Bayesian Methods in Nonlinear Time Series

Oleg Korenok *

September 27, 2007

Abstract

This paper reviews the analysis of the threshold autoregressive, smooth threshold autoregressive, and Markov switching autoregressive models from the Bayesian perspective. For each model we start by describing a baseline model and discussing possible extensions and applications. Then we review the choice of prior, inference, tests against the linear hypothesis, and conclude with models selection. A short discussion of recent progress in incorporating regime changes into theoretical macroeconomic models concludes our survey.

JEL classification: C11, C22, C52

Keywords: Threshold, Smooth Threshold, Markov-switching

*This entry has been prepared for the *Springer Encyclopedia of Complexity and Systems Science* (Robert A. Meyers, Ed.) to be published in 2008. The author is grateful to Bruce Mizrach, the Finance and Econometrics Section editor, as well as to David Harless, Carol S. Lehr, Ming Lo, Stan Radchenko, Philip Rothman, and Tara Sinclair for many helpful comments. Oleg Korenok (korenok@vcu.edu): Department of Economics, VCU School of Business, 1015 Floyd Avenue, Richmond, VA 23284.

1 Introduction

Economic fluctuations display definite nonlinear features. Recessions, wars, financial panics, and varying government policies change the dynamics of almost all macroeconomic and financial time series. In the time series literature, such events are modeled by modifying the standard linear autoregressive (AR) model

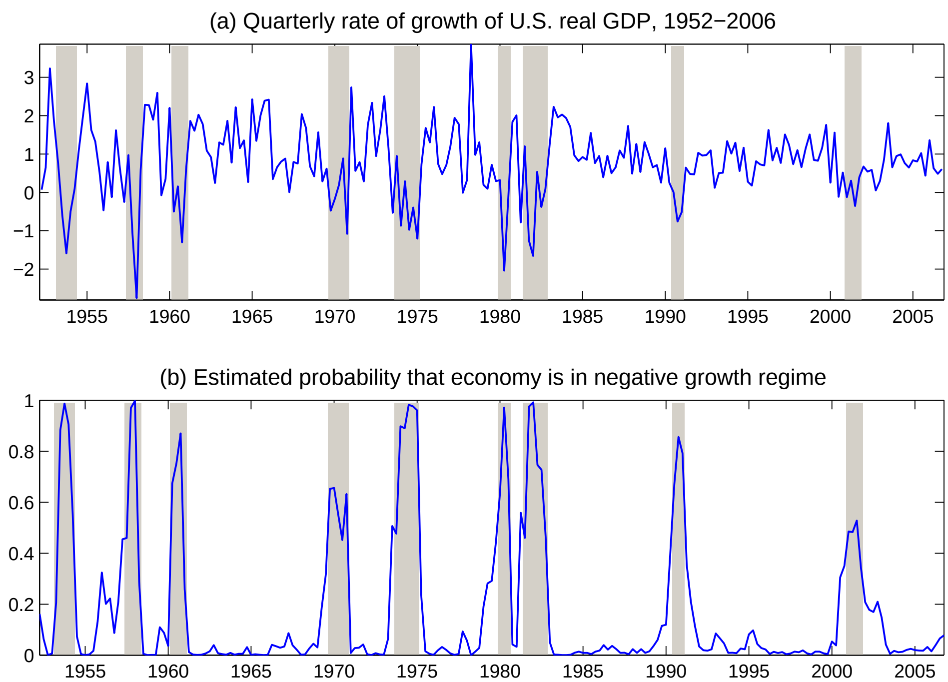
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t,$$

where y_t is a covariance stationary process, ϵ_t is an independent and identically distributed noise process, $\epsilon_t \sim i.i.d.N(0, \sigma^2)$, and the parameters c, ϕ_i, and σ^2 are fixed over time. In particular, the literature assumes that y_t follows two or more regimes. The three most commonly used nonlinear models differ in their description of the transition between regimes. In the threshold autoregressive (TAR) model, regime changes abruptly; in the smooth threshold autoregressive (STAR) model, regime changes slowly. Nevertheless, in both models the regime change depends on the time index or lagged values of y_t. In the Markov switching autoregressive (MAR) model, however, the regime change depends on the past values of an unobserved random variable, the state of the Markov chain, and possibly the lagged values of y_t.

Arguably, the best-known example of the nonlinear time series model is the model of cyclical fluctuations of the U.S. economy. It was first introduced and estimated by Hamilton [40] for quarterly U.S. real Gross National Product over the 1952(II)-1984(IV) period. The model has two discrete regimes. The first regime is associated with a positive 1.2% growth rate and the second regime is associated with a negative -0.4% growth rate. Against his original motivation to find decade-long changes in growth rate trends for the U.S. economy, Hamilton finds that negative growth regimes occur at the business cycle frequency. Positive growth regimes last, on average, 10 quarters, and negative growth regimes last, on average, 4 quarters. Moreover, he finds that the estimated regimes coincide closely with the official National Bureau of Economic Research (NBER) recession dates.

Figure 1 illustrates Hamilton's results for the extended 1952(II)-2006(IV) sample. Panel (a) shows the quarterly growth rate of the U.S. real Gross Domestic Product, currently the more common measure of output; panel (b) plots the estimated probability that the U.S. economy is in a negative growth regime. The shaded regions represent recessionary periods as determined informally and with some delay by the NBER: It took six months for the NBER's Business Cycle Dating Committee to determine the latest peak of the U.S. economy, which occurred in March 2001 but was officially announced in November 2001. Even though the NBER dates were not used in the model, the periods with high probability of a negative growth rate coincide almost perfectly with the NBER dates.

Figure 1: Output Growth and Recession Probabilities in U.S.



In addition to the formal recession dating methodology, Hamilton [40] presents clear statistical evidence for the proposition that the U.S. business cycle is asymmetric: Behavior of output during normal times, when labor, capital, and technology determine long-run economic growth, is distinct from behavior during recessions, when all these factors are underutilized.

Hamilton's paper triggered an explosion of interest in nonlinear time series. The purpose of this paper is to give a survey of the main developments from the Bayesian perspective. The Bayesian framework treats model parameters as random variables and interprets probability as a degree of belief about particular realizations of a random variable conditional on available information. Given the observed sample, the inference updates prior beliefs, formulated before observing the sample, into posterior beliefs using Bayes' theorem

$$p(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)},$$

where y is the sample observations $y = (y_1, ..., y_T)$, θ is the vector of parameters $\theta = (c, \phi_1, ..., \phi_p, \sigma^2)$, $\pi(\theta)$ is the prior distribution that describes beliefs prior to observing the data, $f(y|\theta)$ is the distribution of the sample conditional on the parameters, $f(y)$ is the marginal distribution of the sample, and $p(\theta|y)$ is the posterior distribution that describes the beliefs after observing the sample. Zellner [100], Bauwens, Lubrano, and Richard [83], Koop [90], Lancaster [92], and Geweke [87] cover Bayesian econometrics extensively and provide excellent introductions to relevant computational techniques.

We review the three most commonly used nonlinear models in three separate sections. We start each section by describing a baseline model and discussing possible extensions and applications[1] Then we review the choice of prior, inference, tests against the linear hypothesis, and conclude with models selection. A short discussion of recent progress in incorporating regime changes into theoretical macroeconomic models concludes our survey.

Our survey builds on reviews of the TAR and STAR models in Tong [98], Granger and Terasvirta [86], Terasvirta [96], Bauwens, Lubrano, and Richard [83], Lubrano [93], Potter [94], Franses and van Dijk [85], van Dijk, Terasvirta, and Franses [99], and on reviews of the MAR models in Hamilton [88], Potter [94], and Kim and Nelson [45].

We limit our survey of nonlinear models only to the TAR, STAR, and MAR models. For a reader interested in a wider range of time series models from a Bayesian prospective, we recommend Steel's [95] survey: He overviews linear, as well as nonlinear, and parametric, as well as nonparametric, models.

2 Threshold Autoregressive Model

A threshold regression was introduced by Quandt [63] and was extended to the threshold autoregressive model by Tong [78], [97] and Tong and Lim [79]. Tong [98] had a great impact on popularizing

[1] Matlab implementation of baseline models is available at www.people.vcu.edu/∼okorenok/share/mlab.zip.

TAR models.

We limit our baseline model to a single switching variable z_t. The choice of the switching variable depends on the purpose of the investigation. For the analysis of structural breaks at an unknown point in time, Perron and Vogelsang [59], as well as DeJong [22], among many others, use the time index ($z_t = t$). For the purpose of prediction, Geweke and Terui [35], Chen and Lee [13], and others, use a lagged value of the time series ($z_t = y_{t-d}$), the self-exciting threshold autoregressive (SETAR) model.

In our discussion, the number of lags in the model p and a delay d is fixed. We also limit the baseline model to the homoscedastic case so that the variance of ϵ_t is constant in both regimes.

Introducing a more general notation, $x_t' = (1, y_{t-1}, ..., y_{t-p})$, $\beta' = (c, \phi_1, ..., \phi_p)$, the two-regime TAR model becomes

$$\begin{aligned} y_t &= x_t'\beta_1 + \epsilon_t \quad \text{if } \; z_t < \tau \quad \text{(first regime)}, \\ y_t &= x_t'\beta_2 + \epsilon_t \quad \text{if } \; z_t \geq \tau \quad \text{(second regime)}, \end{aligned}$$

or more succinctly

$$y_t = [1 - I_{[\tau,\infty)}(z_t)]x_t'\beta_1 + I_{[\tau,\infty)}(z_t)x_t'\beta_2 + \epsilon_t, \tag{1}$$

where $I_A(x)$ is an indicator function that is equal to one if $x \in A$, in particular $I_{[\tau,\infty)}(z_t) = 1$ if $z_t \in [\tau, \infty)$. The indicator function introduces the abrupt transition between regimes. It is convenient to rewrite the model in a more compact form

$$y_t = x_t'(\tau)\beta + \epsilon_t, \tag{2}$$

where $x_t'(\tau) = (x_t', I_{[\tau,\infty)}(z_t)x_t')$ and $\beta' = (\beta_1', \delta')$ with $\delta = \beta_2 - \beta_1$.

If the number of observations in regime i is less than or equal to the number of parameters, we cannot estimate parameters, or the model is not identified. In the Bayesian inference, we resolve the identification problem by restricting the region of possible parameter values to the one where the number of observations per regime is greater than the number of regressors.

The baseline model can be extended in several ways. First, we can allow the variance of the error term to differ in each regime. In this case, we rescale the data and introduce an additional parameter $\phi = \frac{\sigma_2^2}{\sigma_1^2}$, as in Lubrano [93]. Second, we can allow the number of lags to differ in each regime. Then p equals to $max\{p_1, p_2\}$.

A more substantial change is required if we want to increase the number of regimes r. We can either use a single transition variable

$$y_t = x_t\beta_i(t) + \sigma_i(t)\epsilon_t,$$

where $i(t) = 1$ if $z_t < \tau_1$, $i(t) = 2$ if $\tau_1 \leq z_t < \tau_2$, ..., $i(t) = r$ if $\tau_r - 1 \leq z_t$; or we can use a combination of two (or more) transition variables as in Astatkie, Watts, and Watt [5], where first stage transition is nested in the second stage transition

$$\begin{aligned} y_t &= [(1 - I_{[\tau_1,\infty)}(z_{1t}))x_t'\beta_1 + I_{[\tau_1,\infty)}(z_{1t})x_t'\beta_2][1 - I_{[\tau_2,\infty)}(z_{2t})] \\ &+ [(1 - I_{[\tau_1,\infty)}(z_{1t}))x_t'\beta_3 + I_{[\tau_1,\infty)}(z_{1t})x_t'\beta_4]I_{[\tau_2,\infty)}(z_{2t}) + \epsilon_t, \end{aligned}$$

nested TAR model.

Also, we can treat either the choice of number of lags, the delay, or the number of regimes as an inference problem. Then p, d, and r are added to the vector of the model parameters, as in Geweke and Terui [35] and Koop and Potter [50].

Finally, the univariate TAR model can be extended to describe a vector of time series as in Tsay [80]. The n dimensional two-regime TAR model can be specified in a manner similar to equation (1) as

$$\begin{aligned} Y_t &= [1 - I_{[\tau,\infty)}(z_t)](C_1 + \Phi_{11}Y_{t-1} + ... + \Phi_{1p}Y_{t-p}) \\ &+ I_{[\tau,\infty)}(z_t)(C_2 + \Phi_{21}Y_{t-1} + ... + \Phi_{2p}Y_{t-p}) + \epsilon_t, \end{aligned}$$

where $Y_t = (y_{1t}, ..., y_{nt})'$ is a $(n \times 1)$ vector, C_1 is a $(n \times 1)$ vector, Φ_{ji}, $j = 1, 2$, $i = 1, ..., p$ are $(n \times n)$ matrices, and $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{nt})$ is a vector of error terms with mean zero and positive definite covariance matrix Σ.

The TAR model has a wide range of applications. Tiao and Tsay [77], Potter [62], Pesaran and Potter [60], Rothman [66], and Koop and Potter [47] demonstrate both statistically significant and economically important nonlinearities in the U.S. business cycle. Pfann, Schotman, and Tschernig [61] find strong evidence of high volatility and low volatility regimes in the behavior of U.S. short-term interest rates. Dwyer, Locke, and Yu [24], Martens, Kofman, and Vorst [55], and Forbes, Kalb, and Kofman [31] describe the relationship between spot and futures prices of the S&P 500 index and model financial arbitrage in these markets as a threshold process. Obstfeld and Taylor [57] study the law of one price and purchasing power parity convergences and find strong evidence of two regimes. They demonstrate fast, months rather than years, convergence when price differences are higher than transaction costs, and slow or no convergence otherwise.

To simplify the exposition, our discussion of inference for all models will be conditional on the initial observations in the sample. We assume that $y_{1-p}, ..., y_0$ are observable. Two alternative treatments are possible. One can treat the initial observations as unobserved random variables and include the marginal density of initial observations into the likelihood. Alternatively, in the Bayesian analysis, one can treat the initial observations as any other parameter and augment the parameter space, θ, with $y_{1-p}, ..., y_0$.

2.1 Prior

The first step in Bayesian inference is to formalize prior beliefs about the model's parameters by choosing functional forms and parameters of prior distributions.

The prior density for τ depends on our choice of z_t. First, we can limit the prior support by the minimum and the maximum of z_t. Second, if $z_t = t$ the threshold is a date, and so the prior density is naturally discrete. If, however, $z_t = y_{t-d}$, the threshold τ is continuous and so is the prior density.

For a model to be identified, we restrict the support of the prior density to the region where the number of observations per regime is greater than the number of regressors. We assign an equal weight to the entire support to get the 'non-informative' prior for τ that is proportional to a constant

$$\pi(\tau) \propto I_{[z_{(k_1)}, z_{(T-k_2)}]}(\tau), \tag{3}$$

where k_1 and k_2 are the number of regressors in the first and second regimes, and the subscript (t) indicates the order in the sample, $z_{(1)} \leq z_{(2)} \leq ... \leq z_{(T)}$. For example, $z_{(1)} = 1$ and $z_{(T)} = T$ if z_t is a time index since the ordering is natural. For an alternative prior distribution of τ see Ferreira

[29].

We assume that the prior density for β and σ^2 is independent of the prior density for τ. Also, because, conditional on τ, the model (2) is linear, we use the natural conjugate prior for β and σ^2

$$\begin{aligned} \pi(\beta|\sigma^2) &= N(\beta|\beta_0, \sigma^2 M_0^{-1}), \\ \pi(\sigma^2) &= IG_2(\sigma^2|\nu_0, s_0), \end{aligned}$$

where $IG_2(.)$ denotes the density of the Inverted Gamma-2 distribution. The functional form of the Inverted Gamma-2 density is given by

$$IG_2(\sigma^2|\nu, s) = \Gamma\left(\frac{\nu}{2}\right)^{-1}\left(\frac{s}{2}\right)^{\frac{\nu}{2}}(\sigma^2)^{-\frac{1}{2}(\nu+2)}exp\left(-\frac{s}{2\sigma^2}\right).$$

The natural conjugate prior allows us to use analytical integration that considerably simplifies the inference.

2.2 Estimation

The next step of the Bayesian analysis is to combine sample information with our prior beliefs to form the posterior beliefs. Given prior distributions, we update prior distributions with the sample likelihood into posterior distributions using Bayes' theorem. The posterior distribution can be further summarized for each parameter with its marginal expectation and variance.

Using the assumption of Normal errors, the likelihood function of the model (2) is

$$f(\beta, \sigma^2, \tau|y) \propto \sigma^{-T} exp\left\{-\frac{1}{2\sigma^2}\sum(y_t - x_t'(\tau)\beta)^2\right\}. \tag{4}$$

The posterior density is a product of the prior and the likelihood

$$p(\beta, \sigma^2, \tau|y) = \pi(\beta|\sigma^2)\pi(\sigma^2)\pi(\tau)f(\beta, \sigma^2, \tau|y). \tag{5}$$

Conditional on the threshold parameter, model (2) is linear. Applying the results from the standard natural conjugate analysis in the linear regression model (for details see Zellner [100]), the posteriors density of β, conditional on threshold and the data, can be obtained by integrating the posterior with respect to σ^2

$$p(\beta|\tau, y) = \int p(\beta, \sigma^2|\tau, y)d\sigma^2 = t(\beta|\beta(\tau), s(\tau), M(\tau), \nu), \tag{6}$$

where $t(.)$ denotes the density of the multivariate Student t-distribution with

$$\begin{aligned}
M(\tau) &= M_0 + \sum x_t(\tau)' x_t(\tau), \\
\beta(\tau) &= M(\tau)^{-1}(\sum x_t(\tau) y_t + M_0 \beta_0), \\
s(\tau) &= s_0 + \beta_0' M_0 \beta_0 + \sum y_t^2 - \beta'(\tau) M(\tau) \beta(\tau), \\
\nu &= \nu_0 + T.
\end{aligned}$$

Further, by integrating equation (6) with respect to β, we obtain the marginal posterior density for τ, which is proportional to the inverse of the integrating constant of $t(\beta|\beta(\tau), s(\tau), M(\tau), \nu)$ times the threshold prior density

$$p(\tau|y) \propto s(\tau)^{-\nu/2} |M(\tau)|^{-1/2} \pi(\tau). \tag{7}$$

Though analytical integration of this function is not available, the fact that it is a univariate function defined on bounded support greatly simplifies the numerical integration.

By integrating numerically the posterior for β conditional on the threshold and the data, we find marginal posterior density for β

$$p(\beta|y) = \int p(\beta|\tau, y) p(\tau|y) d\tau.$$

Finally, using analytical results for the expectation of the conditional density β, we can find the marginal moments of β by integrating only over τ

$$\begin{aligned}
E(\beta|y) &= \int E(\beta|\tau, y) p(\tau|y) d\tau, \\
Var(\beta|y) &= \int Var(\beta|\tau, y) p(\tau|y) d\tau + \int (E(\beta|\tau, y) - E(\beta|y))(E(\beta|\tau, y) - E(\beta|y))' p(\tau|y) d\tau.
\end{aligned}$$

Similarly, applying the results from the standard natural conjugate analysis, we obtain the posterior density of σ^2 conditional on the threshold and the data. Then we integrate out τ numerically to get the marginal posterior density for σ^2

$$p(\sigma^2|y) = \int IG_2(\sigma^2|\nu, s(\tau))p(\tau|y)d\tau,$$

and the marginal moments $E(\sigma^2|y)$ and $Var(\sigma^2|y)$.

2.3 Testing for Linearity and Model Selection

After estimating the TAR model, we might ask whether our data are best characterized by two regimes or a single regime? Model (2) becomes linear when both regimes have identical regression coefficients, so that the difference $\beta_1 - \beta_2 = \delta$ is zero. There are two methods to the test $H_0 : \delta = 0$. The first approach is the Bayesian equivalent of the F-test. Taking into account that β conditional on τ has a Student t-distribution and that the linear transformation of a Student random vector is also a Student, the quadratic transformation of δ

$$\xi(\delta|\tau, y) = (\delta - \delta(\tau))' M_{22.1}(\tau)(\delta - \delta(\tau))\frac{T-k}{k_2 s(\tau)} \tag{8}$$

has a Fisher distribution, where $M_{22.1}(\tau) = M_{22}(\tau) - M_{21}(\tau)M_{11}^{-1}(\tau)M_{12}$, and $\delta(\tau)$ is our estimate. $M(\tau)$ is partitioned by dividing β into β_1 and δ. The posterior 'p-value' of the Bayesian F-test gives the unconditional probability that $\xi(\delta|y)$ exceeds $\xi(\delta = 0|y)$. It can be computed numerically as

$$Pr(\xi(\delta) > \xi(\delta = 0)|y) = \int F(\xi(\delta = 0|y), k_2, T-k)p(\tau|y)d\tau, \tag{9}$$

where $F(\xi(\delta = 0|y), k_2, T-k)$ is the Fisher distribution function with k_2 and $T-k$ degrees of freedom. The null hypothesis is accepted if, for example, $(\xi(\delta) > \xi(\delta = 0)|y)$ is larger than 5%.

The second approach, the posterior odds, is more general, and can also be used to select the number of lags p, the delay parameter d, or the number of regimes r. Koop and Potter [48], [49] advocate and illustrate this approach in the context of the TAR model. To choose between two competing models, m_1 with $\theta_1 = (\beta_1, \delta, \tau, \sigma^2)$ and m_2 with $\theta_2 = (\beta_1, 0, \tau, \sigma^2)$, we calculate the posterior odds ratio

$$po_{12} = \frac{f(y|m_1)\pi(m_1)}{f(y|m_2)\pi(m_2)},$$

where $\pi(m_i)$ is the prior probability for the model i, and $f(y|m_i)$ is the marginal likelihood or marginal density of the sample. Since $f(y|m_i)$ is a normalizing constant of the posterior density, it can be calculated as

$$f(y|m_i) = \int f(y|\theta_i, m_i)\pi(\theta_i|m_i)d\theta_i.$$

With a 'non-informative' prior that assigns equal weight to each model, the posterior odds reduces to the ratio of marginal likelihoods, or the Bayes factor. Again, applying the standard natural conjugate analysis of the linear regression model to the TAR model, the marginal likelihood for model i is

$$f(y|m_i) = \int \frac{\Gamma(\frac{\nu(\tau_i|m_i)}{2})s_0^{\frac{\nu_0}{2}}}{\Gamma(\frac{\nu_0}{2})\pi^{\frac{T}{2}}} s(\tau_i|m_i)^{-\frac{\nu(\tau_i|m_i)}{2}} \left(\frac{|M_0|}{|M(\tau_i|m_i)|}\right)^{\frac{1}{2}} \pi(\tau_i|m_i)d\tau, \tag{10}$$

which can be calculated numerically. The model with the highest marginal likelihood is preferred.

3 Smooth Transition Autoregressive Model

In some applications, imposing an abrupt transition between regimes might be undesirable. For example, if the initial estimate of output is slightly below the threshold, even a small upward revision will result in a substantial change of the forecast in the TAR model. Bacon and Watts [6], in a regression model context, and Chan and Tong [12], in the TAR model context, propose to make the transition between regimes smooth. Terasvirta [75] develops a modeling cycle for the STAR model that includes specification, estimation, and evaluation stages as in the Box and Jenkins [84] modeling cycle for the linear time series model.

In the STAR model, a smooth transition is imposed by replacing the indicator function in equation (1) by the cumulative distribution function

$$y_t = [1 - F(\gamma(z_t - \tau))]x_t'\beta_1 + F(\gamma(z_t - \tau))x_t'\beta_2 + \epsilon_t. \tag{1a}$$

Terasvirta [75] uses the logistic function

$$F(\gamma(z_t - \tau)) = \frac{1}{1 + exp(-\gamma(z_t - \tau))},$$

where $\gamma \in [0, \infty)$ determines the degree of smoothness. As γ increases, smoothness decreases. In the limit, as γ approaches infinity, $F(.)$ becomes an indicator function, with $F(\gamma(z_t - \tau)) \sim 1$ when $z_t \geq \tau$. We can rewrite equation (1a) as

$$y_t = x_t'(\gamma, \tau)\beta + \epsilon_t, \tag{2a}$$

where $x_t'(\gamma, \tau) = (x_t', F(\gamma(z_t - \tau))x_t')$.

Note that the identification problem discussed for the TAR model does not occur in the STAR model. We cannot have fewer observations than regressors because we no longer classify observations into regimes. The new parameter γ, however, introduces a new identification problem. If $\gamma = 0$, the logistic function equals $\frac{1}{2}$ for any value of τ, so τ is not identified. Also $x_t'(\gamma, \tau)$ is perfectly collinear unless the two regimes have no common regressors. Perfect collinearity implies that δ is also not identified. As in the TAR model, we choose such prior densities that resolve the identification problem.

The baseline model can be extended in several directions. Generally, the transition function $F(.)$ is not limited to the logistic function. Any continuous, monotonically increasing function $F(.)$ with $F(-\infty) = 0$ and $F(\infty) = 1$ can be used. For example, the popular alternative to the logistic function is the exponential function

$$F(\gamma(z_t - \tau)) = 1 - exp(-\gamma(z_t - \tau)^2).$$

In the regression model context, Bacon and Watts [6] show that results are not sensitive to the choice of $F(.)$. As in the TAR model, we can increase the number of regimes either with a single transition variable

$$y_t = x_t'\beta_1 + F(\gamma_1(z_t - \tau_1))x_t'(\beta_2 - \beta_1) + ... + F(\gamma_r(z_t - \tau_r))x_t'(\beta_r - \beta_{r-1}) + \epsilon_t,$$

or with a combination of transition variables

$$\begin{aligned} y_t &= [(1 - F(\gamma_1(z_{1t} - \tau_1)))x_t'\beta_1 + F(\gamma_1(z_{1t} - \tau_1))x_t'\beta_2][(1 - F(\gamma_2(z_{2t} - \tau_2)))] \\ &+ [(1 - F(\gamma_1(z_{1t} - \tau_1)))x_t'\beta_3 + F(\gamma_1(z_{1t} - \tau_1))x_t'\beta_4][F(\gamma_2(z_{2t} - \tau_2))] + \epsilon_t. \end{aligned}$$

See van Dijk and Franses [81] for a discussion of the multiple regime STAR model.

Also, we can treat the choice of number of lags p, delay d, or number of regimes r as an inference problem, adding p, d, and r to the vector of parameters in the model. In addition, we can allow the variance of the error term to change between regimes, or more generally, use an autoregressive conditional heteroscedasticity form as in Lundbergh and Terasvirta [53], or a stochastic volatility form as in Korenok and Radchenko [51].

Finally, similar to the TAR model, the univariate STAR model can be extended to model a vector of time series as in Granger and Swanson [37]. The n dimensional two-regime STAR model can be specified as

$$\begin{aligned} Y_t &= [1 - F(\gamma(z_t - \tau))](C_1 + \Phi_{11}Y_{t-1} + ... + \Phi_{1p}Y_{t-p}) \\ &+ F(\gamma(z_t - \tau))(C_2 + \Phi_{21}Y_{t-1} + ... + \Phi_{2p}Y_{t-p}) + \epsilon_t, \end{aligned}$$

where we use the same notation as in the multivariate TAR model.

Applications of the STAR model include models of the business cycles, real exchange rates, stock and futures prices, interest rates, and monetary policy. Terasvirta and Anderson [76] and van Dijk and Franses [81] demonstrate nonlinearities in the U.S. business cycles. Skalin and Terasvirta [70] find similar nonlinearities in Swedish business cycles. Michael, Nobay, and Peel [56], Sarantis [68], and Taylor, Peel, and Sarno [73] show that the real exchange rate nonlinearly depends on the size of the deviation from purchasing power parity; Lundbergh and Terasvirta [54] and Korenok and Radchenko [51] use the STAR model to fit the behavior of exchange rates inside a target zone. Taylor, van Dijk, Franses, and Lucas [74] describe the nonlinear relationship between spot and futures prices of the FTSE100 index. Anderson [1] uses the STAR model to study yield movements in the US Treasury Bill Market. Finally, Rothman, van Dijk, and Franses [67] find evidence of a nonlinear relationship between money and output; Weise [82] demonstrates that monetary policy has a stronger effect on output during recessions.

3.1 Prior

As in the TAR model, the natural conjugate priors for β and σ^2 facilitate analytical integration. Bauwens, Lubrano, and Richard [83] impose the identification at $\gamma = 0$ by modifying the prior density of β

$$\pi(\beta|\sigma^2,\gamma) = N(\beta|0,\sigma^2 M_0^{-1}(\gamma)),$$

where, assuming prior independence between β_1 and δ, M_0 is defined as

$$M_0(\gamma) = \begin{pmatrix} M_{0,11} & 0 \\ 0 & M_{0,22}/exp(\gamma) \end{pmatrix}.$$

As γ gets closer to zero, the prior variance falls, increasing precision around $\delta = 0$. The choice of $\delta = 0$ is consistent with the linear hypothesis, which can be formulated as either $\delta = 0$ or $\gamma = 0$. When γ is positive, prior precision about $\delta = 0$ decreases as variance rises, so more weight is given to the information in the sample. We keep the natural conjugate prior of σ^2 without modifications.

We do not modify the prior for the threshold parameter τ. When γ is large, the smooth transition function is close to the step transition function. Thus, we prefer to limit the prior to the region where the number of observations per regime is greater than the number of regressors to avoid the TAR identification problem.

The prior for the smoothness parameter, γ, cannot be 'non-informative' or flat. As $\gamma \to \infty$ the smooth transition function becomes a step transition with a strictly positive likelihood. This means that the marginal likelihood function of γ is not integrable. To avoid the integration problem, Bauwens, Lubrano, and Richard [83] use the truncated Cauchy density

$$\pi(\gamma) \propto (1+\gamma^2)^{-1} I_{[0,\infty)}(\gamma).$$

3.2 Estimation

Inference in the STAR model follows the TAR methodology, taking into account the additional parameter γ, and the new definitions of $M_0(\gamma)$ and $x_t(\tau,\gamma)$.

In particular, the likelihood function of model (2a) is

$$f(\beta,\sigma^2,\tau,\gamma|y) \propto \sigma^{-T} exp\left\{-\frac{1}{2\sigma^2}\sum(y_t - x_t'(\tau,\gamma)\beta)^2\right\}, \tag{4a}$$

the posterior density is

$$p(\beta, \sigma^2, \tau, \gamma|y) = \pi(\beta|\sigma^2)\pi(\sigma^2)\pi(\tau)\pi(\gamma)f(\beta, \sigma^2, \tau, \gamma|y), \tag{5a}$$

and the joint posterior density of τ and γ is proportional to the inverse of the integrating constant of the Student t-density $t(\beta|\beta(\tau,\gamma), s(\tau,\gamma), M(\tau,\gamma), \nu)$ times the prior densities for c and γ

$$p(\tau, \gamma|y) \propto |s(\tau,\gamma)|^{-(T-k)/2}|M(\tau,\gamma)|^{-1/2}\pi(\tau)\pi(\gamma), \tag{7a}$$

where

$$\begin{aligned} M(\tau,\gamma) &= M_0(\gamma) + \sum x_t(\tau,\gamma)' x_t(\tau,\gamma), \\ \beta(\tau,\gamma) &= M(\tau,\gamma)^{-1}(\sum x_t(\tau,\gamma) y_t + M_0(\gamma)\beta_0), \\ s(\tau,\gamma) &= s_0 + \beta_0' M_0(\gamma)\beta_0 + \sum y_t^2 - \beta'(\tau,\gamma)M(\tau,\gamma)\beta(\tau,\gamma), \\ \nu &= \nu_0 + T. \end{aligned}$$

This function is bivariate and can be integrated numerically with respect to τ and γ. Then, as in the TAR model, we use numerical integration to obtain marginal densities and moments for β and σ^2.

Compared to the TAR model, β_1 and β_2 cannot be interpreted as regression coefficients in regime 1 and regime 2. Smooth transition implies that the effect of change in x_t on y_t is a weighted average of two regimes with weights changing from one observation to the other.

3.3 Testing for Linearity and Model Selection

The STAR model becomes linear when either $\delta = 0$ or $\gamma = 0$. The test for $H_0 : \delta = 0$ is equivalent to the test in the TAR model. The quadratic transformation of δ

$$\xi(\delta|\tau, \gamma, y) = (\delta - \delta(\tau,\gamma))' M_{22.1}(\tau,\gamma)(\delta - \delta(\tau,\gamma))\frac{T-k}{k_2 s(\tau,\gamma)}, \tag{8a}$$

where $M_{22.1}(\tau,\gamma) = M_{22}(\tau,\gamma) - M_{21}(\tau,\gamma)M_{11}^{-1}(\tau,\gamma)M_{12}(\tau,\gamma)$, has a Fisher distribution. We can find the posterior 'p-value' of the Bayesian F-test numerically as

$$Pr(\xi(\delta) > \xi(\delta = 0)|y) = \int\int F(\xi(\delta = 0|y), k_2, T-k)p(\tau,\gamma|y)d\tau d\gamma. \tag{9a}$$

The null hypothesis is accepted, for example, if $(\xi(\delta) > \xi(\delta = 0)|y)$ is larger then 5%.

The test for $H_0 : \gamma = 0$ can be conducted using the 95% highest posterior density interval (HPDI), defined as the smallest interval with 95% probability of γ to be in the interval

$$\begin{aligned} \max_h PDI(h) &= \left\{\gamma | \int p(\tau,\gamma)\pi(\tau)d\tau \geq h\right\}, \\ s.t. \quad & Pr(PDI(h)) \geq 0.95. \end{aligned}$$

The null hypothesis is accepted, for example, if $\gamma = 0$ is inside the 95% HPDI.

As in the TAR model, linearity tests and model selection can be conducted using posterior odds. In the STAR model, the marginal likelihood for model i is given by

$$f(y|m_i) = \int\int \frac{\Gamma(\frac{\nu(\tau_i,\gamma_i|m_i)}{2})s_0^{\frac{\nu_0}{2}}}{\Gamma(\frac{\nu_0}{2})\pi^{\frac{T}{2}}} s(\tau_i,\gamma_i|m_i)^{-\frac{\nu(\tau_i,\gamma_i|m_i)}{2}} \left(\frac{|M_0|}{|M(\tau_i,\gamma_i|m_i)|}\right)^{\frac{1}{2}} \pi(\tau_i|m_i)\pi(\gamma_i|m_i)d\tau_i d\gamma_i, \tag{10a}$$

which can be calculated numerically. The model with the highest marginal likelihood is preferred.

4 Markov-Switching Model

Unlike the threshold models, where the regime transition depends on a time index or on lagged values of y_t, the Markov-switching autoregressive model relies on a random variable, s_t. A Markov-switching regression was introduced in econometrics by Goldfeld and Quandt [36] and was extended to the Markov-switching autoregressive model by Hamilton [40].

As in the threshold models, we limit our baseline MAR model to two regimes that differ only in mean. The variance of the error term is constant. The number of lags p is determined by the model choice. The two-regime MAR model becomes

$$
\begin{aligned}
(y_t - \mu_{s_t}) &= \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu_{s_{t-i}}) + \epsilon_t, \\
\mu_{s_t} &= \mu_0 \qquad\quad \text{if} \;\; s_t = 0 \quad \text{(first regime)}, \\
\mu_{s_t} &= \mu_0 + \mu_1 \quad \text{if} \;\; s_t = 1 \quad \text{(second regime)},
\end{aligned}
\tag{11}
$$

where $\mu_{s_t} = \mu_0 + s_t \mu_1$. An unobserved discreet random variable s_t takes only integer values of 0 or 1. The transition probability $Pr(s_t = j | s_{t-1} = i) = p_{ij}$ that state i will be followed by state j depends only on s_{t-1}, the first order Markov-switching process, with transition probability matrix

$$
P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix}.
$$

Since we have only two possible regimes and $p_{i1} + p_{i2} = 1$, we estimate only two free parameters, the probabilities of remaining in the same regime p_{11} and p_{22}. We also assume that, conditional on previous history of states $s = (s_1, ..., s_T)'$, the transition probabilities are independent of other parameters and the data.

In general, we do not have a clear association between regimes and the state indicator. This introduces an identification problem when we change regime identifiers, 0 and 1, and accordingly change $\mu_0^* = \mu_0 + \mu_1$ and $\mu_1^* = -\mu_1$. For example, if $s_t = 0$ during recessions, then the long run average during recessions is μ_0 and the long-run average during expansions is $\mu_0 + \mu_1$. On the other hand, if $s_t = 0$ during expansions, then the long-run average during expansions is $\mu_0^* = \mu_0 + \mu_1$ and the long-run average during recessions is $\mu_0^* - \mu_1$ or $\mu_1^* = -\mu_1$.

The second identification problem occurs in the MAR model when $\mu_1 = 0$; the model becomes linear. In this case, the conditional mean $E(y_t | s_t = 0) = E(y_t | s_t = 1) = \mu_0$ is independent of the state realizations, s, and transition probability matrix, P. Neither s nor P are identified.

The baseline model can be extended in several directions. The Markov-switching component can be modified by increasing the number of regimes as in Calvet and Fisher [9] and Sims and Zha [69] or by increasing the order of the Markov-switching process so that s_t depends on $s_{t-1}, ..., s_{t-r}$. Both changes can be incorporated by increasing the number of states in the baseline model, as in Hamilton [88].

Diebold, Lee, and Weinbach [20], Filardo [30], and Peria [58] relax the assumption of time invariant Markov-switching by making the transition probabilities depend on lagged values of y_t. In most applications, however, relatively few transitions between regimes makes it difficult to estimate the transition probabilities and restricts model choice to two or three regimes with time-invariant probabilities.

The error term can be modified by introducing regime-switching for the variance of the error term as in Hamilton and Susmel [42], and Cai [8]; by relaxing the assumption of Gaussian density for the error term as in Dueker [23]; or by specifying a general Markov-switching moving average structure for the error term as in Billio, Monfort, and Robert [7].

Finally, the univariate Markov-switching model can be extended to a multivariate model. Diebold and Rudebusch [21] propose a model where a number of time series are driven by a common unobserved Markov-switching variable, the dynamic factor model. The dynamic factor model captures the fact that many economic series show similar changes in dynamic behavior during recessions. Krolzig [91] provides a detailed exposition of how the baseline model can be extended to the Markov-switching vector autoregressive model.

The applications of the MAR model include models of business cycles, interest rates, financial crises, portfolio diversification, options pricing, and changes in government policy. Hamilton [40], Filardo [30], Diebold and Rudebusch [21], Kim and Nelson [45], Kim and Piger [46], and Hamilton [41] find statistically significant evidence that expansionary and contractionary phases of the U.S. business cycle are distinct. Hamilton [39], Cai [8], Garcia and Perron [32], Gray [38], Dueker [23], Smith [71], Hamilton [41], and Dai, Singleton, and Yang [16] describe dramatic changes in interest rate volatility associated with the OPEC oil shocks, the changes in the Federal Reserve operating procedures in 1979-1982, and the stock market crash of October 1987. Ang and Bekaert [3] show a similar increase in volatility in Germany during the reunification period. Jeanne and Masson [43] use the MAR model to describe the crisis of the European Monetary System in 1992-1993; Cerra and Saxena [11] find permanent losses in output after the Asian crisis. Ang and Bekaert [2] report that the correlation between international equity returns is higher during bear markets relative to bull markets. Radchenko [64] shows that gasoline prices respond faster to a permanent oil price change compared to a transitory change. Finally, Sims and Zha [69] document abrupt changes of shocks to U.S. monetary policy, and Davig and Leeper [18] document the regime changes in fiscal policy.

4.1 Prior

As in the threshold models, the natural conjugate priors facilitate considerably the integration of the posterior density. Conditional on s_t, μ_0, and μ_1, the MAR model is linear

$$y_t(s_t) = x_t'(s_t)\tilde{\phi} + \epsilon_t, \tag{12}$$

where $y_t(s_t) = y_t - \mu_{s_t}$, $x_t'(s_t) = (y_{t-1} - \mu_{s_{t-1}}, ..., y_{t-p} - \mu_{s_{t-p}})$, and $\tilde{\phi} = (\phi_1, ..., \phi_p)'$. For the regression coefficient $\tilde{\phi}$ and the variance of the error term σ^2, the natural conjugate prior is given by

$$\begin{aligned} \pi(\tilde{\phi}|\sigma^2) &= N(\tilde{\phi}|\tilde{\phi}_0, \sigma^2 M_{0,\phi}^{-1}) I_A(\tilde{\phi}), \\ \pi(\sigma^2) &= IG_2(\sigma^2|\nu_0, s_0), \end{aligned}$$

where A is a region where the roots of polynomial $1 - \phi_1 L - ... - \phi_p L^p = 0$ lie outside the complex unit circle. This restriction imposes stationarity on $y_t(s_t)$.

Conditional on s_t and $\tilde{\phi}$, the MAR model is also linear

$$y_t(\tilde{\phi}) = x_t'(\tilde{\phi})\tilde{\mu} + \epsilon_t, \tag{13}$$

where $y_t(\tilde{\phi}) = y_t - \sum_{i=1}^p \phi_i y_{t-p}$, $x_t'(\tilde{\phi}) = (1, s_t - \sum_{i=1}^p \phi_i s_{t-p})$, and $\tilde{\mu} = (\mu_0, \mu_1)'$. The natural conjugate prior for $\tilde{\mu}$ is

$$\pi(\tilde{\mu}) = N(\tilde{\mu}|\tilde{\mu}_0, M_{0,\mu}^{-1}) I_{(0,\infty)}(\mu_1),$$

where the indicator function imposes an identification constraint. In particular, we constrain the mean of the second regime to be greater than the mean of the first regime and in this way fix the order of regimes. We also impose that $\mu_1 \neq 0$.

Finally, Kim and Nelson [45] show that the natural conjugate prior for the vector of transition probabilities $\tilde{p} = (p_{11}, p_{22})'$ is

$$\pi(\tilde{p}) = B(p_{11}|\alpha_1, \beta_1) B(p_{22}|\alpha_2, \beta_2),$$

where $B(.)$ denotes the density of Beta distribution defined on the interval $[0, 1]$.

4.2 Estimation

In the Bayesian approach, we add realizations of the vector of states to the model parameters: $\theta = (\mu_0, \mu_1, \phi_1, ..., \phi_p, \sigma, p_{11}, p_{22}, s_1, ..., s_T)'$. Analytical or numerical integration of the posterior density $p(\theta|y)$, where θ is $p + 5 + T \times 1$, may be difficult.

Albert and Chib [4] developed inference methodology that overcomes the curse of dimensionality using Gibbs-sampling, a Markov chain Monte Carlo simulation method of integration. The technique was further refined by Kim and Nelson [44]. Monte Carlo integration takes random draws from the posterior density and, by averaging them, produces estimates of moments. In particular, Gibbs-sampling allows us to generate many draws $\theta^{(g)}, g = 1, ..., G$, from joint density of $p(\theta|y)$ using only conditional densities $p(\theta_i|\theta_{i \neq j}, y)$ either for all i or for blocks of parameters. The joint and marginal distribution of $\theta^{(g)}$ converge at an exponential rate to the joint and marginal distribution of θ under fairly weak conditions. Casella and George [10], Gelfand and Smith [33], and Geweke [34] provide the details.

To implement the Gibbs-sampling simulation, we have to describe the conditional posterior distributions for all parameters or parameter blocks. It is convenient to separate parameters into five blocks: the state vector s, the transition probabilities $\tilde{p}$, the regression coefficients $\tilde{\phi}$ in the conditional linear model (12), the regression coefficients $\tilde{\mu}$ in the conditional linear model (13), and the variance of the error term σ^2.

The state vector s is a first-order Markov process, which implies that given s_{t+1} all information, for example $s_{t+2}, ..., s_T$ and $y_{t+1}, ..., y_T$, is irrelevant in describing s_t. Then the posterior density of s conditional on other parameters becomes

$$p(s|\tilde{p}, \tilde{\phi}, \tilde{\mu}, \sigma^2, y) = p(s_T|\tilde{p}, \tilde{\phi}, \tilde{\mu}, \sigma^2, y) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, \tilde{p}, \tilde{\phi}, \tilde{\mu}, \sigma^2, y^t), \tag{14}$$

where $y^t = (y_1, ..., y_t)'$. The functional form of the posterior density suggests that we can generate draw of the state vector recursively. First we generate the last element s_T. Then, conditional on s_T, we generate s_{T-1}. More generally, conditional on s_{t+1}, we generate s_t for $t = T-1, T-2, ..., 1$.

To generate the state vector, Kim and Nelson [44] use the output from Hamilton's [40] filter.

To facilitate exposition, we suppress the conditioning on parameters and consider first a model without lags.

Hamilton's filter starts from the observation that, before observing the data, the probability of finding the state in regime j, $Pr(s_0 = j|y^0)$, equals the unconditional probability, $Pr(s_t = j)$, which is proportional to the eigenvector of P associated with unitary eigenvalue.

Using transition probabilities and the probability of observing regime j conditional on observations obtained through date t, $Pr(s_t = j|y^t)$, we predict the next period regime

$$Pr(s_{t+1} = j|y^t) = Pr(s_t = 0|y^t)p_{0j} + Pr(s_t = 1|y^t)p_{1j}. \tag{15}$$

Once y_{t+1} is observed, we update the prediction using Bayes rule

$$Pr(s_{t+1} = j|y^{t+1}) = Pr(s_{t+1} = j|y_{t+1}, y^t) = \frac{f(y_{t+1}|s_{t+1} = j, y^t)Pr(s_{t+1} = j|y^t)}{f(y_{t+1}|y^t)}, \tag{16}$$

where the numerator is the joint probability of observing y_{t+1} and $s_{t+1} = j$, which is a product of the probability of observing y_{t+1} given that state s_{t+1} is in regime j (for example $f(y_{t+1}|s_{t+1} = 0, y^t) = N(\mu_0, \sigma^2)$) and our prediction from equation (15). The denominator is the unconditional density of observing y_{t+1}, which is a sum of the numerator over all possible regimes

$$f(y_{t+1}|y^t) = \sum_j f(y_{t+1}|s_{t+1} = j, y^t)Pr(s_{t+1} = j|y^t). \tag{17}$$

Starting from $Pr(s_0 = j|y^0)$, the filter iterates through equations (15) - (17) until we calculate $Pr(s_t = j|y^t)$ for every t and j. As a by-product of the filter we obtain the likelihood function

$$f(\tilde{\phi}, \tilde{\mu}, \tilde{p}, \sigma^2, s|y) = \prod_t f(y_{t+1}|y^t). \tag{18}$$

For the AR(1) model, the filter should be adjusted. Given $Pr(s_t = j|y^t)$, we forecast the next period regime and the previous period regime jointly, taking one summand in equation (15) at a time

$$Pr(s_{t+1} = j, s_t = i|y^t) = p_{ij}Pr(s_t = i|y^t), \tag{15a}$$

for $j = 0, 1$ and $i = 0, 1$. After y_{t+1} is observed, we update our prediction to

$$Pr(s_{t+1} = j, s_t = i|y^{t+1}) = \frac{f(y_{t+1}|s_{t+1} = j, s_t = i, y^t)Pr(s_{t+1} = j, s_t = i|y^t)}{f(y_{t+1}|y^t)}, \tag{16a}$$

where $f(y_{t+1}|s_{t+1} = j, s_t = i, y^t)$ is the density of observing y_{t+1} given that state s_{t+1} is in regime j and state s_t is in regime i (for example $f(y_{t+1}|s_{t+1} = 0, s_t = 0, y^t) = N(\mu_0 + \phi_1(y_t - \mu_0), \sigma^2))$

$$f(y_{t+1}|y^t) = \sum_j \sum_i f(y_{t+1}|s_{t+1} = j, s_t = i, y^t)Pr(s_{t+1} = j, s_t = i|y^t). \tag{17a}$$

Summing (16a) over i,

$$Pr(s_{t+1} = j|y^{t+1}) = \sum_i Pr(s_{t+1} = j, s_t = i|y^{t+1}), \tag{19}$$

finishes the iteration. Iterating through equations (15a) -(17a) and (19) we get $Pr(s_t = j|y^t)$ for every t and j. The extension to a more general AR(p) model is similar.

The output of Hamilton's filter gives only the first term in the product (14), which is sufficient to generate s_T. To generate the other states s_t conditional on y^t and s_{t+1}, $t = T-1, T-2, ..., 1$, we again use Bayes rule

$$Pr(s_t = j|s_{t+1} = i, y^t) = \frac{p_{ji}Pr(s_t = j|y^t)}{\sum_j p_{ji}Pr(s_t = j|y^t)}, \tag{20}$$

where $Pr(s_t = j|y^t)$ is the output from Hamilton's filter. Since s_t is a discrete random variable taking on values 0 and 1, we can generate it by drawing random numbers from uniform distribution between 0 and 1, and comparing them to $Pr(s_t = 1|s_{t+1} = i, y^t)$.

Conditional on other parameters in the model, the likelihood function of transition probabilities reduces to a simple count n_{ij} of transitions from state i to state j

$$f(\tilde{p}|\tilde{\mu}, \tilde{\phi}, \sigma_2, s, y) = p_{11}^{n_{11}}(1 - p_{11})^{n_{12}} p_{22}^{n_{22}}(1 - p_{22})^{n_{21}},$$

which is the product of the independent beta distributions. The posterior distribution for the transition probabilities conditional on the other parameters is a product of independent beta distributions

$$p(\tilde{p}|\tilde{\phi}, \tilde{\mu}, \sigma^2, s, y) = B(\alpha_1 + n_{11}, \beta_1 + n_{12})B(\alpha_2 + n_{22}, \beta_2 + n_{21}).$$

To derive posterior distributions for $\tilde{\phi}$, $\tilde{\mu}$, and σ^2 conditional on other parameters, we use standard results for a linear model with the natural conjugate priors. The natural conjugate priors are reviewed, for example, by Geweke [87], Koop [90], or Lancaster [92]. In particular, the conditional distribution of the regression coefficients is Normal

$$\begin{aligned} p(\tilde{\phi}|\tilde{p}, \tilde{\mu}, \sigma^2, s, y) &= N(\Sigma_\phi(\sigma^{-2}M_{0,\phi}\tilde{\phi}_0 + \sigma^{-2}\sum x_t(s)'y_t(s)), \Sigma_\phi)I_A(\tilde{\phi}), \\ p(\tilde{\mu}|\tilde{p}, \tilde{\phi}, \sigma^2, s, y) &= N(\Sigma_\mu(M_{0,\mu}\tilde{\mu}_0 + \sigma^{-2}\sum x_t(\tilde{\phi})'y_t(\tilde{\phi})), \Sigma_\mu)I_{(0,\infty)}(\mu_1), \end{aligned}$$

where $\Sigma_\phi = \left(\sigma^{-2}M_{0,\phi} + \sigma^{-2}\sum x_t(s)'x_t(s)\right)^{-1}$, $\Sigma_\mu = \left(M_{0,\mu} + \sigma^{-2}\sum x_t(\tilde{\phi})'x_t(\tilde{\phi})\right)^{-1}$. The conditional distribution for the variance of error term is Inverted Gamma-2

$$p(\sigma^2|\tilde{p}, \tilde{\phi}, \tilde{\mu}, s, y) = IG_2\left(s_0 + \sum(y_t(s_t) - x_t'(s_t)\tilde{\phi})^2, \nu_0 + T\right).$$

4.3 Testing for Linearity and Model Selection

Given our prior, the linear model is not nested in the MAR model. To test against a linear model, we use the Bayes factor. We also use the Bayes factor to select the number of regimes and the number of lags.

The Bayes factor is a ratio of marginal likelihoods of the alternative models. To find the marginal likelihood, we need to integrate the product of the likelihood function and the prior density with respect to all parameters. To avoid the curse of dimensionality, Chib [14] shows the marginal likelihood can be computed from the output of the Gibbs sampler requiring only that the integrating constants of the conditional posterior distributions be known. This requirement is satisfied for the natural conjugate priors.

From the Bayes's theorem it follows that the identity

$$f(y) = \frac{f(y|\theta)\pi(\theta)}{p(\theta|y)},$$

holds for any θ. The complete functional form of the numerator is given by the product of the likelihood (18) and the prior densities. Chib suggests evaluating the denominator, the posterior density, at the posterior mode θ^*. Then the posterior density at the posterior mode can be written as

$$p(\theta^*|y) = p(\tilde{\mu}^*|y) \times p(\tilde{\phi}^*|\tilde{\mu}^*, y) \times p(\tilde{\sigma}^{2*}|\tilde{\mu}^*, \tilde{\phi}^*, y) \times p(\tilde{p}^*|y, \mu^*, \tilde{\phi}^*, \sigma^{2*}).$$

The first term

$$p(\tilde{\mu}^*|y) = \int p(\tilde{\mu}^*|\tilde{\phi}, \sigma^2, \tilde{p}, s, y)p(\tilde{\phi}, \sigma^2, \tilde{p}, s|y)d\tilde{\phi}\, d\sigma^2 d\tilde{p}\, ds,$$

can be estimated by averaging over the full conditional density

$$\hat{p}(\tilde{\mu}^*|y) = G^{-1}\sum_{g=1}^{G} p(\tilde{\mu}^*|\tilde{\phi}^{(g)}, \sigma^{2(g)}, \tilde{p}^{(g)}, s^{(g)}, y).$$

This estimate converges at an exponential rate to the true marginal distribution of $\tilde{\mu}$.

In the the second term

$$p(\tilde{\phi}|\tilde{\mu}^*, y) = \int p(\tilde{\phi}^*|\tilde{\mu}^*, \sigma^2, \tilde{p}, s, y)p(\sigma^2, \tilde{p}, s|\tilde{\mu}^*, y)d\sigma^2 d\tilde{p}\, ds,$$

the complete conditional density of $\tilde{\phi}$ cannot be averaged directly because the Gibbs sampler does not provide draws conditional on $\tilde{\mu}^*$. We generate necessary draws by additional G iterations of the original Gibbs sampler, but instead of generating $\tilde{\mu}$ we set it equal to $\tilde{\mu}^*$. Then the estimate of the second term

$$\hat{p}(\tilde{\phi}^*|\tilde{\mu}^*, y) = G^{-1}\sum_{g=G+1}^{2G} p(\tilde{\phi}^*|\tilde{\mu}^*, \sigma^{2(g)}, \tilde{p}^{(g)}, s^{(g)}, y),$$

converges at an exponential rate to the true $p(\tilde{\phi}|\tilde{\mu}^*, y)$. Similarly, by generating additional draws from the Gibbs sampler we compute $\hat{p}(\tilde{\sigma}^{2*}|\tilde{\mu}^*, \tilde{\phi}^*, y)$ and $\hat{p}(\tilde{p}^*|y, \mu^*, \tilde{\phi}^*, \sigma^{2*})$.

Substituting our estimate of posterior density into marginal likelihood results in

$$lnf(y) = lnf(y|\theta^*) + ln\pi(\theta^*) - ln\hat{p}(\tilde{\mu}^*|y) - ln\hat{p}(\tilde{\phi}^*|\tilde{\mu}^*, y) - ln\hat{p}(\tilde{\sigma}^{2*}|\tilde{\mu}^*, \tilde{\phi}^*, y) - ln\hat{p}(\tilde{p}^*|y, \mu^*, \tilde{\phi}^*, \sigma^{2*}).$$

The model with the highest marginal likelihood is preferred.

5 Future Directions

Given the large volume of evidence collected in the nonlinear time series, incorporating regime-switching policies and disturbances into general equilibrium models may lead to a better understanding of monetary and fiscal policies.

Over the years, the time series literature has collected substantial statistical evidence that output, unemployment, and interest rates in the U.S. exhibit different behavior in recessions and expansions. Contrary to the real business cycle models in which short-run and long-run fluctuations have the same origin, the statistical evidence suggests that the forces that cause output to rise may be quite different from those that cause it to fall.

Also, many studies provide evidence that monetary and fiscal policies have changed substantially throughout U.S. history. Taylor [72], Clarida, Gali, and Gertler [15], Romer and Romer [65], and Lubik and Schorfheide [52] show that, since the mid-1980s, the Fed reacted more forcefully to inflation. Favero and Monacelli [28] and Davig and Leeper [18] demonstrate that U.S. fiscal policy has fluctuated frequently responding to wars, recessions, and more generally to the level of debt. Sims and Zha [69], after extensive comparison of 17 regime-switching structural VAR models, report that their best-fitting model requires nine regimes to incorporate the large shocks, for example, generated by the OPEC oil embargo or the Vietnam War. They conclude that, "It is time to abandon the idea that policy change is best modelled as a once-and-for-all, nonstochastic regime switch" (p. 56).

The research by Davig and Leeper [17], [18], [19] and Farmer, Waggoner, and Zha [25], [26], [27] show considerable promise in introducing nonlinear regime-switching components into dynamic stochastic general equilibrium models. For example, Davig and Leeper [18] estimate regime-switching rules for monetary policy and tax policy and incorporate them into the otherwise standard new-Keynesian model. Unlike expansionary fiscal policy in the fixed-regime model, fiscal expansion in the regime-switching model increases inflation and output.

6 Bibliography

Primary Literature

[1] Anderson HM (1997) Transaction costs and nonlinear adjustment towards equilibrium in the US treasury bill market. Oxford Bulletin of Economics and Statistics, Vol 59, pp 465-484.

[2] Ang A and Bekaert G (2002a) International asset allocation with regime shifts. Review of Financial Studies, Vol 15, pp 1137-1187.

[3] Ang A and Bekaert G (2002b) Regime switches in interest rates. Journal of Business and Economic Statistics, Vol 20, pp 163-197.

[4] Albert JH and Chib S (1993) Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business and Economic Statistics, Vol 11, No 1, pp 1-15.

[5] Astatkie T, Watts DG, Watt WE (1997) Nested threshold autoregressive NeTAR models. International Journal of Forecasting, Vol 13, pp 105116.

[6] Bacon DW, Watts DG (1971) Estimating the transition between intersecting straight lines. Biometrica, No 62, pp 525-534.

[7] Billio M, Monfort A, Robert CP (1999) Bayesian estimation of switching ARMA models. Journal of Econometrics, Vol 93, pp 229-255.

[8] Cai J (1994) A Markov model of switching-regime ARCH. Journal of Business and Economic Statistics, Vol 12, pp 309-316.

[9] Calvet L and Fisher A (2004) Regime switching and the estimation of multifractal processes. Journal of Financial Econometrics, Vol 2, pp 49-83.

[10] Casella G and George EI (1992) Explaining the Gibbs sampler. The American Statistician, Vol 46, pp 167-174.

[11] Cerra V and Saxena SC (2005) Did output recover from the Asian crisis? IMF Staff Papers 52, 1-23.

[12] Chan KS and Tong H (1986) On estimating thresholds in autoregressive models. Journal of Time Series Analysis, Vol 7, 178-190.

[13] Chen CWS and Lee JC (1995) Bayesian inference of threshold autoregressive models. Journal of Time Series Analysis, No 16, pp 483-492.

[14] Chib S (1995) Marginal likelihood from the Gibbs output. Journal of the American Statistical Association, Vol 90, pp 1313-1321.

[15] Clarida R, Gali J, and Gertler M (2000) Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics, Vol 115, pp 147180.

[16] Dai Q Singleton KJ, Yang W (2007), Regime shifts in a dynamic term structure model of U.S. treasury bonds. The Review of Financial Studies, forthcoming.

[17] Davig T and Leeper E (2005) Generalizing the Taylor principle. National Bureau of Economic Research, Working Paper No 11874.

[18] Davig T and Leeper E (2006) Fluctuating macro policies and the fiscal theory. In: Acemoglu D, Rogoff, and Woodford M (ed) NBER Macroeconomic Annual. MIT Press: Cambridge.

[19] Davig T and Leeper E (2007) Generalizing the Taylor principle. The American Economic Review, Forthcoming.

[20] Diebold FX, Lee JH, and Weinbach GC (1994) Regime switching with time-varying transition probabilities. In: Hargreaves C (ed) Nonstationary Time Series Analysis and Cointegration, Oxford University Press, Oxford.

[21] Diebold FX and Rudebusch GD (1996) Measuring business cycles: a modern perspective. Review of Economics and Statistics, Vol 78, pp 6777.

[22] DeJong DN (1996) A Bayesian Search for Structural Breaks in U.S. GNP. In: Fomby TB (ed) Advances in Econometrics: Bayesian Methods Applied to Time Series Data, Vol 11, Part B, pp 109-146. JAI Press, Greenwich, Connecticut.

[23] Dueker M (1997) Markov switching in GARCH processes and mean-reverting stock-market volatility. Journal of Business and Economic Statistics, Vol 15, pp 26-34.

[24] Dwyer GP, Locke P, Yu W (1996) Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash. Review of Financial Studies, 9, pp 301332.

[25] Farmer RE, Waggoner DF, and Zha T (2006a) Indeterminacy in a forward looking regime switching model. NBER Working Paper No 12540.

[26] Farmer RE, Waggoner DF, and Zha T (2006b) Minimal state variable solutions to Markov-switching rational expectations models. Unpublished Manuscript.

[27] Farmer RE, Waggoner DF, and Zha T (2007) Understanding the New-Keynesian model when monetary policy switches regimes. Unpublished Manuscript.

[28] Favero CA and Monacelli T (2005) Fiscal policy rules and regime (in)stability: evidence from the U.S. Manuscript, IGIER.

[29] Ferreira PE (1975) A Bayesian analysis of a switching regression model: known number of regimes. Journal of the American Statistical Association, Vol 70, pp 370-374.

[30] Filardo AJ (1994) Business cycle phases and their transitional dynamics. Journal of Business and Economic Statistics, Vol 12, pp 299-308.

[31] Forbes CS, Kalb GRJ, Kofman P (1999) Bayesian arbitrage threshold analysis. Journal of Business and Economic Statistics, Vol 17, pp 364-372.

[32] Garcia R, and Perron P (1996) An analysis of real interest under regime shift. Review of Economics and Statistics, Vol 78, pp 111-125.

[33] Gelfand AE and Smith AFM (1990) Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, Vol 85, No 410, pp 398-409.

[34] Geweke, J (1999) Using simulation methods for Bayesian econometric models: inference, development and communication. Econometric Reviews, Vol 18, pp 1-127.

[35] Geweke J and Terui N (1993) Bayesian threshold autoregressive models for nonlinear time series. Journal of Time Series Analysis, No 14, pp 441-454.

[36] Goldfeld, SM and Quandt RE (1973) A Markov model for switching regressions, Journal of Econometrics, Vol 1, pp 3-16.

[37] Granger CWJ and Swanson NR (1996) Future developments in the study of cointegrated variables. Oxford Bulletin of Economics and Statistics, Vol 58, pp 537553.

[38] Gray SF (1996) Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics, Vol 42, pp 27-62.

[39] Hamilton JD (1988) Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. Journal of Economic Dynamics and Control, Vol 12, 385-423.

[40] Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, Vol 57, No 2, pp 357-384.

[41] Hamilton JD (2005) Whats real about the business cycle? NBER Working Paper No W11161.

[42] Hamilton JD and Susmel R (1994) Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics, Vol 64, 207-333.

[43] Jeanne O and Masson P (2000) Currency crises, sunspots, and Markov- switching regimes. Journal of International Economics 50, 327-350.

[44] Kim CJ and Nelson CR (1998) Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching. Review of Economics and Statistics, Vol 80, No 2, pp 188-201.

[45] Kim CJ and Nelson CR (1999) Friedman's plucking model of business fluctuations: tests and estimates of permanent and transitory components. Journal of Money, Credit, and Banking, Vol 31, pp 317-334.

[46] Kim, CJ and Piger J (2002) Common stochastic trends, common cycles, and asymmetry in economic fluctuations. Journal of Monetary Economics, Vol 49, pp 1189-1211.

[47] Koop G and Potter SM (1999a) Dynamic asymmetries in U.S. unemployment. Journal of Business and Economic Statistics, Vol 17, pp 198-312.

[48] Koop G and Potter SM (1999b) Bayes factors and nonlinearity: evidence from economic time series. Journal of Econometrics, Vol 88, pp 251-281.

[49] Koop G and Potter SM (2000) Nonlinearity, structural breaks or outliers in economic time series? In: Barnett B and Johansen S (ed) Nonlinear Econometric Modeling in Time Series. Cambridge University Press: Cambridge.

[50] Koop G and Potter SM (2003) Bayesian analysis of endogenous delay threshold models. Journal of Business and Economic Statistics, Vol 21, No 1, pp 93-103.

[51] Korenok O and Radchenko R (2005) The smooth transition autoregressive target zone model with the Gaussian stochastic volatility and TGARCH error terms with applications. VCU Economics Department, No 0505.

[52] Lubik TA and Schorfheide F (2004) Testing for indeterminacy: an application to U.S. monetary policy. American Economic Review, Vol 94, pp 190217.

[53] Lundbergh S and Terasvirta T (1998) Modelling economic high-frequency time series with STAR-GARCH models. Working Paper Series in Economics and Finance No. 291, Stockholm School of Economics.

[54] Lundbergh S and Terasvirta T (2006) A time series model for an exchange rate in a target zone with applications. Journal of Econometrics, Vol 131, pp 579-609.

[55] Martens M, Kofman P, Vorst ACF (1998) A threshold error correction for intraday futures and index returns. Journal of Applied Econometrics, 13, pp 245263.

[56] Michael P, Nobay AR, Peel DA (1997) Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation. Journal of Political Economy, Vol 105, 862879.

[57] Obstfeld M, Taylor AM (1997) Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. Journal of the Japanese and International Economies, Vol 11, 441-479.

[58] Peria MSM (2002) A regime-switching approach to the study of speculative attacks: a focus on EMS crises. In: Hamilton JD and Raj B (eds) Advances in Markov-Switching Models, Physica-Verlag, Heidelberg.

[59] Perron P and Vogelsang TJ (1992) Nonstationarity and level shifts with an application to purchasing power parity. Journal of Business and Economic Statistics, Vol 10, pp 301-320.

[60] Pesaran MH and Potter S (1997) A floor and ceiling model of US output. Journal of Economic Dynamics and Control, Vol 21, pp 661-695.

[61] Pfann GA, Schotman PC, Tschernig R (1996) Nonlinear interest rate dynamics and the implications for the term structure. Journal of Econometrics, Vol 74, pp 149-176.

[62] Poter S (1995) A nonlinear approach to US GNP. Journal of Applied Econometrics, Vol 10, pp 109-125.

[63] Quandt RE (1958) The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association, Vol 53, pp 873-880.

[64] Radchenko S (2005) Lags in the response of gasoline prices to changes in crude oil prices: the role of short-term and long-term shocks. Energy Economics, Vol 27, pp 573-602.

[65] Romer CD and Romer DH (2002) A rehabilitation of monetary policy in the 1950s. American Economic Review, Vol 92, pp 121127.

[66] Rothman P (1998) Forecasting asymmetric unemployment rates. Review of Economics and Statistics, Vol 80, pp 164168.

[67] Rothman P, van Dijk D, Franses PH (2001) A multivariate STAR analysis of the relationship between money and output. Macroeconomic Dynamics, Vol 5, pp 506-532.

[68] Sarantis N, (1999) Modeling non-linearities in real effective exchange rates. Journal of International Money and Finance, 18, 2745.

[69] Sims C and Zha T (2006) Were there switches in U.S. monetary policy? American Economic Review, Vol 96, pp 54-81.

[70] Skalin J, Terasvirta T (1999) Another look at swedish business cycles. Journal of Applied Econometrics, Vol 14, 359378.

[71] Smith DR (2002) Markov-switching and stochastic volatility diffusion models of short-term interest rates. Journal of Business and Economic Statistics, Vol 20, pp 183-197.

[72] Taylor JB (1999) An historical analysis of monetary policy rules. In: Taylor JB (ed) Monetary Policy Rules, pp 319341.

[73] Taylor MP, Peel DA, Sarno L (2001) Nonlinear mean-reversion in exchange rate rates: towards a solution to the purchasing power parity puzzles. International Economic Review, Vol 42, pp 1015-1042.

[74] Taylor N, van Dijk D, Franses PH, Lucas A (2000) SETS, arbitrage activity, and stock price dynamics. Journal of Banking and Finance, Vol 24, pp 12891306.

[75] Terasvirta T (1994) Specification, estimation and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association, Vol 89, pp 208-219.

[76] Terasvirta T and Anderson H (1992) Characterising nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics, pp 119-136.

[77] Tiao CG and Tsay RS (1994) Some advances in non linear and adaptive modelling in time series. Journal of Forecasting, Vol 13, pp 109-131.

[78] Tong H (1978) On a threshold model. In: Chan CH (ed) Pattern Recognition and Signal Processing, Sijthoff and Noordhoff, Amsterdam.

[79] Tong H and Lim KS (1980) Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society, B42, pp 245292.

[80] Tsay RS (1998) Testing and modeling multivariate threshold models. Journal of the American Statistical Association, Vol 93, pp 11881202.

[81] van Dijk D and Franses PH (1999) Modeling multiple regimes in the business cycle. Macroeconomic Dynamics, Vol 3, pp 311340.

[82] Weise CL (1999) The Asymmetric effects of monetary policy. Journal of Money, Credit and Banking, Vol 31, pp 85108.

Books and Reviews

[83] Bauwens L, Lubrano M, Richard JF (1999) Bayesian Inference in Dynamic Econometric Models. Oxford University Press, New York.

[84] Box GEP and Jenkins GM (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

[85] Franses PH and van Dijk D (2000) Nonlinear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge.

[86] Granger CWJ and Terasvirta T (1993) Modeling Nonlinear Economic Relationships. Oxford University Press, Oxford.

[87] Geweke J (2005) Contemporary Bayesian Econometrics and Statistics. Wiley.

[88] Hamilton J (1994) Time Series Analysis. Princeton University Press, Princeton, New Jersey.

[89] Kim CJ, Nelson CR (1999) State-Space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications. MIT Press, Cambridge, Massachusetts.

[90] Koop G (2003) Bayesian Econometrics. Wiley.

[91] Krolzig HM (1997) Markov-Switching Vector Autoregressions: Modeling, Statistical Inference, and Application to Business Cycle Analysis. Springer, Berlin.

[92] Lancaster T (2004) An Introduction to Modern Bayesian Econometrics. Blackwell Publishing.

[93] Lubrano M (1999) Bayesian Analysis of Nonlinear Time Series Models with a threshold. In: Barnett WA, Hendry DF, Hylleberg S, Terasvirta T, Tjostheim D, Wurts A (ed) Nonlinear Econometric Modeling, Cambridge University Press, Cambridge, Massachusetts.

[94] Potter SM (1999) Nonlinear time series modelling: an introduction. Journal of Economic Surveys, Vol 13, pp 505528.

[95] Steel M (2007) Bayesian Time Series Analysis. In Durlauf S, Blume L, (ed) The New Palgrave Dictionary of Economics, 2nd ed., Palgrave Macmillan, London, forthcoming.

[96] Terasvirta T (1998) Modelling economic relationships with smooth transition regressions. In Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp 507552.

[97] Tong H (1983) Threshold Models in Non-Linear Time Series Analysis. Lecture Notes in Statistics. No 21, Springer, Heidelberg.

[98] Tong H (1990) Nonlinear Time Series: A Dynamical System Approach. Oxford University Press, Oxford.

[99] van Dijk D, Terasvirta T, Franses PH (2002) Smooth transition autoregressive models - a survey of recent developments. Econometric Reviews, Vol 21, pp 1-47.

[100] Zellner A (1971) An Introduction to Bayesian Inference in Econometrics. Wiley.