Ambiguous Estimation

Raffaella Giacomini, Toru Kitagawa, Harald Uhlig

UCL, UCL, Chicago

New York, June 2017

What we do in a nutshell

- We consider Bayesian estimation under ambiguity ($=$ multiple priors)
- Useful tool for sensitivity analysis in set-identified models

Motivation - the rise of set identification in econometrics

- Choosing identifying assumptions is a crucial and controversial step in causal studies

Motivation - the rise of set identification in econometrics

- Choosing identifying assumptions is a crucial and controversial step in causal studies
- A set of assumptions that everyone can agree on often only set-identifies the object of interest
    - **Impulse response analysis in Structural VARs** with sign restrictions (Uhlig 05 etc) or partial causal ordering
    - **Missing data** with unknown missing mechanism (Manski 89)
    - **Treatment effect model** with monotone selection/instrument (Manski & Pepper 00)
    - **Multiple-equilibria model** with no assumption about equilibrium selection

Motivation - the rise of set identification in econometrics

- Choosing identifying assumptions is a crucial and controversial step in causal studies
- A set of assumptions that everyone can agree on often only set-identifies the object of interest
  - **Impulse response analysis in Structural VARs** with sign restrictions (Uhlig 05 etc) or partial causal ordering
  - **Missing data** with unknown missing mechanism (Manski 89)
  - **Treatment effect model** with monotone selection/instrument (Manski & Pepper 00)
  - **Multiple-equilibria model** with no assumption about equilibrium selection
- Focus here is on estimation and inference about impulse response functions (IRF) in set-identified Structural VARs

Set identification in IRF analysis

- Current approaches:
  - Frequentist estimation and inference for IRF's identified set (estimator is a set)
  - Bayesian estimation and inference for IRF with a single prior (estimator is a point even though parameter is set identified)
  - Robust Bayesian inference (Giacomini & Kitagawa (15), henceforth GK) with multiple priors = estimation and Bayesian inference for the identified set (estimator is a set)
- Vast majority of empirical applications use single-prior Bayesian approach

Drawbacks of single prior Bayesian approach

- Single prior approach: Specifying credible prior not always possible. Because of set identification, limited credibility of the prior affects posterior inference even asymptotically (Poirier, 1989)

Drawbacks of multiple prior Bayesian approach

- GK's multiple priors: Inference not affected by prior choice, but set of priors may contain unrealistic ones, e.g, point mass on extreme scenarios
- For SVARs, it delivers estimates of the identified set, however
  - Set estimates often too wide to say anything meaningful about effect of shocks
  - Empirical researchers may like reporting point estimates of IRFs

This paper

Our proposal is to bridge the gap between the two approaches:
Formally, we consider minimax estimation with a refined set of priors

This paper

Our proposal is to bridge the gap between the two approaches:
Formally, we consider minimax estimation with a refined set of priors

- Consider a set of priors (ambiguous belief) but restrict attention to a Kullback-Leibler (KL) neighborhood of a benchmark prior
- Can report the posterior mean bounds for the corresponding KL-neighborhood set of posteriors
- Can also report minmax estimator that minimizes the worst-case posterior risk over the KL neighborhood

This paper

Our proposal is to bridge the gap between the two approaches:
Formally, we consider minimax estimation with a refined set of priors

- Consider a set of priors (ambiguous belief) but restrict attention to a Kullback-Leibler (KL) neighborhood of a benchmark prior
- Can report the posterior mean bounds for the corresponding KL-neighborhood set of posteriors
- Can also report minmax estimator that minimizes the worst-case posterior risk over the KL neighborhood
- Similar idea to the robust control literature of Hansen & Sargent (01), here applied to estimation

How is this useful?

- Sensitivity analysis: perturb the single prior adopted by current literature, to see how much the posterior conclusions change

How is this useful?

- Sensitivity analysis: perturb the single prior adopted by current literature, to see how much the posterior conclusions change
- Tightening estimates of identified sets: posterior mean bounds of KL-posteriors class can be interpreted as an "identified set" tightened up by the non-dogmatic assumptions contained in the benchmark prior

Related literature

- **Bayesian approach to set identification**: Poirier (97), Moon & Schorfheide (12), Kline & Tamer (16), among others
- **Robust Bayes/sensitivity analysis in econometrics/statistics**: Huber (73), Chamberlain & Leamer (76), Manski (81), Leamer (82), Berger (85), Berger & Berliner (86), Wasserman (90), Chamberlain (00), Geweke (05), Müller (12)
- **Decision under Ambiguity**: Schmeidler (89), Gilboa & Schmeidler (89), Maccheroni, Marinacci & Rustichini (06), Hansen & Sargent (01), Strzalecki (11), among others

Illustrative example

- Simultaneous equation model of labor demand and supply:

$$A\begin{pmatrix}\Delta n_t\\ \Delta w_t\end{pmatrix}=\begin{pmatrix}\epsilon_t^d\\ \epsilon_t^s\end{pmatrix},\quad A=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$$

$(\Delta n_t, \Delta w_t)$: employment and wages growth, $-a_{12}/a_{11}$: labor demand elasticity, $-a_{22}/a_{21}$: labor supply elasticity. Structural shocks $(\epsilon_t^d, \epsilon_t^s) \sim \mathcal{N}(0, I)$

Illustrative example

- Simultaneous equation model of labor demand and supply:

$$A \begin{pmatrix} \Delta n_t \\ \Delta w_t \end{pmatrix} = \begin{pmatrix} \epsilon_t^d \\ \epsilon_t^s \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

  $(\Delta n_t, \Delta w_t)$: employment and wages growth, $-a_{12}/a_{11}$: labor demand elasticity, $-a_{22}/a_{21}$: labor supply elasticity. Structural shocks $(\epsilon_t^d, \epsilon_t^s) \sim \mathcal{N}(0, I)$

- Structural parameters: A

- Parameter of interest: impulse response of Δn_t to ϵ_t^d, $\alpha \equiv$(1,1)-element of A^{-1}

- Reduced-form parameters: Cholesky decomposition of the Var of $(\Delta n_t, \Delta w_t)$, $\phi \equiv vec(\Sigma_{tr}) = (\sigma_{11}, \sigma_{21}, \sigma_{22}) \in \Phi$

Single prior Bayesian estimation

- Bayesian estimation requires specifying priors for all parameters

Single prior Bayesian estimation

- Bayesian estimation requires specifying priors for all parameters
- The reduced form parameter: $\phi = vec(\Sigma_{tr})$

Single prior Bayesian estimation

- Bayesian estimation requires specifying priors for all parameters
- The reduced form parameter: $\phi = vec(\Sigma_{tr})$
- Without enough identifying assumptions, ϕ does not uniquely pin down the structural parameter A
- Instead we have a set of observationally equivalent A: $(A = Q'\Sigma_{tr}^{-1})$, Q orthonormal ("rotation") matrix (the mapping is one-to-many)

Single prior Bayesian estimation

- Bayesian estimation requires specifying priors for all parameters
- The reduced form parameter: $\phi = vec(\Sigma_{tr})$
- Without enough identifying assumptions, ϕ does not uniquely pin down the structural parameter A
- Instead we have a set of observationally equivalent A: $(A = Q'\Sigma_{tr}^{-1})$, Q orthonormal ("rotation") matrix (the mapping is one-to-many)
- Imposing identifying assumptions can be viewed as progressively restricting the space where Q lies until there is a unique Q
- Q is therefore another parameter

Single prior Bayesian estimation

- Bayesian estimation requires specifying priors for all parameters
- The reduced form parameter: $\phi = vec(\Sigma_{tr})$
- Without enough identifying assumptions, ϕ does not uniquely pin down the structural parameter A
- Instead we have a set of observationally equivalent A: $(A = Q'\Sigma_{tr}^{-1})$, Q orthonormal ("rotation") matrix (the mapping is one-to-many)
- Imposing identifying assumptions can be viewed as progressively restricting the space where Q lies until there is a unique Q
- Q is therefore another parameter
- ϕ and Q are fundamentally different: the data is informative about ϕ but not about Q

Single prior Bayesian estimation

- The vast majority of the empirical macro literature adopts the following (single-prior) Bayesian approach:

Single prior Bayesian estimation

- The vast majority of the empirical macro literature adopts the following (single-prior) Bayesian approach:
  - Specify a prior for ϕ and obtain its posterior by combining it with the likelihood for the reduced form model
  - Specify a uniform prior for Q
  - Draw ϕ from its posterior and Q from its prior, compute A and retain only draws that satisfy the identifying restrictions
  - Report the posterior for A (or for the impulse response α)

Single prior Bayesian estimation

- An active literature in macroeconometrics shows that this causes problems (Moon and Schorfheide, 12, Baumeister and Hamilton, 15)

Single prior Bayesian estimation

- An active literature in macroeconometrics shows that this causes problems (Moon and Schorfheide, 12, Baumeister and Hamilton, 15)
- Essentially, you think you are being uninformative by using a uniform prior for Q but you are introducing spurious and non-transparent prior information on the object of interest (A or the impulse response α)

Single prior Bayesian estimation

- An active literature in macroeconometrics shows that this causes problems (Moon and Schorfheide, 12, Baumeister and Hamilton, 15)
- Essentially, you think you are being uninformative by using a uniform prior for Q but you are introducing spurious and non-transparent prior information on the object of interest (A or the impulse response α)
- In practice, this means that your conclusions on the effectiveness of policy can be driven by spurious prior information

Single prior Bayesian estimation

- An active literature in macroeconometrics shows that this causes problems (Moon and Schorfheide, 12, Baumeister and Hamilton, 15)
- Essentially, you think you are being uninformative by using a uniform prior for Q but you are introducing spurious and non-transparent prior information on the object of interest (A or the impulse response α)
- In practice, this means that your conclusions on the effectiveness of policy can be driven by spurious prior information
- This is a problem under set identification because the effect of the prior does not disappear asymptotically as for point identification (Poirier, 89)

GK solution: multiple priors

- Instead of specifying a uniform prior for Q, allow for multiple priors for Q ("ambiguous belief" = lack of prior knowledge about Q)

GK solution: multiple priors

- Instead of specifying a uniform prior for Q, allow for multiple priors for Q ("ambiguous belief" = lack of prior knowledge about Q)
- This is computationally feasible by using an application of linear programming that gives "posterior mean bounds" and an associated credible region

GK solution: multiple priors

- Instead of specifying a uniform prior for Q, allow for multiple priors for Q ("ambiguous belief" = lack of prior knowledge about Q)
- This is computationally feasible by using an application of linear programming that gives "posterior mean bounds" and an associated credible region
- The posterior mean bounds are a (consistent) estimator of the impulse response identified set

Formalization

- **Single prior Bayesian approach**: Write a prior for $\theta = vec(A)$ (Baumeister & Hamilton (15)), or write a prior for $\phi = vec(\Sigma_{tr})$ + uninformative prior for Q (Uhlig (05))
- Report posterior mean for IRF

Formalization

- **GK's multiple prior Bayesian approach**: Write a single prior for ϕ + multiple priors for $\pi_{\theta|\phi}$'s,

$$\Pi_\theta = \left\{ \pi_\theta = \int_\Phi \pi_{\theta|\phi} d\pi_\phi : \pi_{\theta|\phi}(IS_\theta(\phi)) = 1 \ \forall \phi \right\}$$

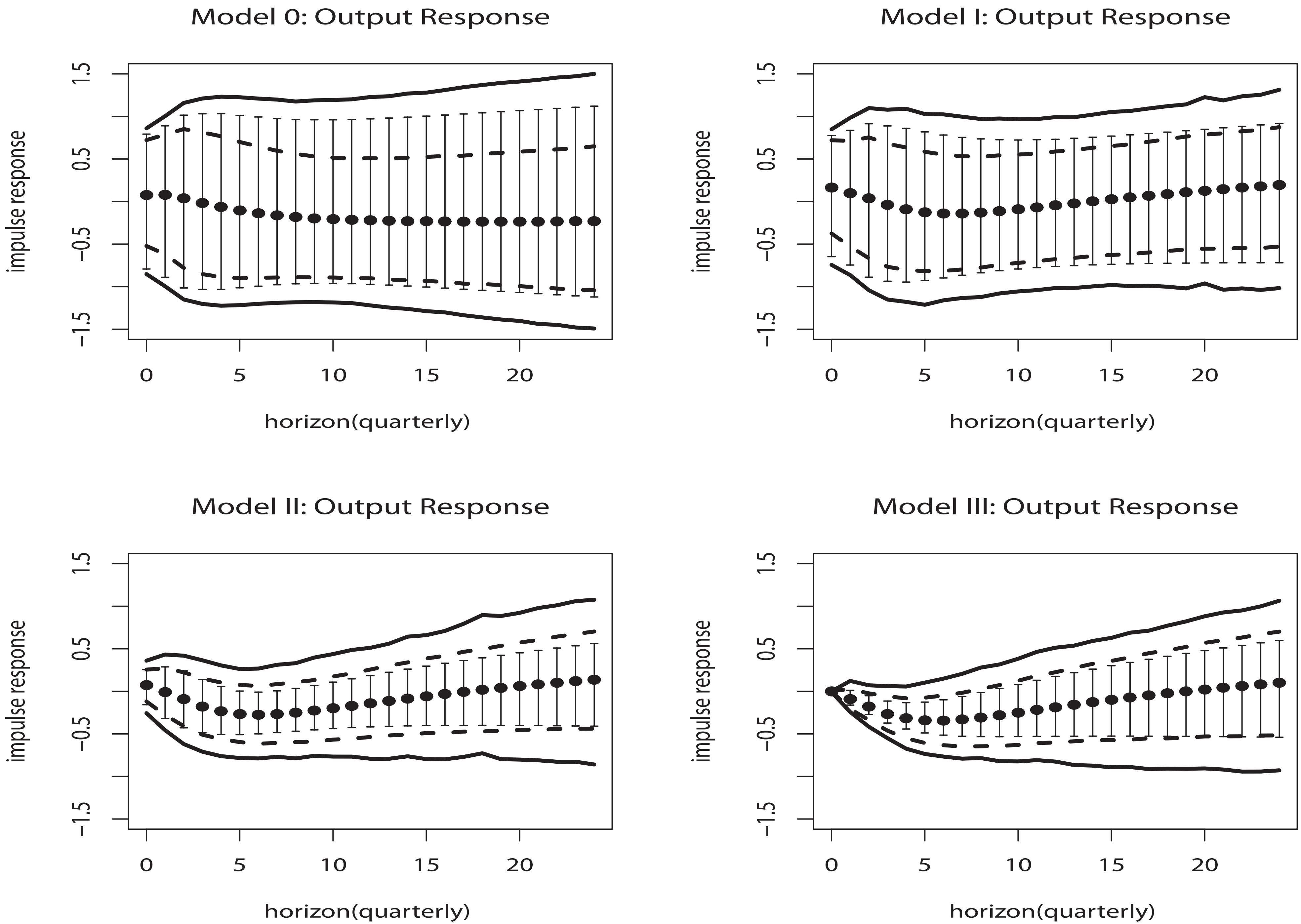
Formalization

- **GK's multiple prior Bayesian approach**: Write a single prior for ϕ + multiple priors for $\pi_{\theta|\phi}$'s,

$$\Pi_\theta = \left\{ \pi_\theta = \int_\Phi \pi_{\theta|\phi} d\pi_\phi : \pi_{\theta|\phi}(IS_\theta(\phi)) = 1 \ \forall \phi \right\}$$

- Obtain corresponding set of posteriors applying Bayes' rule to each prior in the KL class (and then marginalize them for α)

- Report posterior mean bounds (estimator of IRF identified set)



Model 0: Output Response
Model I: Output Response
Model II: Output Response
Model III: Output Response
impulse response
horizon(quarterly)
1.5
0.5
−0.5
−1.5
0
5
10
15
20

This paper's approach

Idea: Consider a set of priors in a neighborhood of the benchmark prior with radius λ,

$$\Pi_\theta^\lambda = \left\{ \pi_\theta = \int_\Phi \pi_{\theta|\phi} d\pi_\phi : \pi_{\theta|\phi} \in \Pi^\lambda(\pi^*_{\theta|\phi}) \right\},$$

where $\pi^*_{\theta|\phi}$ is a benchmark (conditional) prior,

$$\Pi^\lambda(\pi^*_{\theta|\phi}) = \left\{ \pi_{\theta|\phi} : \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) \leq \lambda \right\},$$

$\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) = \int_\Theta \ln \left(\frac{d\pi_{\theta|\phi}}{d\pi^*_{\theta|\phi}} \right) d\pi_{\theta|\phi}$: KL-distance

This paper's approach

Idea: Consider a set of priors in a neighborhood of the benchmark prior with radius λ,

$$\Pi_\theta^\lambda = \left\{ \pi_\theta = \int_\Phi \pi_{\theta|\phi} d\pi_\phi : \pi_{\theta|\phi} \in \Pi^\lambda(\pi^*_{\theta|\phi}) \right\},$$

where $\pi^*_{\theta|\phi}$ is a benchmark (conditional) prior,

$$\Pi^\lambda(\pi^*_{\theta|\phi}) = \left\{ \pi_{\theta|\phi} : \mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) \leq \lambda \right\},$$

$\mathcal{R}(\pi_{\theta|\phi} \| \pi^*_{\theta|\phi}) = \int_\Theta \ln \left(\frac{d\pi_{\theta|\phi}}{d\pi^*_{\theta|\phi}} \right) d\pi_{\theta|\phi}$: KL-distance

- Nice things about the KL-neighborhood set of priors:
    - Rules out point-mass extreme beliefs
    - Invariance to reparametrization and marginalization

Invariance to marginalization

A neighborhood around $\pi^*_{\alpha|\phi}$ or around $\pi^*_{\theta|\phi}$?

Invariance to marginalization

A neighborhood around $\pi^*_{\alpha|\phi}$ or around $\pi^*_{\theta|\phi}$?

Marginalization lemma: Let $\Pi^\lambda(\pi^*_{\alpha|\phi})$ be the KL-neighborhood of the α-marginal of $\pi^*_{\theta|\phi}$. Then, working with $\Pi^\lambda(\pi^*_{\theta|\phi})$ or $\Pi^\lambda(\pi^*_{\alpha|\phi})$ leads to the same range of posteriors for α

Robust Bayes estimator

Statistical decision under the multiple posteriors $\Pi^{\lambda}_{\alpha|X}$?

Robust Bayes estimator

Statistical decision under the multiple posteriors $\Pi^{\lambda}_{\alpha|X}$?

- Let $\delta(x)$ be a decision function and $h(\delta(x), \alpha)$ be the loss function, e.g., $h(\delta(x), \alpha) = (\delta(x) - \alpha)^2$

- Posterior minimax decision:

$$\min_{\delta(x)} \int_{\Phi} \left[\max_{\pi_{\alpha|\phi} \in \Pi^{\lambda}(\pi^{*}_{\alpha|\phi})} \int_{IS_{\alpha}(\phi)} h(\delta(x), \alpha) d\pi_{\alpha|\phi} \right] d\pi_{\phi|X}$$

  In the empirical application below, we compute a point estimator for α for quadratic and absolute error loss

Empirical example

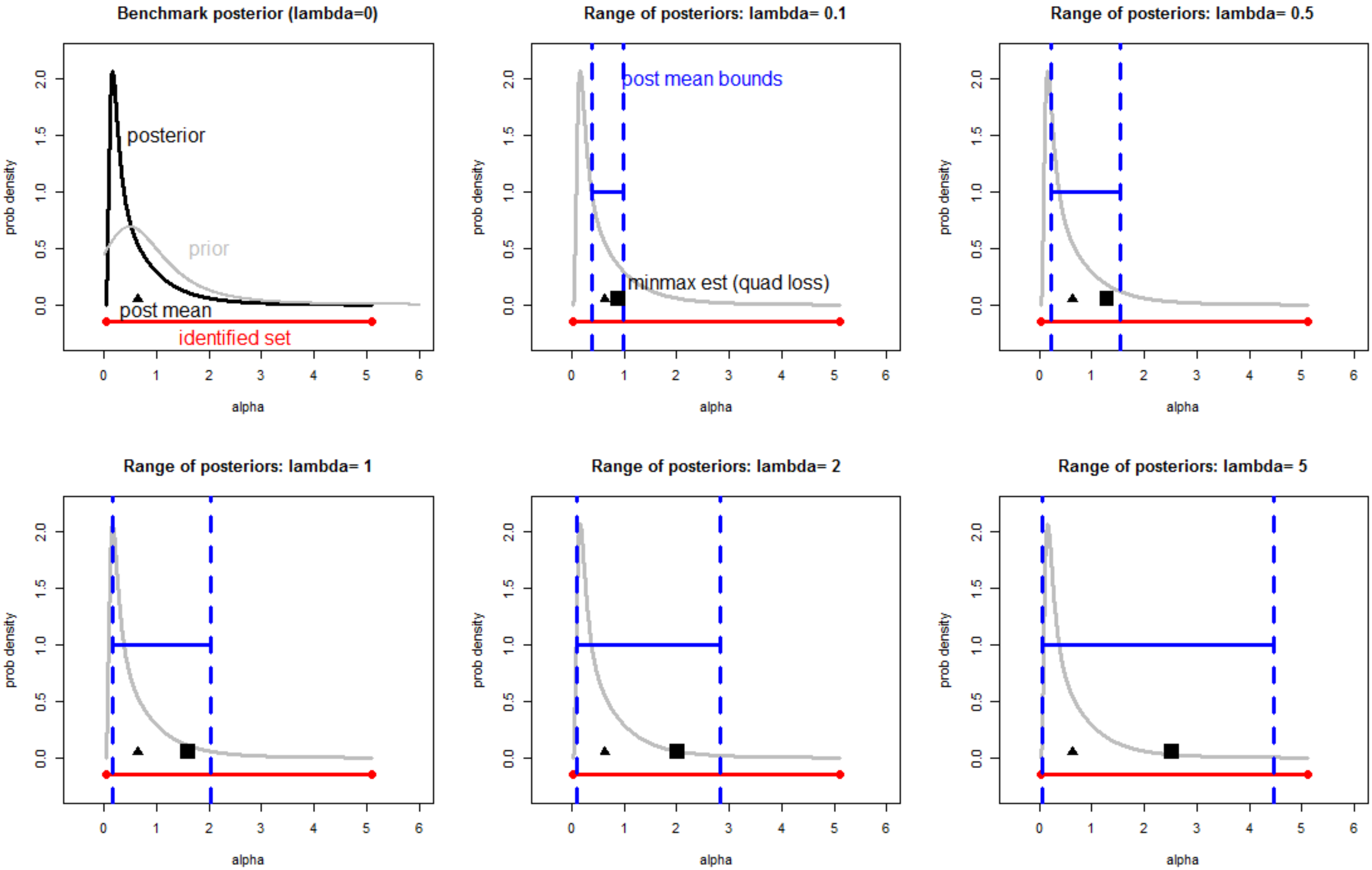
- Baumeister & Hamilton (15) two-variable SVAR(8), quarterly data for 1970 - 2014

$$A_0 \begin{pmatrix} \Delta n_t \\ \Delta w_t \end{pmatrix} = c + \sum_{k=1}^{8} A_k \begin{pmatrix} \Delta n_{t-k} \\ \Delta w_{t-k} \end{pmatrix} + \begin{pmatrix} \epsilon_t^d \\ \epsilon_t^s \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & -\beta \\ 1 & -\alpha \end{pmatrix}, \; (\epsilon_t^d, \epsilon_t^s) \sim \mathcal{N}(0, diag(d_1, d_2))$$

Empirical example

- Set Baumeister & Hamilton (15) prior as a benchmark
  - $\alpha \perp \beta$, $\alpha \sim$ non-centered t-distribution truncated to be positive, $\beta \sim$ non-centered t-distribution truncated to be negative
  - $d_1 \perp d_2 | A_0$. $d_i | A_0 \sim$ *Gamma* with hyper-parameters depending on A_0
  - $(A_1, \ldots, A_8) | (A_0, d_1, d_2) \sim$Multivariate normal
- The benchmark conditional prior $\pi^*_{\alpha|\phi}$ can be obtained analytically
- Based on their meta-analysis of α, they specify the α-prior to assign 90% to $[0.2, 2.0]$

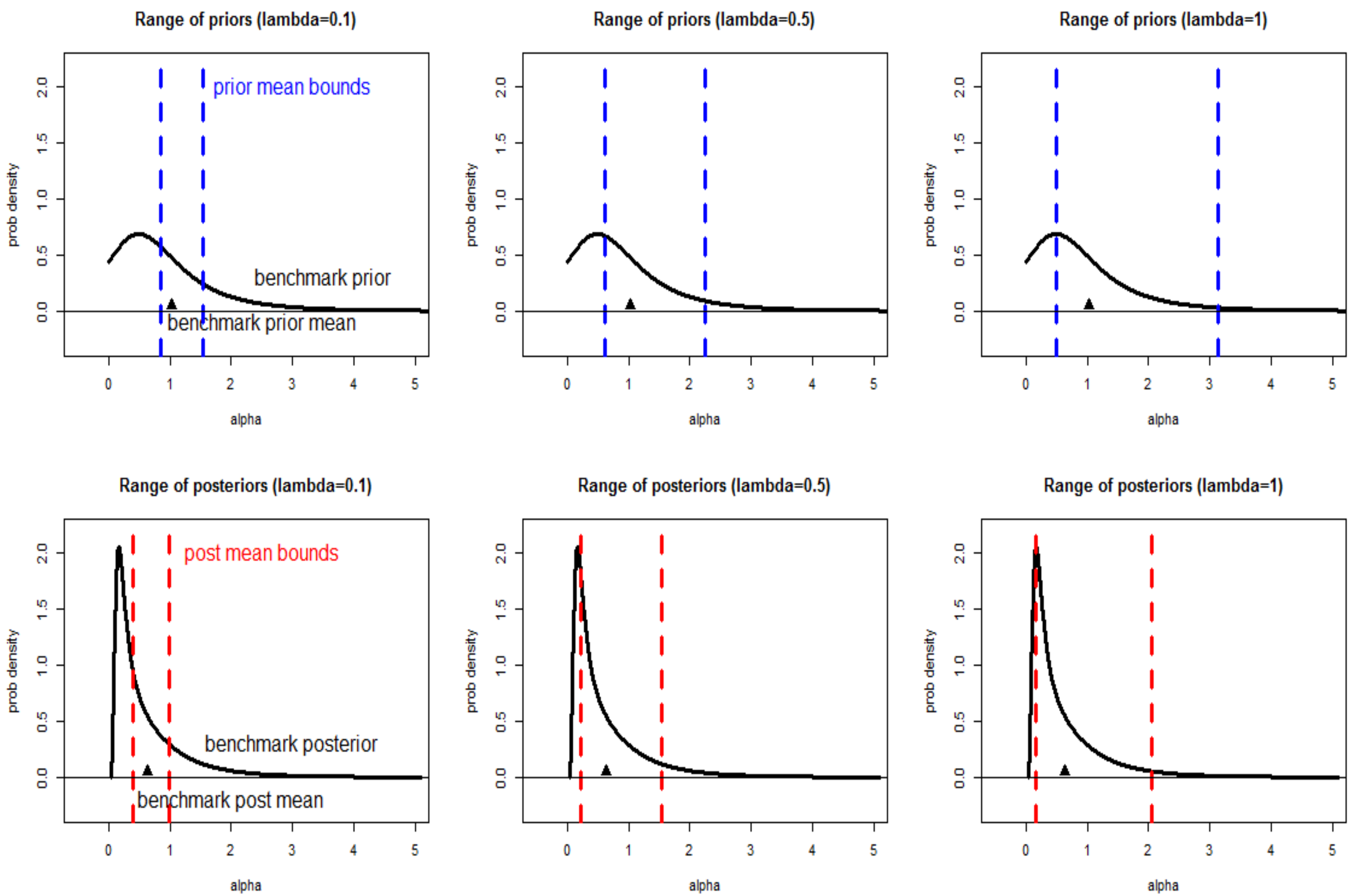


Benchmark posterior (lambda=0)
posterior
prior
post mean
identified set
Range of posteriors: lambda= 0.1
post mean bounds
minmax est (quad loss)
Range of posteriors: lambda= 0.5
Range of posteriors: lambda= 1
Range of posteriors: lambda= 2
Range of posteriors: lambda= 5
prob density
alpha

How to elicit λ?

Given a benchmark prior:

- For each candidate λ, we can compute the range of priors of the parameter of interest
- Based on the ranges of priors computed for several values of λ, choose one that best represents "vague" prior knowledge
- Report both the range of priors and the range of posteriors



Range of priors (lambda=0.1)
prior mean bounds
benchmark prior
benchmark prior mean
Range of priors (lambda=0.5)
Range of priors (lambda=1)
Range of posteriors (lambda=0.1)
post mean bounds
benchmark posterior
benchmark post mean
Range of posteriors (lambda=0.5)
Range of posteriors (lambda=1)
prob density
alpha

Conclusion

- Proposed a middle ground between single- and multiple-prior Bayesian inference, which addresses robustness concerns for the former and excess conservatism for the latter

- General framework can be applied to estimation or general decision problems where available data only set-identify parameters