Reconnecting the Markov switching model with economic fundamentals

Ulf G. Erlandsson*

December 31, 2003

Abstract

This paper seeks to investigate and remedy the apparent inability of Markov regime switching models to predict future states in the medium to long term. We show that projected time varying transition probability series in the model may be biased towards predicting regime switches with high probability in the short run, and as a consequence it is hard or impossible to obtain longer run inference. Also, we show that ordinary likelihood ratio tests of the transition equation variables yield significant results too often in limited samples. We propose a penalized maximum likelihood estimator where non-smoothness in the transition series has negative influence on the likelihood function. This remedies both the short run bias and the spuriousity. In an empirical investigation of U.S. real GDP, the penalized model works better in terms of matching the NBER business cycles as well as for forecasting the probability of contractionary states for horizons longer than 4 quarters.

Keywords: regime switching, transition probability, forecasting
JEL classification: C13; C32; E32

*Department of Economics, Lund University; P.O. Box 7082; 220 07 LUND; email: ulf.erlandsson@nek.lu.se;" home-page: www.nek.lu.se/nekuer

Introduction

Few non-linear time series models have attained the same level of popularity as the the Markov regime switching model of Hamilton (1989) . The division of economic variables into different states, such as contraction and expansions phases of GDP, enjoys a intuitive advantage over more continuos models. A very large part of contemporary research focuses on improving the Hamilton model by including various extensions from other parts of the time-series literature in order to match the moments in the data.

A less probed but equally important area emphasizes the other novel invention following the Hamilton model: the determination of states. We may very well match moments of the model perfectly but if we assume an incorrect or non-optimal structure to determine the probability of states, the performance of the model will suffer. In the plain vanilla Hamilton model, the transition between states is governed by a constant probability. This assumption is rarely questioned, although intuitively we accept that the probability that a bull-market continues is higher after only a few months of rising stock prices than when the bull-market has persisted for a longer period of time. To reflect this in the model, one may let the transition probability be time varying. The first published paper to recognize this possibility is Diebold, Lee and Weinbach (1994).

A number of papers have applied the Diebold et al. methodology: Gray (1996) for interest rates, Tronzanon, Psaradakis and Sola (2003) for target zones and Abiad (2003) in the currency crisis context. Results have been mixed, and it seems that one of the main reasons for the relatively infrequent use of this approach is the difficulty to obtain sensible parameter estimates. The problems are exacerbated in the multivariate transition equation setting. Statistical inference in this environment can be very difficult, since the model in essence tries to estimate a model with two unobserved variables. As always in the Markov regime switching (henceforth denoted MS) model, we try to estimate the unobserved state variable. Moreover, we seek the relation between exogenous variables and the dependent variable in the transition equation which also is unobserved.

This paper seeks to establish that the maximum likelihood estimator will be biased toward finding a parameterization of the model that leads to a projected transition probability series with very abrupt shifts. In general, a parameterization that produces a high probability to switch regimes a short time interval prior to an observed shift will be preferred to a parameterization with a lower probability to switch but during a longer time period, irrespective of the true data generating process.

Although the bias reflects an optimal moment-matching, the use of the model becomes deeply restricted due to this bias. For policy purposes, parameter values may be of lesser importance than the ability to obtain early warnings of an upcoming switch. The exact size of the contraction/expansions trends in a model of GDP are relatively unimportant to a policy maker who intends to take measures that only have lagged effects. The average time before an interest rate change gives effect is estimated to be around 12-24 months. For a central banker, early warnings of the type ""with a 50% probability, GDP growth will enter the" "contraction phase within the next 18 months"" will be preferred to a statement of the type ""with 95% probability, we will enter the contraction phase next month.""

We provide a possible solution to the short run bias problem by introducing a penalty term in the log likelihood function. It penalizes non-smooth behavior of the transition probabilities series with a weight chosen by the researcher. In that sense, it is an ad hoc approach, but our results indicate that this addition remedies a number of problems inherent in the standard estimation procedure. For example, using non-stationary variables in the transition equation introduces spurious significance of these variables. Although the effect is not as severe as in the traditional spurious regression setting, it is too large to be ignored. Using a penalty, it is possible to reduce the spuriousity to nominal levels.

In an empirical excercise, we investigate U.S. real GDP in a regime switching setting. The penalized maximum likelihood method finds a different and much smaller set of variables to include in the final model than the ordinary maximum likelihood does. When forecasting the NBER recessions, the penalized model exhibits better performance than the benchmark models for horizons exceeding 4 quarters, both in- and out-of-sample.

In section 2, we introduce the baseline model and the maximum likelihood estimation procedure. Thereafter we discuss MS models with time varying transition probabilities (TVP) as a proxy for the threshold models. Section 4 is dedicated to a formal proof in a very simple model that the maximum likelihood estimator will select a short-run variable prior to a longer run one, irrespectively of the data generating process (DGP). In section 5, simulation evidence corroborates these results in the stochastic setting. We propose a remedy to the short run bias as well, and analyze the effects of using a penalty term in various settings using Monte Carlo analysis. We apply the proposed method on actual data in section 6. Section 7 concludes.

Model and Estimation Procedure

We base the discussion on a simple form of the Hamilton (1989) Markov regime switching model. The baseline model with constant transition probabilities:

$$\Delta y_t = \mu_{S_t} + \epsilon_t \tag{1}$$

where S_t is a state variable that follows a first order Markov chain with transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{2}$$

where, in turn, p_{ij} denotes the probability to go from state i to state j. The iterative procedure to estimate this kind of model is presented in Hamilton (1994).

The number of extensions made to this simple model, and the combinations thereof, can be counted in the hundreds. Most of these seek to engineer to model in a way as to have a better fit to the data, modifying the elements of equation 1, e.g. by introducing exogenous variables, auto-regressive parameters and ARCH effects. A smaller number of studies, e.g. Diebold, Lee and Weinbach (1994) , have focused on modelling the probability to switch to other regimes, as in equation 2, noting that $\mathbf{P}$ by no means have to be constant. In the general 2 state case:

$$\mathbf{P}_t\left(\mathbf{Z}_t\right) = \begin{bmatrix} h\left(\mathbf{Z}_t^1\right) & 1-h\left(\mathbf{Z}_t^1\right) \\ 1-g\left(\mathbf{Z}_t^2\right) & g\left(\mathbf{Z}_t^2\right) \end{bmatrix} \tag{3}$$

where $f, g \in [0,1]$.[1] This will be referred to the time-varying transition probability (TVP) model. The functional form of f, g is usually chosen to be of probit of logit type. We will assume the logit style functional form for both h, g such that:

$$h\left(x\right) = g\left(x\right) = \frac{\exp\left(x\right)}{1+\exp\left(x\right)} = f(x) \tag{4}$$

In order to estimate the Hamilton Markov regime switching model we iterate on the equations:

$$\xi_{\mathbf{t}|\mathbf{t}} = \frac{\xi_{\mathbf{t}|\mathbf{t-1}} \odot \eta_{\mathbf{t}}}{\mathbf{1}'\left(\xi_{\mathbf{t}|\mathbf{t-1}} \odot \eta_{\mathbf{t}}\right)} \tag{5}$$

and

$$\xi_{\mathbf{t}|\mathbf{t-1}} = \mathbf{P}' \cdot \xi_{\mathbf{t-1}|\mathbf{t-1}} \tag{6}$$

where η_t is a (NxT) matrix of each N states conditional density based on the parameter vector θ. For the 2 state case:

$$\eta_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{\frac{-(y_t-\mu_1)^2}{2\sigma_1^2}\right\} \\ \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{\frac{-(y_t-\mu_2)^2}{2\sigma_2^2}\right\} \end{bmatrix} \tag{7}$$

The log likelihood to be maximized is given by:

$$L(\theta) = \sum_{t=1}^{T} \log \mathbf{1}'\left(\xi_{t|t-1} \odot \eta_{\mathbf{t}}\right) \tag{8}$$

A number of other estimation methods are available, see e.g. Filardo and Gordon (1998) for a Gibbs sampling approach in the time varying transition probability context.

[1] A discussion of the case where the number of states exceeds 2 is saved for later.

Bounded Regime Switching Processes

To be able to conduct simulation excercises, we now introduce a regime switching parameterization where the projected time-series is only dependent upon the parameter vector and a single vector of random disturbances.

The process will be based on the previously considered model in equation 1:

$$\Delta y_t = \mu_{S_t} + \epsilon_t \Rightarrow y_t = y_{t-1} + \mu_{S_t} + \epsilon_t \tag{9}$$

The long run drift of y_t can be calculated using the ergodic (unconditional) probabilities:

$$P\left(S_t = j\right) = \pi = \begin{bmatrix} (1-p_{22})/(2-p_{11}-p_{22}) \\ (1-p_{11})/(2-p_{11}-p_{22}) \end{bmatrix} \tag{10}$$

which can be obtained by solving the eigenvalue problem $|\mathbf{P} - \lambda \mathbf{I}_2| = 0$. Explicitly, the long drift in the N state model becomes

$$\overline{\mu} = \sum_{j=1}^{N} P\left(S_t = j\right) \cdot \mu_j$$

Using the long term drift, it easy to see that in the long-run, this process will mimic a random walk with drift. For many purposes, however, it seems unreasonable that a variable - in the long run - should follow such a process. Examples could be trade balance, debt-to-GDP ratios and real exchange rates. It is likely to be some reversion back to some, yet undefined, mean once we reach a level that is much higher or lower than the posited mean. Assume that $\mu_2 = -\mu_1$ and $\mu_1 \geq 0$. Moreover, assume that there is a bound $a \geq 0$ so that $P\left(S_{t+1} = 2 \middle| y_t \geq a\right) = 1$ and $P\left(S_{t+1} = 1 \middle| y_t \leq a\right) = 1$. In words, if the level process y_t exceeds/goes below $a/-a$, we automatically switch back to a state that reverts the process in the other direction. This is analogously to a Threshold Auto-Regressive (TAR) model. Another way to express this is that the boundary model in effect has time varying transition probabilities. The two transition matrices are:

$$\mathbf{P}_t = \begin{cases} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{if} \quad \|y_t\| > a \\ \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} & \text{if} \quad \|y_t\| \leq a \end{cases} \tag{11}$$

The long run mean $\overline{y}$ of the bounded process can be calculated but most noteworthy here is the fact that the process y_t will never be 'far' away from its long-run mean, which makes an argument for the variable to posses a form of stationarity. For this purpose, we establish the following definition of *global stationarity*:

1. There exists a long-run mean to which the the series returns
2. The long-run variance is time-invariant and finite.
3. The long-run theoretical auto-correlation diminishes when lag-length increases.

An even more flexible, and foremost sensible,[2] version of the bounded model is where the probability to go to the reversion state is dependent upon the distance the process is from its long run mean. Hence, we would posit, with $\|\cdot\|$ denoting an appropriate metric, that $P\left(S_{t+1}=j|S_t=i;\|y_t-\overline{y}t\|\right)$ is large when $\|y_t-\overline{y}t\|$ is. We can translate this to the more general form of the transition matrix in equation 3:

$$\mathbf{P}_t\left(\beta;\|y_t-\overline{y}t\|\right)=\begin{bmatrix} f\left(\beta_1\cdot\|y_t-\overline{y}t\|\right) & 1-f\left(\beta_1\cdot\|y_t-\overline{y}t\|\right) \\ 1-f\left(\beta_2\cdot\|y_t-\overline{y}t\|\right) & f\left(\beta_2\cdot\|y_t-\overline{y}t\|\right) \end{bmatrix} \tag{12}$$

By setting $\beta_1 \neq \beta_2$, the boundaries are allowed to be asymmetric. Once the process y_t goes 'far' off from the long run mean, the probability that we will switch to the state where we revert back increases. In the limit, this switch will happen with probability 1. Hence, the same argument for global stationarity as in the fixed threshold setting applies. We also note that the process does not have to have both an upper and lower bound in order to be globally stationary. If the long run drift term is positive/negative, the process will be globally stationary if there is an upper/lower bound.

Using the baseline model (1) with the associated transition matrix in (12), we see that we have a process that is (1) globally stationary and (2) only a function of the parameter vector $\theta \in \{\mu,\beta\}$ and the vector of disturbances ϵ. Besides the advantage of being parsimonious, this model possesses the quality of being well-suited for simulation excercises, since the sources of variation in the artificial data are reduced to a minimum.

The Short Run Paradox

Now we will consider a simple case of the TVP model which brings about a paradox with serious economic implications. Consider the growth of the debt-to-GDP ratio. We assume that there are two states: the first where debt is growing, and a second where debt is decreasing. Economic theory suggest that there is an upper bound to how much debt in relation to income can grow, since rational lenders will not supply more credit once debt-to-GDP reaches an 'unsustainable' level. The point in time where credit dries up due to unsustainability will be depicted as a regime switch, where the ratio cannot rise anymore,[3] but switches to the decreasing state as the government is forced by creditors to impose policies to this end. There will be some heterogeneity in opinion of what is a sustainable level; hence the timing of the switch is not predictable, although the probability to switch is. One motivation for the sudden shift is that there seems to be herding effects. Once a large investor

[2]The discrete boundaries previously considered cannot be estimated with gradient based optimization algorithms. As in the TAR case, one usually would resort to grid-based estimation procedures.

[3]Of course, GDP can decrease with the effect of the ratio rising, but this is not likely to happen for extended periods of time. We will not consider that case here.

stops buying a certain country's debt, other investors tend to follow to avoid a liquidity squeeze.

To estimate such a model on empirical data, we would use the set-up (state 1 corresponds to the rising debt state):

$$\Delta y_t = \mu_{S_t} + \epsilon_t \tag{13}$$

with

$$\mathbf{P}_t\left(\theta; y_{t-1}\right) = \begin{bmatrix} f\left(\alpha_1 + \beta_1 \cdot y_{t-1}\right) & 1 - f\left(\alpha_1 + \beta_1 \cdot y_{t-1}\right) \\ p_{21} & p_{22} \end{bmatrix}$$

where Δy_t is the growth of the debt-to-GDP ratio. In this case, it seems unlikely that there is a long-run drift of the dependent variably, so $\overline{\mu}$ in equation (12) is set to zero.

In empirical work, however, we do not have the luxury of knowing exactly what variables to include in the transition equations, but have to discriminate between a number of possible candidates. Besides using the dependent variable itself, let us assume we also observe a binary variable called x_t that takes on the value 1 one unit of time prior to the crisis and is zero otherwise. Hence, we estimate the model with the transition matrix:

$$\mathbf{P}_t\left(\theta; y_{t-1}\right) = \begin{bmatrix} f\left(\alpha_1 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot x_{t-1}\right) & 1 - f\left(\alpha_1 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot x_{t-1}\right) \\ p_{21} & p_{22} \end{bmatrix}$$

In this model, we will find that β_2 is very significant and β_1 insignificant, even if the data is generated using the model the model in (13)! The reason for this will be shown below. At this stage, we want to note that for policy purposes, a model that gives us as much advance warning of an oncoming debt crisis as possible will preferred to one that gives us very little time to react. But paradoxically the best econometric fit is obtained with a model that is more or less worthless for polic purposes since it only gives advance warning in the period prior to the crisis.

To see why we by traditional econometric criteria will select the variable with the short duration prior to the shift, we should observe the likelihood function. Suppose there occurs a regime shift at time T and that $S_t = 1$ for $t = 1, 2, ..., T-2, T-1$. We have two binary possible candidates: x^A that produces a low probability $(1-\pi_1) = 1 - f(x^A)$ to switch to regime 2 from time $T - j : T - 1$, $j > 1$; and x^B that produces a very high probability $(1-\pi_2) = 1 - f(x^B)$ to switch at time $T-1$ but a 0 probability otherwise, and $(1-\pi_1) < (1-\pi_2) \quad \Rightarrow \quad \pi_1 > \pi_2$. The corresponding transition matrices are

(A)

$$\mathbf{P}^A_{T-j:T-2} = \begin{bmatrix} \pi_1 & 1-\pi_1 \\ p_{21} & p_{22} \end{bmatrix} \qquad \Rightarrow \qquad \mathbf{P}^A_{T-1} = \begin{bmatrix} \pi_1 & 1-\pi_1 \\ p_{21} & p_{22} \end{bmatrix}$$

and

(B)

$$\mathbf{P}^B_{T-j:T-2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{P}^B_{T-1} = \begin{bmatrix} \pi_2 & 1-\pi_2 \\ p_{21} & p_{22} \end{bmatrix}$$

We furthermore assume that we are certain to have been in regime one for the whole period prior to the switch so that $\xi_{t|t} = [1 \quad 0]'$ for $t = 1, 2..., T-2, T-1$. Because of that, if we disregard the mean zero random part, we can simplify η_t in (7) to

$$\eta_t = [a \quad b]'$$

for $t = 1...T-1$ and we also have that $a > b > 0$. For the time $t = T...T+k$, we have

$$\eta_t = [b \quad a]'$$

Just looking at the time $T-1$ transition matrices, (B) gives us a better explanation of the dependent variable than (A) since To verify this, we note that at time T, the inner part of the likelihood functions is:

$$L_T = \mathbf{1}' \left(\xi_{T|T-1} \odot \eta_T \right) = \mathbf{1}' \left(\mathbf{P}'_{T-1} \xi_{T-1|T-1} \odot \eta_T \right) \tag{14}$$

For the case (A) and (B) this reduces to: $\underbrace{\pi_1 b + (1-\pi_1)a}_{L_T^A}$ and $\underbrace{\pi_2 b + (1-\pi_2)a}_{L_T^B}$.

Setting these equal and solving yields:

$$\underbrace{\pi_1}_{+} \underbrace{(b-a)}_{-} = \underbrace{\pi_2}_{+} \underbrace{(b-a)}_{-}$$

so that

$$L_T^A < L_T^B$$

It is straightforward to verify that also $L_{T-j:T-1}^A < L_{T-j:T-1}^B$. Hence, an ordinary maximum likelihood estimator would prefer variable (B) to variable (A) in this setting. So far, this is not controversial.

But what happens if we try to estimate a model where both x^A, x^B are included in the transition equations? To do this we need to elaborate on the relation between the transition matrix and the functions producing it. Consider that $\pi_1 = f(\alpha_1 - \beta_1 x^A)$, so that a higher value of β_1 means a higher probability to switch regimes, and $\pi_2 = f(\alpha_2 - \beta_2 x^B)$. We assume that both x variables are positive. We first note that a change of β_1 has effects on two likelihood elements. The first is at time $T-1$:

$$\frac{\partial L_{T-1}^A}{\partial \beta_1} = \frac{-x_{T-1}^A \exp\left(\alpha_1 - \beta_1 x_{T-1}^A\right)}{\left[1 + \exp\left(\alpha_1 - \beta_1 x_{T-1}^A\right)\right]^2} \cdot (a-b) < 0 \tag{15}$$

and

$$\frac{\partial L_{T-1}^B}{\partial \beta_2} = 0 \tag{16}$$

In essence, equation (15) means that if there is a probability for a regime switch but none ocurrs, this affects the likelihood value negatively irrespectively of the underlying transition probability process.

The derivative of L^B_{T-1} has been calculated using the fact that $x^B_{T-1} = 0$. Looking at time T derivatives instead we obtain:

$$\frac{\partial L^A_T}{\partial \beta_1} = \frac{-x^A_T \exp\left(\alpha_1 - \beta_1 x^A_T\right)}{\left[1 + \exp\left(\alpha_1 - \beta_1 x^A_T\right)\right]^2} \cdot (b - a) > 0 \tag{17}$$

and

$$\frac{\partial L^B_T}{\partial \beta_2} = \frac{-x^B_T \exp\left(\alpha_2 - \beta_2 x^B_T\right)}{\left[1 + \exp\left(\alpha_2 - \beta_2 x^B_T\right)\right]^2} \cdot (b - a) > 0 \tag{18}$$

Looking at the global likelihood, using equations (15)-(18), we obtain:

$$\frac{\partial L^A}{\partial \beta_1} = \frac{\partial L^A_{T-1}}{\partial \beta_1} + \frac{\partial L^A_T}{\partial \beta_1}$$

and

$$\frac{\partial L^B}{\partial \beta_2} = \frac{\partial L^B_{T-1}}{\partial \beta_2} + \frac{\partial L^B_T}{\partial \beta_2} > 0$$

Note that for when using the variable x^B the maximum likelihood will be found as parameter β_2 goes towards infinity, which is not the case for the x^A.

Now we proceed to the situation where the transition equation consists of both variables:

$$\pi_t = f(\alpha - \beta_1 x^A_t - \beta_2 x^B_t) \equiv f(\Theta)$$

What we will show is that the effect stemming from x^B_T will act very much more strongly than the effect of x^A_T. First, we consider the derivatives of the likelihood function:

$$\frac{\partial L}{\partial \beta_1} = \frac{\partial L_{T-1}}{\partial \beta_1} + \frac{\partial L_T}{\partial \beta_1} \tag{19}$$

and

$$\frac{\partial L}{\partial \beta_2} = \frac{\partial L_{T-1}}{\partial \beta_2} + \frac{\partial L_T}{\partial \beta_2} = 0 + \frac{\partial L_T}{\partial \beta_2} > 0$$

Again, the effect of this will be a solution in which $\beta_2 \to \infty$ and a more ambiguous solution for β_1.

$$\lim_{\beta_2 \to \infty} \frac{\partial L}{\partial \beta_1} = \lim_{\beta_2 \to \infty} \frac{\partial L_{T-1}}{\partial \beta_1} + \lim_{\beta_2 \to \infty} \frac{\partial L_T}{\partial \beta_1} \tag{20}$$

where

$$\lim_{\beta_2 \to \infty} \frac{\partial L_{T-1}}{\partial \beta_1} = \frac{-x^A_{T-1} \exp(\Theta)}{\left[1 + \exp(\Theta)\right]^2} \cdot (a - b) = -x^A_{T-1} \underbrace{\frac{1}{[1 + \exp(\Theta)]}}_{\to 1} \underbrace{\frac{\exp(\Theta)}{[1 + \exp(\Theta)]}}_{\to 0} \cdot (a - b) = 0$$

and

$$\lim_{\beta_2 \to \infty} \frac{\partial L_T}{\partial \beta_1} = \frac{-x^A_T \exp(\Theta)}{\left[1 + \exp(\Theta)\right]^2} \cdot (b - a) = -x^A_T \underbrace{\frac{1}{[1 + \exp(\Theta)]}}_{\to 1} \underbrace{\frac{\exp(\Theta)}{[1 + \exp(\Theta)]}}_{\to 0} \cdot (b - a) = 0$$

Using these results, we see that (20) will converge towards zero. When maximizing the likelihood function, we will see the impact of the variable x_t^A diminish as the more and more weight is put on x_t^B through the parameter β_2. It follows that the standard error of β_1 will become very large.

To summarize, the above discussion has assumed that we have a variable that is a perfect short-run predictor of future state switches. From this it has been shown that any other variable, although better resembling the true data generating process of transition probabilities, will be crowded out and deemed non-signficant in a joint estimation.

Simulation Evidence

In this section, we will consider a case where the short run selection bias prevents a more useful analysis of early warning indicators. Consider the following simple 2 state model:

$$\Delta y_t = \mu_{S_t} + \epsilon_{S,t} \tag{21}$$

where $\overline{\mu} < 0$, $\mu_1 > 0 > \mu_2$, $\epsilon_t \sim N(0, \sigma^2_{S_t})$, $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 1$. The transition matrix is

$$\mathbf{P}_t = \begin{bmatrix} f(\alpha_1) & 1 - f(\alpha_1) \\ 1 - f(\alpha_2 + \beta y_{t-1}) & f(\alpha_2 + \beta y_{t-1}) \end{bmatrix}$$

In words, this process would have a negative drift were it not for the lower probability boundary that reverts the process back into the positive mean state. In figure 1, a simulated series with the parameters $\mu_1 = 0.5$, $\mu_2 = -0.3$, $\alpha = 5$, $\beta = 0.1$ and $\alpha_1 = 3.4761 \quad \Leftrightarrow \quad p_{11} = 0.97$ is plotted.

One note is appropriate here: we see that a pure threshold model would have a hard time to capture the dynamics of the series, since regime switches occur at quite different magnitudes of y_t. The regime switching model can allow for this; although the probability for the first regime switch is quite low, it can be incorporated into the model.

To proceed, we have constructed a binary indicator variable x_t that takes on the value 1 the time period prior to a switch to state 1 and is 0 otherwise. We have then estimated three different setups of transition equations of the model:
(i) $f(\alpha_2 + \beta_1 y_{t-1})$
(ii) $f(\alpha_2 + \beta_1 y_{t-1} + \beta_2 x_{t-1})$
(iii) $f(\alpha_2 + \beta_2 x_{t-1})$

It is apparent from table that the predicted effect of including the binary indicator variable exists in the simulated data, even if the form of the long-run variable is different than from the theoretical set-up. The addition, the difference in the likelihood value between case (ii) and (iii) is virtually zero and the standard error of the parameter β_1 is very high.

The correlation coefficient has been computed as: $\rho = Corr\left[\Delta f(\Theta); \Delta f(\hat{\Theta})\right]$

Figure 1: Simulated data. Solid line indicates the y_t process (right scale); bars indicate the positive mean reversion state and the dotted line indicates the true transition probability process (left scale).

where Δ denotes the first difference operator and $\hat{}$ denotes empirical estimates. It indicates that the model without the short run indicator has a high correlation with the true process, whereas the other models - which should be preferred in terms of statistical significance - has a much lower correlation. The importance of this effect can be seen in figure 2 where the true and projected transition probabilities have been plotted. As can be expected, the model including x_t signals a 0 probability to stay in state 1 one period prior to the actual shift but signals a stay probability of 1 otherwise. Model (i) shows a transition probability pattern similar to the DGP, but is somewhat more extreme in its estimates.

These results indicate that one should simply exclude short-run variables from the regression if one wants to have long-run inference. What is then needed is a sense of how large the short-run bias of the parameter on the long-run variable is. It seems probable that the bias will decrease as the sample size increases.

We should also consider a case where we want to see if another exogenous stochastic variable has an effect on transition probabilities. Suppose we have another economic aggregate z_t such that

$$z_t^i = z_{t-1}^i + \epsilon_t^{z^i}$$

where $\epsilon_t^{z^i} \sim N(0,1)$. The null hypothesis of our test would be $\beta_3 = 0$ when we have replaced the lower row of $\mathbf{P}_t$ with $\left[1 - f(\alpha_2 + \beta_3 z_t^1) \quad f(\alpha_2 + \beta_3 z_t^1)\right]$. This is a regression of a non-stationary variable on a bounded stationary series, since transition probabilities are bounded between $[0,1]$. We conduct a small Monte

Model	(i)		(ii)		(iii)		CTP	
Param.	Value	p	Value	p	Value	p	Value	p
μ_1	0.5967	0.00	0.6040	0.00	0.6040	0.00	0.6009	0.00
μ_2	-0.2255	0.00	-0.2273	0.00	-0.2273	0.00	-0.2260	0.00
σ_1	1.1414	0.00	1.1373	0.00	1.1373	0.00	1.1418	0.00
σ_2	0.4838	0.00	0.4846	0.00	0.4846	0.00	0.4833	0.00
α_1	3.3733	0.00	3.4369	0.00	3.4369	0.00	3.2719	0.00
α_2	21.3593	0.09	25	n.a.	14.6708	0.00	4.3380	0.00
β_1	0.7064	0.13	0.4221	0.94				
β_2			-25	n.a.	-25	n.a.		
LogL	-236.5861		-232.0960		-232.0961		-240.7682	
ρ	0.6311		0.2335		0.2335		n.a.	

Table 1: Parameter values when models (i), (ii) and (iii) are estimated on the simulated data. The optimization procedure has been constrained to not allow parameters in the transition equation to exceed 25 in absolute value. Correlation denotes the correlation coefficient between the simulated TVP series and the empirically projected series.

	ϵ^z		z_t^1		z_t^1, z_t^2	
Mean Likelihood Ratio	1.2202	(1.80)	1.9853	(2.51)	3.9748	(3.36)
LR, 5th percentile	0.0803		0.1663		0.2267	
LR, 10th percentile	0.1345		0.2545		0.3380	
LR, 50th percentile	0.5341		0.6533	0.7440		
ρ^{Δ}	0.0001	(0.06)	0.0001	(0.07)	0.0001	(0.08)

Table 2: Effects of white noise/random walks as explanatory variables in the transition equation. Standard errors in parentheses. 1000 series with 250 observations each have been generated according to the parameterization in the text.

Carlo experiment to this end. The results are presented as case A in table 2.

We note that a larger share of the random walks turn out to be significant than would be expected comparing to a χ^2 distribution with 1 degree of freedom. The corresponding random walks do not exhibit a correlation pattern different from the rest of sample.[4] We draw two conclusions from this: regressing a non-stationary variable in the transition equation does not result in spurious regression results as serious as in the standard case in the literature. But the likelihood ratio statistics are still significant too often. We also note that the LR statistic is somewhat over-sized even for the case where we regress the only the noise term in the transition equation. Moreover, this over-rejection of the null is highly dependent upon the sample size, as illustrated in table 3. Even for a relatively large sample, 500 observations, the empirical size of the likelihood ratio statistic is approximately twice the

[4]Results available upon request.

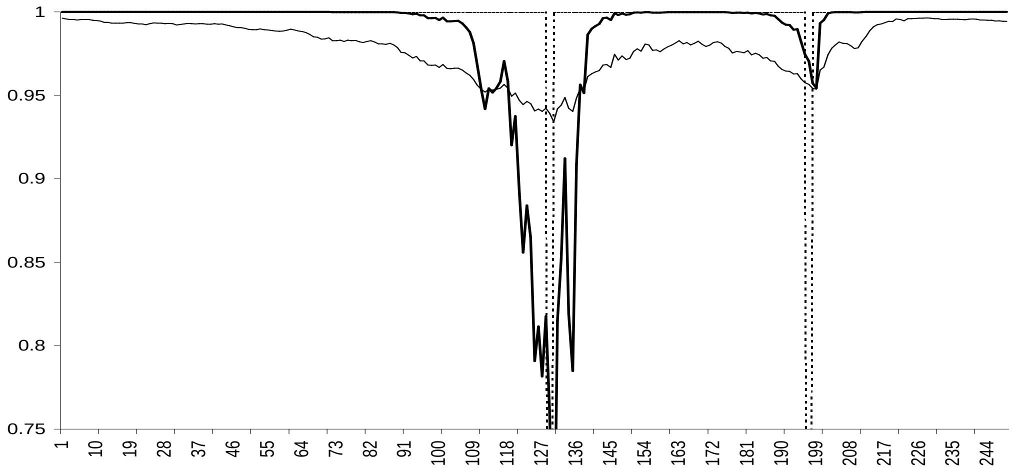


Figure 2: Transition probabilities. Thin line marks the DGP transition probabilities, thick black line marks the projection from model (i) and dotted line marks the projection from model (iii). The projected transition probabilities from model (ii) are virtually identical to those of model (iii).

	Sample size			
	100	250	500	1000
Average p-value	0.2688	0.3470	0.3958	0.4855
5th percentile	0.2509	0.1784	0.1166	0.0565
10th percentile	0.3675	0.2686	0.1873	0.1007

Table 3: Spuriousity and sample size, case z_t^1.

theoretical size.[5]

In order to check the effects of situations where we have two or more possibly non-stationary series in the transition equation, we conduct yet another Monte Carlo experiment where two random walks are included in the transition equation so that the lower row of $\mathbf{P}_t$ is $\left[1 - f(\alpha_2 + \beta_3 z_t^1 + \beta_4 z_t^2) \quad f(\alpha_2 + \beta_3 z_t^1 + \beta_4 z_t^2)\right]$ The results in table 2, case z_t^1, z_t^2 (with the LR statistic distributed according to χ_2^2) indicate that the spuriousity increases as more and more non-stationary variables are included. Using two random walks in the transition equation, we are going to obtain jointly significant likelihood ratios in 23/34% of the cases when applying a 5/10%

[5]Not only the sample size itself but also the stability, i.e. the ""stay"" probabilities, should affect how often we reject the true null. The effective sample size for the transition equation depends on the total number of observed regime switches. The TVP equation effective sample size effect on the over-rejection rate remain to be investigated.

significance level. Again, the effective sample size - that is, the number of observed regime switches - is likely to affect the degree of spurious results. Intuitively, in a sample with few switches, it does not seem very hard to construct a linear combination of two or more non-stationary variables that happens to be high/low relative to the sample mean in periods when switches occur. As is depicted in the bottom row of table 2, these linear combination are going to have virtually zero correlation with the true data generating process.

For empirical purposes, the spurious effects induce a difficult problem: even though the process generating the transition probabilities is globally stationary, we will not be able to regress other possibly globally stationary variables in the transition equation of the model. Inference obtained from such estimations, as measured by the likelihood ratio statistic, is subject to very large positive size distortions.

It seems useful to relate this discussion of spurious regressions to the standard case in the literature. There, one non-stationary variable is regressed on another and the result is a spurious inference on their relation where none exists. Since both variables are directly observable, regressing differences of the variables instead is straightforward. In the transition probability case, the left-hand side variable is non-observable, so we cannot take differences of it directly. Instead, we have to proxy it by the inference we have on states in the empirical data. So, in order to conduct a differencing operation in this set-up, one would regress the (differenced) exogenous variables on the changes in regimes. Assume that we have a globally stationary variable that decreases at a rate δ K periods prior to a regime switch at time T in the dependent variable, and is a noise variable around a given mean μ otherwise. The disturbance term is always $\epsilon \sim (0,1)$. When differenced, the regression would amount to relating δ to the binary variable indicating regime shifts. Hence, the differenced model can only measure a jump in the probability which is equal for all $t = (T-K), (T-K)+1, ..., (T-1)$, and we loose all the information about that the longer we have observed a δ in the series, the more likely we are to switch. Moreover, if the fluctuations in ϵ are large relative to δ, it will be very hard to discern what observations are truly collected from the period $(T-K), (T-K)+1, ..., (T-1)$. Consequently, differencing all variables may lead to a large loss of information which may not be tolerable if the effective sample size is small.

Proposed Remedy

The previous investigation has shown that there is a bias towards selecting variables that induce changes in the transition equation very close and abruptly to a regime switch. We have argued that in empirical work, one may have the opposite objective. Also, including non-stationary variables in the transition equation results in an inference problem similar to that of standard spurious regressions. In this section, we will suggest a simple solution to remedy these problem. To do this, we require the researcher's prior about how important the long-run effects are in relation to the short-run ones.

We begin with the cases considered initially, case (A) and (B). As we saw both in the theoretical and empirical setting, the problem with case (B), was that maximization of the likelihood function led to a corner solution for β_2. A natural way to avoid this is to introduce a penalty in the likelihood function so that there exist a finite solution for (18)= 0. The problem is to decide upon the magnitude and functional form of the penalty. The simple approach suggested here uses a prior about how the projected transition probability series should look. To begin with, we assume that fundamental economic variables evolve slowly over time. Then, if these fundamentals govern the probability to switch economic states, we would expect the series of probabilities to correspondingly move slowly. In a graphical depiction of the probabilities, a smooth series implies slow movements in the underlying variable one measures. For example, the Hodrick-Prescott filter decomposes a time-series into a slowly moving, smooth trend component and a faster moving, non-smooth cyclical component.

Hence, we proxy the prior of slowly moving fundamentals with a term in the likelihood function that penalizes non-smooth behavior. The penalized log likelihood takes on the following form:

$$L(\theta) = \sum_{t=1}^{T} \left\{ \log \mathbf{1'} \left(\xi_{\mathbf{t|t-1}} \odot \eta_{\mathbf{t}} \right) - \mathrm{e}^{\gamma} \mathbf{1'} \left[\mathrm{diag}\left(\mathbf{P}_t \right) - \mathrm{diag}\left(\mathbf{P}_{t-1} \right) \right]^2 \right\} \tag{22}$$

where diag$(\cdot)$ denotes the principal diagonal operator and γ is the weight of the prior given by the econometrician. The first drawback of this approach is obvious: traditional likelihood ratio testing will not possible using the expression in (22) since the penalty term will make it be lower than the the baseline model's log likelihood. However, since this change of the likelihood is always negative, a likelihood ratio statistic based on it will be more conservative in the sense that it rejects too many variables. If this is acceptable, which it often may be in empirical work, one may just use the penalized likelihood instead of the standard one. One way to reduce this drawback is to use the estimation results obtained from maximizing (22) and evaluate the non-penalized likelihood function with the corresponding parameter vector.[6]

Figure 3 shows that applying the penalty results in a more smooth transition probability function. It also shows the trade-off between smoothness and magnitude of the predicted probabilities. Once the functions becomes more smooth, it is less capable of inducing a large transition probability. For penalties of 15 and more, the stay probability is always more than 98%. In table 4, the results show - up to a certain level for the penalty - that the penalized models increase the correlation between the true DGP and the projected stay probabilities. The coefficient estimates are reduced and come closer to their true value as well, and the modified likelihood values decrease as the penalty increases.

From figure 4, where model (ii) has been estimated and used to project stay

[6]The latter case will be denoted with an additional *.

Figure 3: Stay probabilities with different priors ($\gamma = \text{G}$) ; model (ii).

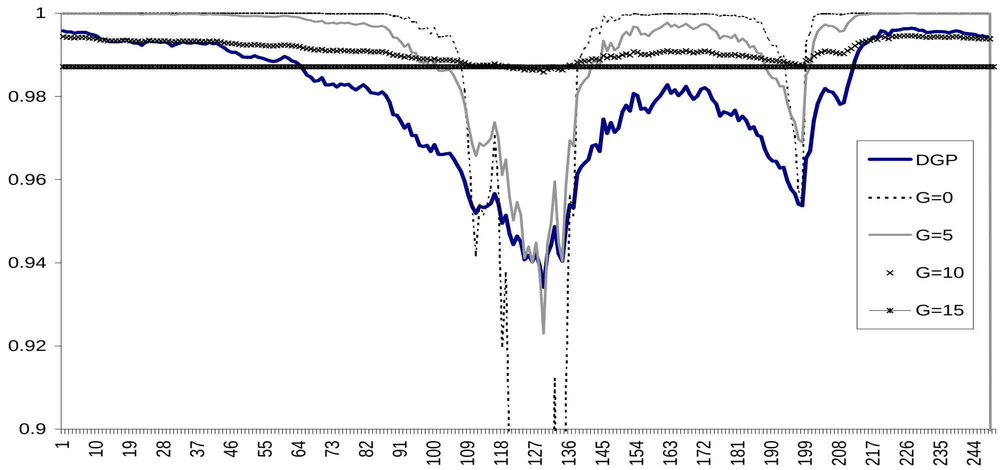
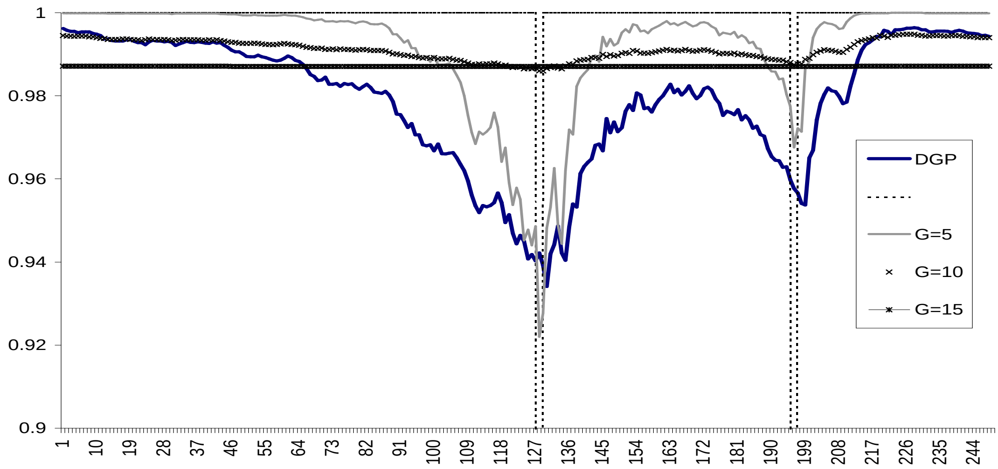


Figure 4: Stay probabilities with different priors ($\gamma = \text{G}$) ; model (ii).



	α_2	β_1	ρ^{Δ}	Log-L*
$\gamma = 0$	19.8316	0.6570	0.7143	-235.7250
$\gamma = 1$	16.7850	0.5375	0.7579	-235.7753
$\gamma = 3$	13.0636	0.3874	0.8323	-236.0339
$\gamma = 5$	9.9710	0.2556	0.9184	-236.5826
$\gamma = 7$	7.8733	0.1574	0.9826	-237.3949
$\gamma = 10$	5.2109	0.0330	0.9739	-239.0265
$\gamma = 15$	4.3393	0.0002	0.9403	-239.7009

Table 4: Diagnostics for different penalty priors; model (i), simulated data set as in figure 1.

	α_2	β_1	β_2	ρ^{Δ}	Log-L*
$\gamma = 0$	23.8403	0.3839	-25.0000	0.1811	-231.0568
$\gamma = 1$	25.0000	0.0000	-24.1682	0.1811	-233.4431
$\gamma = 3$	12.9989	0.3768	-0.6987	0.8039	-235.6277
$\gamma = 5$	10.1913	0.2608	-0.3066	0.9064	-236.3956
$\gamma = 7$	8.0139	0.1612	9 -0.1501	0.9784	-237.2978
$\gamma = 10$	5.2474	0.0339	-0.0320	0.9751	-238.9969
$\gamma = 15$	4.3394	0.0002	-0.0002	0.9410	-239.7007

Table 5: Diagnostics for different penalty priors; model (ii), simulated data set as in figure 1.

probabilities, the importance of the penalty becomes more protruding. The unconstrained model has the binary looking transition series, whereas the prior constrained series exhibit patterns (and by looking in table 5 correlations) closely linked to the true DGP.

In table 6, the effect on the likelihood ratio statistic of using different sizes of penalties on different setups of the model is explored. To begin with, when estimating the model according to the true data generating process, we note that for $0 < \gamma \leq 2$ the LR statistics are just slightly lower than the corresponding statistics for the $\gamma = 0$ case. For higher settings of the penalty, the model finds a distinctly lower number of significant LR statistics than in the non-penalized case. This indicates that one should use caution when setting γ, since too large a value may very well lead to too few rejections of the null hypothesis of no relation between the TVP variable and the true process.[7] Contrary to this, we find the highest correlation between the changes in the true TVP variable and the projected one when higher values of γ are applied. The peak is here obtained when applying

[7]A reduction of this too low rejection rate can be obtained by using the likelihood ratio statistic computed using the penalty coefficient estimates in the non-penalized likelihood function, which yields a higher/more significant statistic. The magnitude of this reduction remains to be studied.

Case	γ 0	1	2	5	7	10	15
y_t (DGP)							
Mean p	0.0772	0.0834	0.0850	0.1173	0.2102	0.5876	0.9629
LR 5%	0.6947	0.6732	0.6499	0.3950	0.0224	0.0000	0.0000
LR 10%	0.8319	0.8240	0.8235	0.6835	0.3165	0.0000	0.0000
LR 50%	0.9720	0.9693	0.9692	0.9636	0.9272	0.3239	0.0000
Mean ρ^{Δ}	0.7807	0.7983	0.8182	0.9033	0.9375	0.8316	0.7109
ϵ_t							
Mean LR	1.3424	0.4587	0.3166	0.0165	0.0020	0.0001	0.0000
LR 5%	0.0986	0.0058	0.0087	0.0000	0.0000	0.0000	0.0000
LR 10%	0.1623	0.0261	0.0203	0.0000	0.0000	0.0000	0.0000
LR 50%	0.5681	0.2928	0.1710	0.0000	0.0000	0.0000	0.0000
Mean ρ^{Δ}	0.0029	0.0023	0.0018	0.0073	-0.0022	0.0025	-0.0010
z_t^1							
Mean LR	1.7328	1.2507	1.0699	0.7004	0.2643	0.0364	0.0002
LR 5%	0.1164	0.0905	0.0474	0.0000	0.0000	0.0000	0.0000
LR 10%	0.2069	0.1509	0.1121	0.0216	0.0000	0.0000	0.0000
LR 50%	0.6293	0.5948	0.5776	0.5043	0.1897	0.0129	0.0000
Mean ρ^{Δ}	0.0038	0.0011	0.0122	0.0092	0.0198	0.0129	0.0107
z_t^1, z_t^2							
Mean LR	4.1401	2.9455	2.5298	1.3919	0.5961	0.0717	0.0005
LR 5%	0.2586	0.0862	0.0474	0.0000	0.0000	0.0000	0.0000
LR 10%	0.3448	0.2328	0.1466	0.0043	0.0000	0.0000	0.0000
LR 50%	0.7241	0.6810	0.6595	0.4655	0.1121	0.0000	0.0000
Mean ρ^{Δ}	-0.0049	0.0046	0.0158	0.0208	0.0264	0.0146	0.0142

Table 6: Likelihood ratios and quantiles when different sizes of penalties are applied. 350 simulations have been used.

$\gamma = 7$. Hence, one has to consider the trade-off between the difficulties using the LR statistic and obtaining the highest possible correlation between projected and true transition probabilities.

In general, the "false" variables as in the 2 middle panels of table 6 are significant about twice as often as they should be at the 5% level, with a somewhat higher over-significance for the non-stationary variable z_t^1. For the noise variable ϵ_t, the effect of the penalty is that we reject the null of no correlation in too many cases as soon as the penalty is applied. For the non-stationary variable, the reduction of the over-significance is gradual, and applying $\gamma = 2$ yields likelihood ratio statistics close to the nominal sizes.

For the case on multiple non-stationary variables as regressors in the TVP equation, we note a large over-significance of likelihood ratio statistics in the non-penalized model. As many as 25% of the statistics are larger than the 5% significance value of a χ^2 distribution with 2 degrees of freedom. Again, setting $\gamma = 2$ yields significance results much closer to the nominal level.

To summarize, the results indicate that applying a penalty is essential to avoid getting too many significant variables in the TVP equation. What the size of the penalty should be is less clear. Setting γ to a high value, one risks obtaining too few signficant variables, but can also get projected transition probabilities closer to the true process and thus have more exact measurement of the impact of one variable on the probability to switch states. Moreover, the effects of γ as depicted in table 6 are very likely to be data dependent. Hence, for a different parameter vector or sample size, the optimal γ could be quite different from what can be inferred from the table. A useful approach for empirical purposes could be to observe the projected transition probabilities and compare them to what seems reasonable for the data. For example, if one reaches a final specification for a set γ that yields transition probabilities that are very certain prior to regime switches that a regime switch will occur, one may suspect that over-fitting has occurred and that γ should be increased. These issues are beyond the scope of this paper.

Empirical Application

The original Hamilton (1989) paper established the usefulness of the Markov regime switching model to replicate business cycles. We apply the same methodology on quarterly real GDP data from 1964:1 to 2002:4 for a total of 150 observations.[8] A similar data-set is studied in Coe (2002). The baseline model is

$$\Delta y_t = \mu_{S_t} + \epsilon_{R_t} \tag{23}$$

where Δy_t is the logarithmic change in real GDP per capita and S_t,R_t are unobserved state variables. The error term ϵ_{R_t} is distributed according to $N(0, \sigma^2_{R_t})$. The

[8]Nominal GDP and inflation as measured by the consumer price index are obtained through the IMF's *International Financial Statistics* database.

	LR	H_0^N	H_0^M
$N = 1$; $M = 2$	33.9970	0.0040	0.5737
$N = 2$; $M = 4^*$	15.8450	0.0080	0.9920

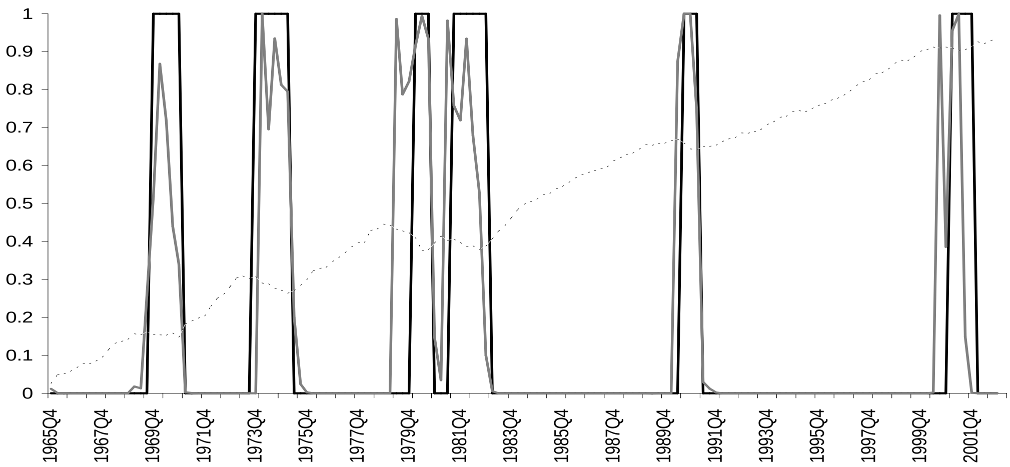
Table 7: Test for Markov switching dynamics. 250 Monte Carlo runs.

first state variable, $S_t \in [S^{Expansion}, S^{Contraction}]$; $S^{Contraction} \prec S^{Expansion}$, governs the intercept and the second one R_t governs volatility. Initially, the transition matrix $\mathbf{P}_t$ is kept constant as in equation (2). The transition matrix $\mathbf{Q}_t$ associated with the R process is assumed to be constant trough out the remainder of the paper.

In order to test for the existence of Markov switching dynamics in the data, we apply the Monte Carlo testing procedure discussed in Cheung and Erlandsson (2003), which is an extension of the Rydén, Teräsvirta and Åsbrink (1998) procedure.
The results in table 7 indicate strong evidence of Markov switching dynamics in the data. Diagnostic testing rejects the hypothesis that variances are equal across states for the standard 2 state setting. Consequently, we also investigate the possibility of variance following a regime switching process of its own so that $S_t \neq R_t$ for some t, thus allowing for a total of 4 states.[9] The results of this test are also clear. We reject the 2 state MS model, but are do not reject the 4^* state counterpart. One could also suspect even more states in the data, but limited computational capacity restricts us from investigating these suspicions. An alternative is to look at the diagnostics of the model in the proposed specification. With 4^* states, neither significant residual autocorrelation as measured by the Ljung-Box Q statistic (p-value 0.074, nor ARCH effects as measured by Engle's LM test (p-value 0.173), is present in the standardized residuals. For the 2 state model the corresponding p-values are 0.001 and 0.311 respectively.

Another diagnostic measure to validate the model is how well it replicates business cycles as measured elsewhere. The by all standards most common benchmark in the literature is the National Bureau of Economic Research (NBER) business cycle dates., which we will denote as $\hat{S}_t$. The smoothed probabilities, computed according to the algorithm of Kim (1994) of the the contractionary state of real-GDP in our model is plotted against the NBER dates in figure 5. As can be seen, the model replicates the dates quite well. Only using 2 state produces a graph that does not resemble the NBER dates. The reason for this seems to be a shift from the low volatility state to the high volatility states in 1984:1-1984:2. The high volatility state seems to be absorbant within the sample, meaning that the volatility process does never return to the low state after 1984. The simpler model produces probabilities that are a mix of the level and volatility states in the more general model.

[9] This parameterization is constricted however so that the 4 state transition matrix is the Kronecker product of the two separate processes' respective transition matrices. Since the resulting transition matrix is constrained, we will denote these 4 states with a subscript $* \rightarrow 4^*$.



3

Figure 5: NBER dates (black) and smoothed probabilities of the contractionary state (grey) using the constant transition probability model. U.S. real GDP is plotted with dots (normalized to 0 in 1965Q4).

To proceed with investigating factors that predict recession, we convert the baseline model to a restricted TVP parameterization. The transition matrix for the S_t process is

$$\mathbf{P}_t = \begin{bmatrix} f(\alpha_1 + \beta \mathbf{X}_t) & 1 - f(\alpha_1 + \beta \mathbf{X}_{t-1}) \\ 1 - f(\alpha_2) & f(\alpha_2) \end{bmatrix}$$

where $\mathbf{X}_{t-1}$ is a set of possible leading indicator candidates. This structure means that the transition probability from the expansionary phase of the economy to the contractionary is time varying, whereas the reverse is constant.

The next step is to specify $\mathbf{X}$. In table 11, we present 31 variables suggested by *Economagic* to be related to the business cycle. Monthly data has been transformed to quarterly by taking the end of quarter monthly value. Each variable in first differences, an 8 quarter moving average, and in levels has been tested individually through likelihood ratio tests,[10] and with differing penalty terms. 12 of the candidates have median p-values below 20%.[11] Of these, 10 are 8 quarter moving averages and 2 are in levels. This should be viewed in the light that 18 variables in first differences are significant at the 10% level when estimating the model without

[10]To obtain better convergence properties all variables have been normalized. The level variable has been calculated as the cumulative sum of normalized first differences so that any time trends have been removed. The moving average has been calculated the same way, but as the average of the 8 last observations.

[11]The median p-values are calculated as the median of the p values for one variable, one transformation and 6 different penalty settings (ranging from 0 to 5).

a penalty term, but all of them turn insignificant once the penalty term is applied.

Given the previous simulation results, setting $\gamma = 2$ seems to strike a good balance between the long-run capability, and the statistical properties. Using this prior, we have conducted a testing down procedure of $\mathbf{X}$. The least significant variable has been removed until the reduction in the likelihood ratio statistic is below 5% level.[12] The final specification that are reached are presented in table 8, column 2. The results from a number of benchmark models are also presented. The first column, $\gamma = 0$, refers to the results when estimating the specification obtained from the penalized setting but setting $\gamma = 0$. The CTP model is the constant transition probability model. In the fourth column, denoted $\gamma = 0(*)$, a final specification has been obtained using the same testing down procedure as above, but with $\gamma = 0$.

The results indicate that all TVP models estimate the contractionary phase as more severe than in the CTP case. When using the penalized version, two variables are found significant as leading indicators of a recession: the seasonally adjusted production price index of finished goods and the number of unemployed civilians.[13] A rise in the production prices decreases the probability to stay in the expansionary state, as does an increase in unemployment. One interpretation of these indicators is that the probability of recessions is strongly linked to shortages in both goods and labor markets.

Using the traditional approach of testing down the model's TVP variables, a very different conclusion is reached. First, many more variables are deemed significant, which also was the prediction of the simulation results in the previous section. Second, for the one variable that the specification have in common, the signs are opposite. The difference is likelihood values is large, even when applying the non-penalized value on the penalized specification.

Using the above specifications of the model and its benchmarks, we register dates t at which the predicted probability to proceed to the contraction state in the time interval $t+1 : t+k$ exceeds a certain threshold level in percent, denoted ω. The forecasted probability for $t+1$ is calculated as:

$$\Pr(S_{t+1} = 1|\Omega_t) = p_{11,t}\Pr(S_t = 1|\Omega_t) + p_{21,t}\Pr(S_t = 2|\Omega_t) \qquad (24)$$

For the $k > 1$ step ahead forecast, we focus on the probability that we will at least one crisis period within the time interval $t+1, t+2, ..., t+k-1, t+k$. This equals 1 minus the probability that we see no crises within the time interval:

$$\Pr(\min(S_{t+1,\ldots,t+k}) = 1|\Omega_t) = 1 - \Pr(S_{t+1,\ldots,t+k} = 2|\Omega_t) = \qquad (25)$$

$$= 1 - \left[p_{12,t}p_{22,t}^{k-1}\Pr(S_t = 1|\Omega_t) + p_{22,t}^{k}\Pr(S_t = 2|\Omega_t)\right]$$

[12]In this non-linear setting, the likelihood ratio statistic has been shown to be more robust than statistics based on the variance-covariance matrix, such as the Wald statistic.

[13]The latter variable reflects the number of unemployed civilians compared to the trend, and is consequently similar to an ordinary unemployment rate figure.

	$\gamma=0$		$\gamma=2$		CTP		$\gamma=0(*)$	
	Value	p	Value	p	Value	p	Value	p
$\mu^{Expansion}$	0.7308	0.00	0.7394	0.00	0.7459	0.00	0.7396	0.00
$\mu^{Contraction}$	-0.7681	0.00	-0.7118	0.07	-0.6064	0.14	-0.9062	0.00
σ_1^2	1.0213	0.00	1.0186	0.00	1.0397	0.00	0.9701	0.00
σ_2^2	0.4866	0.00	0.4818	0.00	0.4765	0.00	0.4759	0.00
α_1^S	100	0.00	3.8072	0.00	2.9896	0.00	100	n.a.
β_{13}^S							-1.6376	0.00
$\beta_3^S 17$							0.6659	0.00
β_{20}^S							-0.4281	0.00
β_{21}^S							-0.5264	0.00
β_{22}^S	-0.743	0.00	-0.0129	0.00				
β_{23}^S							1.4974	0.00
β_{25}^S	0.9836	0.01	0.0157	0.00			-0.2270	0.00
β_{26}^S							-0.5538	0.00
α_2^S	0.2332	0.80	0.7867	0.26	0.9697	0.13	-0.6939	0.68
α_1^R	4.2969	0.00	4.2967	0.00	4.2960	0.00	4.2981	0.00
α_2^R	100	n.a.	100	n.a.	100	0	100	n.a.
LogL	-167.84		-175.77		-181.06		-161.52	
R^2	0.5775		0.5825		0.5494		0.5973	
LR	26.4300	0.00	10.5670	0.01	0	n.a.	39.0743	0.00
LR*	26.4300	0.00	12.9619	0.00	0	n.a.	39.0743	0.00

Table 8: Estimation results. The optimization procedure has been bounded so that $-100 \leq \alpha, \beta \leq 100$. Standard errors of the stay probability parameter in the low volatility state, α_2^R, are not computable since the state is absorbant within the data range. Subscript indices on β refers to the index number of exogenous variables in table 11.

Four cases of signals from the model vis-a-vis the actual development can then be constructed:[14]

1. The model signals a contraction, and a contraction occurs (CE):
   $\Pr(\min(S_{t+1,\ldots,t+k}) = S^{Contraction}|\Omega_t) > \omega$ and $\min(S_{t+1:t+k}) = S^{Contraction}$
2. The model signals a contraction, but no contraction occurs (CE):
   $\Pr(\min(S_{t+1,\ldots,t+k}) = S^{Contraction}|\Omega_t) > \omega$ and $\min(S_{t+1:t+k}) = S^{Expansion}$
3. The model signals no contraction, but a contraction occurs (EC):
   $\Pr(\min(S_{t+1,\ldots,t+k}) = S^{Contraction}|\Omega_t) \leq \omega$ and $\min(S_{t+1:t+k}) = S^{Contraction}$
4. The model signals no contraction, and no contraction occurs (EE):
   $\Pr(\min(S_{t+1,\ldots,t+k}) = S^{Contraction}|\Omega_t) \leq \omega$ and $\min(S_{t+1:t+k}) = S^{Expansion}$

Using these definitions, a number of benchmarks of the model's performance can be constructed. We will focus on two; the first being the ratio of correct signals to the total number of signals $(CC + EE)/(CC + CE + EC + EE)$. In

[14]The first letter in each case's acronym stands for the prediction of the model, the second for the actual development. C refers to at least one contraction within the time interval, E (as in expansions) to a time interval with no contraction.

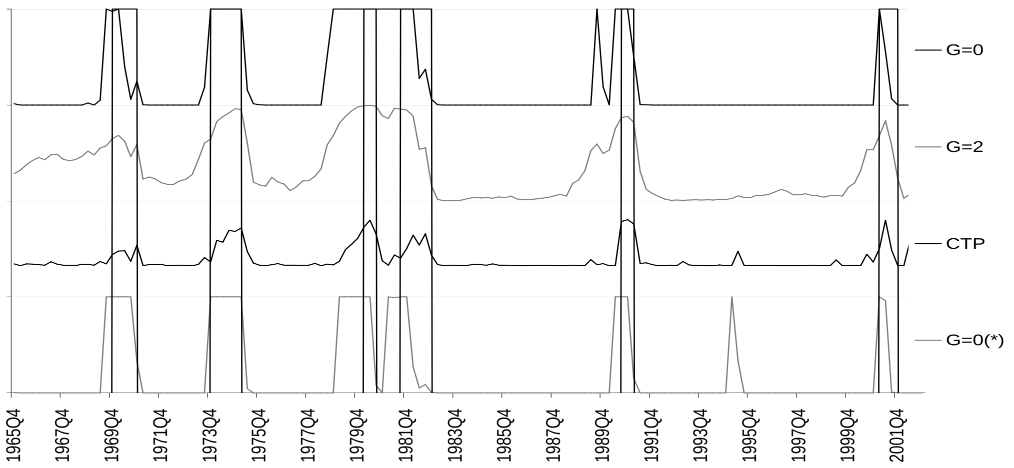


Figure 6: Forecasts of the probability that a contraction will occur within the next 8 quarters.

our setting this benchmark answers the question: "How reliable is the predictions of the model?" The second benchmark, the noise-to-signal ratio, is calculated as $[EC/(EC + EE)]/[CC/(CC + CE)]$, and returns a measure of how strong inference on the true development the model gives. A model that perfectly predicts the future would have a noise-to-signal ratio of 0. The threshold for when a signal is given, ω, may be set to different values than the traditional 50% to reflect different sensitivities to recessions.

Table 9 presents the results of various models. We first note that the only instance where the CTP model outperforms the TVP model in terms of correct forecasts is for the 1 quarter ahead predictions with a threshold of 50%. For prediction horizons of greater than 4 quarters, the CTP model exhibits much worse performance than the TVP models. This is a matter of pure arithmetics: with the 50/25% threshold, the model always predicts the probability to enter the contractionary phase within 12/8 quarters to be greater than the threshold.

The trend for the TVP models based on variable selection found with the penalized likelihood function (i.e. the cases $\gamma = 0$ and $\gamma = 2$) is that as the prediction horizon expands, the better the $\gamma = 2$ model is relative to the $\gamma = 0$ model. This also hold for the relation between the $\gamma = 2$ and $\gamma = 0(*)$ models. Looking at what types of errors the models make, we note that the penalized model for these horizons predicts a larger number of recessions, resulting in fewer case 3 errors.

Figure 6 provides a graphical illustrations on the forecasting performance at the 8 quarter horizon. As has been sought for, the penalized model exhibits a much smoother projection of transition probabilities than the benchmark models. The probabilities also seem to rise earlier prior to recession than for the benchmarks.

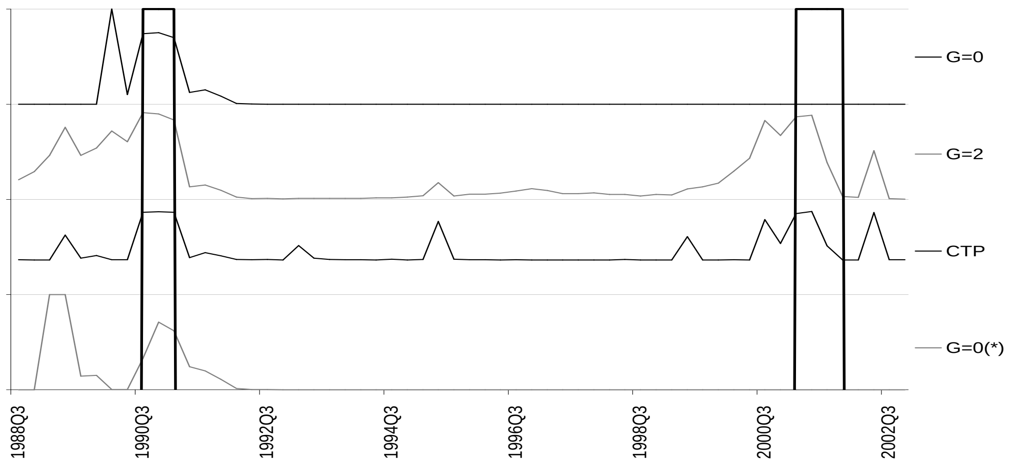


Figure 7: Out-of-sample forecasts of the probability that a contraction will occur within the next 8 quarters.

A more reliable way to evaluate the model's performance is to observe the out-of-sample forecasting properties. If the model reflects a relation that is stable over time, out-of-sample forecasts will resemble the corresponding in-sample forecasts. Otherwise, we have an indication of overfitting. The out-of-sample forecasting perfomance is tabulated in table 10. We have selected to produce forecasts of the two last recessions in the sample, ending the in-sample at 1988:2 and forecasting 1988:3 to 2002:4. Since the latest recession has occured so recently, this means that the performance of the 8 quarter and 12 quarter forecasts cannot be evaluated for that recession.

The penalized model consistently outperforms the benchmark models in this setting at all horizons. When viewing the graphical evidence as in figure 7, it turns out that the penalized model is alone in being able to predict the 2001 recession. It is also more consequent in prediction the 1990 recession, with a gradual increase in probabilities rather than the jagged projection of the other TVP models.

Conclusion

This paper illustrates the inability of the Markov regime switching model to make inference on the probability of states occurring in the medium to long term. For policy purposes, short run predictors of future states may be irrelevant, since many tools such as fiscal policy and interest rate changes only have effects in the medium to long term. Hence, the model has not been a commonly used tool when predicting future states of the economy.

Rather than not being able to make inference in the longer run at all, the model with time varying transition probabilities possesses a bias towards selecting short run variables for predicting future states. This also leads to estimation problems in the maximum likelihood setting, where bounded optimization procedures often has parameter estimates of the TVP variables on the boundaries. Summed together: the estimated model seems to be unable to depict a highly useful dimension of the theoretical model, and it may be hard to obtain estimates at all.

There is also a spuriousity problem in the limited sample setting. We show that this problem may be small when only looking a on variable at a time, but once one tries to specify a model with more possibly non-stationary series, the problem quickly increases.

We propose a simple penalty term, based on the smoothness of the projected time varying probabilities, in the maximum likelihood function which aims to remedy these interconnected problems. Simulation evidence indicates that both estimation and inferential problems are reduced.

In an empirical application, we use a number of suggested leading indicators to predict contractionary states of U.S. real GDP. The standard ML estimates are shown to possess the problems shown in the simulation exercise. A large number of variables show up as significant, and the projected transition probabilities are very non-smooth. Applying the proposes penalized estimator yields a final model specification with fewer variables and a smooth transition probabilities. In the in-sample forecasting excercise, the penalized model performs better for longer (8-12 quarter) horizons. When calculating out-of-sample forecasts, the penalized models exhibits better performance irrespective of the horizon. It is the only model that is able to predict the 2001 recession out-of-sample.

References

Abiad, A. (2003). Early-Warning Systems: A Survey and a Regime-Switching Approach. IMF Working Paper 03/32.

Cheung, Y.-W. and Erlandsson, U. G. (2003). Exchange rates and Markov switching dynamics. In: Regime Switching in Foreign Exchange and Interest Rates, Ph.L. thesis, Ulf G. Erlandsson, Lund University, 2003.

Coe, P. (2002). Power Issues When Testing the Markov Switching Model with the Sup Likelihood Ratio Test Using U.S. Output. *Empirical Economics*, 27:2:395–401.

Diebold, F. X., Lee, J.-H., and Weinbach, G. (1994). Regime switching with time-varying transition probabilities. In Hargreaves, C., editor, *Nonstationary Time Series Analysis and Cointegration*, pages 283–302, Oxford. Oxford University Press.

Filardo, A. J., , and Gordon, S. F. (1998). Business cycle durations. *Journal of Econometrics*, 85:99–123.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57:357–384.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Kim, C.-J. (1994). Dynamic Linear Models with Markov–Switching. *Journal of Econometrics*, 60:1–22.

Rydén, T., Teräsvirta, T., and Åsbrink, S. (1998). Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13:217–244.

Tronzano, M., Psaradakis, Z., and Sola, M. (2003). Target zone credibility and economic fundamentals. *Economic Modelling*, 20:791–807.

Threshold 50%	$\gamma = 0$	$\gamma = 2$	CTP	$\gamma = 0(*)$
$k = 1$				
Correct obs. ratio	0.8919	0.8716	0.9054	0.8581
Noise-to signal ratio	0.0853	0.1491	0.0856	0.1776
# CC/CE/EC/EE	20/9/7/112	13/5/14/116	19/6/8/115	10/4/17/117
$k = 4$				
Correct obs. ratio	0.8621	0.8414	0.8414	0.8483
Noise-to signal ratio	0.1667	0.1912	0.1915	0.1831
# CC/CE/EC/EE	27/2/18/98	24/2/21/98	25/3/20/97	25/2/20/98
$k = 8$				
Correct obs. ratio	0.7376	0.7943	0.6809	0.7092
Noise-to signal ratio	0.3356	0.2770	0.4612	0.3654
# CC/CE/EC/EE	27/2/35/77	37/4/25/75	27/10/35/69	23/2/39/77
$k = 12$				
Correct obs. ratio	0.6131	0.7445	0.5693	0.5839
Noise-to signal ratio	0.5072	0.3943	n.a.	0.5338
# CC/CE/EC/EE	27/2/51/57	48/5/30/54	78/59/0/0	22/2/55/57
Threshold 25%	$\gamma = 0$	$\gamma = 2$	CTP	$\gamma = 0(*)$
$k = 1$				
Correct obs. ratio	0.9054	0.8919	0.8716	0.9122
Noise-to signal ratio	0.0499	0.0853	0.0601	0.0578
# CC/CE/EC/EE	23/10/4/111	20/9/7/112	23/15/4/106	22/8/5/113
$k = 4$				
Correct obs. ratio	0.8690	0.8897	0.7103	0.8414
Noise-to signal ratio	0.1532	0.0562	0.2075	0.1906
# CC/CE/EC/EE	30/4/15/96	41/12/4/88	37/34/8/66	26/4/19/96
$k = 8$				
Correct obs. ratio	0.7589	0.8227	0.4397	0.7021
Noise-to signal ratio	0.3131	0.1938	n.a.	0.3923
# CC/CE/EC/EE	30/2/32/77	50/13/12/66	62/79/0/0	24/4/38/75
$k = 12$				
Correct obs. ratio	0.6423	0.7956	0.5693	0.5912
Noise-to signal ratio	0.4811	0.3031	n.a.	0.5446
# CC/CE/EC/EE	31/2/47/57	62/12/16/47	78/59/0/0	25/3/53/56

Table 9: In-sample prediction results.

Threshold 50%	$\gamma = 0$	$\gamma = 2$	CTP	$\gamma = 0(*)$
$k = 1$				
Correct obs. ratio	0.8772	0.8772	0.8772	0.8421
Noise-to signal ratio	0.1887	0.1224	0.1224	0.4528
# CC/CE/EC/EE	2/2/5/48	4/4/3/46	4/4/3/46	1/3/6/47
$k = 4$				
Correct obs. ratio	0.7963	0.8333	0.7778	0.7222
Noise-to signal ratio	0.2667	0.2029	0.3200	0.9600
# CC/CE/EC/EE	3/1/10/40	6/2/7/39	5/4/8/37	1/3/12/38
$k = 8$				
Correct obs. ratio	0.6800	0.7800	0.7000	0.6800
Noise-to signal ratio	0.4348	0.2744	0.4390	0.4348
# CC/CE/EC/EE	3/1/15/31	8/1/10/31	6/3/12/29	3/1/15/31
$k = 12$				
Correct obs. ratio	0.6522	0.7609	0.5870	0.6522
Noise-to signal ratio	0.4762	0.3041	0.7879	0.4762
# CC/CE/EC/EE	3/1/15/27	8/1/10/27	6/7/12/21	3/1/15/27
Threshold 25%	$\gamma = 0$	$\gamma = 2$	CTP	$\gamma = 0(*)$
$k = 1$				
Correct obs. ratio	0.8772	0.8947	0.8596	0.8596
Noise-to signal ratio	0.1887	0.0750	0.0492	0.2404
# CC/CE/EC/EE	2/2/5/48	5/4/2/46	6/7/1/43	2/3/5/47
$k = 4$				
Correct obs. ratio	0.7963	0.9074	0.7778	0.7407
Noise-to signal ratio	0.2667	0.0636	0.2404	0.5612
# CC/CE/EC/EE	3/1/10/40	11/3/2/38	8/7/5/34	2/3/11/38
$k = 8$				
Correct obs. ratio	0.6800	0.8800	0.3600	0.7000
Noise-to signal ratio	0.4348	0.1496	n.a.	0.3889
# CC/CE/EC/EE	3/1/15/31	13/1/5/31	18/32/0/0	4/1/14/31
$k = 12$				
Correct obs. ratio	0.6522	0.8043	0.3913	0.6739
Noise-to signal ratio	0.4762	0.2514	n.a.	0.4268
# CC/CE/EC/EE	3/1/15/27	10/1/8/27	18/28/0/0	4/1/14/27

Table 10: Out-of-sample prediction results.

Indicator	Reporting frequency	Index number
Total private: Indexes of Aggregate Weekly Hours, SA	M	1
Average Weekly Hours; Private Nonagricultural Establishments; SA	M	2
Total Borrowings at Federal Reserve Banks; Billions of Dollars; NSA	M	3
Change in Business Inventories; SAAR Billions of Dollars	Q	4
Corporate Profits After Tax with IVA and CCAdj; Billions; SAAR	Q	5
Consumer Price Index All Urban Consumers: Total; 1982-84=100; SA	Q	6
Consumer Price Index All Urban Consumers: Less Food and Energy; 1982-84=100, SA	M	7
Employment Ratio; Civilian Employment/Civilian Non. Inst. Pop.; Percent SA	M	8
Gross Savings; Billions of Dollars SAAR	Q	9
Index of Help Wanted Advertising; in Newspapers; 1987=100; SA	M	10
Total Industrial Production Index; 1992=100 SA	M	11
M2 Money Stock; Billions of Dollars; SA	M	12
Bank Prime Loan Rate	M	13
Nonfarm Business Sector: Output Per Hour of All Persons; SA, 1992=100	Q	14
Private Business Sector: Output Per Hour of All Persons; SA, 1992=100	Q	15
Payroll Employment; of Wage and Salary Workers; Thousands; SA	M	16
Personal Consumption Expenditures; Billions of Dollars SAAR	M	17
Personal Income; Billions of Dollars SAAR	M	18
PPI - Capital Equipment; 1982=100 SA	M	19
PPI - Crude Materials for Further Processing; 1982=100 SA	M	20
PPI - Finished Consumer Foods; 1982=100 SA	M	21
PPI - Finished Goods; 1982=100 SA	M	22
PPI - Intermediate Materials; 1982=100 SA	M	23
Personal Saving; Billions of Dollars SAAR	Q	24
Civilian Unemployed for 15 Weeks and Over; Thousands; SA	M	25
Manufacturing Sector: Unit Labor Cost; SA, 1992=100	Q	26
Nonfarm Business Sector: Unit Labor Cost; SA, 1992=100	Q	27
Consumer Sentiment; University of Michigan; 1966Q1=100; NSA	Q	28
Unemployment Level; All Civilian Workers; Thousands; SA	Q	29
Capacity Utilization: Manufacturing (SIC); SA	M	30
Industrial Production Index: Consumer goods; 1997=100; SA	M	31

Table 11: Evaluated predictors of the transition probability to the contraction state.