State-Space Time Series Modeling of Structural Breaks

J. Huston McCulloch

Economics Department

Ohio State University

Working Paper # 00-11

June 18, 1997

Rev. Aug. 23, 2000

The author is grateful to participants in the Ohio State University Economics Department Econometrics seminar for helpful comments and suggestions. Author's e-mail: mcculloch.2@osu.edu

INTRODUCTION

The parameters of time series models often appear to change over time. A popular method of dealing with these changes is to introduce "structural breaks" or "regime shifts." These "breaks" are assumed to occur infrequently, but when they do happen, one or more parameter is customarily permitted to undergo a complete break with its past value. (See, e.g. Perron 1994; Bai, Lumsdaine and Stock 1998).

This note shows, using standard state-space time series modeling techniques, that if such a model is taken literally and estimated by maximum likelihood (ML), it does not give the desired results: The assumption that no new value of the parameter (or parameters) in question is any more likely than any other implies that the ML estimator of the parameter in question should ignore the possibility of structural breaks, no matter how evident they may be in the data. Furthermore, such a model has no predictive power, since the parameter(s) in question could make another complete break with its past at any moment, and take the time series anywhere.

The solution to this problem is to explicitly model the breaks as coming from a distribution that is more likely to give a new value in some vicinity of the old value than arbitrarily far from it. Once this is done, the rate of occurrence of the breaks and/or the parameter(s) of the break-generating distribution may be estimated by ML, and the time-changing parameter(s) estimated by an appropriate smoothing algorithm. The time-changing parameter(s) will make the future course of the time series more unpredictable than it would otherwise be, but at least this uncertainty will be probabilistically quantifiable.

A SIMPLE "STRUCTURAL BREAK" MODEL

The simplest model that gives rise to the possibility of "structural breaks" is one in which an observed time series y_t has an unobserved conditional mean x_t that occasionally undergoes permanent shifts. Let

$$\begin{aligned} y_t &= x_t + e_t, \qquad e_t \sim f(e_t), \\ x_t &= x_{t-1} + h_t, \quad h_t \sim g(h_t), \end{aligned} \tag{1}$$

where the observation errors ε_t and the random shifts η_t have mean zero and are serially and mutually independently distributed. The shift η_t may be thought of as representing a "break" if with some small positive probability λ it is drawn from some distribution with density $h(\eta_t)$, and otherwise, with probability $(1 - \lambda)$, is zero. This makes $g(\eta_t)$ a compound distribution, with a mass point with weight $(1 - \lambda)$ at 0, and density $\lambda\, h(\eta_t)$ everywhere else.[1]

In general, system (1) can be estimated using the recursive filtering algorithm due to Sorenson and Alspach (1971) (see also Kitagawa 1987, Harvey 1989: 162-165). The implicit hyperparameter vector θ includes λ and any other parameters of $f(\cdot)$ and $h(\cdot)$, such as their variances and any shape parameters. The filter is initialized with

$$p(x_1|y_1) = f(y_1 - x_1). \tag{2}$$

Then, given last period's filter density $p(x_{t-1}|Y_{t-1})$, the one-step-ahead predictive density is given by

$$\begin{aligned} p(x_t \mid Y_{t-1}) &= \int_{-\infty}^{\infty} p(x_t \mid x_{t-1})\, p(x_{t-1} \mid Y_{t-1}) dx_{t-1} \\ &= \int_{-\infty}^{\infty} g(x_t - x_{t-1})\, p(x_{t-1} \mid Y_{t-1}) dx_{t-1} \\ &= (1-l)\, p(x_{t-1} \mid Y_{t-1})\Big|_{x_{t-1}=x_t} + l \int_{-\infty}^{\infty} h(x_t - x_{t-1})\, p(x_{t-1} \mid Y_{t-1}) dx_{t-1} \end{aligned} \tag{3}$$

[1] If desired, $g(\eta)$ may be written $(1-\lambda)\delta(\eta) + \lambda h(\eta)$, where $\delta(\cdot)$ is the Dirac delta function.

where $Y_t = \{y_1, y_2, \ldots y_t\}$. The new filter density is then given by

$$\begin{aligned} p(x_t \mid Y_t) &= p(y_t \mid x_t)\, p(x_t \mid Y_{t-1}) / p(y_t \mid Y_{t-1}) \\ &= f(y_t - x_t)\, p(x_t \mid Y_{t-1}) / p(y_t \mid Y_{t-1}), \end{aligned} \tag{4}$$

where the denominator, which is the probability of y_t conditional on Y_{t-1}, and therefore the likelihood contribution of y_t, is given by the integral of the numerator:

$$\begin{aligned} p(y_t \mid Y_{t-1}) &= \int_{-\infty}^{\infty} p(y_t \mid x_t)\, p(x_t \mid Y_{t-1}) dx_t \\ &= \int_{-\infty}^{\infty} f(y_t - x_t)\, p(x_t \mid Y_{t-1}) dx_t . \end{aligned} \tag{5}$$

The first moment of the filter density gives the expectation of x_t, conditional on data up to time t, while the hyperparameter vector θ may be estimated by numerically maximizing the likelihood, or equivalently, the probability of the entire sample through the last observed time T, conditional on the first observation:

$$L(q \mid Y_T) = p(Y_T \mid y_1) = \prod_{t=2}^{T} p(y_t \mid Y_{t-1}) . \tag{6}$$

The "smoother" density, which gives the distribution of x_t conditional on all the data in both directions, may be obtained by backward recursion using the following formula:

$$\begin{aligned} p(x_t \mid Y_T) &= p(x_t \mid Y_t) \int_{-\infty}^{\infty} p(x_{t+1} \mid Y_T)\, p(x_{t+1} \mid x_t) / p(x_{t+1} \mid Y_t) dx_{t+1} \\ &= p(x_t \mid Y_t) \int_{-\infty}^{\infty} p(x_{t+1} \mid Y_T)\, g(x_{t+1} - x_t) / p(x_{t+1} \mid Y_t) dx_{t+1} \\ &= p(x_t \mid Y_t) \left((1 - l) \frac{p(x_{t+1} \mid Y_T)\big|_{x_{t+1} = x_t}}{p(x_{t+1} \mid Y_t)\big|_{x_{t+1} = x_t}} + l \int_{-\infty}^{\infty} \frac{p(x_{t+1} \mid Y_T) h(x_{t+1} - x_t) dx_{t+1}}{p(x_{t+1} \mid Y_t)} \right) \end{aligned} \tag{7}$$

The expectation of this smoother density and its quantiles provide an ex post point estimate and confidence intervals for x_t. The estimated value of λ times T gives a point estimate of the number of "breaks" that are present in the sample.

The Conventional "Structural Break" Model

The conventional "structural break" literature makes no assumption that any value of η_t, apart from 0, is more likely than any other value. The only way this can be true is if $h(\eta)$ is everywhere 0. This tacit assumption implies that $g(\eta)$ integrates to $1 - \lambda$, which must be less than unity if breaks are a true possibility, and therefore that $g(\eta)$ is an improper density.[2]

Theorem 1: If the observation errors ε_t in (1) are drawn from a normal distribution with mean 0 and some variance σ^2, i.e.

$$f(e) = n(e;0,s^2), \tag{8}$$

where

$$n(x;m,s^2) = \frac{1}{\sqrt{2p}\,s}\exp\left(\frac{-(x-m)^2}{2s^2}\right), \tag{9}$$

and the parameter shifts η_t equal 0 with probability $1 - \lambda$, and have zero or essentially zero density elsewhere, then

i. $$p(x_t|Y_t) = n(x_t;\bar{y}_t,s^2/t), \tag{10}$$

ii. $$p(x_t|Y_T) = n(x_t;\bar{y},s^2/T), \tag{11}$$

iii. $$L(q\mid Y_T) = p(Y_T\mid y_1) = \frac{1}{\sqrt{T}}\left(\frac{1-l}{\sqrt{2p}\,s}\right)^{T-1}\exp\left(-\frac{1}{2s^2}\sum_{t=1}^{T}(y_t-\bar{y})^2\right), \tag{12}$$

where

$$\bar{y}_t = \frac{1}{t}\sum_{i=1}^{t} y_i\,, \tag{13}$$

[2] Alternatively, $h(\eta)$ may be thought of as having say a normal distribution, with a very large yet finite variance. In this case, $g(\eta)$ is technically proper yet has a density that is essentially 0 except for its mass point at the origin.

and $\bar{y} = \bar{y}_T$.

Proof:

See Appendix.

Corollary:

It follows immediately from parts i) and ii) of the Theorem that

$$E(x_t|Y_t) = \bar{y}_t, \tag{14}$$

and

$$E(x_t|Y_T) = \bar{y}, \tag{15}$$

regardless of the value of λ. In other words, the appropriate ex post estimate of the mean of the distribution that y_t is being drawn from is simply the average of all the observations, regardless of any appearance of structural breaks. Furthermore, the likelihood iii) above is maximized by

$$\hat{l}_{ML} = 0, \tag{16}$$

again regardless of any appearance of structural breaks.[3]

DISCUSSION AND CONCLUSION

It has been demonstrated that proper ML estimation of a traditional "structural break" model, taken literally, leads to an estimator of the relevant parameter that ignores the possibility of breaks, and to an estimator of the probability of breaks that makes them impossible, no matter how numerous or obvious "regime shifts" may be in the data. This problem arises because of the tacit assumption that any particular sized non-zero regime shift has zero density. This makes a

[3] It is interesting that the ML estimator of the variance using (6) is the sum of squared residuals about the mean divided by (T-1), rather than by T.

finite regime shift infinitely less likely than a draw of long runs of ε's at different levels, no matter how astronomically unlikely the latter may be.

The solution to this impasse is to explicitly model the breaks as coming from a distribution that gives finite density to $h(\eta_t)$. Once this is done, the appropriate estimator of the "breaking" parameter in question is the mean of its smoother density, as computed by structural time series methods.

If λ is simply assumed to be 1 and $f(\eta_t)$ and $g(\eta_t)$ are both taken to be Gaussian with mean 0 and some variances σ^2 and τ^2, a "regime shift" of sorts occurs in every period, with Gaussian density $h(\eta_t) = g(\eta_t)$. In this case, the Sorenson/Alspach filter (4) reduces to the familiar Kalman filter, which may be estimated quickly in closed form. The two variances may then be estimated by ML. Unfortunately, this most easily computed case does not allow dramatic regime shifts of the type that are desired.

Once either $f(\varepsilon_t)$ or $g(\eta_t)$ is non-Gaussian, as is the case if $\lambda < 1$, there is no computational advantage to a Gaussian assumption for the remainder of the system, since the Sorenson/Alspach filter and/or Kitagawa smoother must still be computed numerically. A Laplace (back-to-back exponential) distribution for the breaks, conditional on their occurrance, is easily computed and would allow a wider variety of break sizes to take place than if $h(\eta)$ were taken to be Gaussian. Naturally, if the observed breaks are sufficiently few in number, it will be nearly impossible to determine anything about the precise shape of their distribution empirically. Nevertheless, any proper distribution is better than $h(\eta) \equiv 0$.[4]

[4] Another approach that allows occasional dramatic regime shifts, alongside continual minor regime shifts, is to set $\lambda = 1$ but then to model $g(\eta_t) = h(\eta_t)$ as a leptokurtic distribution such as the symmetric stable or Student. The degree of leptokurtosis, i.e. the stable characteristic exponent or Student degrees of freedom, is then an additional hyperparameter that needs to be estimated, but this is offset by the fact that λ does not. Such a model, with stable

Once system (1) is properly estimated, a proper out-of-sample predictive densities for x_t may be obtained by repeated application of (3) for t > T, so long as $h(\eta_t)$ is non-zero. The predictive density for y_t is then given by

$$\begin{aligned} p(y_t \mid Y_T) &= \int_{-\infty}^{\infty} p(y_t \mid x_t)\, p(x_t \mid Y_T)\, dx_t \\ &= \int_{-\infty}^{\infty} f(y_t - x_t)\, p(x_t \mid Y_T)\, dx_t . \end{aligned} \tag{17}$$

This of course reflects uncertainty that increases with horizon, and that is greater than would be the case if x_t were a constant. But if $h(\eta_t)$ is tacitly assumed to be 0, and the ML estimator of λ overriden with a value λ^* greater than 0,[5] the predictive densities for out-of-sample values of x_t and therefore y_t become improper and increasingly uninformative, with a probability equal to only $(1-\lambda^*)^{(t-T)}$ of any finite value being observed. The mean of such an improper distribution is necessarily undefined.

Having adopted a proper model and estimated it, the presence of breaks can be tested for by constraining λ to be 0 and/or $h(\eta)$ to have zero scale, and constructing a likelihood ratio statistic for this null hypothesis. Unfortunately, this null hypothesis is on the boundary of the parameter space, and there may also be unidentified nuisance parameters under the null, so that the LR statistic will not necessarily have its customary $c^2_{(1)}$ distribution. Nevertheless, its distribution may be ascertained by Monte Carlo simulation.[6]

If breaks are present, they may be approximately dated by examining the quantiles of the computed smoother densities. However, the smoother density will not pinpoint the breaks, and,

---

distributions for both $f(\varepsilon_t)$ and $g(\eta_t)$, has been implemented numerically for bond returns by Oh (1994) and for U.S. inflation by Bidarkota and McCulloch (1998).

[5] Equal, say, to the number of apparent breaks divided by T.

like the filter density, it will in fact ordinarily widen in the vicinity of apparent breaks.[7] If desired, a binary break indicator may be added to the system as an additional state variable, and the probability of a break in each period in addition to the filter and smoother densities for the breaking parameter.

The present note deals exclusively with the simple local level model (1), with a single additive unobserved state variable. Nevertheless, its considerations extend directly to more complicated models, in which one or more parameter may undergo permanent or semi-permanent shifts. If there is more than one unobserved continuous state variable, however, direct numerical integration quickly becomes intractable, since it requires evaluating a multiple integral for each point of a multi-dimensional grid at each t = 1, … T, and then repeating this for each iteration in the likelihood maximization search. In such a case, Monte Carlo integration along the lines proposed by Kitagawa (1996) is the most promising avenue at present.[8]

---

[6] Cp. McCulloch (1997), where LR critical values are simulated for the null hypothesis that the stable distribution characteristic exponent α takes on its Gaussian boundary value of 2.

[7] Cp. Bidarkota and McCulloch (1998).

[8] A binary additional state variable, as suggested in the preceding paragraph, would not require an unwieldy double integral, but merely a manageable pair of single integrals.

APPENDIX

Lemma 1 (elementary):

The product of two normal densities is a third normal density, times a scaling factor. Specifically,

$$n(x;m_1,s_1^2)\,n(x;m_2,s_2^2)=\frac{s_3}{\sqrt{2p}\,s_1 s_2}\exp\left(-\frac{1}{2}\left(\frac{m_1^2}{s_1^2}+\frac{m_2^2}{s_2^2}-\frac{m_3^2}{s_3^2}\right)\right)n(x;m_3,s_3^2), \qquad \text{(A.1)}$$

where

$$\frac{1}{s_3^2}=\frac{1}{s_1^2}+\frac{1}{s_2^2}, \qquad m_3=s_3^2\left(m_1\frac{1}{s_1^2}+m_2\frac{1}{s_2^2}\right) \qquad \text{(A.2)}$$

Proof of Lemma 1:

$$n(x;m_1,s_1^2)\,n(x;m_2,s_2^2)=\frac{1}{2ps_1s_2}\exp\left(-\frac{1}{2}\left(\frac{(x-m_1)^2}{s_1^2}+\frac{(x-m_2)^2}{s_2^2}\right)\right) \qquad \text{(A.3)}$$

while

$$\begin{aligned}\frac{(x-m_1)^2}{s_1^2}+\frac{(x-m_2)^2}{s_2^2}&=x^2\left(\frac{1}{s_1^2}+\frac{1}{s_2^2}\right)-2x\left(\frac{m_1}{s_1^2}+\frac{m_2}{s_2^2}\right)+\frac{m_1^2}{s_1^2}+\frac{m_2^2}{s_2^2}\\&=\frac{(x-m_3)^2}{s_3^2}+\frac{m_1^2}{s_1^2}+\frac{m_2^2}{s_2^2}-\frac{m_3^2}{s_3^2}.\end{aligned} \qquad \text{(A.4)}$$

Multiplying and dividing the RHS of (A.3) by $\sqrt{2p}\,s_3$ and substituting in (A.4) yields (A.1).

///

Proof of Theorem 1:

We have

$$p(x_1|y_1)=n(x_1;y_1,s^2) \qquad \text{(A.5)}$$

This satisfies i) for t = 1. Assume tentatively that i) is true for t-1. The predictive density (3) is then

$$p(x_t \mid Y_{t-1}) = (1-l)p(x_{t-1} \mid Y_{t-1}) \\ = (1-l)n(x_t; \bar{y}_{t-1}, \frac{s^2}{t-1}) \tag{A.6}$$

By (4), the new filter density for time t is

$$p(x_t \mid Y_t) = n(y_t; x_t, s^2)(1-l)n(x_t; \bar{y}_{t-1}, \frac{s^2}{t-1}) / p(y_t \mid Y_{t-1}). \tag{A.7}$$

By Lemma 1 above,

$$n(y_t; x_t, s^2)n(x_t; \bar{y}_{t-1}, \frac{s^2}{t-1}) = n(x_t; y_t, s^2)n(x_t; \bar{y}_{t-1}, \frac{s^2}{t-1}) \\ = \left(\frac{t-1}{2\pi t s^2}\right)^{1/2} \exp\left(-\frac{1}{2s^2}\left(y_t^2 + (t-1)\bar{y}_{t-1}^2 - t\bar{y}_t^2\right)\right) n(x_t; \bar{y}_t, \frac{s^2}{t}). \tag{A.8}$$

Since the third normal density integrates to unity, it follows that the likelihood contribution is just (1-λ) times the scaling factor:

$$p(y_t \mid Y_{t-1}) = (1-l)\left(\frac{t-1}{2\pi t s^2}\right)^{1/2} \exp\left(-\frac{1}{2s^2}\left(y_t^2 + (t-1)\bar{y}_{t-1}^2 - t\bar{y}_t^2\right)\right), \tag{A.9}$$

and that the filter density is this third normal density itself:

$$p(x_t \mid Y_t) = n(x_t; \bar{y}_t, \frac{s^2}{t}). \tag{A.10}$$

This completes the proof of i) by induction. Part iii) follows immediately from (5).

The proof of ii) proceeds by a backward induction that begins by confirming that the proposition in question is valid is true for t = T, and then demonstrates that if it is valid for any t+1, it is also valid for t. By i), ii) is valid for t = T. Suppose that ii) is valid for t+1. Then by (7) with $h(\cdot) \equiv 0$, and using (A.6),

$$p(x_t \mid Y_T) = n(x_t; \bar{y}_t, \frac{s^2}{t})(1-l)\frac{n(x_t; \bar{y}, \frac{s^2}{T})}{(1-l)n(x_{t+1}; \bar{y}_t, \frac{s^2}{t})\Big|_{x_{t+1}=x_t}} \qquad \text{(A.11)}$$

$$= n(x_t; \bar{y}, \frac{s^2}{T}).$$

This completes the proof of ii), and therefore of Theorem 1.

///

REFERENCES

Bai, Jushan, Robin L. Lumsdaine, and James H. Stock (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* **63** : 395-432.

Bidarkota, Prasad V., and J. Huston McCulloch (1998). Optimal univariate inflation forecasting with symmetric stable shocks. *Journal of Applied Econometrics* **13** : 659:70.

Harvey, Andrew C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Kitagawa, Genshiro (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association* **82**: 1032-41. See also comments and rejoinder, pp. 1041-63.

_____ (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* **5**: 1-25.

McCulloch, J. Huston (1997). Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business & Economic Statistics* **15**: 74-81.

Oh, Chang-Seok (1994). *Estimation of Time Varying Term Premia of U.S. Treasury Securities: Using a STARCH Model with Stable Distributions.* Ph.D. dissertation, Ohio State Univ.

Perron, Pierre (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* **57**: 1361-1401.

Sorenson, H.W., and D.L. Alspach (1971). Recursive Bayesian estimation using Gaussian sums. *Automatica* **7**: 465-79.