Multiple Structural Breaks, Forecasting and Present Value Calculations

Hashem Pesaran
Faculty of Economics, Cambridge University
Centre for International Macroeconomics and Finance

December 2006.

Declaration

This presentation is based on the following two papers:

FORECASTING TIME SERIES SUBJECT TO MULTIPLE STRUCTURAL BREAKS, *Review of Economic Studies*, 2006.

and

LEARNING, STRUCTURAL INSTABILITY AND PRESENT VALUE CALCULATIONS, *Econometric Reviews*, forthcoming.

Google “**pesaran**”

Introduction

- Present value relations play a key role in economics and finance and are used in testing the permanent income hypothesis, in standard inventory models and to calculate the present value of assets such as stocks and bonds.
- Computing present values requires forecasting a stream of future values of the variable of interest at horizons that can be long, but finite (as in the case of bonds) or even infinitely long (as in the case of stocks).
- It is customary in such calculations to assume that the underlying driving process follows a simple ARMA process with **stable** and **known** parameter values.
- We wish to **dispense with the assumption of known parameters and to consider the possibility of past and future breaks** in the data generating process of the driving variable.
- How to compute expected values of future dividends in the

presence of the considerable uncertainty surrounding not just dividends in the near future but dividends at very distant future points in time.

Key questions in forecasting:

- How will future values be affected by breaks?
- Frequency of breaks
- Size of breaks
- Break point dependencies

Alternative approaches to modeling breaks and allowing for breaks in forecasting:

- Deterministic. Bai and Perron (1998, 2003) - does not help in forecasting
- Stochastic recurring. Hamilton (1989) - often too restrictive
- Time Varying Parameter (TVP) Models - too many changes
- Use post-break data only. ROC Pesaran-Timmermann (2002, Journal of Empirical Finance) - does not allow for breaks in future
- Pool forecats across different estimation windows. Pesaran and Timmermann (2006, Journal of Econometric, forthcoming).
- Stochastic Break Process (Draws from a meta distribution)
  - Break Points
  - Size of parameter change

How much can be learned about future breaks from past breaks?

Related to how similar the parameters are across various break segments

- Pooled scenario: Narrow dispersion of the distribution of parameters across breaks - parameters from previous break segments contain considerable information on parameters after future breaks
- Regime specific scenario: wider spread across breaks, less commonality and more uncertainty

Hierarchical Hidden Markov Chain Approach

- Parameters within each break segment drawn from common meta distribution
- Dependencies in parameters across neighbouring regimes also allowed
- Forecasts embody information on the size and frequency of past breaks instead of discarding observations prior to the most recent break point
- As new regimes occur, the priors of the meta distribution are updated using Bayes' rule
- Uncertainty about the number of in-sample breaks integrated out by BMA

Modeling the Break Process

Related work: Gelman et al (2002), Inclan (1994), Kim, Nelson and Piger (2004), Koop (2003), Koop and Potter (2004a,b), McCulloch and Tsay (1993), Chib (1998)

Hidden Markov Chain (HMC) multiple change point model - Chib (1998)

Breaks captured by $S_t = 1, 2, \ldots, K+1$

$s_t = l : y_t$ has been drawn from $f(y_t | \Upsilon_{t-1}, \boldsymbol{\theta}_l)$

$Y_t = \{y_1, \ldots, y_t\}$: current information set

$\theta_l = \left[\boldsymbol{\beta}_l, \sigma_l^2 \right]$: location, scale param. in regime l :

$\theta_t = \theta_l$ if $\tau_{l-1} \leq t \leq \tau_l$

$\Upsilon_K = \{\tau_0, \ldots., \tau_{K+1}\}$: collection of break points

S_t : first order Markov process with constrained transition probability matrix

At each point in time, S_t can either remain in the current state or jump to the next state.

Conditional on K in-sample breaks, transition probability matrix is

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & 0 & \dots & 0 \\ 0 & p_{22} & p_{23} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & p_{KK} & p_{K,K+1} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Other specifications are possible.... cf Levels and variances-shift AR model of McCulloch and Tsay (1993), and duration dependence model of Koop and Potter (2004a,b).

Transition Probabilities and Meta Distributions

Assume that p_{ii} are independent drawns from a beta distribution

$$p_{ii} \sim Beta(\underline{a}, \underline{b}), \text{ for } i = 1, 2 \ldots, K$$

Joint density of $p = (p_{11}, \ldots, p_{KK})'$ is

$$\pi(\mathbf{p}) = c_K \prod_{i=1}^{K} p_{ii}^{(\underline{a}-1)} (1 - p_{ii})^{(\underline{b}-1)}$$

$$c_K = \{\Gamma(\underline{a} + \underline{b}) / \Gamma(\underline{a}) \Gamma(\underline{b})\}^K$$

For forecasting, we propose a hierarchical break point formulation by making use of meta distributions for the unknown parameters.

Assumption that parameters are drawn from a meta distribution is not very restrictive:

- pooled scenario (all parameters are identical across regimes)
- regime-specific scenario (each regime has different (own) parameters)

Hierarchical prior places some structure on the relation between regime coefficients:

$$\boldsymbol{\beta}_j \overset{iid}{\sim} N(\mathbf{b}_0, \mathbf{B}_0)$$

$$\sigma_j^{-2} \overset{iid}{\sim} G(v_0, d_0)$$

Which scenario most closely represents the data can be inferred from the estimates of $\mathbf{B}_0$ and d_0

At the next level of the hierarchy we assume

$$\mathbf{b}_0 \sim N\left(\underline{\boldsymbol{\mu}_\beta}, \underline{\boldsymbol{\Sigma}_\beta}\right)$$

$$\mathbf{B}_0^{-1} \sim W\left(\underline{\nu_\beta}, \underline{\mathbf{V}_\beta^{-1}}\right)$$

$W(.)$: Wishart distribution

$\underline{\boldsymbol{\mu}_\beta}$, $\underline{\boldsymbol{\Sigma}_\beta}$, $\underline{\nu_\beta}$, $\underline{\mathbf{V}_\beta^{-1}}$: hyperparameters

ν_0, d_0 : error term precision hyperparameters:

$$\nu_0 \sim Exp\left(\underline{\rho_0}\right)$$

$$d_0 \sim Gamma(\underline{c_0}, \underline{d_0})$$

$\underline{\rho_0}$, $\underline{c_0}$, $\underline{d_0}$: hyperparameters

We simulate the hierarchical HMC model by means of a Gibbs sampler.

Forecasting - Posterior Predictive Distributions

Conditional on Y_T, forecasts of $y_{T+h},$ can be made under a range of out-of-samlpe scenarios:

- No new break: forecast y_{T+h} using only posterior distn of parameters from last regime, $\{\beta_{K+1}, \sigma^2_{K+1}\}$
- New break: need to compute
  - probabilities of all possible break dates from $p_{K+1,K+1}$, $p_{K+2,K+2}, \ldots$
  - regression parameters, $\beta_j \sim (b_0, B_0)$
  - error term precisians, $h_j \sim (v_0, d_0)$

From Markov chain property, the probability of a break at time $T+j$ is

$$Pr(s_{T+h} = K+2 | \tau_{K+1} = T+j, s_T = K+1) = (1 - p_{K+1,K+1}) p^{j}_{K+1,K+1}$$

Second-stage out-of-sample forecast replaces the earlier conditional P by

$$\tilde{\mathbf{P}} = \begin{pmatrix} p_{11} & p_{12} & 0 & \dots & 0 & & \\ 0 & p_{22} & p_{23} & \dots & 0 & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ 0 & \dots & 0 & p_{KK} & p_{K,K+1} & & \\ 0 & 0 & \dots & 0 & p_{K+1,K+1} & p_{K+1,K+2} & \\ 0 & 0 & \dots & & 0 & p_{K+2,K+2} & \\ & & & & & & \ddots \end{pmatrix}$$

Combining the Different Forecasts using Bayesian Model Averaging

- True # of in-sample breaks is unknown
- Integrate out uncertainty using BMA
- Compute predictive density as a weighted average of the predictive densities, each of which conditions on a given value of K, using model posteriors as (relative) weights
  $p_K(y_{T+h}|\Upsilon_T) \equiv p(y_{T+h}|s_T = K+1, \Upsilon_T)$: posterior density conditional on K in-sample breaks

$$p_K(y_{T+h}|\mathrm{Y}_T) = \Pr(s_{T+h} = K+1|\mathrm{Y}_T) \times p(y_{T+h}|s_T = K+1, s_{T+h} = K+1, \mathrm{Y}_T)$$

$$\sum_{j=1}^{h} \Pr(s_{T+h} = K+2|\tau_{K+1} = T+j, s_T = K+1)$$

$$\times\, p(y_{T+h}|\tau_{K+1} = T+j, s_T = K+1, s_{T+h} = K+2, \mathrm{Y}_T) +$$

$$\sum_{j=1}^{h-m+1} \sum_{l=j+1}^{h-m+2} \cdots \sum_{m=l+1}^{h}$$

$$\Pr(s_{T+h} = K+m+1|\tau_{K+m} = T+l, \ldots, \tau_{K+1} = T+j, s_T = K+1)$$

$$\times\, p(y_{T+h}|\tau_{K+m} = T+l, \ldots, \tau_{K+1} = T+j, \mathrm{Y}_T).$$

M_k : model assuming $k-1$ breaks at time T (i.e., $s_T = k$)

BMA forecast:

$$p(y_{T+h}|\mathrm{Y}_T) = \sum_{k=1}^{\bar{K}} p_k(y_{T+h}|\mathrm{Y}_T)p(M_k|\mathrm{Y}_T),$$

Weights are given by the posterior model probs:

$$p(M_k|y) \propto f(y|M_k)p(M_k)$$

$f(y|M_k)$: marginal likelihood

$p(M_k)$: prior for model M_k

Parameter Uncertainty and Present Value Calculations

Known μ and Known σ^2

$$y_T = \lim_{H\to\infty}\left\{\sum_{h=1}^{H}\delta^h E(x_{T+h}|\mathrm{I}_T)\right\},$$

$$\Delta \ln x_{t+1} = \mu + \sigma\varepsilon_{t+1},$$

$$E(x_{T+h}|\mathrm{I}_T;\mu,\sigma) = [e^{\mu}M_{\varepsilon}(\sigma)]^h,$$

where $M_{\varepsilon}(\sigma)$ is the moment generating function of ε_t, assuming that it exists. Suppose $\delta e^{\mu}M_{\varepsilon}(\sigma) < 1$

$$y_T = \left[\frac{\delta e^{\mu}M_{\varepsilon}(\sigma)}{1-\delta e^{\mu}M_{\varepsilon}(\sigma)}\right]x_T.$$

In the case of normally distributed errors $M_\varepsilon(\sigma) = \exp(0.5\sigma^2)$.

However, $M_\varepsilon(\sigma)$ does not exists if the innovations are t-distributed! Geweke (1991).

Unknown μ with a Known σ^2

Based on the past observations, (m size of estimation window) $\mathbf{X}_{m,T} = (x_{T-m+1}, x_{T-m+2}, \ldots, x_T)'$, with a Gaussian prior:

$$\mu \sim N\left(\underline{\mu}, \underline{\sigma}_\mu^2\right), \ \underline{\sigma}_\mu^2 > 0.$$

$$\bar{\mu} = \bar{\sigma}_\mu^2 \left(\frac{\underline{\mu}}{\underline{\sigma}_\mu^2} + \frac{m\,\bar{x}_{m,T}}{\sigma^2} \right),$$

$\bar{x}_{m,T} = m^{-1} \sum_{t=T-m+1}^{T} x_t$, and

$$\bar{\sigma}_\mu^2 = \left(\frac{1}{\underline{\sigma}_\mu^2} + \frac{m}{\sigma^2} \right)^{-1}$$

$$y_{T:T+H} = x_T \sum_{h=1}^{H} \left(\delta e^{\bar{\mu} + ½ h \bar{\sigma}_\mu^2 + ½ \sigma^2} \right)^h = x_T \sum_{h=1}^{H} [\rho(h,m)]^h,$$

where

$$\rho(h,m) = e^{-\ln(1+r)+\bar{\mu}+½h\bar{\sigma}_{\mu}^{2}+½\sigma^{2}}.$$

But

$$\bar{\mu} = \bar{x}_{m,T} + \left(\frac{\sigma^2}{m}\right)\left(\frac{\underline{\mu} - \bar{x}_{m,T}}{\underline{\sigma}_{\mu}^{2}}\right) + O\left(\frac{1}{m^2}\right),$$

$$\bar{\sigma}_{\mu}^{2} = \frac{\sigma^2}{m} + O\left(\frac{1}{m^2}\right),$$

and

$$y_{T:T+H} \approx x_T \sum_{h=1}^{H} \left(e^{-\ln(1+r)+½\sigma^2(1+h/m)+\bar{x}_{m,T}}\right)^{h}.$$

The factor, $1 + h/m$, **is due to estimation uncertainty and for a fixed estimation window** (m), **explods as** $h \to \infty$.

Even if $\hat{\lambda}_{m,T} = \exp(-\ln(1+r) + \bar{x}_{m,T} + 0.5\sigma^2) < 1$ (namely even if the estimated certainty equivalence convergence condition holds), $y_{T:T+H}$ will be divergent as $H \to \infty$.

Unknown μ and σ^2

- The non-convergence problem is accentuated σ is also unknown.
- Using conjugate priors for μ and σ^2 the posterior distribution of μ will be t –distributed and $E\left(e^{h\mu} \middle| \mathrm{I}_T; \underline{\mu}, \underline{\sigma}_{\mu}^2, \underline{\sigma}^2, \underline{v}\right)$ ceases to exist for any $h > 0$, where $\underline{\sigma}^2$ and $\underline{v}$ are the parameters of the gamma prior density assumed for σ^2 :

$$\frac{\underline{\sigma}^2}{\sigma^2} | \underline{\sigma}^2, \underline{v} \sim \chi^2(\underline{v}).$$

- The use of non-conjugate priors for μ and σ^2 does help in resolving the non-existence of $E\left(e^{h\mu} \middle| \mathrm{I}_T; \underline{\mu}, \underline{\sigma}_{\mu}^2, \underline{\sigma}^2, \underline{v}\right)$.
- BUT use of non-conjugate priors does not resolve the non-convergence of the infinite sums that are involved in present value calculations.

Trend Stationary Log-linear Driving Processes

The non-convergence problem continues to be present if the geomtric random walk model is replaced by trend stationary process:

$$\Delta[\ln x_{t+1} - a - \mu(t+1)]$$
$$= -(1-\rho)(\ln x_t - a - \mu t) + \sigma\varepsilon_{t+1},$$

where $|\rho| < 1$, and as before μ represents the average growth of the logarithm of the driving process, x_t. In the case of this process

$$\ln(x_{T+h}/x_T) = -(1-\rho^h)(\ln x_T - a - \mu T) + \mu h + \sigma \sum_{j=1}^{h} \rho^j \varepsilon_{T+j},$$

and

$$E(x_{T+h}|\mathrm{I}_T; a, \mu, \rho, \sigma) = e^{-(1-\rho^h)(\ln x_T - a - \mu T)}(e^{h\mu})\left[\prod_{j=1}^{h} M_\varepsilon(\sigma\rho^j)\right],$$

It is clear that the various issues discussed for the unit root case readily apply here. Even if ε_{t+1} has a moment generating function, the present value is unlikely to exist if μ is not known with certainty. For example, suppose a, ρ and σ are known and μ is estimated based on the regression of $\ln x_t - \rho \ln x_{t-1} - a(1-\rho)$ on $(1-\rho)t + \rho$. Assuming, as before, that conditional on a, ρ and σ the prior probability distribution of μ is Gaussian and given by (ref: gp1), then the posterior distribution of μ will be given by

$$\mu \Big| \mathbf{X}_T, a, \rho, \sigma, \underline{\mu}, \underline{\sigma}_\mu^2 \sim N(\bar{\mu}, \bar{\sigma}_\mu^2),$$

where

$$\bar{\mu} = \bar{\sigma}_\mu^2\left(\frac{\underline{\mu}}{\underline{\sigma}_\mu^2} + \frac{\hat{\mu}_T}{\hat{\sigma}_T^2}\right),$$

$$\hat{\mu}_T = \frac{\sum_{t=1}^{T} [\ln x_t - \rho \ln x_{t-1} - a(1-\rho)][(1-\rho)t + \rho]}{\sum_{t=1}^{T} [(1-\rho)t + \rho]^2},$$

$$\hat{\sigma}_T^2 = \frac{\sigma^2}{\sum_{t=1}^{T} [(1-\rho)t + \rho]^2}, \text{ and } \bar{\sigma}_\mu^2 = \left(\frac{1}{\underline{\sigma}_\mu^2} + \frac{1}{\hat{\sigma}_T^2} \right)^{-1}.$$

Hence

$$E(x_{T+h} | \mathrm{I}_T; a, \rho, \sigma) = \left[e^{-(1-\rho^h)(\ln x_T - a)} \prod_{j=1}^{h} M_\varepsilon(\sigma \rho^j) \right] E_\mu \left(e^{\left[h + (1-\rho^h)T \right] \mu} \right),$$

$$= \left[e^{-(1-\rho^h)(\ln x_T - a)} \prod_{j=1}^{h} M_\varepsilon(\sigma \rho^j) \right] e^{\left[h + (1-\rho^h)T \right] \bar{\mu} + ½ \left[h + (1-\rho^h)T \right]^2 \bar{\sigma}_\mu^2}$$

and for a fixed T its rate of expansion is governed by the term $\exp(.5\bar{\sigma}_{\mu}^{2}h^{2})$.

- Under the unit root process the precision of μ is of order T^{-1}, while when $\ln x_t$ is trend stationary it is given by $T^{-3/2}$.

Present Values with a Stochastic Discount Factor: The Lucas Tree Model

In the context of a representative agent model with the utility function, $u(c_t)$

$$y_T = \lim_{H\to\infty}\left\{\sum_{h=1}^{H}\delta^h E\left(\frac{u'(c_{t+h})}{u'(c_t)}x_{T+h}|\mathrm{I}_T\right)\right\},$$

Certain analytical results can be obtained for the Lucas's tree model where consumption is equal to dividends ($c_{T+h} = x_{T+h}$) and the utility function is

$$u(c) = (1-\gamma)^{-1}(c^{1-\gamma}-1), \quad (\gamma \neq 1)$$

In this case,

$$y_T = \lim_{H\to\infty}\left\{\sum_{h=1}^{H}E(e^{-h\ln(1+r)+(1-\gamma)(\ln x_{T+h}-\ln x_T)}|\mathrm{I}_T)\right\},$$

and under the geometric random walk model with a known mean and variance we have

$$E(y_T|\mu,\sigma^2) = x_T \lim_{H\to\infty}\left\{\sum_{h=1}^{H} E\left(e^{-h\ln(1+r)+(1-\gamma)\mu h+0.5(1-\gamma)^2\sigma^2 h}|\mu,\sigma^2,\mathrm{I}_T\right)\right\},$$

which is convergent for known values of μ and σ^2 so long as

$$-\ln(1+r)+(1-\gamma)\mu+0.5(1-\gamma)^2\sigma^2 < 0.$$

Consider the case of unknown μ and a known σ^2.

$$E\left(e^{(1-\gamma)\mu h+0.5(1-\gamma)^2\sigma^2 h}|\sigma^2,\mathrm{I}_T\right)$$

$$= e^{(1-\gamma)\bar{\mu}h+0.5(1-\gamma)^2\sigma^2 h+0.5(1-\gamma)^2h^2\bar{\sigma}_\mu^2},$$

where $\bar{\mu}$ and $\bar{\sigma}_\mu^2$ are the posterior mean and variance of μ given before. Hence, the series expansion of $E(y_T/x_T|\sigma^2)$ eventually will be dominated by terms $e^{0.5(1-\gamma)^2h^2\bar{\sigma}_\mu^2}$, $h = 1,2,\ldots$ and the PV will be divergent unless $\gamma = 1$.

Possible Solutions to the Non-Convergence Problem

- Suppose that over the forecast horizon $T+1, T+2, \ldots, T+H$, μ can take any one of the values $\mu_1, \mu_2, \ldots, \mu_m$ with probabilities $\pi_1, \pi_2, \ldots, \pi_m$ where $\Sigma_{i=1}^{m} \pi_i = 1$, and $1 > \pi_i > 0$.
- To simplify the analysis also assume that σ^2, μ_i and π_i are known at time T. Under this example, the present value is given by

$$y_T = x_T \left\{ \sum_{i=1}^{m} \pi_i \lim_{H \to \infty} \sum_{h=1}^{H} e^{-h \ln(1+r) + h\mu_i} [M_\varepsilon(\sigma)]^h \right\}.$$

  Since $1 > \pi_i > 0$, y_T exists if $\delta e^{\mu_i} M_\varepsilon(\sigma) < 1$ for *all* i.
- Contrast this result with the associated certainty equivalent expression:

$$y_T^{CE} = x_T \left\{ \lim_{H\to\infty} \sum_{h=1}^{H} e^{-h\ln(1+r)+h\bar{\mu}_\pi} [M_\varepsilon(\sigma)]^h \right\},$$

where

$$\bar{\mu}_\pi = \Sigma_{i=1}^m \pi_i \mu_i.$$

The condition for y_T^{CE} to exist is given by $\delta e^{\bar{\mu}_\pi} M_\varepsilon(\sigma) < 1$.

- Clearly, it is possible for the latter to be satisfied without $\delta e^{\mu_i} M_\varepsilon(\sigma) < 1$ being satisfied for all i.
- A sufficiently large μ_i, even if extremely unlikely (π_i very close to zero), can result in divergence of y_T, although for all other outcomes that are much more likely the associated infinite sums could be convergent.
- The non-convergence of the present value arises from the particular combinations of
  - (i) a geometric random walk driving process,
  - (ii) an infinite horizon

- (iii) constant, but unknown parameters.

Use of Linear Driving Processes.

$$x_t = \mu + x_{t-1} + \sigma\varepsilon_t.$$

Then $E(x_{T+h}|\mathrm{I}_T) = x_T + \mu h$, and

$$y_T = \lim_{H\to\infty}\left\{\sum_{h=1}^{H}\delta^h E(x_{T+h}|\mathrm{I}_T)\right\}$$

$$= \frac{x_T}{1-\delta} + \mu\sum_{\tau=1}^{\infty} h\delta^h,$$

or

$$y_T = \frac{x_T}{1-\delta} + \frac{\mu\delta}{(1-\delta)^2}.$$

Use of Finite Horizons

$$y_T(\tilde{H}) = \sum_{h=1}^{\tilde{H}} \delta^h E(x_{T+h}|\mathrm{I}_T),$$

Choice of $\tilde{H}$ could be problematic.

One could assume that (for $0 < \theta < 1$)

$$\Pr(\tilde{H} = s) = \frac{(1-\theta)\theta^s}{\theta\left(1-\theta^{\bar{H}}\right)}, \text{ for } s = 1, 2, \ldots, \bar{H}$$

$$= 0, \text{ for } s > \bar{H},$$

Yaari (1965), Cass and Yaari (1967) and Blanchard and Fischer (1989) consider the case where $\bar{H} \to \infty$.

$$E_{\tilde{H}}\left[y_T(\tilde{H})\right] = \sum_{s=1}^{\bar{H}} \frac{1-\theta^{\bar{H}-s+1}}{\theta\left(1-\theta^{\bar{H}}\right)} (\theta\delta)^s E(x_{T+s}|\mathrm{I}_T).$$

In this set up the choice of $\bar{H}$ is of secondary importance.

However, as $\bar{H} \to \infty$

$$\lim_{\bar{H} \to \infty} E_{\tilde{H}}\left[y_T(\tilde{H}) \right] = \theta^{-1} \sum_{s=1}^{\infty} (\theta\delta)^s E(x_{T+s} | \mathrm{I}_T),$$

which returns to the infinite horizon problem.

A finite $\bar{H}$ would still be required in general.

Simulating Present Values under Alternative Scenarios

$$\lim_{H\to\infty} y_T = \lim_{H\to\infty} \sum_{h=1}^{H} \exp(-h\ln(1+r)) \int x_{T+h} p(x_{T+h}|\mathrm{I}_T) dx_{T+h},$$

$\mathrm{I}_T = \{x_1, \ldots, x_T\}$.

- $p(x_{T+h}|\mathrm{I}_T)$ is the predictive density of dividends at time $T+h$ conditional on I_T.

Consider three different scenarios capturing different assumptions about the forecaster's beliefs:

1. Allow for parameter estimation uncertainty but ignores past and future breaks:

$$p(x_{T+h}|S_{T+h} = 1, \mathrm{I}_T) = \int p(x_{T+h}|\Theta_T, S_{T+h} = 1, \mathrm{I}_T)$$

$$\pi(\Theta_T|S_{T+h} = 1, \mathrm{I}_T) d\Theta_T,$$

where Θ are the constant model parameters whose posterior distribution given the data at time T is $\pi(\Theta|S_{T+h} = 1, \mathrm{I}_T)$.

2. Accounts for historical breaks but ignores the possibility of future breaks:

$$p(x_{T+h}|S_{T+h} = K+1, \mathrm{I}_T) = \int p(x_{T+h}|\Theta_{K+1}, S_{T+h} = S_T = K+1, \mathrm{I}_T)$$

$$\pi(\Theta_{K+1}|\mathrm{H}, p, \mathrm{I}_T)d\Theta_{K+1},$$

where Θ_{K+1} are the parameters in the last regime (labelled $K+1$), while H is the set of hyper parameters.

3. A model that accounts for parameter estimation uncertainty as well as past and future breaks:

$$\int p(x_{T+h}|S_{T+h} = K+n_b+1, S_T = K+1, \mathrm{I}_T),$$

where n_b is the maximum number of out-of-sample breaks so the predictive density can be calculated (integrating out uncertainty about the dates of the breaks,

$\tau_{K+1} = T + j_1, \ldots, \tau_{K+n_b} = j_{n_b}$) as

$$p(x_{T+h} | S_{T+h} = K + n_b + 1, S_T = K + 1, \mathrm{I}_T) = \sum_{j_1=1}^{h-n_b+1} \cdots \sum_{j_{n_b}=j_{n_b-1}+1}^{h} \int \cdots \int$$

$$p(x_{T+h}, \Theta_{K+2}, \ldots, \Theta_{K+n_b+1}, \mathrm{H}, S_{T+h} = K + n_b + 1,$$

$$\tau_{K+1} = T + j_1, \ldots, \tau_{K+n_b} = T + j_{n_b}, S_T = K + 1, \mathrm{I}_T)$$

$$\times \pi(\tau_{K+1} = T + j_1, \ldots, \tau_{K+n_b} = t + j_{n_b} | S_{T+h} = K + n_b + 1, S_T = K + 1)$$

$$\times \pi(\Theta_{K+2}, \ldots, \Theta_{K+n_b+1}, \mathrm{H} | \mathrm{I}_T) d\Theta_{K+2} \ldots d\Theta_{K+n_b+1} d\mathrm{H}.$$

$H = 1000$, $r = 10\%$ per annum.

The parameters of the prior were as follows: $p_{ii} \sim Beta(\underline{a}, \underline{b})$ with $\underline{a} = \underline{b} = 0.5$. We assume an uninformative prior for the parameters of the conditional mean of the dividend process by setting $\underline{\boldsymbol{\mu}_\beta} = \mathbf{0}_{2\times1}$, $\underline{\mathbf{V}_\beta} = 1000\mathbf{I}_2$

The hyperparameters determining the error term precision are $\underline{c_0} = 1; \underline{d_0} = 1/100; \underline{\rho_0} = 100$, while the prior for the transition probability matrix is assumed to be drawn from a $Gamma(\underline{a}_0, \underline{b}_0)$ distribution with $\underline{a}_0 = 1; \underline{b}_0 = 1/10$.

Empirical Results

Present Value Calculations

We use data from Shiller (2000) available at http://www.econ.yale.edu/~shiller/data.htm.

Monthly observations on real dividend and real share prices over the period 1871:03 to 2003:09.

Based on the posterior modes for the probability of a shift in the state variable, S_t, the five breaks are estimated to have occurred in 1911, 1922, 1930, 1952 and in 1960.

A Dividend Model with Breaks

$$\begin{aligned}
\Delta \ln(x_{t+1}) - \mu_1 &= \beta_1(\Delta \ln(x_t) - \mu_1) + \sigma_1 \varepsilon_{t+1}, & \tau_0 &\le t \le \tau_1 \\
\Delta \ln(x_{t+1}) - \mu_2 &= \beta_2(\Delta \ln(x_t) - \mu_2) + \sigma_2 \varepsilon_{t+1}, & \tau_1 + 1 &\le t \le \tau_2 \\
&\vdots & &\vdots \\
\Delta \ln(x_{t+1}) - \mu_{K+1} &= \beta_{K+1}(\Delta \ln(x_t) - \mu_{K+1}) + \sigma_{K+1} \varepsilon_{t+1}, & \tau_K + 1 &\le t \le T
\end{aligned}$$

Concluding Remarks

Introduced hierarchical hidden Markov chain approach to model the meta distribution for the parameters of the stochastic process underlying structural breaks

This allowed us to forecast economic time series subject to structural breaks

- We showed that stock prices can be quite sensitive to the nature of the assumptions concerning uncertainty and instability of the parameters of the dividend process.
- These findings suggest that our understanding of the dynamics in stock prices can be improved by focusing on the uncertainty surrounding the underlying fundamentals process.

Why should the approach work empirically?

- paucity of identifiable breaks in most samples
- breaks important but the parameters of the break process are

imprecisely estimated

- Shrinkage towards prior - Diebold and Pauly (1990), Stock and Watson (2003), Garratt, Lee, Pesaran and Shin (2003), Aiolfi and Timmermann (2004)

When can we expect the approach to work?

- Long term forecasting
- Short post-break samples
- Time series particularly sensitive to breaks

Extensions

- Time-varying state transitions - requires variables explaining regime shifts (and a model for predicting their future values)
- Alternative priors - Koop and Potter (2004a,b)
- Other Present Value problems - energy and environment, life cycle models, cost benefit analysis of large projects....