Projection Inference for Set-Identified Svars

Bulat Gafarov (PSU), Matthias Meier (University of Bonn), and José-Luis Montiel-Olea (Columbia)

September 21, 2016

Introduction: Set-id. SVARs

⋆ SVAR: Theoretical restrictions $\mathcal{R}$ imposed on a VAR.
(Sims [1980, 1986])

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \eta_t, \quad \Sigma = \mathbb{E}[\eta_t \eta_t']$$

⋆ Goal of the restrictions: $(A_1, \ldots A_p, \Sigma) \mapsto_{\mathcal{R}} \text{IRF}_{k,i,j}$
(response of variable i to a j-th 'structural shock' at horizon k)

⋆ Map '$\mapsto_{\mathcal{R}}$' can be 1-to-1 (point id.) or 1-to-many (set id.).
(set i.d. SVARs have become popular in applied macro work)

⋆ Common practice: set-identify SVARs with $\geq$ / $=$ restrictions.
(Faust [1998]; Canova and De Nicolo [2002]; Uhlig[2005])

Motivation

- ⋆ Most empirical studies report Bayesian *credible* sets for $\text{IRF}_{k,i,j}$
  (Bayesian Inference depends on the specification of prior beliefs)

- ⋆ Practical concern: prior beliefs are not 'dominated' by the data
  (results are sensitive to the choice of priors even if $T \to \infty$)

- ⋆ Theoretical Critique: Coverage and 'Robust'credibility $\to 0$.
  (as $T \to \infty$; Moon & Schorfheide [2012], Kitagawa [2012])

- ⋆ Recent work on non-Bayesian Inference for set-id. SVARs.
  (MSG [2013]-Freq. Inference; GK [2014]-Robust Bayes)

- ⋆ Is there a simple way to conduct inference in set-id. SVARs?
  (that pleases both a frequentist and a robust Bayesian, and that is general and computationally feasible?)

Description of the Inference Problem

$$\mathrm{IRF}_{k,i,j} \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu) \subseteq \left[\underline{v}_{k,i,j}(\mu)\ ,\ \overline{v}_{k,i,j}(\mu)\right],$$

$$\mu \equiv (A, \Sigma).$$

This paper

- ⋆ Studies the properties of 'projection inference' for set i.d. SVARs.
  (Scheffé [1953]; Dufour [1990]; Dufour and Taamouti [2005])

- ⋆ We collect $\text{IRF}_{k,i,j}$'s in a $1-\alpha$ Wald Ellipsoid for $\mu \equiv (A, \Sigma)$.
  (that is, we 'project' a nominal $1-\alpha$ Wald Ellipsoid)

- ⋆ Strategy: focus on the endpoints of the identified set for $\text{IRF}_{k,i,j}$.
  (the maximum and minimum response, $\overline{v}_{k,i,j}(\mu)$, $\underline{v}_{k,i,j}(\mu)$)

$$\left[\inf_{\mu \in CS_T(1-\alpha)} \underline{v}_{k,i,j}(\mu), \sup_{\mu \in CS_T(1-\alpha)} \overline{v}_{k,i,j}(\mu) \right]$$

- ⋆ Our projection region has coverage and RB credibility $\geq 1-\alpha$.
  (for any vector of IRFs! Thus providing simultaneous inference)

Pros & Cons

Pros:

- ⋆ Generality: can handle the typical application in applied work (+/0 restrictions on IRFs, long-run restrictions, elasticity bounds)
- ⋆ Feasibility: solve two nonlinear optimization problems per $\text{IRF}_{k,i,j}$ (we use state-of-the-art solution algorithms for these problems)

Cons:

- ⋆ Projection is conservative for a frequentist and a Robust Bayesian (coverage and robust credibility are strictly above 1-α.)
- ⋆ We 'calibrate' projection to remove the excess of Robust Cred. ($= 1 - \alpha$ and not $> 1 - \alpha$. Calibration based on KMS[2016])

Outline

1. Model and Main Definitions
2. Assumptions and Results
3. Implementation and Illustrative Example
4. Conclusion

1. Model and Main Definitions

SVAR(P)

- Structural VAR for the n-dimensional vector Y_t:

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + B\varepsilon_t, \quad \Sigma \equiv BB'$$

- Vector of reduced form parameters is:

$$\mu = (\text{vec}(A_1, A_2, \ldots, A_p)', \text{vech}(\Sigma)')' \in \mathbb{R}^d$$

- Coefficients of the Structural Impulse Response Function:

$$\text{IRF}^H = \{\text{IRF}_{k_h,i_h,j_h}(A,B)\}_{h=1}^H, \quad \text{IRF}_{k_h,i_h,j_h}(A,B) = \underbrace{e'_{i_h} C_{k_h}(A)}_{1\times n} B_{j_h}.$$

- Interested in simulatenous inference about $\lambda^H \equiv \text{IRF}^H$.
  (Inoue and Kilian [2013,2016] and Lütkepohl et. al [2016])

Restrictions $\mathcal{R}(\mu)$ on B

- $\star$ Identified set for λ^H:

$$\mathcal{I}_H^{\mathcal{R}}(\mu) \equiv \left\{ \lambda \in \mathbb{R}^H \,\middle|\, \lambda_h = \mathrm{IRF}_{k_h, i_h, j_h} \text{ s.t. } BB' = \Sigma,\ B \in \mathcal{R}(\mu),\ \forall h \right\}$$

- $\star$ $\pm/0$ restrictions on IRFs: $e_{i'}' C_{k'}(A) B_{j'} \geq 0$
  (e.g. Sims [1980], Uhlig [2005])

- $\star$ $\pm/0$ long-run restrictions: $e_{i'}'(\mathbb{I}_n - A(1))^{-1} B_{j'} \geq 0$
  (e.g. Blanchard, Quah [1989], Gali [1999])

- $\star$ Elasticity bounds: $(e_{i'}' B_{j'})/(e_i' B_{j'}) \in [c, d]$
  (e.g. Kilian, Murphy [2012], Baumeister, Hamilton [2015])

Bounds on the Identified Set: Max and Min Response

- ⋆ The endpoints of the identified set for each $\text{IRF}_{k,i,j}$:

$$\overline{v}_{k,i,j}(\mu) \equiv \sup_{B} \text{IRF}_{k,i,j}(A,B) \quad \text{s.t.} \quad BB' = \Sigma,\; B \in \mathcal{R}(\mu)$$

$$\underline{v}_{k,i,j}(\mu) \equiv \inf_{B} \text{IRF}_{k,i,j}(A,B) \quad \text{s.t.} \quad BB' = \Sigma,\; B \in \mathcal{R}(\mu)$$

- ⋆ Nonlinear, possibly nondifferentiable transformations of μ.

- ⋆ Obviously ...

$$\mathcal{I}_H^{\mathcal{R}}(\mu) \subseteq \times_{h=1}^{H} \left[\underline{v}_{k_h,i_h,j_h}(\mu)\;,\;\overline{v}_{k_h,i_h,j_h}(\mu)\right].$$

- ⋆ No need to assume the i.d. set is connected.

Projection region for λ^H

- Let $\text{CS}_T(1-\alpha;\mu)$ be the (typical) Wald ellipsoid for μ.
- Let $CS_T(1-\alpha;\text{IRF}_{k,i,j})$ be the interval defined by:

$$\left[\inf_{\mu\in CS_T(1-\alpha;\mu)} \underline{v}_{k,i,j}(\mu), \sup_{\mu\in CS_T(1-\alpha;\mu)} \overline{v}_{k,i,j}(\mu)\right]$$

- The projection region for $\lambda^H = \{\text{IRF}_{k_h,i_h,j_h}(A,B)\}_{h=1}^{H}$ is:

$$\text{CS}_T(1-\alpha;\lambda^H) \equiv$$

$$CS_T(1-\alpha;\text{IRF}_{k_1,i_1,j_1}) \times \ldots \times CS_T(1-\alpha;\text{IRF}_{k_H,i_H,j_H})$$

- We now present the properties of $\text{CS}_T(1-\alpha;\lambda^H)$ as $T\to\infty$

2. Assumptions and Results 1 to 4

Result 1: Frequentist Coverage

- ⋆ Let P be a DGP for the data. Parameterized by (A, B, F).
- ⋆ We want projection to be valid over a class $\mathcal{P}$ of DGPs:
- ⋆ **A1:** Suppose the class of DGPs $\mathcal{P}$ is such that

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

- ⋆ **R1:** Under Assumption A1:

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

Proof: Straightforward Projection Argument

Suppose that $H = 1$. For any $\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))$:

$$P\Big(\lambda \in \Big[\inf_{\mu \in CS_T(1-\alpha)} \underline{v}_{k,i,j}(\mu), \sup_{\mu \in CS_T(1-\alpha)} \overline{v}_{k,i,j}(\mu)\Big]\Big)$$

$$\geq$$

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

$$\Big(\text{as } \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P)) \subseteq [\underline{v}_{k,i,j}(\mu(P)), \overline{v}_{k,i,j}(\mu(P))]\Big)$$

$$\geq$$

$$P\Big(\mu(P) \in CS_T(1-\alpha)\Big).$$

Robust Bayes Framework

⋆ Let P^* be a prior for the structural parameters (A, B).
(F is now a fixed known distribution; we use $\mathcal{N}(0, \mathbb{I}_n)$)

⋆ Represent the prior P^* in terms of $(P^*_\mu, P^*_{Q|\mu})$, $Q \equiv \Sigma^{-1/2}B$.
(Orthogonal reduced-form parameterization Arias et. al [2014])

⋆ Let $\mathcal{P}(P^*_\mu)$ denote the class of priors such that $\mu \sim P^*_\mu$.

⋆ The *robust credibility* of $\text{CS}_T(1-\alpha, \lambda^H)$ is defined as:

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

Result 2: Robust Bayesian credibility

- ⋆ We can view robust credibility as a random variable (as it depends on the data $\mathbf{Y}_T$)

- ⋆ **A2**: Suppose that P^* is such that whenever $\mathbf{Y}_T \sim f(\mathbf{Y}_T|\mu_0)$:

$$P^*(\mu(A,B) \in \mathrm{CS}_T(1-\alpha;\mu) \,|\, \mathbf{Y}_T) = 1 - \alpha + o_p(\mathbf{Y}_T|\mu_0).$$

- ⋆ This is implied by the Bernstein von-Mises Theorem for μ.

- ⋆ **R2**: Under Assumption 2:

$$\inf_{P^* \in \mathcal{P}(\mathcal{P}^*_\mu)} P^*\left(\lambda^H \in CS_T(1-\alpha;\lambda^H) \,\middle|\, \mathbf{Y}_T\right) \geq 1 - \alpha + o_p(\mathbf{Y}_T|\mu_0)$$

- ⋆ Proof: Another embarrassingly simple projection argument!

Calibrated Projection

- ⋆ Yes: We know that projection inference is conservative!
  (both in terms of frequentist coverage and a robust credibility)

- ⋆ In theory, it is conceptually simple to remove ‘projection bias’
  (project a smaller Wald ellipsoid as suggested by KMS[2016])

- ⋆ In practice, removing the excess of robust Bayesian credibility is much easier than removing the excess of frequentist coverage.

- ⋆ We suggest an algorithm to ‘calibrate’ robust credibility.

- ⋆ The algorithm also removes the excess of frequentist coverage
  (provided the bounds of i.d. set are differentiable)

Result 3: Calibrated Robust Credibility

- ⋆ Our calibration algorithm is based on the following result.
- ⋆ Suppose there is a nominal level $1 - \alpha^*(\mathbf{Y}_T)$ s.t:

$$P^*_\mu\Big(\times_{h=1}^{H} \left[\underline{v}_{k_h,i_h,j_h}(\mu), \overline{v}_{k_h,i_h,j_h}(\mu)\right] \subseteq \text{CS}_T(1 - \alpha^*(\mathbf{Y}_T), \lambda^H) \mid \mathbf{Y}_T\Big) = 1 - \alpha$$

- ⋆ **R3**: Then, for every data realization:

$$\inf_{P^* \in \mathcal{P}(P^*_\mu)} P^*\left(\lambda^H(A, B) \in CS_T(1 - \alpha^*(\mathbf{Y}_T); \lambda^H) \,\middle|\, \mathbf{Y}_T\right) = 1 - \alpha,$$

- ⋆ Proof: slightly more involved. See Appendix A.2.

Calibration Algorithm

- ⋆ Take M draws (μ_m^*) from the posterior distribution of μ (or from its asymptotic approximation based on BvM)

- ⋆ For each $h = 1, \ldots H$ and each $m = 1, \ldots, M$ evaluate:

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

(we use nonlinear numerical solvers to evaluate the bounds)

- ⋆ Fix a confidence level $1 - \alpha_s < 1 - \alpha$. Count how often:

$$[\underline{v}_{k_h,i_h,j_h}(\mu_m^*)\ ,\ \overline{v}_{k_h,i_h,j_h}(\mu_m^*)] \subseteq \mathrm{CS}_T(1 - \alpha_s; \lambda_{k_h,i_h,j_h})$$

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

- ⋆ If RBC is not in between $[1 - \alpha - \eta, 1 - \alpha + \eta]$, change α_s. (η is a tolerance level for the excess of robust credibility)

Result 4: Coverage of Calibrated Projection

- ⋆ Question: Suppose we have found $\alpha^*(Y_T)$. Is it true that:

$$P_{\mu_0}\Big([\ \underline{v}_h(\mu_0), \overline{v}_h(\mu_0)\ ] \subseteq \text{CS}_T(1-\alpha^*(\mathbf{Y}_T);\lambda_h),\ \forall h = 1,\ldots,H\Big) \to 1-\alpha\ ?$$

- ⋆ Answer: Yes. The following regularity conditions suffice:
- ⋆ $\underline{v}_h(\mu_0)$, $\overline{v}_h(\mu_0)$ are differentiable at μ_0 for each $h = 1,\ldots,H$.
- ⋆ $\sqrt{T}(\widehat{\mu}-\mu_0) \overset{d}{\to} N(0,\Omega)$, $\widehat{\Omega}_T \overset{p}{\to} \Omega$, and BvM holds.
- ⋆ Proof: Under differentiability $\text{CS}_T(1-\alpha^*(Y_T);\lambda_h)$ is approximately:

$$\left[\underline{v}_h(\widehat{\mu}_T) - \frac{r^*_T \underline{\sigma}_h(\mu_0)}{\sqrt{T}}\ ,\ \overline{v}_h(\widehat{\mu}_T) + \frac{r^*_T \overline{\sigma}_h(\mu_0)}{\sqrt{T}}\right]$$

3. Implementation and illustrative example

Projection as an Optimization Problem

- ⋆ To implement projection we need to solve the program:

$$\sup_{\mu \in \mathrm{CS}_T(1-\alpha;\mu)} \overline{v}_{k,i,j}(\mu)$$

- ⋆ The confidence set for the reduced-form parameters is taken as:

$$\mathrm{CS}_T(1-\alpha) \equiv \left\{ \mu \in \mathbb{R}^d \,\middle|\, T(\widehat{\mu}_T - \mu)'\widehat{\Omega}_T^{-1}(\widehat{\mu}_T - \mu) \leq \chi^2_{1-\alpha,d} \right\}$$

- ⋆ Thus, the program of interest becomes:

$$\sup_{\mu \in \mathrm{CS}_T(1-\alpha)} \left(\sup_{B \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu)} e_i' C_k(A) B_j \right)$$

- ⋆ Which is a twice differentiable, non-convex, nonlinear program:

$$\sup_{\mu, B} e_i' C_k(A) B_j \quad \text{s.t. } B \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu) \text{ and } \mu \in \mathrm{CS}_T(1-\alpha)$$

Solution Algorithm

- ⋆ State-of-the-art nonlinear, non-convex, large-scale problems?
- ⋆ Large literature on local and global optimization algorithms
  (Local: AL, SQP, IP; Global: Multistart, GSearch, GAs)
- ⋆ Since the problem is non-convex, local solvers are not enough.
  (we use a two-phase solution algorithm: local+global)
- ⋆ SQP/IP 'most powerful local algorithm for nonlinear prog.'
  (Nocedal and Wright [2006], p. 253; implemented in `fmincon`)
- ⋆ For the global stage we use `Multistart`, `GlobalSearch`, `ga`
  (take the local solution as an input for global algorithm)

Example: Demand-Supply SVAR BH(2015), ECMA

⋆ Effect of a structural shock on labor demand over wages/emp? (2-SVAR, 6 lags (1-AIC, 1-BIC), Q1-70/Q2-14: $\Delta w_t, \Delta\eta_t$.)

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

⋆ Sign restrictions set-identify structural demand and supply shocks:

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

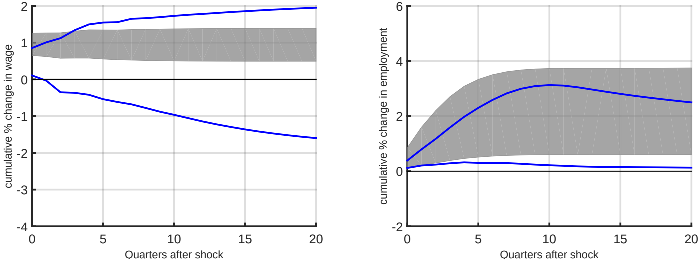
⋆ Elasticities of supply and demand: $\alpha \equiv b_2/b_1 \quad \beta \equiv b_4/b_3$.

Table: Additional Identifying Restrictions

Motivation	**BH(2015)**	**This paper**
Empirical studies report $\alpha \in [.27, 2]$	$\alpha \sim \max\{.6 + .6t_3, 0\}$	$.27 \leq \alpha \leq 2$
Empirical studies $\beta \in [-2.5, -.15]$	$\beta \sim \min\{-.6 + .6t_3, 0\}$	$-2.5 \leq \beta \leq -.15$
$\gamma = 0$ is too strong	$\gamma \sim \mathcal{N}(0, V)$	$-2V \leq \gamma \leq 2V$

$$\gamma \equiv e_2'(\mathbb{I}_n - A_1 - A_2 - \ldots - A_6)^{-1} B_1$$

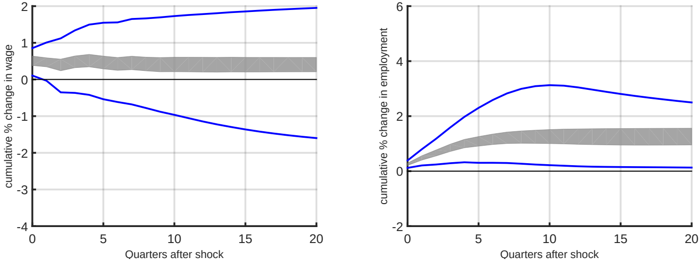
68% Projection Region and 68% Credible Set.



Expansionary Demand Shock

(BH priors)

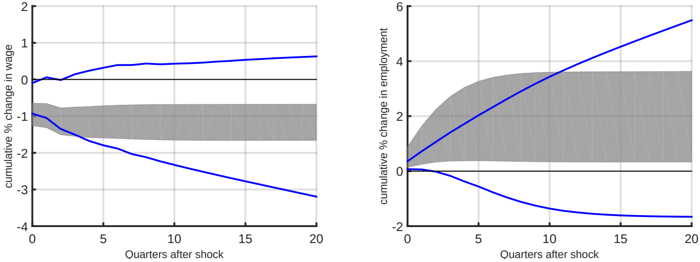
68% Projection Region and 68% Credible Set.



Expansionary Demand Shock

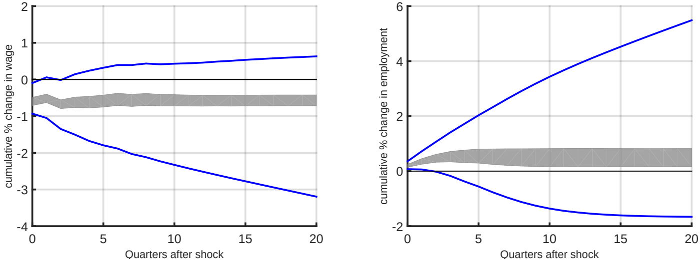
(Uhlig's priors)

68% Projection Region and 68% Credible Set.



Expansionary Supply Shock
(BH priors)

68% Projection Region and 68% Credible Set.



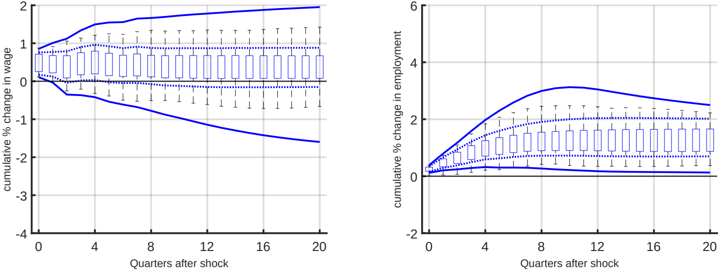
Expansionary Supply Shock

(Uhlig's priors)

Robustness or Conservativeness?

- ⋆ Projection is informative about the effects of ϵ^d on η_t ...
- ⋆ But not very informative about the rest of the dynamic effects.
- ⋆ This could be a consequence of the 'robustness'of projection,
- ⋆ Or a consequence of its conservativeness ($> 1 - \alpha$).
- ⋆ To separate these effects, we report the calibrated projection. (represented as dotted line in the following figures)

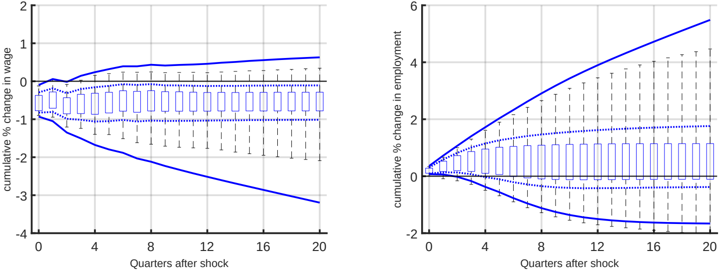
68% Projection Region and 68% Calibrated Projection



Expansionary Demand Shock

Boxes: horizon-by-horizon 68% robust Bayesian credible set
Whiskers: minimum/maximum IRF (100,000 draws)

68% Projection Region and 68% Calibrated Projection



Expansionary Supply Shock

Boxes: horizon-by-horizon 68% robust Bayesian credible set
Whiskers: minimum/maximum IRF (100,000 draws)

Comments

- ⋆ Credible sets differ substantially depending on the prior beliefs.
  (compare BH with Uhlig priors)

- ⋆ Prior-free, 'projected' region: qualitatively different inference.
  (only the employment response to demand shock significant)

- ⋆ SQP/IP: 12 min; Uhlig: 38 min; BH: 66min; Global: 9hrs.
  (see Table III, p. 20 in the paper)

- ⋆ Global algorithms do not improve the local solution.
  (see Appendix B, p. 48 in the paper)

- ⋆ Calibration of RCS takes around 3 minutes for $M = 1,000$.
  (and around 5 hours with $M = 100,000$ and 50 parallel workers)

4. Conclusion

Main Messages from our Paper

- ⋆ We studied the properties of projection inference for SVARs.
  (delivers frequentist and Robust Bayes interpretation)
- ⋆ We have emphasized the generality of projection inference.
  (can handle typical applications in applied macro work)
- ⋆ We thought seriously about computational feasibility.
  (implementation requires solving two mathematical programs)
- ⋆ We showed how to calibrate the RB credibility of projection.
  (which also calibrates coverage under some reg. conditions)

Bottom Line

We think that projection is a simple way to conduct inference in set-id SVARs.

(it has both frequentist coverage and Robust Bayes credibility)
(it is also general, feasible, and delivers simultaneous inference)

Thanks very much for listening!