Lectures on Structural Change

Eric Zivot
Department of Economics, University of Washington

April 4, 2001

1 Overview of Testing for and Estimating Structural Change in Econometric Models

2 Some Preliminary Asymptotic Theory

Reference: Stock, J.H. (1994) "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics, Vol. IV.*

3 Tests of Parameter Constancy in Linear Models

3.1 Rolling Regression

Reference: Zivot, E. and J. Wang (2002). *Modeling Financial Time Series with S-PLUS*. Springer-Verlag, New York.

For a window of width $k < n < T$, the rolling linear regression model is

$$\underset{(n\times 1)}{\mathbf{y}_t(n)} = \underset{(n\times k)}{\mathbf{X}_t(n)}\underset{(k\times 1)}{\boldsymbol{\beta}_t(n)} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}_t(n)},\ t = n, \ldots, T$$

- Observations in $y_t(n)$ and $X_t(n)$ are n most recent values from times $t - n + 1$ to t
- OLS estimates are computed for sliding windows of width n and increment m
- Poor man's time varying regression model

3.1.1 Application: Simulated Data

Consider the linear regression model

$$\begin{aligned} y_t &= \alpha + \beta x_t + \varepsilon_t,\ t = 1, \ldots, T = 200 \\ & x_t \sim iid\ N(0, 1) \\ & \varepsilon_t \sim iid\ N(0, \sigma^2) \end{aligned}$$

No structural change parameterization: $\alpha = 0, \beta = 1, \sigma = 0.5$

Structural change cases

- Break in intercept: $\alpha = 1$ for $t > 100$
- Break in slope: $\beta = 3$ for $t > 100$
- Break in variance: $\sigma = 0.25$ for $t > 100$
- Random walk in slope: $\beta = \beta_t = \beta_{t-1} + \eta_t$, $\eta_t \sim iid\ N(0, 0.1)$ and $\beta_0 = 1$.

(show simulated data)

3.1.2 Application: Exchange rate regressions

References:

1. Bailey, R. and T. Bollerslev (???)
2. Sakoulis, G. and E. Zivot (2001). "Time Variation and Structural Change in the Forward Discount: Implications for the Forward Rate Unbiasedness Hypothesis," unpublished manuscript, Department of Economics, University of Washington.

Let

$$\begin{aligned} s_t &= \text{log of spot exchange rate in month } t \\ f_t &= \text{log of forward exchange rate in month } t \end{aligned}$$

The forward rate unbiased hypothesis is typically investigated using the so-called differences regression

$$\begin{aligned} \Delta s_{t+1} &= \alpha + \beta(f_t - s_t) + \varepsilon_{t+1} \\ f_t - s_t &= \text{forward discount} \end{aligned}$$

If the forward rate f_t is an unbiased forecast of the future spot rate s_{t+1} then we should find

$$\alpha = 0 \text{ and } \beta = 1$$

The forward discount is often modeled as an AR(1) model

$$f_t - s_t = \delta + \phi(f_{t-1} - s_{t-1}) + u_t$$

Statistical Issues

- Δs_{t+1} is close to random walk with large variance
- $f_t - s_t$ behaves like highly persistent AR(1) with small variance
- $f_t - s_t$ appears to be unstable over time

Tasks:

- compute rolling regressions for 24-month windows incremented by 1 month
- compute rolling regressions for 48-month windows incremented by 12 months

(insert rolling regression graphs here)

3.2 Chow Forecast Test

Reference: Chow, G.C. (1960). "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 52, 211-22.

Consider the linear regression model with k variables

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta} + u_t,\ u_t \sim (0, \sigma^2),\ t = 1, \ldots, n \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \end{aligned}$$

Parameter constancy hypothesis

$$H_0 : \boldsymbol{\beta} \text{ is constant}$$

Chow forecast test intuition

- If parameters are constant then out-of-sample forecasts should be unbiased (forecast errors have mean zero)

Chow forecast test construction:

- Split sample into $n_1 > k$ and $n_2 = n - n_1$ observations

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix}$$

- Fit model using first n_1 observations

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ \hat{\mathbf{u}}_1 &= \mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 \\ \hat{\sigma}_1^2 &= \hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1/(n_1 - k) \end{aligned}$$

- Use $\hat{\boldsymbol{\beta}}_1$ and $\mathbf{X}_2$ to predict $\mathbf{y}_2$ using next n_2 observations

$$\hat{\mathbf{y}}_2 = \mathbf{X}_2\hat{\boldsymbol{\beta}}_1$$

- Compute out-of-sample prediction errors

$$\hat{\mathbf{u}}_2 = \mathbf{y}_2 - \hat{\mathbf{y}}_2 = \mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_1$$

Under $H_0 : \boldsymbol{\beta}$ is constant

$$\hat{\mathbf{u}}_2 = \mathbf{u}_2 - \mathbf{X}_2(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})$$

and

$$\begin{aligned} E[\hat{\mathbf{u}}_2] &= \mathbf{0} \\ var(\hat{\mathbf{u}}_2) &= \sigma^2 \left(\mathbf{I}_{n_2} + \mathbf{X}_2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_2'\right) \end{aligned}$$

Further, If the errors u are Gaussian then

$$\hat{\mathbf{u}}_2 \sim N(\mathbf{0}, var(\hat{\mathbf{u}}_2))$$
$$\hat{\mathbf{u}}_2' var(\hat{\mathbf{u}}_2)^{-1}\hat{\mathbf{u}}_2 \sim \chi^2(n_2)$$
$$(n_1 - k)\hat{\sigma}_1^2/\sigma^2 \sim \chi^2(n_1 - k)$$

This motivates the Chow forecast test statistic

$$Chow_{FCST}(n_2) = \frac{\hat{\mathbf{u}}_2' \left(\mathbf{I}_{n_2} + \mathbf{X}_2(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_2'\right)\hat{\mathbf{u}}_2}{n_2\hat{\sigma}_1^2} \sim F(n_2, n_1 - k)$$

Remarks:

- Test is a general specification test for unbiased forecasts
- Implementation requires *a priori* splitting of data into fit and forecast samples

3.2.1 Application: Exchange rate regression cont'd

(show Eviews output for Chow forecast test)

3.3 CUSUM and CUSUMSQ Tests

Reference: Brown, R.L., J. Durbin and J.M. Evans (1975). "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society*, Series B, 35, 149-192.

3.3.1 Recursive least squares estimation

The *recursive least squares* (RLS) estimates of $\boldsymbol{\beta}$ are based on estimating

$$y_t = \boldsymbol{\beta}_t'\mathbf{x}_t + \xi_t, \; t = 1, \ldots, n$$

by least squares recursively for $t = k+1, \ldots, n$ giving $n-k$ least squares (RLS) estimates $(\hat{\boldsymbol{\beta}}_{k+1}, \ldots, \hat{\boldsymbol{\beta}}_T)$.

- RLS estimates may be efficiently computed using the Kalman Filter
- If $\boldsymbol{\beta}$ is constant over time then $\hat{\boldsymbol{\beta}}_t$ should quickly settle down near a common value.
- If some of the elements in $\boldsymbol{\beta}$ are not constant then the corresponding RLS estimates should show instability. Hence, a simple graphical technique for uncovering parameter instability is to plot the RLS estimates $\hat{\beta}_{it}$ $(i = 1, 2)$ and look for instability in the plots.

3.3.2 Recursive residuals

Formal tests for structural stability of the regression coefficients may be computed from the standardized $1 - step$ ahead recursive residuals

$$\begin{aligned} w_t &= \frac{v_t}{\sqrt{f_t}} = \frac{y_t - \hat{\boldsymbol{\beta}}_{t-1}'\mathbf{x}_t}{\sqrt{f_t}} \\ f_t &= \hat{\sigma}^2 \left[1 + \mathbf{x}_t'(\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{x}_t\right] \end{aligned}$$

Intuition:

- If β changes in the next period then the forecast error will not have mean zero

3.3.3 CUSUM statistic

The CUSUM statistic of Brown, Durbin and Evans (1975) is

$$\begin{aligned} CUSUM_t &= \sum_{j=k+1}^{t} \frac{\hat{w}_j}{\hat{\sigma}_w} \\ \hat{\sigma}_w^2 &= \frac{1}{n-k} \sum_{t=1}^{n} (w_t - \bar{w})^2 \end{aligned}$$

Under the null hypothesis that $\boldsymbol{\beta}$ is constant, $CUSUM_t$ has mean zero and variance that is proportional to $t - k - 1$.

3.3.4 Application: Simulated Data

(insert graphs here)

Remarks

- Ploberger and Kramer (1990) show the CUSUM test can be constructed with OLS residuals instead of recursive residuals
- CUSUM Test is essentially a test to detect instability in intercept alone
- CUSUM Test has power only in direction of the mean regressors
- CUSUMSQ test does not have good power for changing coefficients but has power for changing variance

3.3.5 Application: Exchange Rate Regression cont'd

(show recursive estimates, CUSUM and CUSUMSQ plots)

3.4 Nyblom's Parameter Stability Test

Reference:NYBLOM, J. (1989). "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84 (405), 223-230.

Consider the linear regression model with k variables

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$$

The time varying parameter (TVP) model assumes

$$\boldsymbol{\beta} = \boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t,\ \boldsymbol{\eta}_t \sim (0, \sigma_\eta^2)$$

Note: the TVP model nests the single break model by setting $\eta_t = \gamma, t = r + 1$ and $\eta_t = 0$ otherwise.

The hypotheses of interest are then

$$\begin{aligned} H_0 &: \boldsymbol{\beta} \text{ is constant } \Leftrightarrow \sigma_\eta^2 = 0 \\ H_1 &: \sigma_\eta^2 > 0 \end{aligned}$$

Nyblom (1989) derives the locally best invariant test as the Lagrange multiplier test

$$\begin{aligned} L &= tr\left[\mathbf{S}_n^{-1}\sum_{j=1}^{n}\left(\sum_{k=j}^{n}\hat{\varepsilon}_k\mathbf{x}_k\right)\left(\sum_{k=j}^{n}\hat{\varepsilon}_k\mathbf{x}_k'\right)\right]/\hat{\sigma}_\varepsilon^2 > c \\ \hat{\varepsilon}_t &= y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}},\ \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \mathbf{S}_n &= n^{-1}\mathbf{X}'\mathbf{X} \end{aligned}$$

Under mild assumptions regarding the behavior of the regressors, the limiting distribution of L under the null is a Camer-von Mises distribution:

$$\begin{aligned} L &\Rightarrow \int_0^1 \mathbf{B}_k^\mu(\lambda)\mathbf{B}_k^\mu(\lambda)d\lambda \\ \mathbf{B}_k^\mu(\lambda) &= \mathbf{W}_k(\lambda) - \lambda\mathbf{W}_k(1) \\ \mathbf{W}_k(\lambda) &= \text{k dimensional Brownian motion} \end{aligned}$$

Remarks:

- Test statistic is a sum of cumulative sums of weighted full sample residuals
- Test is for constancy of all parameters
- Test is applicable for models estimated by methods other than OLS
- Distribution of L is different if x_t is non-stationary (unit root, deterministic trend)

3.4.1 Application: Simulated Data

Nyblom Test

Model	L_c
No SC	.332
Mean shift	13.14^{***}
Slope shift	14.13^{***}
var shift	.351
RW slope	9.77^{***}

3.4.2 Application: Exchange rate regression cont'd

Nyblom Test

Model	L_c
AR(1)	1.27^{***}
Diff reg	.413

3.5 Hansen's Parameter Stability Tests

References

1. Hansen, B.E. (1992). "Testing for Parameter Instability in Linear Models "*Journal of Policy Modeling*, 14(4), 517-533.

2. Hansen, B.E. (1992). "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321-336.

Same set-up as Nyblom's LM test. Under the null of constant parameters, the score vector from the linear model with Gaussian errors is based on the normal equations

$$\begin{aligned}
\sum_{t=1}^{n} x_{it}\hat{\varepsilon}_t &= 0,\ i=1,\ldots,k \\
\sum(\hat{\varepsilon}_t^2 - \hat{\sigma}^2) &= 0 \\
\hat{\varepsilon}_t &= y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}} \\
\hat{\sigma}^2 &= n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t^2
\end{aligned}$$

Define

$$\begin{aligned}
f_{it} &= \begin{cases} x_{it}\hat{\varepsilon}_t & i=1,\ldots k \\ \hat{\varepsilon}_t^2 - \hat{\sigma}^2 & i=k+1 \end{cases} \\
S_{it} &= \sum_{j=1}^{t} f_{ij},\ i=1,\ldots,k+1
\end{aligned}$$

Note that

$$\sum_{i=1}^{n} f_{it} = 0, \; i = 1, \ldots, k+1$$

Hansen's LM test for

$$H_0 : \beta_i \text{ is constant, } i = 1, \ldots, k$$

and for

$$H_0 : \sigma^2 \text{ is constant}$$

is

$$\begin{aligned} L_i &= \frac{1}{nV_i} \sum_{t=1}^{n} S_{it}^2, \; i = 1, \ldots, k \\ V_i &= \sum_{t=1}^{n} f_{it}^2 \end{aligned}$$

Under the null of no structural change

$$L_i \Rightarrow \int_0^1 B_1^{\mu}(\lambda) B_1^{\mu}(\lambda) d\lambda$$

For testing the joint hypothesis

$$H_0 : \boldsymbol{\beta} \text{ and } \sigma^2 \text{ are constant}$$

define the $(k+1) \times 1$ vectors

$$\begin{aligned} \mathbf{f}_t &= (f_{1t}, \ldots, f_{k+1,t})' \\ \mathbf{S}_t &= (S_{1t}, \ldots, S_{k+1,t})' \end{aligned}$$

Hansen's LM statistic for testing the constancy of all parameters is

$$\begin{aligned} L_c &= \frac{1}{n} \sum_{t=1}^{n} \mathbf{S}_t' \mathbf{V}^{-1} \mathbf{S}_t = \frac{1}{n} tr \left(\mathbf{V}^{-1} \sum_{t=1}^{n} \mathbf{S}_t \mathbf{S}_t' \right) \\ \mathbf{V} &= \sum_{t=1}^{n} \mathbf{f}_t \mathbf{f}_t' \end{aligned}$$

Under the null of no-structural change

$$L_c \Rightarrow \int_0^1 \mathbf{B}_{k+1}^{\mu}(\lambda) \mathbf{B}_{k+1}^{\mu}(\lambda) d\lambda$$

Remarks

- Tests are very easy to compute and are robust to heteroskedasticity

- Null distribution is non-standard and depends upon number of parameters tested for stability
- 5% critical value for individual tests is 0.470
- Tests are not informative about the date of structural change
- Hansen's L_1 test for constancy of intercept is analogous to the CUSUM test
- Hansen's L_{k+1} test for constancy of variance is analogous to CUSUMSQ test
- Hansen's L_c test for constancy of all parameters is similar to Nybolom's test
- Distribution of tests is different if data are nonstationary (unit root, deterministic trend) - see Hansen (1992), JBES.

3.5.1 Application: Simulated Data

Hansen Tests

Model	α	β	σ^2	Joint
No SC	.179	.134	.248	.503
Mean shift	13.19***	.234	.064	13.3***
Slope shift	.588	5.11***	.067	5.25***
var shift	.226	.119	.376*	.736
RW slope	.253	4.08***	.196	4.4***

3.5.2 Application: Exchange rate regression cont'd

Hansen Tests

Model	intercept	slope	variance	Joint
AR(1)	.382	.147	2.94***	3.90***
Diff reg	.104	.153	.186	.520

4 Tests for Single Structural Change

Consider the linear regression model with k variables

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_t + \varepsilon_t,\ t = 1 \ldots, n$$

No structural change null hypothesis

$$H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}$$

Single break date alternative hypothesis

$$
\begin{aligned}
H_1 \quad &: \quad \begin{cases} \boldsymbol{\beta}_t = \boldsymbol{\beta},\ t \leq m = \text{break date} \\ \boldsymbol{\beta}_t = \boldsymbol{\beta} + \boldsymbol{\gamma},\ t > m \text{ and } \boldsymbol{\gamma} \neq \mathbf{0} \end{cases} \\
k \quad &< \quad m < T - k \\
\lambda \quad &= \quad \frac{m}{n} = \text{ break fraction}
\end{aligned}
$$

Remarks:

- Under no break null $\boldsymbol{\gamma} = \mathbf{0}$.
- Pure structural change model: all coefficients change ($\gamma_i \neq 0$ for $i = 1, \ldots, k$)
- Partial structural change model: some coefficients change ($\gamma_i \neq 0$ for some i)

4.1 Chow's Test with Known Break Date

Assume: m or λ is known

For a data interval $[r, \ldots, s]$ such that $s - r > k$ define

- $\hat{\boldsymbol{\beta}}_{r,s}$ = OLS estimate of $\boldsymbol{\beta}$
- $\hat{\boldsymbol{\varepsilon}}_{r,s}$ = OLS residual vector
- $SSR_{r,s} = \hat{\boldsymbol{\varepsilon}}_{r,s}'\hat{\boldsymbol{\varepsilon}}_{r,s}$ = sum of squared residuals

Chow's breakpoint test for testing H_0 vs. H_1 with m known is

$$
F_n\left(\frac{m}{n}\right) = F_n(\lambda) = \frac{(SSR_{1,n} - (SSR_{1,m} + SSR_{m+1,n}))/k}{(SSR_{1,m} + SSR_{m+1,n})/(n-2k)}
$$

The Chow test may also be computed as the F-statistic for testing $\boldsymbol{\gamma} = \mathbf{0}$ from the dummy variable regression

$$
\begin{aligned}
y_t &= \mathbf{x}_t'\boldsymbol{\beta} + D_t\mathbf{x}_t'\boldsymbol{\gamma} + \varepsilon_t \\
D_t &= 1 \text{ if } t > m; 0 \text{ otherwise}
\end{aligned}
$$

Under $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ with m known

$$
F_n(\lambda) \xrightarrow{d} \chi^2(k)
$$

4.1.1 Application: Simulated Data cont'd

(insert table here)

4.2 Quandt's LR Test with Unknown Break Date

References:

1. Quandt, R.E. (1960). "Tests of Hypotheses that a Linear System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

2. Davies, R.A. (1977). "Hypothesis Testing When a Nuisance Parameter is Present only Under the Alternative," *Biometrika*, 64, 247-254.

3. Kim, H.-J., and D. Siegmund (1989). "The Likelihood Ratio Test for a Change-Point in Simple Linear Regression," *Biometrika*, 76, 3, 409-23.

Assume: m or λ is unknown.

Quandt considered the LR statistic for testing $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$ when m is unknown. This turns out to be the maximal $F_n(\lambda)$ statistic over a range of break dates $m_0, \ldots, m_1$:

$$
\begin{aligned}
QLR &= \max_{m \in [m_0, m_1]} F_n\left(\frac{m}{n}\right) = \max_{\lambda \in [\lambda_0, \lambda_1]} F_n(\lambda) \\
\lambda_i &= \frac{m_i}{n},\ i = 0, 1
\end{aligned}
$$

Remarks

- Implicitly, the break data m and break fraction λ are estimated using

$$
\begin{aligned}
\hat{m} &= \arg\max_{m}\ F_n\left(\frac{m}{n}\right) \\
\hat{\lambda} &= \arg\max_{\lambda}\ F_n(\lambda)
\end{aligned}
$$

- Under the null, m defined under the alternative is not identified. This is an example of the "Davies problem".

- Davies (1977) showed that that if parameters are unidentified under the null, standard χ^2 inference does not obtain.

Under $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, Kim and Siegmund (1989) showed

$$
\begin{aligned}
QLR &\Rightarrow \sup_{\lambda \in [\lambda_0, \lambda_1]} \frac{\mathbf{B}_k^{\mu}(\lambda)\mathbf{B}_k^{\mu}}{\lambda(1-\lambda)} \\
\mathbf{B}_k^{\mu}(\lambda) &= \mathbf{W}_k(\lambda) - \lambda \mathbf{W}_k(1) = \text{ Brownian Bridge}
\end{aligned}
$$

Remarks

- Distribution of QLR is non-standard and depends on the number of variables k and the trimming parameters λ_0 and λ_1

- Critical values computed by simulation are given in Andrews (1993), and are larger than $\chi^2(k)$ critical values:

5% critical values		
k	$\chi^2(k)$	QLR
1	3.84	8.85
10	18.3	27.03

4.2.1 Application: Simulated Data

Construction of F-statistics

(insert graphs here)

4.3 Andrew's Tests with Unknown Break Date

Reference: Andrews, D.W.K. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 59, 817-858.

4.3.1 Application: AR(1) process with change in AR coefficient cont'd

4.4 Empirical Application

Reference: Stock and Watson (199?), " " *Journal of Business and Economic Statistics*.

5 Day 2: Estimation of Models with One Structural Change

Reference: Bai, J. (1994). "Least Squares Estimation of a Shift in Linear Process," *Journal of Time Series Analysis*

5.1 Abstract

- This paper considers a mean shift with an unknown shift point in a linear process and estimates the unknown shift point by the method of least squares.
- Pre-shift and post-shift means are estimated concurrently with the change point.
- The consistency and the rate of convergence for the estimated change point are established.

- The asymptotic distribution for the change point estimator is obtained when the magnitude of shift is small. It is shown that serial correlation affects the variance of the change point estimator via the sum of the coefficients of the linear process.

5.2 The Mean Shift Model

Consider a time series Y_t that undergoes a mean shift at an unknown time:

$$\begin{aligned} Y_t &= \mu_t + X_t,\ t = \ldots, -2, -1, 0, 1, 2, \ldots \\ X_t &= a(L)\varepsilon_t \\ a(L) &= \sum_{j=0}^{\infty} a_j L^j,\ a(1) \neq 0 \end{aligned}$$

where

$$\mu_t = \begin{matrix} \mu_1 \text{ if } t \leq k_0 \\ \mu_2 \text{ if } t > k_0 \end{matrix}$$

and μ_1, μ_2 and k_0 are unknown parameters and k_0 is the change point.

Estimation problem: estimate μ_1, μ_2 and k_0 given T observations on Y_t. If X_t has an ARMA representation then it is also of interest to estimate the ARMA parameters. Define

$$\begin{aligned} k_0 &= [\tau T],\ 0 < \tau < 1 \\ \lambda &= \mu_2 - \mu_1 \end{aligned}$$

5.3 Results for Model with Known Break Date

If $k = k_0$ were known, then μ_1 and μ_2 could be consistently estimated by least squares ignoring the dynamics in X_t by solving

$$\min_{\mu_1, \mu_2} \left\{ \sum_{t=1}^{k_0} (Y_t - \mu_1)^2 + \sum_{t=k_0+1}^{T} (Y_t - \mu_2)^2 \right\}$$

Then μ_1 and μ_2 are consistent with limiting distributions

$$\begin{aligned} T^{1/2}(\widehat{\mu}_1 - \mu_1) &\xrightarrow{d} N\left\{0, \tau^{-1} a(1)^2 \sigma^2\right\} \\ T^{1/2}(\widehat{\mu}_2 - \mu_2) &\xrightarrow{d} N\left\{0, (1-\tau)^{-1} a(1)^2 \sigma^2\right\} \end{aligned}$$

5.4 Least Squares Estimation

The least squares (LS) estimator $\widehat{k}$ of the change point k_0 ignoring the dynamics in X_t is defined as

$$\begin{aligned} \widehat{k} &= \arg\min_{k} \left[\min_{\mu_1, \mu_2} \left\{ \sum_{t=1}^{k} (Y_t - \mu_1)^2 + \sum_{t=k+1}^{T} (Y_t - \mu_2)^2 \right\} \right] \\ \widehat{\tau} &= \widehat{k}/T \end{aligned}$$

i.e., the shift point is estimated by minimizing the sum of squares of residuals among all possible sample splits *ignoring the dynamics* in X_t. The LS residual is defined as

$$\widehat{X}_t = Y_t - \widehat{\mu}_1 - (\widehat{\mu}_2 - \widehat{\mu}_1) I_{[t > \widehat{k}]}$$

where $I_{[\cdot]}$ is the indicator function.

The following assumptions are made:

- **Assumption A**:

$$\varepsilon_t \sim iid(0, \sigma^2) \text{ or } \varepsilon_t \sim mds(0, \sigma^2)$$

- **Assumption B**:

$$\sum_{j=0}^{\infty} j |a_j| < \infty$$

5.4.1 Consistency of $\widehat{\tau}$

Proposition 1 ***Theorem 2*** *Under Assumptions A and B, the estimator $\widehat{\tau}$ satisfies*

$$|\widehat{\tau} - \tau| = O_p\left(\frac{1}{T\lambda^2}\right).$$

so that $\hat{\tau}$ is a consistent estimator of τ.

Since $\hat{k} = [\hat{\tau} T]$ it follows that

$$\widehat{k} - k = O_p(\lambda^{-2})$$

so that $\widehat{k}$ is not a consistent estimator for k.

5.5 Limiting Distribution of Break Fraction

- To construct a managable asymptotic distribution for $\widehat{\tau}$ it is necessary to assume that λ depends on T and diminishes as T increases.
- If λ is kept fixed independent of T then the limiting distribution of $\widehat{\tau}$ depends on the distribution of ε_t and on λ in a complicated way.
- The asymptotic distribution for $\widehat{\tau}$ is used to construct an asymptotic confidence interval for τ or k.

The dependence of λ on T is functionalized as
Assumption C:

$$\lambda_T \longrightarrow 0, \ \frac{T^{1/2}\lambda_T}{(\log T)^{1/2}} \longrightarrow \infty$$

.

Theorem 3 *Under Assumptions A, B and C, for every $M < \infty$,*

$$T\lambda_T^2(\widehat{\tau} - \tau) \xrightarrow{d} a(1)\sigma^2 \arg\max_v \left\{ W(v) - \frac{1}{2}|v| \right\}$$

and $W(v)$ is a two-sided Brownian motion on $\Re$.

Remarks:

- τ converges to τ at rate $T \Rightarrow \tau$ is super-consistent
- The asymptotic distribution of τ is non-standard (not normal)
- Scale of limiting distribution depends on autocorrelation in X_t through $\sigma^2 a(1)$
- Confidence intervals for τ may be computed from limiting distribution
- Asymptotic distribution may not be accurate if $\lambda = \mu_2 - \mu_1$ is large

5.6 Limiting Distribution of Estimated Shift Coefficients

Proposition 4 ***Theorem 5***

$$T^{1/2}(\widehat{\mu}_1 - \mu_1) \xrightarrow{d} N\left\{0, \tau^{-1}a(1)^2\sigma^2\right\}$$
$$T^{1/2}(\widehat{\mu}_2 - \mu_2) \xrightarrow{d} N\left\{0, (1-\tau)^{-1}a(1)^2\sigma^2\right\}$$

- Since $\widehat{\tau}$ converges at rate T the limiting distribution of $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are the same as if k_0 were known

5.7 Simulation Results

The simulation model is

$$\begin{aligned} Y_t &= \mu + \lambda I_{[t \geq k_0]} + X_t, \ t = 1, \ldots, T \\ X_t &= \rho X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \\ & \quad \varepsilon_t \sim iid\ N(0,1) \end{aligned}$$

where $T = 100, k_0 = 0.5T, \lambda = 2, \rho = -0.6, 0.0, 0.6$ and $\theta = -0.5, 0.5$. The main results based on $N = 100$ simulations are:

- For fixed θ, the range of $\widehat{k}$ becomes larger as ρ varies from -0.6 to 0.6.
- The range of $\widehat{k}$ is smaller for $\theta = -0.5$ than for $\theta = 0.5$ for every given ρ as predicted by theory.
- The ignored dynamics of the error does not have much effect on the point estimates of the break date.

6 Estimation of Models with Multiple Structural Change

6.1 Estimating Multiple Breaks One-at-a-Time

Reference: Bai, J. (1997). "Estimating Multiple breaks One at a Time," *Econometric Theory.*

- Sequential (one-by-one) rather than simultaneous estimation of multiple breaks is investigated.
- The number of least squares regressions required to compute all of the break points is of order T, the sample size.
- Each estimated break point is shown to be consistent for one of the true ones despite underspecification of the number of breaks.
- The estimated break points are shown to be $T-$consistent, the same rate as the simultaneous estimation.
- Unlike simultaneous estimation, the limiting distributions are generally not symmetric and are influenced by regression parameters of all regimes.
- A simple method is introduced to obtain break point estimators that have the same limiting distributions as those obtained via simultaneous estimation.
- A procedure is proposted to consistently estimate the number of breaks.

6.1.1 The Model

All of the results can be determined from the simple two break model

$$\begin{aligned} Y_t &= \mu_1 + X_t, \text{ if } t \leq k_1^0 \\ Y_t &= \mu_2 + X_t \text{ if } k_1^0 + 1 \leq t \leq k_2^0 \\ Y_t &= \mu_3 + X_t \text{ if } k_2^0 + 1 \leq t \leq T \end{aligned}$$

where μ_i is the mean of regime i and X_t is a linear process of martingale differences such that $X_t = a(L)\varepsilon_t$ and k_1^0 and k_2^0 are the unknown break points.

6.1.2 Estimation of One Break

The idea of sequential estimation is to consider one break at a time: that is, the model is treated as if there were only one break point and this break point is estimated using least squares. The break point is estimated by minimizing the sum of squared residuals allowing for one break

$$S_T(k) = \sum_{t=1}^{k}(Y_t - \overline{Y}_k)^2 + \sum_{t=k+1}^{T}(Y_t - \overline{Y}_k^*)^2$$

where $\overline{Y}_k$ is the mean of the first k observations and $\overline{Y}_k^*$ is the mean of the last $T-k$ observations. The minimization is achieved by searching over all possible break points and picking the break point that gives the smallest $S_T(k)$. The break point estimator is denoted $\widehat{k}$.

In addition to the break date, the break fraction $\tau = k/T$ is of interest. The estimated break fraction is $\widehat{\tau} = \widehat{k}/T$.

6.1.3 Asymptotic Results

- In a multiple break model, $\widehat{\tau}$ is $T-$consistent for one of the true breaks $\tau_i^0 = k_i^0/T$
- However, $\widehat{k}$ is not consistent for any k_i^0.
- If the first break dominates in terms of the relative span of regimes and the magnitude of shifts, i.e.

$$\frac{\tau_1^0}{\tau_2^0}(\mu_1 - \mu_2)^2 > \frac{1-\tau_2^0}{1-\tau_1^0}(\mu_2 - \mu_3)^2,$$

  then $\widehat{\tau} \longrightarrow \tau_1^0$. Otherwise, $\widehat{\tau} \longrightarrow \tau_2^0$.
- $\widehat{\tau}$ is $T-$consistent for one of the true breaks.
- If the first and second breaks are equivalent in terms of the relative span of regimes and the magnitude of shifts then $\widehat{\tau}$ converges to a random variable with mass only on τ_1^0 and τ_2^0.

6.1.4 Sequential Estimation

- When $\widehat{\tau} = \widehat{k}/T$ is consistent for τ_1^0, an estimate for τ_2^0 can be obtained by applying the same technique to the subsample $[\widehat{k}, T]$. Let $\widehat{k}_2$ denote the resulting estimator. Then $\widehat{\tau}_2 = \widehat{k}_2/T$ is $T-$consistent for τ_2^0 because in the subsample $[\widehat{k}, T]$, k_2^0 is the dominating break and the limiting distribution of $\widehat{k}_2$ is the same as in the single break model.
- For fixed magnitudes of shifts the limiting distributions depend on the unknown distribution of the data and on the unknown magnitudes of the shifts. To get limiting distributions that are invariant the shifts need to shrink as the sample size increases.
- The limiting distribution from sequential estimation is not symmetric about zero in general. In particular, if $\mu_2 - \mu_1$ and $\mu_3 - \mu_2$ have the same sign then the distribution of $\widehat{k}$ will have a heavy right tail, reflecting a tendency to overestimate the break point relative to simultaneous estimation.

6.1.5 Repartition

In general the sequential estimation method has a tendency to either under or over-estimate the true location of a break point when there are multiple breaks. A simple re-estimation method called repartition can eliminate this asymmetry from the asymptotic distribution and works as follows:

- Start with $T-$ consistent estimates of k_i such as from sequential estimation. Call these $\widehat{k}_i$.
- Reestimate k_1 using the subsample $[1, \widehat{k}_2]$ and reestimate $\widehat{k}_2$ using the subsample $[\widehat{k}_1, T]$. Denote the resulting estimators $\widehat{k}_1^*$ and $\widehat{k}_2^*$.
- Because $\widehat{k}_i$ is close to k_i^0 the subsample $[k_{i-1}^0, k_{i+1}^0]$ is effectively used to estimate k_i^0 and so $\widehat{k}_i^*$ is $T-$ consistent with a limiting distribution equivalent to that for a single break model (or for a model with multiple breaks estimated by the simultaneous method)

6.1.6 More than Two Breaks

The general multiple break model with m breaks is

$$\begin{aligned} Y_t &= \mu_1 + X_t, \text{ if } t \leq k_1^0 \\ Y_t &= \mu_2 + X_t \text{ if } k_1^0 + 1 \leq t \leq k_2^0 \\ &\vdots \\ Y_t &= \mu_{m+1} + X_t \text{ if } k_m^0 + 1 \leq t \leq T. \end{aligned}$$

- A subsample $[k, l]$ is said to contain a nontrivial break point if both k and l are bounded away from a break point for a positive fraction of observations. Assuming knowledge of the number of breaks as well as the existence of a nontrivial break in a given subsample, all the breaks can be identified and all the estimated break fractions are $T-$consistent.
- In practice, a problem arises immediately as to whether a subsample contains a nontrivial break, which is clearly related to the determination of the number of breaks. It is suggested that the decision be made based on testing the hypothesis of parameter constancy for the subsample using a sup-Wald test. Such a decision rule leads to a consistent estimate of the number of breaks.
- The number of breaks is determined using a squential estimation procedure coupled with hypothesis testing.
  - First test the entire sample for parameter constancy using the sup-Wald test.

- If parameter constancy is rejected identify the first break as the value that maximizes the sup-Wald statistic. When the first break is identified, the whole sample is divided into two subsamples with the first subsample consisting of the first $\widehat{k}$ observations and the second subsample consisting of the rest of the observations.
- Then use the sup-Wald test on the two subsamples and estimate a break date on the subsample where the test fails. Divide the corresponding subsample in half at the new break date and continue with the process.
- Stop when the sup-Wald test does not reject on any subsample. The number of break points is equal to the number of subsamples minus 1.

- Bai and Perron (1997) suggest an alternative but related procedure to determine the number of breaks.
- Yet another procedure to select the number of breaks is to use the Schwarz BIC. This has be advocated by Yao (1988), Kim and Maddala (1991) and Nunes et. al. (1997), Lui, Wu and Ziduck (2000) and Wang and Zivot (2001)

6.1.7 Simulation Results

The basic simulation model is the three break (4 regime) in mean model:

$$
\begin{aligned}
Y_t &= \mu_1 + X_t, \ t \leq k_1^0 \\
Y_t &= \mu_2 + X_t, \ k_1^0 + 1 \leq t \leq k_2^0 \\
Y_t &= \mu_3 + X_t, \ k_2^0 + 1 \leq t \leq k_3^0 \\
Y_t &= \mu_4 + X_t, \ k_3^0 + 1 \leq t \leq T
\end{aligned}
$$

The design parameters are

$$
\begin{aligned}
\mu &= (1.0, 2.0, 1.0, 0.0)': \text{ design 1} \\
&= (1.0, 2.0, -1.0, 1.0)': \text{ design 2} \\
&= (1.0, 2.0, 3.0, 4.0)': \text{ design 3} \\
T &= 160 \\
k &= (40, 80, 120)' \\
& X_t \ \tilde{} \ N(0, 1)
\end{aligned}
$$

and the number of simulations is $5,000$. In design 1, the magnitude of the breaks are the same. In design 2, the middle break is the largest

Experiment 1: Estimate the break point assuming the number of breaks is known using sequential, repartition and simultaneous estimation methods. Use designs I and II only

- For design I, the distribution of the three break points are similar since the magnitude of the breaks are identical. The distribution of sequential estimates is slightly asymmetric whereas the distributions of the repartition and simultaneous estimates are symmetric and almost identical.
- For design II, the distribution of the three break points are not identical since the middle break is the most pronounced. For all estimation methods, the middle break has the most concentrated distribution followed by the third and then first break. For the sequential methods, the first and third breaks have the same distribution as the simultaneous estimation. The middle break has an asymmetric distribution for the sequential method and the asymmetry is removed by repartition.

Experiment 2: Determine the number of breaks using the sequential method with hypothesis testing and the Schwarz BIC model selection criterion.

- The sequential method has a tendency to underestimate the true number of breaks. The problem is caused in part by the inconsistent estimation of the error variance (for the no structural change test) in the presence of multiple breaks. When multiple breaks exist and only one is allowed in estimation, the error variance cannot be consistently estimated, is biased upward and thus decreases the power of the structural change test.
- The problem of the bias in the estimation of the error variance can be overcome using a two-step method. In the first step, a consistent (or less biased) estimate for the error variance is obtained by allowing more breaks. Let m ($= 4$ in the simulations) denote the fixed number of breaks imposed for the purpose of estimating the error variance. The error variance can be estimated via simultaneous or the "one additional" sequential procedure. In the second step, the number of breaks is determined by the sequential procedure coupled with hypothesis testing where the structural change test usings the first step estimation of the error variance.
- For design I, BIC does better at determining the true number of breaks than the two-step method and the two-step method often only finds 1 break.
- For design II, the two methods are comparable.
- For design III, the two-step method outperforms BIC.
- The two-step method can be improved by using the Bai and Perron sup $F(l)$-test for testing multiple breaks instead of the Andrews sup-Wald test for a single break.

6.2 Estimating Linear Models with Multiple Structural Change

Reference: Bai, J. (1997). "Estimation of a Change Point in Multiple Regression," *Review of Economics and Statistics*.

- This paper studies the least squares estimation of change pointz in linear regression models
- Extends the results of Bai (1997) Econometric Theory to linear regression models
- The model allows for lagged dependent variables and trending regressors.
- The error process can be dependent and heteroskedastic.
- For nonstationary regressors or disturbances the asymptotic distribution is shown to be skewed.
- The analysis applies to both pure and partial changes.
- A sequential method for estimating multiple breaks is described as well as methods for constructing confidence intervals.

6.3 Estimation of Multiple Breaks Simultaneously

References:

1. Bai, J. and P. Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78
2. Bai, J. and P. Perron (2002). "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*.

6.3.1 Application: Exchange Rate Regressions

Reference: Sakoulis, G. and E. Zivot (2001). "Time-Variation and Structural Change in the Forward Discount: Implications for the Forward Rate Unbiasedness Hypothesis," unpublished manuscript, Department of Economics, University of Washington.

7 Estimation of Time Varying Parameter Models

References:

1. Durbin, J. and S.-J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford
2. Koopman, S.-J., N. Shephard, and J.A. Doornik (2001). "Statistical Algorithms for State Space Models Using SsfPack 2.2," *Econometrics Journal*, 2, 113-166.

7.1 Linear Gaussian State Space Models

$$\begin{aligned}
\underset{m\times 1}{\boldsymbol{\alpha}_{t+1}} &= \underset{m\times 1}{\mathbf{d}_t} + \underset{m\times m}{\mathbf{T}_t} \cdot \underset{m\times 1}{\boldsymbol{\alpha}_t} + \underset{m\times r}{\mathbf{H}_t} \cdot \underset{r\times 1}{\boldsymbol{\eta}_t} \\
\underset{N\times 1}{\mathbf{y}_t} &= \underset{N\times 1}{\mathbf{c}_{\mathbf{t}}} + \underset{N\times m}{\mathbf{Z}_{\mathbf{t}}} \cdot \underset{m\times 1}{\boldsymbol{\alpha}_{\mathbf{t}}} + \underset{N\times N}{\mathbf{G}_t} \cdot \underset{N\times 1}{\boldsymbol{\varepsilon}_t}
\end{aligned}$$

where $t = 1, \ldots, n$ and

$$\begin{aligned}
&\boldsymbol{\alpha}_1 \sim N(\mathbf{a}, \mathbf{P}), \boldsymbol{\eta}_t \sim iid\ N(0, \mathbf{I}_r), \boldsymbol{\varepsilon}_t \sim iid\ N(\mathbf{0}, \mathbf{I}_N) \\
&E[\boldsymbol{\varepsilon}_t \boldsymbol{\eta}_t'] = \mathbf{0}
\end{aligned}$$

Compact notation used by `SsfPack`

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{\alpha}_{t+1} \\ \mathbf{y}_t \end{pmatrix} &= \underset{(m+N)\times 1}{\boldsymbol{\delta}_t} + \underset{(m+N)\times m}{\boldsymbol{\Phi}_t} \cdot \underset{m\times 1}{\boldsymbol{\alpha}_t} + \underset{(m+N)\times 1}{\mathbf{u}_t}, \\
&\boldsymbol{\alpha}_1 \sim N(\mathbf{a}, \mathbf{P}) \\
&\mathbf{u}_t \sim iid\ N(\mathbf{0}, \boldsymbol{\Omega}_t)
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\delta}_t &= \begin{pmatrix} \mathbf{d}_t \\ \mathbf{c}_t \end{pmatrix}, \ \boldsymbol{\Phi}_t = \begin{pmatrix} \mathbf{T}_t \\ \mathbf{Z}_t \end{pmatrix}, \ \mathbf{u}_t = \begin{pmatrix} \mathbf{H}_t \boldsymbol{\eta}_t \\ \mathbf{G}_t \boldsymbol{\varepsilon}_t \end{pmatrix}, \\
\boldsymbol{\Omega}_t &= \begin{pmatrix} \mathbf{H}_t \mathbf{H}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_t \mathbf{G}_t' \end{pmatrix}
\end{aligned}$$

Initial value parameters

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{P} \\ \mathbf{a}' \end{pmatrix}$$

Note: For multivariate models, i.e. $N > 1$, $\mathbf{G}_t \mathbf{G}_t'$ is assumed diagonal.

7.2 Initial Conditions

Initial state variance is assumed to be of the form

$$\mathbf{P} = \mathbf{P}_* + \kappa \mathbf{P}_\infty$$

$$\kappa = 10^7$$

P_* is for stationary state components

P_∞ is for non-stationary state components

Note: When some elements of state vector are nonstationary, the `SsfPack` algorithms implement an "exact diffuse prior" approach.

7.3 Regression Model with Time Varying Parameters

Time varying parameter regression

$$\begin{aligned} y_t &= \beta_{0,t} + \beta_{1,t} x_t + \nu_t,\ \nu_t \sim N(0, \sigma_\nu^2) \\ \beta_{0,t+1} &= \beta_{0,t} + \xi_t,\ \xi_t \sim N(0, \sigma_\xi^2) \\ \beta_{1,t+1} &= \beta_{1,t} + \varsigma_t,\ \varsigma_t \sim N(0, \sigma_\varsigma^2) \end{aligned}$$

Let $\boldsymbol{\alpha}_t = (\beta_{0,t}, \beta_{1,t})'$, $\mathbf{x}_t = (1, x_t)'$, $\mathbf{H}_t = diag(\sigma_\xi, \sigma_\varsigma)'$ and $G_t = \sigma_\nu$. The state space form is

$$\begin{pmatrix} \boldsymbol{\alpha}_{t+1} \\ y_t \end{pmatrix} = \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{x}_t' \end{pmatrix} \boldsymbol{\alpha}_t + \begin{pmatrix} \mathbf{H}\boldsymbol{\eta}_t \\ G\varepsilon_t \end{pmatrix}$$

and has parameters

$$\boldsymbol{\Phi}_t = \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{x}_t' \end{pmatrix},\ \boldsymbol{\Omega} = \begin{pmatrix} \sigma_\xi^2 & 0 & 0 \\ 0 & \sigma_\varsigma^2 & 0 \\ 0 & 0 & \sigma_\nu^2 \end{pmatrix}$$

The initial state matrix is

$$\boldsymbol{\Sigma} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}$$

7.4 Kalman Filter and Smoother

The Kalman filter is a recursive algorithm for the evaluation of moments of the normally distributed state vector $\boldsymbol{\alpha}_{t+1}$ conditional on the observed data $\mathbf{Y}_t = (y_1, \ldots, y_t)$ and the state space model parameters. Let $\mathbf{a}_t = E[\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1}]$ and $\mathbf{P}_t = var(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1})$

- The *filtering* or *updating* equations compute

$$\begin{aligned} \mathbf{a}_{t|t} &= E[\boldsymbol{\alpha}_t | \mathbf{Y}_t], \\ \mathbf{P}_{t|t} &= var(\boldsymbol{\alpha}_t | \mathbf{Y}_t), \\ \mathbf{v}_t &= \mathbf{y}_t - \mathbf{c}_t - \mathbf{Z}_t \mathbf{a}_t \text{ (prediction error)}, \\ \mathbf{F}_t &= var(\mathbf{v}_t) \text{ (prediction error variance)} \end{aligned}$$

- The *prediction* equations of the Kalman filter compute $\mathbf{a}_{t+1}$ and $\mathbf{P}_{t+1}$

The *Kalman smoothing* algorithm is a backward recursion which computes the mean and variance of specific conditional distributions based on the full data set $\mathbf{Y}_n = (y_1, \ldots, y_n)$.

- The smoothed estimates of the state vector $\boldsymbol{\alpha}_t$ and its variance matrix are denoted

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_t &= \mathbf{a}_{t|n} = E[\boldsymbol{\alpha}_t | \mathbf{Y}_n] \\ \mathbf{P}_{t|n} &= var(\hat{\boldsymbol{\alpha}}_t | \mathbf{Y}_n) \end{aligned}$$

The smoothed estimate $\hat{\boldsymbol{\alpha}}_t$ is the optimal estimate of $\boldsymbol{\alpha}_t$ using all available information $\mathbf{Y}_n$.

- The smoothed estimate of the response $\mathbf{y}_t$ and its variance are computed using

$$\begin{aligned} \hat{\mathbf{y}}_t &= \mathbf{c}_t + \mathbf{Z}_t \hat{\boldsymbol{\alpha}}_t \\ var(\hat{\mathbf{y}}_t | \mathbf{Y}_n) &= \mathbf{Z}_t var(\hat{\boldsymbol{\alpha}}_t | \mathbf{Y}_n) \mathbf{Z}_t' \end{aligned}$$

- The smoothed disturbance estimates are the estimates $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ based on all available information $\mathbf{Y}_n$, and are denoted

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}_t &= \varepsilon_{t|n} = E[\boldsymbol{\varepsilon}_t | \mathbf{Y}_n] \\ \hat{\boldsymbol{\eta}}_t &= \eta_{t|n} = E[\boldsymbol{\eta}_t | \mathbf{Y}_n] \end{aligned}$$

7.5 Prediction Error Decomposition of Log-Likelihood

The *prediction error decomposition* of the log-likelihood function for the unknown parameters $\boldsymbol{\varphi}$ of a state space model is

$$\begin{aligned} \ln L(\boldsymbol{\varphi}|Y_n) &= \sum_{t=1}^{n} \ln f(\mathbf{y}_t | \mathbf{Y}_{t-1}; \boldsymbol{\varphi}) \\ &= -\frac{nN}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \left(\ln |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right) \end{aligned}$$

where $f(\mathbf{y}_t | \mathbf{Y}_{t-1}; \boldsymbol{\varphi})$ is a conditional Gaussian density implied by the state space model and the vector of prediction errors $\mathbf{v}_t$ and prediction error variance matrices $\mathbf{F}_t$ are computed from the Kalman filter recursions.

7.6 Example: Testing the forward rate unbiasedness hypothesis