Efficient Estimation of the Parameter Path in Unstable Time Series Models

Ulrich K. Müller and Philippe Petalas
Princeton University

March 10, 2007

Motivation

Time series models have potentially time varying parameters

$\Rightarrow$ Parameter path estimator in unstable models?

- descriptive tool that helps determine the source of the instability
- interesting for forecasting
- sometimes structural interpretation

Parameter Path in Parametric Model

- Stable and stationary parametric model with log-likelihood function $\sum_{t=1}^{T} l_t(\theta)$, where $\theta \in \Theta \subset \mathbb{R}^k$

- Likelihood function of unstable model: $\sum_{t=1}^{T} l_t(\theta_t)$

- Parametrize parameter path as

$$\{\theta_t\}_{t=1}^{T} = \{\theta + \delta_t\}_{t=1}^{T} \quad \text{with} \quad \sum_{t=1}^{T} \delta_t = 0$$

  so that θ is benchmark value and $\delta = (\delta_1', \cdots, \delta_T')' \in \mathbb{R}^{Tk}$ are the deviations

Inference in Linear Gaussian Model

- Consider the Gaussian model

$$\begin{aligned} &Y_0 = \theta + T^{-1/2}\nu_0 \\ &Y_t = \delta_t + \nu_t \qquad t = 1, \cdots, T \\ &\delta \sim \mathcal{N}(0, \Sigma_\delta) \ \text{ independent of } \{\nu_t\} \end{aligned}$$

  with Y_0 and $Y = (Y_1', \cdots, Y_T')'$ observed and $\nu_t \sim i.i.d.\mathcal{N}(0, \Omega)$

- Under symmetric loss, efficient estimator of δ is $\hat{\delta}^* = E[\delta|Y] = \Sigma Y$, where Σ is a function of Σ_δ and Ω, and efficient estimator of path $\{\theta_t\}_{t=1}^T$ is $\{Y_0 + \hat{\delta}_t^*\}_{t=1}^T$.

- Efficient test for parameter stability $H_0 : \delta = 0$ against $H_1 : \delta \sim \mathcal{N}(0, \Sigma_\delta)$ is based on $Y'\Sigma Y = (\Sigma Y)'\Sigma^{-1}\Sigma Y = \hat{\delta}^{*\prime}\Sigma^{-1}\hat{\delta}^*$

- When $\{\delta_t\}_{t=1}^T$ is a demeaned Gaussian Random Walk, then estimator $\hat{\delta}^* = \Sigma Y$ can be computed from variants of Kalman smoothing

Optimality Properties

- The model

$$
\begin{aligned}
&Y_t = \delta_t + \nu_t \qquad t = 1, \cdots, T \\
&\delta = \mathcal{N}(0, \Sigma_\delta) \;\; \text{independent of } \{\nu_t\}
\end{aligned}
$$

posits δ to be random: Assumption $\delta \sim \mathcal{N}(0, \Sigma_\delta)$ may be viewed as prior in a Bayesian analysis

- If δ is viewed as fixed but unknown, then $\hat{\delta}^* = \Sigma Y$ minimizes Weighted Average Risk for symmetric loss functions L

$$
WAR(\hat{\delta}) = \int E[L(\hat{\delta}, \delta)]dQ(\delta)
$$

when Q is proportional to the distribution $\mathcal{N}(0, \Sigma_\delta)$

- Similarly, a parameter stability test based on $Y'\Sigma Y$ maximizes Weighted Average Power

$$
WAP(\varphi) = \int P(\varphi \text{ rejects} | \delta = d)dQ(d)
$$

Contribution

- Considers general time series likelihood model (nonlinear, non-Gaussian) with Gaussian parameter evolution/weighting function.

- Focusses on parameter variations whose presence cannot be detected with probability one, even in the limit.

- Shows that sample information is efficiently summarized by linear Gaussian pseudo model, with score vectors as the observations
$\Rightarrow$ asymptotically weighted average risk minimizing path estimators and weighted average power maximizing test statistic are straightforward to compute

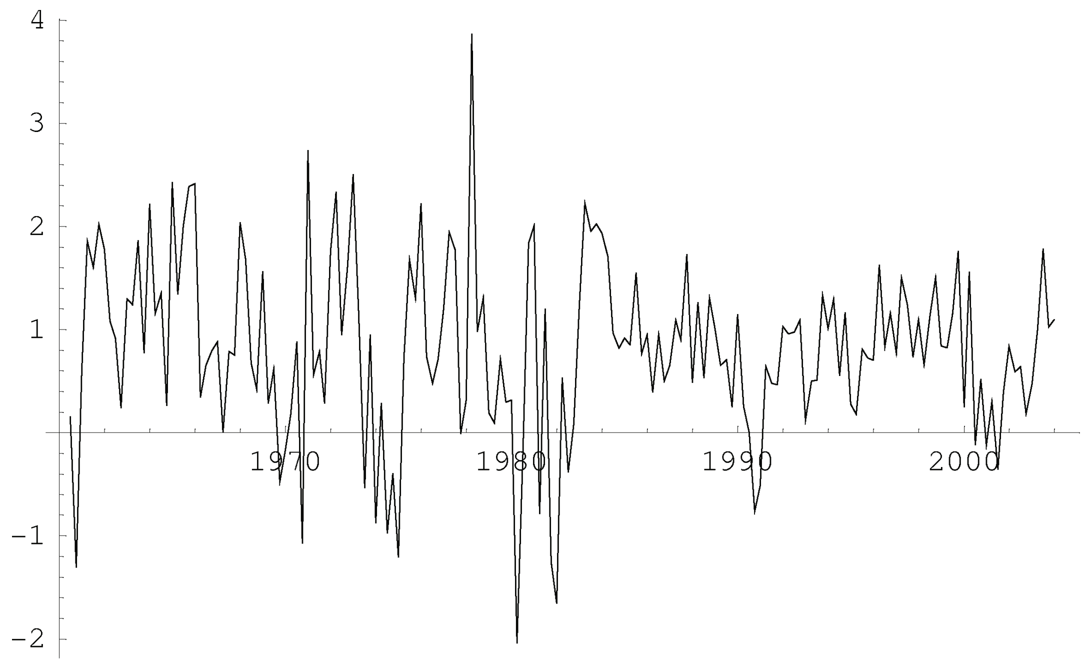
- Idea: quadratic approximation to log-likelihood.

Example: US GDP Growth

Time varying innovation variance of AR(2)

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \sqrt{\sigma_t^2} e_t$$

where $\{e_t\}$ is i.i.d.$\mathcal{N}(0, 1)$



WAR and WAP in General Model

- General unstable model has likelihood $\sum_{t=1}^{T} l_t(\theta_t)$, where $\{\theta_t\}_{t=1}^{T} = \{\theta + \delta_t\}_{t=1}^{T}$

- Compare path estimators $\hat{a} = (\hat{a}_1', \cdots, \hat{a}_T')' \in \mathbb{R}^{Tk}$ by Weighted Average Risk

$$WAR(\hat{a}) = \int w(\theta) \int E[L_T(\hat{a}, \theta, \delta)] dQ_T(\delta) d\theta$$

where L_T is a bounded loss function

- Compare tests of parameter stability $H_0 : \delta = 0$ by Weighted Average Power

$$WAP(\varphi) = \int P(\varphi \text{ rejects} | \delta = d, \theta = \theta_0) dQ_T(d)$$

Assumption on Weights

- Weighting function Q_T on $\delta = (\delta_1', \cdots, \delta_T')'$ is distribution of

  $$\{T^{-1/2}(G(t/T) - \bar{G}_T)\}_{t=1}^T$$

  where $\bar{G}_T = T^{-1}\sum_{t=1}^T G(t/T)$ and $G(\cdot)$ is a continuous Gaussian process on the unit interval.

  Example: $G(\cdot) = \Upsilon^{1/2}W(\cdot)$, where $W(\cdot)$ is standard $k \times 1$ Wiener process: Q_T is the distribution of a demeaned Gaussian Random Walk

- Weighting function w does not depend on T and is continuous and integrable

Motivation for Weighting Function Q_T

- Motivation for persistent parameter variation
  - plausible for many causes of instability (drifts in preferences, institutions, etc.)
  - popular in economic modelling
- Motivation for small parameter variation
  - corresponds to local neighborhood in which tests of parameter stability have nontrivial power
  - focusses on parameter paths which are difficult to determine, even asymptotically
  - 'small' time variation empirically common (somewhat significant p-values of stability tests)

Notation

- Maximum likelihood estimator $\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{T} l_t(\theta)$

- Score vectors $s_t(\theta) = \partial l_t(\theta)/\partial\theta$, $t = 1, \cdots, T$

- Hessians $h_t(\theta) = -\partial s_t(\theta)/\partial\theta'$, $t = 1, \cdots, T$

- average Hessians $\hat{H} = T^{-1} \sum_{t=1}^{T} h_t(\hat{\theta})$

Main Result

Under weak regularity conditions on the likelihood of the *stable* model, asymptotically efficient inference on θ and $\{\delta_t\}_{t=1}^T$ can be carried out *as if* the sample information was given by the pseudo Gaussian model

$$\begin{aligned}&\hat{\theta} = \theta + T^{-1/2}\nu_0\\&\hat{H}^{-1}s_t(\hat{\theta}) = \delta_t + \nu_t, \qquad t = 1, \cdots, T\end{aligned}$$

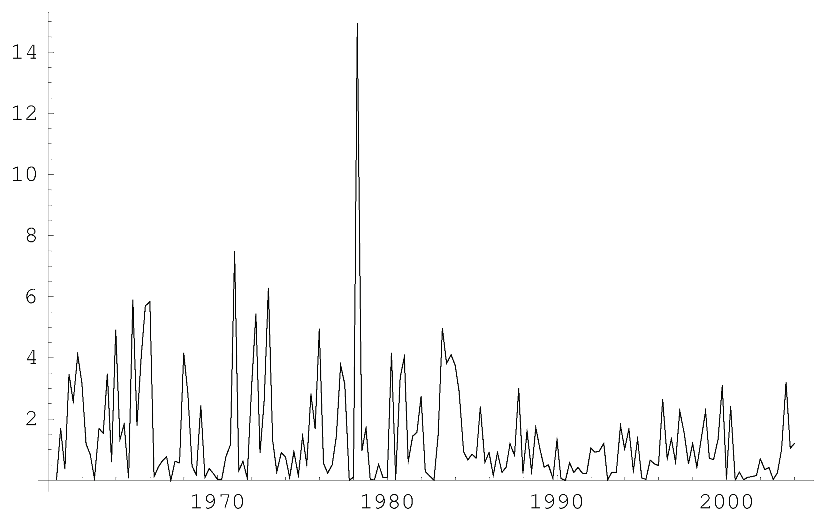
where $\nu_t \sim i.i.d.\mathcal{N}(\hat{H}^{-1})$.

1. For bounded and symmetric loss functions, asymptotically efficient estimator of δ is $\Sigma\hat{S}$, and asymptotically efficient test of parameter constancy is $\hat{S}'\Sigma\hat{S}$, where $\hat{S} = (\hat{H}^{-1}s_1(\hat{\theta})', \cdots, \hat{H}^{-1}s_T(\hat{\theta})')'$

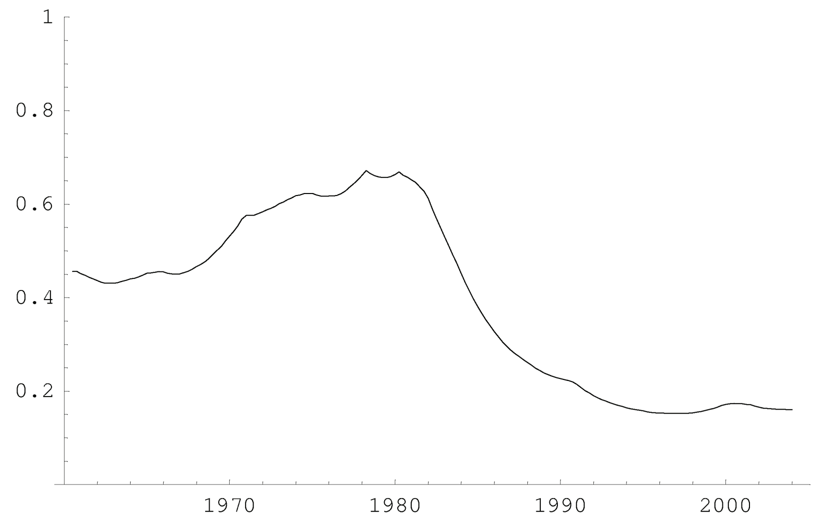
2. When $G(\cdot) = \Upsilon^{1/2}W(\cdot)$, the efficient path estimator $\Sigma\hat{S}$ amounts to Kalman smoothing the scores.

Application to GDP Growth

$\{e_t^2\}$



$\{\hat{\sigma}_t^2\}$



Heuristic Derivation

- The sample information about the θ and $\{\delta_t\}_{t=1}^T$ is fully contained in the function $\sum l_t(\theta + \delta_t)$.

- Idea: quadratic approximation of $\sum l_t(\theta + \delta_t)$ by second-order Taylor expansion.

- Relies heavily on 'Local Law of Large Numbers' (LLLN)

$$T^{-1} \sum h_t(\theta_t) \xrightarrow{p} H$$

for all sequences $\{\theta_t\}_{t=1}^T$ s.t. $\sup_t ||\theta_t - \theta_0|| \to 0$ (reasonable, as $T^{-1} \sum h_t(\theta_t) = T^{-1} \sum [h_t(\theta_0) + R_t]$ where R_t is small when Hessian is differentiable)

$\Rightarrow$ in particular, $\hat{H} = T^{-1} \sum h_t(\hat{\theta}) \xrightarrow{p} H$, since $\hat{\theta}$ is $\sqrt{T}$ consistent

Properties of Smooth Averages

- Let $\{\gamma_t\}$ be a step function of order $T^{-1/2}$ with a single step, i.e. $\gamma_t = T^{-1/2} g_0 \mathbf{1}[t \leq \pi T] + T^{-1/2} g_1 \mathbf{1}[t > \pi T]$. Then with $\sup_t ||\theta_t - \theta_0|| \to 0$,

$$\begin{aligned} \sum_{t=1}^{T} \gamma_t' h_t(\theta_t) \gamma_t &= T^{-1} \sum_{t=1}^{[\pi T]} g_0' h_t(\theta_t) g_0 + T^{-1} \sum_{t=[\pi T]+1}^{T} g_1' h_t(\theta_t) g_1 \\ &\approx \pi g_0' H g_0 + (1-\pi) g_1' H g_1 \\ &\approx \sum_{t=1}^{T} \gamma_t' H \gamma_t \approx \sum_{t=1}^{T} \gamma_t' \hat{H} \gamma_t \end{aligned}$$

- Any smooth process $\{\gamma_t\}$ of order $T^{-1/2}$ can be approximated by a step function (with many steps), so be the same argument,

$$\sum_{t=1}^{T} \gamma_t' h_t(\theta_t) \gamma_t \approx \sum_{t=1}^{T} \gamma_t' \hat{H} \gamma_t$$

Taylor Expansion

- By T exact Taylor expansions

$$\sum l_t(\theta+\delta_t) - \sum l_t(\hat\theta) = \sum s_t(\hat\theta)'(\theta+\delta_t-\hat\theta) - \frac{1}{2}(\theta+\delta_t-\hat\theta)'h_t(\tilde\theta_t)(\theta+\delta_t-\hat\theta)$$

where $\tilde\theta_t$ is between $\theta_0+\delta_t$ and $\hat\theta$.

- But $\{\theta+\delta_t-\hat\theta\}_{t=1}^T$ is smooth and of order $T^{-1/2}$, so that

$$\sum(\theta+\delta_t-\hat\theta)'h_t(\tilde\theta_t)(\theta+\delta_t-\hat\theta) \approx \sum(\theta+\delta_t-\hat\theta)'\hat H(\theta+\delta_t-\hat\theta)$$

- Also, $\sum s_t(\hat\theta)=0$ by FOC of MLE, and $\sum \delta_t = 0$ by construction.

Completing the Square

- We find

$$\begin{aligned} &\sum(l_t(\theta+\delta_t)-l_t(\hat\theta))-\tfrac{1}{2}\sum s_t(\hat\theta)'\hat H^{-1}s_t(\hat\theta) \\ \approx\ & -\tfrac{1}{2}\sum(\hat H^{-1}s_t(\hat\theta)-\delta_t)'\hat H(\hat H^{-1}s_t(\hat\theta)-\delta_t)-\tfrac{1}{2}T(\theta-\hat\theta)'\hat H(\theta-\hat\theta) \end{aligned}$$

$\Rightarrow$ Ignoring constants, this is the log-likelihood of the (demeaned) Gaussian random variables $\{\hat H^{-1}s_t(\hat\theta)\}_{t=1}^T$ with mean $\{\delta_t\}$ and covariance matrix $\hat H^{-1}$, and the Gaussian random variable θ with mean $\hat\theta$ and covariance matrix $T^{-1}\hat H^{-1}$

$\Rightarrow$ Sample information can be approximated by the observations

$$\begin{aligned} \hat\theta &= \theta + T^{-1/2}\nu_0 \\ \hat H^{-1}s_t(\hat\theta) &= \delta_t+\nu_t \qquad t=1,\cdots,T \end{aligned}$$

where $\nu_t \sim i.i.d.\mathcal{N}(\hat H^{-1})$.

Extensions

1. Approximation for nonstationary stable models

   $\Rightarrow$ sample information summarized by linear Gaussian model

   $$s_t(\hat{\theta}) + \tilde{h}_t\hat{\theta} = \tilde{h}_t(\delta_t + \theta) + \nu_t, \quad \nu_t \sim \text{independent } \mathcal{N}(0, \tilde{h}_t)$$

   where $\tilde{h}_t$ is any sequence of positive definite matrices satisfying $\sup_{\lambda\in[0,1]} \left\| T^{-1} \sum_{t=1}^{[\lambda T]} (\tilde{h}_t - h_t(\theta_0)) \right\| \xrightarrow{p} 0$

2. Weighting functions Q_T that are mixtures of distribution of Gaussian stochatic processes $G_i(\cdot)$

   $\Rightarrow$ useful for, say, making scale of parameter variability data induced

Conclusion

1. Sample information is efficiently summarized by Gaussian pseudo model, in the limit.

2. Computationally straightforward formulae for efficient estimator of the parameter path.

3. Unifying framework for efficient estimators of parameter path and efficient tests of parameter stability.