Estimating HANK for Central Banks

Sushant Acharya, William Chen, Marco Del Negro, Keshav Dogra, Aidan Gleich
Shlok Goyal, Ethan Matlin, Donggyu Lee, Reca Sarfati, Sikata Sengupta

Bank of Canada, Federal Reserve Bank of New York, Harvard, MIT

XXV Annual Conference of the Central Bank of Chile
"Heterogeneity in Macroeconomics: Implications for Monetary Policy"

Goals of the paper

1. Describe a **set of tools** that central banks can use both to estimate HANK on a routine basis and to assess their ability to fit and forecast objects of interest
    - Forecasting comparisons (*making them feasible computationally*)
    - DSGE-VARs
2. Use these tools to kick the tires of a frontier HANK model

Kicking HANK's tires

- Bayer, Born, and Luetticke (2020, 2022; BBL) "Shocks, Frictions, and Inequality in US Business Cycles" is a riff off Smets and Wouters (2007; SW) "Shocks and Frictions in US Business Cycles," in the sense that it provides HANK's interpretation of the business cycle—and contrasts it with RANK's interpretation
- But SW "validated" their interpretation by showing that their model could forecast business cycle variables as well as if not better than reduced form models such as VARs
  - ... and this validation was a key reason why SW-style DSGE models became used in central banks

Kicking HANK's tires

- Bayer, Born, and Luetticke (2020, 2022; BBL) "Shocks, Frictions, and Inequality in US Business Cycles" is a riff off Smets and Wouters (2007; SW) "Shocks and Frictions in US Business Cycles," in the sense that it provides HANK's interpretation of the business cycle—and contrasts it with RANK's interpretation
- But SW "validated" their interpretation by showing that their model could forecast business cycle variables as well as if not better than reduced form models such as VARs
  - ... and this validation was a key reason why SW-style DSGE models became used in central banks
- How does BBL's HANK fare in terms of forecasting accuracy compared to SW?
- Does incorporating distributional data/measures of inequality help or hinder the forecast accuracy for business cycle variables?

- To be clear, even if forecasting is not the objective *per se*, assessing a model's forecasting accuracy is a way to learn about its successes and failures, and figure out where it can be improved

- Regardless of how well HANK forecasts business cycle variables, heterogeneity/inequality is important for policy makers, and RANK models are silent about it

The model: Bayer, Born, and Luetticke (2020)

- **Heterogeneous agents** version of **Smets & Wouters**
- Heterogeneity arises from i) (persistently) different productivity across workers, and ii) (randomly à la Calvo) being workers or entrepreneurs (get profits from monopolistically competitive firms)
- Agents trade *liquid bonds* (government plus borrowing households) and *illiquid capital* (illiquidity modeled à la Calvo)
  - Households borrow at a penalty rate

  → Heterogeneity in MPCs as some agents are at (or close to) the borrowing constraint (including some owning illiquid assets)

The model: Bayer, Born, and Luetticke (2020)

+ (most of) Smets & Wouters' bells and whistles:

real rigidities	*nominal rigidities*
investment adjustment costs	price stickiness
variable capital utilization	wage stickiness
	partial indexation to lagged inflation

- Key difference with SW is that the representative agent's Euler equation determining aggregate consumption is replaced by heterogeneous consumption decisions by agents, which changes the transmission mechanism of a variety of shocks (see the original HANK: Kaplan, Moll, and Violante, 2018)
- Shocks: SW shocks (tfp, mon. pol., MEI, risk premium, price and wage markup) + income risk shock (volatility of productivity) + tax level and progressivity shocks + deficit + meas. errors for non-SW observables

Estimation: Time series data

- With HANK models we have two datasets that "inform" the model's parameters:
  - Time series ($Y^{(ts)}$) and cross-sectional moments ($Y^{(m)}$)

$Y^{(ts)}$: *macro time series* ($Y^{(ts)} = y_{1:T}^{(ts)}$)

- in BBL, same observables as Smets and Wouters (2007): output, consumption, investment, and wage growth, total hours worked, inflation, and the federal funds rate, for the period 1954Q3-2015Q4
- plus: federal tax receipts, idiosyncratic income uncertainty (1983Q1-2013Q1), highest bracket of the US individual income tax (tax progressivity, 1954-2015 annual), wealth and income shares of the top 10% (1954-2014, annual)

$\Rightarrow$ Likelihood of time series (as in standard DSGE models estimation):

$$p(y_{1:T}^{(ts)}|\theta)$$

Cross-sectional and other steady state moments $Y^{(m)}$

- in BBL, a few *moments*/targets from *steady state* distribution (vector $Y^{(m)}$): Top 10 wealth share, fraction of borrowers, liquid assets/GDP, and illiquid assets/GDP

⇒ Penalty function for micro moments (quasi-likelihood of micro data):

$$log\ p(Y^{(m)}|\theta) = -\frac{1}{2}\ (log\,\bar{m}(\theta) - log Y^{(m)})'\ \Sigma_d^{-1}\ (log\,\bar{m}(\theta) - log Y^{(m)})$$

where $\bar{m}(\theta)$ are the model implied (steady state) moments

Combining time series and moments

- Posterior:

$$p(\theta|Y^{(ts)}, Y^{(m)}) \;\propto\; \underbrace{p(y^{(ts)}_{1:T}|\theta)}_{\text{Likelihood}} \;\; \underbrace{p(Y^{(m)}|\theta)\; p(\theta)}_{\text{Prior}}$$

- $p(Y^{(m)}|\theta)$ is viewed as a <u>prior</u> (Del Negro and Schorfheide, 2008, "Forming priors for DSGE models"): $p(Y^{(m)}|\theta) \propto p(\theta|Y^{(m)})$ implicitly generates a prior for all parameters affecting the steady state

- $p(\theta)$ is the "standard" prior (generally same as in Smets and Wouters) for those parameters that do not affect the steady state, or that enter the RANK version of the model

Combining time series and moments

- Posterior:

$$p(\theta|Y^{(ts)}, Y^{(m)}) \propto \underbrace{p(y_{1:T}^{(ts)}|\theta)}_{\text{Likelihood}} \; \underbrace{p(Y^{(m)}|\theta)\; p(\theta)}_{\text{Prior}}$$

- $p(Y^{(m)}|\theta)$ is viewed as a prior (Del Negro and Schorfheide, 2008, "Forming priors for DSGE models"): $p(Y^{(m)}|\theta) \propto p(\theta|Y^{(m)})$ implicitly generates a prior for all parameters affecting the steady state

- $p(\theta)$ is the "standard" prior (generally same as in Smets and Wouters) for those parameters that do not affect the steady state, or that enter the RANK version of the model

- Note, Liu and Plagborg-Møller, 2022 propose an approach involving fitting the *time series* of the *entire* cross sectional distribution

Is a trade-off between fitting macro and micro data?

- In order to formally investigate whether there is a trade-off between fitting macro and micro data one can introduce a parameter Υ_d that controls the weight on the prior $p(Y^{(m)}|\theta)$:

$$p(\theta|Y^{(ts)}, Y^{(m)}) \propto \underbrace{p(y_{1:T}^{(ts)}|\theta)}_{\text{Likelihood}} \underbrace{p(Y^{(m)}|\theta)^{\Upsilon_d} \, p(\theta)}_{\text{Prior}}$$

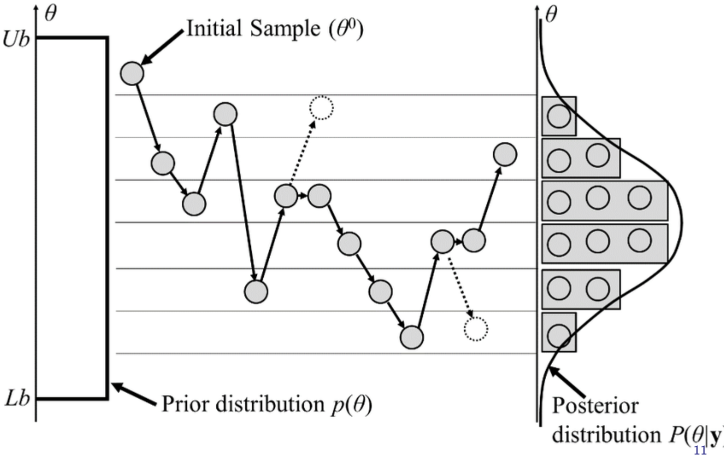
- As $\Upsilon_d \to \infty$ we force the model to meet the micro targets $Y^{(m)}$ (equivalent to $\Sigma_d \to 0$ in the penalty function)
- BBL indeed use $\Upsilon_d \to \infty$ and also use a degenerate prior (Dirac distribution) for all steady state parameters, so the steady state does not need to be recomputed when estimating the model
- But in future research (after figuring out how to deal with the computational challenge of steady steady computations) it would be interesting to investigate whether there is a trade-off

Making (repeated) estimation feasible

- The posterior $p(\theta | Y^{(ts)}, Y^{(m)})$ does not have a known form $\rightarrow$ Monte Carlo methods
- Standard approach to obtaining draws from the *posterior distribution* in DSGE estimation: *Markov Chain* Monte Carlo (Random Walk Metropolis Hastings; e.g., Dynare)

- Start with one particle θ and let it travel the posterior distribution (always accept moves "up" and only sometimes accept moves "down")

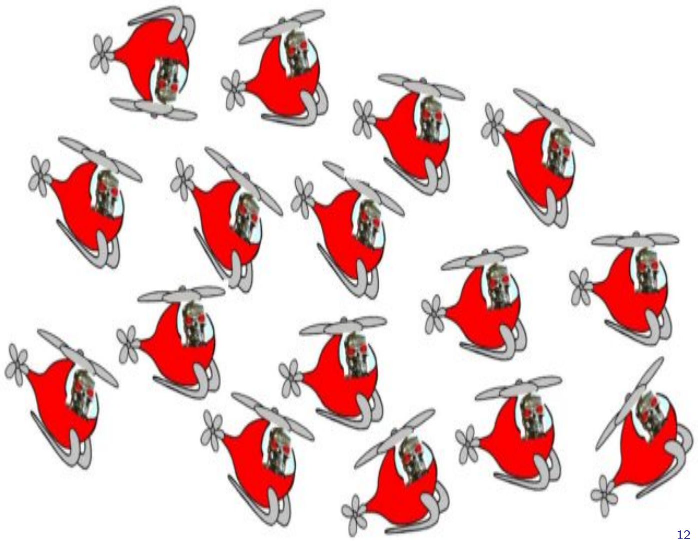
- Problem for HANK: *It is difficult to parallelize* (it's Markov!) ... and it can get stuck!



Different approach: Sequential Monte Carlo

- Relatively "new" to the DSGE estimation literature (Creal, 2007; Herbst and Schorfheide, 2014, 2015); old for the statistics literature (Gordon et al., 1993; Chopin, 2002, ...)

- Start with a swarm of particles

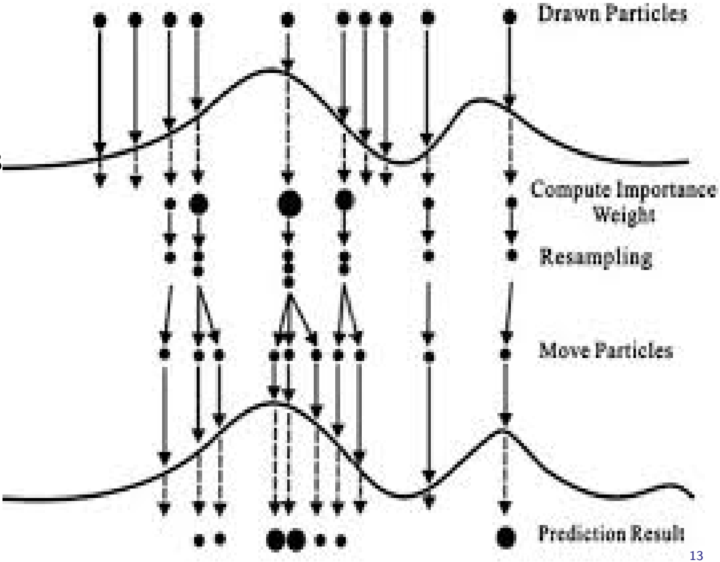


"New" approach: Sequential Monte Carlo

- "New" to the DSGE estimation literature (Creal, 2007, Herbst and Schorfheide, 2014, 2015); old for the statistics literature (Gordon et al., 1993, Chopin, 2002, ...)

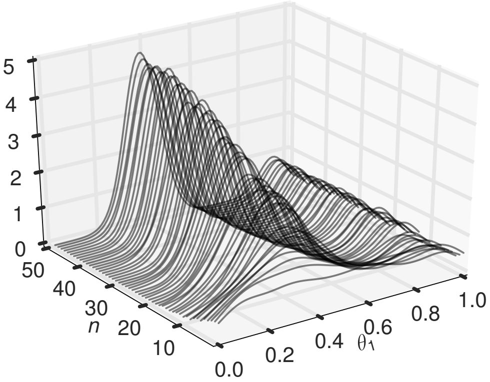
- Start with a <u>swarm</u> of particles

- ... and let them all travel and "adapt" to the posterior



SMC in a nutshell

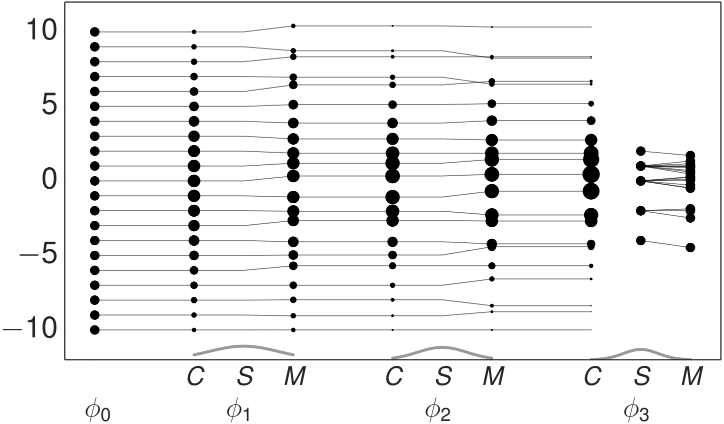
- *Sequential*/"incremental" importance sampling using *likelihood tempering*
- Importance sampling: get a bunch of draws $\{\theta^i\}_{i=1}^N$ from a proposal distribution $q(\theta)$ and compute the associated weights $W_n^i \propto \pi(\theta^i)/q(\theta^i)$
- Problem: effective sample size $ESS = N/\left(\frac{1}{N}\sum_{i=1}^N (W_n^i)^2\right) << N$ if the proposal is "bad"



$$\pi_n(\theta) \propto p(y_{1:T}|\theta)^{\phi_n} p(\theta)$$

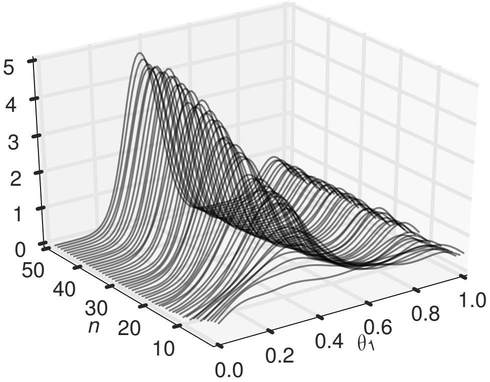
$$\phi_n = 0 \rightarrow 1$$

SMC: A graphical illustration



- $\pi_n(\theta)$ is represented by a swarm of particles $\{\theta_n^i, W_n^i\}_{i=1}^N$
- C is Correction; S is Selection; and M is Mutation.

How fast does $\phi_n \to 1$?

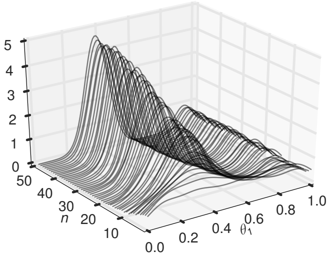


$$\pi_n(\theta) \propto p(y_{1:T}|\theta)^{\phi_n} p(\theta)$$

$$\phi_n = 0 \to 1$$

- If ϕ_n increases rapidly, ESS deteriorates quickly
- Fixed schedule (Herbst and Schorfheide, 2014): $\phi_n = \left(\frac{n}{N_\phi}\right)^\lambda, \ \lambda \sim 2$

Adaptive likelihood tempering



$$\pi_n(\theta) \propto p(y_{1:T}|\theta)^{\phi_n} p(\theta)$$

$$\phi_n = 0 \to 1$$

- Choose ϕ_n to target a desired level of ESS decrease:

$$f(\phi_n) = \widehat{ESS}(\phi_n) - \alpha \widehat{ESS}_{n-1} = 0$$

- See also Jasra et al., 2011, Del Moral et al., 2012, Schafer and Chopin, 2013, Geweke and Frischknecht, 2014, and Zhou et al., 2015

Generalized tempering/Online estimation

- The initial proposal distribution does not have to be the prior!
- **It can be some other** distribution, e.g., some other **posterior**: $\tilde{p}(\tilde{Y}|\theta)p(\theta)$

$$\pi_n(\theta) \propto p(y_{1:T}|\theta)^{\phi_n}\tilde{p}(\tilde{Y}|\theta)^{1-\phi_n}p(\theta)$$

- If it is the posterior from a shorter sample: e.g., $\tilde{p}(\tilde{Y}|\theta) = p(y_{1:T_0}|\theta),\ T_0 < T \rightarrow$ *data* tempering (but smoother!)
    - Very useful for forecasting, as **you do not have to start from scratch**
    - ... and the *adaptive* tempering (unlike in standard data tempering) assures that the particles survive

Generalized tempering/Online estimation

- The initial proposal distribution does not have to be the prior!
- **It can be some other** distribution, e.g., some other **posterior**: $\tilde{p}(\tilde{Y}|\theta)p(\theta)$

$$\pi_n(\theta) \propto p(y_{1:T}|\theta)^{\phi_n} \tilde{p}(\tilde{Y}|\theta)^{1-\phi_n} p(\theta)$$

- If it is the posterior from a shorter sample: e.g., $\tilde{p}(\tilde{Y}|\theta) = p(y_{1:T_0}|\theta)$, $T_0 < T \rightarrow$ *data* tempering (but smoother!)
  - Very useful for forecasting, as **you do not have to start from scratch**
  - ... and the *adaptive* tempering (unlike in standard data tempering) assures that the particles survive
- But it can be something else entirely, e.g., estimation obtained using a slightly different model, a different prior, a coarser solution method ...

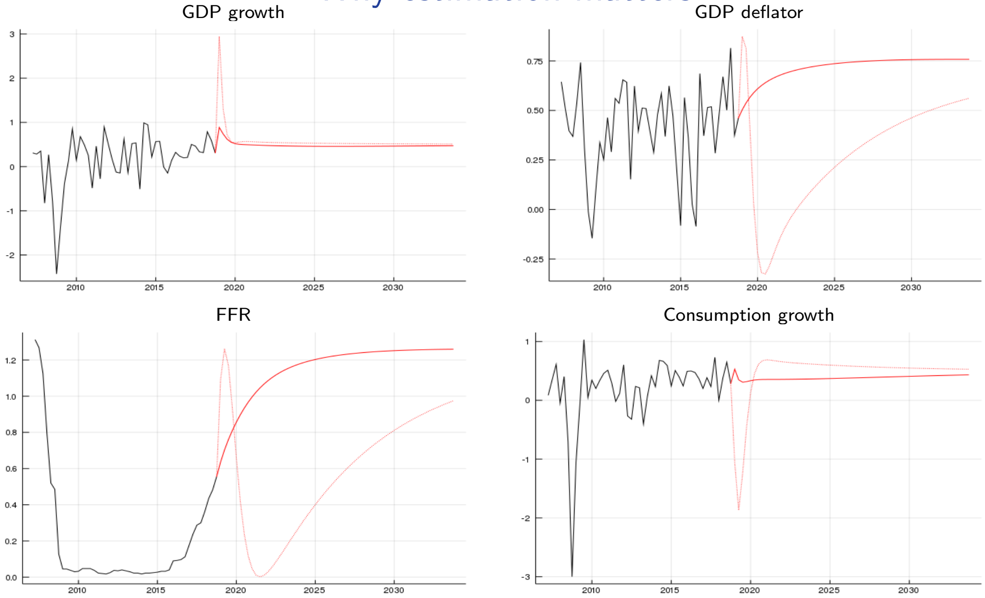
Summing up

- Reasons to use SMC for HA models in particular, and in general models whose likelihood is costly to evaluate

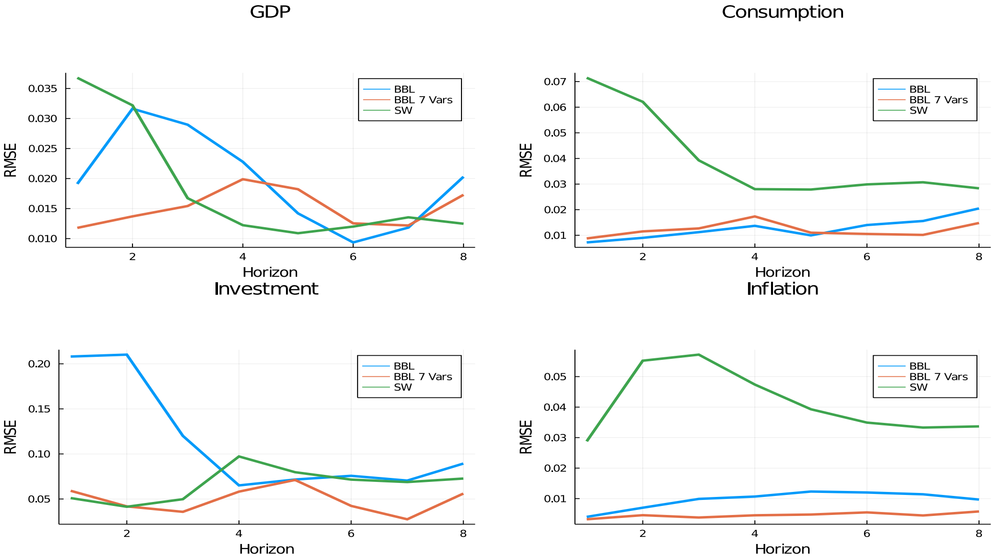
  1. *It can be parallelized*
  2. *Robust* to multimodality
  3. *Previous estimations* (swarm of particles) *can be re-used* as a bridge for new estimations ("online" estimation)
     - new data → **routine estimation** (and forecasting evaluation exercises) becomes feasible

- "Online estimation of DSGE models" Cai, Del Negro, Herbst, Matlin, Sarfati, Schorfheide, 2019; see also our blog and our Julia SMC package on GitHub

Why estimation matters



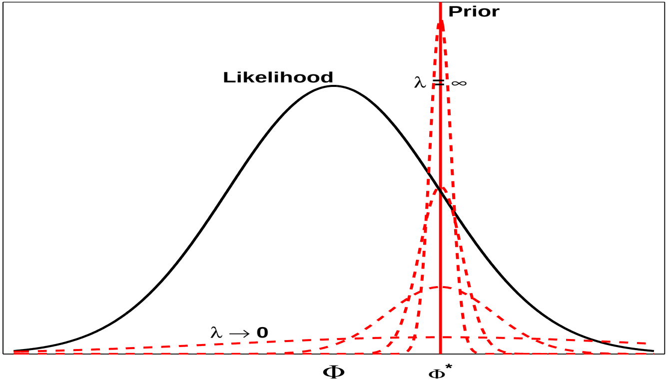
RMSEs: BBL vs SW and vs BBL w/o distributional data



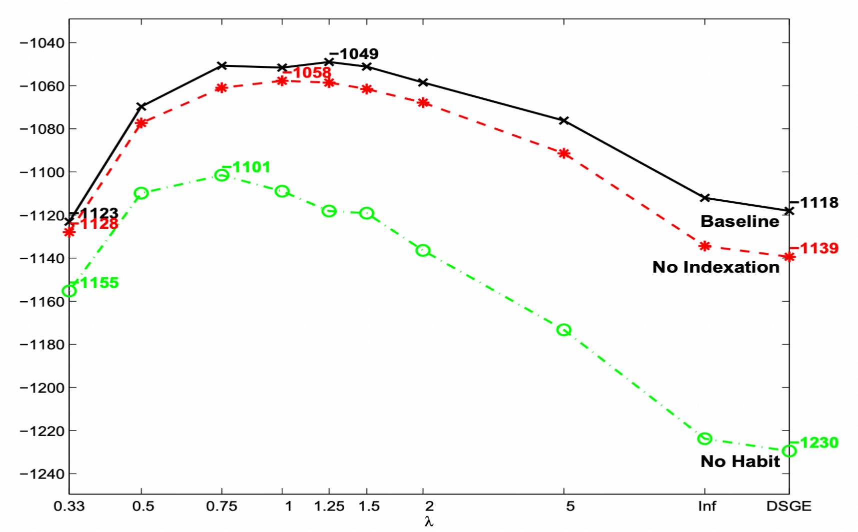
DSGE-VAR as a measure of fit

- How misspecified is my model? A linear DSGE implies restricted VARs. The question about misspecification becomes: How binding are these cross-equation restrictions?
- Think of these restrictions as a *prior* for the VAR parameters, whose tightness is controlled by the hyperparameter λ ($\lambda = \infty$ dogmatically imposes the cross-eq. restrictions).
- Then assessing misspecification boils down to the optimal choice of λ

VAR Likelihood and DSGE-based prior:



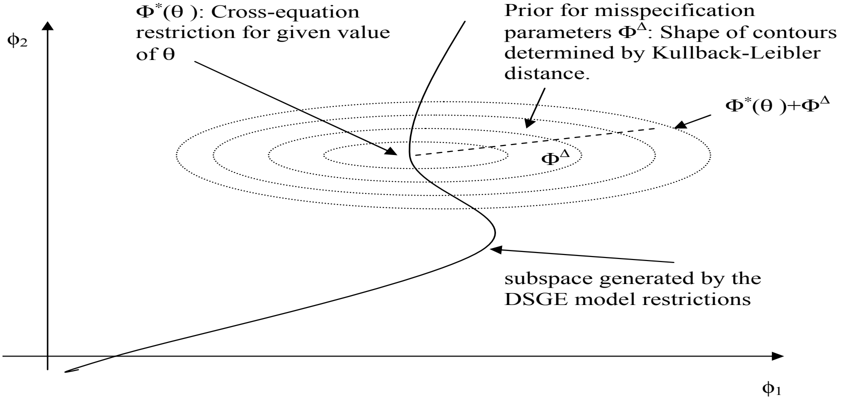
Marginal likelihood as a function of λ for the SW model



Learning about deviations from the cross-equation restrictions

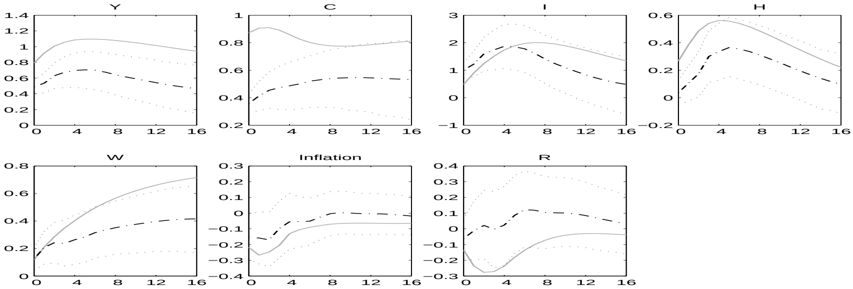
- Given the optimal λ, posterior estimates for the VAR parameters provide information on Φ^{Δ}: How dynamics in the data differ from dynamics in the model

$$\Phi = \Phi^*(\theta) + \Phi^{\Delta}$$



- Use DSGE-VAR($\hat{\lambda}$) as a benchmark: From the IRFs comparison for DSGE-VAR($\hat{\lambda}$) and DSGE we can learn about Φ^{Δ}
  - CEE have a favorite VAR, and see if the DSGE model can replicate the "VAR facts".
  - Here, we have a candidate DSGE model and ask: How much does relaxing the cross-equation restrictions change the fit and IRFs?

No Habit Formation Model Technology Shock IRFs: DSGE-VAR(∞) vs. DSGE-VAR($\hat{\lambda}$)



Conclusions

1. We described a **set of tools** that central banks can use both to estimate HANK on a routine basis and to assess their ability to fit and forecast objects of interest

    - Forecasting comparisons (*making them feasible computationally*)
    - DSGE-VARs

2. We used these tools to kick the tires of a frontier HANK model: **So far, the tires look pretty good**