Non-nested model selection

in unstable environments

Raffaella Giacomini

UCLA

(with Barbara Rossi, Duke)

Motivation

- The problem: select between two competing models, based on how well they fit the data

- Both models are possibly misspecified

- Widely documented instability in macro and financial time series $\Longrightarrow$ the relative performance of the models may also be unstable

- Existing techniques for model selection compare the *average* performance of the models over historical samples $\Longrightarrow$ loss of information if relative performance varies over time

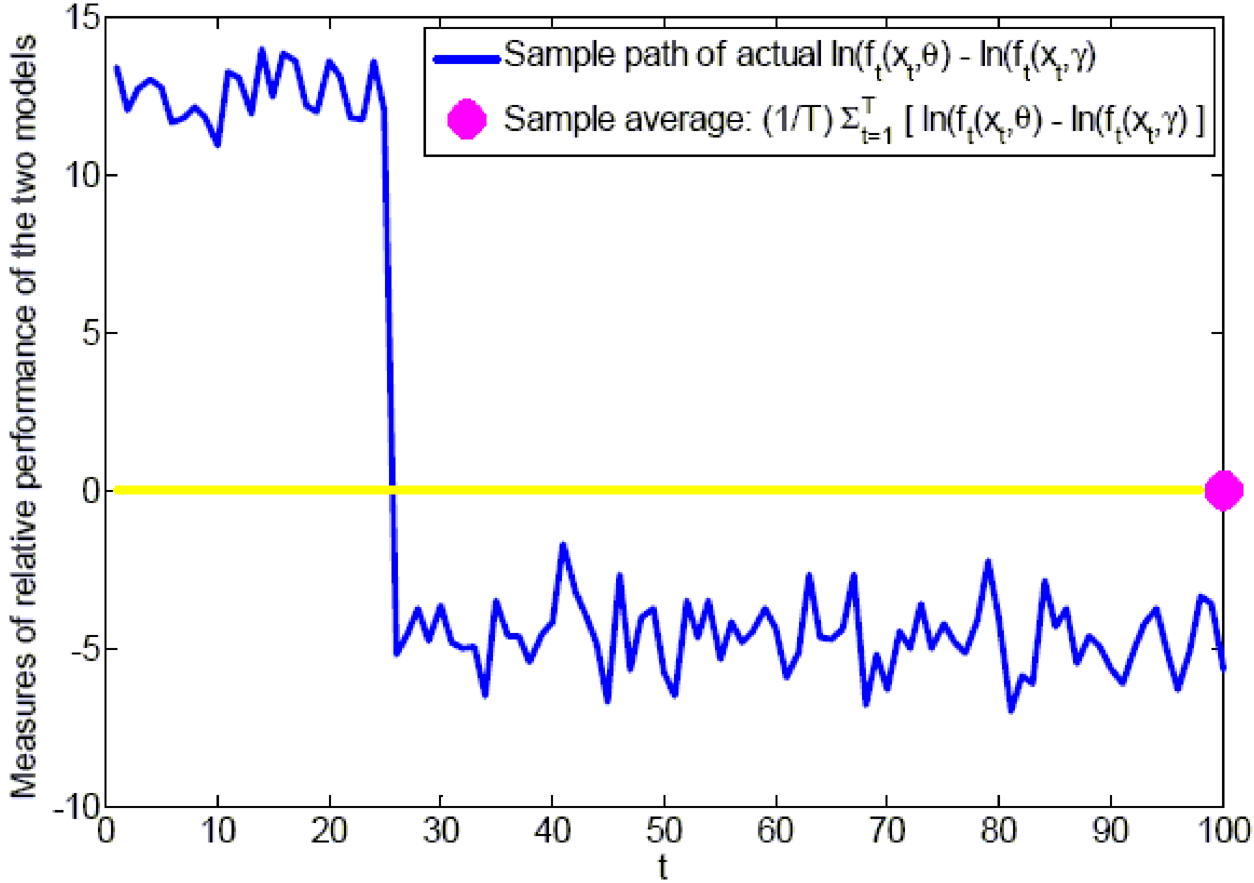
- Our goal: propose formal techniques to test whether the relative performance of two models is stable

Example

- Model 1 is better than model 2 ($=$ higher likelihood) at the beginning of the sample and then the order switches.

- One would like to choose model 2 at the end of the sample, but a measure of average performance may find that the models are equivalent.

Example (cont.)



Contributions

- We propose new model selection tests to evaluate whether the relative performance of two competing models varies over time. Two tests:

  - **Fluctuation test** to analyze the evolution of the models' relative performance over **historical** samples

  - **Sequential test** to monitor the models' relative performance in **real time**, as new data becomes available

- Valid under general conditions: data instability; models can be nonlinear, multivariate and misspecified; ML estimation

- Empirical application: evaluate time-variation in performance of Smets and Wouters' (2003) DSGE model relative to BVARs.

Outline of the talk

- $\boxed{\text{Fluctuation test}}$

- Sequential test

- Monte Carlo evidence

- Empirical application to DSGE vs. BVAR

- Conclusion

Related literature

Fluctuation test:

- Vuong, 1989 $\Longrightarrow$ test for equal full-sample *average* relative performance of non-nested, misspecified models (iid data)

- Rivers and Vuong, 2002 $\Longrightarrow$ heterogeneous + dependent data but tests that models are asymptotically equivalent

- Diebold- Mariano, 1995; West, 1996; McCracken 2000 etc. $\Longrightarrow$ test for equal out-of-sample *average* relative performance of non-nested, misspecified models

- Giacomini and White, 2006 $\Longrightarrow$ test if out-of-sample relative performance is related to explanatory variables

Related literature (cont.)

- Rossi, 2005 $\Longrightarrow$ test if parameters are stable $+$ significant ($\sim$ nested model selection under instability, but assumes correct specification)
  - We test a different H_0: the relative performance is *equal at each point in time* $+$ consider non-nested, possibly misspecified models

- Tests for parameter instability (Brown, Durbin and Evans, 1975, Andrews, 1993, Bai, 1998 etc. etc.)
  - We borrow from this literature, but focus on stability of relative performance rather than model's parameters (relative performance may be stable even though parameters change)

Related literature

Sequential test:

- Chu, Stinchcombe and White (1996) $\Longrightarrow$ real-time *parameter* stability within one model

- Inoue and Rossi (2005) $\Longrightarrow$ assess if variable has real-time predictive content for another variable

  - We ask a different question (compare non-nested models based on measures of fit) and have more general assumptions (e.g., both models misspecified)

Fluctuation test - Set-up

- Select a model for vector y_t using other variables z_t. Let $x_t = (y_t', z_t')'$. Historical sample $(x_1, ..., x_T)$.

- Two competing models, estimated by ML.

- Idea: re-estimate models recursively starting from observation $R < T$ using either an expanding sample ("recursive scheme") or a rolling samples of size R ("rolling scheme")

- At each time t, measure relative performance as $Q_t\left(\hat{\theta}_t\right) - Q_t\left(\hat{\gamma}_t\right)$ where $Q_t\left(\hat{\theta}_t\right)$ is the average log-likelihood over the estimation sample

Fluctuation test - Recursive scheme

Performance measure for model 1: $Q_t\left(\hat{\theta}_t\right) = t^{-1}\sum_{j=1}^{t} \ln f(x_j; \hat{\theta}_t)$

_ _ _ _ _ _ _ $\rightarrow$ $Q_R\left(\hat{\theta}_R\right) - Q_R\left(\hat{\gamma}_R\right)$

1 R

_ _ _ _ _ _ _ _ $\rightarrow Q_{R+1}\left(\hat{\theta}_{R+1}\right) - Q_{R+1}\left(\hat{\gamma}_{R+1}\right)$

1 R+1

_ $\rightarrow Q_T\left(\hat{\theta}_T\right) - Q_T\left(\hat{\gamma}_T\right)$

1 T

Fluctuation test - Rolling scheme

Performance measure for model 1: $Q_t\left(\hat{\theta}_t\right) = R^{-1}\sum_{j=t-R+1}^{t} \ln f(x_j;\hat{\theta}_t)$

$\_\_\_\_\_\_\_ \quad \rightarrow \quad Q_R\left(\hat{\theta}_R\right) - Q_R\left(\hat{\gamma}_R\right)$

1 R

$\_\_\_\_\_\_\_ \quad \rightarrow Q_{R+1}\left(\hat{\theta}_{R+1}\right) - Q_{R+1}\left(\hat{\gamma}_{R+1}\right)$

2 R+1

$\_\_\_\_\_\_ \quad \rightarrow Q_T\left(\hat{\theta}_T\right) - Q_T\left(\hat{\gamma}_T\right)$

T-R+1 T

Fluctuation test - Null hypothesis

- Null hypothesis: $H_0 : EQ_t\left(\theta_t^*\right) - EQ_t\left(\gamma_t^*\right) = 0$ for all $t = 1, ..., T$,

- θ_t^* and γ_t^* are the pseudo-true parameters. E.g., for the recursive scheme,

$$\theta_t^* = \arg\max_\theta E\left(t^{-1}\sum_{j=1}^{t} \ln f\left(x_j;\theta\right)\right)$$

- Note that the parameters of the models may be unstable under the null hypothesis

Fluctuation test - Implementation

- Compute sequences of statistics for $t = R, ..., T$

$$\begin{aligned} F_t^{rec} &= \hat{\sigma}_t^{-1}\sqrt{t}\left(Q_t(\hat{\theta}_t) - Q_t(\hat{\gamma}_t)\right) \\ F_t^{roll} &= \hat{\sigma}_t^{-1}\sqrt{R}\left(Q_t(\hat{\theta}_t) - Q_t(\hat{\gamma}_t)\right), \end{aligned}$$

  $\hat{\sigma}_t^2$ is a HAC estimator of the asymptotic variance $\sigma_t^2 = var(\sqrt{t}\,(Q_t(\theta_t^*) - Q_t(\gamma_t^*)))$

- Does the sample path of F_t significantly deviate from its hypothesized value of 0?

Fluctuation test - Recursive scheme

- Intuition: if the models are non-nested, for a particular t, F_t^{rec} can be approximated by a $N(0,1)$ under $H_0 \Longrightarrow$ the sample path of F_t^{rec} behaves like a Brownian motion that starts at zero at time $t = R$.

  $\Longrightarrow$ can derive boundary lines that are crossed with given probability α. For each $t = R, ..., T$ and significance level α:

$$c_{\alpha,t}^{rec} = \pm\, k_\alpha^{rec}\sqrt{\frac{T-R}{t}}\left(1 + 2\frac{t-R}{T-R}\right)$$

  (Same as for the CUSUM test of Brown, Durbin and Evans, 1975)

- Typical values of (α, k_α^{rec}) are $(0.01, 1.143)$, $(0.05, 0.948)$ and $(0.10, 0.850)$

Fluctuation test - Rolling scheme

- For the rolling scheme, the sample path of F_t^{rol} behaves like the increments of a Brownian bridge $\Longrightarrow$ can derive boundary lines that are crossed with given probability α by simulation

$$c_{\alpha,t}^{roll} = \pm\, k_{\alpha}^{roll}$$

- k_{α}^{roll} depends on R/T. We give a table with k_{α}^{roll} for typical values of a and R/T

Fluctuation test - Assumptions

- Standard assumptions that guarantee that a FCLT holds for $T^{-1/2}\sum_{j=1}^{t} \ln f(x_j;\theta)$

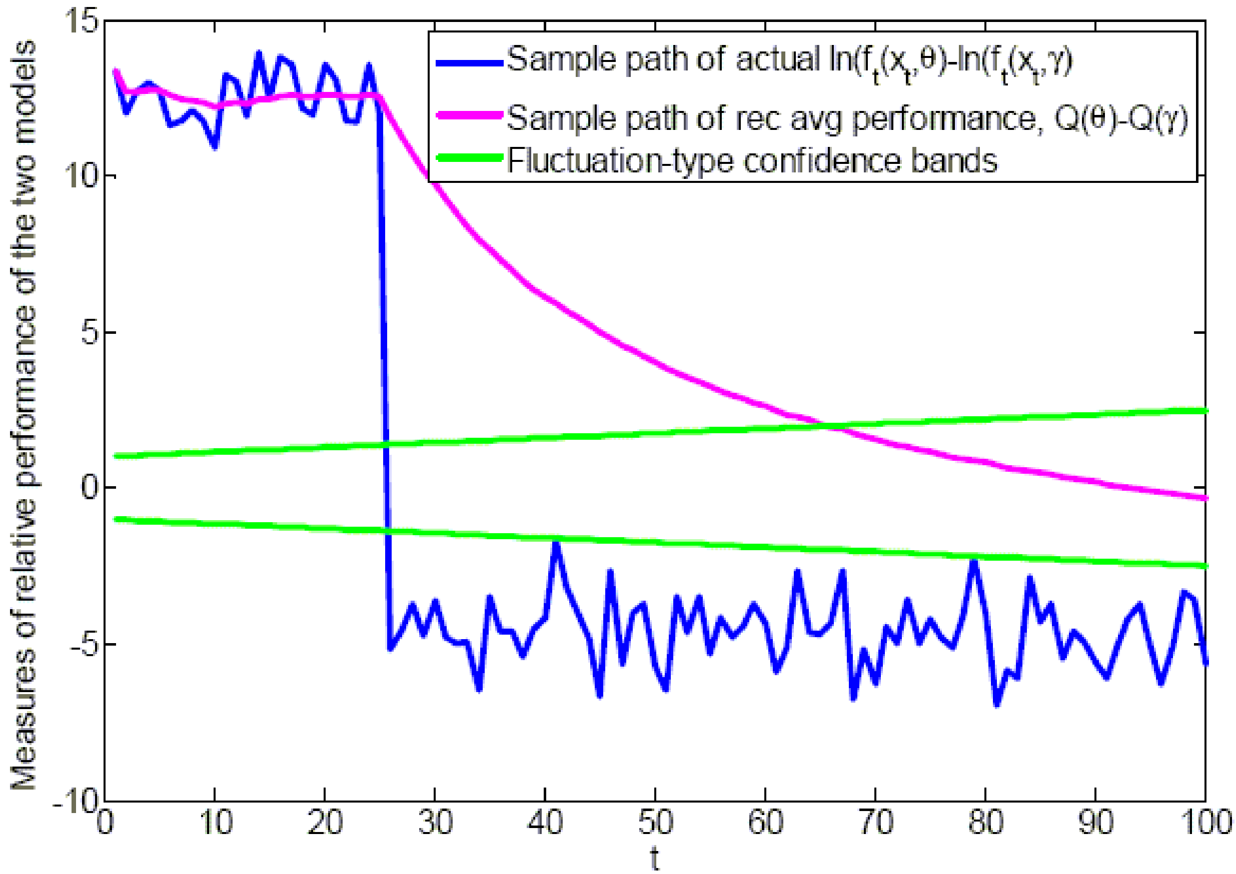
- Parameter and data instability allowed under H_0

- Data have short memory ($\Longrightarrow$ no unit roots)

- σ_t^2 is not $o(1)$ (rules out nested models)

- $R/T \to \rho$ finite and positive

Fluctuation test in practice (recursive)



Outline of the talk

- Fluctuation test
- $\boxed{\text{Sequential test}}$
- Monte Carlo evidence
- Empirical application to DSGE vs. BVAR
- Conclusion

Sequential test

- Monitor the model-selection decision in the post-historical sample period

- Suppose model 1 was best over the historical sample up to time T :

$$EQ_T\left(\theta_T^*\right) - EQ_T\left(\gamma_T^*\right) > 0$$

- Test hypothesis that model 1 continues to be the best for all subsequent periods:

$$H_0 : EQ_t(\theta_t^*) - EQ_t(\gamma_t^*) \geq 0 \text{ for } t = T+1, T+2, \ldots,$$

  against $H_1 : EQ_t(\theta_t^*) - EQ_t(\gamma_t^*) < 0$ at some $t \geq T$.

- Doing a sequence of full-sample Vuong's (1989) tests rejects too often $\Longrightarrow$ find critical values that control the overall size of the procedure

Sequential test

- Compute sequence of test statistics for $t = T + 1, T + 2, ...$

$$J_t = \hat{\sigma}_t^{-1} \sqrt{t} \left(Q_t(\hat{\theta}_t) - Q_t(\hat{\gamma}_t) \right)$$

- As in Chu, Stinchcombe and White (1995), the critical value at time t for a level α test is:

$$c_{\alpha,t} = -\sqrt{k_\alpha + \ln(t/T)}$$

- Typical values of (α, k_α) are $(0.05, 2.7955)$ and $(0.10, 2.5003)$.

Outline of the talk

- Fluctuation test

- Sequential test

- $\boxed{\text{Monte Carlo evidence}}$

- Empirical application to DSGE vs. BVAR

- Conclusion

Monte Carlo evidence

- Compare our test to Vuong's (1989) in-sample test and to West's (1996) out-of-sample test

- DGP with parameter variation:

$$
\begin{aligned}
y_t &= \beta_t x_t + \gamma_t z_t + \varepsilon_t,\ t = 1, 2, ...T,\ T = 400 \\
\beta_t &= 1 + \beta \cdot 1\,(200 < t \leq 250) + (1 - \beta) \cdot 1\,(t > 250) \\
\gamma_t &= 1 + \gamma \cdot 1\,(200 < t \leq 250) + (1 - \gamma) \cdot 1\,(t > 250)\,.
\end{aligned}
$$

- Model 1: $y_t = \beta x_t + u_{1t}$. Model 2: $y_t = \gamma z_t + u_{2t}$.

- Size: $\beta = \gamma = 0.5 \Longrightarrow$ models are equally good.

- Power: $\beta = 0.95, \gamma = 0.4 \Longrightarrow$ time variation in relative performance

Monte Carlo evidence

Table 2. Empirical rejection frequencies of nominal 5% tests.

		F_t^{rec}	Vuong	West
(a) Historical sample	Size	0.051	0.047	0.044
	Power	0.449	0.047	0.026
(b) Post-historical sample	t/T	J_t	Vuong	
	1.5	0.010	0.121	
	1.75	0.020	0.152	
	2	0.032	0.179	

Outline of the talk

- Fluctuation test

- Sequential test

- Monte Carlo evidence

- $\boxed{\text{Empirical application to DSGE vs. BVAR}}$

- Conclusion

Application: SW's DSGE vs. BVAR

- Smets and Wouters (2003) (SW): "An estimated DSGE model of the Euro Area": estimation of a 7-equation linearized DSGE model with sticky prices and wages, habit formation, capital adjustment costs and variable capacity utilization.

- Their findings:
  - the DSGE model has comparable fit to that of atheoretical VARs
  - the parameter estimates have the expected sign

Open questions

- Have the parameters been stable? Perhaps not. Possible *structural changes* in the economy (European union introduction, productivity changes, etc.)

- If the parameters have changed $\Longrightarrow$ the performance of the DSGE model may have changed too... so SW's result only holds on average

- Can we say that the DSGE's performance has been *stable over time*, and *significantly better* than that of the VAR?

Application: SW's DSGE vs. BVAR

- SW sample: quarterly data 1970:2-1999:4 on DGP, consumption, investment, prices, real wages, employment, real interest rate

- Two questions:

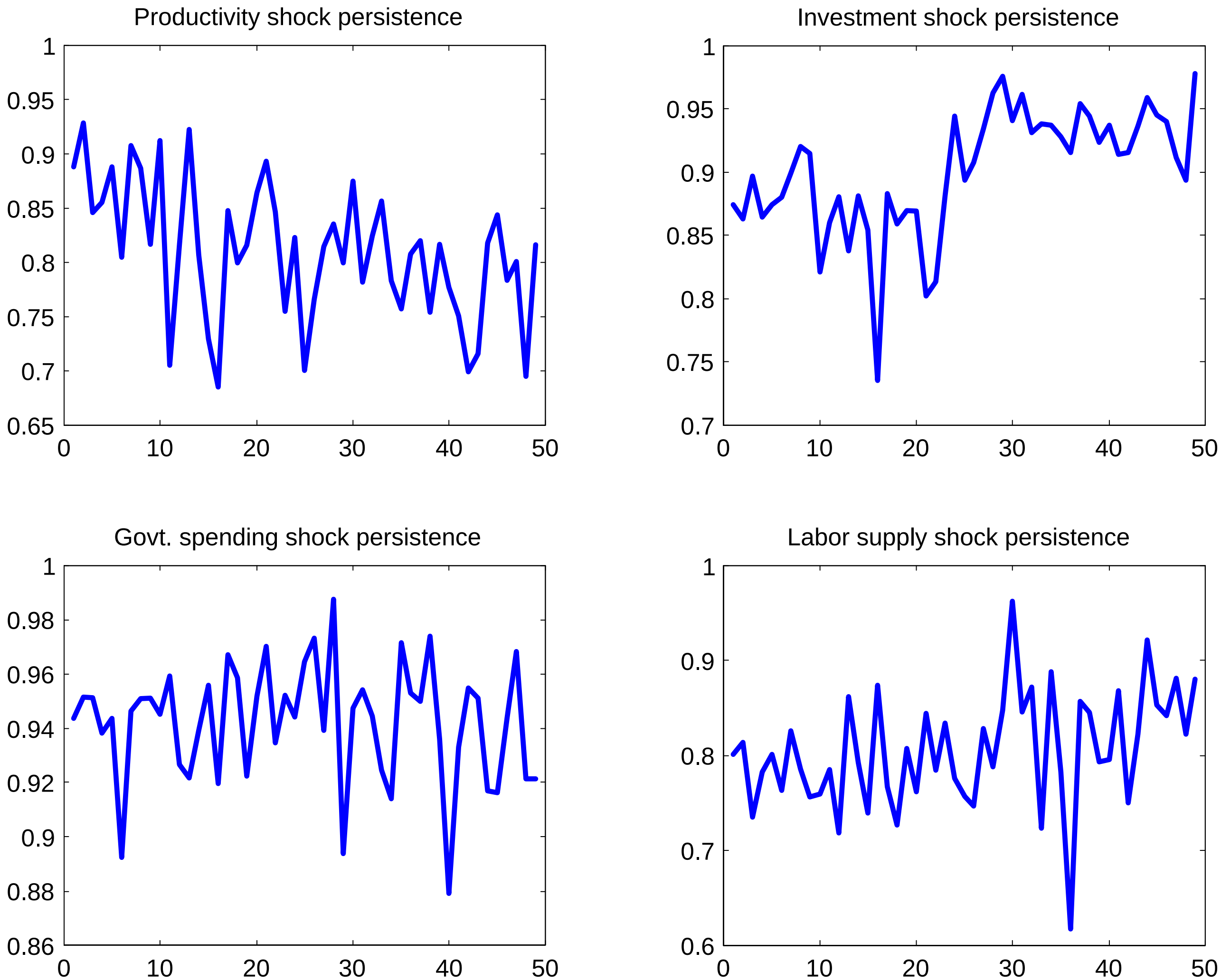
  1. Is there evidence of parameter variation in the DSGE parameters?

  2. Was the DSGE *consistently* and *significantly* better than atheoretical Bayesian VARs over the sample?

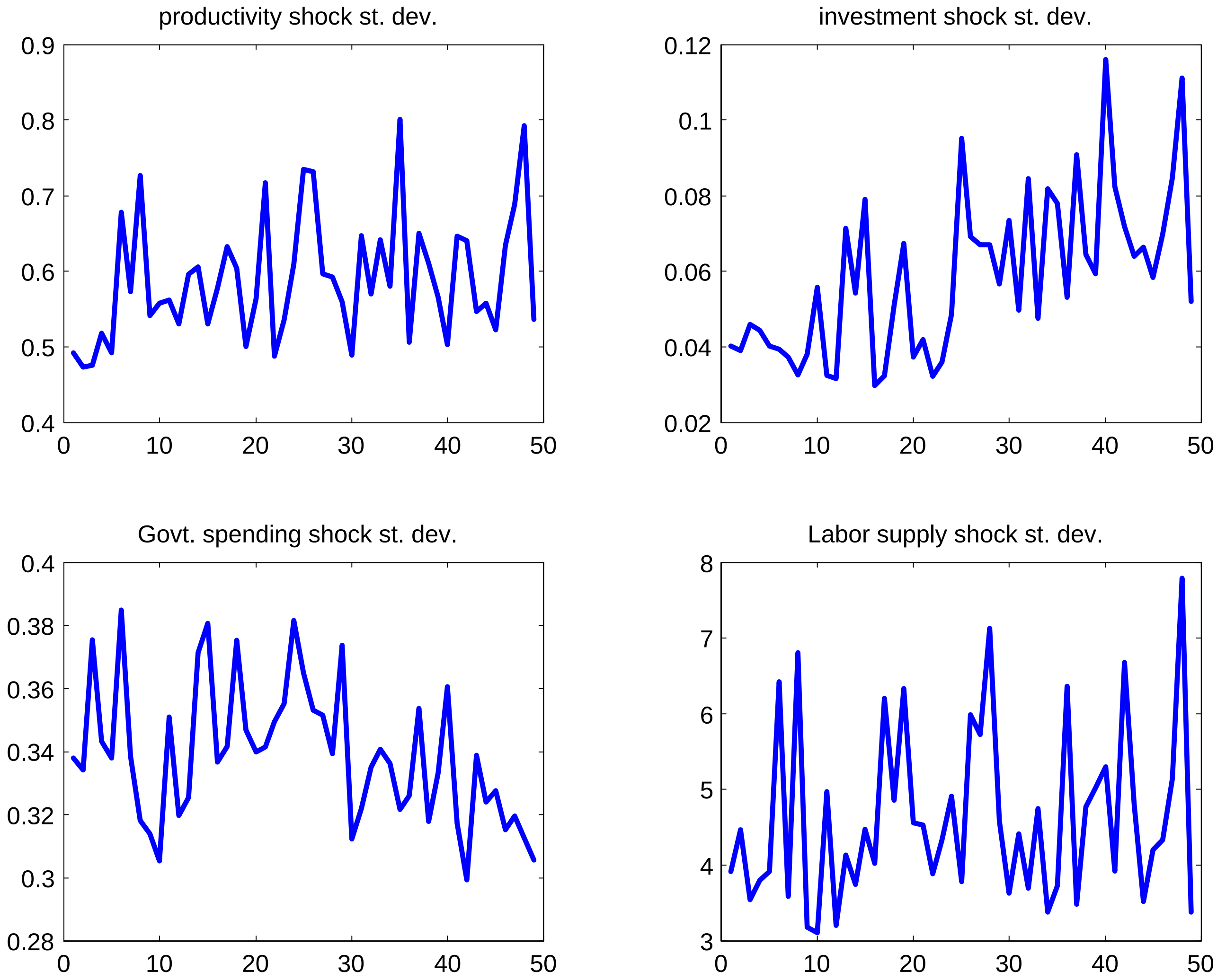
Q#1: time-variation in DSGE parameters

- Estimate DSGE model recursively using rolling samples of size $R = 70$ and plot posterior mode
  - persistence of shocks
  - standard deviation of shocks
  - monetary policy parameters

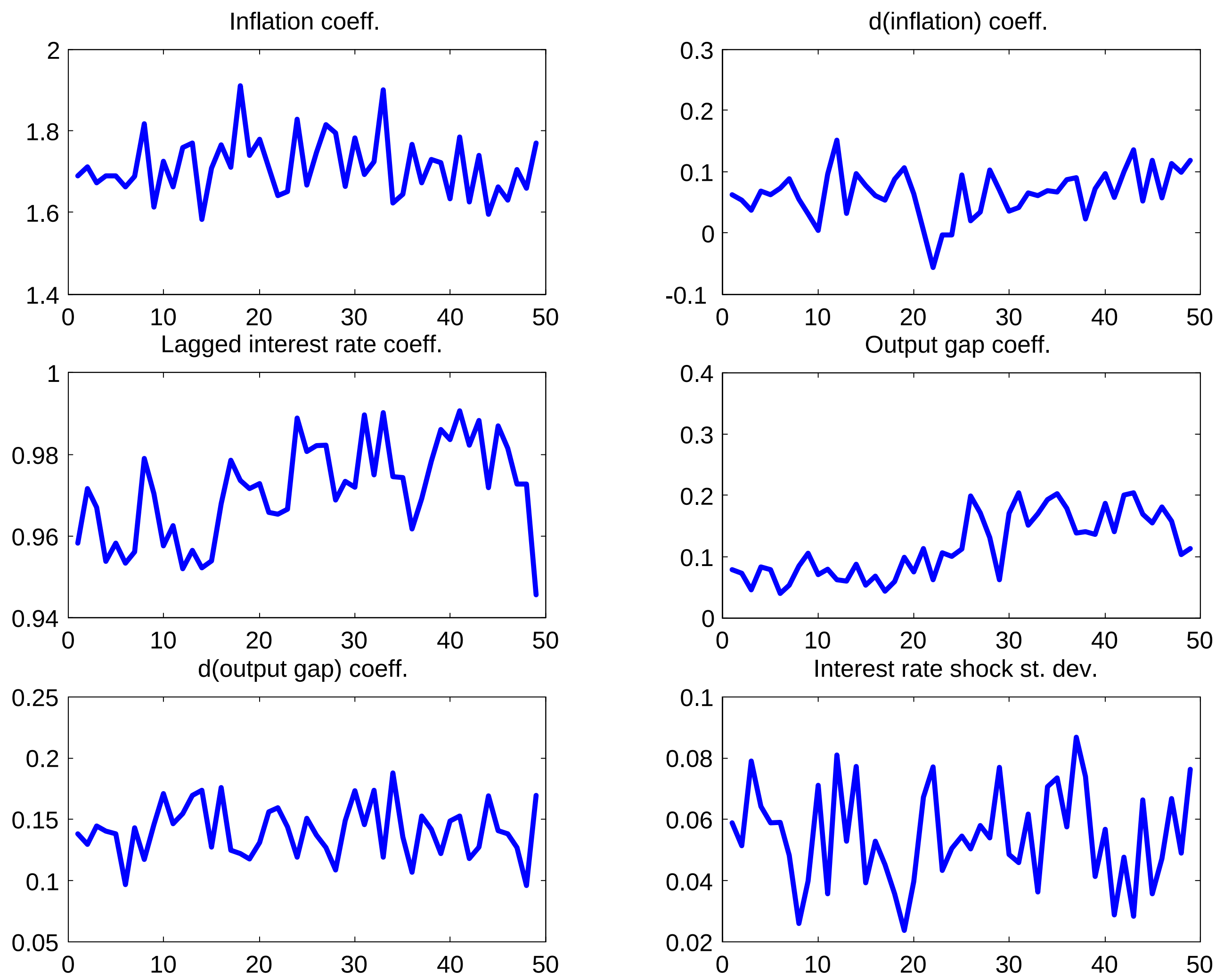
Parameter variation - shocks' persistence



Parameter variation - standard deviation of shocks



Parameter variation – monetary policy parameters



Summary: time-variation in structural parameters

- Moderate evidence of parameter variation in the DSGE model:
  - persistence of productivity shock $\downarrow$
  - persistence and standard deviation of investment shock $\uparrow$
  - monetary policy coefficients fairly stable

Q#2: time-varying performance of DSGE vs BVAR

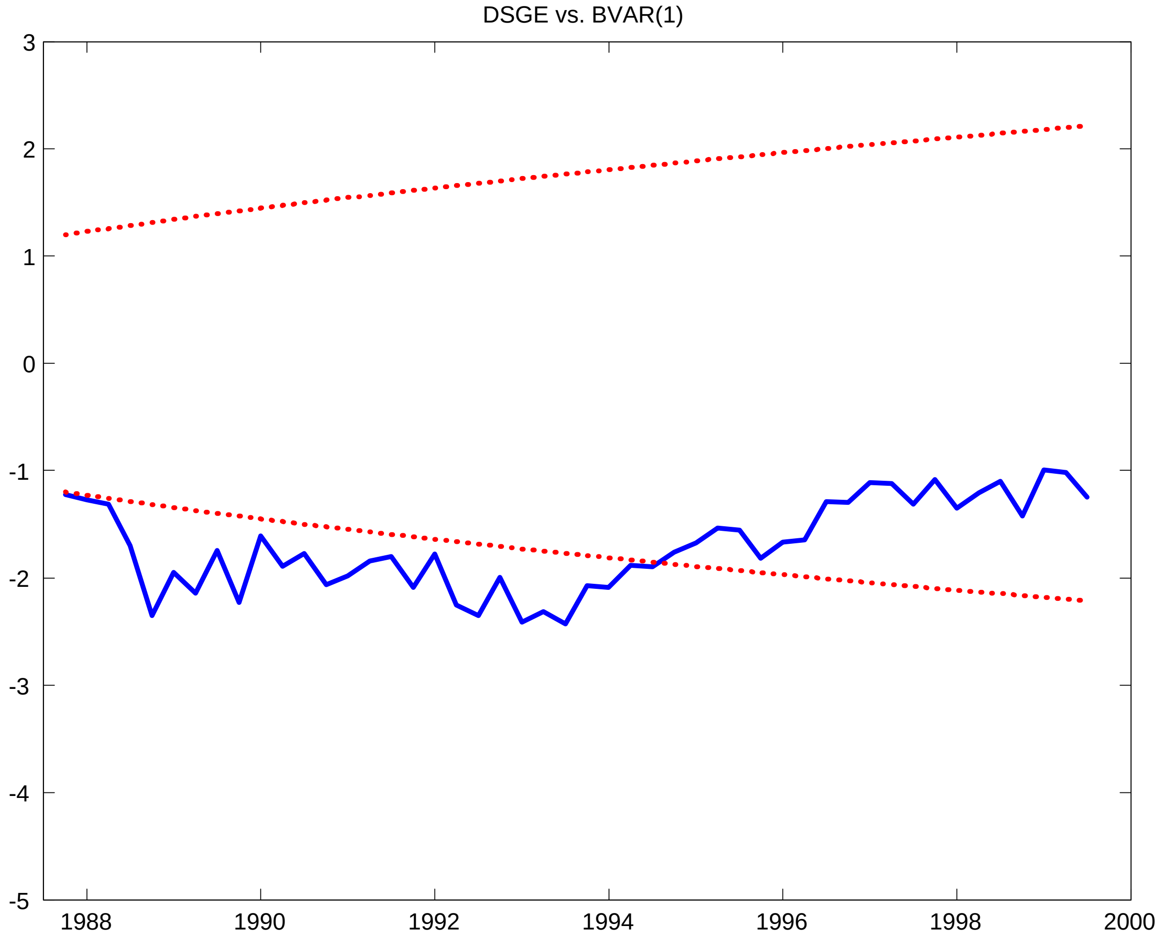
- Recursive and rolling fluctuation tests. Total sample $T = 118$. Initial estimation sample $R = 70$

- DSGE vs. BVAR(1) with Minnesota priors

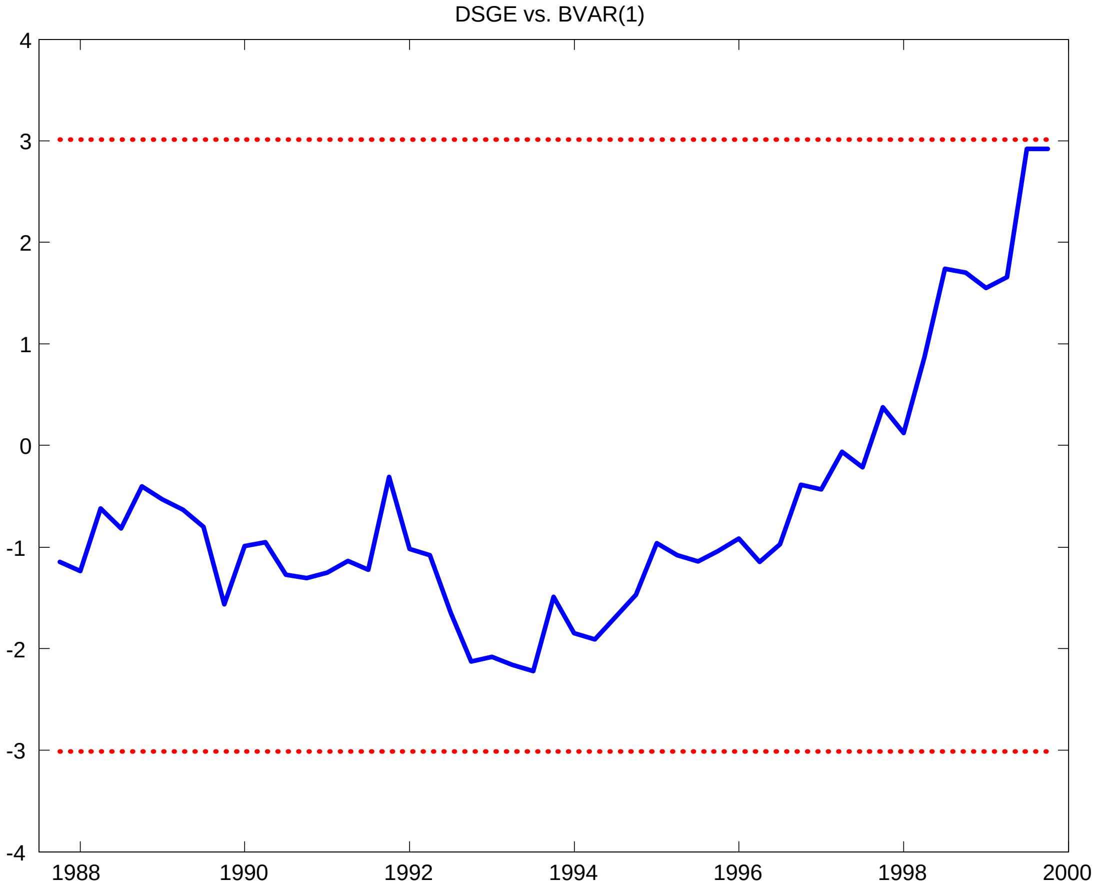
- Compute sequences of difference in average log-likelihoods $Q_t\left(\widehat{\theta}_t\right) - Q_t\left(\widehat{\gamma}_t\right)$ for $t = R, ..., T$

- $\widehat{\theta}_t$ is the posterior mode (a consistent estimator of θ_t^*)

Fluctuation test - recursive



Fluctuation test - rolling



Conclusion and extensions

- Proposed a formal method for evaluating time-variation in relative performance of misspecified non-nested models

- Empirical application confirmed SW's result that a DSGE has comparable performance to a BVAR in recent years

- Extension: optimal fluctuation test (against random walk alternative)