PROBABILITY MODELS FOR MONETARY POLICY DECISIONS

CHRISTOPHER A. SIMS
PRINCETON UNIVERSITY
SIMS@PRINCETON.EDU

1. Motivation and Outline

- For a 2002 Brookings paper I spoke to staff at a number of central banks and looked over written material related to their models. The idea was to describe and criticize how models are used in the monetary policy process.
- The paper was critical of the current state of policy modeling, arguing that in certain respects there has been regress, not progress, though the paper also was critical of academic econometric and macroeconomic literature for not paying attention to the problems faced by real-time policy makers.
- The reaction of some to the paper was that its positive recommendations, such as they were, were awfully vague in comparison to its negative criticisms.

2. Summary of the criticisms: Not probability models

- The models in use have completely abandoned the scientific program laid out by Haavelmo (1944), in a paper called "The Probability Approach in Econometrics".
- Haavelmo criticized the cobbling together of equation-by-equation OLS estimates and the notion, then not uncommon, that quantitative modeling could never "test" a theory.
- He argued that equation *systems* should be regarded as asserting multivariate distributions for observed data, and that once we had done this, we could confront theory with data, not merely calibrate it.

3. A crude time line of econometric policy modeling

- Tinbergen cobbles together OLS equations
- Brookings-SSRC, FRB-Penn-MIT try to follow the Haavelmo program
- People find in practice that FIML, 3SLS, 2SLS either make little difference or produce weird estimates, and are time consuming.
- People find that if equations adapt over time to past errors, in a decentralized way (equations worked on by "sectoral" experts), the system starts to get unreasonably large and to misbehave in long run projections.
- There is a reform, perhaps instigated by the Bank of Canada and its QPM, in which

Date: September 3, 2003.

(a) The model's long run is pinned down to be reasonable, and insulated from the data – either by appeal to cointegration theory or to "theory".
(b) Dynamic adjustments are estimated, almost entirely by single equation methods.
(c) There is no attempt to form a likelihood for the system or to use system methods of estimation, and no one trusts the model and its disturbances as a realistic characterization of uncertainty.

- "Econometrics" is seen as irrelevant to policy discussions about uncertainty, because it appears to be concerned with testing whether models are true, not about characterizing the kind of odds-quoting uncertainty that comes naturally in policy discussion.
- Flexible inflation targeting leads to increased interest in means to generate internally consistent quantitative projections of the effects of policy and in means to characterize uncertainty about those projections.

4. WHERE DO WE GO FROM HERE?

- Bayesian methods and language. Haavelmo swallowed Neyman-Pearson testing theory uncritically — not surprisingly, as it was frontier statistical method at the time. But this approach, which forswears ever putting probability on "parameters" or "models", integrates at best very awkwardly into a decision-theoretic framework. It also has no language for discussion of integration of uncertain prior information with the implications of data, which is fundamental to success with large economic models.
- DSGE models that aim to fit, not just tell stories. This means more shocks, more flexible dynamics, and with that more effort to characterize results and tell stories with the model.

5. EXISTING WORK IN THESE DIRECTIONS

- Statisticians and researchers from other disciplines, as well as econometricians, continue to work on MCMC computational methods.
- Smets and Wouters, Christiano et al in the work described yesterday, are working with "serious" DSGE's, aimed at fitting multivariate time series well.
- There is an interesting and enlightening branch of work (Schorfheide, 2000; Geweke, 1999) that evaluates use of "unserious" DSGE's — with too few shocks, or too many ad hoc restrictions on dynamics and functional form — from the perspective of Bayesian inference. This is well worth reading as an antidote to the notion that Bayesian methods work just fine when all the models under consideration are known to be false. But as we discover that serious DSGE's are not that hard to handle, the value of this branch of work for practical policy advice seems limited.
- Brock, Durlauf, and West (2003) have recently written on the importance for policy of characterizing uncertainty across models.

- Perturbation methods promise to let us investigate issues that depend on nonlinearity in these models, and are on the verge of being practically useful.

6. Obstacles ahead

- Our models of monetary non-neutrality are not trustworthy for welfare evaluation.
- To make DSGE models fit, we have to load them up with inertia, and then add difficult-to-interpret shocks to soak up variable-specific non-smooth randomness that the inertial mechanisms say should not exist.
- Stochastic volatility, time-varying coefficients, non-Gaussianity. These are central focuses of empirical work in finance, and they are clearly present in macro data. A really serious model can't ignore them.
- Bayesian model comparison is not as straightforward as it looks.

7. Odds Ratios and Model Averaging

- Practical experience with Bayesian approaches to handling multiple models has frequently turned out to be disappointing or bizarre. The general phenomenon of conflicts between Bayesian model comparison results and classical tests is labeled the "Lindley paradox".
- One standard applied Bayesian textbook (Gelman, Carlin, Stern, and Rubin, 1995) has no entry for "model selection" in its index, only an entry for "model selection, why we do not do it".

8. Pathologies of Bayesian model comparison

- Results are sensitive to prior distributions on parameters within each model's parameter space, even when the priors attempt to be "uninformative".
- Results can be sensitive to seemingly minor aspects of model specification.
- Results tend to be implausibly sharp, with posterior probabilities of models mostly very near zero or one.

9. Pathologies don't arise from the methodology itself

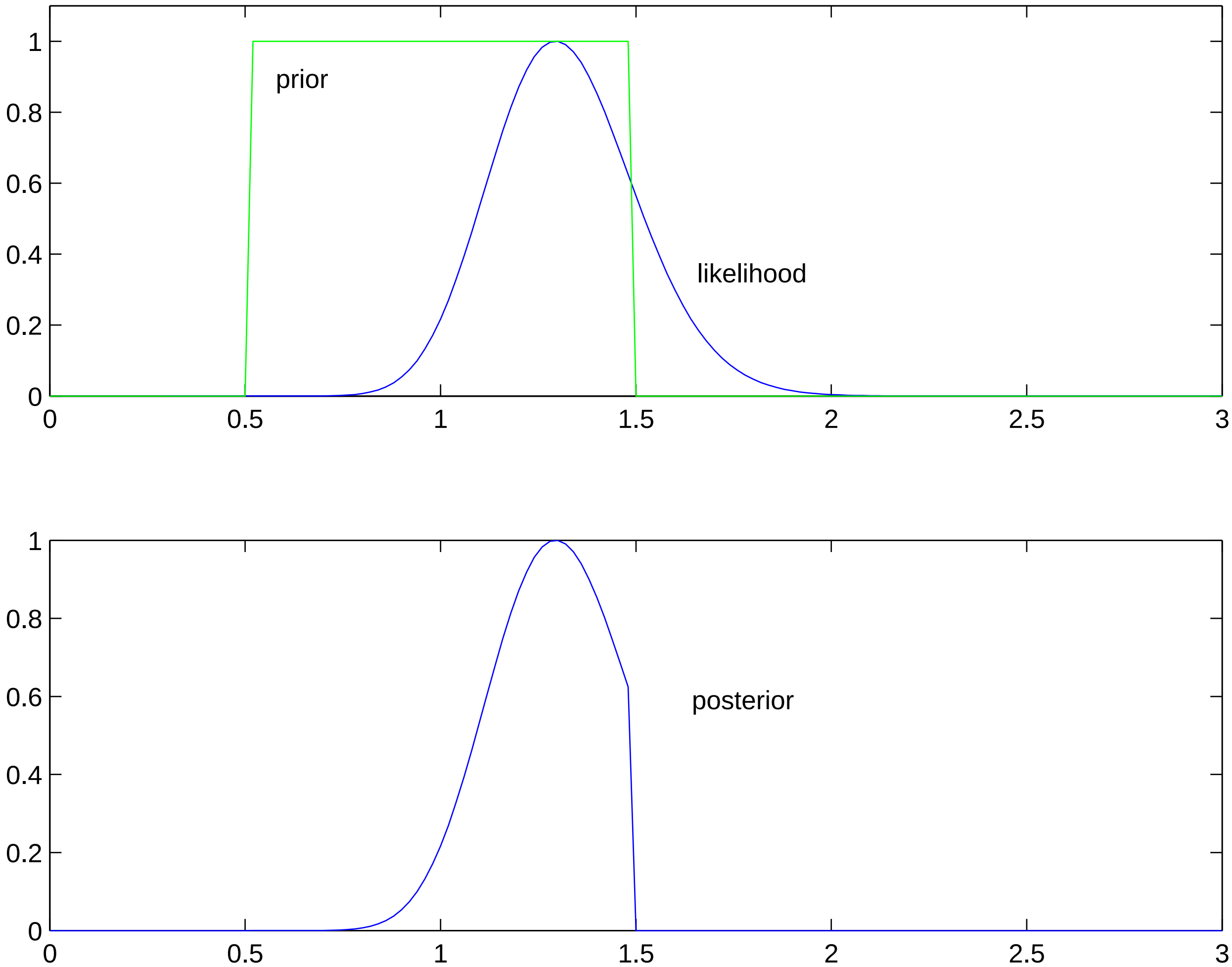
They arise from the ways we generate and interpret collections of parametric models.

Once this is understood, the model comparison methodology can be useful, but as much for guiding the process of generating and modifying our collections of models as for choosing among or weighting a given collection of models.

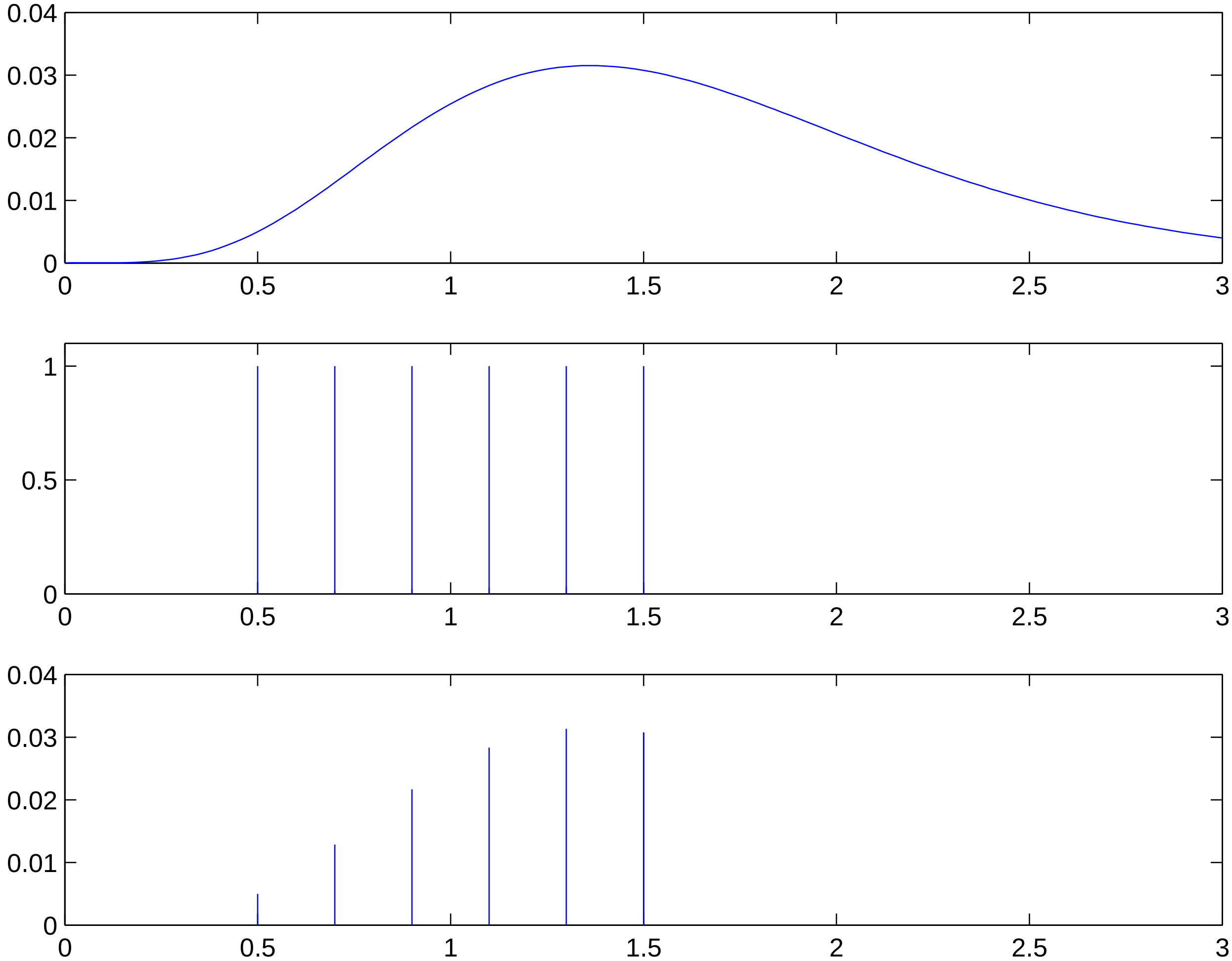
10. When the Set of Models is Too Sparse

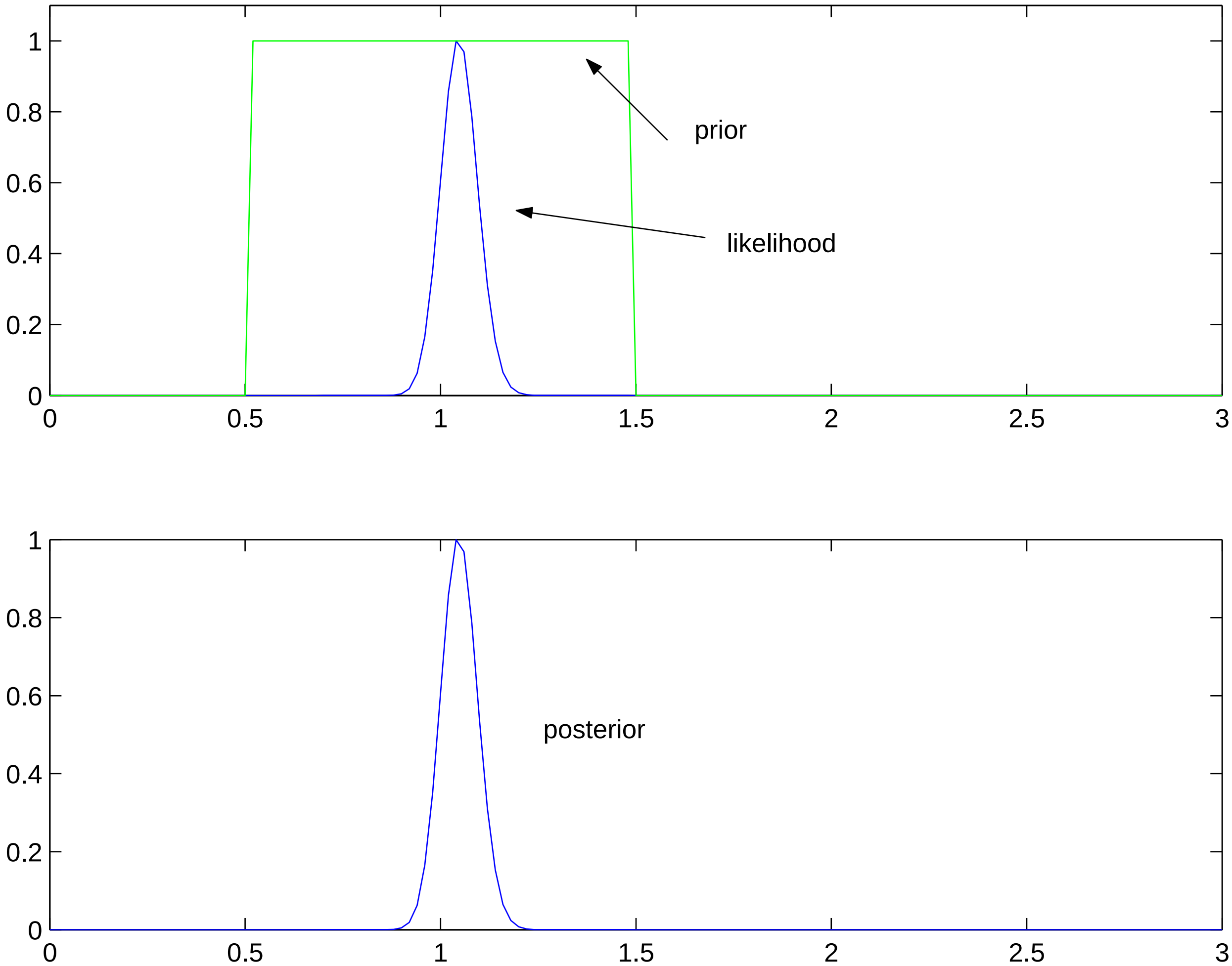
This is the kind of example that Gelman, Carlin, Stern, and Rubin (1995) have in mind when they cite, as a condition for posterior probabilities on models being "useful", the requirement that "each of the discrete models makes scientific sense, and there are no obvious scientific models in between".

11.

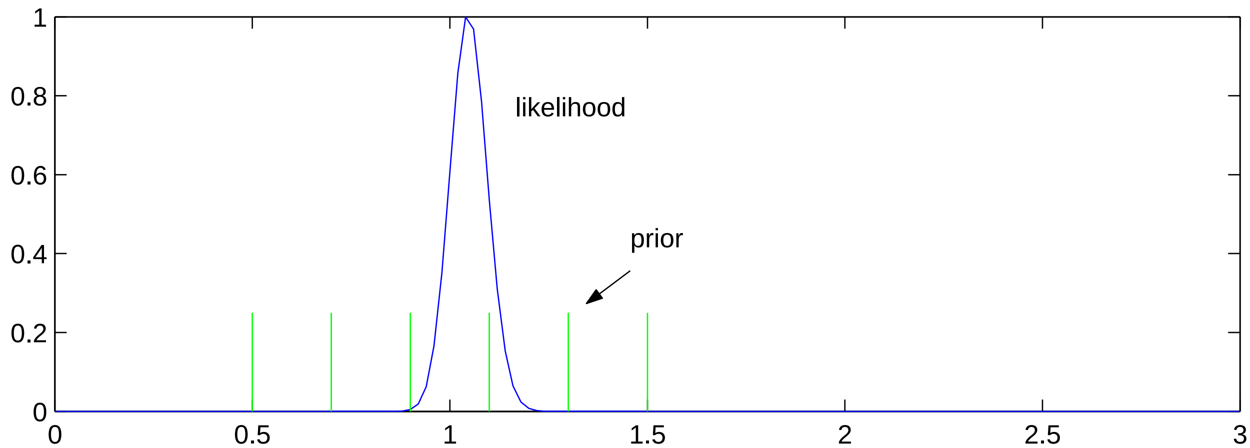


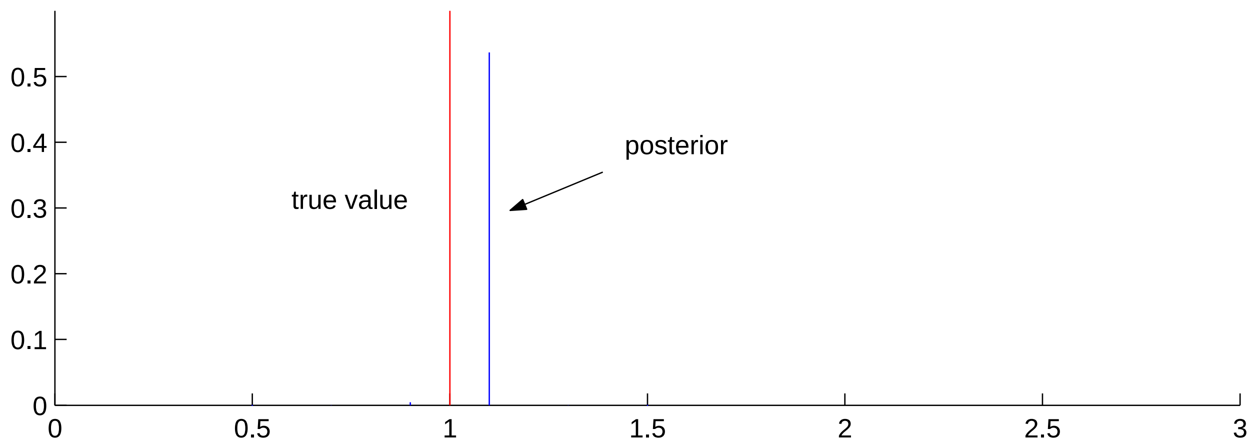
12.


0.04
0.03
0.02
0.01
0
0.5
1
1.5
2
2.5
3
0.5



14.





15. Given a sample, bounds on posterior odds

$$w_i(X) = \int_{\Theta_i} p_i(X \mid \theta_i)\pi_i(\theta_i)\, d\theta_i \tag{1}$$

posterior probability on model i is $w_i / \sum_j w_j$. Therefore

- By concentrating π_i near the peak of the model i likelihood, we can push w_i toward its upper bound, the LH max.
- If the parameter space is unbounded and the LH is integrable, LH will be arbitrarily small somewhere in the parameter space. We can make w_i arbitrarily small by concentrating the prior where LH is small.
- We can also make w_i arbitrarily small by making the prior very *flat* (e.g., $N(0, \Sigma)$ with $\Sigma \to \infty$), because LH approaches zero at infinity and "flatness" in this sense implies probability concentrated at infinity.

16. Implication for practice

- To avoid paradoxical results from model comparison exercises it is best where possible to think of the models as representative points in a larger continuous parameter space.
- Results that show overwhelming odds in favor of one model in the collection then imply that the collection is not well chosen: some rethinking of specification and filling in of the space of models is in order.

17. Applying these points to Smets and Wouters

- Posterior probability on DSGE in US data, as we saw yesterday, is 1.0.
- Posterior probability on DSGE in Euro area data in first paper was .07.
- Posterior probability on DSGE in current Euro area paper, with more data, is 0.00.
- As Peter Ireland was suggesting yesterday, the collection of models being considered is too sparse.

18. Trends and means

- Smets and Wouters, in their EMU work, take out mean and trend from the logged data, separately for each variable, and using the entire sample — including the "forecast period".
- For the US work, the trend is estimated in the DSGE model, but the data set is still "de-meaned" based on the whole sample, including forecast period. This could be particularly important for inflation.
- The DSGE "knows" the data are zero-mean and stationary. The VAR's have priors that focus on the possibility of unit root nonstationarity, which is by construction not there.

19. Training samples

$$p(Y_T \mid \theta)\pi(\theta) = [p(Y_{T_1} \mid \theta, Y_{T_0})] \cdot [p(Y_{T_0} \mid \theta)\pi(\theta)]\,, \tag{2}$$

The training sample method uses

$$\frac{p(Y_{T_0} \mid \theta)\pi(\theta)}{\int p(Y_{T_0} \mid \theta)\pi(\theta)\,d\theta} \tag{3}$$

as if it were the prior pdf and $p(Y_{T_1} \mid \theta, Y_{T_0})$ as if it were the likelihood.

20. Potential problems with training samples

- It "levels the playing field" if all models are handled this way. But if only some are (as in Smets and Wouters), it is arbitrary.

Model	Base	Training	Detrended	S&W detrend
VAR(3)		−330.21		
BVAR(6)	−280.87			
BVAR(5)	−277.85			
BVAR(4)	−281.23		−292.19	
BVAR(3)	−280.15	−251.26	−290.66	−266.71
Martingale	−312.06	−311.2	−272.17	
S&W DSGE				−269.2

TABLE 1. Marginal Data Densities (w_i's)

- The part of the sample used may be unrepresentative. In the US data, e.g., inflation in the 1947-57 training sample behaves very differently from any other period of postwar US history, with bursts of high, very short-lived inflation. Only the VAR models are given the "benefit" of a prior based on this period.
- Large models and small models are not in fact treated equally by this procedure. In fact, if a fixed fraction of the sample is used as training sample while $T \to \infty$, Bayesian model choice loses its consistency property when a restricted model is compared to a higher-dimensional one that embeds it.

21. (PARTIALLY) UNRAVELING THE EFFECTS OF PRIORS AND DATA FILTERS ON THE SMETS AND WOUTERS RESULTS

REFERENCES

BROCK, W. A., S. N. DURLAUF, AND K. D. WEST (2003): "Policy Evaluation in Uncertain Economic Environments," *Brookings Papers on Economic Activity*, (1), 1–67.

GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN (1995): *Bayesian Data Analysis*. Chapman and Hall, London.

GEWEKE, J. (1999): "Computational Experiments and Reality," Discussion paper, University of Iowa.

HAAVELMO, T. (1944): "The Probability Approach in Econometrics," *Econometrica*, 12(supplement), iii–vi+1–115.

SCHORFHEIDE, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15(6), 645–670.