Empirical Evaluation and Estimation of Large-Scale, Nonlinear Economic Models

Michal Andrle, RES

March 2013

Outline

Introduction to the project of empirical evaluation of FSGM/GIMF.

1. Motivation & Options
2. BSMM/SP Estimation
3. Example: MC experiment with GIMF2

Why?

Motivation:

1. Need for **system** evaluation or estimation of FSGM/GIMF
2. Explicit account of expectation dynamics
3. Challenge of joining estimation with complex prior on model's properties

   ⋮

4. Not tried before

Estimation by the Analogy Principle…

"…the **analogy principle** of estimation …proposes that population parameters are estimated by sample statistics which have the same property in the sample as the parameters do in the population"
(Goldberger, 1968)

"I have found that the analogy principle disciplines econometric research by focusing attention on estimation problems rather than on methods. …One can define an analog estimator only after one has stated the estimation problem of interest."
(Manski, 1988)

BSMM/SP Estimation Method

Bayesian **S**imulated **M**ethod of **M**oments with **S**ystem **P**riors

1. **Bayesian**
   Includes priors and boundaries on parameters to regularize the problem.
2. **Simulated**
   Uses the law of motion of the model to evaluate moments, and treat small-sample bias.
3. **Method of Moments**
   Minimizes distance between sample and model-implied statistics.
4. **System Priors**
   Introduces sophisticated economic priors on moment system properties, inducing complex interdependent joint prior distribution of parameters.

BSMM/SP: Requirements

Requirements:

1. A fully simmulable, possibly nonlinear, model is required.
2. Having some data is handy. . . (mixed frequency is fine)
3. Time. . .

Note, even the Google PageRank algorithm runs just once a month. . .

BSMM/SP: Why NOT Maximum Likelihood?

Maximum Likelihood Estimation (MLE) or Bayesian MLE is a popular and frequently used option for DSGE model estimation.

MLE is a problematic option for large, nonlinear models:

1. MLE is very non-robust to misspecification compared to GMM/SMM
2. MLE suffers from stochastic-singularity problems
3. Given the size, not feasible in non-linear form
4. In linearized model, the likelihood evaluation is very ill-conditioned and costly (large state-space)
5. Implicitly optimizes one-step ahead projection performance
6. More demanding data requirements
7. One can use ‘priors’ also with other estimation methods

BSMM/SP: Criterion function

The criterion function has three components

$$p(\theta|X, Z, \gamma) \propto Q_N(\theta|X) \times \pi_S(\theta|Z) \times \pi_M(\theta|\gamma) \tag{1}$$

1. $Q_N(\theta|X)$: exponential χ-square loss function
2. $\pi_S(\theta|Z, \gamma)$: system priors
3. $\pi_M(\theta|\gamma)$: marginal independent priors

Normalizing constant is ignored, as it is irrelevant for both optimization of the posterior mode and posterior sampling.

BSMM/SP: Marginal prior – $\pi_M(\theta|\gamma)$

The marginal-independent prior is specified by $\pi_M(\theta|\gamma)$:

- Includes prior distribution and boundaries for parameters in $\theta \in \Theta$
- The priors are specified by a hyper-parameter vector γ
- Priors are specified as independent, marginal distributions

Evaluating $\pi_M(\theta|\gamma)$ is trivial and fast...

BSMM/SP: System priors – $\pi_S(\theta|Z)$

The system priors are specified by $\pi_S(\theta|Z)$:

- Places a priori constraints on system properties of the model
- The priors are specified by a hyper-parameter vector Z
- Priors induce complex global interdependencies among parameters

System prior examples (see Andrle and Beneš, 2012):

- value of the sacrifice ratio
- duration of inflation deviation from target after a demand shock
- cumulative response of output to a productivity shock
- ...

Evaluating $\pi_{S|Z}$ amounts to solving/simulating the model.

BSMM/SP: Criterion function (a)

The χ-square criterion function is:

$$\log Q_N(\theta|X) = (\mu(X) - \mu(\theta))' W_{R,T} (\mu(X) - \mu(\theta)) \quad (2)$$
$$= g(X,\theta)' W_{R,T} g(X,\theta). \quad (3)$$

Here X is $N \times T$ matrix of sample data, $\mu(X)$ is the conditional or unconditional moments in the data, $\mu(\theta)$ is the model implied moments vector, $L \times 1$.

$W_{R,T}$ is a positive definite weighting matrix, chosen to maximize efficiency of the estimator.

BSMM/SP: Criterion function (b)

Weighting matrix is determined to get efficient criterion.

- continuous-updating of the $W_{R,T|\theta}$
- exact finite-sample distribution of $g(\theta_i)$ for T periods

Under the null-hypothesis, given θ_i, we simulate R times and select T periods to calculate the moments, obtaining their finite sample distribution.

$$W_{R,T} = \mathbb{E}_{|R}\left[(g(\theta_i) - \bar{g}(\theta_i)(g(\theta_i) - \bar{g}(\theta_i)'\right] \tag{4}$$

Thus, the moments with large short-sample variance are penalized during the estimation.

Yet, this is not in line with 'minimal econometric interpretation' of Geweke.

BSMM/SP: Parameter Estimates

What can be in the moment vector $\mu(.)$?

Absolutely anything you can compute using the model...

(i) Unconditional moments
(means, variances, correlations, higher order...)

(ii) Conditional moments (kernel estimation)
data is the limiting factor and dimensionality

BSMM/SP: Parameter Estimates

Obtaining posterior estimates of θ:

- search for posterior mode $\hat{p}(\theta|X, Z, \gamma)$ using optimization
- use Adaptive Random-Walk Metropolis algorithm to reveal the full posterior distribution

TBW:

Issues

Important choices to make:

1. Type of simulation
2. Treatment of parameters affecting steady-state
3. Number of shocks employed for the estimation
4. Specification of robust system priors

Progress have already been made on all issues!

Issues: Simulation Method

Speed $\times$ Accuracy $\times$ Nonlinearities Trade-offs:

(i) point-linearization (fast)
(ii) path-linearization (default, reasonably fast)
(iii) full nonlinear stochastic simulation (slow, convergence problems)

Path-linearization and Stochastic simulations

(i) Simulate the *full nonlinear* model for a shock of a size ε_0/U
(ii) Rescale the response of the model back by U
(iii) Sample $R \times \tau T$ matrix of shocks
(iv) Permute the response into a stochastic simulation for K selected variables

Fast for small number of shocks and high U.

Issues: Parameters affecting steady-state

Three elementary strategies used:

- use of “greedy” homotopies
- two-step estimation with a pre-fetch discretization
- pre-fetch discretization with homotopy

Default approach being tested is greedy homotopies.

Issues: Number of Shocks

Parsimony is a virtue, limits are computational:

1. Stochastic singularity is a good thing! (Andrle, 2012)
2. Demand shocks dominate business cycle frequencies (Andrle, Bruha 2013)
3. Robust estimation with persistent data (Gorodnichenko et al, 2011)
4. Small sample size due to annual frequency, identification problems
5. Trend missspecification propgates the whole spectrum (Cogley, 2001, Andrle, 2008)
6. Selection of moments/shocks obviously matters for identification (Canova and Sala, 2010)
7. Proper modeling of trends is missing from GIMF, GEM and FSGM/G20MOD

Monte Carlo Test using GIMF (a)

Tests of the method initially using GIMF:

- Integrated in GIMF-Dynare infrastructure
- Uses GIMF2 (US, RW) model
- Path-linearization simulation
- Homotopy for steady-state solution

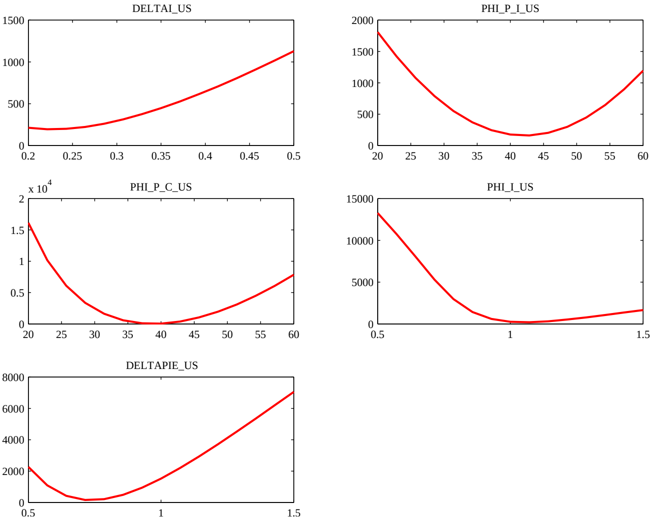
Parameter estimated using the data simulated from the model, so true values of coefficients were known:

Experiments carried out:

i) Cumulative local-identification paths
ii) Identification analysis using the Fischer Information Matrix
iii) Estimation of the posterior mode

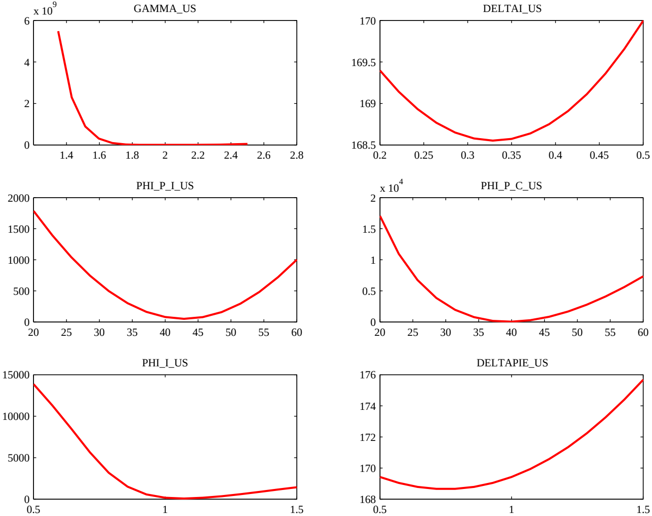
Monte Carlo Test using GIMF (b-i)

Local conditional identification (no priors)



Monte Carlo Test using GIMF (b-ii)

Local conditional identification (with priors)



Monte Carlo Test using GIMF (c)

Results:

Name	γ_{US}	$\delta_{I,US}$	$\phi_{P\_I,US}$	$\phi_{P\_C,US}$	$\phi_{I,US}$	$\delta_{\pi,US}$
True	2.000	0.330	40.00	40.00	1.000	0.750
Estimate	2.004	0.330	39.42	40.13	1.006	0.752
Std. error	0.034	0.096	0.968	0.486	0.022	0.017

Timing:

1. Non-optimized evaluation of the loss w/o sstate fun.: 6.5 sec
   (Optimized evaluation: 3.2 sec)
2. Hessian using central differences: 1347.38 sec
3. Posterior mode: 6278.72 sec, 28 iterations (with steady-state param)
4. Posterior mode: 1347.38 sec, 16 iterations (no steady-state param)
5.

Moments:

1. 17 moment statistics (unconditional)
2. vairables used: GDP, consumption, investment, core inflation, interest rates

Monte Carlo Test using GIMF (c)

Identification using the Fisher Information Matrix:
TBW