Macroeconometrics - Discussion of Presentations by Ulrich Müeller and Harald Uhlig[1]

Frank Schorfheide

University of Pennsylvania, CEPR, and NBER

August 17, 2015

[1] Thanks to Paul Sangrey (Penn) for excellent research assistance.

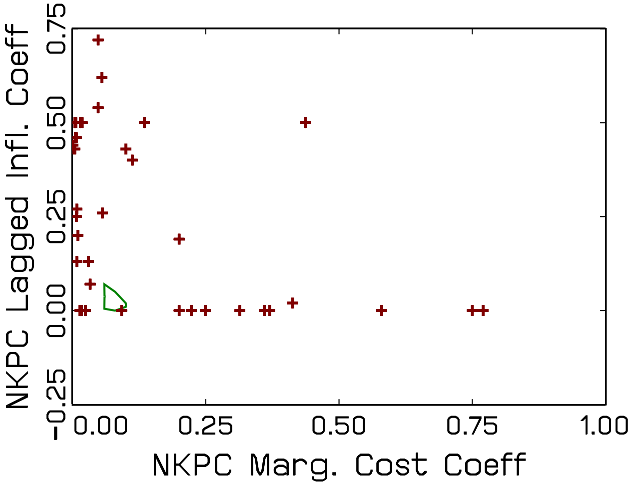
Introduction

- We try to predict:
  - GDP growth, Inflation over next five years
  - Effect of raising the federal funds rate from essentially zero to fifty basis points
- An answer could be:
  - 2 (percent);
  - between 0 and 4 (percent);
  - between -1 and 5 (percent)
- Key challenge for macroECONOMETRICIANS: accurate characterization of uncertainty.

We Are Not Always Successful - A Slide from ESWC 2010

NK Phillips Curve

$$\tilde{\pi}_t = \gamma_b \tilde{\pi}_{t-1} + \gamma_f \mathbb{E}_t[\tilde{\pi}_{t+1}] + \kappa \widetilde{MC}_t$$



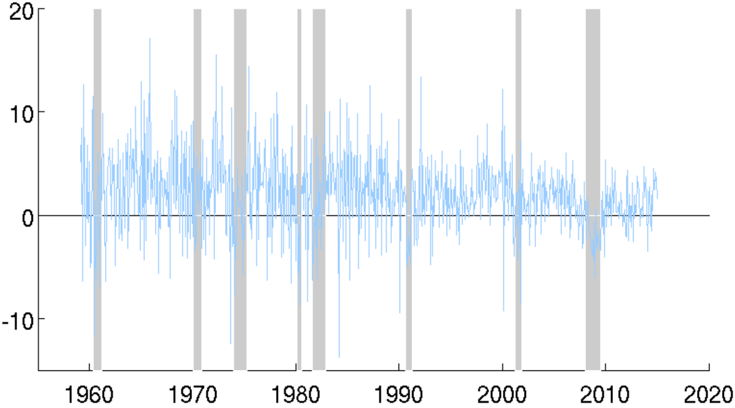
Today's Presentations

- Ulrich Müller: measures of uncertainty related to low frequency features of macroeconomic and financial data
- Harald Uhlig: uncertainty associated with predictions of policy effects derived from VARs.

Why Should We Care About Low Frequency Econometrics (MW)?

- We might be interested in long horizon forecasts.
- Agents in our models engage in long-horizon forecasting, e.g., asset pricing in Bansal-Yaron long-run risks model:
  - Exogenous cashflow processes:
$$\begin{aligned} c_{t+1} &= \mu + x_t + \sigma_{c,t}\epsilon_{c,t+1} \\ d_{t+1} &= \mu_d + \phi x_t + \pi \sigma_{c,t}\epsilon_{d,t} + \sigma_{d,t}\epsilon_{d,t+1} \end{aligned}$$
  - Common predictable component:
$$x_{t+1} = \rho x_t + \sigma_{x,t}\eta_{t+1}$$
  - How large is ρ?
- Many other reasons...

Consumption Growth Data – Can You See x_t?



Parametric Approach - See Schorfheide, Song, Yaron (2014)

- Estimate local level model using Bayesian techniques:

$$
\begin{aligned}
c_t &= \mu_c + x_t + \sigma u_t + \text{measurement errors} \\
x_t &= \rho x_t + \sqrt{1-\rho^2}(\varphi_x \sigma)\eta_t
\end{aligned}
$$

- Advantage: we get a model that fits the data at all frequencies.
- Disadvantage:
  - requires careful modeling of measurement errors;
  - data contains more info at high frequencies;
  - model extrapolates high-freq properties to low-freq properties;
  - mis-specified high-freq dynamics contaminate inference about low-freq component;
  - might understate uncertainty about low-freq dynamics.

Müller-Watson Approach - A Test Drive

- With a bit more time and effort, I could have done

$$\begin{aligned} c_t &= \mu + \frac{g}{T} x_t + \sigma \epsilon_t \\ x_t &= (1 - \delta/T) x_{t-1} + \sigma \eta_t \end{aligned}$$

and estimated $\rho = 1 - \delta/T$ as in long-run-risks literature.

- But, thus far, I only did

$$\begin{aligned} c_t &= \mu + \frac{g}{T} x_t + \sigma \epsilon_t \\ x_t &= x_{t-1} + \sigma \eta_t \end{aligned}$$

- Can't estimate ρ with that, so focus on prediction of average consumption growth:

$$\bar{c}_{T:T+H} = \frac{1}{H} \sum_{h=1}^{H} c_{T+h}$$

Müller-Watson Approach - A Test Drive

Step 1. Project data $\{c_t\}$ onto cosine functions $\cos(\pi jt/T)$, $j = 1, \ldots, q$:

- (Standardized) regression coefficients are given by

$$C_j = T^{-1/2} \sum_{t=1}^{T} \sqrt{2} \cos(\pi jt/T) c_t, \quad j = 1, \ldots, q$$

- Sample average:

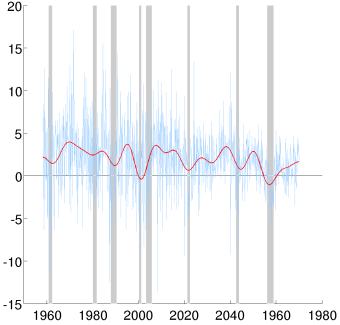
$$C_0 = T^{-1/2} \sum_{t=1}^{T} (c_t - \mu)$$

- Treat regression coefficients as low frequency data:

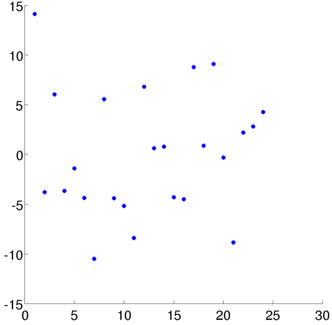
$$C_0,\ C_1,\ \ldots,\ C_q,$$

Raw versus Transformed Data

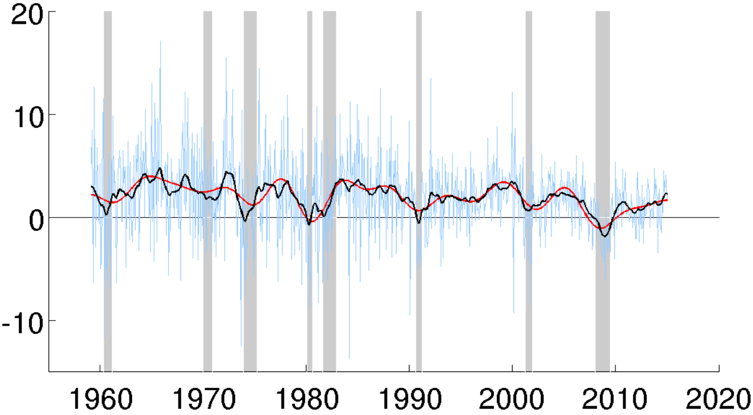
Before



After ($q = 24$)



Smoothed x_t versus Cosine Projections ($q = 24$)



Müller-Watson Approach - A Test Drive

Step 2. Use asymptotics to construct likelihood

$$C_j \implies \sigma \int_0^1 \Psi_j(s) dW_u(s) + \sigma g \int_0^1 \Psi_j(s) W_\eta(s) ds$$

$$C_0 \implies \sigma \int_0^1 dW_u(s) + \sigma g \int_0^1 W_\eta(s) ds$$

Use asymptotics to approximate distribution of forecast object:

$$\frac{1}{\lambda}\sqrt{T}(\bar{c}_{T+1:T+\lfloor \lambda T \rfloor} - \mu)$$

$$\implies \frac{\sigma}{\lambda} W_u^+(\lambda) + g\sigma W_\eta(1) + \frac{g\sigma}{\lambda} \int_0^\lambda W_\eta^+(s) ds$$

Müller-Watson Approach - A Test Drive

Step 3. These calculations imply

$$\big(C_0, C_1 \ldots, C_q, \text{Forecast}\big)\big|(\mu, g, \sigma) \Longrightarrow N(0, \sigma^2\Sigma(g))$$

Step 4. MW have developed really sophisticated inference procedures for the parameters. I simply do quasi-Bayesian inference.

- We just showed that $p\big(C_0, \ldots, C_q \big| \mu, g, \sigma\big)$ is approximately normal $\Longrightarrow$ Gaussian likelihood.
- Pick prior $p(\mu, g, \sigma)$.
- Sample from posterior

$$p\big(\mu, g, \sigma \big| C_0, \ldots, C_q\big).$$

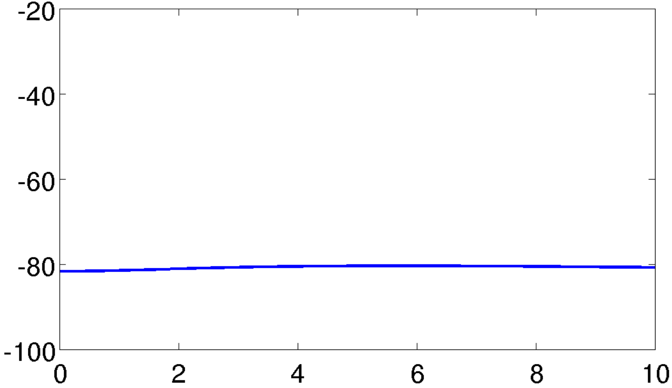
- Sample form predictive distribution

$$\begin{aligned} & p(\bar{c}|C_0, \ldots, C_q) \\ & \quad = \int p(\bar{c}|C_0, \ldots, C_q, \mu, g, \sigma) p(\mu, g, \sigma|C_0, \ldots, C_q) d(\mu, g, \sigma). \end{aligned}$$

Some Intuition

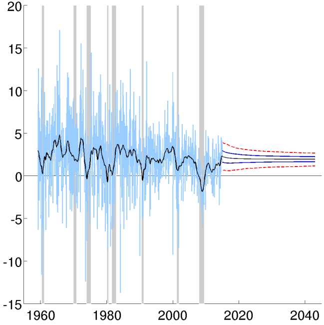
- $g^2/(1+g^2)$ is "local-level-to-noise" ratio;
- If you remove the noise from the data, you can't estimate this ratio anymore...
- Very little information about g in transformed data.
- But... for forecasting long-run means we only care about local-level process, not it's relative variance.

Estimation Results – Concentrated Likelihood for g

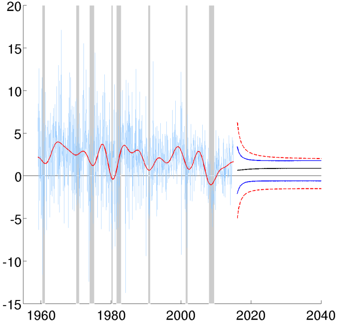


Müller-Watson Approach - A Test Drive

Parametric $\rho \approx 0.95$



Müller-Watson $\rho = 1$



Score Card – Müller-Watson

- If you have fourty years of data, you don't have much information about cycles that last at least five years... small sample inference!
- Interesting small sample inference problems.
- MW approach is appealing if you ask questions that are ONLY related to low frequency properties of the data (not for g, not for medium-run forecasting).
- User has to select a spectral band... what are the principles to do so? What if conclusions are sensitive to spectral band?
- Important to make it user-friendly:
  - derivation of likelihood can be tedious and prone to errors;
  - seperate basic idea of data transformation from complicated non-standard inference (I used basic MCMC!)
- Multivariate problems: I did not get to

$$
\begin{aligned}
c_{t+1} &= \mu + x_t + \sigma_{c,t}\epsilon_{c,t+1} \\
d_{t+1} &= \mu_d + \phi x_t + \pi\sigma_{c,t}\epsilon_{d,t} + \sigma_{d,t}\epsilon_{d,t+1}
\end{aligned}
$$

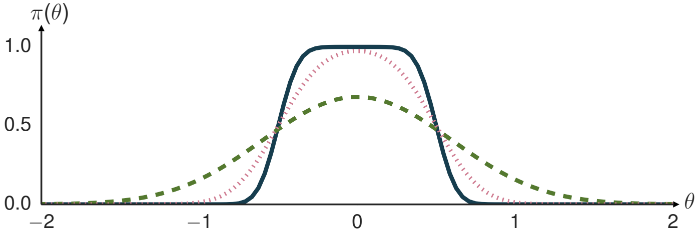
Structural VARs and Identification – Uhlig

- Structural analysis with VARs requires identification assumptions.
- Critical review of sign-restriction literature that Harald Uhlig pioneered.
- The Uhlig Principles:
  - If you know it, impose it!
  - If you do not know it, do not impose it!
- Another Principle:
  - If you are not sure, use probability distributions to characterize your state of knowledge / ignorance
- Leads to: Bayesian analysis of set (or partially) identified models

Set-Identified Models

- Simple example:
  - Reduced form parameter ϕ: reduced-form VAR coefficients
  - Structural parameter θ: structural impulse response
  - Identified set $\theta \in \Theta(\phi)$
- We can learn ϕ from the data, which determines $\Theta(\phi)$, ...
- but not the location of θ in $\Theta(\phi)$.
- Bayesian analysis:
  - Likelihood $p(Y|\phi)$
  - Prior $p(\phi)$ and $p(\theta|\phi)$
  - Note that $p(\theta|\phi)$ does not get updated!

Example



- Suppose $\Theta(\phi) = [\phi, \phi + 1]$.
- $\theta | \phi \sim U[\phi, \phi + 1]$
- $\phi | Y \sim N(\bar{\phi}, \bar{V}_\phi)$
- The figure depicts the posterior distribution $\pi(\theta)$ for $\bar{\phi} = -0.5$ and $\bar{V}_\phi$ equal to $1/4$ (dotted), $1/20$ (dashed), and $1/100$ (solid).

Principle of Good Reporting (I)

- θ is impulse response.
- Priors are typically not specified on $\theta|\phi$.
- Instead, they are specified on columns of a rotation matrix, e.g., VAR(1):

$$y_t = \Phi y_{t-1} + Aq, \quad \theta = f(\Phi, A, q) \tag{1}$$

  where $\|q\| = 1$ and $\Sigma = AA'$.
- Benchmark prior: Haar measure on q (or Ω), "uniform" under rotations.
- Leads to non-uniform prior on $\Theta(\phi)$.
- Not a big deal if you like the resulting prior, but...

Principle of Good Reporting (II)

- Yet Another Principle: if you work with set-identified models, show
  - an estimate of the identified set $\Theta(\hat{\phi})$;
  - the distribution of $p(\theta|\hat{\phi})$
- Why?
  - Audience wants to get a sense of size of identified set relative to uncertainty with respect to its location;
  - audience wants to see the prior $p(\theta|\phi)$ in relevant regions of parameter space.
- Also, be clear whether inference is on
  - the parameter θ (most of VAR literature);
  - the identified set $\Theta(\phi)$.

Computational Issues

- Rejection sampler: draw q from "uniform" distribution, reject draws that violate sign restrictions.
- "Principle 8: when you reject a lot of draws, it means you have sharp identification! Good!"
- "Principle 8a: when you reject a lot of draws, it means you are using an inefficient sampler! Bad!"
- This problem has not been fully addressed in the literature – construct a better proposal distribution for the sampler to handle "small" identified sets.

Where do Sign Restrictions Come From?

- DSGE models imply (more than) sign restrictions but typically not zero restrictions.
- We probably don't believe them literally: it is not inconceivable that prices fall in response to an expansionary monetary policy shock.
- but we might think that they should hold approximately: if prices fall, they do so initially and by a small amount.
- Restricting A or A^{-1}?
  - DSGE model view of the world: it's natural to restrict A. One thinks about how the economy responds to shocks.
  - Systems-of-equations view: it's natural to restrict A^{-1}. Equations are interpreted as monetary policy rule, aggregated demand, aggregate supply...

Sign Restrictions - Score Card

- If you have better identifying information – USE IT.
- Could reconcile / encompass results generated with more restrictive identification schemes: Good.
- More generally:
  - Literature contains hundreds of IRFs to monetary policy shocks;
  - many papers focus on qualitative instead of quantitative aspects of the IRFs (price puzzle yes or no);
  - having a meta study that aggregates quantitative results would be very useful.
  - We have a hard time measuring the effects of conventional monetary policy interventions on aggregate prices and output; claims that we have precise knowledge of the effects of unconventional monetary policy make me cringe!

In Closing

- Key challenge for macroECONOMETRICIANS: deliver tools that provide a “good” characterization of uncertainty, associated with quantitative statements.
- The papers presented in this session successfully confronted this challenge.
- Macroeconometrics is Well and Alive!
- As always – much more work can / should be done.
- To all the young folks in the audience: Do MACROeconometrics / time series analysis (instead of microeconometrics)!!!