PROJECTION INFERENCE FOR SET-IDENTIFIED SVARS.[1]

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

We study the properties of projection inference for set-identified Structural Vector Autoregressions. A nominal $1-\alpha$ projection region collects the structural parameters that are compatible with a $1-\alpha$ Wald ellipsoid for the model's reduced-form parameters (autoregressive coefficients and the covariance matrix of residuals).

We show that projection inference can be applied to a general class of stationary models, is computationally feasible, and—as the sample size grows large—it produces regions for the structural parameters and their identified set with both frequentist coverage and *robust* Bayesian credibility of at least $1-\alpha$.

A drawback of the projection approach is that both coverage and robust credibility may be strictly above their nominal level. Following the work of Kaido, Molinari, and Stoye (2016), we 'calibrate' the radius of the Wald ellipsoid to guarantee that—for a given posterior on the reduced-form parameters—the robust Bayesian credibility of the projection method is exactly $1-\alpha$. If the bounds of the identified set are differentiable, our calibrated projection also covers the identified set with probability $1-\alpha$.

We illustrate the main results of the paper using the demand/supply-model for the U.S. labor market in Baumeister and Hamilton (2015).

Keywords: Sign-restricted SVARs, Set-Identified Models, Projection Method.

1. INTRODUCTION

A Structural Vector Autoregression (SVAR) (Sims (1980, 1986)) is a time series model that brings theoretical restrictions into a linear, multivariate autoregression. The theoretical restrictions are used to transform *reduced-form* parameters (regression coefficients and the covariance matrix of residuals) into *structural* parameters that are more amenable to policy interpretation. Depending on the restrictions imposed, the map between reduced-form and structural parameters can be one-to-one (a point-identified SVAR) or one-to-many (a set-identified SVAR).

[1] We owe special thanks to Toru Kitagawa, Patrik Guggenberger, Francesca Molinari, and Lutz Kilian for comments and suggestions. We are also grateful to seminar participants at the 2015 Conference for Young Econometricians at Cornell University, the 2015 Interactions Conference, the econometrics workshop at Columbia University, the Federal Reserve Bank of Cleveland, and the University of Pennsylvania. This draft: June 30th, 2016. First draft: October 23th, 2015.

[2] Penn State University and National Research University Higher School of Economics.

[3] University of Bonn, Department of Economics.

[4] Columbia University, Department of Economics.

It is now customary for empirical macroeconomic studies to impose sign and/or exclusion restrictions on structural dynamic responses in SVARs in order to set-identify the model, as in the pioneering work of Faust (1998) and Uhlig (2005). The vast majority of these studies use numerical Bayesian methods to construct posterior credible sets for the coefficients of the structural impulse-response function.

Despite the popularity of the Bayesian approach, a practical concern is the fact that posterior inference for the structural parameters continues to be influenced by prior beliefs even if the sample size is infinite. This point has been documented—in detail and generality—in the work of Poirier (1998), Gustafson (2009), and Moon and Schorfheide (2012). More recently, Baumeister and Hamilton (2015) provided an explicit characterization of the influence of prior beliefs on posterior distributions for structural parameters in set-identified SVARs.

This paper studies the properties of the *projection method*—which does not rely on the specification of prior beliefs for set-identified parameters—to conduct *simultaneous* inference about the coefficients of the structural impulse-response function (and their identified set). The proposal is to 'project' a typical Wald ellipsoid for the reduced-form parameters of a VAR. The suggested nominal $1-\alpha$ projection region consists of all the structural parameters of interest compatible with the reduced-form parameters in a nominal $1-\alpha$ Wald ellipsoid.

The attractive features of the projection approach—explained in more detail in the next section—are its general applicability, its computational feasibility, and the fact that a nominal $1-\alpha$ projection region has—asymptotically and under mild assumptions—both frequentist coverage and *robust* Bayesian credibility of at least $1-\alpha$.[1] Moreover, building on Kaido et al. (2016), we show that our baseline projection can be 'calibrated' to eliminate excessive robust Bayesian credibility.

The remainder of the paper is organized as follows. Section 2 presents an overview of the projection approach. Section 3 presents the SVAR model and establishes the frequentist coverage of projection. Section 4 establishes the asymptotic robust Bayesian credibility of the projection region. Section 5 presents the 'calibration' algorithm designed to eliminate the excess of robust Bayesian credibility and shows that, under regularity conditions, our calibration also removes the excess of frequentist coverage. Section 6 discusses the implementation of projection in the context of the demand/supply SVAR for the U.S. labor market. Section 7 concludes.

[1]The robustness is relative to the choice of prior for the so-called 'rotation' matrix as in the recent work of Giacomini and Kitagawa (2015).

2. OVERVIEW AND RELATED LITERATURE

2.1. *Overview*

Let μ denote the parameters of a reduced-form vector autoregression; i.e., the slope coefficients in the regression model and the covariance matrix of residuals. Let λ denote the structural parameter of interest; i.e., the response of some variable i to a structural shock j at horizon k (or a vector of responses). In set-identified SVARs there is a known map between μ and the lower and upper bound for λ; see Uhlig (2005). Consequently, the smallest and largest value of a particular structural coefficient of interest can be written, simply and succinctly, as $\underline{v}(\mu)$ and $\overline{v}(\mu)$.

Our projection region for λ (and for its identified set) is based on a straightforward application of the classical idea of *projection* inference; see Scheffé (1953), Dufour (1990), and Dufour and Taamouti (2005, 2007). Let $\widehat{\mu}_T$ denote the sample least squares estimator for μ and let $\mathrm{CS}_T(1-\alpha;\mu)$ denote its nominal $1-\alpha$ Wald confidence ellipsoid. If, asymptotically, $\mathrm{CS}_T(1-\alpha;\mu)$ covers the parameter μ with probability $1-\alpha$, then, asymptotically, the interval

$$\mathrm{CS}_T(1-\alpha;\lambda) \equiv \Big[\inf_{\mu\in \mathrm{CS}_T(1-\alpha,\mu)} \underline{v}(\mu)\,,\ \sup_{\mu\in \mathrm{CS}_T(1-\alpha,\mu)} \overline{v}(\mu) \Big] \tag{2.1}$$

covers the set-identified parameter λ (and its identified set) with probability at least $1-\alpha$ (uniformly over a large class of data generating processes).[2,3]

In many applications there is interest in conducting *simultaneous* inference on h structural parameters; for example, if one wants to analyze the response of variable i to a structural shock j for all horizons ranging from period 1 to h as in Jordà (2009), Inoue and Kilian (2013, 2016), and Lütkepohl, Staszewska-Bystrova, and Winker (2015). In this case, the projection region given by:

$$\mathrm{CS}_T(1-\alpha;(\lambda_1,\ldots,\lambda_h)) \equiv \mathrm{CS}_T(1-\alpha;\lambda_1)\times\ldots\times \mathrm{CS}_T(1-\alpha;\lambda_h), \tag{2.2}$$

covers the structural coefficients $(\lambda_1,\ldots\lambda_h)$ and their identified set with probability at least $1-\alpha$ as the sample size grows large. The only assumption required to

[2]Formally, we show that the confidence interval described in (2.1) is *uniformly consistent of level* $1-\alpha$ for the structural parameter λ (and its identified set) over some class of data generating processes.

[3]The application of projection inference to SVARs was first suggested by Moon and Schorfheide (2012) (p. 11, NBER working paper 14882). The projection approach is also briefly mentioned in the work Kline and Tamer (2015) (Remark 8) in the context of set-identified models. None of these papers established the properties for projection inference discussed in our work.

guarantee the frequentist coverage of our projection region is the asymptotic validity of the confidence set for the reduced-form parameters, μ.

General Applicability: The validity of our projection method requires no regularity assumptions (like continuity or differentiability) on the bounds of the identified set $\underline{v}(\cdot)$ and $\overline{v}(\cdot)$. This means we can handle the typical application of set-identified SVARs in the empirical macroeconomics literature (exclusion restrictions on contemporaneous coefficients, long-run restrictions, elasticity bounds, and of course sign/zero restrictions on the responses of different variables at different horizons for different shocks).

Computational Feasibility: The implementation of our projection approach requires neither numerical inversion of hypothesis tests nor sampling from the space of rotation matrices. Instead, we use state-of-the-art optimization algorithms to solve for the maximum and minimum value of a mathematical program to compute the two end points of the confidence interval in (2.1).

Robust Bayesian Credibility: In the spirit of making our results appealing to Bayesian decision makers, we show that our suggested nominal $1-\alpha$ projection region will have—as the sample size grows large—robust Bayesian credibility of at least $1-\alpha$. This means that the asymptotic posterior probability that the vector of structural parameters of interest belongs to the projection region will be at least $1-\alpha$; for a fixed prior on the reduced-form parameters, μ, and for *any prior* on the set-identified parameters. A sufficient condition to establish the robust Bayesian credibility of projection is that the prior for μ used to compute credibility satisfies the *Bernstein-von Mises* theorem.

'Calibrated' Projection: Despite the features highlighted above, projection inference is *conservative* both for a frequentist and a robust Bayesian. That is, both the asymptotic confidence level and the asymptotic robust credibility of projection can be strictly above $1-\alpha$. Kaido et al. (2016) [henceforth, KMS] refer to the excess of frequentist coverage as *projection conservatism* and develop an innovative *calibration* approach to eliminate it.[4]

The calibration exercise in KMS requires, in the SVAR context, the computation of Monte-Carlo coverage probabilities for the projection region over an exhaustive grid of values for the reduced-form parameters, μ. In several SVAR applications, the

[4]Another recent paper proposing a procedure to eliminate the frequentist excess coverage in moment-inequality models is Bugni, Canay, and Shi (2014). Adapting their *profiling* idea to our set-up could be of theoretical interest and of practical relevance. We leave this question open for future research.

dimension of μ compromises the construction of an exhaustive grid.

Instead of insisting on removing excessive frequentist coverage, we suggest practitioners to calibrate projection to attain a robust Bayesian credibility of exactly $1-\alpha$. The calibration of robust credibility is computationally feasible even if μ is of large dimension, as no exhaustive grid for μ is needed. We provide a detailed description of our calibration procedure in Section 5. Broadly speaking, the calibration consists of drawing μ from its posterior distribution (or a suitable large-sample Gaussian approximation); evaluating the functions $\underline{v}(\mu), \overline{v}(\mu)$ for each draw of μ; and decreasing the radius defining the projection region until it contains exactly $(1-\alpha)\%$ of the values of $\underline{v}(\mu), \overline{v}(\mu)$ (for different horizons and different shocks if desired).[5] We show that if $\underline{v}(\mu), \overline{v}(\mu)$ are differentiable, our suggested calibration also removes the excessive frequentist coverage *of the identified set*.

Illustrative Example: The illustrative example in this paper is a simple demand and supply model of the U.S. labor market. We estimate standard Bayesian credible sets for the dynamic responses of wages and employment using the Normal-Wishart-Haar prior specification in Uhlig (2005) and also the alternative prior specification recently proposed by Baumeister and Hamilton (2015). The main set-identifying assumptions are sign restrictions on contemporaneous responses: an expansionary structural demand shock increases wages and employment upon impact; an expansionary structural supply shock decreases wages but increases employment, also upon impact.[6]

The Bayesian credible sets for this application illustrate the attractiveness of set-identified SVARs. The data, combined with prior beliefs, and with the (set)-identifying assumptions imply that the initial responses to demand and supply shocks persist in the medium-run, which was not restricted ex-ante.

The Bayesian credible sets for this application also illustrate how the quantitative results in set-identified SVARs could be affected by the prior specification. For example, under the prior in Baumeister and Hamilton (2015) the 5-year ahead response of employment to a demand shock could be as large as 4%; whereas under the priors in Uhlig (2005) the same effect is at most 2%.

Our baseline projection approach (which takes around 15 minutes) allows us to

[5]In Section 6 we provide more detials on the computation time of our calibrated projection (which is around 5 hours in our illustrative example).

[6]Following Baumeister and Hamilton (2015) we also consider elasticity bounds on the wage elasticity of both labor demand and labor supply, and also bounds on the long-run impact of a demand shock over employment.

get a prior-free assessment about the magnitude (and direction) of the structural responses of interest. For example, the largest value in our projection region for the 5-year response of employment to a structural demand shock is around 2.5%. This effect is larger than the one implied by the prior in Uhlig (2005), but smaller than the one implied by the priors in Baumeister and Hamilton (2015).

Our baseline projection approach—though informative about the effects of demand shocks—is not conclusive about the medium-run effects of structural supply shocks on wages and employment (the projection region allows for both positive and negative responses). This could be a consequence of either the robustness of projection or its conservativeness. To disentangle these effects, we calibrate projection to guarantee that it has exact robust Bayesian credibility. The calibrated projection shows that an expansionary supply shock will decrease wages in each quarter over a 5 year horizon. However, the qualitative effects of supply shocks on employment remain undetermined. The simple SVAR for the labor market illustrates the usefulness of both the baseline and the calibrated projection to analyze the robustness of quantitative and qualitative results in SVARs to prior beliefs.

2.2. ***Related Literature***

There has been recent interest in departing from the standard Bayesian analysis of set-identified SVARs in an attempt to provide robustness to the choice of priors. Below we provide a short description of the similarities and differences between our projection approach and three alternative methods available in the literature. It is worth mentioning that our baseline projection approach is the only procedure (among the three alternative methods discussed) that has both a frequentist and a robust Bayes interpretation. In addition, none of the other approaches allow for simultaneous inference on a vector of impulse-response coefficients (a feature that has been deemed desirable in point-identified SVARs). Our baseline projection achieves all these properties while retaining computational tractability (we solve two mathematical programs per coefficient of interest).

a) In a pioneering paper, Moon, Schorfheide, and Granziera (2013) [MSG] proposed both projection and Bonferroni frequentist inference using a moment-inequality, minimum distance framework based on Andrews and Soares (2010). In terms of applicability, their procedures are designed for set-identified SVARs that impose restrictions on the dynamic responses of only one structural shock. It is possible to extend their approach to the same class of modes that we consider; there is, however, a serious issue regarding computational feasibility. Specifically, both the projection and Bonferroni approaches require the researcher to compute—by simulation—a critical value for each single orthogonal matrix of dimension $n \times n$, where n is the dimension of the SVAR. Our baseline implementation of the projection method does not require any type of grid over the space of orthogonal matrices and does not require the simulation of any critical value.

b) Giacomini and Kitagawa (2015) [GK] develop a novel and generally applicable robust Bayesian approach to conduct inference about a specific coefficient of the impulse-response function in a set-identified SVAR. In terms of our notation, their procedure can be described as follows. One takes posterior draws from μ and evalutes, at each posterior draw, the functions $\underline{v}(\mu), \overline{v}(\mu)$ by solving a nonlinear program. Their credible set is a numerical (grid-search) approximation to the *smallest* interval that covers $100(1-\alpha)\%$ of the posterior realizations of the identified set.

GK and Baseline projection: In terms of properties, our baseline projection is shown to admit both a frequentist and a robust Bayes interpretation (the GK approach has only been shown to admit the latter). In terms of implementation, GK solve as many nonlinear programs as posterior draws for μ. This means that our baseline procedure will be typically faster to implement than the GK robust procedure (since our baseline projection only needs to solve two nonlinear programs). The price to pay for the reduced computational cost is the excess of robust Bayesian credibility.

GK and Calibrated projection: Our calibrated projection requires a similar amount of work as the GK robust method (both procedures evaluate the bounds of the identified set for each posterior draw). There are two differences remaining between the two approaches. First, our calibrated projection allows for *simultaneous* credibility statements over different horizons, different variables, and different shocks. Second, our calibrated projection is guaranteed to have correct frequentist coverage whenever the bounds of the identified set are differentiable in μ.

c) Gafarov, Meier, and Montiel Olea (2015) [GMM1] establish the differentiability of the bounds of $\underline{v}(\mu), \overline{v}(\mu)$ for a class of SVAR models that impose restrictions only on the responses to one structural shock. Based on the differentiability results, they propose a 'delta-method' confidence interval given by the plug-in estimators of the bounds plus/minus r times standard errors. In Appendix C we show that, in large samples, the 'delta-method' procedure in GMM1 is equivalent to a projection region based on a Wald ellipsoid for μ with radius r^2.

3. BASIC MODEL, MAIN ASSUMPTIONS, AND FREQUENTIST RESULTS

3.1. *Model*

This paper studies the n-dimensional Structural Vector Autoregression with p lags; i.i.d. structural innovations—denoted ε_t—distributed according to F; and unknown $n \times n$ structural matrix B:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + B\varepsilon_t, \;\; \mathbb{E}_F[\varepsilon_t] = 0_{n\times 1}, \;\; \mathbb{E}_F[\varepsilon_t \varepsilon_t'] \equiv \mathbb{I}_n. \tag{3.1}$$

see Lütkepohl (2007), p. 362.

The *reduced-form parameters* of the SVAR model are defined as the vectorized autoregressive coefficients and the half vectorized covariance matrix of reduced-form residuals:

$$\mu \equiv (\text{vec}(A)', \text{vech}(\Sigma)')' \in \mathbb{R}^d, \quad \text{where} \quad A \equiv (A_1, A_2, \ldots, A_p), \quad \Sigma \equiv BB'.$$

In applied work, these reduced-form parameters are estimated directly from the data using least squares. That is:

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

where

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

and

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

A common formula for the asymptotic variance of $\widehat{\mu}_T$ in stationary models is:

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

where

$$V_T \equiv \begin{pmatrix} \mathbb{I}_n \otimes \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} & \mathbf{0} \\ \mathbf{0} & L_n \end{pmatrix},$$

and L_n is the matrix of dimension $n(n+1)/2 \times n^2$ such that $\text{vech}(\Sigma) = L_n \text{vec}(\Sigma)$, see Lütkepohl (2007), p. 662 equation A.12.1.

3.2. *Assumptions for frequentist inference*

The SVAR parameters $(A_1, \ldots, A_p, B, F)$ define a probability measure, denoted P, over the data observed by the econometrician. The measure P is assumed to belong to some class $\mathcal{P}$ which we describe in this section.

We state a simple high-level assumption concerning the asymptotic behavior of the $1-\alpha$ Wald confidence ellipsoid for μ, which is defined as:

$$CS_T(1-\alpha;\mu) \equiv \Big\{\mu \in \mathbb{R}^d \quad | \quad T(\widehat{\mu}_T - \mu)'\widehat{\Omega}_T^{-1}(\widehat{\mu}_T - \mu) \leq \chi^2_{d,1-\alpha}\Big\}.^7 \tag{3.2}$$

The first assumption requires the *uniform consistency in level* (over the class $\mathcal{P}$) of the Wald confidence set for the reduced-form parameters. This is:

Assumption 1 $\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} P\Big(\mu(P) \in CS_T(1-\alpha;\mu)\Big) \geq 1-\alpha.$

Assumption 1 holds if the class $\mathcal{P}$ under consideration contains only *uniformly* stable VARs where the error distributions, F, have *uniformly* bounded fourth moments.[8] Assumption 1 turns out to be sufficient to conduct frequentist inference on the structural parameters of a set-identified SVAR, defined as follows.

Coefficients of the Structural Impulse-Response Function: Given the autoregressive coefficients $A \equiv (A_1, A_2, \ldots, A_p)$ define, recursively, the nonlinear transformation

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

where $C_0 = \mathbb{I}_n$ and $A_m = 0$ if $m > p$; see Lütkepohl (1990), p. 116.

Definition (Coefficients of the Structural IRF) The (k,i,j)-*coefficient of*

[7]The radius $\chi^2_{d,1-\alpha}$ in equation (3.2) denotes the $1-\alpha$ quantile of a central χ^2 distribution with d degrees of freedom.

[8]A class $\mathcal{P}$ that satisfies Assumption 1 could be written by using a uniform version of the conditions in Lütkepohl (2007), p. 73. This is, there are positive constants c_1, c_2, c_3, c_4 such that:

$$\begin{aligned}\mathcal{P} \;=\; \{(A_1, A_2, \ldots, A_p, B, F) \mid \;& \det(\mathbb{I}_n - A_1 z - \ldots A_p z^p) \notin (-c_1, c_1) \text{ for } z \in \mathbb{C}, |z| \leq 1; \\ & B \text{ is such that } 0 < c_2 < \text{eigmin}(BB') < \text{eigmax}(BB') < c_3; \text{and } \mathbb{E}_F[|\varepsilon_{n_1,t}\varepsilon_{n_2,t}\varepsilon_{n_3,t}\varepsilon_{n_4,t}|] < c_4 \\ & \text{for all } t \text{ and } n_1, n_2, n_3, n_4 \in \{1, \ldots n\}, \text{ and } \mathbb{E}_F[\varepsilon_t] = \mathbf{0_{n\times 1}},\ \mathbb{E}_\mathbf{F}[\varepsilon_\mathbf{t}\varepsilon'_\mathbf{t}] = \mathbb{I}_\mathbf{n} \quad \}.\end{aligned}$$

Other possible definitions of $\mathcal{P}$ can be given by generalizing Theorem 3.5 in Chen and Fang (2015) to either multivariate linear processes with i.i.d. innovations or to martingale difference sequences.

the structural impulse-response function is defined as the scalar parameter:

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

where e_i and e_j denote the i-th and j-th column of the identity matrix $\mathbb{I}_n$.

3.3. *Main result concerning frequentist inference*

In this section we show that, under Assumption 1, it is possible to 'project' the $1-\alpha$ Wald confidence set for μ to conduct frequentist inference about the coefficients of the structural impulse-response function and the function itself in set-identified models.

Set-Identified SVARs: As mentioned in the introduction, the SVAR allows researchers to transform the reduced-form parameters, $\mu \equiv (\text{vec}(A)', \text{vech}(\Sigma)')'$, into the structural parameters of interest, $\lambda_{k,i,j}(A,B)$. The parameter μ determines a unique value of A; however, several values of B are compatible with Σ (any B such that $BB' = \Sigma$). This indeterminacy of B implies there are multiple values of $\lambda_{k,i,j}(A,B)$ that are compatible with one value of μ.

The Identified Set and its Bounds: It is common in applied macroeconomic work to impose restrictions on the matrix $B \in \mathbb{R}^{n\times n}$ in order to limit the range of a structural coefficient of interest, $\lambda_{k,i,j}$ (taking μ as given). Mathematically, a set of restrictions on B—that we denote as $\mathcal{R}(\mu)$—can be interpreted as a subset of $\mathbb{R}^{n\times n}$. This leads to the following definition:

Definition (**Identified Set and its bounds**) Fix a vector of reduced-form parameters, μ, and a set of restrictions $\mathcal{R}(\mu)$ on B.

a) The *identified set* for the structural parameter $\lambda_{k,i,j}(A,B)$ is defined as:

$$\mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu) \equiv \Big\{ v \in \mathbb{R} \;\Big|\; v = \lambda_{k,i,j}(A,B), \;\; BB' = \Sigma, \;\text{ and } B \in \mathcal{R}(\mu) \Big\}. \tag{3.3}$$

b) The *upper bound* of the identified set $\overline{v}_{k,i,j}(\mu)$ is defined as the value function of the program:

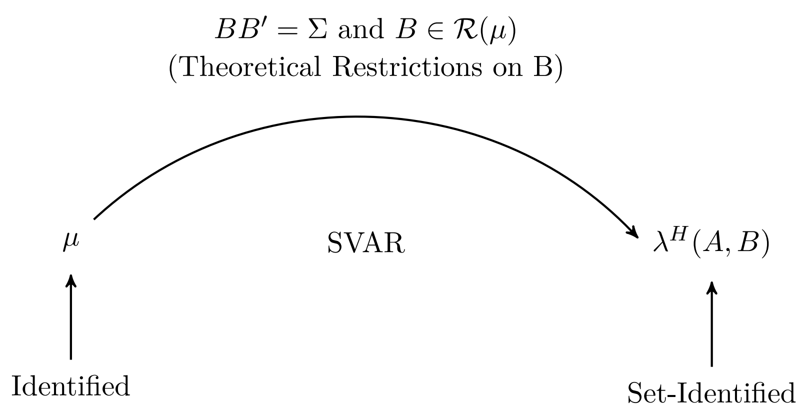
$$\overline{v}_{k,i,j}(\mu) \equiv \sup_{B \in \mathbb{R}^{n\times n}} e_i' C_k(A) B e_j, \quad \text{s.t.} \quad BB' = \Sigma, \text{ and } B \in \mathcal{R}(\mu). \tag{3.4}$$

The *lower bound*, $\underline{v}_{k,i,j}(\mu)$, is defined analogously.

c) Consider any collection $\lambda^H \equiv \{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ of structural coefficients and let its identified set be given by:

$$\mathcal{I}_H^{\mathcal{R}}(\mu) \equiv \Big\{(v_1,\ldots,v_H) \in \mathbb{R}^H \;\Big|\; v_h = \lambda_{k_h,i_h,j_h}(A,B),\ BB' = \Sigma,\ \text{ and } B \in \mathcal{R}(\mu)\Big\}.$$

The main elements in the previous definition can be illustrated as follows:



$$\mathcal{I}_H^{\mathcal{R}}(\mu) \subseteq \mathbb{R}^H : \text{ The identified-Set for } \lambda^H(A,B).$$

Table I presents a list of the most common restrictions, $\mathcal{R}(\mu)$, used in SVAR analysis (all of which can be handled by our frequentist approach described below).

Projection Approach: A key feature of set-identified SVARs is that the bounds of the identified set depend on a finite-dimensional parameter. 'Projecting' down the $1-\alpha$ Wald ellipsoid for μ seems a natural approach to conduct inference on the structural impulse response function. The first result in this paper establishes the frequentist uniform validity of projection inference.

Result 1 (Frequentist Coverage of Projection Inference for λ^H) *Consider the projection region for the collection of structural coefficients $\lambda^H \equiv \{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ given by:*

$$CS_T(1-\alpha,\lambda^H) \equiv CS_T(1-\alpha,\lambda_{k_1,i_1,j_1}) \times \ldots \times CS_T(1-\alpha,\lambda_{k_H,i_H,j_H}) \subseteq \mathbb{R}^H, \tag{3.5}$$

where

$$CS_T(1-\alpha;\lambda_{k,i,j}) \equiv \Big[\inf_{\mu \in CS_T(1-\alpha,\mu)} \underline{v}_{k,i,j}(\mu)\, , \sup_{\mu \in CS_T(1-\alpha,\mu)} \overline{v}_{k,i,j}(\mu) \Big], \tag{3.6}$$

and $CS_T(1-\alpha;\mu)$ is the $1-\alpha$ Wald confidence ellipsoid for μ. If the class of data generating processes $\mathcal{P}$ satisfies Assumption 1, then:

$$\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} \inf_{\lambda^H \in \mathcal{I}_H^{\mathcal{R}}(\mu(P))} P\Big(\lambda^H \in CS_T(1-\alpha;\lambda^H)\Big) \geq 1-\alpha.$$

That is, the projected confidence interval in (3.5) covers the vector of structural coefficients λ^H with probability at least $1-\alpha$, uniformly over the class $\mathcal{P}$.

Proof: The proof of Result 1 uses a standard and conceptually straightforward projection argument. Take an element $P \in \mathcal{P}$ and let $\lambda^H \in \mathbb{R}^H$ be any given element of the identified set $\mathcal{I}_H^{\mathcal{R}}(\mu(P))$. Note that:

$$P\Big(\lambda^H \in CS_T(1-\alpha;\lambda^H)\Big)$$

$$\begin{aligned}
= \;& P\Big((\lambda_{k_1,i_1,j_1},\ldots,\lambda_{k_H,i_H,j_H}) \in CS_T(1-\alpha;\lambda_{k_1,i_1,j_1}) \times \ldots \times CS_T(1-\alpha;\lambda_{k_H,i_H,j_H})\Big) \\
& \Big(\text{by definition of our confidence interval for } \lambda^H\Big) \\
\geq \;& P\Big([\underline{v}_{k_h,i_h,j_h}(\mu(P))\, , \, \overline{v}_{k_h,i_h,j_h}(\mu(P))] \subseteq \Big[\inf_{\mu \in \mathrm{CS}_T(1-\alpha,\mu)} \underline{v}_{k_h,i_h,j_h}(\mu)\, , \sup_{\mu \in \mathrm{CS}_T(1-\alpha,\mu)} \overline{v}_{k_h,i_h,j_h}(\mu) \Big] \\
& \quad \forall h = 1,\ldots,H\Big), \\
& \Big(\text{since } \lambda_{k_h,i_h,j_h} \in [\underline{v}_{k_h,i_h,j_h}(\mu(P), \overline{v}_{k_h,i_h,j_h}(\mu(P))]\Big) \\
\geq \;& P\Big(\mu(P) \in CS_T(1-\alpha;\mu)\Big).
\end{aligned}$$

The desired result follows directly from Assumption 1. This shows that the projection region for λ^H is uniformly consistent in level. *Q.E.D.*

Table I: Common Restrictions Used in Set-Identified SVARs

(i denotes the variable, j denotes the shock, and k the horizon)

Restrictions	**Description**	**Notation**	**Examples**
Short-run	Exclusion Restrictions imposed on B or B'^{-1}	$e_i' B e_j = 0$ or $e_i' B'^{-1} e_j = 0$ (Note that $B'^{-1} = \Sigma^{-1} B$)	Sims (1980) Christiano et al. (1996) Rubio-Ramirez et al. (2015)
Long-run	A zero constraint on the long-run impact matrix	$e_i' (\mathbb{I}_n - A_1 - A_2 - \ldots A_p)^{-1} B e_j = 0$	Blanchard and Quah (1989)
Sign	Sign restrictions on IRFs	$e_i' C_k(A) B e_j \geq, \leq 0$	Uhlig (2005) Mountford and Uhlig (2009)
Elasticity Bounds	Bounds on the elasticity of a variable	$\frac{e_i' C_k(A) B e_j}{e_{\tilde{i}}' C_k(A) B e_j} \geq, \leq c, \; \tilde{i} \neq i$	Kilian and Murphy (2012)
Shape Constraints	Shape constraints on IRFs (e.g., monotonicity)	e.g., $e_i' C_k(A) B e_j \leq e_i' C_{k+1}(A) B e_j$	Scholl and Uhlig (2008)
Other	Sign Restrictions on Long-Run Impacts Noncontemporaneous Zero Restrictions General equalities/inequalities on B	$e_i' (\mathbb{I}_n - A_1 - A_2 - \ldots A_p)^{-1} B e_j \geq, \leq 0$ $e_i' C_k(A) B e_j = 0$ $g(B, \mu) \geq, \leq, = 0$	

The projection approach can handle SVAR models with any of the restrictions described on this table (imposed on one or multiple shocks).

REMARK 1: The idea of 'projecting' a confidence set for a parameter μ to conduct inference about a lower dimensional parameter λ has been used extensively in econometrics; see Scheffé (1953), Dufour (1990), and Dufour and Taamouti (2005, 2007) for some examples. In addition to its conceptual simplicity, one advantage of the projection approach is that its validity does not require special conditions on the identifying restrictions that can be imposed by practitioners.[9]

REMARK 2: The problem of conducting inference on the whole impulse-response function (and not only on one specific coefficient) has been a topic of recent interest, both from the Bayesian and frequentist perspective.

For Bayesian set-identified SVARs with only sign restrictions, Inoue and Kilian (2013) report the vector of structural impulse-response coefficients with highest posterior density (based on a prior on reduced-form parameters and a uniform prior on rotation matrices). They propose a Bayesian credible set (represented by *shotgun* plots) that characterizes the joint uncertainty about a given collection of structural impulse-response coefficients.

For frequentist point-identified SVARs, Inoue and Kilian (2016) propose a bootstrap procedure that allows the construction of asymptotically valid confidence regions for any subset of structural impulse responses. To the best of our knowledge, our projection approach is the first frequentist procedure for set-identified SVARs that provide confidence regions for any collection of structural coefficients (response of different variables, to different shocks, over different horizons).

It is important to note that Uhlig (2005)'s approach to conduct inference on set-identified SVARs does not provide credible sets for vectors of the structural parameters. The same is true for the Bayesian approaches described in the recent work of Arias, Rubio-Ramirez, and Waggoner (2014) and Baumeister and Hamilton (2015), as well as the approaches of Moon et al. (2013) and Giacomini and Kitagawa (2015).

REMARK 3: A common concern in set-identified models is whether the suggested inference approach is valid only for the identified parameter, λ^H, or also for its identified set $\mathcal{I}_H^{\mathcal{R}}(\mu)$. Note that the second to last inequality in the proof of Result 2 imply that our projection region covers the identified set of any vector of coefficients λ^H.

[9]For instance, we do not need to assume that $\underline{v}_{k,i,j}(\cdot)$ and $\overline{v}_{k,i,j}(\cdot)$ are continuous or differentiable functions of the reduced-form parameters.

4. ROBUST BAYESIAN CREDIBILITY

This section analyzes the *robust credibility* of projection as the sample size grows large.

Bayesian Set-up: In a Bayesian SVAR the distribution of the structural innovations is fixed and treated as a known object. A common choice—which we follow in this section—is to assume that $F \sim \mathcal{N}_n(0, \mathbb{I}_n)$. We discuss how to relax this restriction after stating Assumption 2.

Let P^* denote some prior for the structural parameters $(A_1, \ldots, A_p, B)$ and let $\lambda^H(A, B) \in \mathbb{R}^H$ denote the vector of structural coefficients of interest. For a given square root of $\Sigma \equiv BB'$ define the 'rotation' 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 rotation matrix, conditional on μ.[10] Following this notation, let $\mathcal{P}(P^*_\mu)$ denote the class of prior distributions such that $\mu \sim P^*_\mu$.

We are interested in characterizing the smallest posterior probability that the set $\text{CS}_T(1-\alpha; \lambda^H)$ could receive, allowing the researcher to vary the prior for Q:

$$\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A, B) \in CS_T(1-\alpha; \lambda^H) \, \Big| \, Y_1, \ldots, Y_T\Big). \tag{4.1}$$

The event of interest is whether the structural coefficients $\lambda^H(A, B)$ (treated as random variables in the Bayesian Set-up) belong to the projection region, after conditioning on the data. This event would typically be referred to as the credibility of $\text{CS}_T(1-\alpha; \lambda^H)$ (see Berger (1985), p. 140). We would like to find the smallest credibility of projection when different priors over Q are considered as in the pioneering work of Kitagawa (2012). We follow the recent work of Giacomini and Kitagawa (2015) and refer to (4.1) as the *robust Bayesian credibility* of the set $\text{CS}_T(1-\alpha, \lambda^H)$.

Let $f(Y_1, ..., Y_T|\mu)$ denote the Gaussian statistical model for the data (which depends solely on the reduced-form parameters) and let $o_p(1; Y_1, \ldots Y_T|\mu)$ denote a random variable such that $\lim_{T\to\infty} P_{Y_1,\ldots,Y_T|\mu}(|o_p(1; Y_1, \ldots Y_T|\mu)| > \epsilon) = 0$ for all $\epsilon > 0$ when the distribution of the data is conditioned on μ.

Main Assumption for Bayesians: Robust credibility can be viewed as a random variable (as it depends on $Y_1, \ldots, Y_T$). We use the following high-level assumption to characterize its asymptotic behavior:

[10] Arias et al. (2014) refer to this parameterization of the SVAR model as the orthogonal reduced-form.

Assumption 2 Whenever $Y_1, \ldots, Y_T \sim f(Y_1, \ldots, Y_T|\mu_0)$, the prior P^* is such that:

$$P^*\Big(\mu(A,B) \in CS_T(1-\alpha;\mu) \,\Big|\, Y_1, \ldots, Y_T\Big) = 1 - \alpha + o_p(1; Y_1, \ldots, Y_T|\mu_0).$$

Assumption 2 requires the prior over the reduced-form parameters (and the statistical model) to be regular enough to guarantee that the asymptotic Bayesian credibility of the $1-\alpha$ Wald ellipsoid converges in probability to $1-\alpha$. Thus, our high-level assumption is implied by the Bernstein von-Mises Theorem (DasGupta (2008), p. 291) for the reduced-form parameter μ.

Since the Gaussian statistical model $f(Y_1, \ldots Y_T|\mu_0)$ can be shown to be Locally Asymptotically Normal (LAN) whenever A_0 is stable and Σ_0 has full rank, Theorem 1 and 2 in Ghosal, Ghosh, and Samanta (1995) (GGS) imply that Assumption 2 will be satisfied whenever P^*_μ has a continuous density at μ_0 with polynomial majorants.[11] In fact, the same theorems could be used to establish Assumption 2 for non-Gaussian SVARs that are LAN and satisfy the regularity conditions of Ibragimov and Has' minskii (2013) (IH), as long as $\text{CS}_T(1-\alpha;\mu)$ is centered at the Maximum Likelihood estimator of μ and $\widehat{\Omega}_T$ is replaced by the model's inverse information matrix. An alternative approach to establish Assumption 2 using a different set of primitive conditions can be found in the recent work of Connault (2016).

We now establish the robust Bayesian credibility of projection as $T \to \infty$.

Result 2 *[**Asymptotic Robust Bayesian Credibility of Projection**] Suppose that the prior P^* for (A,B) satisfies Assumption 2 at μ_0. Then:*

$$\inf_{P^* \in \mathcal{P}^*(\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha;\lambda^H) \,\Big|\, Y_1, \ldots, Y_T\Big) \geq 1 - \alpha + o_p(1; Y_1, \ldots Y_T|\mu_0).$$

Proof: Note that:

$$P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha;\lambda^H) \,\Big|\, Y_1, \ldots Y_T\Big)$$

[11] In Appendix A.1 we verify an 'almost sure' version of Assumption 2 for a Gaussian SVAR for the Normal-Wishart priors suggested in Uhlig (1994) and Uhlig (2005) and a confidence set for μ based on the formula for the asymptotic variance $\widehat{\Omega}_T$ that obtains in the Gaussian model [Lütkepohl (2007) p. 93].

$$= P^*\Big(\lambda_{k_h,i_h,j_h}(A,B) \in CS_T(1-\alpha;\lambda_{k_h,i_h,j_h})\; \forall\, h = 1\ldots,H \;\Big|\; Y_1,\ldots,Y_T\Big)$$
$$\Big(\text{by definition of the projection region for } \lambda^H\Big)$$

$$\geq P^*\Big([\underline{v}_{k_h,i_h,j_h}(\mu(A,B))\,,\,\overline{v}_{k_h,i_h,j_h}(\mu(A,B))] \in CS_T(1-\alpha;\lambda_{k_h,i_h,j_h}) \quad \forall\, h = 1\ldots, H \;\Big|\; Y_1,\ldots,Y_T\Big),$$
$$\Big(\text{since } \lambda_{k_h,i_h,j_h}(A,B) \in [\underline{v}_{k_h,i_h,j_h}(\mu(A,B)),\overline{v}_{k_h,i_h,j_h}(\mu(A,B))] \text{ for any } A,B\Big)$$
$$\geq P^*\Big(\mu(A,B) \in CS_T(1-\alpha;\mu) \;\Big|\; Y_1,\ldots,Y_T\Big).$$

This implies that in any finite sample:

$$\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha;\lambda^H) \;\Big|\; Y_1,\ldots,Y_T\Big)$$

is at least as large as

$$P^*\Big(\mu(A,B) \in CS_T(1-\alpha;\mu) \;\Big|\; Y_1,\ldots,Y_T\Big).$$

Assumption 2 gives the desired result. *Q.E.D.*

This means that—given any prior that satisfies Assumption 2—our projection region can be interpreted, in large samples, as a robust $1-\alpha$ credible region for the impulse-response function and its coefficients.

5. CALIBRATED PROJECTION FOR A ROBUST BAYESIAN

The projection approach generates *conservative* regions for both a frequentist and a robust Bayesian. For a frequentist, the large-sample coverage may be strictly above the desired confidence level. For a robust Bayesian, the asymptotic robust credibility of the nominal $1-\alpha$ projection region may be strictly above $1-\alpha$.

This section applies the approach in Kaido et al. (2016) to eliminate the excess of robust Bayesian credibility in a computationally tractable way. We focus on calibrating the robust credibility of our projection region to be exactly equal to $1-\alpha$ (either in a finite sample for a given prior on μ, or in large samples for a large class of priors on μ).[12]

Given a vector $\Lambda^H = \{\lambda_{k_h,i_h,j_h}\}_{h=1}^H$ of structural coefficients of interest and its corresponding nominal $1-\alpha$ projection region, the calibration exercise is based on the following result:

Result 3 ***(Calibration of Robust Credibility)*** *Let P^*_μ denote a prior for the reduced-form parameters. Suppose there is a nominal level $1-\alpha^*(Y_1,\ldots,Y_T)$ such that for every data realization:*

$$P^*_\mu\left(\times_{h=1}^H[\underline{v}_{k_h,i_h,j_h}(\mu),\overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq CS_T(1-\alpha^*(Y_1,\ldots,Y_T),\lambda^H)|Y_1,\ldots,Y_T\right)$$

equals $1-\alpha$. Then, for every data realization:

$$\inf_{P^*\in\mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B)\in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H)\;\Big|\;Y_1,\ldots,Y_T\Big) = 1-\alpha.$$

Proof: See Appendix A.2. *Q.E.D.*

This means that in order to calibrate the robust credibility of projection, it is sufficient to choose $1-\alpha^*(Y_1,\ldots,Y_T)$ to guarantee that exactly $\alpha\%$ of the bounds of the identified set for the different structural coefficients in λ^H fall outside the projection region.

Calibration Algorithm: The calibration algorithm we propose consists in finding a nominal level $1-\alpha^*(Y_1,\ldots,Y_T)$ such that the posterior probability of the event:

$$[\underline{v}_{k_1,i_1,j_1}(\mu),\overline{v}_{k_1,i_1,j_1}(\mu)]\times\ldots\times[\underline{v}_{k_h,i_h,j_h}(\mu),\overline{v}_{k_h,i_h,j_h}(\mu)]\subseteq \mathrm{CS}_T(1-\alpha^*,\lambda^H)$$

[12] We also discuss the calibration of projection in SVARs from the frequentist perspective (see Appendix B). We argue that the computational feasibility of the frequentist calibration might be compromised when μ is of large dimension.

equals $1 - \alpha$ under the posterior distribution associated to the prior P^*_μ or under a suitable large-sample approximation for the posterior such as $\mu | Y_1, \ldots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T / T)$.[13]

The calibration algorithm is the following:

1. Generate M draws (for example, $M = 1,000$) from the posterior of the reduced-form parameters. If desired, one could use the large-sample approximation of the posterior given by:

$$\mu^*_m \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T / T).$$

2. Let $\lambda^H = \{\lambda_{k_h, i_h, j_h}\}_{h=1}^H$ denote the structural coefficients of interest. For each $h = 1, \ldots H$ and for each $m = 1, \ldots M$ evaluate:

$$[\underline{v}_{k_h, i_h, j_h}(\mu^*_m), \overline{v}_{k_h, i_h, j_h}(\mu^*_m)],$$

   as defined in equation (3.4). We provide Matlab code to evaluate these bounds.

3. Fix an element α_s on the interval $(\alpha, 1)$. Set a tolerance level $\eta > 0$.
4. For each $m = 1, \ldots M$ generate the indicator function z_m that takes the value of 0 whenever there exists an index $h \in \{1, \ldots H\}$ such that:

$$[\underline{v}_{k_h, i_h, j_h}(\mu^*_m), \overline{v}_{k_h, i_h, j_h}(\mu^*_m)] \notin \text{CS}_T(1 - \alpha_s, \lambda_{k_h, i_h, j_h}).$$

   The projection region $\text{CS}_T(1 - \alpha_s, \lambda_{k_h, i_h, j_h})$ is defined in equation (3.6) in Result 3 and implemented using the SQP/IP algorithm that will described in the next section (Section 6).

5. Compute the robust credibility of the nominal $1 - \alpha_s$ projection as:

$$RC_T(\alpha_s) = \frac{1}{M} \sum_{m=1}^{M} z_m.$$

   If such quantity is in the interval $[1 - \alpha - \eta, 1 - \alpha + \eta]$ stop the algorithm. If $\text{RC}_T(\alpha_s)$ is strictly above $1 - \alpha + \eta$, go back to Step 3 and choose a larger value of α_s. If $\text{RC}_T(\alpha_s)$ is strictly below $1 - \alpha - \eta$ go back to Step 3 and choose a smaller value of α_s.

[13]The Gaussian approximation for the posterior will eliminate projection bias asymptotically provided a Berstein von-Mises Theorem for μ holds. We establish this result in Appendix A.4.

We now show that whenever the bounds of the identified set for each λ_h are differentiable, our calibration algorithm also removes the excess of frequentist coverage.

Result 4 (**Robust Bayes Calibration and the Frequentist Coverage of the Identified Set**) *Suppose that for each $h = 1, \ldots H$ the bounds of the identified set $\underline{v}_h(\mu)$ and $\overline{v}_h(\mu)$ are differentiable at μ_0. Suppose in addition that at μ_0:*

1. $\sqrt{T}(\widehat{\mu} - \mu_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$,
2. $\widehat{\Omega}_T \xrightarrow{p} \Omega$, *where Ω is positive definite,*
3. *The prior for the reduced-form parameters used in the calibration satisfies the Bernstein von-Mises Theorem in Ghosal et al. (1995):*

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

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

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq CS_T(1 - \alpha^*(Y_1, \ldots, Y_T); \lambda_h), \ \forall h = 1, \ldots H\right) \to 1 - \alpha.$$

Proof: See Appendix A.3. The heuristic argument behind this result is the following. We show that the differentiability of $\underline{v}_h$ and $\overline{v}_h$ at μ_0 implies that:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \mathrm{CS}_T(1 - \alpha^*(Y_1, \ldots, Y_T); \lambda_h), \ \forall h = 1, \ldots H\right)$$

is approximately the same as:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \left[\underline{v}_h(\widehat{\mu}_T) - \frac{r_T^* \underline{\sigma}_h(\mu_0)}{\sqrt{T}} \, , \, \overline{v}_h(\widehat{\mu}_T) + \frac{r_T^* \overline{\sigma}_h(\mu_0)}{\sqrt{T}} \right], \ \forall h \leq H \right).$$

where r_T^* be the radius that calibrates robust Bayesian credibility. The Bernstein-von Mises Theorem implies that such probability is approximately the same as:

$$P^*\left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \left[\underline{v}_h(\widehat{\mu}_T) - \frac{r_T^* \underline{\sigma}_h(\mu_0)}{\sqrt{T}} \, , \, \overline{v}_h(\widehat{\mu}_T) + \frac{r_T^* \overline{\sigma}_h(\mu_0)}{\sqrt{T}} \right], \ \forall h \leq H \right),$$

which, by the calibration of robust Bayesian credibility is approximately $1 - \alpha$. *Q.E.D.*

6. IMPLEMENTATION OF BASELINE AND CALIBRATED PROJECTION

6.1. *Projection as a mathematical optimization problem*

This subsection discusses the implementation of the baseline projection region:

$$\text{CS}_T(1-\alpha;\lambda_{k,i,j}) \equiv \Big[\inf_{\mu \in \text{CS}_T(1-\alpha,\mu)} \underline{v}_{k,i,j}(\mu) \, , \sup_{\mu \in \text{CS}_T(1-\alpha,\mu)} \overline{v}_{k,i,j}(\mu) \Big].$$

We note that both the upper bound and lower bound of this confidence interval can be thought of as solutions to a pair of 'nested' optimization problems.

The first optimization problem—that we refer to as the *inner* optimization—solves for $\overline{v}_{k,i,j}(\mu)$ and $\underline{v}_{k,i,j}(\mu)$. These functions correspond to the largest and smallest value of the structural impulse response $\lambda_{k,i,j}$ given a set of restrictions and a vector of reduced-form parameters μ.

The second optimization problem—that we refer to as *outer* optimization—solves for the maximum value of $\overline{v}_{k,i,j}(\cdot)$ and the minimum value of $\underline{v}_{k,i,j}(\cdot)$ over the $(1-\alpha)$ Wald Confidence ellipsoid, $\text{CS}_T(1-\alpha,\mu)$.

Implementation: Our proposal is to combine the inner and outer problem into a *single* mathematical program that gives the bounds of the projection confidence interval directly. The upper bound can be found by solving:

$$\sup_{A,\Sigma,B} e_i' C_k(A) B e_j \quad \text{subject to} \quad BB' = \Sigma, \quad B \in \mathcal{R}(\mu), \text{ and} \qquad (6.1)$$
$$T(\widehat{\mu}_T - \mu(A,\Sigma))' \widehat{\Omega}_T^{-1} (\widehat{\mu}_T - \mu(A,\Sigma)) \leq \chi^2_{d,1-\alpha}.$$

The lower bound of the projection confidence interval can be found analogously. Importantly, the simple reformulation in (6.1) allows us to base the implementation of our projection region upon state-of-the-art solution algorithms for optimization problems. Our suggestion is to use a simple SQP/IP algorithm.

6.2. *Solution algorithms to implement baseline projection*

The nature of the optimization problem: The nonlinear mathematical program in (6.1) has two challenging features. On the one hand, the optimization problem is non-convex; this complicates the task of finding a global minimum with algorithms designed to detect local optima. On the other hand, the number of optimization arguments and constraints increases quadratically in the dimension of the SVAR; this compromises the feasibility of some optimization routines designed to detect global optima (for example, brute-force grid search on $\mathrm{CS}_T(1-\alpha,\mu)$ to optimize $\underline{v}_{k,i,j}(\mu)$ and $\overline{v}_{k,i,j}(\mu)$).

Our Approach: Taking these two features into consideration, we first implemented projection by running a local optimization algorithm followed by a global algorithm that used the local solution as an input. The algorithms and the functions used to implement the projection confidence interval are described below. In the application analyzed in this paper, the global stage of the algorithm did not have any impact on the local solution. We thus suggest researchers to implement our approach using only the SQP/IP routine described below.

Local Algorithms: Although no standard classification exists for local optimization algorithms, the most common procedures are often grouped as follows: penalty and Augmented Lagrangian Methods; Sequential Quadratic Programming (SQP); and Interior Point Methods (IP); see p. 422 of Nocedal and Wright (2006) for more details.

Within this class of algorithms, we focus on the IP and SQP algorithms, both of which are considered as the "*most powerful algorithms for large-scale nonlinear programming*", Nocedal and Wright (2006), p. 563.[14] Conveniently, IP and SQP are included in Matlab®'s `fmincon` function, which comes with the Optimization toolbox. We run the SQP algorithm—which is usually faster than IP—and in case it does not find a solution, we switch to IP, an algorithm which we denote by *SQP/IP*.

Global Algorithms: IP and SQP are well adjusted to handle various degeneracy problems in order to find a local minimum for large-scale non-convex problems. There is now a large body of literature on global optimization strategies; see Horst and Pardalos (1995) and Romeijn and Pardalos (2013). Popular global optimization algorithms include adaptive stochastic search; branch and bound methods; homotopy methods; Genetic algorithms (GA); simulated annealing and two-phase algorithms such as *MultiStart* and *GlobalSearch*.[15] We focus on the two-phase algorithms MultiStart, GlobalSearch and on the genetic algorithm available in Matlab.[16]

[14] Furthermore, these algorithms exploit the existence of second-order derivatives which are well-defined in our problem.

[15] For a more detailed list and classification of global methods see p. 519 of Chapter 15 in Romeijn and Pardalos (2013). For a description of two-phase algorithms see Chapter 12 in Romeijn and Pardalos (2013).

[16] Genetic algorithms are a well developed field of computing and they have been used in many applications; see the introduction to Chapter 9 in Romeijn and Pardalos (2013). A very interesting application in economics that motivated our focus on GA is given in Qu and Tkachenko (2015).

6.3. *Implementing baseline projection in an example*

As an example, we consider the demand-supply SVAR model studied in Section 5 of Baumeister and Hamilton (2015) [henceforth, BH]. We fit a 6-lag VAR to U.S. data on growth rates of real labor compensation, Δw_t, and total employment, Δn_t, from 1970:Q1 to 2014:Q2.[17]

Using our notation, the demand-supply SVAR can be written as:

$$\begin{pmatrix}\Delta w_t \\ \Delta \eta_t\end{pmatrix} = A_1 \begin{pmatrix}\Delta w_{t-1} \\ \Delta \eta_{t-1}\end{pmatrix} + \ldots + A_6 \begin{pmatrix}\Delta w_{t-6} \\ \Delta \eta_{t-6}\end{pmatrix} + B \begin{pmatrix}\epsilon_t^d \\ \epsilon_t^s\end{pmatrix},$$

BH set-identify an expansionary demand and supply shock by means of the following sign restrictions:

$$B \equiv \begin{pmatrix} b_1 & b_3 \\ b_2 & b_4 \end{pmatrix} \quad \text{satisfies} \quad \begin{bmatrix} + & - \\ + & + \end{bmatrix}.$$

The sign restrictions state that a demand shock increases both real labor compensation and total employment, while a supply shock lowers wages but raises employment.

In this model, the short-run wage elasticity of labor supply (identified from a demand shock) is defined as:

$$\alpha \equiv b_2/b_1$$

Likewise, the short-run wage elasticity of labor demand (identified from a supply shock) is defined as:

$$\beta \equiv b_4/b_3$$

Finally, the long-run impact of a demand shock over employment is given by:

$$\gamma \equiv e_2'(\mathbb{I}_n - \sum_{p=1}^{6} A_p)^{-1} B e_1.$$

BH impose three additional restrictions. The first two of them are elasticity bounds motivated by the findings of different empirical studies. Hamermesh (1996), Akerlof and Dickens (2007), Lichter, Peichl, and Siegloch (2014) provide bounds on

[17] Our selection is based on the fact that 6 is the smallest number of lags such that CS(68%; μ) does not contain unstable VAR coefficients and non-invertible reduced-form covariance matrices. 68% confidence sets correspond to a single standard deviation and are frequently used in applied macroeconomic research. The Bayes Information Criteria and the Information Criteria both select less than six lags.

the wage elasticity of labor demand. Chetty, Guren, Manoli, and Weber (2011), Reichling and Whalen (2012) provide bounds on the wage elasticity of labor supply. The third and final restriction arises from imposing lower and upper bounds on the long-run impact of a demand shock on employment.

BH incorporate the restrictions in the form of priors on the structural parameters, but we treat the constraints as additional sign restrictions. Let t_v denote the standard t distribution with v degrees of freedom. The following table summarizes the way in which BH incorporate prior information:

TABLE II
Additional Identifying Restrictions

Restrictions	**Motivation**	**BH**	**This paper**
Bounds on α	Empirical studies report $\alpha \in [.27, 2]$	$\alpha \sim \max\{.6 + .6t_3, 0\}$	$.27 \leq \alpha \leq 2$
Bounds on β	Empirical studies $\beta \in [-2.5, -.15]$	$\beta \sim \min\{-.6 + .6t_3, 0\}$	$-2.5 \leq \beta \leq -.15$
Bounds on γ	$\gamma = 0$ is too strong	$\gamma \sim \mathcal{N}(0, V)$	$-2V \leq \gamma \leq 2V$

Thus, summarizing, our version of the BH model has 10 sign restrictions:

$$
\begin{aligned}
\text{Demand and Supply Shocks:} \quad &: \quad b_1 \geq 0, b_2 \geq 0, -b_3 \geq 0, b_4 \geq 0,\\
\text{Elasticity Bounds} \quad &: \quad 2b_1 - b_2 \geq 0, b_2 - .27b_1 \geq 0,\\
& \qquad b_4 + .15b_3 \geq 0, -2.5b_3 - b_4 \geq 0,\\
\text{Long-Run} \quad &: \quad e_2'(\mathbb{I}_n - \sum_{p=1}^{6} A_p)^{-1} B e_1 + 2V \geq 0,\\
& - \quad e_2'(\mathbb{I}_n - \sum_{p=1}^{6} A_p)^{-1} B e_1 + 2V \geq 0,
\end{aligned}
$$

where the parameter V is allowed to take the values $\{.01, .1, 1\}$ as in p. 1992 of BH.

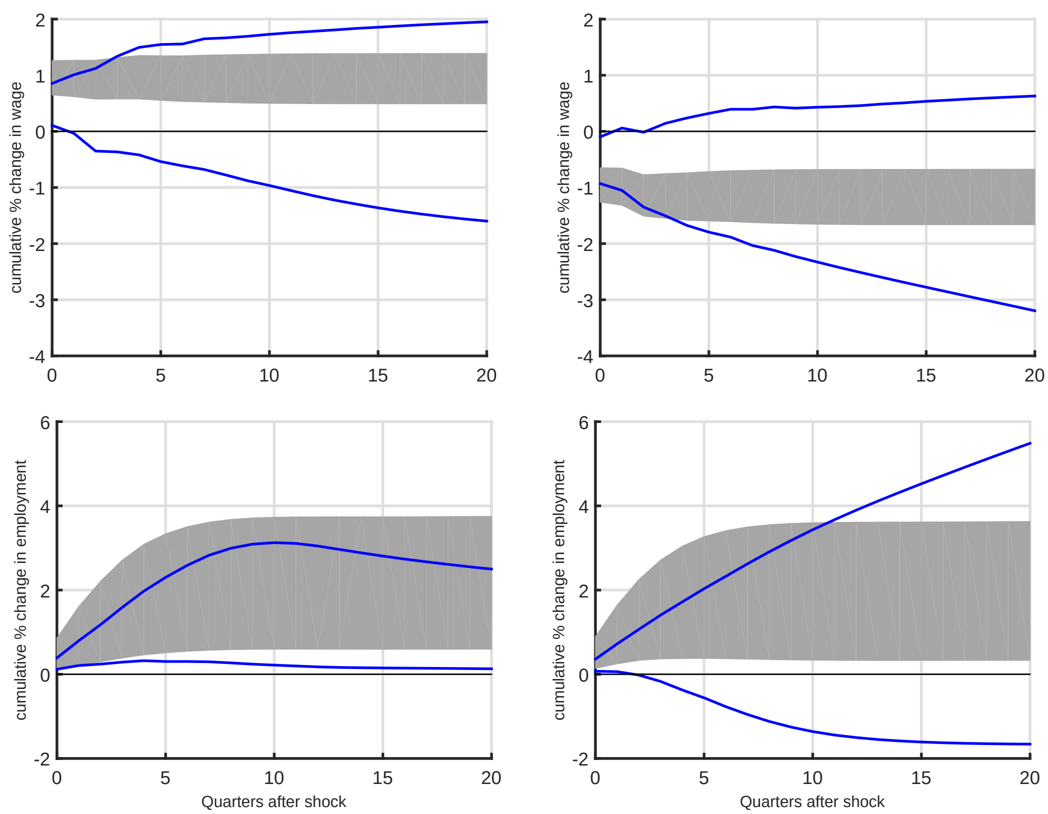
6.4. ***Results of the implementation of baseline projection***

Using our SQP/IP local solution algorithm, we compute the 68% projection confidence intervals for the cumulative response of wages and employment to the structural shocks in the model (20 consecutive quarters and setting $V = 1$). In addition to the projection region, we compute the 68% Bayesian credible set following the implementation in both Uhlig (2005) and BH.

Figure 1 shows the projection region as solid blue line and the standard Bayesian credible set (based on BH priors) as a grey-shaded area.

Figure 1: 68% Projection Region and 68% Credible Set.

(Baumeister and Hamilton (2015) priors)



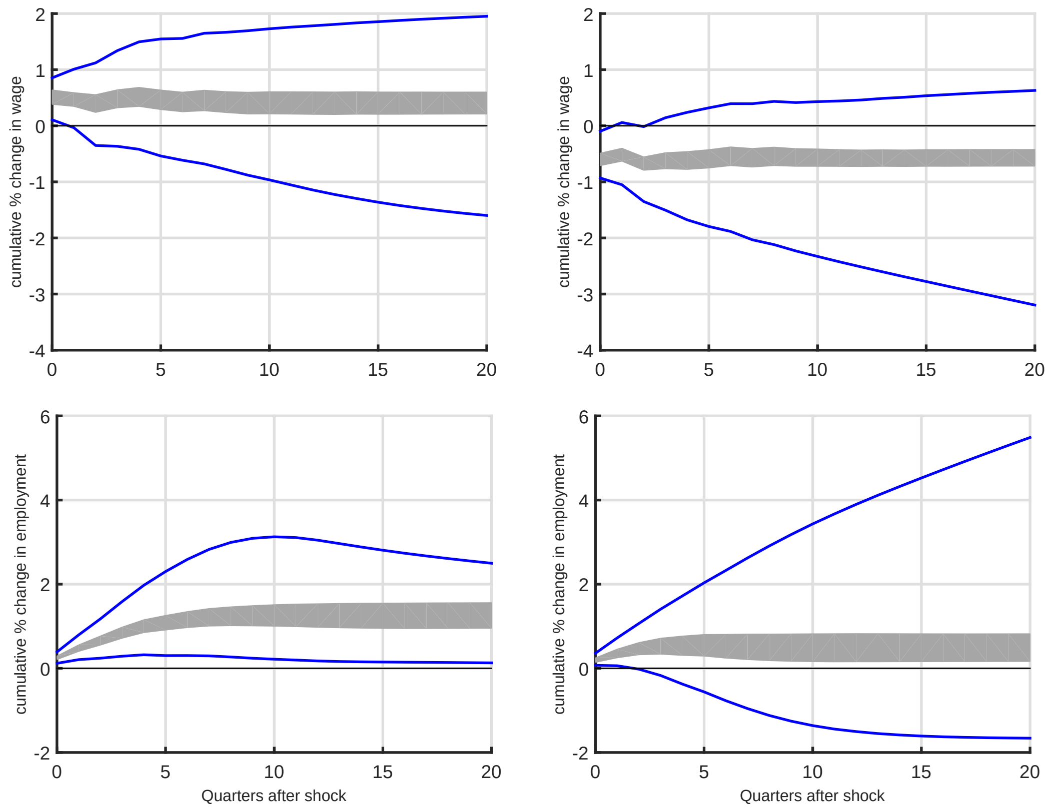
(a) Expansionary Demand Shock (b) Expansionary Supply Shock

(Solid, Blue Line) 68% Projection Region; (Shaded, Gray Area) 68% Bayesian Credible Set based on the priors in Baumeister and Hamilton (2015).

Figure 2 shows the boundaries of the projection region as solid blue line and the Bayesian credible set based on Uhlig (2005)'s priors as a grey-shaded area.

Figure 2: 68% Projection Region and 68% Credible Set.

(Uhlig (2005) priors)



(a) Expansionary Demand Shock (b) Expansionary Supply Shock

(Solid, Blue Line) 68% Projection region; (Shaded, Gray Area) 68% Bayesian Credible Set based on the Nornal-Wishart-Haar priors suggested in Uhlig (2005) and the inequality constraints summarized below Table II. The credible set is implemented following Arias et al. (2014).

Comment about Credible Sets: The 68% credible sets differ substantially depending on the specification of prior beliefs. Such sensitivity is the main motivation for our projection approach. In this example, the length of the credible sets for the cumulative response of employment seems to differ by a factor of at least two. The projection region seems quite large compared to the credible sets. This could be a consequence of either the robustness of projection or its conservativeness. To disentangle these effects, we calibrate projection to guarantee that it has exact robust Bayesian credibility in the next subsection.

Concrete comments regarding computational feasibility: Table III compares computing time for the projection (which has both a frequentist and a Robust Bayes interpretation) and the standard Bayesian methods.[18] Since the global methods are initialized at the local solution, these procedures take as least as much time as SQP/IP. Among the three global methods considered, the Genetic Algorithm takes the longest. Brute-force grid search (which refers to grid search on $\text{CS}_T(1-\alpha,\mu)$ to optimize $\underline{v}_{k,i,j}(\mu)$ and $\overline{v}_{k,i,j}(\mu)$) with only 1,000 draws from $\mu \in \mathbb{R}^{27}$ takes about 6 times longer than the baseline SQP/IP and generates substantially smaller bounds (see Appendix D.2).[19]

TABLE III
Computational time in seconds

Algorithm	Details	Time
SQP/IP		734
SQP/IP + MultiStart	100 initial points	33,314
SQP/IP + GlobalSearch	100 trial points (20 in Stage 1)	1,359
Genetic Algorithm	population of 100, 500 generations	76,863
Grid Search on $\text{CS}_T(1-\alpha,\mu)$	1,000 draws from μ	4,548
Bayesian, BH	1,000,000 Metropolis-Hastings draws	3,992
Bayesian, Uhlig	100,000 accepted posterior draws	2,338

Notes: Laptop @2.4GHz IntelCore i7.

Comments Regarding Local and Global algorithms: Figure 6 in Appendix D.1 compares the bounds of the projection confidence interval for the first four algorithms listed in Table III. For this application, it seems that none of the global algorithms improve on the local solution obtained from SQP/IP.[20]

[18]To get a fair sense of the computational cost, none of the global algorithms were parallelized.

[19]Instead of pseudo-random draws from the multi-variate normal distribution, we use quasi-random Sobol sequences, which have the property of being a low-discrepancy sequence in the hypercube. We translate the sequence into multivariate-normal draws using Cholesky decomposition. In our experience, this improves the performance of grid search substantially for a given number of grid points.

[20]In our Matlab code to implement projection we take SQP/IP as the default algorithm to

6.5. ***Implementing Calibrated projection in our example***

The key restriction used to set-identify an expansionary demand shock in the illustrative example is that it must increase wages and employment, upon impact. According to the credible sets in Figures 1 and 2, the expansionary shock has—in fact—noncontemporaneous effects over these two variables (every quarter over a 5 year horizon). Our calibrated projection confirms that there are medium-run effects of demand shocks over employment, but suggests that the non-zero effects over wages beyond the first two quarters could be an artifact of prior beliefs.

A similar observation is true for supply shocks. Our calibrated projection suggests that the decrease in wages five years after an expansionary supply shock is robust to the choice of prior on the set-identified parameters. The medium-run effects of supply shocks over employment lack this robustness.

Implementation of our Calibrated Projection: We close this subsection providing further details about the computational demands of our calibration exercise.

Instead of working with a specific posterior for μ, we calibrated projection relying on the large-sample approximation $\mu|Y_1, \ldots, Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T/T)$. Taking draws from this model is straightforward and does not require any special sampling technique (as a Monte-Carlo Markov Chain). Figure 3 used M=100,000 draws.

As described in our calibration algorithm, for each of the draws of μ (denoted μ^*_m), and for each horizon $k \in \{0, 1, 2, \ldots 20\}$, variable $i \in \{\text{wage}, \text{employment}\}$ and shock $j \in \{\text{demand shock}, \text{supply shock}\}$ we solved two mathematical programs to generate:

$$[\underline{v}_{k,i,j}(\mu^*_m), \overline{v}_{k,i,j}(\mu^*_m)].$$

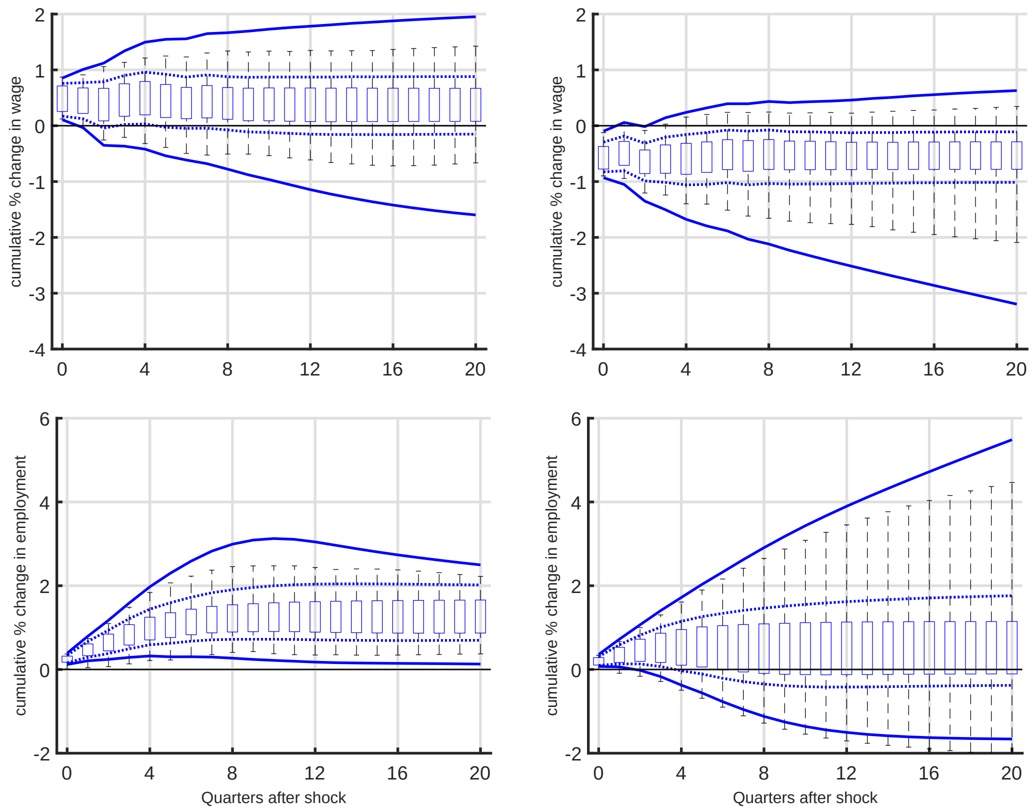
Computing the bounds of the identified set for all the combinations (k, i, j) given μ^*_m took approximately 9 seconds. Generating the boxes and the black dashed lines in Figure 3 took approximately 5 hours using 50 parallel Matlab 'workers' on a computer cluster at the University of Bonn.[21] Notice that we choose M=100,000 for illustrative purposes and the calibration results are barely different for M=1,000, which takes 3 minutes using the same computer cluster (or 2.5 hours not using parallelization at all).

After generating the bounds of the identified set, the calibration exercise adjusts

construct the projection region.

[21] Calibrating projection to guarantee frequentist coverage at one point in the parameter space took us 76 hours using the 50 parallel Matlab workers in the same computer cluster.

Figure 3: 68% Projection Region and 68% Calibrated Projection.



(a) Expansionary Demand Shock (b) Expansionary Supply Shock

(Solid Line) 68% Projection region; (Dotted Line) 68% Projection region *calibrated* to guarantee 68% robust Bayesian credibility of the IRF functions jointly (100,000 draws from the Gaussian approximation to the posterior of μ); (Box) 68% Projection region *calibrated* horizon by horizon and shock by shock; (Black Dashed Line) Support of the bounds of the identified set given the 100,000 posterior draws.

the nominal level of projection to simulatenously contain 68% of the draws from the bounds of the identified set for each combination (k, i, j).[22] The calibrated confidence level for the Wald ellipsoid is $1.85 \cdot 10^{-4}\%$ instead of the original 68%. This means that instead of projecting a Wald ellipsoid with radius $\chi^2_{68\%,27}$ we are using a $\chi^2_{68\%,4.5}$.

[22]To do this, we ran the baseline projection SQP/IP algorithm for different nominal confidence levels. An efficient calibration algorithm that requires only few iterations over the nominal level is the combination of bisection with secant and interpolation as provided by Matlab's `fzero` function. For reasonably low tolerance of $\eta = 0.001$, we need 15 iteration steps. With each step taking about 734 seconds, see Table III, steps 3 through 5 take about 1 hour.

7. CONCLUSION

A practical concern regarding standard Bayesian inference for set-identified Structural Vector Autoregressions is the fact that prior beliefs continue to influence posterior inference even when the sample size is infinite. Motivated by this observation, this paper studied the properties of projection inference for set-identified SVARs.

A nominal $1-\alpha$ projection region collects all the structural parameters of interest that are compatible with the VAR reduced-form parameters in a nominal $1-\alpha$ Wald ellipsoid. By construction, projection inference does not rely on the specification of prior beliefs for set-identified parameters.

We argued that the projection approach is general, computationally feasible, and—under mild assumptions concerning the asymptotic behavior of estimators and posterior distributions for the reduced-form parameters—produces regions with frequentist coverage and asymptotic *robust* Bayesian credibility of at least $1-\alpha$.

The main drawback of our projection region is that it is conservative. For a frequentist, the large-sample coverage is strictly above the desired confidence level. For a robust Bayesian, the asymptotic robust credibility of the nominal $1-\alpha$ projection region is strictly above $1-\alpha$.

We used the calibration idea described in Kaido et al. (2016) to eliminate the excess of robust Bayesian credibility. The calibration procedure consists of drawing the reduced-form parameters, μ, from its posterior distribution (or a suitable large-sample Gaussian approximation); evaluating the functions $\underline{v}(\mu), \overline{v}(\mu)$ for each draw of μ; and, finally, decreasing the nominal level of the projection region until it contains exactly $(1-\alpha)\%$ of the values of $\underline{v}(\mu), \overline{v}(\mu)$. The calibration exercise required more work than the baseline projection, but it is computationally feasible (and easily parallelizable). Moreover, if the bounds of the identified set are differentiable, our calibrated projection covers the identified set with probability $1-\alpha$.

We implemented our projection confidence set in the demand/supply SVAR for the U.S. labor market. The main set-identifying assumptions were sign restrictions on contemporaneous responses. Standard Bayesian credible sets suggested that the medium-run response of wages and employment to structural shocks behave in the same way as the contemporaneous responses. Our projection region (baseline and calibrated) showed that only the qualitative effects of demand shocks over employment and the qualitative effects of supply shocks over wages are robust to the choice of prior. Our projection approach is a natural complement for the Bayesian credible sets that are commonly reported in applied macroeconomic work.

REFERENCES

Akerlof, G. A. and W. T. Dickens (2007): "Unfinished Business in the Macroeconomics of Low Inflation: A Tribute to George and Bill by Bill and George," *Brookings Papers on Economic Activity*, 31–46.

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

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

Baumeister, C. and J. Hamilton (2015): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," *Econometrica*, 5, 1963–1999.

Berger, J. (1985): *Statistical decision theory and Bayesian analysis*, Springer.

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

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

Chen, L. H. and X. Fang (2015): "Multivariate normal approximation by Stein's method: The concentration inequality approach," *arXiv preprint arXiv:1111.4073.*

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *The American Economic Review, Papers and Proceedings*, 101, 471–475.

Christiano, L. J., M. Eichenbaum, and C. Evans (1996): "The effects of monetary policy shocks: some evidence from the flow of funds," *The Review of Economics and Statistics*, 78, 16–34.

Connault, B. (2016): "A weakly-dependent Bernstein-von Mises Theorem," *Working Paper, University of Pennsylvania.*

DasGupta, A. (2008): *Asymptotic theory of statistics and probability*, Springer Verlag.

Dufour, J.-M. (1990): "Exact tests and confidence sets in linear regressions with autocorrelated errors," *Econometrica: Journal of the Econometric Society*, 475–494.

Dufour, J.-M. and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

——— (2007): "Further results on projection-based inference in IV regressions with weak, collinear or missing instruments," *Journal of Econometrics*, 139, 133–153.

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

Gafarov, B., M. Meier, and J. L. Montiel Olea (2015): "Delta-Method inference for a class of set-identified SVARs," *Working paper, New York University*.

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*.

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.

Hamermesh, D. S. (1996): *Labor demand*, Princeton University Press.

Horst, R. and P. M. Pardalos (1995): *Handbook of global optimization, Nonconvex Optimization and its Applications*, Springer Science & Business Media.

Ibragimov, I. d. A. and R. Z. Has' minskii (2013): *Statistical estimation: asymptotic theory*, vol. 16, Springer Science & Business Media.

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

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

——— (2016): "Joint confidence sets for structural impulse responses," *Journal of Econometrics*, 192, 421–432.

Jordà, Ò. (2009): "Simultaneous confidence regions for impulse responses," *The Review of Economics and Statistics*, 91, 629–647.

Kaido, H., F. Molinari, and J. Stoye (2016): "Inference for Projections of Identified Sets," *Working Paper, Boston University*.

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. (2012): "Estimation and inference for set-identified models using lower posterior probability," *Working Paper, University College London*.

Kline, B. and E. Tamer (2015): "Bayesian Inference in a class of Partially Identified Models," *Working paper, Harvard University*.

Lichter, A., A. Peichl, and S. Siegloch (2014): "The own-wage elasticity of labor demand: A meta-regression analysis," *ZEW-Centre for European Economic Research Discussion Paper*.

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.

Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2015): "Confidence Bands for Impulse Responses: Bonferroni vs. Wald," *Oxford Bulletin of Economics and Statistics*, 77, 800–821.

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

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

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

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

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

Qu, Z. and D. Tkachenko (2015): "Global Identification in DSGE Models Allowing for Indeterminacy," *Working Paper, Boston University*.

Reichling, F. and C. Whalen (2012): "Review of estimates of the frisch elasticity of labor supply," *Working Paper, Congressional Budget Office*.

Romeijn, H. E. and P. M. Pardalos (2013): *Handbook of global optimization*, vol. 2, Springer Science & Business Media.

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*.

Scheffé, H. (1953): "A method for judging all contrasts in the analysis of variance*," *Biometrika*, 40, 87–110.

Scholl, A. and H. Uhlig (2008): "New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates," *Journal of International Economics*, 76, 1–13.

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

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

Uhlig, H. (1994): "What macroeconomists should know about unit roots: a Bayesian perspective," *Econometric Theory*, 10, 645–671.

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

Appendix A

APPENDIX A: PROOF OF MAIN RESULTS

A.1. *Verification of Assumption 2 for the Gaussian SVAR with a Normal-Wishart Prior.*

Consider the SVAR in (3.1) and assume that $F \sim \mathcal{N}(0, \mathbb{I}_n)$. Let P^* denote a prior on the SVAR parameters (A, B).

Note first that Assumption 2 depends only on the distribution that P^* induces over the reduced-form parameters, μ. Thus, we abuse notation and refer to P^* as the prior distribution on (A, Σ).

The analysis in this section focuses on the *Normal-Wishart* prior P^* used in Gaussian SVAR analysis. We establish an almost sure version of Assumption 2.

Prior for μ: Consider the hyper-parameters:

$$\bar{A}_0 \in \mathbb{R}^{n \times np},\ S_0 \in \mathbb{R}^{n \times n},\ N_0 \in \mathbb{R}^{np \times np},\ v_0 \in \mathbb{R}.$$

Definition The Normal-Wishart Prior P^* over the parameters $(\text{vec}(A), \text{vech}(\Sigma))$—defined by hyper parameters $(\bar{A}_0, S_0, N_0, v_0)$—is given by:

$$\text{vec}(A)|\Sigma \sim \mathcal{N}\Big(\text{vec}(\bar{A}_0)\ ,\ N_0^{-1} \otimes \Sigma\Big),$$

and

$$\Sigma^{-1} \sim \text{Wishart}_n\Big(S_0^{-1}/v_0\ ,\ v_0\Big).$$

Posterior in the Gaussian SVAR: Let

$$Q_T \equiv \frac{1}{T}\sum_{t=1}^{T} X_t X_t',$$

and define the updated hyperparameters:

$$\begin{aligned}
\bar{A}_T &= \widehat{A}_T Q_T \Big(\frac{N_0}{T} + Q_T\Big)^{-1} + \bar{A}_0 \frac{N_0}{T}\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \\
S_T &= \frac{v_0}{T + v_0} S_0 + \frac{T}{T + v_0}\widehat{\Sigma}_T + \frac{1}{T + v_0}\Big(\bar{A}_T - \bar{A}_0\Big) N_0 \Big(\frac{N_0}{T} + Q_T\Big)^{-1} Q_T \Big(\bar{A}_T - \bar{A}_0\Big)'
\end{aligned}$$

where $\widehat{A}_T$ and $\widehat{\Sigma}_T$ are the ordinary least squares estimators for A and Σ defined in Section 3.1.

From p. 410 in Uhlig (1994) and p. 410 in Uhlig (2005) the posterior distribution for the vector $(\text{vec}(A)', \text{vech}(\Sigma)')'$ can be written as:

$$\text{vec}(A)|Y_1, \ldots, Y_T = \text{vec}(\bar{A}_T) + \Big[\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \otimes \frac{\Sigma}{T}\Big]^{1/2} W, \quad W \sim \mathcal{N}_{n^2 p}(\mathbf{0}, \mathbb{I}_{n^2 p}),$$

$$\Sigma|Y_1, \ldots, Y_T = S_T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1} S_T^{1/2}, \quad Z_t \sim \mathcal{N}_n(\mathbf{0}\ ,\ \mathbb{I}_n), \text{i.i.d},$$

where both random vectors are independent of the data and $\{Z_t\}_{t=1}^T$ independent of W. Note that for a given data realization, the posterior distribution of (A, Σ) is a measurable function of

$\mathcal{W} \equiv (W, Z_1, \ldots Z_T)$. We use the term $o_{\mathcal{W}}(1)$ to denote any sequence that converges to zero as $T \to \infty$ for almost every realization of $\mathcal{W}$.

Asymptotic Behavior of the posterior for μ: We now show that all of the Normal-Wishart priors in the Gaussian model satisfy our Assumption 2. Note first that for almost every data realization $(Y_1, \ldots, Y_T)$ and almost every realization of the random vector Z_t we have that

$$\Sigma - \widehat{\Sigma}_T \to 0,$$

by applying the strong of large numbers to $(1/T)\sum_{t=1}^{T} Z_t Z_t'$. Consequently:

$$\begin{aligned}
\sqrt{T}(\text{vec}(A) - \text{vec}(\widehat{A}_T)) &= \widehat{A}_T\sqrt{T}\Big(Q_T\Big(\frac{N_0}{T} + Q_T\Big)^{-1} - \mathbb{I}_{n^2p}\Big) + \bar{A}_0\frac{N_0}{\sqrt{T}}\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \\
&+ \Big[\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \otimes \widehat{\Sigma}_T\Big]^{1/2} W + o_{P^*|Y_1,\ldots Y_T}(1), \\
&= \widehat{A}_T\sqrt{T}\Big(Q_T\Big(Q_T^{-1} + Q_T^{-1}\frac{N_0}{T}Q_T^{-1} + O(1/T^2)\Big) - \mathbb{I}_{n^2p}\Big) \\
&\quad \text{(by a first-order Taylor expansion)} \\
&+ \bar{A}_0\frac{N_0}{\sqrt{T}}\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \\
&+ \Big[\Big(\frac{N_0}{T} + Q_T\Big)^{-1} \otimes \widehat{\Sigma}_T\Big]^{1/2} W + o_{\mathcal{W}}(1), \\
&= \Big[Q_T^{-1} \otimes \widehat{\Sigma}_T\Big]^{1/2} W + o_{\mathcal{W}}(1).
\end{aligned}$$

This implies that the posterior distribution of $\sqrt{T}(\text{vec}(A) - \text{vec}(\widehat{A}_T))$ converges in distribution, for almost every data realization $(Y_1, \ldots, Y_T)$, to the random vector:

$$[Q_T^{-1/2} \otimes \widehat{\Sigma}_T^{1/2}]W, \text{ where } W \sim \mathcal{N}_{n^2p}(\mathbf{0}, \mathbb{I}_{n^2p}). \tag{A.1}$$

Note now that

$$\begin{aligned}
\sqrt{T}(\text{vech}(\Sigma) - \text{vech}(\widehat{\Sigma}_T)) &= \sqrt{T}\text{vech}\Big(S_T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1} S_T^{1/2} - \widehat{\Sigma}_T\Big), \\
&= \sqrt{T}\text{vech}\Big(\widehat{\Sigma}_T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1}\widehat{\Sigma}_T^{1/2} + O(1/T) - \widehat{\Sigma}_T\Big), \\
&= \sqrt{T}\text{vech}\Big(\widehat{\Sigma}_T^{1/2}\Big[\Big(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'\Big)^{-1} - \mathbb{I}_n\Big]\widehat{\Sigma}_T^{1/2}\Big) + o(1).
\end{aligned}$$

This implies that the posterior distribution of $\sqrt{T}(\text{vech}(\Sigma) - \text{vech}(\widehat{\Sigma}_T))$ converges in distribution, for almost every data realization $(Y_1, \ldots, Y_T)$, to the random vector:

$$\Big(2D^+(\widehat{\Sigma}_T \otimes \widehat{\Sigma}_T)D^+\Big)^{1/2} Z, \text{ where } Z \sim \mathcal{N}_{n(n+1)/2}(\mathbf{0}, \mathbb{I}_{n(n+1)/2}),\ Z \perp W, \tag{A.2}$$

and $D^{+} \equiv (D'D)^{-1}D'$ is the Moore-Penrose inverse of the duplication matrix D such that $\text{vec}(\Sigma) = D\text{vech}(\Sigma)$.

Now, assume that the confidence set for the reduced-form parameters is constructed using the Gaussian Maximum Likelihood asymptotic variance of $\widehat{\mu}_T$ as in p.93 of Lütkepohl (2007); that is:

$$\widehat{\Omega}_T \equiv \begin{pmatrix} Q_T^{-1} \otimes \widehat{\Sigma}_T & \mathbf{0}_{n^2p\times(n(n+1)/2)} \\ \mathbf{0}_{(n(n+1)/2)\times n^2p} & 2D^{+}(\widehat{\Sigma}_T \otimes \widehat{\Sigma}_T)D^{+\prime} \end{pmatrix}. \tag{A.3}$$

Let G denote the joint distribution of (W, Z), which is a standard multivariate normal independently of the data. Then, combining (A.1), (A.2), (A.3)

$$\begin{aligned} P^{*}\Big(\mu \in \text{CS}_T(1-\alpha,\mu)|(Y_1,\ldots,Y_T)\Big) &= P^{*}\Big(\sqrt{T}(\mu-\widehat{\mu}_T)'\widehat{\Omega}_T^{-1}\sqrt{T}(\mu-\widehat{\mu}_T) \leq \chi^2_{d,1-\alpha}|(Y_1,\ldots Y_T)\Big) \\ &\to G\Big(\begin{pmatrix} W \\ Z \end{pmatrix}'\begin{pmatrix} W \\ Z \end{pmatrix} \leq \chi^2_{d,1-\alpha} \mid Y_1,\ldots,Y_T\Big) \text{for a.e. data realization} \\ &= G\Big(\begin{pmatrix} W \\ Z \end{pmatrix}'\begin{pmatrix} W \\ Z \end{pmatrix} \leq \chi^2_{d,1-\alpha}\Big) \\ &= (1-\alpha). \end{aligned}$$

A.2. *Proof of Result 3 (Finite-Sample Calibration for a Robust Bayesian)*

PROOF: The proof of Result 2 has already established that for any data realization:

$$\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big).$$

is at least as large as:

$$P^*_\mu\left(\times_{h=1}^{H}[\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq \mathrm{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T),\lambda^H)|Y_1,\ldots,Y_T\right).$$

Hence, it is sufficient to show that for any data realization:

$$\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big) \leq 1-\alpha.$$

In order to establish this upper bound for each data realization, we will find a prior on Q (conditional on μ) that gives credibility of exactly $1-\alpha$ to the calibrated projection region. Fix the data, and denote the set $\mathrm{CS}_T(1-\alpha(Y_1,\ldots,Y_T);\lambda^H)$ simply by $\mathcal{C}(Y^T)$. Before the realization of the data, the set $\mathcal{C}(Y^T)$ is just some subset of $\mathbb{R}^H$, so the prior can depend on this set. Let $\overline{v}_h(\mu)$ abbreviate $\overline{v}_{k_h,i_h,j_h}(\mu)$ and define $\underline{v}_h(\mu)$ analogously. Let $Q_{\max}(\mu;h)$ denote the rotation matrix for which the structural parameter achieves its upper bound; i.e., $\lambda(\mu, Q_{\max}(\mu;h)) = \overline{v}_h(\mu)$ (the matrix $Q_{\min}$ is defined analogously).

For each μ such that $\times_{h=1}^{H}[\underline{v}_h(\mu),\overline{v}_h(\mu)] \notin \mathcal{C}(Y^T)$, let $\overline{h}(\mu)$ denote the smallest horizon for which $\overline{v}_{\overline{h}(\mu)}(\mu)$ is not contained in the $h(\mu)$-th coordinate of the region $\mathcal{C}(Y^T)$. If no upperbound falls outside $\mathcal{C}(Y^T)$ set $\overline{h}(\mu)=0$. Define $\underline{h}(\mu)$ analogously. Consider the following prior for $Q|\mu$ that depends on the set $\mathcal{C}_T(Y^T)$:

$$Q|\mu = \begin{cases} Q_{\max}(\mu;1) & \text{if} \quad \times_{h=1}^{H}[\underline{v}_h(\mu),\overline{v}_h(\mu)] \subseteq \mathcal{C}_T(Y^T), \\ Q_{\max}(\mu,\overline{h}(\mu)) & \text{if} \quad \times_{h=1}^{H}[\underline{v}_h(\mu),\overline{v}_h(\mu)] \nsubseteq \mathcal{C}(Y^T) \text{ and } \overline{h}(\mu) \geq \underline{h}(\mu), \\ Q_{\min}(\mu,\underline{h}(\mu)) & \text{if} \quad \times_{h=1}^{H}[\underline{v}_h(\mu),\overline{v}_h(\mu)] \nsubseteq \mathcal{C}(Y^T) \text{ and } \overline{h}(\mu) < \underline{h}(\mu), \end{cases}$$

Finally, let P^{**} denote the prior induced by P^*_μ and $Q|\mu$ as defined above. Note that for each data realization $(Y_1,\ldots,Y_T)$:

$$\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha(Y_1,\ldots Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big)$$

is—by definition of infimum—smaller than or equal

$$P^{**}\Big(\lambda^H(\mu,Q) \in CS_T(1-\alpha(Y_1,\ldots,Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big).$$

By construction, the prior for $Q|\mu$ is such that $\lambda^H(\mu,Q) \in CS_T(1-\alpha(Y_1,\ldots,Y_T);\lambda^H)$ if and only if $\times_{h=1}^{H}[\underline{v}_h(\mu),\overline{v}_h(\mu)] \subseteq \mathcal{C}_T(Y^T)$. To see this, note that whenever the bounds of the identified set $\times_{h=1}^{H}[\underline{v}_h(\mu),\overline{v}_h(\mu)] \nsubseteq \mathcal{C}_T(Y^T)$, either $\overline{h}(\mu) \neq 0$ or $\underline{h}(\mu) \neq 0$ implying that the structural parameter $\lambda_h(\mu,Q)$ takes the value of $\overline{v}_{\overline{h}(\mu)}(\mu)$ or $\underline{v}_{\overline{h}(\mu)}(\mu)$ (whichever horizon is largest). Since these bounds are not contained in $\mathcal{C}_T(Y^T)$:

$$P^{**}\Big(\lambda^H(\mu,Q) \in CS_T(1-\alpha(Y_1,\ldots,Y_T);\lambda^H) \,\Big|\, Y_1,\ldots,Y_T\Big).$$

equals

$$P^*_\mu \left(\times_{h=1}^{H} [\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \in \text{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T),\lambda^H) | Y_1,\ldots,Y_T \right) = 1-\alpha.$$

This means that:

$$1-\alpha \leq \inf_{P^* \in \mathcal{P}(P^*_\mu)} P^* \left(\lambda^H(A,B) \in CS_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda^H) \;\middle|\; Y_1,\ldots,Y_T \right) \leq 1-\alpha.$$

Q.E.D.

A.3. *Proof of Result 4 (Robust Bayesian Calibration and Frequentist Coverage)*

Let r_T^* be the radius that calibrates robust Bayesian credibility; i.e., the radius corresponding to the calibrated nominal level $1-\alpha^*(Y_1,\ldots,Y_T)$. In a slight abuse of notation we replace $\text{CS}_T(1-\alpha^*(Y_1,\ldots,Y_T);\lambda_h)$ by $\text{CS}_T(r_T^*;\lambda_h)$. The proof of this result is based on Lemma 1,2, and 3 in Appendix C.

PROOF: We show that for every $\epsilon > 0$ there is $T(\epsilon)$ such that $T > T(\epsilon)$ implies that:

$$1-\alpha-\epsilon \leq P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \text{CS}_T(r_T^*;\lambda_h),\ \forall h = 1,\ldots H\right) \leq 1-\alpha+\epsilon.$$

Let $M_1(\epsilon)$ and $M_2(\epsilon)$ be two real-valued (nonnegative) functions. Define the events:

$$\begin{aligned}
A_1(\epsilon) =\quad & \{(Y_1,\ldots,Y_T) \quad | \ ||\sqrt{T}(\widehat{\mu}_T-\mu_0)|| \leq M_1(\epsilon)\},\\
A_2(\epsilon) =\quad & \{(Y_1,\ldots Y_T) \quad | \text{ the largest eigenvalue of } \widehat{\Omega}_T \text{ is smaller than } M_2(\epsilon)\},\\
A_3(\eta) =\quad & \{(Y_1,\ldots Y_T) \quad | \ |\widehat{\overline{\sigma}}_h(\mu_0)-\overline{\sigma}_h(\mu_0)| < \eta/(2\chi^2_{d,1-\alpha}) \text{ and } |\widehat{\underline{\sigma}}_h(\mu_0)-\underline{\sigma}_h(\mu_0)| < \eta/(2\chi^2_{d,1-\alpha})\},\\
A_4(\epsilon) =\quad & \{(Y_1,\ldots Y_T) \quad | \sup_{B\in\mathcal{B}(\mathbb{R}^d)} \left|P^*\left(\sqrt{T}(\mu^*-\widehat{\mu}_T)\in B \mid Y_1,\ldots,Y_T\right)-\mathbb{P}\left(Z\in B\right)\right| \leq \epsilon/8,\\
& \qquad \text{where } Z\sim\mathcal{N}_d(\mathbf{0},\Omega),\}.\\
A_5(\epsilon) =\quad & \{(Y_1,\ldots Y_T) \quad | \ r_T^* < \chi^2_{d,1-\alpha}+\epsilon\}.
\end{aligned}$$

where the standard errors $\widehat{\overline{\sigma}}_h(\mu_0)$, $\overline{\sigma}_h(\mu_0)$, $\widehat{\underline{\sigma}}_h(\mu_0), \underline{\sigma}_h(\mu_0)$ are defined in Lemma 1 of Appendix C.1. We first show that the probability of these events can be made arbitrarily close to 1 for a large enough sample size.

To see this, note that Assumption 1 of Result 4 (convergence in distribution of $\sqrt{T}(\widehat{\mu}_T-\mu_0)$) implies there exists a function $M_1(\epsilon)$ and a large enough sample size $T_1(\epsilon)$ such that for $T \geq T_1(\epsilon)$, $P_{\mu_0}(A_1(\epsilon)) > 1-\epsilon/25$. Assumption 2 of Result 4 (convergence in probability of $\widehat{\Omega}_T$) implies there exists a function $M_2(\epsilon)$ and a large enough sample size $T_2(\epsilon,\eta)$ such that for $T \geq T_2(\epsilon,\eta)$, $P_{\mu_0}(A_2(\epsilon)) > 1-\epsilon/25$ and $P_{\mu_0}(A_3(\eta)) > 1-\epsilon/25$. Assumption 3 of Result 4 (Bernstein von-Mises Theorem in total variation) implies that there is $T_4(\epsilon)$ such that $T \geq T_4(\epsilon,\eta)$, $P_{\mu_0}(A_4(\epsilon)) > 1-\epsilon/25$. Finally, since Assumption 3 of Result 4 implies the assumption of Result 2 (the baseline projection has robust credibility of at least $1-\alpha$ with high probability) implies there is $T_5(\epsilon)$ such that $T \geq T_5(\epsilon)$, $P_{\mu_0}(A_5(\epsilon)) > 1-\epsilon/25$

This means that for $T \geq \max\{T_1(\epsilon), T_2(\epsilon,\eta), T_4(\epsilon), T_5(\epsilon)\}$,

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \text{CS}_T(r_T^*;\lambda_h),\ \forall h = 1,\ldots H\right)$$

$$\leq$$

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \text{CS}_T(r_T^*;\lambda_h),\ \forall h = 1,\ldots H\ \cap A(\epsilon,\eta)\right) + \epsilon/5, \tag{A.4}$$

where

$$A(\epsilon,\eta) \equiv A_1(\epsilon) \cup A_2(\epsilon) \cup A_3(\eta) \cup A_4(\epsilon) \cup A_5(\epsilon).$$

Define now, given $Z \sim \mathcal{N}(0,\Omega)$, the quantile $r_{1-\alpha+\epsilon}$ by the equation:

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)r_{1-\alpha+\epsilon} \leq \dot{\underline{v}}_h(\mu_0)'Z, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r_{1-\alpha+\epsilon}, \ \forall h = 1,\ldots H\right) = 1-\alpha+\epsilon.$$

Lemma 3 in Appendix C.3 has shown that there exists $T_6(\epsilon)$ such that :

$$(Y_1,\ldots,Y_T) \in A(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha})) \implies r^*_T \leq r_{1-\alpha+\epsilon/5} \equiv \overline{r},$$

where the function $\eta(\epsilon/10,\chi^2_{d,1-\alpha}))$ is defined as in Lemma 2 in Appendix C.2. This implies that whenever $T \geq \max\{T_1(\epsilon), T_2(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha})), T_4(\epsilon), T_5(\epsilon), T_6(\epsilon)\}$, Equation (A.4) is bounded above by:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \text{CS}_T(\overline{r};\lambda_h), \ \forall h = 1,\ldots H \ \cap A(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha}))\right) + \epsilon/5. \tag{A.5}$$

Lemma 1 in Appendix C.1 has shown that $(Y_1,\ldots,Y_T) \in A(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha}))$ implies that the projection region of radius $r_{1-\alpha+\epsilon/5}$ is contained in the delta-method interval $DM^h_T(r,\eta)$ (with radius $r = r_{1-\alpha+\epsilon/5}$ and expansion $\eta = \eta(\epsilon/10, \chi^2_{d,1-\alpha})$):

$$\left[\underline{v}_h(\widehat{\mu}_T) - \frac{(\overline{r}+\eta(\epsilon/10,\chi^2_{d,1-\alpha}))\underline{\sigma}_h(\mu_0)}{\sqrt{T}}\ , \ \overline{v}_h(\widehat{\mu}_T) + \frac{(\overline{r}+\eta(\epsilon/10,\chi^2_{d,1-\alpha}))\overline{\sigma}_h(\mu_0)}{\sqrt{T}}\right],$$

for every $h = 1,\ldots H$. This means that Equation (A.5) is bounded above by:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq DM^h_T(\overline{r}, \eta(\eta/10, \chi^2_{d,1-\alpha})), \ \forall h = 1,\ldots H\ \right) + \epsilon/5. \tag{A.6}$$

An application of the delta-method implies there is $T_7(\epsilon)$ larger than

$$T \geq \max\{T_1(\epsilon), T_2(\epsilon, \eta(\epsilon/10, \chi^2_{d,1-\alpha})), T_4(\epsilon), T_5(\epsilon), T_6(\epsilon)\}$$

such that Equation A.6 is bounded above by $2\epsilon/5$ plus:

(A.7)

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)(\overline{r}+\eta(\epsilon/10,\chi^2_{d,1-\alpha})) \leq \dot{\underline{v}}_h(\mu_0)'Z, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)(\overline{r}+\eta(\epsilon/10,\chi^2_{d,1-\alpha}), \ \forall h = 1,\ldots H\right),$$

which, by definition of $\eta(\epsilon, \chi^2_{d,1-\alpha})$, imploes that the latter equation is bounded above by

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)\overline{r} \leq \dot{\underline{v}}_h(\mu_0)'Z, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)\overline{r}, \ \forall h = 1,\ldots H\right) + 4\epsilon/5.$$

Using the definition of $\overline{r}$, we conclude that there is $T(\epsilon)$ such that for $T \geq T(\epsilon)$:

$$P_{\mu_0}\left([\underline{v}_h(\mu_0), \overline{v}_h(\mu_0)] \subseteq \text{CS}_T(r^*_T;\lambda_h), \ \forall h = 1,\ldots H\right) \leq 1-\alpha+\epsilon.$$

The lower bound is derived analgously.

Q.E.D.

A.4. *Asymptotic Calibration for a Robust Bayesian* $(\mu|Y_1,\dots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T/T))$

We now show that whenever $\alpha^*_T \equiv \alpha(Y_1,\dots,Y_T)$ is calibrated to guarantee that

$$\mathbb{P}_T\Big(\times_{h=1}^{H} [\underline{v}_{k_1,i_1,j_1}(\mu), \overline{v}_{k_1,i_1,j_1}(\mu)] \times \ldots \times [\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq \text{CS}_T(1-\alpha^*_T, \lambda^H) \,|\, Y_1,\dots Y_T \Big)$$

equals $1-\alpha$ whenever $\mu|Y_1,\dots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T/T)$, then one can guarantee asymptotic robust credibility of $1-\alpha$ for a large class of priors on μ. This is formalized below.

Let $f(Y_1,\dots Y_T\,|\,\mu_0)$ denote the Gaussian density for the VAR data and let $\Omega \in \mathbb{R}^{d\times d}$ denote the probability limit of $\widehat{\Omega}_T$. Let G_Ω denote a Gaussian measure centered at $\mathbf{0}_d$ with covariance matrix Ω. Let $\mathcal{B}(d)$ denote Borel sets in $\mathbb{R}^d$.

Result 5 *Let $Y_1,\dots Y_T \sim f(Y_1,\dots Y_T\,|\,\mu_0)$ and suppose that the prior P^*_μ is such that:*

$$\sup_{A\in\mathcal{B}(d)} \left| P^*_\mu(\sqrt{T}(\mu-\widehat{\mu}_T) \in A \,|\, Y_1,\dots Y_T) - G_\Omega(A) \right| = o_p(Y_1,\dots Y_T;\mu_0).$$

Then,

$$\inf_{P^*\in\mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in CS_T(1-\alpha^*_T,\lambda^H) \,|\, Y_1,\dots Y_T\Big) = 1-\alpha + o_p(Y_1,\dots Y_T;\mu_0).$$

Proof: Result 3 has shown that for any $\alpha(Y_1,\dots,Y_Y)$

$$\inf_{P^*\in\mathcal{P}(P^*_\mu)} P^*\Big(\lambda^H(A,B) \in \text{CS}_T(1-\alpha^*_T,\lambda^H) \,|\, Y_1,\dots Y_T\Big) = P^*_\mu\left(\mu \in A^*_T \,|\, Y_1,\dots Y_T\right),$$

where $A^*_T \subseteq \mathbb{R}^d$ is defined as:

$$\{\mu\in\mathbb{R}^d \mid \times_{h=1}^{H} [\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq \text{CS}_T(1-\alpha^*_T,\lambda^H)\}.$$

Note that

$$\begin{aligned} P^*_\mu\left(\mu\in A^*_T \,|\, Y_1,\dots Y_T\right) &= P^*_\mu\Big(\sqrt{T}(\mu-\widehat{\mu}_T) \in \sqrt{T}(A^*_T-\widehat{\mu}_T)\ :|\, Y_1,\dots Y_T\Big) - G_\Omega(\sqrt{T}(A^*_T-\widehat{\mu}_T)) \\ &+ G_\Omega(\sqrt{T}(A^*_T-\widehat{\mu}_T)) - G_{\widehat{\Omega}_T}(\sqrt{T}(A^*_T-\widehat{\mu}_T)) \\ &+ G_{\widehat{\Omega}_T}(\sqrt{T}(A^*_T-\widehat{\mu}_T)) \end{aligned}$$

We make three observations:

1. Note first that:

$$P^*_\mu\Big(\sqrt{T}(\mu-\widehat{\mu}_T) \in \sqrt{T}(A^*_T-\widehat{\mu}_T)\ :|\, Y_1,\dots Y_T\Big) - G_\Omega(\sqrt{T}(A^*_T-\widehat{\mu}_T))$$

is smaller than or equal

$$\sup_{A\in\mathcal{B}(d)} \left| P^*_\mu(\sqrt{T}(\mu-\widehat{\mu}_T) \in A \,|\, Y_1,\dots Y_T) - G_\Omega(A) \right|,$$

which is, by assumption, $o_p(Y_1,\dots Y_T;\mu_0)$.

2. Note then that

$$|G_{\widehat{\Omega}_T}(\sqrt{T}(A_T^* - \widehat{\mu}_T)) - G_{\Omega}(\sqrt{T}(A_T^* - \widehat{\mu}_T))| = o_p(Y_1, \dots, Y_T; \mu_0)$$

since $\widehat{\Omega}_T \xrightarrow{p} \Omega$ and G is the Gaussian measure centered at zero.

3. Finally, note that $G_{\widehat{\Omega}_T}(\sqrt{T}(A_T^* - \widehat{\mu}_T))$ is the same as is the same as

$$\mathbb{P}(N(\widehat{\mu}_T, \widehat{\Omega}_T/T) \in A_T^* | Y_1, \dots, Y_T),$$

which, by definition of A_T^*, is the same as:

$$\mathbb{P}_T\left(\times_{h=1}^{H}[\underline{v}_{k_1,i_1,j_1}(\mu), \overline{v}_{k_1,i_1,j_1}(\mu)] \times \dots \times [\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)] \subseteq \mathrm{CS}_T(1-\alpha_T^*, \lambda^H) | Y_1, \dots Y_T\right)$$

where $\mu | Y_1, \dots Y_T \sim \mathcal{N}_d(\widehat{\mu}_T, \widehat{\Omega}_T/T)$.

We conclude that:

$$|\inf_{P^* \in \mathcal{P}(P_\mu^*)} P^*\left(\lambda^H(A,B) \in \mathrm{CS}_T(1-\alpha_T^*, \lambda^H) \mid Y_1, \dots Y_T\right) - (1-\alpha)| \leq o_p(Y_1, \dots Y_T; \mu_0),$$

which implies the desired result. *Q.E.D.*

Appendix B and C

APPENDIX B: FREQUENTIST CALIBRATION OF PROJECTION

We have shown that projection can be calibrated to achieve exact robust Bayesian credibility for a given prior on the reduced-form parameters. We now discuss the extent to which projection can be calibrated to achieve large-sample frequentist coverage of $1-\alpha$.

Frequentist calibration requires either an exact or an approximate statistical model for the data. We assume that: $\widehat{\mu}_T \sim \mathbb{P}_\mu \equiv \mathcal{N}_d(\mu, \widehat{\Omega}_T/T)$, where μ belongs to some set $\mathcal{M} \subseteq \mathbb{R}^d$ and $\widehat{\Omega}_T$ is treated as a non-stochastic matrix.

Let λ be some structural coefficient of interest. The frequentist calibration exercise consists in finding a radius, $r_T(\alpha)$, for the Wald ellipsoid such that:

$$\inf_{\mu \in \mathcal{M}} \inf_{\lambda \in \mathcal{I}^{\mathcal{R}}(\mu)} \mathbb{P}_\mu \Big(\lambda \in CS_T(r_T(\alpha); \lambda) \Big) = 1-\alpha.$$

An algorithm to Calibrate Projection over a grid G: Let d denote the dimension of μ and let $1-\alpha$ be the desired confidence level.

1. Generate a grid of S scalars $\{r_1, r_2, \ldots, r_S\}$ on the interval $[0, \sqrt{\chi^2_{d,1-\alpha}}]$. Each of these values will serve as the potential 'radius' of the Wald ellipsoid for μ. Fix one element r_s.
2. Generate a grid of I values $G \equiv \{\mu_1, \mu_2, \ldots, \mu_I\} \in \mathcal{M} \subseteq \mathbb{R}^d$. Fix an element $\mu_i \in G$.
3. Generate M i.i.d. draws from the model

$$\widehat{\mu}^i_{T,m} \sim \mathcal{N}_d(\mu_i, \widehat{\Omega}_T/T).$$

   Let $\mathrm{CS}^m_T(r_s, \lambda)$ denote the confidence interval for λ associated to $\widehat{\mu}^i_{T,m}$ with radius r_s. Note that in order to compute the confidence interval for λ, $\widehat{\Omega}_T$ is fixed across all draws.
4. Generate a grid of size K $\{\lambda^i_1, \lambda^i_2, \ldots, \lambda^i_K\}$ from the identified-set for λ given μ_i, denoted $\mathcal{I}^{\mathcal{R}}(\mu_i)$.
5. For each μ_i compute:

$$CP_T(\mu_i; r_s, \widehat{\Omega}_T) \equiv \min_{k \in K} \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\Big\{ \lambda_k \in CS^m_T(r_s; \lambda) \Big\}.$$

6. Report the approximate confidence level of the projection confidence interval with radius r_s as:

$$\mathrm{ApproxCL}_T(r_s) \equiv \min_{i \in I} CP_T(\mu_i; r_s, \widehat{\Omega}_T)$$

7. Find the value in the grid

$$\{\mathrm{ApproxCL}_T(r_1), \ldots \mathrm{ApproxCL}_T(r_S)\}.$$

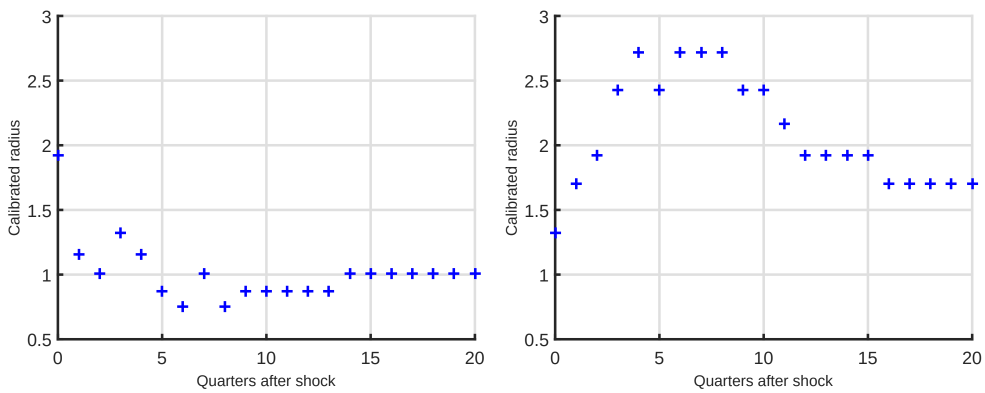
   that is the closest to the desired confidence level $1-\alpha$. Denote this value by $r^*_T(\alpha, G)$.
8. The radius $r^*_T(\alpha, G)$ obtained in Step 6 approximates the value $r_T(\alpha)$ that calibrates frequentist projection.

In our application $\mu \in \mathbb{R}^{27}$, which means that constructing an exhaustive grid for μ is computationally infeasible. To illustrate the computational demands of frequentist calibration in the SVAR exercise, consider a grid G that contains only $\widehat{\mu}_T$. We follow Step 1 to 5 to adjust the confidence set for the responses of wages and employment to a structural demand shock (the first column of Figure 1).

Figure 4 below reports our calibrated radii, horizon by horizon, for the responses of wages and employment to an expansionary demand shock. Note that the default radius used by our projection method is $\chi^2_{27,68\%} = 29.87$.

Figure 4: Calibrated Radii for the 68% Projection Region; $G = \{\widehat{\mu}_T\}$

(Responses to an Expansionary Demand Shock)



(a) Radii for Wages (b) Radii for Employment

(Blue Pluses) For each horizon k and each variable i the blue markers in Panel a) and b) correspond to the calibrated radius $r_T(\alpha, G)$ for $\lambda_{k,i,j}$ (as computed in Step 1 to 5). Each radius is computed using a grid of 16 points ranging from .5 to 5 ($S = 16$ in Step 1); a grid G containing only $\widehat{\mu}_T$ ($I = 1$ in Step 2); 1,000 draws for the reduced-form parameters ($J = 1,000$ in Step 3); and a grid of $1,000$ points for $\lambda_{k,i,j}$ ($K = 1,000$ in Step 4). Generating this figure took approximately 76 hours using 50 parallel Matlab 'workers' on a computer cluster at Bonn University.

Calibrating coverage for a coefficient or a vector of coefficients: One could modify Step 4 in the algorithm to cover a vector of impulse-response functions, as opposed to one particular coefficient. In our application, this alternative calibrated radius (over the grid that contains only $\widehat{\mu}_T$) is 4.21. This radius is designed to cover the vector of responses for wages and employment to a structural demand shock over the 20 quarters under consideration. Calculating this radius took approximately 57 hours using 50 Matlab workers on a private computer cluster at Bonn University.[23]

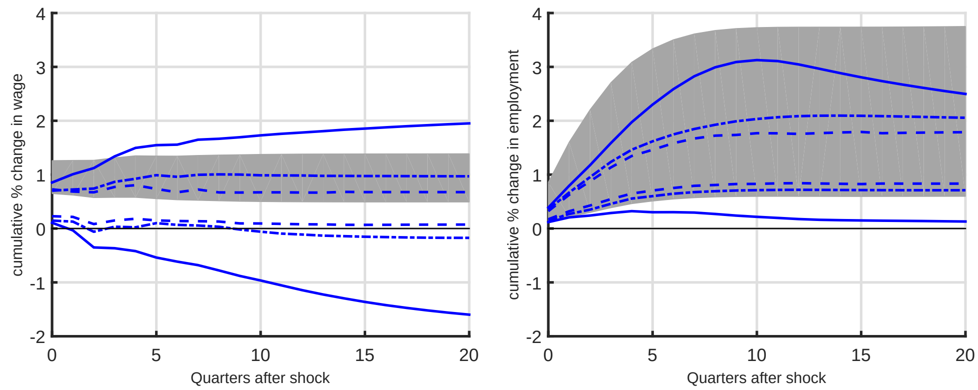
The following figure compares the calibrated projection using the horizon by horizon calibrated radii against the calibrated projection using a radius of 4.21. The calibration over $\mathcal{G}$ implies that the true calibrated radius $r_T(\alpha)$ designed to cover the impulse-response functionshould be larger

[23]The cluster consists of 16 worker-nodes, where each node comprises 8 virtual CPUs and 32 GB virtual RAM, that is a maximum of 8 workers. Each virtual CPU is the core of a Xeon E7-8837@2.67GHz-processor.

than 4.21.

Figure 5: 68% Calibrated Projection for a Frequentist; $G = \{\widehat{\mu}_T\}$

(Responses to an Expansionary Demand Shock)



(a) Cumulative Response of Wages (b) Cumulative Response of Employment

(Solid, Blue Line) 68% Projection region using the default radius $\chi^2_{27,68\%} = 29.87$; (Dash-Dotted, Blue Line) 68% Calibrated Projection Region using the radius 4.21; (Dashed, Blue Line) 68% Calibrated Projection Confidence Region based on the radii in Figure 4; (Shaded, Gray Area) 68% Bayesian Credible Set based on the priors in Baumeister and Hamilton (2015).

APPENDIX C: PROJECTION REGION UNDER DIFFERENTIABILITY

Let h denote some triplet (k,i,j). This section studies the solution to the mathematical program defining the projection region whenever the bounds $\underline{v}_h$, $\overline{v}_h$ are differentiable at a point μ in the parameter space (and the derivative at this point is bounded away from zero). We show that a projection region for the (k,i,j) coefficient of the impulse-response function—indexed by h—is approximately equal to the delta-method confidence interval suggested in Gafarov et al. (2015):

$$\Big[\underline{v}_h(\widehat{\mu}_T) - \frac{r\widehat{\underline{\sigma}}_h}{\sqrt{T}}\, ,\, \overline{v}_h(\widehat{\mu}_T) + \frac{r\widehat{\overline{\sigma}}_h}{\sqrt{T}}\Big].$$

This result has two important consequences:

i) Under differentiability, *the frequentist calibration of projection is straightforward*: it is sufficient to use the square of the $(1-\alpha)$ quantile of a standard normal as the radius of the Wald ellipsoid for the reduced-form parameters. For example, if the desired confidence level is 95%, the radius of the Wald ellipsoid can be set to $(1.64)^2$.

ii) Under differentiability, *the radius that removes excess robust Bayesian credibility also eliminates excess frequentist coverage*. Thus, the robust Bayes calibration is also a frequentist calibration.

C.1. *Lemma 1: Projection Region and Delta-Method confidence interval*

Notation for this Lemma: For $\mu \in \mathbb{R}^p$ define the 'delta-method' type standard errors as:

$$\widehat{\underline{\sigma}}_h(\mu) \equiv \Big(\dot{\underline{v}}_h(\mu)'\widehat{\Omega}_T\dot{\underline{v}}_h(\mu)\Big)^{1/2}, \quad \widehat{\overline{\sigma}}_h(\mu) \equiv \Big(\dot{\overline{v}}_h(\mu)'\widehat{\Omega}_T\dot{\overline{v}}_h(\mu)\Big)^{1/2}.$$

Define also their population counterparts as:

$$\underline{\sigma}_h(\mu) \equiv \Big(\dot{\underline{v}}_h(\mu)'\Omega\dot{\underline{v}}_h(\mu)\Big)^{1/2}, \quad \overline{\sigma}_h(\mu) \equiv \Big(\dot{\overline{v}}_h(\mu)'\Omega\dot{\overline{v}}_h(\mu)\Big)^{1/2},$$

where $\dot{\overline{v}}_h(\mu)$ denotes the derivative of $\overline{v}_h$ at μ. The function $\dot{\underline{v}}_h$ is defined analogously.

For a positive (or negative) $\delta \in \mathbb{R}$, consider the expansion (or contraction) of the delta-method type confidence interval:

$$\text{(C.1)} \qquad \mathrm{DM}^h_T(r,\delta) \equiv \Big[\underline{v}_h(\widehat{\mu}_T) - \frac{(r+\delta)\underline{\sigma}_h(\mu_0)}{\sqrt{T}}\, ,\, \overline{v}_h(\widehat{\mu}_T) + \frac{(r+\delta)\overline{\sigma}_h(\mu_0)}{\sqrt{T}}\Big],$$

where the standard errors evaluated at $\mu_0 \in \mathbb{R}^d$ (the 'true' reduced-form parameter). Note that—up to the term δ—the interval in (C.1) can be interpreted as a 'delta-method' plug-in version of the Imbens and Manski (2004) confidence interval for a set-identified scalar parameter. The following result—which is taken from the recent working paper of Montiel-Olea and Plagborg-Møller (2016) (henceforth, MOPM16)—establishes the relation between the projection region and the confidence interval in (C.1):

Lemma 1 ***(Projection and delta-method confidence interval)*** *Suppose that $\underline{v}_h$ and $\overline{v}_h$ are differentiable at μ_0 with nonzero derivative. Fix constants $\delta > 0, M_1 > 0$, $M_2 > 0$ and $M_3 > 0$. Suppose that the data $(Y_1, \ldots Y_T)$ is such that:*

a) $||\widehat{\mu}_T - \mu_0|| \leq M_1/\sqrt{T}$,

b) The largest eigenvalue of $\widehat{\Omega}_T$, *denoted* $\lambda_{\max}(\widehat{\Omega}_T)$, *is smaller than* M_2.

c) $|\widehat{\overline{\sigma}}_h(\mu_0) - \overline{\sigma}_h(\mu_0)| < \delta/(2M_3)$ *and* $|\widehat{\underline{\sigma}}_h(\mu_0) - \underline{\sigma}_h(\mu_0)| < \delta/(2M_3)$.

Then, there exists $T(\delta, M_1, M_2, M_3)$ *such that for any* $T \geq T_1(\delta, M_1, M_2, M_3)$ *and* $0 < r \leq M_3$*:*

$$DM_T^h(r, -\delta) \subseteq CS_T(r; \lambda_h) \subseteq DM_T^h(r, \delta),$$

where $CS_T(r, \lambda_h)$ *is the projection region for the parameter* λ_h *based on the Wald ellipsoid of radius* r^2*. This means that our projection region with radius* r^2 *is approximately equal—in large samples and under differentiability at* μ_0*—to the delta-method confidence interval in equation (C.1).*

Proof of Lemma 1: The proof of this Lemma is based on the theoretical comparison of Bonferroni and (rectangular) Wald projection confidence regions established in MOPM16. We reproduce their argument for the sake of exposition. Note first that one can write $\overline{v}_h(\mu)$ as:

$$\begin{aligned}
\overline{v}_h(\mu) &= \overline{v}_h(\mu) - \overline{v}_h(\widehat{\mu}_T) + \overline{v}_h(\widehat{\mu}_T) \\
&= \overline{v}_h(\mu) - \overline{v}_h(\mu_0) - (\overline{v}_h(\widehat{\mu}_T) - \overline{v}_h(\mu_0)) + \overline{v}_h(\widehat{\mu}_T).
\end{aligned}$$

For any $\mu \in \mathbb{R}^d$ define the function $\Delta(\mu; \mu_0) \equiv \overline{v}_h(\mu) - \overline{v}_h(\mu_0) - \dot{\overline{v}}_h(\mu_0)'(\mu - \mu_0)$. Therefore:

$$\begin{aligned}
\overline{v}_h(\mu) &= \Delta(\mu; \mu_0) - \Delta(\widehat{\mu}_T; \mu_0) + \dot{\overline{v}}_h(\mu_0)(\mu - \widehat{\mu}_T) + \overline{v}_h(\widehat{\mu}_T) \\
&= \frac{\Delta(\mu; \mu_0)}{||\mu - \mu_0||}||\mu - \mu_0|| - \frac{\Delta(\widehat{\mu}_T; \mu_0)}{||\widehat{\mu}_T - \mu_0||}||\widehat{\mu}_T - \mu_0|| + \dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) + \overline{v}_h(\widehat{\mu}_T).
\end{aligned}$$

Note now that for any μ in the Wald ellipsoid of radius $r \leq M_3$ (i.e., $(\widehat{\mu}_T - \mu)'\widehat{\Omega}_T^{-1}(\widehat{\mu}_T - \mu) \leq r^2/T$) we have that:

$$\begin{aligned}
||\mu - \mu_0|| &\leq ||\mu - \widehat{\mu}_T|| + ||\widehat{\mu}_T - \mu_0|| \\
&\leq ||\mu - \widehat{\mu}_T|| + M_1/\sqrt{T} \\
&\quad \text{(by Assumption a) of the Lemma)} \\
&\leq \max_{\{\mu \,|\, ||\widehat{\Omega}^{-1/2}(\mu - \widehat{\mu}_T)|| \leq r/\sqrt{T}\}} ||\mu - \widehat{\mu}_T|| + M_1/\sqrt{T} \\
&= \lambda_{\max}(\widehat{\Omega}_T) r/\sqrt{T} + M_1/\sqrt{T} \\
&\leq (M_2 r + M_1)/\sqrt{T} \\
&\quad \text{(by Assumption b) of the Lemma)} \\
&\leq (M_2 M_3 + M_1)/\sqrt{T} \\
&\quad \text{(since } r \leq M_3).
\end{aligned}$$

By the differentiability of $\overline{v}_h$ at μ_0, there exists $T_1(\delta, M_1, M_2, M_3)$ such that $T \geq T_1(\delta, M_1, M_2, M_3)$ implies that:

$$\left|\frac{\Delta(\mu; \mu_0)}{||\mu - \mu_0||}\right| \leq \frac{\delta \overline{\sigma}_h(\mu_0)}{4(M_2 M_3 + M_1)} \quad \text{and} \quad \left|\frac{\Delta(\widehat{\mu}_T; \mu_0)}{||\widehat{\mu}_T - \mu_0||}\right| \leq \frac{\delta \overline{\sigma}_h(\mu_0)}{4M_1}.$$

This implies that for such large T we have that for any μ in the Wald ellipsoid of radius $r \leq M_3$, denoted $\mathrm{CS}_T(r;\mu)$:

$$\overline{v}_h(\widehat{\mu}_T) + \dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) - \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} \leq \overline{v}_h(\mu) \leq \overline{v}_h(\widehat{\mu}_T) + \dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T) + \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T}.$$

Consequently:

$$\begin{aligned}
\overline{v}_h(\widehat{\mu}_T) + \widehat{\overline{\sigma}}_h(\mu_0)r/\sqrt{T} - \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} &= \overline{v}_h(\widehat{\mu}_T) + \max_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T)\right] - \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} \\
&\quad \text{(since } \max_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T)\right] = \widehat{\overline{\sigma}}_h(\mu_0)r/\sqrt{T}\text{)} \\
&\leq \max_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \\
&\quad \text{(since we have bounded } \overline{v}_h(\mu) \text{ from below)} \\
&\leq \overline{v}_h(\widehat{\mu}_T) + \max_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T)\right] + \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} \\
&\quad \text{(since we have bounded } \overline{v}_h(\mu) \text{ from above)} \\
&= \overline{v}_h(\widehat{\mu}_T) + \widehat{\overline{\sigma}}_h(\mu_0)r/\sqrt{T} + \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T},
\end{aligned}$$

and, likewise,

$$\begin{aligned}
\overline{v}_h(\widehat{\mu}_T) - \widehat{\overline{\sigma}}_h(\mu_0)r/\sqrt{T} - \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} &= \overline{v}_h(\widehat{\mu}_T) + \min_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T)\right] - \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} \\
&\leq \min_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \\
&\leq \overline{v}_h(\widehat{\mu}_T) + \min_{\mu \in \mathrm{CS}_T(r;\mu)} \left[\dot{\overline{v}}_h(\mu_0)'(\mu - \widehat{\mu}_T)\right] + \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T} \\
&= \overline{v}_h(\widehat{\mu}_T) - \widehat{\overline{\sigma}}_h(\mu_0)r/\sqrt{T} + \delta\overline{\sigma}_h(\mu_0)/2\sqrt{T}.
\end{aligned}$$

Finally, note that Assumption c) of the Lemma implies that:

$$\overline{v}_h(\widehat{\mu}_T) + \overline{\sigma}_h(\mu_0)(r-\delta)/\sqrt{T} \leq \max_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \leq \overline{v}_h(\widehat{\mu}_T) + \overline{\sigma}_h(\mu_0)(r+\delta)/\sqrt{T},$$

and

$$\overline{v}_h(\widehat{\mu}_T) - \overline{\sigma}_h(\mu_0)(r+\delta)/\sqrt{T} \leq \min_{\mu \in \mathrm{CS}_T(r;\mu)} \overline{v}_h(\mu) \leq \overline{v}_h(\widehat{\mu}_T) - \overline{\sigma}_h(\mu_0)(r-\delta)/\sqrt{T}.$$

An analogous argument applied to $\underline{v}_h(\mu)$ gives the desired result.

C.2. *Lemma 2: Delta-Method Interval and a Bernstein-von Mises result*

Lemma 1 in the previous subsection will be used to show that calibrating robust Bayesian credibility also calibrates frequentist coverage. We need an additional Lemma before establishing the main result.

We want to show that the posterior probability:

$$P^* \left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \left[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0) r}{\sqrt{T}} \ , \ \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0) r}{\sqrt{T}} \right] \ , \ \forall h = 1, \dots H \middle| Y_1, \dots, Y_T \right),$$

can be approximated by the probability that $2H$ correlated normals with unit variance fall below the threshold r. We now show that this approximation result can be applied as a consequence of the Berstein-von Mises Theorem for the reduced-form parameter μ.

Notation for this Lemma: Let P^* denote the prior distribution over μ. Let μ^* denote a random variable with such distribution. For each index $h = 1, \dots H$ define:

$$\overline{Z}_h^* = \sqrt{T}(\overline{v}_h(\mu^*) - \overline{v}_h(\widehat{\mu}_T)), \quad \underline{Z}_h^* = \sqrt{T}(\underline{v}_h(\mu^*) - \underline{v}_h(\widehat{\mu}_T)).$$

Let the remainder of the first-order Taylor approximation of $\overline{v}_h$ and $\underline{v}_h$ be defined as:

$$\overline{\Delta}_h(\mu, \mu_0) \equiv \overline{v}_h(\mu) - \overline{v}_h(\mu_0) - \dot{\overline{v}}_0(\mu_0)'(\mu - \mu_0),$$

and

$$\underline{\Delta}_h(\mu, \mu_0) \equiv \underline{v}_h(\mu) - \underline{v}_h(\mu_0) - \dot{\underline{v}}_0(\mu_0)'(\mu - \mu_0).$$

Let $Z \sim \mathcal{N}_d(\mathbf{0}, \Omega)$. Given a radius $r > 0$ and a constant $\eta > 0$ define the function:

$$\begin{aligned} \Gamma_H(r, \eta) \quad \equiv \quad & \mathbb{P}\left(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \leq r + \eta, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \leq r + \eta, \ \forall h = 1, \dots H\right) \\ - \quad & \mathbb{P}\left(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \leq r - \eta, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \leq r - \eta, \ \forall h = 1, \dots H\right). \end{aligned}$$

For a given $\epsilon > 0$ and $M > 0$ define $\eta(\epsilon, M)$ as the real-valued function such that:

$$\max_{0 \leq r \leq M} \left| \Gamma_H(r, \eta(\epsilon, M)) \right| < \epsilon.$$

Lemma 2 *Let $\mathcal{B}(\mathbb{R}^d)$ denote the collection of all Borel sets in $\mathbb{R}^d$. Fix $\epsilon > 0, M > 0$. Let $M_1 \equiv M_1(\epsilon)$ be such that:*

$$\mathbb{P}(||Z|| \leq M_1(\epsilon)) = \epsilon/4, \quad Z \sim \mathcal{N}_d(\mathbf{0}, \Omega).$$

Suppose that for every $h = 1, \dots H$, the bounds $\overline{v}_h$ and $\underline{v}_h$ are differentiable at μ_0 with nonzero derivative. Suppose that the data $(Y_1, Y_2, \dots, Y_T)$ is such that:

1.

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} \left| P^* \left(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y_1, \dots, Y_T \right) - \mathbb{P}\left(Z \in B\right) \right| \leq \epsilon/8, \text{ where } Z \sim \mathcal{N}_d(\mathbf{0}, \Omega),$$

Then for any $0 \leq r \leq M$, there is $T_2(\epsilon, M)$ such that for any $T \geq T_2(\epsilon, M)$ the absolute value of the difference between:

$$P^*\left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \left[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right],\ \forall h = 1, \ldots H \middle| Y_1, \ldots, Y_T\right)$$

and

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)r \leq \dot{\underline{v}}_h(\mu_0)'Z,\ \textit{and}\ \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r,\ \forall h = 1, \ldots H\right), Z \sim \mathcal{N}_d(\boldsymbol{0}, \Omega), \tag{C.2}$$

is smaller than ϵ*. This means that the credibility of the delta-method region can be approximated by Equation* (C.2)*.*

Proof: We prove the lemma in two parts. The first part establishes an upper bound and the second one establishes a lower bound.

Part 1: We are interested in the posterior probability:

$$P^*\left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \left[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right], \forall h \leq H \text{ and } ||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T}\ \middle|\ Y_1, \ldots, Y_T\right),$$

which is the same as:

$$P^*\left(-r\underline{\sigma}_h(\mu_0) \leq \underline{Z}_h^* \text{ and } \overline{Z}_h^* \leq r\overline{\sigma}_h(\mu_0),\ \forall h \leq H \text{ and } ||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T}\ \middle|\ Y_1, \ldots, Y_T\right),$$

Write:

$$\begin{aligned} \overline{v}_h(\mu^*) - \overline{v}_h(\mu_0) &= \overline{\Delta}(\mu^*, \mu_0) - \overline{\Delta}(\widehat{\mu}_T, \mu_0) + \dot{\overline{v}}_h(\mu_0)'(\mu^* - \widehat{\mu}_T) \\ &= \frac{\overline{\Delta}(\mu^*, \mu_0)}{||\mu^* - \mu_0||}||\mu^* - \mu_0|| - \frac{\overline{\Delta}(\widehat{\mu}_T, \mu_0)}{||\widehat{\mu}_T - \mu_0||}||\widehat{\mu}_T - \mu_0|| + \dot{\overline{v}}_h(\mu_0)'(\mu^* - \widehat{\mu}_T) \end{aligned}$$

Note that if $||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T}$, then $||\mu^* - \mu_0|| \leq 2M_1/\sqrt{T}$. The differentiability assumption implies that there is $T_2(\epsilon, M)$ large enough such that for $T \geq T_2(\epsilon, M)$:

$$||\mu - \mu_0|| \leq 2M_1/\sqrt{T} \implies \left|\overline{\Delta}_h(\mu, \mu_0)/||\mu - \mu_0||\right| < \eta(\epsilon/2, M)\overline{\sigma}_h(\mu_0)/(4M_1)$$

for all $h = 1, \ldots H$, and

$$||\mu - \mu_0|| \leq 2M_1/\sqrt{T} \implies \left|\underline{\Delta}_h(\mu, \mu_0)/||\mu - \mu_0||\right| < \eta(\epsilon/2, M)\underline{\sigma}_h(\mu_0)/(4M_1)$$

for all $h = 1, \ldots H$. Therefore:

$$-\eta(\epsilon/2, M)\overline{\sigma}_h(\mu_0) + \dot{\overline{v}}_h(\mu_0)'\sqrt{T}(\mu^* - \widehat{\mu}_T) \leq \overline{Z}_n^*.$$

An analogous argument for $\underline{v}_h(\mu^*)$ implies that

$$\underline{Z}_n^* \leq \eta(\epsilon/2, M)\underline{\sigma}_h(\mu_0) + \dot{\underline{v}}_h(\mu_0)'\sqrt{T}(\mu^* - \widehat{\mu}_T).$$

Consequently the posterior probability we are interested in, which can be written as:

$$P^*\left(-r\underline{\sigma}_h(\mu_0) \leq \underline{Z}_h^* \text{ and } \overline{Z}_h^* \leq r\overline{\sigma}_h(\mu_0),\ \forall h \leq H \text{ and } ||\mu^* - \widehat{\mu}_T|| \leq M_1/\sqrt{T}\ \middle|\ Y_1, \ldots, Y_T\right),$$

is smaller than or equal:

$$P^*\left((-r-\eta(\epsilon/2,M))\underline{\sigma}_h(\mu_0) \le \dot{\underline{v}}_h(\mu_0)'(\mu^*-\widehat{\mu}_T) \text{ and } \dot{\overline{v}}_h(\mu_0)'(\mu^*-\widehat{\mu}_T) \le (r+\eta(\epsilon/2,M))\overline{\sigma}_h(\mu_0),\ \forall h \le H \,\middle|\, Y_1,\ldots,Y_T\right).$$

By Assumption 1 of the Lemma 2, the latter probability is at most:

$$\mathbb{P}\left(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \le r+\eta(\epsilon/2,M), \text{ and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \le r+\eta(\epsilon/2,M)\right)+\epsilon/8.$$

Assumption 4 of the Lemma (and a further application of Assumption 3 to $P^*(||\sqrt{T}\mu^*-\widehat{\mu}_T|| \le M_1 \mid Y_1,\ldots,Y_T)$) implies that:

$$P^*\left([\underline{v}_h(\mu^*),\overline{v}_h(\mu^*)] \subseteq \left[\underline{v}_h(\widehat{\mu})-\frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu})+\frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right], \forall h \le H \,\middle|\, Y_1,\ldots,Y_T\right),$$

$$\le$$

$$\mathbb{P}\left(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \le r+\eta(\epsilon/2,M), \text{ and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \le r+\eta(\epsilon/2,M)\right)+\epsilon/2. \tag{C.3}$$

This gives an upper bound to the 'credibility'of the delta-method region.

Part 2: We now derive a lower bound. Start with the probability

$$\mathbb{P}\left(-\dot{\underline{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0) \le r-\eta(\epsilon/2,M), \text{ and } \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0) \le r-\eta(\epsilon/2,M), \forall h \le H\right)-\epsilon/2.$$

Note that the latter probability is smaller than or equal $-3\epsilon/8$ plus:

$$P^*\left(-\frac{\dot{\underline{v}}_h(\mu_0)}{\underline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^*-\widehat{\mu}_T) \le (r-\eta(\epsilon/2,M)), \text{ and } \frac{\dot{\overline{v}}_h(\mu_0)}{\overline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^*-\widehat{\mu}_T) \le (r-\eta(\epsilon/2,M)), \forall h \le H \,\middle|\, Y_1,\ldots,Y_T\right),$$

by an application of Assumption 3 of the Lemma. Moreover, the latter probability is bounded above by:

$$P^*\Bigg(-\frac{\dot{\underline{v}}_h(\mu_0)}{\underline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^*-\widehat{\mu}_T) \le (r-\eta(\epsilon/2,M)), \text{ and } \frac{\dot{\overline{v}}_h(\mu_0)}{\overline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^*-\widehat{\mu}_T) \le (r-\eta(\epsilon/2,M)), \forall h \le H, \text{ and}$$
$$||\mu^*-\widehat{\mu}_T|| \le \frac{M_1}{\sqrt{T}} \,\Bigg|\, Y_1,\ldots,Y_T\Bigg),$$

by an application of Assumptions 3 and 4 of the Lemma and the monotonicity of probability measures. Finally, Assumption 1 and the differentiability of $\overline{v}_h$ at μ_0 implies that

$$\overline{Z}^*_n \le \dot{\overline{v}}_h(\mu_0)'\sqrt{T}(\mu^*-\widehat{\mu}_T)+\eta(\epsilon/2,M)\overline{\sigma}_h(\mu_0).$$

Assumption 2 and analogous argument for $\underline{v}_h(\mu^*)$ implies that

$$-\eta(\epsilon/2,M)\underline{\sigma}_h(\mu_0)+\dot{\underline{v}}_h(\mu_0)'\sqrt{T}(\mu^*-\widehat{\mu}_T) \le \underline{Z}^*_n.$$

Consequently,

$$P^*\left(-\frac{\underline{\dot{v}}_h(\mu_0)}{\underline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^*-\widehat{\mu}_T)\leq(r-\eta(\epsilon/2,M)),\ \text{and}\ \frac{\dot{\overline{v}}_h(\mu_0)}{\overline{\sigma}_h(\mu_0)}'\sqrt{T}(\mu^*-\widehat{\mu}_T)\leq(r-\eta(\epsilon/2,M)),\right.$$
$$\left.\forall h\leq H,\ ||\mu^*-\widehat{\mu}_T||\leq\frac{M_1}{\sqrt{T}}\ \middle|\ Y_1,\ldots,Y_T\right),$$

is bounded above by:

$$P^*\left(-r\underline{\sigma}_h(\mu_0)\leq\underline{Z}_h^*\ \text{and}\ \overline{Z}_h^*\leq r\overline{\sigma}_h(\mu_0),\ \forall h\leq H\ \middle|\ Y_1,\ldots,Y_T\right).$$

We conclude that:

$$\mathbb{P}\left(-\underline{\dot{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0)\leq r-\eta(\epsilon/2,M),\ \text{and}\ \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0)\leq r-\eta(\epsilon/2,M)\right)-\epsilon/2 \tag{C.4}$$
$$\leq$$
$$P^*\left([\underline{v}_h(\mu^*),\overline{v}_h(\mu^*)]\subseteq\left[\underline{v}_h(\widehat{\mu})-\frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu})+\frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right],\forall h\leq H\ \middle|\ Y_1,\ldots,Y_T\right).$$

Conclusion: Equations (C.3) and (C.4) imply that

$$\mathbb{P}\left(-\underline{\dot{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0)\leq r-\eta(\epsilon/2,M),\ \text{and}\ \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0)\leq r-\eta(\epsilon/2,M)\right)-\epsilon/2$$
$$\leq$$
$$P^*\left([\underline{v}_h(\mu^*),\overline{v}_h(\mu^*)]\subseteq\left[\underline{v}_h(\widehat{\mu})-\frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu})+\frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right],\forall h\leq H\ \middle|\ Y_1,\ldots,Y_T\right)$$
$$\leq$$
$$\mathbb{P}\left(-\underline{\dot{v}}_h(\mu_0)'Z/\underline{\sigma}_h(\mu_0)\leq r+\eta(\epsilon/2,M),\ \text{and}\ \dot{\overline{v}}_h(\mu_0)'Z/\overline{\sigma}_h(\mu_0)\leq r+\eta(\epsilon/2,M)\right)+\epsilon/2.$$

These bounds, the monotonicity of the probability measure, and the definition of $\eta(\epsilon/2,M)$ imply the desired result. *Q.E.D.*

C.3. ***Lemma 3: Asymptotic Behavior of the radius that calibrates r_T^* Bayesian credibility***

The previous lemma showed that for any $0 \leq r \leq M$ the probability

$$P^*\left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \left[\underline{v}_h(\widehat{\mu}) - \frac{\underline{\sigma}_h(\mu_0)r}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu}) + \frac{\overline{\sigma}_h(\mu_0)r}{\sqrt{T}}\right],\ \forall h = 1, \ldots H \middle| Y_1, \ldots, Y_T\right)$$

is approximately the same as

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)r \leq \dot{\underline{v}}_h(\mu_0)'Z,\ \text{and}\ \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r,\ \forall h = 1, \ldots H\right), Z \sim \mathcal{N}_d(\mathbf{0}, \Omega).$$

Let $r_{1-\alpha-\epsilon}$ denote the 'critical value' such that:

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)r_{1-\alpha-\epsilon} \leq \dot{\underline{v}}_h(\mu_0)'Z,\ \text{and}\ \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r_{1-\alpha-\epsilon},\ \forall h = 1, \ldots H\right) = 1 - \alpha - \epsilon,$$

and let $r_{1-\alpha+\epsilon}$ be defined analogously. We now show that if the sample size is large enough and data satisfies the Assumptions of Lemma 1 and Lemma 2 then:

$$r_{1-\alpha-\epsilon} \leq r_T^* \leq r_{1-\alpha+\epsilon}.$$

Lemma 3 *Fix $\epsilon > 0$, $M_1, M_2, M_3 > 0$. Suppose that the data $(Y_1, \ldots, Y_T)$ is such that the Assumptions of Lemma 1 and Lemma 2 are satisfied:*

i) $||\widehat{\mu}_T - \mu_0|| \leq M_1/\sqrt{T}$,

ii) The largest eigenvalue of $\widehat{\Omega}_T$, denoted $\lambda_{\max}(\widehat{\Omega}_T)$, is smaller than M_2.

iii) $|\widehat{\overline{\sigma}}_h(\mu_0) - \overline{\sigma}_h(\mu_0)| < \eta(\epsilon/2, M_3)/(2M_3)$ *and* $|\widehat{\underline{\sigma}}_h(\mu_0) - \underline{\sigma}_h(\mu_0)| < \eta(\epsilon/2, M_3)/(2M_3)$, *for all $h = 1, \ldots H$; where $\eta(\cdot)$ is defined as in Lemma 2.*

iv)

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} \left| P^*\left(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y_1, \ldots, Y_T\right) - \mathbb{P}\left(Z \in B\right) \right| \leq \epsilon/8,\ \text{where}\ Z \sim \mathcal{N}_d(\mathbf{0}, \Omega),$$

Then, there is $T_3(\epsilon, M_1, M_2, M_3)$ such that for $T \geq T_3(\epsilon, M_1, M_2, M_3)$:

$$r_{1-\alpha-\epsilon} \leq r_T^* \leq r_{1-\alpha+\epsilon},$$

provided $r_{1-\alpha+\epsilon} < M_3$.

Proof: Without loss of generality, we can assume that $r_{1-\alpha+\epsilon} + \eta(\epsilon/2, M_3) < M_3$. Note that Lemma 1 implies that for $T \geq T_1(\eta(\epsilon/2, M_3), M_1, M_2, M_3)$:

$$P^*\left(\times_{h=1}^H [\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq \text{CS}_T(r_{1-\alpha-\epsilon}, \lambda^H) \;\middle|\; Y_1, \ldots, Y_T\right)$$

minus

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)r_{1-\alpha-\epsilon} \leq \dot{\underline{v}}_h(\mu_0)'Z,\ \text{and}\ \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)r_{1-\alpha-\epsilon},\ \forall h = 1, \ldots H\right)$$

is bounded above by the sum of two terms. The first term is the difference between

$$P^*\left([\underline{v}_h(\mu^*), \overline{v}_h(\mu^*)] \subseteq DM_T^h(r_{1-\alpha-\epsilon}\ ,\ \eta(\epsilon/2\ ,\ M_3)),\ \forall h \leq H \middle| Y_1, \ldots, Y_T\right)$$

and

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon}+\eta(\epsilon/2,M_3)) \leq \dot{\underline{v}}_h(\mu_0)'Z, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon}+\eta(\epsilon/2,M_3)), \ \forall h=1,\dots H\right)$$

The second term is the difference between the latter probability and

$$\mathbb{P}\left(-\underline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon}) \leq \dot{\underline{v}}_h(\mu_0)'Z, \text{ and } \dot{\overline{v}}_h(\mu_0)'Z \leq \overline{\sigma}_h(\mu_0)(r_{1-\alpha-\epsilon}), \ \forall h=1,\dots H\right).$$

Lemma 2 implies that the magnitude of the first term is bounded above by $\epsilon/2$ if $T \geq T_2(\epsilon/2, M_3)$. The definition of $\eta(\cdot)$ implies that the second term is bounded above by $\epsilon/2$. Therefore, we conclude that:

$$P^*\left(\times_{h=1}^{H}\left[\underline{v}_h(\mu^*),\overline{v}_h(\mu^*)\right] \subseteq \text{CS}_T(r_{1-\alpha-\epsilon},\lambda^H) \,\middle|\, Y_1,\dots,Y_T\right) - (1-\alpha-\epsilon) \leq \epsilon,$$

which implies that for $T \geq \max\{T_1(\eta(\epsilon/2,M_3),M_1,M_2,M_3),T_2(\epsilon/2,M_3)\}$:

$$r_{1-\alpha-\epsilon} \leq r_T^*,$$

as the credibility of the projection region is monotone in its radius. An analogous argument implies that for $T \geq T_1(\eta(\epsilon/2,M_3),M_1,M_2,M_3)$:

$$r_{1-\alpha-\epsilon} \leq r_T^* \leq r_{1-\alpha+\epsilon}.$$

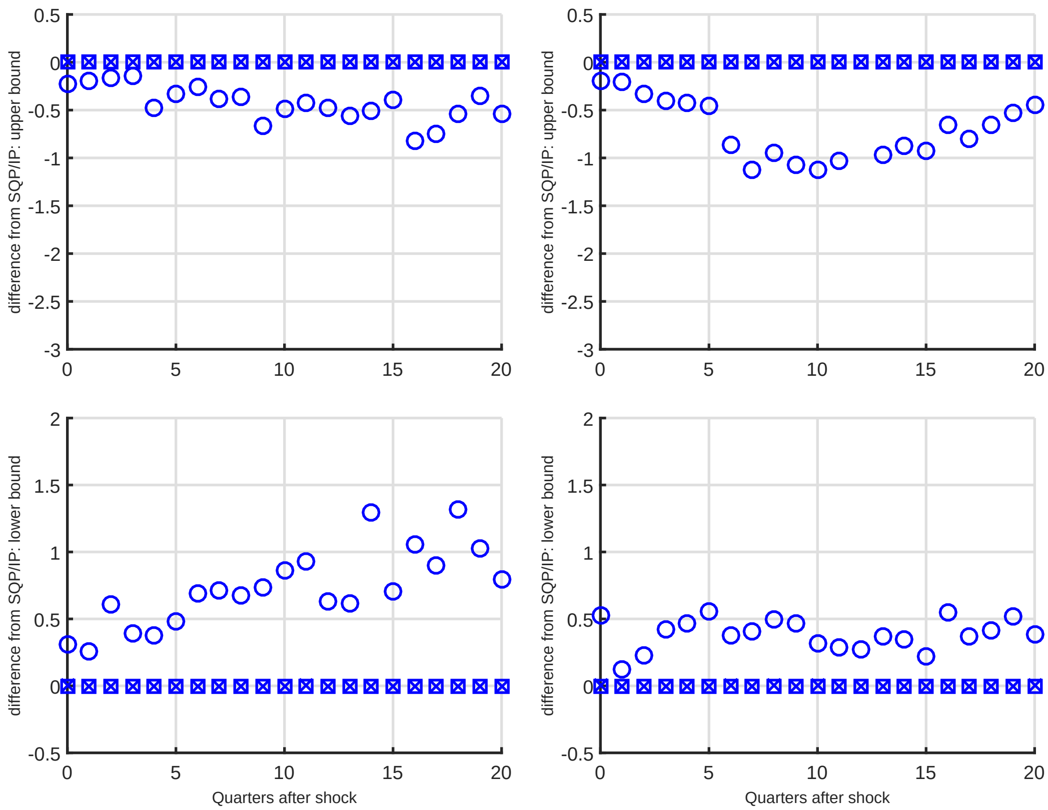
Q.E.D.

C.4. *Proof of Result*

APPENDIX D: ADDENDA FOR IMPLEMENTATION

D.1. *SQP/IP vs. Global Methods*

Figure 6: Accuracy of SQP/IP for a demand shock

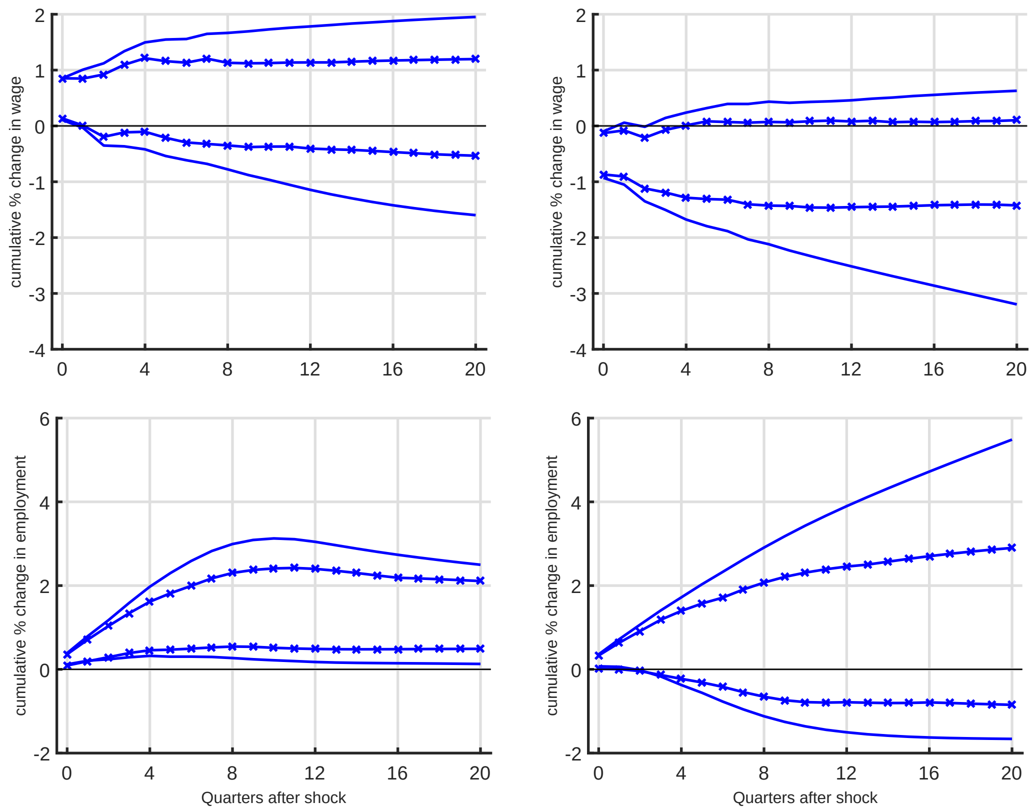


(a) Wage Response (b) Employment Response

(Square, Blue) Optimal Value reported by SQP/IP minus Optimal Value reported by SQP/IP + `Multistart`; (Cross, Blue) Optimal Value reported by SQP/IP minus Optimal Value reported by SQP/IP + `Global Search`; (Circle, Blue) Optimal Value reported by SQP/IP minus Optimal Value reported by `ga`.

D.2. *SQP/IP vs. Grid Search on* $CS_T(1-\alpha, \mu)$

Figure 7: Simulation error in Projection region.



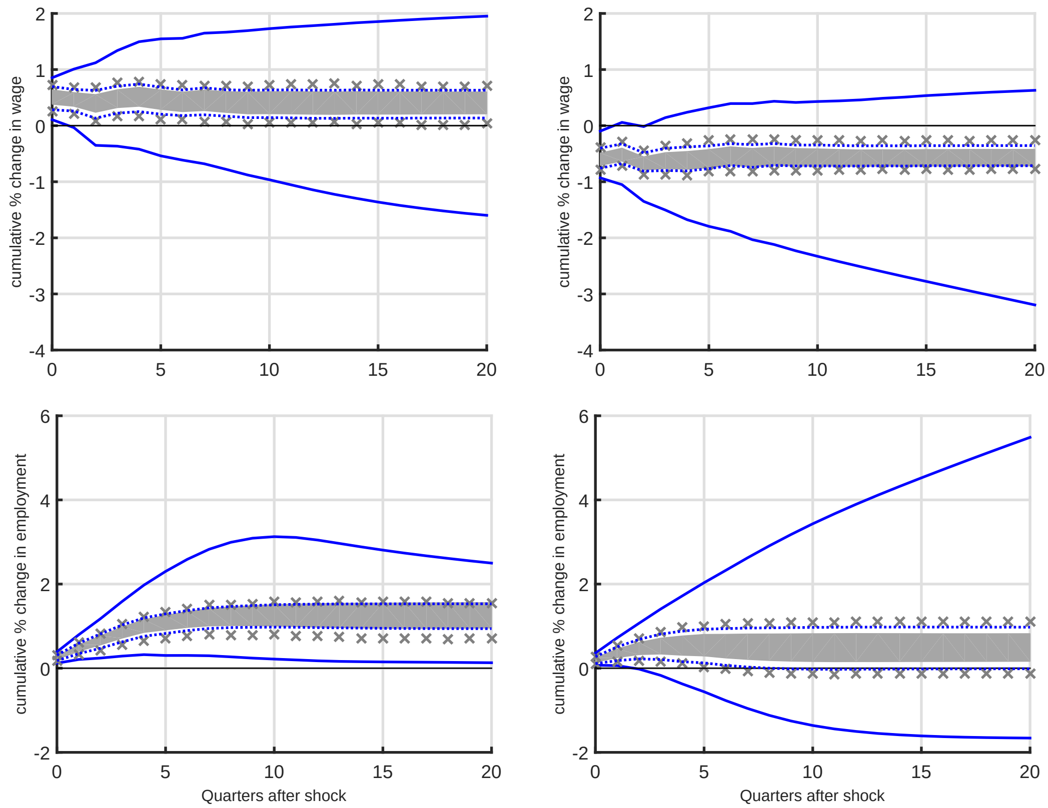
(a) Expansionary Demand Shock (b) Expansionary Supply Shock

(Solid Line) 68% Projection region using the SQP/IP algorithm described in Section 4; (Connected, Solid Line) 68% Projection region using a two-step algorithm: 1) Sample M=100,000 reduced form parameters that satisfy the 68% Wald ellipsoid constraint. 2) For each draw, solve for the identified set. The smallest and largest value of the identified set is the simulation-based approximation of the Projection region.

D.3. *Comparison with the credible set in Giacomini and Kitagawa (2015)*

Figure 8: 68% Differentiable Projection and 68% GK Robust Credible Set.

(Uhlig (2005) priors)



(a) Expansionary Demand Shock

(b) Expansionary Supply Shock

(Solid, Blue Line) 68% Frequentist Projection Confidence Interval; (Shaded, Gray Area) 68% Bayesian Credible Set based on the priors in Uhlig (2005); (Dotted, Blue Line) 68% Calibrated Projection Confidence Interval. The calibration is implemented assuming differentiability of the bounds of the identified set and strict set-identification of the structural parameter; (Crosses, Gray) 68% Robust Credible Set based on Giacomini and Kitagawa (2015) using the priors for the reduced-form parameters described in Uhlig (2005).