Estimation and Inference in Unstable Nonlinear Least Squares Models

Otilia Boldea & Alastair Hall

Department of Economics
North Carolina State University

December 11, 2006
How do changes in policy regimes or economic environments affect economic decisions at a macroeconomic level?

Answer in Econometric Literature:

- linear or nonlinear models;
- stationary and nonstationary models;
- known or unknown break;
- single break or multiple breaks;
- single estimating equation or system of equations;
- estimation of breaks or tests for structural change.
Tests in linear models

- stationary: Bai and Perron (1998); 
  \((\text{sup}) \ F\text{-tests}\) - initially developed by Quandt (1960).

  \((\text{inf}) \ t\text{-tests}\).

Tests in nonlinear models

  (sup, average, exponential) Wald, LM LR-type tests.

Break-point estimation in linear models

- level shifts in mean: Yao and Au (1988);
- general classes of linear models Bai and Perron (1998), Hall and Han (2005), Perron and Qu (2006).
How do changes in policy regimes or economic environments affect economic decisions at a macroeconomic level?

Answer:

- **nonlinear models** that can be estimated via nonlinear least squares (NLS)
- allow for **multiple parameter changes**
- assume changes occur at **unknown dates**

- provide an **estimation method for change points and parameters**, derive their asymptotic distributions
- propose **several stability tests**.
Consider the nonlinear model: \( Y = f(X, \theta) + U; \)

**Example 1:** in representative agent models, \( Y \) could be the consumption growth, \( X \) might include income growth and interest rates, and \( \theta \) could include a tax parameter that changes over time;

**Example 2:** in partial adjustment models such as inventory models, \( Y \) could be the current change in inventories, \( X \) might include the gap between desired and actual past inventories, and \( \theta \) could include the accelerator parameter that might be unstable over time.
1. Model
2. Assumptions
3. Asymptotics
4. Stability Tests
5. Simulations
6. Conclusions
Model

\[ y_t = f(x_t, \theta_{i+1}^0) + u_t \quad t = [T_i^0 + 1, T_{i+1}^0] \quad i = 0, 1, \ldots, m. \]

- \( T_i^0 \) are unknown, \( m = \) known
- \( \theta_i^0 \) is a \( p \times 1 \) vector
- \( E[u_t|x_t] = 0, \ u_t \) i.i.d.

Estimation of break dates and parameters:

- as in Bai and Perron (1998), minimize sum of squared residuals
- over all possible partitions of the \([1, T]\) interval
- and over the parameters defined for each partition.
Estimators:

- \((\hat{T}_1, \ldots, \hat{T}_m) = \operatorname{argmin}_{(T_1, \ldots, T_m)} \operatorname{argmin}_{(\theta_1, \ldots, \theta_{m+1})} T^{-1} \sum_{i=0}^{m} \sum_{t=T_i+1}^{T_{i+1}} [y_t - f(x_t, \theta_{i+1})]^2\)

where \((T_1, \ldots, T_m)\) are possible \(m\)-partitions of the \([1, T]\) interval.

- \(\hat{\theta}_i = \hat{\theta}_i(\hat{T}_1, \ldots, \hat{T}_m)\)

where \(\hat{\theta}_i(\hat{T}_1, \ldots, \hat{T}_m)\) are NLS estimators for a given partition.
Assumptions

- **A1: Break Fractions**
  \[ T_i^0 = [T \lambda_i^0], \text{ where } 0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1. \]

- **A2: Parameter Space**
  \( \Theta \) is a compact and convex subset of \( \mathbb{R}^p \).

- **A3: Underlying Memory of Processes**
  Let \( f_t(\theta) = f(x_t, \theta) \) and \( \psi_t(\theta) = u_t f_t(\theta) \).

  - (i) Assume \( f_t(\theta) \) and \( \psi_t(\theta) \) are strictly stationary processes, \( \beta \)-mixing, where the \( \beta \)-mixing coefficients, \( \beta_1 \) for \( f_t(\theta) \) and \( \beta_2 \) for \( \psi_t(\theta) \), satisfy \( \beta_i(s) \leq D_i s_i^{-A} \), with \( D_i > 0 \) and also \( A_i > 2 + 4 \xi_i \), for some \( \xi_i > 0 \), \( i = 1, 2 \);

  - (ii) \( \sup_{\theta} E|f_t(\theta)|^{2+\delta_1} < \infty \), \( \sup_{\theta} E|\psi_t(\theta)|^{2+\delta_2} < \infty \) for some \( \delta_i > 0 \), \( i = 1, 2 \).
Assumptions

Question Model Assumptions Asymptotics Tests Simulations Conclusions

- **A4: Smoothness**
  
  Let \( F_t(\theta) = \partial f(x_t, \theta) / \partial \theta \).

  - (i) \( f(x_t, \theta) \) is twice continuously differentiable in \( \Theta \), for each \( x_t \), where \( E[\sup_\theta f_t(\theta)]^2 \), \( E[\sup_\theta F_t(\theta)] \) and \( E[\sup_\theta \partial F_t(\theta)/\partial \theta'] \) exist and are bounded;

  - (ii) \( T^{-1} \sum_{t=1}^{[Tr]} F_t(\theta) F_t(\theta)' \xrightarrow{p} rW(\theta) \), a positive definite matrix of constants, uniformly in \( \theta \times r \).

- **A5: Error Process**

  Let \( u_t(\theta) = y_t - f_t(\theta) \).

  - (i) \( E[\sup_\theta |u_t(\theta)| \text{ given } x_t] < \infty \);

  - (ii) \( u_t \ i.i.d. \);

  - (iii) \( E[u_t | x_t] = 0 \ and \ Var[u_t | x_t] = \sigma^2 < \infty \);

  - (iv) \( E[\sup_t |u_t|] < \infty \).
**A6: Break Identification**

\[ E[f_t(\theta_j^0)] \neq E[f_t(\theta_{j+1}^0)] \text{ for each } j = 1, 2, \ldots, m. \]

**A7: Parameter Identification**

\[ \bar{S}(\theta_1, \ldots, \theta_{m+1}) = (m + 1)\sigma^2 + \sum_{i=1}^{m+1} [\lambda_i^0 - \lambda_{i-1}^0] E[f_t(\theta_i) - f_t(\theta_i^0)]^2 \]

has a unique minimizer at \((\theta_1^0, \ldots, \theta_{m+1}^0)\).
Linear vs. nonlinear models:

- similarity:
  \[ \text{OLS} = (X'X)^{-1}X'y; \quad \text{NLS} = (F'F)^{-1}F'y + o_p(T^{-1/2}). \]
  where \( F = \frac{\partial f(X, \theta^0)}{\partial \theta}. \)

- difference in our setting:
  the above approximation cannot be legitimately performed prior to obtaining \( T \)-rate consistent estimators of the change points.

MAIN RESULT 1: Consistency of Break Fractions

Let the estimated break-fractions \( \hat{\lambda}_i \) be such that \( \hat{T}_i = [T\hat{\lambda}_i] \).

- Under A1-A5: \( \hat{\lambda}_i \xrightarrow{p} \lambda^0_i \), for \( i = 1, \ldots, m. \)
Consistency of Break Fraction Estimates:

- By means of two lemmas:

- Let \( d_t = \hat{\lambda}_t - \lambda_t \), and use inequality:

\[
T^{-1} \sum_{t=1}^{T} \hat{\lambda}_t^2 = T^{-1} \sum_{t=1}^{T} \lambda_t^2 + T^{-1} \sum_{t=1}^{T} d_t^2 + 2T^{-1} \sum_{t=1}^{T} d_t \lambda_t \leq T^{-1} \sum_{t=1}^{T} \lambda_t^2
\]

- Lemma 1. \( T^{-1} \sum_{t=1}^{T} d_t \lambda_t \overset{p}{\to} 0 \).

- Lemma 2. If estimated break fraction \( \hat{\lambda}_j \overset{p}{\to} \lambda_j^0 \) for some \( j \)

\[
\Rightarrow \limsup P \left[ T^{-1} \sum_{t=1}^{T} d_t^2 > C \right] > \epsilon, \text{ for some } C, \epsilon > 0.
\]
Lemma 1. \( T^{-1} \sum_{t=1}^{T} d_t u_t \xrightarrow{p} 0. \)

- Note that:

\[
T^{-1} \sum_{t=1}^{T} u_t d_t = T^{-1} \sum_{i=0}^{m} \sum_{T_i}^{T_i+1} u_t f(x_t, \theta_i^0) - T^{-1} \sum_{i=0}^{m} \sum_{\hat{T}_i}^{\hat{T}_i+1} u_t f(x_t, \hat{\theta}_i)
\]

\[A
\]

\[B\]

- \( A = o_p(1) \) by pointwise laws of large numbers.
- to show \( B = o_p(1) \), we borrow a proof by Caner (2005), that uses empirical process theory.
**Lemma 2.** If estimated break fraction $\hat{\lambda}_j \overset{p}{\rightarrow} \lambda_j^0$ for some $j$

$\Rightarrow \lim \sup P \left[ T^{-1} \sum_{t=1}^{T} d_t^2 > C \right] > \epsilon$, for some $C, \epsilon > 0$.

From inequality below, Lemma 2 follows:

$$T^{-1} \sum_{t=1}^{T} d_t^2 \geq T^{-1} \sum_{1}^{T} [f_t(\hat{\theta}_k) - f_t(\theta_j^0)]^2 + T^{-1} \sum_{2}^{T} [f_t(\hat{\theta}_k) - f_t(\theta_{j+1})]^2.$$ 

By contradiction, we get consistency of break-point estimates.
MAIN RESULT 2: T-Rate Convergence of Break Fractions

Under A1-A6, for every $\eta > 0$, there exists a finite $C > 0$, such that for all large $T$, $P(| T(\hat{\lambda}_k - \lambda^0_k) | > C) < \eta$ ($k = 1, \ldots, m$).

- crucial result because we will encounter in all estimation / inference procedures sums of the form:

\[
T^{-1} \sum_{[T\hat{\lambda}_i] + 1}^{[T\hat{\lambda}_{i+1}]} O_p(1), \quad \text{but we need} \quad T^{-1} \sum_{[T\lambda^0_i] + 1}^{[T\lambda^0_{i+1}]} O_p(1).
\]

- the result above allows us to approximate the first sum with the second

- if we couldn’t do so, then the difference between those two sums would not disappear in the limit as $T$ grows large.
MAIN RESULT 3: Asymptotic Normality

Under A1-A7, \( T^{1/2}(\hat{\theta} - \theta^0) \xrightarrow{d} \mathcal{N}(0, [W(\theta^0)]^{-1}) \), where \([W(\theta^0)]\) is a block diagonal matrix whose i-i-th block is \( \sigma^2[\lambda_i^0 - \lambda_{i-1}^0] E[F_t(\theta_i^0)'F_t(\theta_i^0)] \).

- we obtain normality of parameters if we consistently estimate the break fractions, at a T-rate;
- to show this result, we use mean value expansions of partial sums of squares, where the end points of these sums are the estimated change points;
- the exciting part is that given the T-rate convergence of break fractions, we can replace the estimated change points with the true ones;
- even in unstable nonlinear models of type NLS, we can find the breaks and estimate the parameters we need.
Similar to Bai and Perron (1998), we propose 3 classes of test that detect instability.

These are hypotheses of interest:

- **Test 1**: no breaks vs. a known number of breaks;

- **Test 2**: no breaks vs. an unknown number of breaks;

- **Test 3**: l vs. l+1 breaks.

Tests have non-standard distributions, but they carry over from the linear setting of Bai and Perron (1998), where critical values can be found.
Computation:

- as Bai and Perron (2003) show, independent of the number of breaks, we only need to search over $\frac{T(T+1)}{2}$ partitions of the sample;

- furthermore, we need to bound the candidate change points away from the end-points of the sample (cut-offs usually 5%-15%);

- by doing so, we further reduce the number of partitions we need to search over.
1. A Test of No Break vs. A Known Number of Breaks

- **Hypothesis:**
  \[ H_0 : m = 0 \; \text{vs.} \; H_A : m = k \]

- **Sup F-type Test:**
  \[
  \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} F_T(k; p) = \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} \frac{(SSR_0 - SSR_k)/kp}{SSR_k/[T - (k+1)p]}
  \]

  where \( \Lambda_\epsilon = \{ (\lambda_1, \ldots, \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon \} \)

- **Distribution under the Null:**
  \[
  \frac{1}{kp} \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} \sum_{i=1}^k \frac{\|\lambda_i W_p(\lambda_{i+1}) - \lambda_{i+1} W_p(\lambda_i)\|^2}{\lambda_i \lambda_{i+1} (\lambda_{i+1} - \lambda_i)}
  \]

- The test is consistent for its alternative;
- **it does not depend on nuisance parameters.**
2. A Test of No Break vs. A Unknown Number of Breaks

■ Hypothesis:
\[ H_0 : m = 0 \quad \text{vs.} \quad H_A : m \ \text{unknown}, \ m < M, M \ \text{fixed} \]

■ Double Maximum Test:
\[ D \max F_T(M, a_1, \ldots, a_M) = \max_{1 \leq m \leq M} a_m \sup_{\Lambda_\epsilon} F_T(m; p) \]

■ Distribution under the Null:
\[ \max_{1 \leq m \leq M} \frac{a_m}{kp} \sup_{(\lambda_1, \ldots, \lambda_k) \in \Lambda_\epsilon} \sum_{i=1}^k \frac{\|\lambda_i W_p(\lambda_{i+1}) - \lambda_{i+1} W_p(\lambda_i)\|^2}{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)} \]

■ Test is consistent for its alternative;
■ depends on choice of weights.
2. A Test of No Break vs. A Unknown Number of Breaks

Choice of weights:

- equal weights over the possible number of breaks;
- give more weight to some number of breaks according to some prior;
- since for any fixed number of parameters $p$, the critical values of $\sup_{(\lambda_1,\ldots,\lambda_k) \in \Lambda_\epsilon} F_T(m; p)$ decreases as $m$ increases, this implies that if we have a large number of breaks, we may get a test with low power, because the marginal p-values decrease with $m$;
- one way to keep marginal p-values of the tests equal across number of breaks is to use weights that depend on $p$ and the significance level of the test
- for example, let $c(p, \alpha, m)$ be the asymptotic critical value of the test $\sup_{(\lambda_1,\ldots,\lambda_k) \in \Lambda_\epsilon} F_T(m; p)$ and assign:
  
  $a_1 = 1$ and $a_m = c(p, \alpha, 1) / c(p, \alpha, m)$ for $1 < m \leq M$. 

3. Test for an Additional Break

- **Hypothesis:**
  \[ H_0 : m = l \quad \text{vs.} \quad H_A : m = l + 1. \]

- **Test each \((l+1)\)-segment for an additional break by means of:**
  \[
  \left\{ S_T(\hat{T}_1, \ldots, \hat{T}_l) - \min_{1 \leq i \leq l+1} \inf_{\tau} S_T(\hat{T}_1, \ldots, \hat{T}_{i-1}, \tau, \hat{T}_i, \ldots, \hat{T}_l) \right\} / \hat{\sigma}^2
  \]

- **Distribution under the Null:**
  \[
  \lim P(F_T(l + 1|l) \leq x) = G_{p,\eta}^{l+1}
  \]
  where \( G_{p,\eta} \) is the cdf of \( \sup_{\eta \leq \mu \leq 1-\eta} \frac{\|W_q(\mu) - \mu W_q(1)\|^2}{\mu(1-\mu)} \).

- **Test is consistent and provides insight for constructing sequential rather than global methods for estimation.**
3. Sequential Estimation of Break-Points

- if there is evidence for one break, then estimate it and split sample into 2 parts;

- if there is evidence of an additional break, then search each sub-sample to find the estimated second break;

- iterate procedure until there is no evidence of an additional break.
Model:

\[
f(x_t, \theta) = \theta_1^t + \theta_2^t e^{x_t \theta_3^t} \quad t \in [T^0_i, T^0_{i+1}].
\]

Table 1: \(m = 1, 100\) simulations, break fraction: 0.4

<table>
<thead>
<tr>
<th>T</th>
<th>True breaks</th>
<th>MC Est Breaks</th>
<th>(\hat{\theta}_1) before</th>
<th>(\hat{\theta}_1) after</th>
<th>(\hat{\theta}_2) before</th>
<th>(\hat{\theta}_2) after</th>
<th>(\hat{\theta}_3) before</th>
<th>(\hat{\theta}_3) after</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>12</td>
<td>12.06</td>
<td>1.21 (1.12)</td>
<td>-1.05 (.74)</td>
<td>-10.14 (1.58)</td>
<td>10.08 (.81)</td>
<td>1.12 (1.33)</td>
<td>-.99 (.05)</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>20.09</td>
<td>1.05 (.91)</td>
<td>-1.12 (.76)</td>
<td>-9.93 (.92)</td>
<td>10.06 (.61)</td>
<td>1.00 (.06)</td>
<td>-.99 (.02)</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>40.01</td>
<td>1.00 (.42)</td>
<td>-.99 (.35)</td>
<td>-9.98 (.44)</td>
<td>10.02 (.37)</td>
<td>1.00 (.02)</td>
<td>-.99 (.02)</td>
</tr>
</tbody>
</table>
Model:

\[ f(x_t, \theta) = \theta_i^1 + \theta_i^2 e^{x_t \theta_i^3} \quad t \in [T_i^0, T_{i+1}^0]. \]

Table 2: \( m = 2 \), 100 simulations, break fractions: [0.4, 0.7]

<table>
<thead>
<tr>
<th>T</th>
<th>True breaks</th>
<th>MC Estimated breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>[0.40, 0.70]</td>
<td>0.399 0.70</td>
</tr>
<tr>
<td>50</td>
<td>0.40</td>
<td>0.40 0.70</td>
</tr>
<tr>
<td>50</td>
<td>0.40</td>
<td>0.40 0.70</td>
</tr>
</tbody>
</table>
■ we provided a comprehensive treatment of estimation and inference in NLS models similar to Bai and Perron (1998) results for linear models;

■ **key difference** comes from using a mean value expansion rather than an exact formula for parameter estimates;

■ as a consequence of nonlinearity, we use *nonlinear asymptotics* and *empirical process theory*;

■ the method we develop is useful for detecting breaks in nonlinear economic models, by means of proposed tests;

■ it also offers a solution on estimating the model when breaks are present.
derive asymptotic distributions of change point estimates and Wald-like tests;

study finite sample behavior of estimates and tests;

extend to more general nonlinear models;

use it for estimating consumption models in the presence of tax changes and inventory models in the presence of flexible accelerator instability.
Thank You!