)})$. Assumption 2 will be satisfied as long as μ is such that $A_{1,(2,j)} \neq 0$ for all $j = 1, \ldots 4$, which means that each of the entries in the first lag of Y_{t-1} has predictive power on Y_t after controlling for the rest of the lags.

[9] The formula for the maximum (minimum) k-th period ahead *cumulative* IRF replaces $C_k(\widehat{A}_T)$ by $C_0(\widehat{A}_T) + C_1(\widehat{A}_T) +, \ldots, +C_k(\widehat{A}_T)$.

Figure 1: Identified Set for the Cumulative Impulse Response Functions to a one standard deviation UMP shock (given $\widehat{\mu}_T$) for two different identification schemes



(Solid, Blue Line) Endpoints of the identified set for the cumulative responses given $\widehat{\mu}_T$ and the equality/inequality restrictions in Table I. (Blue, Crosses) Endpoints of the identified set with the additional restriction that the cumulative response of IP to a UMP shock one month after impact is non-negative, $e_2'(C_0 + C_1(A))B_1 \geq 0$. Note that the upper bounds of the identified sets under the two identification schemes almost overlap.

4.2. *Theorem 2*

In this section we show that the endpoints of the identified set—$\underline{v}_{k,i,j}(\cdot)$ and $\overline{v}_{k,i,j}(\cdot)$—are directionally differentiable functions of the reduced-form parameter μ. This result is the basis of our delta-method approach to conduct inference in set-identified SVARs.

4.2.1. *Assumptions*

In order to establish our differentiability result we need an additional regularity condition. Our key assumption is as follows:

Assumption 3 The matrices $Z(\cdot)$ and $S(\cdot)$ are differentiable at μ.

We are not aware of equality/inequality restrictions in the SVAR literature that do not satisfy this property. In particular, all the examples given in Subsection 2.2 of this paper satisfy Assumption 3 for every value of $\mu \in \mathcal{M}$.

4.2.2. *Directional Differentiability*

We will continue working with the auxiliary function $v_{k,i,j}(\mu; r(\mu))$, where we now explicitly acknowledge the possible dependence of r on μ. Lemma 2 in Appendix A.3 shows that if Assumptions 1-3 hold and $v_{k,i,j}(\mu; r(\mu)) \neq 0$, then the function $v_{k,i,j}(\mu; r(\mu))$ is differentiable with respect to $(\text{vec}(A)', \text{vec}(\Sigma)')'$ with the derivative $\dot{v}_{k,i,j}(\mu; r(\mu))$ given by:

$$\begin{bmatrix} \frac{\partial v_{k,i,j}(\mu; r(\mu))}{\partial \text{vec}(A)} \\ \\ \frac{\partial v_{k,i,j}(\mu; r(\mu))}{\partial \text{vec}(\Sigma)} \end{bmatrix} = \begin{bmatrix} \frac{\partial \text{vec}(C_k(A))}{\partial \text{vec}(A)}(x^*(\mu; r(\mu)) \otimes e_i) - \sum_{k=1}^{l} w_k^* \frac{\partial \text{vec}(r_k(\mu))}{\partial \text{vec}(A)} x^*(\mu; r(\mu)) \\ \\ \lambda^* \Sigma^{-1} x^*(\mu; r(\mu)) \otimes \Sigma^{-1} x^*(\mu; r(\mu)) - \sum_{k=1}^{l} w_k^* \frac{\partial \text{vec}(r_k(\mu))}{\partial \text{vec}(\Sigma)} x^*(\mu; r(\mu)) \end{bmatrix},$$

where $r_k(\mu)$ denotes the k-th column of $r(\mu)$,

$$x^*(\mu; r(\mu)) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2} r(\mu)}\Big)\Sigma^{1/2} C_k(A)' e_i \Big/ v_{k,i,j}(\mu; r(\mu)),$$

$$\lambda^* \equiv \frac{1}{2} v_{k,i,j}(\mu; r(\mu)), \quad w^* \equiv [r(\mu)'\Sigma r(\mu)]^{-1} r(\mu)'\Sigma C_k(A) e_i,$$

and w_k^* is the k-th component of the vector w^*.[10]

[10]The *envelope theorem* sheds light on the derivative formula provided in Lemma 2. Note first that, by definition,

$$v_{k,i,j}(\mu; r(\mu)) = \max_{x \in \mathbb{R}^n} e_i' C_k(A) x \quad \text{s.t.} \quad x'\Sigma^{-1} x = 1 \text{ and } r'(\mu) x = \mathbf{0}_{l \times 1}.$$

The auxiliary *Lagrangian function* of this problem is given by

$$\mathcal{L}(x; \mu, r(\mu)) = (x' \otimes e_i')\text{vec}(C_k(A)) - \lambda\Big((x' \otimes x')\text{vec}(\Sigma^{-1}) - 1\Big) - w'(r(\mu)'x),$$

where λ is the Lagrange multiplier corresponding to the quadratic equality restriction and $w \in \mathbb{R}^l$ is the

We now state the definition of directional differentiability and present our second Theorem.

Definition 3: We say that the real-valued function v with domain $\mathcal{M} \subseteq \mathbb{R}^d$ is *directionally differentiable* at μ if for any $h \in \mathbb{R}^d$, any sequence $t_N \to \infty$, and any sequence $h_N \in \mathbb{R}^d$ such that $h_N \to h$ $(\mu + t_N h_N \in \mathcal{M})$, there exists a continuous function $\dot{v}_\mu : \mathbb{R}^d \to \mathbb{R}$ such that:

$$t_N \Big(v(\mu + h_N/t_N) - v_{k,i,j}(\mu) \Big) \to \dot{v}_\mu(h).$$

We refer to the function $\dot{v}_\mu$ as the directional derivative of $v(\cdot)$ at μ.[11]

Let $X^*(\mu)$ denote the argmax of program (2.5). For $x \in X^*(\mu)$ let $r(\mu;x)$ denote the matrix that collects *all* elements of $Z(\mu)$ and $S(\mu)$ that are active at x.

Theorem 2 *Suppose that Assumptions 1-3 hold. Suppose in addition $\overline{v}_{k,i,j}(\mu) > 0$. Then $\overline{v}_{k,i,j}$ is a directionally differentiable function of the reduced-form parameter μ with the directional derivative given by*

$$\max_{x \in X^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r(\mu;x))' h \Big]. \tag{4.2}$$

Whenever $X^*(\mu) = \{x^*\}$ is a singleton, the value function $\overline{v}_{k,i,j}(\mu)$ is fully differentiable with the derivative $\dot{v}_{k,i,j}(\mu; r(\mu; x^*))$.[12]

Proof: See Appendix A.4.

Theorem 4.2, p. 223 in Fiacco and Ishizuka (1990) and Theorem 4.24, p. 280 in the book of Bonnans and Shapiro (2000) present a generalized version of the envelope theorem. They show that—under suitable constraint qualifications—the directional derivative (in direction h and evaluated at parameter μ) of the largest and smallest value in a mathematical program with equality and inequality constraints is given by

$$\sup_{x \in X^*(\mu)} \Big[\nabla_\mu \mathcal{L}(x;\mu) h \Big],$$

vector of Lagrange multipliers corresponding to the l equality restrictions. The envelope theorem suggests that $\dot{v}_{k,i,j}(\mu; r(\mu))$ is given by the formula in Lemma 2. This intuition is confirmed in the proof of Lemma 2 provided $v_{k,i,j}(\mu; r(\mu)) \neq 0$.

[11] See p.172 in Shapiro (1991).

[12] If $\overline{v}_{k,i,j}(\mu) < 0$ the directional derivative simply becomes

$$\max_{x \in X_*(\mu)} \Big[-\dot{v}_{k,i,j}(\mu; r(\mu;x))' h \Big],$$

and

$$\inf_{x \in X_*(\mu)} \Big[\nabla_\mu \mathcal{L}(x;\mu) h \Big],$$

provided there is a unique set of Lagrange Multipliers supporting the optimal solutions. Theorem 2 uses the results in Lemma 1 and Lemma 2 to verify this formula.

Delta-Method vs. Bootstrap: We also note that directionally differentiable functions have been a topic of recent research. Fang and Santos (2015) show that the standard bootstrap is not consistent when applied to parameters of the form $v(\mu)$, where v is a directionally differentiable function. Kitagawa, Payne, and Montiel Olea (2017) show that Bayesian credible sets based on the quantiles of the posterior distribution of $v(\mu)$ will be asymptotically equivalent to the frequentist bootstrap (which is not consistent in this case). These results imply that typical frequentist and Bayesian inference for directionally differentiable functions is not guaranteed to be consistent.

The next section shows that the special form of the directional derivative that arises in the class of SVAR models studied in this paper allows the researcher to conduct (computationally convenient) delta-method inference, with a slight adjustment on the standard errors. We note that the recent paper of Hong and Li (2017) provides an alternative frequentist point-wise consistent inference procedure for directionally differentiable functions of general form. Such an approach, however, has two drawbacks compared to our delta method. First, implementing the numerical delta-method in Hong and Li (2017) requires a user specified tunning parameter. Second, their procedure requires the evaluation of the value function for a large number of re-sampled values of μ (whereas our delta-method only requires the evaluation of the value functions at $\widehat{\mu}$).

4.3. ***Theorem 3***

This section proposes a delta-method interval of the form

$$\text{CS}_T(1-\alpha) \equiv \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\, \widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\, \widehat{\sigma}_{(k,i,j),T}/\sqrt{T} \Big],$$

where

$$\widehat{\mu}_T \equiv (\text{vec}(\widehat{A}_T)', \text{vec}(\widehat{\Sigma}_T)'),$$

is the LS estimator for μ defined as

$$\widehat{A}_T \equiv \Big(\frac{1}{T}\sum_{t=1}^{T} Y_t X_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\Big)^{-1}, \quad \widehat{\Sigma}_T \equiv \frac{1}{T-np-1}\sum_{t=1}^{T} \widehat{\eta}_t \widehat{\eta}_t',$$

with

$$X_t \equiv (Y'_{t-1}, \dots, Y'_{t-p})', \quad \widehat{\eta}_t \equiv Y_t - \widehat{A}_T X_t.$$

We work under the assumption that $\sqrt{T}(\widehat{\mu}_T - \mu)$ is asymptotically normal with some covariance matrix Ω.[13] We use the results in Theorem 2 and the asymptotic normality of $\widehat{\mu}_T$ to suggest the following formula for $\widehat{\sigma}_{(k,i,j),T}$:

$$\widehat{\sigma}_{(k,i,j),T} \equiv \max_{r \in R(\widehat{\mu}_T)} \Big(\dot{v}_{k,i,j}(\widehat{\mu}_T; r)' \widehat{\Omega}_T \dot{v}_{k,i,j}(\widehat{\mu}_T; r) \Big)^{\frac{1}{2}}, \tag{4.3}$$

where $R(\widehat{\mu}_T)$ is the set of *all* possible active constraints in program (2.5) evaluated at $\widehat{\mu}_T$. Note that our procedure does not attempt to estimate neither the argmax nor the argmin of program (2.5).

Frequentist Coverage: Let P denote the data generating process and let $\mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))$ denote the identified set for the structural parameter $\lambda_{k,i,j}$ given the equality/inequality restrictions in $\mathcal{R}(\mu)$. This section shows that under our proposed specification of $\widehat{\sigma}_{(k,i,j),T}$,

$$\liminf_{T \to \infty} \inf_{\lambda \in \mathcal{I}^{R}_{k,i,j}(\mu(P))} P\Big(\lambda \in \mathrm{CS}_T(1-\alpha)\Big) \geq 1 - \alpha.$$

Consequently, the delta-method interval presented in this paper is *point-wise consistent in level.*

Robust Bayesian Credibility: We also show that under some regularity conditions our delta-method interval has, asymptotically, robust Bayesian credibility of at least the nominal level. To formalize this statement, let P^* denote some prior for the structural parameters $(A_1, \dots A_p, B)$ and let $\lambda_{k,i,j}(A, B) \in \mathbb{R}$ denote the structural coefficient of interest. For a given square root of $\Sigma \equiv BB'$ define the orthogonal matrix $Q \equiv \Sigma^{-1/2}B$. It is well known that a prior P^* can be written as $(P^*_\mu, P^*_{Q|\mu})$, where P^*_μ is a prior on the reduced-form parameters, and $P^*_{Q|\mu}$ is a prior on the orthogonal matrix, conditional on μ. Following this notation, let $\mathcal{P}(P^*_\mu)$ denote the class of prior distributions such that $\mu^* \sim P^*_\mu$.

Define the *Robust Bayes Credibility* of our delta-method region as

$$RBC(Y_1, \dots, Y_T) \equiv \inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda(A,B) \in \mathrm{CS}_T(1-\alpha) \,\Big|\, Y_1, \dots, Y_T\Big). \tag{4.4}$$

[13] A common formula for $\widehat{\Omega}$ based on the assumption of uncorrelated, possibly heteroskedastic structural innovations is given by

$$\widehat{\Omega}_T \equiv \Big(\frac{1}{T}\sum_{t=1}^{T} \mathrm{vec}\Big([\widehat{\eta}_t X'_t, \widehat{\eta}_t \widehat{\eta}'_t - \widehat{\Sigma}_T]\Big) \mathrm{vec}\Big([\widehat{\eta}_t X'_t, \widehat{\eta}_t \widehat{\eta}'_t - \widehat{\Sigma}_T]\Big)'.$$

Our delta-method approach is also valid under the presence of time-series dependence in η_t (we only need a heteroskedasticity and autocorrelation robust estimator of Ω).

We show that if the prior for the reduced-form parameters μ satisfies the *Bernstein-von Mises Theorem* and the bounds of the identified set are *differentiable* then for any $\epsilon > 0$:

$$\lim_{T\to\infty} P(RBC(Y_1,\ldots,Y_T) < 1-\alpha-\epsilon) = 0$$

Thus, our delta-method interval has *asymptotic* robust Bayesian credibility of at least $1-\alpha$.

We now describe the main large-sample assumptions used to establish the frequentist coverage and the robust Bayesian credibility of our delta-method interval.

4.3.1. ***Assumptions***

The SVAR parameters $(A_1,\ldots,A_p,B,F)$ define a probability distribution, denoted P, over the data observed by the econometrician. Our main assumptions concerning P are as follows. First, we assume that the LS estimator $\widehat{\mu}_T$ is asymptotically normal with a covariance matrix that can be estimated consistently.

Assumption 4 (**Asymptotic Normality of** $\widehat{\mu}_T$) The data generating process P is such that for $\mu(P) \in \mathbb{R}^d$:

$$\sqrt{T}(\widehat{\mu}_T - \mu(P)) \xrightarrow{d} \zeta(P) \sim \mathcal{N}_d\Big(\mathbf{0}\,,\,\Omega(P)\Big),$$

and

$$\widehat{\Omega}_T \xrightarrow{p} \Omega(P).$$

Second, we will assume that the prior P^*_μ used to compute robust Bayesian credibility and the data generating process P satisfy the Bernstein von-Mises Theorem in Ghosal, Ghosh, and Samanta (1995). More precisely, we assume that:

Assumption 5 **(Bernstein-von Mises Theorem)**

$$\sup_{B\in\mathcal{B}(\mathbb{R}^d)} \Big| P^*_\mu\Big(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y_1,\ldots,Y_T\Big) - \mathbb{P}\left(\zeta(P)\in B\right)\Big| \xrightarrow{p} 0,$$

where $\zeta(P) \sim \mathcal{N}_d(\mathbf{0},\Omega(P))$, and $\mathcal{B}(\mathbb{R}^d)$ is the set of all Borel measurable sets in $\mathbb{R}^d$.

Assumption 5 is satisfied for Normal-Inverse Wishart prior (see Uhlig (2005)) in a VAR model with Gaussian i.i.d. errors (see Gafarov, Meier, and Montiel Olea (2016)). More generally, if the VAR reduced-form errors are i.i.d. Gaussian, Theorem 1 and 2 in Ghosal et al. (1995) imply that Assumption 5 will be satisfied whenever P^*_μ has a continuous density at μ with polynomial majorants.

4.3.2. ***Large-sample coverage and robust Bayesian credibility***

Dümbgen (1993), Shapiro (1991), and Fang and Santos (2015) have shown if v is a directionally differentiable function with directional derivative $\dot{v}_\mu(h)$ (in direction h evaluated at μ) then:

$$\sqrt{T}(v(\widehat{\mu}_T) - v(\mu)) \xrightarrow{d} \dot{v}_\mu(\zeta),$$

whenever Assumption 4 holds. Theorem 2 in the previous section established that the directional derivative of $\overline{v}_{k,i,j}$—in direction h evaluated at μ—is given by

$$\max_{x \in X^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r(\mu; x))' h \Big],$$

where $X^*(\mu)$ is the argmax of program (2.5) at μ. Thus, Theorem 2 and Assumption 4 imply that

$$\sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T) - \overline{v}_{k,i,j}(\mu)) \xrightarrow{d} \max_{x \in X^*(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r(\mu; x))' \zeta \Big],$$

where

$$\dot{v}_{k,i,j}(\mu; r(\mu; x))' \zeta \sim \mathcal{N}_1\Big(0, \dot{v}_{k,i,j}(\mu; r(\mu; x))' \Omega \dot{v}_{k,i,j}(\mu; r(\mu; x))\Big).$$

Our suggestion—which exploits the specific form of the directional derivative in the SVAR context—is to consider:

$$\widehat{\sigma}_{(k,i,j),T} \equiv \max_{r \in R(\widehat{\mu}_T)} \Big(\dot{v}_{k,i,j}(\widehat{\mu}_T; r)' \widehat{\Omega}_T \dot{v}_{k,i,j}(\widehat{\mu}_T; r) \Big)^{\frac{1}{2}},$$

where $R(\widehat{\mu}_T)$ is the set of *all* the different collections of active constraints evaluated at $\widehat{\mu}_T$. The idea is that $\widehat{\sigma}_{(k,i,j),T}$ converges in probability to

$$\max_{r \in R(\mu)} \Big(\dot{v}_{k,i,j}(\mu; r)' \Omega \dot{v}_{k,i,j}(\mu; r) \Big)^{\frac{1}{2}},$$

which is larger than or equal to

$$\max_{x \in X^*(\mu)} \Big(\dot{v}_{k,i,j}(\mu; r(\mu, x))' \Omega \dot{v}_{k,i,j}(\mu; r(\mu, x)) \Big)^{\frac{1}{2}}.$$

Thus, our formula for the standard error implies that there is no need to estimate neither the argmax nor the argmin of the program defining $\overline{v}(\mu)$. The suggested confidence interval is shown to be point-wise consistent in level.[14] We also show that our delta-method interval

[14]The question of how to build a *uniformly consistent in level*, delta-method confidence set for a set-identified parameter is beyond the scope of this paper. For the readers interested in uniform inference for set-identified parameters in SVARs our suggestion is to apply the projection approach developed in Gafarov et al. (2016). In comparison, the delta-method approach described in this paper is faster to implement.

has, asymptotically, robust Bayesian credibility of at least the nominal level (provided some regularity conditions are satisfied). These two properties are formalized in the following theorem.

Theorem 3 *Let $\widehat{\sigma}_{(k,i,j),T}$ be defined as in (4.3). Suppose that the asymptotic variance of the candidate value functions in $X^*(\mu)$ and $X_*(\mu)$ are strictly positive; that is*

$$\min_{x \in X_*(\mu(P)) \cup X^*(\mu(P))} ||\Omega^{1/2}(P)\dot{v}_{k,i,j}(\mu(P); r(\mu(P); x))|| > 0.$$

a) If Assumptions 1-4 are satisfied at $\mu = \mu(P)$, then

$$\liminf_{T\to\infty} \inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in CS_T(1-\alpha)\Big) \geq 1-\alpha.$$

b) If in addition Assumption 5 holds and $X^*(\mu(P))$ and $X_*(\mu(P))$ are both singletons, then for any $\epsilon > 0$:

$$\lim_{T\to\infty} P\left(\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda(A,B) \in CS_T(1-\alpha) \,\Big|\, Y_1,\ldots,Y_T\Big) < 1-\alpha-\epsilon\right) = 0.$$

Proof: See Appendix A.5.

Note that we have assumed that the identified set is non-empty at μ, and we have also showed that under Assumptions 1-4 the probability of observing an empty identified set at $\widehat{\mu}_T$ converges to zero as the sample size grows to infinity. It is of course still possible to observe an empty identified set at a given realization of $\widehat{\mu}_T$. In this case, our algorithm will report a maximum response equal to $-c$ and a minimum response equal to c.[15]

4.3.3. ***Monte-Carlo Evidence***

Frequentist Coverage: We conduct a simple Monte-Carlo exercise to study the coverage probability of our delta-method interval. We set $(1-\alpha) = .68$ implying that $z_{1-\alpha/2} = .9945$. Instead of generating new draws of $(Y_1,\ldots,Y_T)$, we generate $10,000$ draws of $\widehat{\mu}_T$ directly from its asymptotic normal distribution $\mathcal{N}_d(\mu, \Omega/T)$ (where we fix the values of μ and Ω at its estimated values in the UMP example). We decided to proceed in this way in order to 'enforce' the asymptotic normality assumption for $\widehat{\mu}_T$ (which is the key requirement in part a) of Theorem 3). We set $T = 342$ which corresponds to the number of periods in our empirical application.

For each 'draw' of $\widehat{\mu}_T$ (denoted μ^*) we compute the interval

$$\Big[\underline{v}_{k,i,j}(\mu^*) - .9945\,\sigma^*_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\mu^*) + .9945\,\sigma^*_{(k,i,j),T}/\sqrt{T}\Big],$$

[15] In our Matlab implementation, this will generate a warning message asking the user to drop restrictions.

where we treat Ω as known to compute the formula for the standard error $\widehat{\sigma}_{(k,i,j),T}$. We do this to assume away any problem concerning the estimation of Ω (as Theorem 3 assumes that we have a consistent estimator for the asymptotic variance of $\widehat{\mu}_T$).

Finally, we check whether $[\underline{v}_{k,i,j}(\widehat{\mu}_T), \overline{v}_{k,i,j}(\widehat{\mu}_T)]$ is contained in the confidence interval corresponding to each draw μ^* from $\mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T)$. The estimated probability provides a lower bound on the coverage of the identified parameter. The results are reported in Figure 2. We note that the simulated coverage probability lies between 68% and 84% (except for the contemporaneous IRF for FFR which is equal to zero by assumption). The simulated coverage probability is higher than the nominal size of 68%. This is consistent with our theorem, as we are using a standard error that protects against potential violations of full differentiability (even when the function is differentiable at μ).[16]

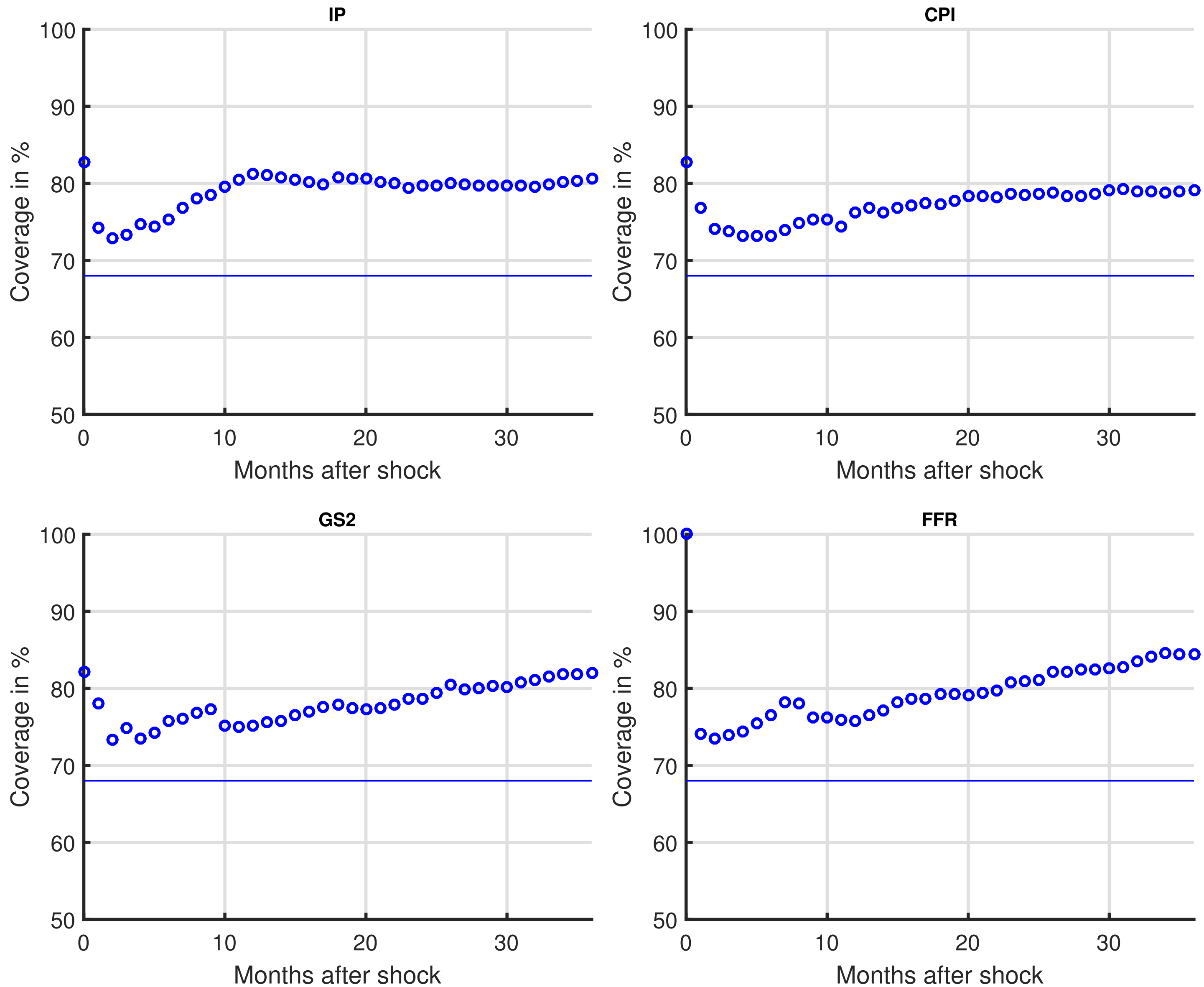
Robust Bayesian Credibility in the UMP application: We also compute the robust Bayesian credibility of our delta-method interval based on an uninformative Normal-Inverse Wishart prior on μ (following Uhlig (2005)). Namely, we generate 10,000 draws of μ^* from the posterior distribution and report the share of draws for which $[\underline{v}_{k,i,j}(\mu^*), \overline{v}_{k,i,j}(\mu^*)]$ is contained in

$$\left[\underline{v}_{k,i,j}(\widehat{\mu}_T) - .9945\, \widehat{\sigma}_{(k,i,j),T}/\sqrt{T},\ \ \overline{v}_{k,i,j}(\widehat{\mu}_T) + .9945\, \widehat{\sigma}_{(k,i,j),T}/\sqrt{T}\right].$$

The results are provided in Figure 3. The simulated credibility is larger or close to the nominal level of 68%, which is consistent with part b of Theorem 3. We also report the robust Bayesian credibility based on the asymptotic normal approximation in Figure 5 in Appendix B.1.

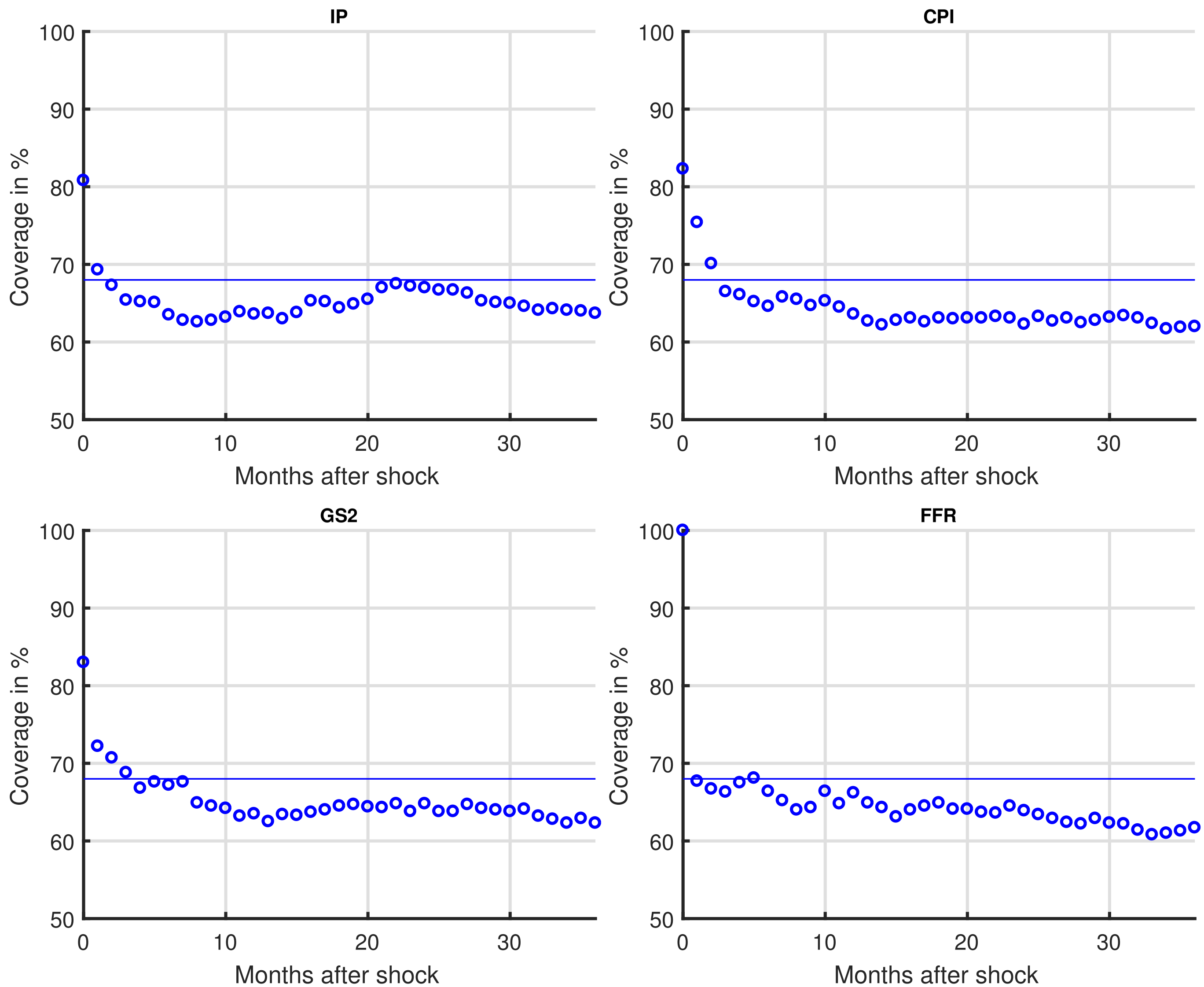
[16] One can use the ideas of Freyberger and Horowitz (2015) to propose an alternative estimator for the standard error which could deliver yet tighter CS. We leave this extension for further research.

Figure 2: Monte-Carlo coverage probability based on the model $\mu^* \sim \mathcal{N}(\widehat{\mu}_T, \widehat{\Omega}_T/T)$, $T = 342$.



(Circles) Monte-Carlo estimate of the probability $P\Big([\underline{v}_{k,i,j}(\widehat{\mu}_T), \overline{v}_{k,i,j}(\widehat{\mu}_T)] \subset [\underline{v}_{k,i,j}(\mu^*) - .9945\, \sigma^*_{(k,i,j),T}/\sqrt{T},\ \overline{v}_{k,i,j}(\mu^*) + .9945\, \sigma^*_{(k,i,j),T}/\sqrt{T}]\Big)$ for the model $\mu^* \sim \mathcal{N}(\widehat{\mu}_T, \widehat{\Omega}_T)$, with $T = 342$. The values $\widehat{\mu}_T$ and $\widehat{\Omega}_T$ correspond, respectively, to the estimators of the reduced-form parameter and its asymptotic covariance matrix in the UMP application. (Solid Line) Nominal confidence level for the delta-method confidence interval (68%).

Figure 3: Robust Bayesian credibility of the delta-method interval based on the posterior distribution corresponding to an uniformative Normal-Inverse Wishart prior on μ^* (as in Uhlig (2005)), $T = 342$.



(Circles) Monte-Carlo estimate of the probability $P^*_\mu\Big([\underline{v}_{k,i,j}(\mu^*), \overline{v}_{k,i,j}(\mu^*)] \subset [\underline{v}_{k,i,j}(\widehat{\mu}_T) - .9945\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}, \overline{v}_{k,i,j}(\widehat{\mu}_T) + .9945\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}] | Y_1, \ldots, Y_T\Big)$ based on the posterior distribution associated to an uniformative Normal-Inverse Wishart prior on μ^* (as in Uhlig (2005)) with $T = 342$. The values $\widehat{\mu}_T$ and $\widehat{\Omega}_T$ correspond, respectively, to the estimators of the reduced-form parameter and its asymptotic covariance matrix in the UMP application. (Solid Line) Nominal level of the delta-method interval (68%).

5. UNCONVENTIONAL MONETARY POLICY SHOCKS

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 2 year Treasury bond rate fell by 10 basis points.

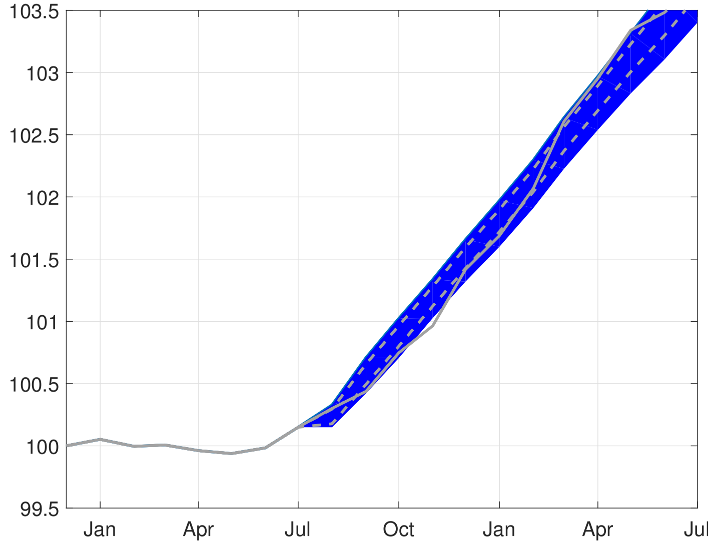
Figure 4 uses our delta-method approach to construct confidence bands for the evolution of the levels of the four variables in the monetary SVAR. We fix all the variables at their level on July 2010 and we trace their evolution (over a 12-month window) according to the confidence set for their cumulative responses. The motivation for this exercise is as follows. Suppose that—back in August 2010—an econometrician is asked to provide confidence bands for the evolution of IP, CPI, 2YTB, and FF after the August 2010 announcement of the Federal Open Market Committee (FOMC). The econometrician observes the realization of the macroeconomic variables from July 1979 until August 2010, but decides to deliberately ignore the two years of data after the crisis (to avoid introducing structural changes, stochastic volatility, or any other feature that will complicate the estimation of the VAR).

The econometrician uses the data until December 2007—one semester before the financial crisis—to conduct delta-method inference on the cumulative responses to a one standard deviation unconventional monetary policy shock. The econometrician then uses these cumulative responses to get a rough idea of the evolution of the variables (in levels) following the announcement of the Federal Reserve in August 2010. The econometrician assumes there is a linear trend for CPI/IP, and ignores sampling uncertainty coming from the trend estimation in reporting the bands.

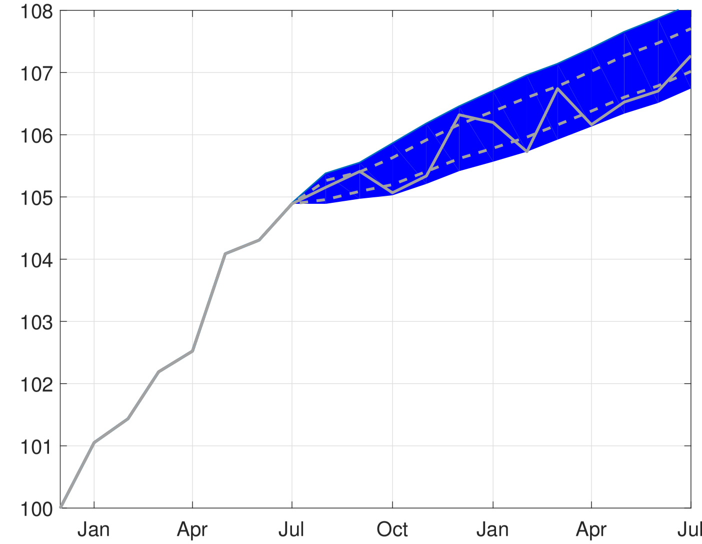
An ex-post evaluation of this exercise (over a window of 12 months) is reported in Figure 4.[17] We note that the observed dynamics for CPI, IP, GS2, and FFR from August 2010 to July 2011 fall within the bounds motivated by our delta-method interval. We also note that our delta-method interval misses the observed value at most three out of 12 months, which means that our 68% confidence set covers each of these variables at least 75% of the time. We also report the 68% Bayesian credible sets.

[17]The reason to focus in a 12-month window is to cover the period between the QE2 announcement and the announcement of the so-called "Operation Twist" in September 2011. See `http://www.federalreserve.gov/newsevents/press/monetary/20110921a.htm`.

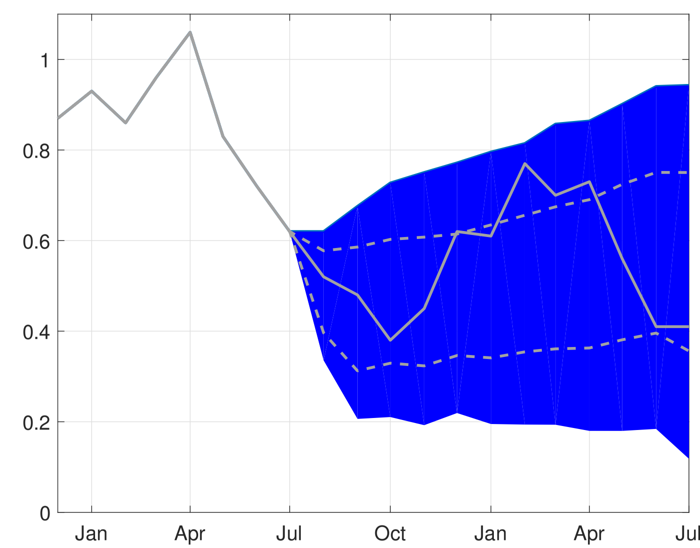
Figure 4: Delta-Method Interval for CPI, IP, 2yTB, FF after the August 2010 announcement



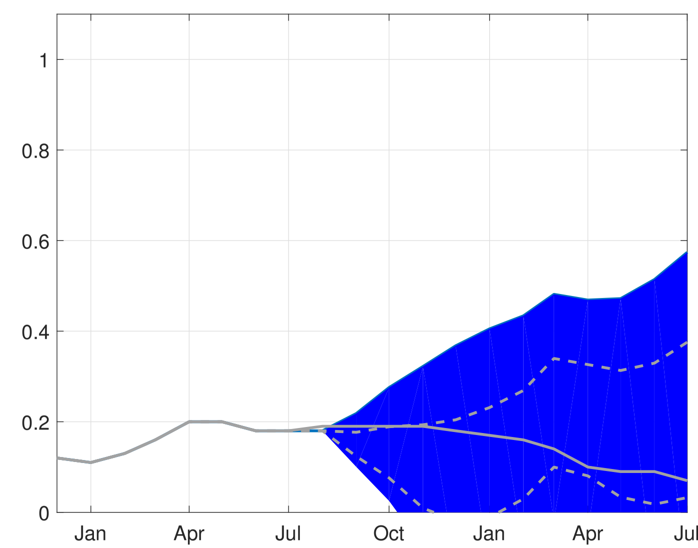
(a) Consumer Price Index



(b) Industrial Production Index



(c) 2-year Treasury Bond Rate



(d) Federal Funds Rate

(Shaded Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Dashed Line) Evolution of the Levels CPI, IP, 2yTB, and FF based on the 68% credible set constructed using the priors in Uhlig (2005).

Computational Cost: We close this section with some comments regarding the computational cost of our delta-method procedure. Most of the work to compute the endpoints of the identified set and its derivatives is analytical. Consequently, practitioners can expect the computational burden of our procedure to be low. We note that the implementation of our delta-method interval in the running example takes only around .15 seconds (using a standard Laptop @2.4GHz IntelCore i7).

Comparison with the Projection Approach: Figure 6 in Appendix B.1 presents a comparison between the delta-method approach and the *projection* approach recently proposed by Gafarov et al. (2016) [GMM16]. The projection approach has two theoretical properties that we were not able to verify for the delta-method. First, projection is consistent in level *uniformly* over a reasonable class of data generating processes. Second, projection yields valid *simultaneous* inference; that is, it covers the whole impulse-response function (across different horizons and different variables) and not only its scalar coefficients.[18] We note that in our application the projection confidence interval (which is wider than the delta-method bands) contains the realized value of IP, CPI, 2YTB, and FF for every horizon under consideration.

Comparison with GK Robust Approach: Figure 7 in the Appendix reports the robust-Bayesian credible set in Giacomini and Kitagawa (2015). The implementation of the robust-Bayes credible set (based on 10,000 posterior draws and using our algorithm to evaluate the endpoints) took around 9,106 seconds.[19]

Comparison with GSM: Figure 9 in the Appendix reports the 68% Bonferroni confidence set of Granziera et al. (2017).[20] Appendix A.7.1 describes the algorithm and related computational issues. The computational cost is approximately 4,000 seconds on a single core machine for 10,000 grid points.

It is hard to provide a general theoretical comparison of the length of the Bonferroni CS and the delta method. The efficiency ranking of the two procedures is likely depend on the particular DGP. One can see that, in our illustrative example, the 68% delta method CS is tighter than the corresponding Bonferroni CS with the same nominal level for almost all combinations of the horizons and time series. One possible explanation behind the larger length of Granziera et al. (2017) is that their procedure is *uniformly* consistent in level over the class of GDPs for which the reduced form impulse response functions converge to a

[18]While our paper focuses on point-wise inference, it is straightforward to provide joint inference by applying Bonferroni correction to the significance level. Figure 8 compares confidence sets that cover not only a single impulse response but the impulse response functions of all variables and all horizons of interest. We compare our Delta-method results when using a Bonferroni-correction with Inoue and Kilian (2013)'s joint Bayes credible set for impulse response functions using the priors for the reduced-form parameters in Uhlig (2005). See Appendix A.7.2.

[19]Out of which 1,266 seconds were used just to compute the identified set for each posterior draw, and the remaining time to translate the posterior bounds into the GK robust bounds

[20]Granziera et al. (2017) also propose a projection-based CS which is a special case of the Bonferroni CS. There is no clear theoretical ranking of the various CS proposed in that paper so we chosen the least computationally intensive variation.

normal distribution. We note that our delta-method is not guaranteed to have this property.

6. CONCLUSION

This paper focused on set-identified structural VAR models that impose equality and inequality restrictions to set-identify only one structural shock. For this class of models, the endpoints of the identified set have special properties that allow an intuitive and computationally simple approach to conduct frequentist and (asymptotic) robust Bayes inference. Specifically, the paper made three contributions:

(i) We presented an algorithm to compute—for each horizon, each variable, a fixed vector of reduced-form parameters, and a given collection of equality and/or inequality restrictions—the largest and smallest value of the coefficients of the structural IRF (see Theorem 1). Our algorithm did not require random sampling from the space of orthogonal matrices or unit vectors. Instead, we treated the bounds of the identified set as the *maximum and minimum value* of a mathematical program whose solutions we were able to characterize analytically. Our algorithm can be used outside our delta-method framework (for example, in computing the maximum and minimum response for the Giacomini and Kitagawa (2015) robust Bayes approach).

(ii) We provided sufficient conditions under which the largest and smallest value of the structural parameters are directionally differentiable functions of the reduced-form parameters (see Theorem 2). This result also seems to be of interest in its own right and could be used to explore the frequentist properties of the robust-Bayesian procedure in Giacomini and Kitagawa (2015).

(iii) Finally, we proposed a computationally convenient delta-method approach to conduct inference for the set-identified coefficients of the structural IRF. We presented sufficient conditions to guarantee the point-wise consistency in level and asymptotic robust Bayes credibility of our suggested inference approach. We note that the delta-method in this paper exploited the structure of the directional derivative.

We illustrated our results by set-identifying the responses of different U.S. macroeconomic variables to an unconventional monetary policy shock. We used the theory and methods developed in this paper to assess the effects of the announcement of the second part of the Quantitative Easing program in August 2010.

ACKNOWLEDGEMENTS

We would like to thank seminar participants at the econometrics workshop at Bonn, Brown, Cornell, ITAM, Michigan, Ohio State University, Pennsylvania State University, Rutgers, UCSD, Vanderbilt, Wisconsin-Madison, the 2014 NSF-Time Series Conference, and the 2016 summer meetings of the Econometric Society (Europe and North America) for helpful comments on an earlier draft of this paper. We would also like to thank Lutz Kilian, Oliver Linton, Andres Santos, and two anonymous referees for extremely helpful comments and suggestions. Bulat Gafarov gratefully acknowledges support from the Basic Research Program of the National Research University Higher School of Economics. The usual disclaimer applies.

REFERENCES

An, L. and J. Wang (2012): "Exchange rate pass-through: Evidence based on vector autoregression with sign restrictions," *Open Economies Review*, 23, 359–380.

Arias, J., J. F. Rubio-Ramirez, and D. F. Waggoner (2017): "Inference Based on SVAR Identified with Sign and Zero Restrictions: Theory and Applications," *Working paper, Emory 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 (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 5, 1963–1999.

Beaudry, P., D. Nam, and J. Wang (2011): "Do mood swings drive business cycles and is it rational?" *NBER Working Paper (w17651).*

Benkwitz, A., M. H. Neumann, and H. Lütekpohl (2000): "Problems related to confidence intervals for impulse responses of autoregressive processes," *Econometric Reviews*, 19, 69–103.

Bhattarai, S., G. B. Eggertsson, and B. Gafarov (2015): "Time Consistency and the Duration of Government Debt: A Signalling Theory of Quantitative Easing," *NBER Working Paper (w21336).*

Blanchard, O. J. and D. Quah (1989): "The dynamic effects of aggregate demand and supply disturbances," *The American Economic Review*, 79, 655–673.

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

Dolado, J. J. and H. Lütkepohl (1996): "Making Wald tests work for cointegrated VAR systems," *Econometric Reviews*, 15, 369–386.

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

Fang, Z. and A. Santos (2015): "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.

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. L. Horowitz (2015): "Identification and shape restrictions in nonparametric instrumental variables estimation," *Journal of Econometrics*, 189, 41–53.

Fujita, S. (2011): "Dynamics of worker flows and vacancies: evidence from the sign restriction approach," *Journal of Applied Econometrics*, 26, 89–121.

Gafarov, B., M. Meier, and J. L. Montiel Olea (2016): "Projection Inference for Set-Identified SVARs," *Working paper, Columbia University.*

Gertler, M. and P. Karadi (2015): "Monetary Policy Surprises, Credit Costs and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Ghosal, S., J. K. Ghosh, and T. Samanta (1995): "On convergence of posterior distributions," *The Annals of Statistics*, 2145–2152.

Giacomini, R. and T. Kitagawa (2015): "Robust Inference about partially identified SVARs," *Working Paper, University College London.*

Granziera, E., H. R. Moon, and F. Schorfheide (2017): "Inference for VARs Identified with Sign Restrictions," .

Gustafson, P. (2009): "What are the limits of posterior distributions arising from nonidentified models, and why should we care?" *Journal of the American Statistical Association*, 104, 1682–1695.

Hong, H. and J. Li (2017): "The numerical delta-method," *Working Paper, Stanford University.*

Inoue, A. and L. Kilian (2002): "Bootstrapping autoregressive processes with possible unit roots," *Econometrica*, 70, 377–391.

——— (2013): "Inference on impulse response functions in structural VAR models," *Journal of Econometrics*, 177, 1–13.

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.

Kitagawa, T., J. Payne, and J. L. Montiel Olea (2017): "Posterior distribution of nondifferentiable functions," *Working paper, Columbia University.*

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

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.

——— (2007): *New Introduction to Multiple Time Series Analysis*, Springer.

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

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

Nocedal, J. and S. Wright (2006): *Numerical optimization*, Springer Science & Business Media, second edition ed.

Ok, E. A. (2007): *Real analysis with economic applications*, vol. 10, Princeton University Press.

Poirier, D. J. (1998): "Revising beliefs in nonidentified models," *Econometric Theory*, 14, 483–509.

Rubio-Ramirez, J., D. Caldara, and J. Arias (2015): "The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure," *Working Paper, Board of Governors of the Federal Reserve*.

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

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in linear time series models with some unit roots," *Econometrica*, 113–144.

Toda, H. Y. and T. Yamamoto (1995): "Statistical inference in vector autoregressions with possibly integrated processes," *Journal of Econometrics*, 66, 225–250.

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.

Vargas-Silva, C. (2008): "Monetary policy and the US housing market: A VAR analysis imposing sign restrictions," *Journal of Macroeconomics*, 30, 977–990.

APPENDIX A: MAIN RESULTS

A.1. *Lemma 1*

We now show that Assumptions 1 and 2 imply that given a collection $r \in \mathbb{R}^{n \times m}$ of 'active' constraints ($m \leq n-1$) the maximum response is determined in closed-form (and up to sign) by the Karush-Kuhn-Tucker conditions of programs (2.5) and (2.6). The following Lemma constitutes the basis of Theorem 1.

Lemma 1 *Suppose that Assumptions 1 and 2 hold. Let r be a matrix of dimension $n \times m$ collecting the gradients of the 'active' (binding) constraints at a solution $x^*(\mu)$ of the mathematical program (2.5), then :*

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

$$\overline{v}_{k,i,j}(\mu) = \Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}, \tag{A.1}$$

or

$$\overline{v}_{k,i,j}(\mu) = -\Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}, \tag{A.2}$$

where

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

b) If in addition $\overline{v}_{k,i,j}(\mu) \neq 0$, then $x^*(\mu)$ is given by

$$x^*(\mu) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \overline{v}_{k,i,j}(\mu).$$

Consequently, the sign of $\overline{v}_{k,i,j}(\mu)$ depends on which of the two values of $x^*(\mu)$ in the equation above (the one with (A.1) in the denominator or the one with (A.2)) satisfies the sign restrictions that are not in r.

Proof: Let $S(\mu)$ denote the $n \times m_s$ matrix of m_s 'sign' restrictions and let $Z(\mu)$ denote the $n \times m_z$ matrix of 'zero' restrictions. For notational simplicity, we deliberately ignore the dependence of the equality/inequality restrictions on μ. The problem in equation (2.5) is equivalent to

$$\overline{v}_{k,i,j}(\mu) \equiv \max_{x \in \mathbb{R}^n} e_i'C_k(A)x \quad \text{subject to} \;\; x'\Sigma^{-1}x = 1, \quad S'x \geq \mathbf{0}_{m_s \times 1}, \quad Z'x = \mathbf{0}_{m_z \times 1}. \tag{A.3}$$

The auxiliary Lagrangian function is given by

$$\mathcal{L}(x, \lambda, w_1, w_2; \mu) = e_i'C_kx - \lambda(x'\Sigma^{-1}x - 1) - w_1'(S'x) - w_2'(Z'x).$$

Assumptions 1–2 imply that we can characterize the maximum response using the Karush-Kuhn-Tucker conditions for the mathematical program in (2.5). The Karush-Kuhn-Tucker necessary conditions for this problem are as follows:

$$\begin{aligned}
\text{Stationarity} \quad &: \quad C_k'(A)e_i - 2\lambda\Sigma^{-1}x - Sw_1 - Zw_2 = \mathbf{0}_{n \times 1}, \\
\text{Primal Feasibility} \quad &: \quad x'\Sigma^{-1}x = 1, \\
& \quad\;\; S'x \geq \mathbf{0}_{m_s \times 1}, \\
& \quad\;\; Z'x = \mathbf{0}_{m_z \times 1}, \\
\text{Complementary Slackness} \quad &: \quad w_{1i}(e_i'Sx) = 0 \quad \forall \quad i = 1 \ldots m_s,
\end{aligned}$$

plus the additional dual feasibility constraint requiring the Lagrange multipliers, w_{1i}, to be smaller than or equal to zero.

Let $x^*(\mu)$ be one (out of possibly many) maximizers of the program of interest and suppose that the $n \times m$ matrix r collects all the restrictions that are active (binding). Because of Assumption 1 and 2 , the matrix r is of full rank m and m must be smaller than or equal $n-1$. Using Stationarity, Primal Feasibility, and Complementary Slackness at x^* we get

$$\begin{aligned} 0 = x^{*\prime}[C_k(A)'e_i - 2\lambda^*\Sigma^{-1}x^* - Sw_1 - Zw_2] &= x^{*\prime}C_k(A)'e_i - 2\lambda^* x^{*\prime}\Sigma^{-1}x^* - x^{*\prime}Sw_1 - x^{*\prime}Zw_2 \\ &= x^{*\prime}C_k(A)'e_i - 2\lambda^* - x^{*\prime}Sw_1 - x^{*\prime}Zw_2 \\ &\quad \text{(where we have used } x^{*\prime}\Sigma^{-1}x^* = 1) \\ &= x^{*\prime}C_k(A)'e_i - 2\lambda^* \\ &\quad \text{(where we have used } x^{*\prime}Z = \mathbf{0}_{m_z\times 1} \text{ and complementary slackness)} \\ &= \overline{v}_{k,i,j}(\mu) - 2\lambda^*, \end{aligned}$$

where $\overline{v}_{k,i,j}(\mu)$ denotes the value of the maximum response when the constraints in r are active. Thus, the Lagrange multiplier λ^* is unique and given by

$$\lambda^* = \frac{1}{2}\overline{v}_{k,i,j}(\mu).$$

Note also that $\lambda^* \neq 0$ if and only if $\overline{v}_{k,ij}(\mu; r) \neq 0$. We now show that there are unique w_1^* and w_2^* that satisfy the Karush-Kuhn Tucker conditions. Let w^* denote the nonzero components of w_1^* and all the components of w_2^*. Note that left multiplying the stationarity condition by Σ and rearranging the terms we have :

$$2\lambda^* x^{*\prime} = \Big(C_k(A)'e_i - rw^*\Big)'\Sigma,$$

$$\begin{aligned} \Big(C_k(A)'e_i - rw^*\Big)'\Sigma\Big(C_k(A)'e_i - rw^*\Big) &= 4(\lambda^*)^2 x'\Sigma^{-1}x \\ &= 4(\lambda^*)^2 \\ &\quad \text{(where we have used } \; x^{*\prime}\Sigma^{-1}x^* = 1) \\ &= 4\Big(\frac{1}{2}\overline{v}_{k,i,j}(\mu)\Big)^2 \\ &= \overline{v}_{k,i,j}(\mu)^2. \end{aligned} \tag{A.4}$$

Consequently the value function given active constraints r is given by either

$$\overline{v}_{k,i,j}(\mu) = \Big[\Big(C_k(A)'e_i - rw^*\Big)'\Sigma\Big(C_k(A)'e_i - rw^*\Big)\Big]^{1/2},$$

or

$$\overline{v}_{k,i,j}(\mu) = -\Big[\Big(C_k(A)'e_i - rw^*\Big)'\Sigma\Big(C_k(A)'e_i - rw^*\Big)\Big]^{1/2}.$$

We will use the first order conditions to find the vector of Lagrange multipliers w^* and show that they are unique. Note that

$$\begin{aligned} 0 = 2\lambda^* r'x^* &= \Big[r'\Sigma(C_k(A)'e_i - rw^*)\Big] \\ &= \Big[r'\Sigma C_k(A)'e_i - r'\Sigma r w^*\Big]. \end{aligned}$$

Under the assumptions of the lemma, r is of rank m. The equation above holds if and only if

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

Consequently, the Lagrange multipliers for the active restrictions are unique. To conclude the proof, we get an explicit expression of the value function in terms of μ. To do so, note that

$$\begin{aligned}
\Sigma^{1/2}\Big(C_k(A)'e_i - rw^*\Big) &= \Sigma^{1/2}C_k(A)'e_i - \Sigma^{1/2}rw^* \\
&= \Sigma^{1/2}C_k(A)'e_i - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma C_k(A)'e_i \\
&= \Big(\mathbb{I}_n - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma^{1/2}\Big)\Sigma^{1/2}C_k(A)'e_i \\
&= \Big(\mathbb{I}_n - P_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \\
&= M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i.
\end{aligned}$$

Therefore, the equation above and (A.4) imply that either

$$\overline{v}_{k,i,j}(\mu) = \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}$$

or

$$\overline{v}_{k,i,j}(\mu) = -\Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}.$$

Furthermore, since any solution for which r is the set of binding constraints satisfies $2\lambda^* x^{*\prime} = (C_k(A)'e_i - rw^*)'\Sigma$, then for any $\overline{v}_{k,i,j}(\mu) \neq 0$ the solution x^* should be given by either

$$x^* = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2},$$

or

$$x^* = -\Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}.$$

In any case the Lagrange multipliers for the active constraints are given (as shown above) by,

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

A.2. *Proof of Theorem 1*

The choice set of program (2.5) is non-empty (by Assumption 1) and compact (because of the ellipsoid constraint $BB' = \Sigma$). Hence, the maximum exists. Let $x^* \in \mathbb{R}^n$ be a solution and let r^* be the set of constraints that are active at x^*.

Step 1: We show first that

$$\overline{v}_{k,i,j}(\mu) \geq \max_{r\in R}\Big(\max\{f^+_{max}(\mu; r), f^-_{max}(\mu; r)\}\Big).$$

We do so by considering two different cases.

Case 1.1: Take any $r \in R$, and assume first that $v_{k,i,j}(\mu; r) \neq 0$. If $\mathbf{1}_{m_s}(x^*_+(\mu; r)) = 0$, then

$$f^+_{max}(\mu; r) = v_{k,i,j}(\mu; r) - 2c \leq c - 2c = -c < \overline{v}_{k,i,j}(\mu),$$

where the first equality above follows from the definition of f^{+}_{max} and the two remaining inequalities follow from the definition of the penalty term c.

Note, however, that if $r \in R$ is such that $\mathbf{1}_{m_s}(x^{*}_{+}(\mu;r)) = 1$, then $x^{*}_{+}(\mu;r)$ satisfies all the equality and inequality restrictions in (2.5) and, by construction, also satisfies the ellipsoid constraint

$$x^{*}_{+}(\mu;r)'\Sigma^{-1}x^{*}_{+}(\mu;r) = 1.$$

Consequently, $\overline{v}_{k,i,j}(\mu) \geq f^{+}_{max}(\mu;r)$ for all $r \in R$. An analogous argument shows that $\overline{v}_{k,i,j}(\mu) \geq f^{-}_{max}(\mu;r)$. This implies that

$$\overline{v}_{k,i,j}(\mu) \geq \max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\},$$

for all $r \in \mathbb{R}$ such that $v_{k,i,j}(\mu;r) \neq 0$.

Case 1.2: Consider now any r such that $v_{k,i,j}(\mu;r) = 0$. If there is no feasible point x^{*} that gives such a value, then $f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = -2c < \overline{v}_{k,i,j}(\mu)$. If there is such a feasible point $x^{*} \neq 0$ then $f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = 0$. Since x^{*} is in the choice set of the program (2.5), then $f^{+}_{max}(\mu;r) = f^{-}_{max}(\mu;r) = 0 \leq \overline{v}_{k,i,j}(\mu)$.

Therefore, Case 1.1 and 1.2 imply that

$$\overline{v}_{k,i,j}(\mu) \geq \max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\} \quad \text{for all } r \in R.$$

Step 2: We now show that

$$\overline{v}_{k,i,j}(\mu) \leq \max_{r \in R}\Big(\max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\}\Big).$$

Again, we consider two cases.

Case 2.1: Assume first that $\overline{v}_{k,i,j}(\mu) \neq 0$. Without loss of generality, let us assume that $\overline{v}_{k,i,j}(\mu) > 0$ (the case in which $\overline{v}_{k,i,j}(\mu) < 0$ is completely analogous). Let $r^{*} \in R$ denote the set of active restrictions (which by Assumptions 1 and 2 has at most $n-1$ columns) at the solution x^{*} (this is one out of the potentially many solutions to the program). By Lemma 1 we know that

$$\overline{v}_{k,i,j}(\mu) = \Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r^{*}}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2},$$

and

$$x^{*}(\mu;r^{*}) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r^{*}}\Big)\Sigma^{1/2}C_k(A)'e_i\Big/\Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r^{*}}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}.$$

Since this point satisfies the sign restrictions not in r^{*} (because it is a solution), then

$$\Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r^{*}}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2} = f^{+}_{max}(\mu;r^{*}).$$

Consequently,

$$\overline{v}_{k,i,j}(\mu) = f^{+}_{max}(\mu;r^{*}) \leq \max_{r \in R}\Big(\max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\}\Big).$$

Case 2.2: If $\overline{v}_{k,i,j}(\mu) = 0$, there is an $x^{*} \neq 0$ in the choice set. Hence, the Karush-Kuhn-Tucker conditions imply that $C_k(A)'e_i$ is a linear combination of the active constraints that generate the value of zero (which means, by definition of the algorithm, that there is an r^{*} such that $f^{+}_{max}(\mu;r^{*}) = f^{-}_{max}(\mu;r^{*}) = 0$). Therefore, $\overline{v}_{k,i,j}(\mu) = f(\mu;r^{*}) \leq \max_{r \in R}\Big(\max\{f^{+}_{max}(\mu;r), f^{-}_{max}(\mu;r)\}\Big)$.

As the result, the value function $\overline{v}_{k,i,j}(\mu)$ is obtained by computing the Karush-Kuhn-Tucker points in Lemma 1 for each r, penalizing the value $\overline{v}_{k,i,j}(\mu; r)$ if not feasible, and maximizing over all the possible values of r.

The proof for the lower bound is analogous;

$$\underline{v}_{k,i,j}(\mu) = \min_{r\in R}\Bigg(\min\{f^{+}_{min}(\mu;r), f^{-}_{min}(\mu;r)\}\Bigg),$$

with:

$$\begin{aligned} f^{+}_{min}(\mu;r) &\equiv v_{k,i,j}(\mu;r) + 2(1-\mathbf{1}_{m_s}(x^*_+(\mu;r)))c, \\ f^{-}_{min}(\mu;r) &\equiv -v_{k,i,j}(\mu;r) + 2(1-\mathbf{1}_{m_s}(x^*_-(\mu;r)))c. \end{aligned}$$

A.3. *Lemma 2*

Lemma 2 *Suppose that Assumptions 1-3 hold. Let $r(\mu)$ be a matrix of dimension $n \times l$ collecting the gradients of the 'active' (binding) constraints at a solution $x^*(\mu)$ of the mathematical program (2.5) such that $v_{k,i,j}(\mu; r(\mu)) \neq 0$. Then $v_{k,i,j}(\mu; r(\mu))$ is differentiable with respect to μ with the derivative $\dot{v}_{k,i,j}(\mu; r(\mu))$ given by*

$$\begin{bmatrix} \frac{\partial v_{k,i,j}(\mu;r(\mu))}{\partial vec(A)} \\ \\ \frac{\partial v_{k,i,j}(\mu;r(\mu))}{\partial vec(\Sigma)} \end{bmatrix} = \begin{bmatrix} \frac{\partial vec(C_k(A))}{\partial vec(A)}(x^*(\mu;r(\mu))\otimes e_i) - \sum_{k=1}^{l} w^*_k \frac{\partial vec(r_k(\mu))}{\partial vec(A)} x^*(\mu;r(\mu)) \\ \\ \lambda^*(\Sigma^{-1}x^*(\mu;r(\mu))\otimes \Sigma^{-1}x^*(\mu;r(\mu))) - \sum_{k=1}^{l} w^*_k \frac{\partial vec(r_k(\mu))}{\partial vec(\Sigma)} x^*(\mu;r(\mu)) \end{bmatrix},$$

where $r_k(\mu)$ denotes the k-th column of $r(\mu)$,

$$x^*(\mu;r(\mu)) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r(\mu)}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu;r(\mu)),$$

$$\lambda^* \equiv \frac{1}{2}v_{k,i,j}(\mu;r(\mu)), \quad w^* \equiv [r(\mu)'\Sigma r(\mu)]^{-1} r(\mu)'\Sigma C_k(A)e_i,$$

and w^*_k is the k-th component of the vector w^*.

Proof: Note first that Assumption 3 implies that $r \equiv r(\mu)$ is differentiable with respect to μ. Moreover, since $v_{k,i,j}(\mu; r) \neq 0$ the function

$$v_{k,i,j}(\mu;r) = \Big(e_i' C_k(A)\Sigma^{1/2} M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}$$

is differentiable as well. Moreover, the function

$$x^*(\mu;r) \equiv \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu;r)$$

is also differentiable. Therefore,

$$\begin{aligned} \frac{dv_{k,i,j}(\mu;r)}{d\mu} &= \frac{d[e_i'C_k(A)x^*(\mu;r)]}{d\mu} \\ &\quad \text{(since } v_{k,i,j}(\mu;r) = e_i'C_k(A)x^*(\mu;r)\text{)} \\ &= \frac{dx^*(\mu;r)}{d\mu}C_k'(A)e_i + \frac{d(x^*(\mu;r)'\otimes e_i')\text{vec}(C_k(A))}{d\mu}, \\ &\quad \text{(where we have re-written } e_i'C_k(A)x^* \text{ as } (x^{*\prime}\otimes e_i')\text{vec}(C_k(A))\text{)} \\ &= \frac{dx^*(\mu;r)}{d\mu}C_k'(A)e_i + \frac{d\text{vec}(C_k(A))}{d\mu}(x^*(\mu;r)\otimes e_i) \\ &\quad \text{(where we have applied the chain rule for matrix derivatives).} \end{aligned}$$

We now use the envelope theorem to compute this derivative. Note that —using Assumptions 1 and 2— Lemma 1 shows the existence of unique multipliers $\lambda^* \in \mathbb{R}$ and $w^* \in \mathbb{R}^l$ such that

$$C_k(A)' e_i = \lambda^* 2\Sigma^{-1} x^*(\mu; r) + r w^*.$$

Therefore,

$$\frac{dv_{k,i,j}(\mu; r)}{d\mathrm{vec}(A)} = \frac{dx^*(\mu; r)}{d\mathrm{vec}(A)} \Big[\lambda^* 2\Sigma^{-1} x^*(\mu; r) + r w^* \Big] + \frac{d\mathrm{vec}(C_k(A))}{d\mathrm{vec}(A)} (x^*(\mu; r) \otimes e_i)$$

and

$$\frac{dv_{k,i,j}(\mu; r)}{d\mathrm{vec}(\Sigma)} = \frac{dx^*(\mu; r)}{d\mathrm{vec}(\Sigma)} \Big[\lambda^* 2\Sigma^{-1} x^*(\mu; r) + r w^* \Big] + \frac{d\mathrm{vec}(C_k(A))}{d\mathrm{vec}(\Sigma)} (x^*(\mu; r) \otimes e_i).$$

Note also that because $x^*(\mu, r)$ satisfies the ellipsoid constraint

$$0 = \frac{dx^*(\mu; r)' \Sigma^{-1} x^*(\mu; r)}{d\mathrm{vec}(A)} = 2 \frac{dx^*(\mu; r)}{d\mathrm{vec}(A)} \Sigma^{-1} x^*(\mu; r)$$

and, also, since the equality constraints are met,

$$\begin{aligned} 0 &= \frac{dr(\mu)' x^*(\mu; r(\mu))}{d\mathrm{vec}(A)} \\ &= \frac{dx^*(\mu; r)}{d\mathrm{vec}(A)} r(\mu) + \left(\tfrac{dr_1(\mu)}{d\mathrm{vec}(A)} x^*(\mu; r(\mu)), \quad \cdots \quad , \tfrac{dr_l(\mu)}{d\mathrm{vec}(A)} x^*(\mu; r(\mu)) \right), \end{aligned}$$

where $r_k(\mu)$ denotes the k-th column of $r(\mu)$. Consequently,

$$\frac{dv_{k,i,j}(\mu; r)}{d\mathrm{vec}(A)} = \frac{d\mathrm{vec}(C_k(A))}{d\mathrm{vec}(A)} (x^*(\mu; r) \otimes e_i) - \sum_{k=1}^{l} w_k^* \frac{d\mathrm{vec}(r_k(\mu))}{d\mathrm{vec}(A)} x^*(\mu; r),$$

where w_k^* is the k-th entry of the vector of lagrange multipliers w^*. This gives the partial derivative of $v_{k,i,j}(\mu; r_l(\mu))$ with respect to vec(A). We note that this derivative can also be written as

$$\frac{dv_{k,i,j}(\mu; r)}{d\mathrm{vec}(A)} = \frac{d\mathrm{vec}(C_k(A))}{d\mathrm{vec}(A)} (x^*(\mu; r) \otimes e_i) - \frac{d\mathrm{vec}(r(\mu)')}{d\mathrm{vec}(A)} (x^*(\mu; r) \otimes \mathbb{I}_l) w^*,$$

which is the expression given in the overview. Finally, to get the derivative with respect to vec(Σ) we note that

$$0 = \frac{dx^*(\mu; r)' \Sigma^{-1} x^*(\mu; r)}{d\mathrm{vec}(\Sigma)} = 2 \frac{dx^*(\mu; r)}{d\mathrm{vec}(\Sigma)} \Sigma^{-1} x^*(\mu; r) - (\Sigma^{-1} x^*(\mu; r) \otimes \Sigma^{-1} x^*(\mu; r)),$$

and

$$\begin{aligned} 0 &= \frac{dr(\mu)' x^*(\mu; r(\mu))}{d\mathrm{vec}(\Sigma)} \\ &= \frac{dx^*(\mu; r(\mu))}{d\mathrm{vec}(\Sigma)} r(\mu) + \left(\tfrac{dr_1(\mu)}{d\mathrm{vec}(\Sigma)} x^*(\mu; r(\mu)), \quad \cdots \quad , \tfrac{dr_l(\mu)}{d\mathrm{vec}(\Sigma)} x^*(\mu; r(\mu)) \right). \end{aligned}$$

Consequently,

$$\frac{dv_{k,i,j}(\mu; r)}{d\mathrm{vec}(\Sigma)} = \lambda^* (\Sigma^{-1} x^*(\mu; r) \otimes \Sigma^{-1} x^*(\mu; r)) - \sum_{k=1}^{l} w_k^* \frac{d\mathrm{vec}(r_{k,l}(\mu))}{d\mathrm{vec}(\Sigma)} x^*(\mu; r).$$

A.4. *Proof of Theorem 2*

Structure of the proof: The proof proceeds in five steps. First, we show that Assumptions 1 and 2 imply that the choice set of program (2.5) is non-empty for any $\tilde{\mu}$ in a neighborhood of μ. Second, we show that the choice set of program (2.5) is both lower and upper-hemicontinuous correspondence at μ. Third, we use the continuity of the choice set and the Maximum theorem to establish continuity of $\overline{v}_{k,i,j}(\cdot)$ at μ. Fourth, we use Lemma 1 and the continuity of $\overline{v}_{k,i,j}(\cdot)$ to show that

$$\max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\} \leq \liminf_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

Finally, we use Lemma 1, Theorem 1, and Lemma 2 to show (by contradiction) that

$$\limsup_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) \leq \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}.$$

Step 1 : By Assumption 1 there is a point $x^* \in \mathbb{R}^n$ that belongs to the choice set of program (2.5). Let $Z^*(\mu) \in \mathbb{R}^{n\times m_e}$ denote the restrictions in program (2.5) that are active at x^*. By Assumption 2, we know that $m_e \leq n-1$. Let $S^*(\mu) \in \mathbb{R}^{n\times m_i}$ denote all the other restrictions that are not in $Z^*(\mu)$. This means that $S^*(\mu)'x^* > 0_{m_i\times 1}$ (since these restrictions are not in $Z^*(\mu)$). Note first that Assumption 2 implies there is $\epsilon_1 > 0$ such that $\lambda_{min}(\mu) \equiv \min \text{eig}(Z^*(\mu)'Z^*(\mu)) > \epsilon_1$. Since x^* is feasible we can also pick ϵ_2 such that $(s_m^*(\mu)/||s_m^*(\mu)||)'x^*(\mu)$ is larger than ϵ_2 for each $m \in \{1, 2, \ldots, m_i\}$. Define

$$U(\mu) \equiv \{\tilde{\mu} \quad | \quad \lambda_{min}(\tilde{\mu}) > \epsilon_1, \quad (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'x^* > \epsilon_2 \;\; \forall\, m, \quad ||Z^*(\tilde{\mu})'x^*|| < \sqrt{\epsilon_1}\epsilon_2/2\} \cap \mathcal{M}.$$

By construction $\mu \in U(\mu)$. Moreover, the continuity of $Z(\cdot)$ and $S(\cdot)$ and openness of $\mathcal{M}$ implies that $Z^*(\cdot)$ and $S^*(\cdot)$ are continuous and therefore $U(\mu)$ is open. We now show that for every $\tilde{\mu} \in U(\mu)$ there is $\tilde{x} \in \mathbb{R}^d$ that satisfies the equality restrictions in $Z^*(\tilde{\mu})$ and also the inequalities in $S^*(\tilde{\mu})$ with slack. To formalize this point, define

$$\tilde{x} \equiv \tilde{x}(\tilde{\mu}, \mu) \equiv x^* - N_{Z^*(\tilde{\mu})}x^*, \tag{A.5}$$

where $N_{Z^*(\tilde{\mu})} \equiv Z^*(\tilde{\mu})(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}Z^*(\tilde{\mu})'$ is well defined because $\lambda_{min}(\tilde{\mu}) > \epsilon_1$. Note first that, by construction,

$$Z^*(\tilde{\mu})'\tilde{x} = Z^*(\tilde{\mu})'x^* - Z^*(\tilde{\mu})'N_{Z^*(\tilde{\mu})}x^* = Z^*(\tilde{\mu})'x^* - Z^*(\tilde{\mu})'Z^*(\tilde{\mu})(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}Z^*(\tilde{\mu})'x^* = 0_{m_e\times 1},$$

implying that the equality restrictions at $Z^*(\tilde{\mu})$ are satisfied by $\tilde{x}$. Thus, we only need to show that the inequalities in $s_m^*(\tilde{\mu})$ are satisfied with slack (after normalizing by its norm). To see this, simply note that

$$\begin{aligned}
(s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'\tilde{x} &= (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'(\tilde{x} - x^*) + (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'x^* \\
&> (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'(\tilde{x} - x^*) + \epsilon_2 \\
&\geq -|s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'(\tilde{x} - x^*)| + \epsilon_2 \\
&\geq -||(\tilde{x} - x^*)|| + \epsilon_2.
\end{aligned}$$

But

$$\begin{aligned}
||\tilde{x} - x^*|| &= (x^{*\prime}Z^*(\tilde{\mu})(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}Z^*(\tilde{\mu})x^*)^{1/2} \\
&\leq \sup_{\omega \text{ s.t. } ||\omega||=1} (\omega(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}\omega)^{1/2}||Z^*(\tilde{\mu})x^*|| \\
&= ||Z^*(\tilde{\mu})x^*||\sqrt{\lambda_{\min}(\tilde{\mu})} \\
&\leq (\sqrt{\epsilon_1}\epsilon_2/2\sqrt{\epsilon_1}) \\
&= \epsilon_2/2.
\end{aligned}$$

This implies that $s_m^*(\tilde{\mu})'\tilde{x} > 0$ for every $m \in \{1, 2, \ldots, m_i\}$. This shows that for every $\tilde{\mu} \in U(\mu)$, $\tilde{x} \in \mathcal{R}(\tilde{\mu})$. To complete Step 1, notice that $x^\dagger \equiv \tilde{x}/(\tilde{x}'\tilde{\Sigma}^{-1}\tilde{x}) \in \mathcal{R}(\tilde{\mu})$. By construction, $x^\dagger$ is in the choice set of program 2.5 evaluated at $\tilde{\mu}$.

Step 2 : Let the multivalued correspondence $\Gamma(\cdot) : \mathcal{M} \to \mathbb{R}^n$ be defined as the choice set of program (2.5). We show continuity of this correspondence at μ by showing that it is both lower and upper hemicontinuous.

To establish upper hemicontinuity, pick any sequence $\mu_N \in \mathcal{M}$ s.t. $\mu_N \to \mu$ and any converging sequence $x_N \in \Gamma(\mu_N)$ s.t. $x_N \to x^*$. Consider any sign restriction $s(\mu_N)$. By construction, $s(\mu_N)'x_N \geq 0$. By continuity of $s(\mu_n)$, we get in the limit $s(\mu)'x^* \geq 0$. Similarly, $(x^*)'\Sigma^{-1}(x^*) = 1$ and for any zero restriction $z(\mu)'x^* = 0$. The set $\Gamma(\mu)$ is compact, so by Theorem 2 on p. 218 in Ok (2007), $\Gamma(\cdot)$ is upper hemicontinuous at μ.

To establish lower hemicontinuity, consider any sequence $\mu_N \in \mathcal{M}$ s.t. $\mu_N \to \mu$ and any point $x^* \in \Gamma(\mu)$. Then, by Step 1, the elements of the sequence defined as

$$x_N \equiv \tilde{x}(\mu_N, \mu)/(\tilde{x}(\mu_N, \mu)'\Sigma^{-1}\tilde{x}(\mu_N, \mu))$$

belong to $\Gamma(\mu_N)$. By continuity of $Z^*(\cdot)$ and Σ^{-1} at $\mu \in \mathcal{M}$ (implied by Assumption 3) and using the invertibility of the matrices $(Z^*(\tilde{\mu_N})'Z^*(\tilde{\mu_N}))$ for N large enough (implied by Assumption 2) we have $x_N \to x^*$. By Proposition 4 on p. 224 in Ok (2007), $\Gamma(\mu)$ is lower hemicontinuous. By definition, it is continuous at μ.

Step 3 : Let $(\Theta, \rho) \equiv (U(\mu), \rho)$ be a metric space with Euclidean metric $\rho(\cdot)$. By Steps 1 and 2, the choice set of the program in (2.5) is a non-empty, compact-valued, continuous correspondence at μ. By the Maximum theorem, see p. 229 in Ok (2007), $\overline{v}_{k,i,j}(\cdot)$ is continuous at μ.

Additional Notation: Consider any sequence $\mu_N = (\text{vec}A_N{}', \text{vec}\Sigma_N{}')'$ such that

$$\mu_N = \mu + h_N/t_N,$$

where $h_N \to h \in \mathbb{R}^d$, $t_N \to \infty$ and such that μ_N belongs to the parameter space $\mathcal{M}$ for N large enough. By Step 1 there exists N^* large enough such that the choice set of the program in (2.5) at μ_N is non-empty for every $N \geq N^*$. Thus, $\overline{v}_{k,i,j}(\mu_N)$ is well-defined for N large enough. Moreover, the continuity of the value function established in Step 3 implies that we can assume that $\overline{v}_{k,i,j}(\mu_N) \neq 0$ for N large enough. In fact, it is without loss of generality to assume that $\overline{v}_{k,i,j}(\mu_N) > 0$ for N large enough.

Let $X^*(\mu)$ denote the argmax of program (2.5) at μ. By Theorem 1—and using the fact that $\overline{v}_{k,i,j}(\mu) \neq 0$—$X^*(\mu)$ has a finite number of elements. Assume then that the argmax has L elements and denote them as $x_1^*(\mu), x_2^*(\mu), \ldots, x_L^*(\mu)$.

For each $l \in \{1, 2, \ldots, L\}$, let $r_l^*(\mu)$ denote the $n \times m_{z_l}$ matrix of *all* active restrictions at a solution $x_l^*(\mu)$. Likewise, let $S_l^*(\mu)$ be the matrix of dimension $n \times m_{s_l}$ that collects *all* slack restrictions at $x_l^*(\mu)$. Consequently, for each solution $x_l^*(\mu)$ there are unique matrices $r_l^*(\mu)$ and $S_l^*(\mu)$ such that

$$r_l^*(\mu)'x_l^*(\mu) = \mathbf{0}_{m_{z_l} \times 1}, \quad S_l^*(\mu)'x_l^*(\mu) > \mathbf{0}_{m_{s_l} \times 1}.$$

Define

$$R^*(\mu) \equiv \{r_1^*(\mu), r_2^*(\mu), \ldots, r_l^*(\mu)\}.$$

Proof of differentiability: We establish the differentiability of the value function in two sub-steps.

Step 4: First, we show that

$$\text{(A.6)} \qquad \max_{r_l^*(\mu) \in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\} \leq \liminf_{N \to \infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

PROOF: Take any $r_l^*(\mu) \in R^*(\mu)$. By definition of $r_l^*(\mu)$ all the columns of $S(\mu)$ that are not contained in $r_l^*(\mu)$ are slack (at μ). Consider then the candidate solution $x_+^*(\mu, r_l^*(\mu))$. This candidate solution is continuous at μ (which follows from the formula in Lemma 1 and the fact that $v_{k,i,j}(\mu, r_l^*(\mu)) = \overline{v}_{k,i,j}(\mu) > 0$). Therefore, for N large enough this candidate solution $x_+^*(\mu_N, r_l^*(\mu_N))$ is in the choice set of program (2.5) at μ_N, which implies that

$$v_{k,i,j}(\mu_N, r_l^*(\mu_N)) \leq \overline{v}_{k,i,j}(\mu_N).$$

Hence, the inequality above implies that for any $r_l^*(\mu) \in R^*(\mu)$ we have that

$$t_N(v_{k,i,j}(\mu_N, r_l^*(\mu_N)) - v_{k,i,j}(\mu, r_l^*(\mu))) \leq t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

Lemma 2 thus implies that for any $r_l^*(\mu) \in R^*(\mu)$,

$$\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h \leq \liminf_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)),$$

which establishes equation (A.6).

Step 5: Now we show that

$$\limsup_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) \leq \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}. \tag{A.7}$$

Proof: We prove the statement above by contradiction. So, suppose that (A.7) does not hold. Then, there exists $\epsilon_0 > 0$ and a subsequence μ_{N_k} such that for every $r_l^*(\mu) \in R^*(\mu)$,

$$\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h + \epsilon_0 \leq t_N(\overline{v}_{k,i,j}(\mu_{N_k}) - \overline{v}_{k,i,j}(\mu)). \tag{A.8}$$

We will show that assuming the existence of $\epsilon_0 > 0$ and a subsequence μ_{N_k} will lead to a contradiction.

Additional Notation: Let $x^*_{N_k}$ be any element in the argmax of program (2.5) at μ_{N_k}. Let $r^*_{N_k}(\mu_{N_k})$ be the matrix that collects all active restrictions at $x^*_{N_k}$ and let $S^*_{N_k}(\mu_{N_k})$ be the matrix that collects all of the sign restrictions that are slack at $x^*_{N_k}$; i.e, $S^*_{N_k}(\mu_{N_k})'x^*_{N_k} > \mathbf{0}$. Let

$$R_+(\mu) \equiv \{r \in R(\mu) \mid v_{k,i,j}(\mu; r(\mu)) > 0\}.$$

Partition the set $R_+(\mu)$ into the following four disjoint sets:

i) $R^*(\mu)$,

ii) The restrictions $r(\mu) \in R_+(\mu)/R^*(\mu)$ for which $x_+(\mu; r(\mu))$ belongs to $X^*(\mu)$,

iii) The restrictions $r(\mu) \in R_+(\mu)$ that do not fall in neither i) nor ii) and for which some sign restriction not included in $r(\mu)$ is violated,

iv) The restrictions $r(\mu) \in R(\mu)$ that do not fall in i), ii), iii) for which $x_+(\mu, r(\mu))$ is feasible but $v_{k,i,j}(\mu, r(\mu)) < \overline{v}_{k,i,j}(\mu)$.

Proof of A.7): Note that the restrictions of Type i) cannot be satisfied by $x^*_{N_k}$ infinitely often. In other words, there is no $l = 1, \ldots L$ such that

$$r^*_{N_k}(\mu_{N_k}) = r_l^*(\mu_{N_k}), \text{ and } S^*_{N_k}(\mu_{N_k}) = S_l^*(\mu_{N_k})$$

for infinitely many values of k. If this were the case, there would be a further subsequence N_{K_T} for which $\overline{v}_{k,i,j}(\mu_{N_{K_T}}) = v_{k,i,j}(\mu_{N_{K_T}}, r_l(\mu_{N_{K_T}}))$. Thus, equation (A.8) would contradict the differentiability of $v_{k,i,j}(\mu, r_l(\mu))$.

Restrictions of Type iii) cannot be satisfied infinitely often by $x^*_{N_k}$. This follows from the fact that if $r^*_{N_k}(\mu_{N_K})$ were equal to some $r(\mu_{N_K})$ for $r(\mu)$ of type iii) infinitely often, then we could always find some large k for which $x^*_{N_k}$ is the form $x_+(\mu_{N_k}, r(\mu_{N_K}))$. Such candidate solution will eventually violate a sign restriction, contradicting the fact that $x^*_{N_k}$ is in fact a solution.

Restrictions of Type iv) cannot be satisfied infinitely often by $x^*_{N_k}$. If this were the case, then we could always find some large k for which

$$
\begin{aligned}
\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h + \epsilon_0 &\leq t_{N_k}(\overline{v}_{k,i,j}(\mu_{N_k}) - \overline{v}_{k,i,j}(\mu)) \\
&= t_{N_k}(v_{k,i,j}(\mu_{N_k}, r_l(\mu_{N_k})) - \overline{v}_{k,i,j}(\mu)) \\
&= t_{N_k}(v_{k,i,j}(\mu_{N_k}, r_l(\mu_{N_k})) - v_{k,i,j}(\mu, r_l(\mu))) \\
&+ t_{N_k}(v_{k,i,j}(\mu_{N_k}, r_l(\mu)) - \overline{v}_{k,i,j}(\mu))
\end{aligned}
$$

(where $r_l(\cdot)$ is some set of restrictions of type iv)). But the fact that $(v_{k,i,j}(\mu_{N_k}; r_l(\mu)) - \overline{v}_{k,i,j}(\mu) < 0)$ contradicts the definition of the subsequence μ_{N_k}.

Finally, we show that if r is a restriction of Type ii) it cannot be the case that

$$r^*_{\mu_{N_k}}(\mu_{N_p}) = r(\mu_{N_p})$$

infinitely often. To establish this claim, suppose that there is a restriction r of Type ii) such that

$$r(\mu_{N_p})'x^*(\mu_{N_p}) = \mathbf{0}$$

infinitely often. This means we can construct a further subsequence $\mu_{N_{p_q}}$ for which (by Lemma 1)

$$\overline{v}_{k,i,j}(\mu_{N_{p_q}}) = v_{k,i,j}(\mu_{N_{p_q}}, r(\mu_{N_{p_q}})).$$

Therefore, by equation (A.8) we must have that for every $r_l^*(\mu) \in R^*(\mu)$,

$$
\begin{aligned}
\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h + \epsilon_0 &\leq t_{N_{p_q}}(v_{k,i,j}(\mu_{N_{p_q}}, r(\mu_{N_{p_q}})) - \overline{v}_{k,i,j}(\mu)) \\
&= t_{N_{p_q}}(v_{k,i,j}(\mu_{N_{p_q}}, r(\mu_{N_{p_q}})) - v_{k,i,j}(\mu, r(\mu))) \\
&+ t_{N_{p_q}}(v_{k,i,j}(\mu, r(\mu)) - \overline{v}_{k,i,j}(\mu)) \\
&= t_{N_{p_q}}(v_{k,i,j}(\mu_{N_{p_q}}, r(\mu_{N_{pp_q}})) - v_{k,i,j}(\mu, r(\mu))),
\end{aligned}
$$

where the last line follows from the fact that $r(\mu)$ is of Type ii) and, hence, leads to a candidate solution $x_+(\mu; r(\mu))$ that equals $x_l^*(\mu)$ for some l, which we will assume (without loss of generality) to be equal to 1. The differentiability result in Lemma 2 implies that for every $l = 1, \ldots, L$,

$$\dot{v}_{k,i,j}(\mu; r_l^*(\mu))'h + \epsilon_0 \leq \dot{v}_{k,i,j}(\mu, r(\mu))'h.$$

We show that this last inequality leads to a contradiction as we must have

$$\dot{v}_{k,i,j}(\mu, r_1^*(\mu))'h = \dot{v}_{k,i,j}(\mu; r(\mu))'h. \tag{A.9}$$

To see this, note first that $r_1^*(\mu)$ must contain all the columns of $r(\mu)$ as

$$r(\mu)'x_1^*(\mu) = \mathbf{0},$$

and, by definition, $r_1^*(\mu)$ contains all the constraints that are active at $x_1^*(\mu)$. Thus, we can write $r_1^*(\mu)$ as

$$r_1^*(\mu) = [r(\mu), \tilde{r}(\mu)],$$

where $r(\mu)$ and $\tilde{r}(\mu)$ are linearly independent. Our formula for $\dot{v}_{k,i,j}$ in Lemma 2 implies that (A.9) will hold if the Lagrange multipliers corresponding to the constraints in $\tilde{r}(\mu)$ are zero. To see that this is indeed the case, note that by the argument used in the proof of Lemma 2, the Karush-Kuhn-Tucker conditions for the program that only imposes $r(\mu)$ as equality conditions (along with the ellipsoid constraint) imply that

$$C_k'(A)e_i = v_{k,i,j}(\mu; r(\mu))\Sigma^{-1}x_+(\mu; r(\mu)) + r(\mu)w_1.$$

The analogous conditions for the program that imposes $r_1^*(\mu)$ as constraints imply that

$$C_k'(A)e_i = v_{k,i,j}(\mu; r_1^*(\mu))\Sigma^{-1}x_+(\mu; r_1^*(\mu)) + r_1^*(\mu)w_1^*.$$

Therefore—since by assumption $x_+(\mu; r_1^*(\mu)) = x_+(\mu; r(\mu))$—it has to be the case that

$$r(\mu)w_1 - r_1^*(\mu)w_1^* = \mathbf{0}_{n\times 1}.$$

Partitioning $w_1^* = [{w_{1,1}^*}', {w^*}_{1,2}']'$ according to $r(\mu) = [r(\mu), \tilde{r}(\mu)]$, we have that

$$r(\mu)(w_1 - w_{1,1}^*) + \tilde{r}(\mu)w_{1,2}^* = \mathbf{0}_{n\times 1}.$$

Assumption 2 implies that the latter equality holds if and only if $w_1 = w_{1,1}^*$ and $w_{1,2}^* = \mathbf{0}$. Therefore we conclude that equation (A.9) must holds. This leads to a contradiction as $\epsilon_0 > 0$ and

$$\dot{v}_{k,i,j}(\mu; r_1^*(\mu))'h + \epsilon_0 \leq \dot{v}_{k,i,j}(\mu; r_1^*(\mu))'h.$$

Summary of Step 4: Step 4.1 showed that

$$\max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu; r_l^*(\mu))'h\} \leq \liminf_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

Step 4.2 showed that

$$\limsup_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) \leq \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}.$$

We conclude that

$$\lim_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) = \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}.$$

A.5. *Proof of Theorem 3 Part a)*

Let P denote the data generating process. For notational simplicity we write μ instead of $\mu(P)$ and Ω instead of $\Omega(P)$ whenever convenient. Note first that

$$P\left(\lambda \in \left[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T}\right]\right) \tag{A.10}$$

is bounded from below by

$$P\left(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T) - \underline{v}_{k,i,j}(\mu)) \leq z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T} \text{ and } \ -z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \leq \sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T) - \overline{v}_{k,i,j}(\mu))\right),$$

which is itself bounded from below by

$$P\left(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T) - \underline{v}_{k,i,j}(\mu)) \leq z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \text{ and } -z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \leq \sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T) - \overline{v}_{k,i,j}(\mu)),\ \text{ and } ||\sqrt{T}(\widehat{\mu}_T - \mu)|| \leq M_\epsilon\right),$$

where M_ϵ is such that

$$P\Big(||\zeta(P)|| > M_\epsilon\Big) \leq \epsilon.$$

By Theorem 2, both $\underline{v}_{k,i,j}(\cdot)$ and $\overline{v}_{k,i,j}(\mu)$ are directionally differentiable function with directional derivatives denoted by $\underline{\dot{v}}_{k,i,j;\mu}(\cdot), \dot{\overline{v}}_{k,i,j;\mu}(\cdot)$. The directional differentiability implies that for any $\delta > 0$ there is T large enough such that for any $h \in \mathbb{R}^d$ such that $||h|| \leq M_\epsilon$,

$$-\delta \leq \sqrt{T}(\underline{v}_{k,i,j}(\mu + h/\sqrt{T}) - \underline{v}_{k,i,j}(\mu)) - \underline{\dot{v}}_{k,i,j;\mu}(h) \leq \delta$$

and

$$-\delta \leq \sqrt{T}(\overline{v}_{k,i,j}(\mu + h/\sqrt{T}) - \overline{v}_{k,i,j}(\mu)) - \dot{\overline{v}}_{k,i,j;\mu}(h) \leq \delta.$$

Therefore, for T large enough

$$\inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T}\Big]\Big)$$

is bounded from below by

$$P\Big(\delta + \underline{\dot{v}}_{k,i,j;\mu}(\sqrt{T}(\widehat{\mu}_T - \mu)) \leq z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T} \text{ and}$$

$$-z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \leq \dot{\overline{v}}_{k,i,j;\mu}(\sqrt{T}(\widehat{\mu}_T - \mu)) - \delta, \text{ and } ||\sqrt{T}(\widehat{\mu}_T - \mu)|| \leq M_\epsilon\Big),$$

which, by Assumption 4 (and using the continuity of the directional derivative), converges in distribution to

$$P\Big(\delta + \underline{\dot{v}}_{k,i,j;\mu}(\zeta(P)) \leq z_{1-\alpha/2}\,\sigma \text{ and}$$

$$-z_{1-\alpha/2}\sigma \leq \dot{\overline{v}}_{k,i,j;\mu}((\zeta(P)) - \delta, \text{ and } ||\zeta(P)|| \leq M_\epsilon\Big),$$

where σ is the probability limit of $\widehat{\sigma}_{k,i,j;T}$,

$$\sigma \equiv \max_{r \in R(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r)'\Omega \dot{v}_{k,i,j}(\mu; r)\Big].$$

Consequently, for every $\delta > 0$,

$$\liminf_{T\to\infty} \inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T}\Big]\Big)$$

is larger than or equal

$$1 - P\Big(\underline{\dot{v}}_{k,i,j;\mu}(\zeta(P)) > z_{1-\alpha/2}\,\sigma - \delta\Big) - P\Big(\dot{\overline{v}}_{k,i,j;\mu}(\zeta(P)) < -z_{1-\alpha/2}\,\sigma + \delta\Big)$$

$$-P\Big(||\zeta(P)|| > M_\epsilon\Big).$$

Take some $x \in X_*(\mu)$ for which $\underline{\sigma}(x) \equiv \dot{v}_{k,i,j}(\mu; r(\mu;x))'\Omega \dot{v}_{k,i,j}(\mu; r(\mu;x)) > 0$ (one such x must exist by the assumption of this theorem). The fact that $\zeta(P)$ is symmetric and using our formula for the directional derivative of $\underline{v}_{k,i,j}$ we have that

$$\begin{aligned} P\Big(\underline{\dot{v}}_{k,i,j;\mu}(\zeta(P)) > z_{1-\alpha/2}\,\sigma - \delta\Big) &\leq P\Big(\dot{v}_{k,i,j}(\mu; r(\mu;x))'\zeta(P) > z_{1-\alpha/2}\sigma - \delta\Big) \\ &\leq \mathbb{P}\Big(N(0,1) > z_{1-\alpha/2}\frac{\sigma}{\underline{\sigma}(x)} - \frac{\delta}{\underline{\sigma}(x)}\Big), \\ &\leq \mathbb{P}\Big(N(0,1) > z_{1-\alpha/2} - \frac{\delta}{\underline{\sigma}(x)}\Big), \end{aligned}$$

for any $\delta > 0$ (since $\sigma \geq \underline{\sigma}(x)$).

Now, take some $x \in X^*(\mu)$ for which $\overline{\sigma}(x) \equiv \dot{v}_{k,i,j}(\mu; r(\mu;x))'\Omega \dot{v}_{k,i,j}(\mu; r(\mu;x)) > 0$. Note that

$$\begin{aligned} P\Big(\dot{\overline{v}}_{k,i,j;\mu}(\zeta(P)) < -z_{1-\alpha/2} + \delta\,\sigma\Big) &\leq P\Big(\dot{v}_{k,i,j}(\mu; r(\mu,x))'\zeta(P) < -z_{1-\alpha/2}\sigma + \delta\Big) \\ &\leq \mathbb{P}\Big(N(0,1) < -z_{1-\alpha/2}\frac{\sigma}{\overline{\sigma}(x)} + \frac{\delta}{\overline{\sigma}(x)}\Big), \\ &\leq \mathbb{P}\Big(N(0,1) < -z_{1-\alpha/2} + \frac{\delta}{\overline{\sigma}(x)}\Big), \end{aligned}$$

for any $\delta > 0$ (since $\sigma > \overline{\sigma}(x)$). We conclude that for any $\epsilon > 0$ and $\delta > 0$

$$\liminf_{T\to\infty} \inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T}\Big]\Big)$$

is bounded from below by

$$\Phi\left(z_{1-\alpha/2} - \frac{\delta}{\underline{\sigma}(x)}\right) - \Phi\left(-z_{1-\alpha/2} + \frac{\delta}{\overline{\sigma}(x)}\right) - \epsilon,$$

where $\Phi(\cdot)$ is the standard normal c.d.f. Since $\epsilon > 0$, $\delta > 0$ are arbitrary and $\Phi(\cdot)$ is continuous, the desired result follows.

A.6. *Proof of Theorem 3 Part b)*

PROOF: We would like to show that for every $\epsilon > 0, \eta > 0$ there is $T^*(\epsilon, \eta)$ such that for $T \geq T^*(\epsilon, \eta)$ we have that

$$P(RBC(Y_1, \ldots, Y_T) < 1 - \alpha - \epsilon) < \eta.$$

We divide the proof into 5 steps.

Step 1 (Definitions of $M_{\epsilon,\eta}$, δ_ϵ): Let ζ be a $\mathcal{N}_d(\mathbf{0}, \Omega(P))$ random vector. For given $\epsilon > 0, \eta > 0$ define $M_{\epsilon,\eta} \in \mathbb{R}$ as the scalar such that

$$\mathbb{P}(||\zeta|| > M_{\epsilon,\eta}) < \min\{\epsilon/3, \eta/4\}.$$

Let $\Phi(\cdot)$ denote the standard normal c.d.f. Define $\delta_\epsilon > 0$ to be any scalar such that

$$|\Phi(z_{1-\alpha/2} - \delta_\epsilon/\underline{\sigma}(\mu)) - \Phi(-z_{1-\alpha/2} + \delta_\epsilon/\overline{\sigma}(\mu)) - (1-\alpha)| < \epsilon/3.$$

Such a scalar exists by the continuity of $\Phi(\cdot)$ and the fact that $\underline{\sigma}(\mu)$ and $\overline{\sigma}(\mu)$ are positive.

Step 2 (Definitions of $A_T(\epsilon), B_T(\epsilon), C_T(\epsilon)$). Let

$$Y^T \equiv (Y_1, \ldots, Y_T)$$

denote the data. In a slight abuse of notation, let $\widehat{\sigma}_T$ abbreviate $\widehat{\sigma}_{k,i,j}$ and let σ denote the probability limit of $\widehat{\sigma}_T$. Define the events:

$$\begin{aligned}
A_T(\epsilon,\eta) &\equiv \left\{ Y^T \;\middle|\; ||\sqrt{T}(\widehat{\mu}_T - \mu)|| > M_{\epsilon,\eta} \right\},\\
B_T(\epsilon) &\equiv \left\{ Y^T \;\middle|\; \sup_{B \in \mathcal{B}(\mathbb{R}^d)} |P^*_\mu\left(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y^T\right) - \mathbb{P}\left(\zeta \in B\right)| > \frac{\epsilon}{3} \right\},\\
C_T(\epsilon) &\equiv \left\{ Y^T \;\middle|\; |\widehat{\sigma} - \sigma| > \frac{\delta_\epsilon}{2 z_{1-\alpha/2}} \right\}.
\end{aligned}$$

We will show that if the Robust Bayes Credibility of our delta-method interval falls below $1-\alpha-\epsilon$ then one of the events above occurs *a fortiori*. We will then argue that our assumptions imply that the probability of each of these events becomes arbitrarily small for large T (implying the event in which the Robust Bayes Credibility is below $1-\alpha-\epsilon$ happens with an arbitrarily small probability).

Note that the CLT for $\widehat{\mu}_T$ (Assumption 4) implies that for any $\epsilon > 0$ and any $\eta > 0$

(A.11) $\qquad P(A_T(\epsilon,\eta)) \to \mathbb{P}(||\zeta|| > M_{\epsilon,\eta}).$

The Bernstein von-Mises Theorem for μ^* (Assumption 5) implies that for any $\epsilon > 0$

(A.12) $\qquad P(B_T(\epsilon)) \to 0.$

Finally, the definition of probability limit implies that

$$P(C_T(\epsilon)) \to 0. \tag{A.13}$$

Therefore, for any $\epsilon > 0, \eta > 0$ there exists $T_1(\epsilon, \eta)$ such that for any $T \geq T_1(\epsilon, \eta)$

$$|P(A_T(\epsilon,\eta)) - P(||\zeta|| > M_{\epsilon,\eta})| < \eta/4, \quad |P(B_T(\epsilon))| < \eta/4, \quad P(C_T(\epsilon)) < \eta/4. \tag{A.14}$$

Step 3 (First order approximations of the bounds of the identified set). Let μ denote the true parameter and define $Z_T^* \equiv \sqrt{T}(\mu^* - \widehat{\mu}_T)$ and $Z_T \equiv \sqrt{T}(\widehat{\mu}_T - \mu)$. Let $\underline{v}(\cdot)$ abbreviate $\underline{v}_{k,i,j}(\cdot)$ and, likewise, let $\overline{v}(\cdot)$ abbreviate $\overline{v}_{k,i,j}(\cdot)$. Note that

$$\sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)) = \sqrt{T}(\underline{v}(\mu + Z_T^*/\sqrt{T} + Z_T/\sqrt{T}) - \underline{v}(\mu)) - \sqrt{T}(\underline{v}(\mu + Z_T/\sqrt{T}) - \underline{v}(\mu)).$$

The differentiability of $\underline{v}(\cdot)$ at μ (which follows from Theorem 2 and the fact that $X_*(\mu)$ is a singleton) implies that whenever $||Z_T^*|| \leq M_\epsilon$ and $||Z_T|| \leq M_\epsilon$ there is $T_2(\epsilon, \eta)$ such that for $T \geq T_2(\epsilon, \eta)$,

$$|\sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)) - \dot{\underline{v}}_\mu(Z_T^* + Z_T) - \dot{\underline{v}}_\mu(Z_T)| = |\sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)) - \dot{\underline{v}}_\mu(Z_T^*)| < \delta_\epsilon/2.$$

Analogously, we can find $T_3(\epsilon, \eta)$ such that for $T \geq T_3(\epsilon, \eta)$ we have

$$|\sqrt{T}(\overline{v}(\mu^*) - \overline{v}(\widehat{\mu}_T)) - \dot{\overline{v}}_\mu(Z_T^*)| < \delta_\epsilon/2.$$

Step 4 (Lower bound on the Robust Bayesian Credibility of a set). Define the posterior probability that the bounds of the identified set are contained in our delta-method interval as

$$c(Y^T) \equiv P_\mu^* \left([\underline{v}(\mu^*), \overline{v}(\mu^*)] \subseteq \left[\underline{v}(\widehat{\mu}_T) - z_{1-\alpha/2}\, \widehat{\sigma}/\sqrt{T},\ \overline{v}(\widehat{\mu}_T) + z_{1-\alpha/2}\, \widehat{\sigma}/\sqrt{T} \right] |Y^T \right).$$

Note that for every data realization

$$c(Y^T) \leq RBC(Y^T),$$

which follows from the fact that for any (A, B) we have that $\lambda(A, B) \in [\underline{v}(\mu), \overline{v}(\mu)]$. Therefore for any $\epsilon > 0$

$$P(RBC(Y^T) < 1 - \alpha - \epsilon) \leq P(c(Y^T) < 1 - \alpha - \epsilon) \tag{A.15}$$

Thus, to establish Theorem 4 it suffices to show that for any $\epsilon > 0$

$$\lim_{T\to\infty} P(c(Y^T) < 1 - \alpha - \epsilon) = 0.$$

We establish such a result in the following step.

Step 5: We now show that for any $\epsilon > 0$, $\eta > 0$ there is T large enough such that

$$P(c(Y^T) < 1 - \alpha - \epsilon) \leq P(A_T(\epsilon, \eta) \cup B_T(\epsilon) \cup C_T(\epsilon)),$$

or equivalently, that

$$P(A_T^c(\epsilon) \cap B_T^c(\epsilon) \cap C_T^c(\epsilon)) \leq P(c(Y^T) \geq 1 - \alpha - \epsilon)$$

for T large enough. We start by re-writing $c(Y^T)$ as

$$P_\mu^* \Big(- z_{1-\alpha/2}\widehat{\sigma} \leq \sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)), \text{ and } \sqrt{T}(\overline{v}(\mu^*) - \overline{v}(\widehat{\mu}_T)) \leq z_{1-\alpha/2}\widehat{\sigma} |Y^T \Big),$$

and noting that

$$c(Y^T) \geq \tag{A.16}$$

$$P_\mu^* \Big(-z_{1-\alpha/2}\widehat{\sigma} \leq \sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)), \text{ and } \sqrt{T}(\overline{v}(\mu^*) - \overline{v}(\widehat{\mu}_T)) \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||\sqrt{T}(\mu^* - \widehat{\mu}_T)|| \leq M_{\epsilon,\eta} |Y^T \Big).$$

Take $T^*(\epsilon,\eta)=\max\{T_1(\epsilon,\eta),T_2(\epsilon,\eta),T_3(\epsilon,\eta)\}$. From Equation (A.16) and Step 2 it follows that

$$Y^T \in A_T^c(\epsilon,\eta) \implies$$

$$c(Y^T) \geq P_\mu^*\Big(-z_{1-\alpha/2}\widehat{\sigma} \leq \underline{\dot{v}}_\mu(Z_T^*)-\delta_\epsilon/2, \text{ and } \dot{\overline{v}}_\mu(Z_T^*)+\delta_\epsilon/2 \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||Z_T^*|| \leq M_{\epsilon,\eta}|Y^T\Big) \tag{A.17}$$

for $T \geq T^*$. In addition,

$$Y^T \in C_T^c(\epsilon)$$

implies that the right-hand side of equation (A.17) is larger than or equal

$$P_\mu^*\Big(-z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(Z_T^*)-\delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(Z_T^*)+\delta_\epsilon \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||Z_T^*|| \leq M_{\epsilon,\eta}|Y^T\Big).$$

This means that for $T \geq T^*(\epsilon,\eta)$

$$Y^T \in A_T^c(\epsilon)\cap C_T^c(\epsilon) \implies$$

$$c(Y^T) \geq P_\mu^*\Big(-z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(Z_T^*)-\delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(Z_T^*)+\delta_\epsilon \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||Z_T^*|| \leq M_{\epsilon,\eta}|Y^T\Big). \tag{A.18}$$

Define the set

$$B=\Big\{z\in\mathbb{R}^d \;\Big|\; -z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(z)-\delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(z)+\delta_\epsilon \leq z_{1-\alpha/2}\sigma, \text{ and } ||z|| \leq M_{\epsilon,\eta}\Big\}.$$

By definition, $\underline{\dot{v}}_\mu(\cdot)$ and $\dot{\overline{v}}_\mu(\cdot)$ are linear and thus measurable functions. This means that B is a Borel Set (as it is the inverse image of a Borel subset on the real line under a measurable function). Consequently,

$$Y^T \in A_T^c(\epsilon,\eta)\cap B_T^c(\epsilon)\cap C_T^c(\epsilon)$$

implies that

$$\begin{aligned} C(Y^T) &\geq \mathbb{P}\Big(-z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(\zeta)-\delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(\zeta)+\delta_\epsilon \leq z_{1-\alpha/2}\sigma, \text{ and } ||\zeta|| \leq M_{\epsilon,\eta}\Big)-\epsilon/3 \\ &= \mathbb{P}\Big(-\underline{\dot{v}}_\mu(\zeta) \leq z_{1-\alpha/2}\sigma-\delta_\epsilon, \text{ and } -z_{1-\alpha/2}\sigma+\delta_\epsilon \leq -\dot{\overline{v}}_\mu(\zeta), \text{ and } ||\zeta|| \leq M_{\epsilon,\eta}\Big)-\epsilon/3. \end{aligned}$$

Note further that because the distribution of ζ is the same as that of $-\zeta$ and because $\underline{\dot{v}}_\mu(\cdot)$, $\dot{\overline{v}}_\mu(\cdot)$ are linear functions (by definition of derivative) we have that

$$\begin{aligned} C(Y^T) &\geq \mathbb{P}\Big(\underline{\dot{v}}_\mu(\zeta) \leq z_{1-\alpha/2}\sigma-\delta_\epsilon, \text{ and } -z_{1-\alpha/2}\sigma+\delta_\epsilon \leq \dot{\overline{v}}_\mu(\zeta), \text{ and } ||\zeta|| \leq M_\epsilon\Big)-\epsilon/3 \\ &\geq 1-\mathbb{P}\Big(\underline{\dot{v}}_\mu(\zeta) > z_{1-\alpha/2}\sigma-\delta_\epsilon\Big)-\mathbb{P}\Big(-z_{1-\alpha/2}\sigma+\delta_\epsilon > \dot{\overline{v}}_\mu(\zeta)\Big)-2\epsilon/3 \\ &= 1-\mathbb{P}\Big(N(0,1) > z_{1-\alpha/2}\frac{\sigma}{\underline{\sigma}(\mu)}-\frac{\delta_\epsilon}{\underline{\sigma}(\mu)}\Big)-\mathbb{P}\Big(-z_{1-\alpha/2}\frac{\sigma}{\overline{\sigma}(\mu)}+\frac{\delta_\epsilon}{\overline{\sigma}(\mu)} > N(0,1)\Big)-2\epsilon/3 \\ &\geq 1-\mathbb{P}\Big(N(0,1) > z_{1-\alpha/2}-\frac{\delta_\epsilon}{\underline{\sigma}(\mu)}\Big)-\mathbb{P}\Big(-z_{1-\alpha/2}+\frac{\delta_\epsilon}{\overline{\sigma}(\mu)} > N(0,1)\Big)-2\epsilon/3 \\ &\geq \Phi\Big(z_{1-\alpha/2}-\frac{\delta_\epsilon}{\underline{\sigma}(\mu)}\Big)-\Phi\Big(-z_{1-\alpha/2}+\frac{\delta_\epsilon}{\overline{\sigma}(\mu)}\Big)-2\epsilon/3 \\ &\geq 1-\alpha-\epsilon. \end{aligned}$$

Thus, we have shown that if $T \geq T^*(\epsilon,\eta)$, then

$$Y^T \in A_T^c(\epsilon,\eta)\cap B_T^c(\epsilon)\cap C_T^c(\epsilon) \implies c(Y^T) \geq 1-\alpha-\epsilon.$$

This means that if $T \geq T^*(\epsilon,\eta)$, then

$$
\begin{aligned}
P\Big(c(Y^T) < 1-\alpha-\epsilon\Big) &\leq P(A_T(\epsilon,\eta)) + P(B_T(\epsilon)) + P(C_T(\epsilon)) \\
&\leq |P(A_T(\epsilon)) - \mathbb{P}(||\zeta|| > M_{\epsilon,\eta})| + \mathbb{P}(||\zeta|| > M_{\epsilon,\eta}) \\
&+ P(B_T(\epsilon)) + P(C_T(\epsilon)) \\
&\leq 4(\eta/4) \quad \text{(by equation (A.14))} .
\end{aligned}
$$

Therefore, for any $\epsilon > 0, \eta > 0$ there is $T^*(\epsilon, \eta)$ such that

$$P(RBC(Y^T) < 1-\alpha-\epsilon) \leq P(c(Y^T) < 1-\alpha-\epsilon) < \eta.$$

A.7. *Implementation details*

A.7.1. *Bonferroni confidence set*

This section describes the implementation of the Bonferroni-type method proposed by Granziera et al. (2017). The following algorithm is a variation of the algorithm outlined on p.17 in Granziera et al. (2017). There is one minor difference between the algorithms. To avoid dealing with degenerate $\hat{D}$, we add Step 3.c instead of implicitly adjusting the criterion function as proposed in Section 4.2 of Granziera et al. (2017). The rate of the sequence $\underline{\sigma}_T$ guaranties that the additional noise $\underline{\sigma}_T \epsilon_b$ does not affect the asymptotic distribution of $\mathcal{G}(\xi_g)$.

1. Generate N_B draws $\left\{\mu_b^*\right\}_{b=1}^{N_B} \sim N\left(\widehat{\mu}_T, \widehat{\Omega}_T\right)$.
2. Generate N_G grid points $\{x_g\}_{g=1}^{N_g}$ on a unit $d-$sphere distributed uniformly using the algorithm from Uhlig (2005).
3. For every grid point x_g, we implement the following statistical test (of size $1-\alpha/2$) of whether $B_{1g} = \widehat{\Sigma}_T^{1/2} x_g$ satisfies all identification restrictions. This is done by following steps a) to g) below.
   (a) Compute estimated residuals[21],
   $$\xi_g = \left(S'\left(\hat{\mu}_T\right), Z'\left(\hat{\mu}_T\right)\right)' B_{1g}.$$
   (b) Compute re-centered bootstrap residuals $\left\{\xi_{g;b}^*\right\}_{b=1}^{N_B}$,
   $$\tilde{\xi}_{g;b}^\star = \left(S'\left(\mu_b^*\right), Z'\left(\mu_b^*\right)\right)' \Sigma_b^{*1/2} x_g - \xi_g.$$
   (c) Add independent normally distributed noise with $\epsilon_b \sim N(0, I)$ and $\underline{\sigma}_T = 10^{-6}/\sqrt{T\ln(\ln T)}$,
   $$\xi_{g;b}^* = \tilde{\xi}_{g;b}^\star + \underline{\sigma}_T \epsilon_b.$$
   (d) Compute standard errors for $\left\{\xi_{g;b}^*\right\}_{b=1}^{N_B}$. The diagonal matrix $\hat{D}^{1/2}$ has the corresponding standard errors on the diagonal.
   (e) Select binding inequities as inequalities corresponding to the components ℓ of ξ_g such that
   $$e_\ell' \hat{D}^{-1/2} \xi_g \leq \kappa_T = 1.96 \ln(\ln T).$$
   (f) Compute the criterion function $\mathcal{G}(\xi_g)$ and $\left\{\mathcal{G}\left(\xi_{g;b}^*\right)\right\}_{b=1}^{N_B}$ which includes only the equalities

[21] We only compute matrices $\left(S'\left(\hat{\mu}_T\right), Z'\left(\hat{\mu}_T\right)\right)' \sqrt{\widehat{\Sigma}_T}$ and $\left(S'\left(\mu_b^*\right), Z'\left(\mu_b^*\right)\right)' \sqrt{\Sigma_b^*}$ once to speed up the costly matrix multiplication.

and the binding inequalities, where

$$\mathcal{G}\left(\xi_{g;b}^{*}\right)=\sum_{\ell=1}^{m_z}\left(e_\ell'\hat{D}^{-1/2}\xi_{g;b}^{*}\right)^2+\sum_{\ell=m_z+1}^{m_s+m_s}\left(e_\ell'\hat{D}^{-1/2}\xi_{g;b}^{*}\right)^2\mathbf{1}\left\{e_\ell'\hat{D}^{-1/2}\xi_g\leq\kappa_T\right\} \tag{A.19}$$

(g) Grid point x_g passes the test if $\mathcal{G}\left(\xi_g\right)$ is less that $1-\alpha/2$ sample quantile of $\left\{\mathcal{G}\left(\xi_{g;b}^{*}\right)\right\}_{b=1}^{N_B}$.

4. If x_g passes the test in Step 3, compute $\underline{\lambda}_{k,i,j}^{(g)}$ and $\bar{\lambda}_{k,i,j}^{(g)}$ as $\alpha/4$ and $1-\alpha/4$ sample quantiles of $\left\{e_i'C_k\left(A_b^*\right)\sqrt{\Sigma_b^*}x_g\right\}_{b=1}^{N_B}$ correspondingly. Otherwise set $\underline{\lambda}_{k,i,j}^{(g)}=+\infty$ and $\bar{\lambda}_{k,i,j}^{(g)}=-\infty$.

5. Report

$$\mathrm{CS}_T^{GSM}\left(1-\alpha\right)=\left[\min_{g=1,N_G}\underline{\lambda}_{k,i,j}^{(g)},\max_{g=1,N_G}\bar{\lambda}_{k,i,j}^{(g)}\right].$$

Our implementation corresponds to a generalized version of the criterion function considered in Section 6 of Granziera et al. (2017). This generalized criterion function can potentially be applied to a combination of zero and sign restrictions. In our baseline empirical application, however, the acceptance rate of Step 3 is so low that we could not find a single point out of 10000 grid points that would pass the test. For this reason, we report the results for the alternative identification scheme with the zero restriction on the FFR being replaced by a negative sign restriction.

The number of grid points that pass Step 4 of the algorithm depends crucially on the number of the identifying restrictions imposed. In our experiment, every additional sign restriction reduces the acceptance rate almost by half and, correspondingly, requires twice more grid points and computational time to achieve the same level of accuracy. For the UMP example with 4 sign restrictions the acceptance rate is 9.1%.

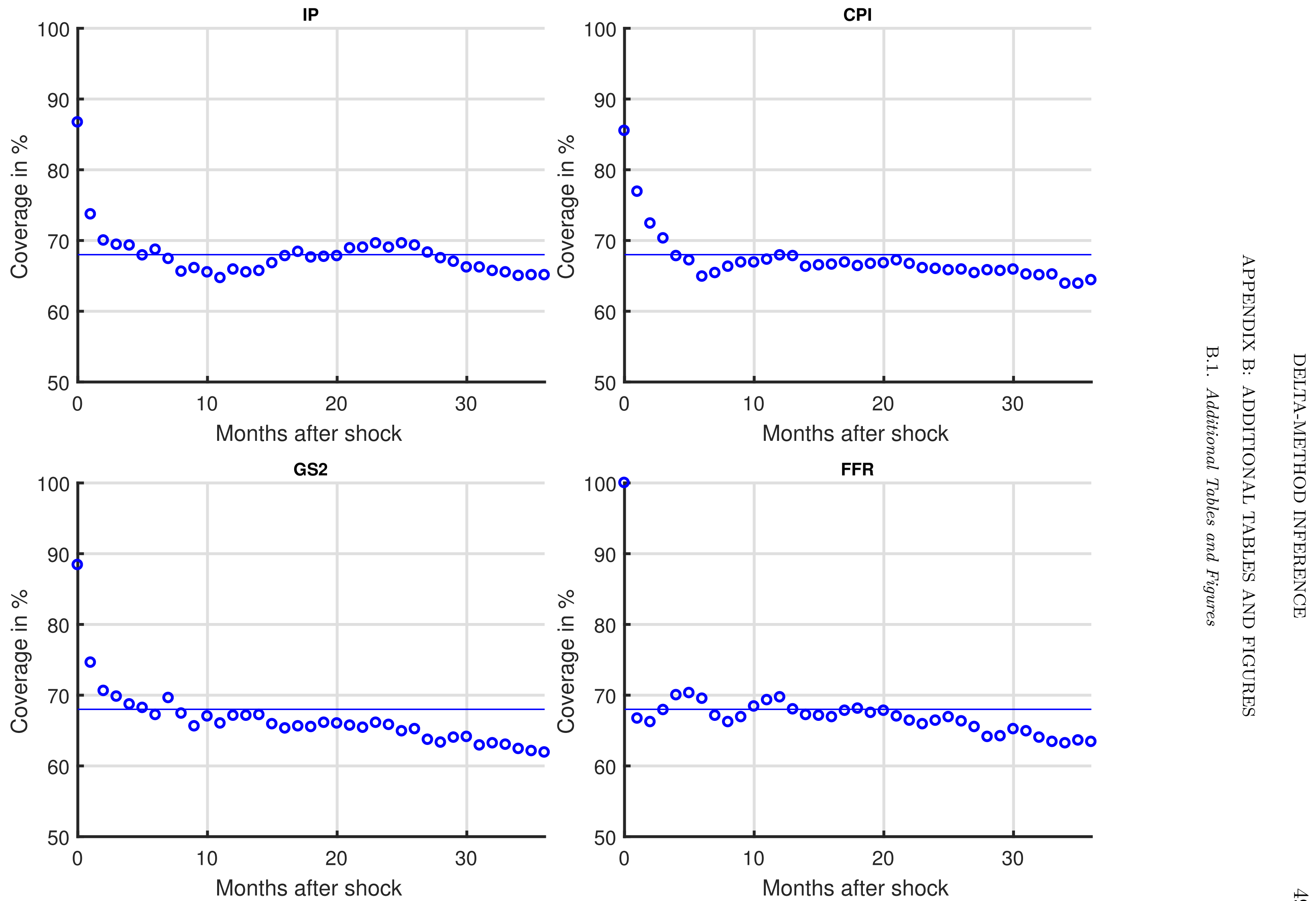
A.7.2. *Joint Confidence Sets*

To implement Inoue and Kilian (2013)'s algorithm, we first sample 10,000 joint draws from the posterior of reduced-form parameters and structural coefficients that satisfy all identification restriction. We use those draws to compute 10,000 structural impulse response function. Second, we sample 20,000 draws of reduced-form parameters to compute the marginal posterior density for each structural response. Third, we compute a joint 68% credible set by keeping all of structural responses which have marginal density higher than the lowest 32%. The second step is computationally costly. In our implementation it takes 2.5 hours when using 50 parallel workers in Matlab.

APPENDIX B: ADDITIONAL TABLES AND FIGURES

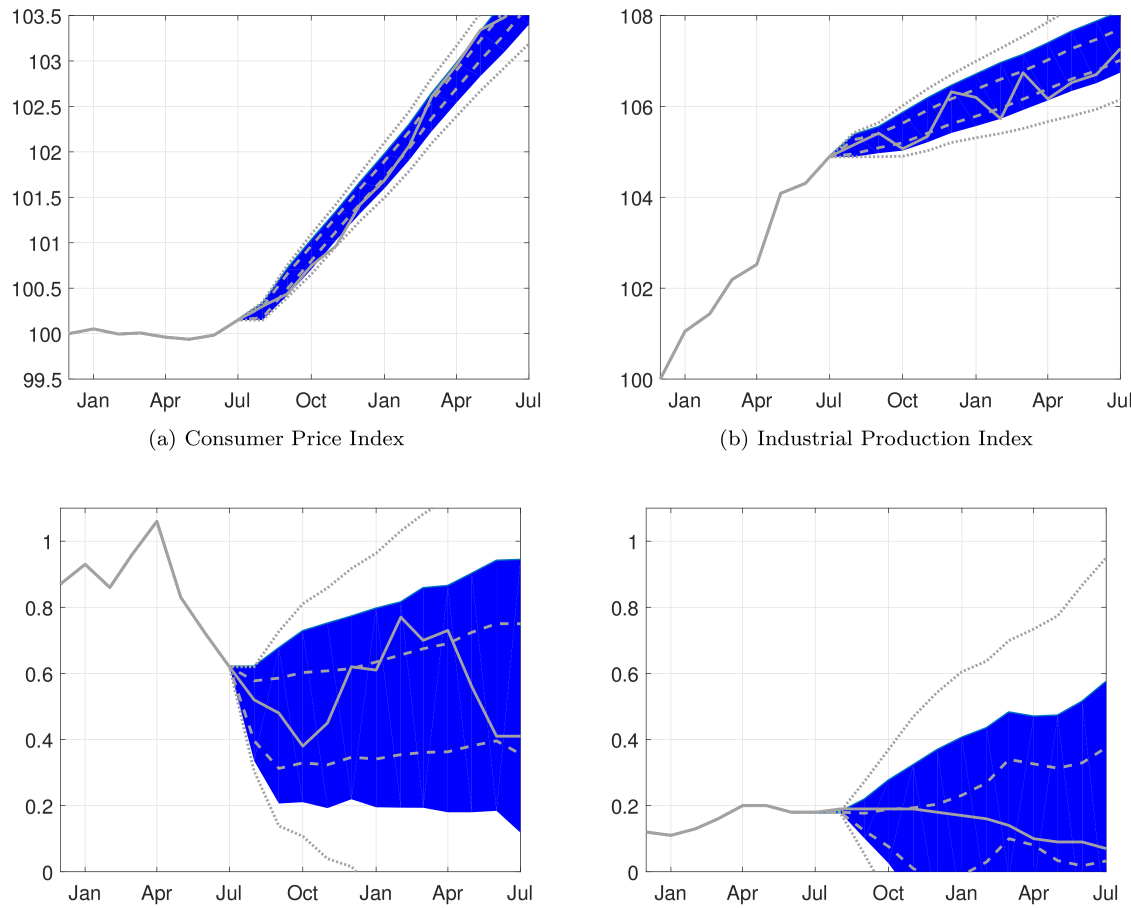
B.1. *Additional Tables and Figures*

Figure 5: Robust Bayesian credibility of the delta-method interval based on the posterior distribution $\mu^* \sim \mathcal{N}(\widehat{\mu}_T, \widehat{\Omega}_T/T)$, $T = 342$.



(Circles) Monte-Carlo estimate of the probability $P^*_\mu\left([\underline{v}_{k,i,j}(\mu^*), \overline{v}_{k,i,j}(\mu^*)] \subset [\underline{v}_{k,i,j}(\widehat{\mu}_T) - .9945\widehat{\sigma}_{(k,i,j),T}/\sqrt{T}, \overline{v}_{k,i,j}(\widehat{\mu}_T) + .9945\widehat{\sigma}_{(k,i,j),T}/\sqrt{T} | Y_1, \ldots, Y_T]\right)$ for the posterior distribution $\mu^* \sim \mathcal{N}(\widehat{\mu}_T, \widehat{\Omega}_T)$, with $T = 342$. The values $\widehat{\mu}_T$ and $\widehat{\Omega}_T$ correspond, respectively, to the estimators of the reduced-form parameter and its asymptotic covariance matrix in the UMP application. (Solid Line) Nominal credibility for the delta-method confidence interval (68%).

Figure 6: Projection Confidence Interval for CPI, IP, 2yTB, FF after the August 2010 announcement



(a) Consumer Price Index

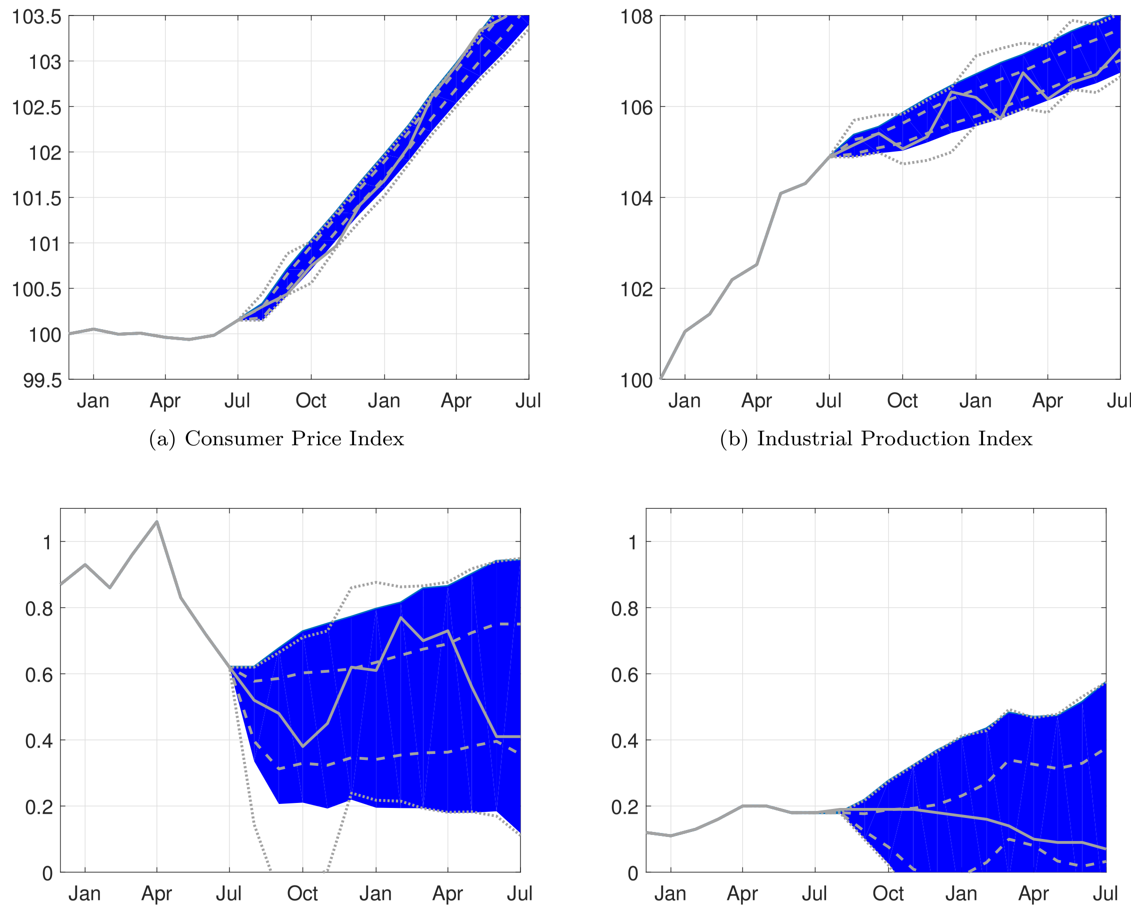
(b) Industrial Production Index

(c) 2-year Treasury Bond Rate

(d) Federal Funds Rate

(Shaded Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Dashed Line) 68% credible set constructed using the priors in Uhlig (2005). (Gray, Dotted Line) Gafarov et al. (2016)'s 68% confidence interval based on the projection approach.

Figure 7: Robust Credible Set for CPI, IP, 2yTB, FF after the August 2010 announcement



(a) Consumer Price Index

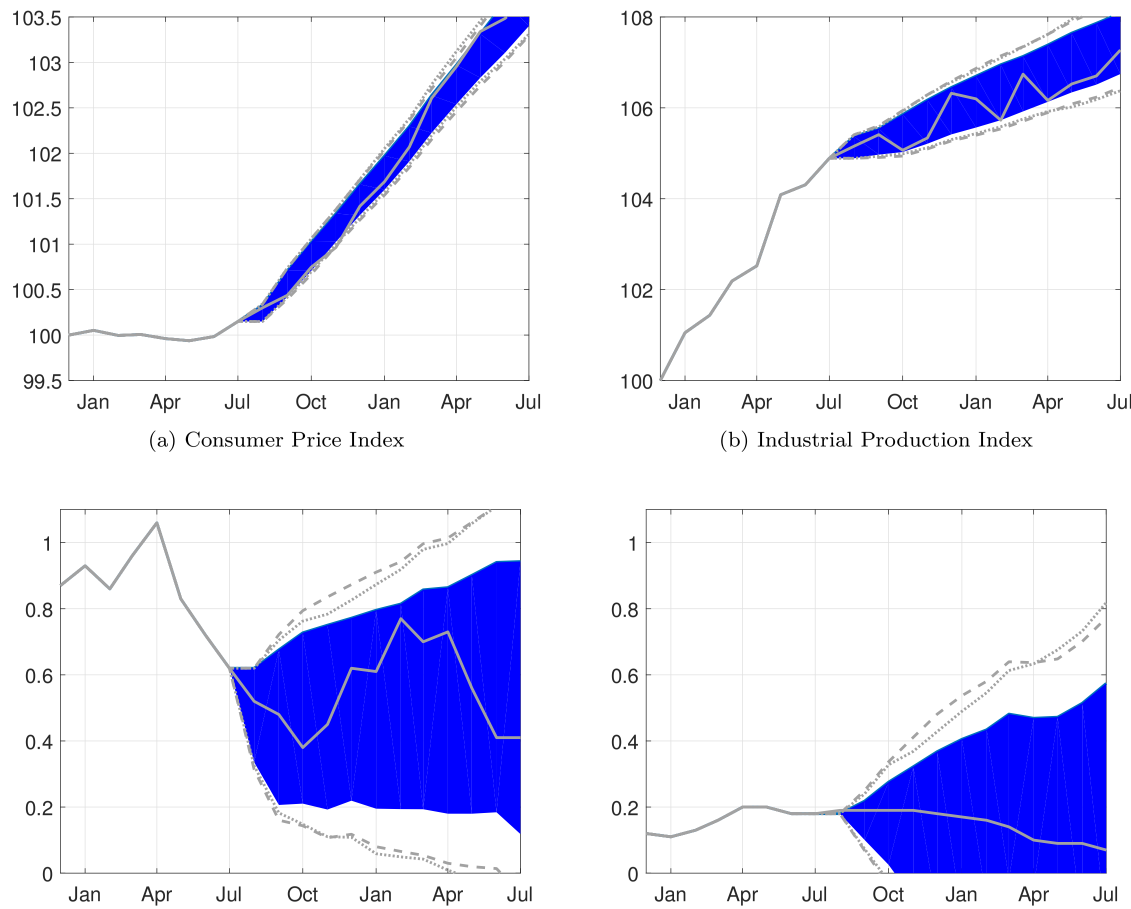
(b) Industrial Production Index

(c) 2-year Treasury Bond Rate

(d) Federal Funds Rate

(Shaded Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Dashed Line) 68% credible set constructed using the priors in Uhlig (2005). (Dotted Line) Giacomini and Kitagawa (2015)'s 68% robust-Bayes credible set constructed using the priors for the reduced-form parameters in Uhlig (2005)

Figure 8: Joint Credible Set (corresponding to all 4 variables and 36 horizons) for impulse response functions of CPI, IP, 2yTB, FF after the August 2010 announcement



(a) Consumer Price Index

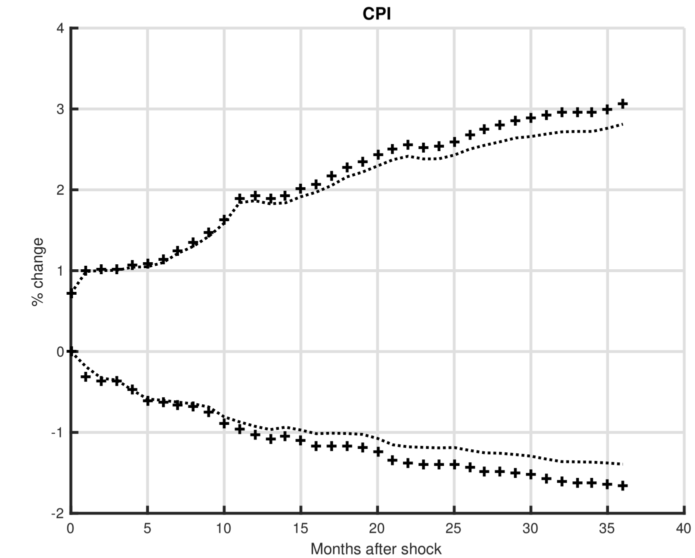
(b) Industrial Production Index

(c) 2-year Treasury Bond Rate

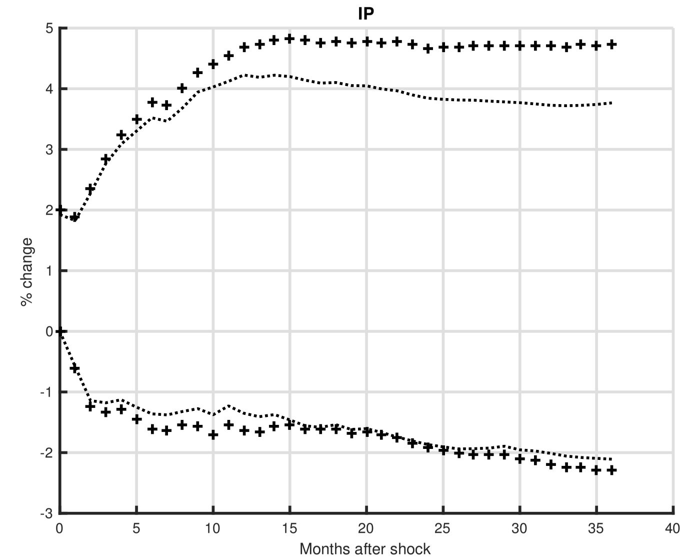
(d) Federal Funds Rate

(Shaded Area) Evolution of the Levels CPI, IP, 2yTB, and FF based on our 68% delta method confidence bands for the coefficients of Cumulative Impulse-Response Functions. (Solid Line) Observed Levels of CPI, IP, 2yTB, and FF from December 2009 to July 2011. Both the CPI index and the IP index were normalized to have a starting value of 100. (Dashed Line) Bonferroni-corrected joint 68% delta method confidence bands. (Dotted Line) Inoue and Kilian (2013)'s joint 68% Bayes credible set for impulse response functions using the priors for the reduced-form parameters in Uhlig (2005).

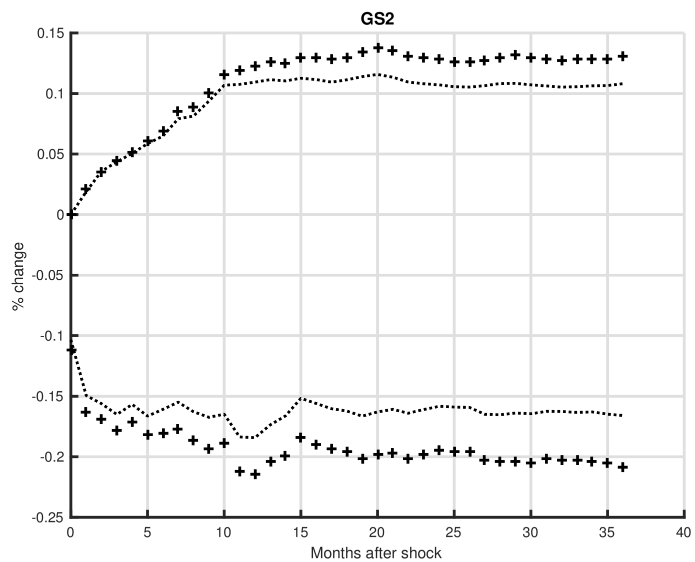
Figure 9: Confidence set for impulse response functions of CPI, IP, 2yTB, FF to a UMP shock under alternative identification scheme



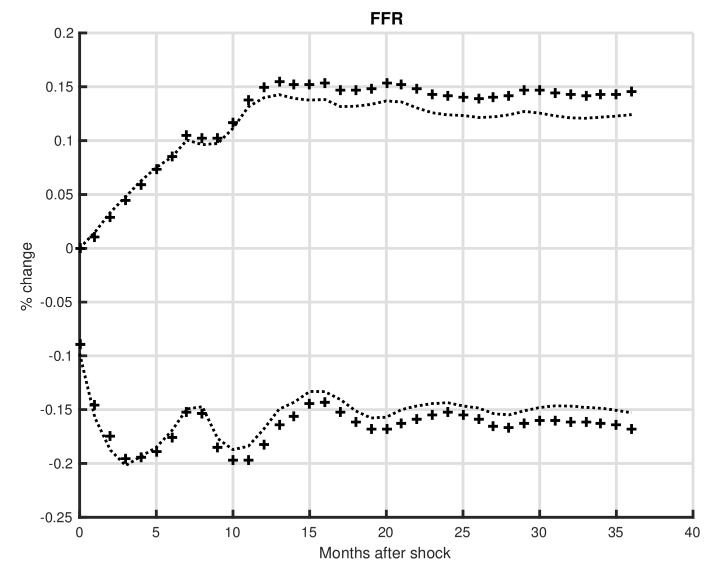
(a) Consumer Price Index



(b) Industrial Production Index



(c) 2-year Treasury Bond Rate



(d) Federal Funds Rate

(Pluses) 68% Bonferroni-type confidence bounds of Granziera et al. (2017). See details on implementation in A.7.1 . (Dots) 68% delta method confidence bands

Note: the identification scheme used to produce these plots differs from the one used for Figure 1. The zero restriction on the response of the FFR is replaced with a negative sign restriction to improve the acceptance rate of the algorithm for Bonferroni-type CS.

Figure 10: The mathematical program defining $\overline{v}_{k,i,j}(\mu)$ $(n = 3)$ with one zero restriction.

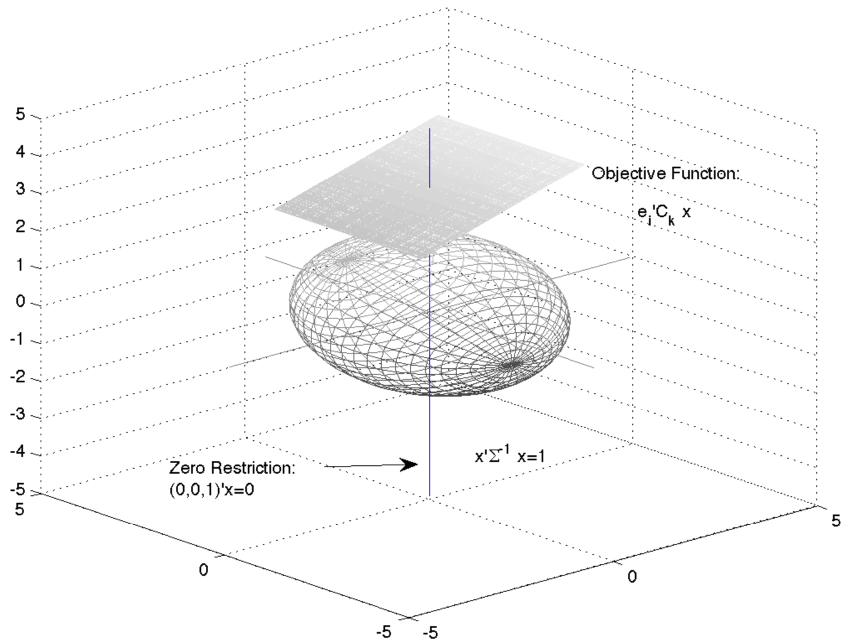


Figure 10 provides a graphical representation of the mathematical program (2.5), where $BB' = \Sigma$ has been replaced by the 'ellipsoid' constraint $x'\Sigma^{-1}x = 1$, $x \equiv B_j \in \mathbb{R}^3$. The objective function corresponds to the hyperplane with the normal vector $C_k(A)'e_i \in \mathbb{R}^3$. In this example, there is only one equality restriction with the normal vector given by the solid line. This restriction requires the contemporaneous impact of the j-th shock on the third variable to be zero. Note that without the equality restriction the maximizer and minimizer will be given by the point at which the hyperplane is tangent to the ellipsoid.

Figure 11: Solving for $\overline{v}_{k,i,j}(\mu)$ ($n = 3$, $\Sigma = \mathbb{I}_3$) with one equality restriction.

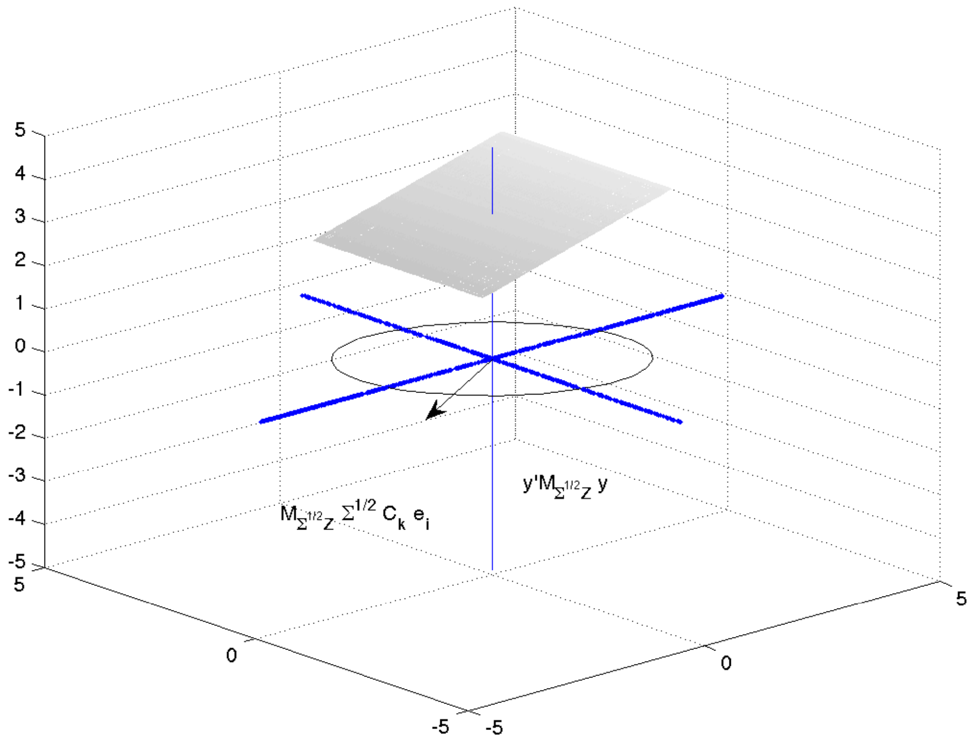


Figure 11 provides a graphical representation of the solution to the mathematical program (2.5) when $\Sigma = \mathbb{I}_3$ and there is only one zero restriction. The solution to the program must lie in the orthogonal complement of Z (the thin solid line). In this picture, the orthogonal complement corresponds to the space spanned by the 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 ellipsoid. One can find the value of this problem by projecting the gradient of the objective function on the orthogonal complement of $\Sigma^{1/2}z$ (the arrow) and selecting a direction in the ellipsoid collinear to it. The value function $\overline{v}_{k,ij}(\mu)$ will be given by the norm of the arrow.

Suppose there are only equality constraints. Note that $Z'B_j = \mathbf{0}_{m\times 1}$ implies that the re-parameterized choice variable $\widetilde{x} \equiv \Sigma^{-1/2}B_j$ 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 $\widetilde{x}'\widetilde{x} = 1$. 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 $(A.1)$ gives the maximum response in (2.5).