# An Econometric Analysis of Some Models for Constructed Binary Time Series[*]

Don Harding[†]and Adrian Pagan[‡]

January 3, 2009

## Abstract

Macroeconometric and financial researchers often use *secondary* or *constructed* binary random variables that differ in terms of their statistical properties from the *primary* random variables used in microeconometric studies. One important difference between primary and secondary binary variables is that, while the former are, in many instances, independently distributed (i.d.), the latter are rarely i.d. We show how popular rules for constructing the binary states interact with the stochastic processes for of the variables they are constructed from, so that the binary states need to be treated as Markov processes. Consequently, one needs to recognize this when performing analyses with the binary variables, and it is not valid to adopt a model like static Probit which fails to recognize such dependence. Moreover, these binary variables are often censored, in that they are constructed in such a way as to result in sequences of them possessing the same sign. Such censoring imposes restrictions upon the DGP of the binary states and it creates difficulties if one tries to utilize a dynamic Probit model with them. Given this we describe methods for modeling with these variables that both respects their Markov process nature and which

1

explicitly deals with any censoring constraints. An application is provided that investigates the relation between the business cycle and the yield spread.

Key Words: Business cycle; binary variable, Markov process, Probit model, yield curve

JEL Code C22, C53, E32, E37

# 1    Introduction

Macroeconometric and financial econometric research often feature binary random variables. We will designate such a random variable as $S_t$, and assume that it takes the values of unity and zero. Such binary random variables arise in a number of ways, although they differ in their origin. Because of this it is useful to distinguish between binary random variables that are *primary* and those that are *secondary* or *constructed*. In the first set one would include most of those that arise in micro-econometrics. If a time series is involved there will generally be a panel of data on whether an individual makes a particular decision. In these cases the binary variable is often thought of as deriving from an underlying *continuous latent variable* (as in the Probit model). Also in this set would be cases where a continuous random variable - on which there are realizations - depends upon a latent binary random variable. The clearest example of the latter would be Markov Switching (MS) models - Hamilton (1989). In contrast to those cases, this paper is concerned with secondary binary random variables which are constructed from the realizations of a continuous random variable (or variables) $y_t$. This case does not seem to have been studied much, a notable exception being Kedem (1980). However, as we will try to illustrate, quite a few interesting econometric issues arise when such variables are used in empirical work.

Although there are many examples of constructed binary time series, a selective account of these follows.

1. Cycles in economic activity are often described with a binary random variable. Here a series $y_t$ is chosen to represent economic activity and a cycle in it involves expansions, $S_t = 1$, and contractions, $S_t = 0$, which are extracted from the $y_t$ using some rule. In the event that the series $y_t$ represents the level of economic activity then it is the *business cycle* that is being isolated. If a permanent component is taken away from $y_t$ we are investigating the *growth cycle*. In some instances a number

of variables representing economic activity are utilized to construct the $S_t$ e.g. those provided by the NBER.

2. Bull and bear markets are often described with a binary variable. The underlying variable here will be some asset price e.g. the Dow-Jones or the S&P500. Similar sets of rules are used to construct the binary variables as adopted in dating business cycles. A number of series exist for such $S_t$ e.g. Pagan and Sossounov (2003) and Candelon et al (2008) for Asian stock markets.

3. Financial crises. Here a unity indicates that a crisis is occurring while a zero indicates that this is not a crisis period — see Eichengreen et al (1995), Kaminsky and Reinhart (1999), Claessens et al (2008) and Bordo et al (2001). The latter state (p 55) that "We construct the familiar index of exchange market pressure (calculated as a weighted average of exchange rate change, short-term interest rate change, and reserve change...). A crisis is said to occur when this index exceeds a critical threshold". Models are then constructed to see if an outcome of $S_t = 1$ can be predicted, and these are the basis of the "vulnerability" or "early warning systems" literature.

4. IPO markets are often classified as hot ($S_t = 1$) and cold ($S_t = 0$) - see Ibbotson and Jaffe (1975), and Brailsford et al (2001) - depending upon either the volume of new offers or the excess returns earned on them.

5. Commodity and real estate markets are often classified as booms and slumps depending upon movements in the underlying prices e.g. Cashin et al. (2002).

6. In recent times a literature has emerged which looks at contagion and which constructs binary variables that aim to capture features of the joint movement in the extreme values found in two series $x_t, y_t$. These might be equity returns in two countries. This literature is often referred to as the study of "co-exceedances". It involves the construction of a binary variable $S_t = 1(\ y_{1t} > c, x_{1t} > c)$ - see Bae et al (2003). Sometimes this is augmented with a quantitative measure of the extremes such as $z_t = \min(y_t, x_t) S_t$ -see Baur and Schulze (2005). After construction, the $S_t$ and $z_t$ are often used as the dependent variable

in some regression model e.g. Bae et al (2003) fit a logistic regression while Baur and Schulze use a quantile regression.

One could continue on in this vein but, as the examples above indicate, there are many situations in which binary random variables are constructed from some observed continuous random variable and then utilized in some way e.g. in a regression. This therefore raises the question of whether any econometric problems arise from such operations.

The next section outlines some of the rules used to construct $S_t$ from $y_t$. As might be expected the DGP of $S_t$ will be determined by the interaction of whatever dating rule is adopted and the DGP for $y_t$. Section 3 shows that we would expect that the resulting $S_t$ will exhibit time series dependence i.e. they will follow a Markov process (MP), and we will describe simple regression methods for modeling the $S_t$. A failure to make an allowance for the Markov nature of $S_t$ when it is used in empirical work leads to potentially invalid inferences and biases. Yet it has been a common assumption within the literature that the $S_t$ have no dependence, as seen in the work of Birchenall et al (1999), Chin et al (2000), and Estrella and Mishkin (1998) who all advocate and work with a static discrete choice model (which we take to be Probit for illustrative purposes but this is not important). Whether a *dynamic* Probit model would be sufficient to capture this dependence is something we also take up in section 3. A crucial element in this discussion will be the fact that the rules adopted to construct $S_t$ from $y_t$ are often designed to ensure that certain patterns are seen in the history of $S_t$. We demonstrate that such *censoring* imposes restrictions upon the Markov process generating $S_t$ which simple dynamic Probit models cannot easily capture, whereas our regression methods automatically adapt to the phenomenon. Section 4 then considers how one would model and/or use the $S_t$ when there are other covariates determining it. Some parametric and non-parametric methods are described for modelling the relationship between $S_t$ and the covariates. Finally, Section 5 provides an example of the methods that focusses on the question of whether the probability of a recession depends on the yield spread.

## 2  Constructing the States

*Rules* for determining the states can be of two types, depending on whether they emphasise *turning points* in the underlying series or focus on sequences

of observations that *terminate* a phase such as an expansion or a recession. A turning point rule finds local maxima and minima in the series $y_t$. These represent (say) the peaks and troughs of a business cycle. A termination rule prescribes an event which would cause a change in the value of the state $S_t$. In turn termination rules could either be *non-parametric* or derive from a *parametric* model of $y_t$.

To illustrate these distinctions suppose we consider defining a business cycle. Perhaps the simplest definition is what might be termed the *calculus rule*. This says that a peak in a series on activity, $y_t$, occurs at time $t$ if $\Delta y_t > 0$ and $\Delta y_{t+1} < 0$. Thought of in this way it is a turning point rule. The reason for the name is the result in calculus that identifies a maximum with a change in sign of the first derivative from being positive to negative. A trough (or local minimum) can be found using the outcomes $\Delta y_t < 0$ and $\Delta y_{t+1} > 0$. The states $S_t$ are simply defined in this case as $S_t = 1(\Delta y_t > 0)$, so that $S_t$ depends only on contemporaneous information. This rule has been popular for defining a business cycle when $y_t$ is yearly data, see Cashin and McDermott (2002 ) and Neftci (1984). Note that we might also think of this rule as a termination rule since recessions terminate when growth is positive, showing that the two types of rule are not always distinct.

In practice one does not define a recession in this way. When data occurs at (say) the quarterly or monthly frequency one needs to recognize that common usage of a word like "recession" would identify it with a *sustained* decline in the level of economic activity i.e. something that lasts for several periods. If one applied the calculus rule most likely there would be too many turning points, since the growth rate might often switch sign between one period and the next. Visualizing a peak in a series leads one to the idea that a local peak in $y_t$ occurs at time $t$ if $y_t$ exceeds values $y_s$ for $s$ in a window $t-k < s < t+k$, and where $k$ is chosen in some way. One can define a trough in a similar way. By making $k$ large enough we also capture the idea that the level of activity has declined (or increased) in a sustained way. Of course we need to limit the window in time over which this test is applied. For later reference we note that, in this instance, $S_t$ depends upon $y_{t\pm j}, j = 0, ..., k$ and so future values of $y_t$ are needed to determine the value of $S_t$ i.e. to know whether a turning point occurred at time $t$ we need to know the future behavior of $y_t$.

A turning point rule based on a non-zero window is the basis of the NBER business cycle dating procedures summarized in the Bry and Boschan (1971) dating algorithm. In that program, designed for the analysis of monthly data,

$k = 5$. However, because much analysis is conducted with quarterly data, an analogue with such data would seem to be $k = 2$. We will refer to this latter rule as the BBQ rule. It is an automated dating rule and therefore differs from the NBER Dating Committee's decisions since the latter utilizes a number of series for $y_t$ and exercises some judgement. But the correspondence in dates produced by the automated procedure (BBQ) and the NBER choices is close, and the situation is therefore reminiscent of the popular use of an interest rate rule to describe the Fed's setting of the Federal Funds rate. It captures the essence of decisions without the fine detail. These turning point rules have been used in other contexts than the business cycle e.g. the dating of bull and bear markets in equity prices by Pagan and Sussonov (2003), Bordo and Wheelock (2006) and Claessens et al (2008).

A termination rule that is often cited in the financial press is that a recession can be identified by a "two quarters rule" summarized as (assuming that it is symmetric when defining expansions):

$$
\begin{aligned}
S_t &= 1 \text{ if } (\Delta y_{t+1} > 0, \Delta y_{t+2} > 0 | S_{t-1} = 0). \\
S_t &= 0 \text{ if } 1(\Delta y_{t+1} < 0, \Delta y_{t+2} < 0 | S_{t-1} = 1) \\
S_t &= S_{t-1} \text{ otherwise.}
\end{aligned}
\tag{1}
$$

This rule is non-parametric in the sense that it looks for patterns in the data without making any assumptions about the DGP of $y_t$. Lunde and Timmermann (2004) used a variant of this non-parametric rule for stock prices while hot and cold markets for IPO's were identified by Ibbotson and Jaffee (1975), with a hot market being signalled by whether excess returns and their changes for two periods exceed the median values. Eichengreen et al. (1995) and Classens et al (2008) employ rules of this type to establish the location of crises in time.

Parametric (model-based) termination rules proceed by working with a parametric model of $\Delta y_t$. Perhaps the best known of these arises by assuming that $\Delta y_t$ is a function of a latent binary variable $\xi_t$ that follows a Markov process, and to then construct a series of binary states using the *MS rule* $S_t = 1\{\Pr(\xi_t = 1|F_t) - 0.5\}$, where $F_t$ is a set containing either the past history of the observed random variable $\Delta y_t$ or perhaps the complete sample of observations - see Hamilton (1989). Of course one could use other parametric models of $\Delta y_t$ to produce $S_t$ e.g. a SETAR model, and the threshold for $\Pr(\xi_t = 1|F_t)$ could be set differently to 0.5. In all cases like this a clas-

sification into binary outcomes $S_t$ is produced which will effectively involve checking if movements in some function of the $\Delta y_t$ (and its lags) exceeds a threshold. Notice that in the MS case the binary states that define recessions and expansions etc are the $S_t$ *not the* $\xi_t$. There may be no simple relation between these two binary random variables and it will be rare for $S_t = \xi_t$. Others who have used parametric termination rules are Maheu and McCurdy(2000) (an MS model for stock prices), Brailsford et al (2001) ( an MS model for IPO's) and Abiad (2003) (an MS model to establish crises).

Another important feature of many constructed states is that extra censoring rules are applied based on imposing a minimum or maximum time that can be spent in a particular state. Thus, for the business cycle dates as published by the NBER, recessions and expansions must be five months long and a complete cycle must last for 15 months. In quarterly terms these are best interpreted as requiring a two quarter minimum phase length and 5 quarters for a complete cycle. To implement these restrictions one generally proceeds in a two stage fashion. In the first stage turning points are established with the basic rule. Then if (say) there is some recession which only lasted one quarter, the turning points which demarcate that recession are deleted and the revised set of turning points will not have such a recession embodied in them. Consequently, the final turning points are a sub-set of the original ones i.e. the original ones are *censored*. We might term this a "hard" censoring constraint since it explicitly over-rides the original set of dates. However, it may also be the case that "soft" censoring is present which stems from the nature of the basic rule that determines turning points, phase changes etc. An example would be the "two quarters" rule, as this automatically makes recessions and expansions have a minimum duration of two quarters. It also seems likely that, for non-parametric dating rules that utilize a threshold on $y_t$ for deciding on the value of $S_t$, the threshold is often implicitly chosen in order to produce such an outcome. Minimum duration of phases is certainly evident in a lot of the $S_t$ that have been constructed.[1]

---

[1]In many instances use has been made of the Bry and Boschan algorithm for detecting turning points in series such as stock prices. Thus in these cases the binary random variables must obey the NBER censoring constraints.

# 3  The DGP of The Binary States

## 3.1  Interactions of Rules and the DGP of $\Delta y_t$

Because the $S_t$ are binary time series they need to be thought of as coming from a Markov Process (MP). The issue then becomes what order of MP they are likely to be and how the underlying data $y_t$ and rules interact to determine this. It is impossible to provide a general account of this for any given $S_t$, partly due to the difficulty of agreeing on a process for $\Delta y_t$ ( and perhaps even the nature of $y_t$) and partly because the rule used for constructing the $S_t$ may not be fully disclosed e.g. as with the NBER Business Cycle Dating Committee. This will necessitate working with models of $S_t$ that adapt to the data. Nevertheless, in order to gain an understanding of how the DGP of $\Delta y_t$ interacts with a rule, it will be useful to run through a number of simple cases in which one of these is varied and the other is held constant.

We begin by studying what type of MP eventuates when $y_t$ is generated as a random walk with drift

$$\Delta y_t = \mu + \sigma e_t, \tag{2}$$

where $e_t$ is $i.i.d(0, \sigma^2)$, and the calculus rule is employed for dating i.e. $S_t = 1(\Delta y_t > 0)$. We will tentatively assume that this generates a first order MP - termed MP(1). Hamilton (1994 p684) shows that if the $S_t$ is an $MP(1)$ the following identity holds:

$$S_t = p_{01} + (1 - p_{01} - p_{10})S_{t-1} + \eta_t, \tag{3}$$

where $\eta_t$ is discrete and conditionally heteroskedastic ( since it depends upon $S_{t-1}$) and

$$p_{jk} = \Pr(S_{t+1} = k | S_t = j). \tag{4}$$

We therefore need to evaluate the $p_{jk}$ under the chosen scenario. This is straightforward since

$$
\begin{aligned}
p_{10} &= \Pr(S_{t+1} = 0 | S_t = 1) \\
&= \Pr(\Delta y_{t+1} < 0 | \Delta y_t > 0) \\
&= \Pr(\Delta y_{t+1} < 0) = \psi,
\end{aligned}
$$

due to independence of $\Delta y_t$. In the same way $p_{01} = 1 - \psi$ and, from (3),

$$S_t = 1 - \psi + (0 \times S_{t-1}) + \eta_t, \tag{5}$$

showing that there is no serial correlation in the states $S_t$ i.e. it is an $MP(0)$.

What happens if one relaxes the assumption that $y_t$ follows a random walk with drift ( and Gaussian innovations) but retains the calculus rule? Kedem(1980, p34) sets out the relation between the autocorrelations of the $\Delta y_t$ and $S(t)$ processes. Letting $\rho_{\Delta y}(k) = corr\left(\Delta y_t, \Delta y_{t-k}\right)$, and $\rho_S(k) = corr\left(S_t, S_{t-k}\right)$, he determines that

$$\rho_S(k) = \frac{2}{\pi}\arcsin\left(\rho_{\Delta y}(k)\right). \tag{6}$$

Thus an AR(k) process for $\Delta y_t$ will mean an MP(k) for the $S_t$ process. It is probably not surprising that the DGP of $\Delta y_t$ affects the DGP of $S_t$.

Now let us return to the case where $y_t$ is a random walk with drift and use the "two quarters rule" for dating phase shifts rather than the calculus rule. Proceeding as before we take the $S_t$ process to be an MP(1).[2] Then the appendix shows that

$$p_{10} = \frac{\psi^2}{(1+\psi)}, \; p_{01} = \frac{(1-\psi)^2}{2-\psi} \tag{7}$$

$$p_{11} = \frac{1+\psi-\psi^2}{(1+\psi)}, \; p_{00} = \frac{1+\psi-\psi^2}{2-\psi}. \tag{8}$$

Hence, using (3), we will have

$$S_t = \frac{(1-\psi)^2}{2-\psi} + [1 - \frac{(1-\psi)^2}{2-\psi} - \frac{\psi^2}{(1+\psi)}]S_{t-1} + \eta_t, \tag{9}$$

and this example shows that the dating rule employed will make the $S_t$ process at least an MP(1).

## 3.2  Estimating Models of $S_t$

As these examples show it will be rare for the $S_t$ to have no dependence. In general it will follow an MP of non-zero order also so we now briefly look at

---

[2]The process might actually be of second or higher order but, for the purpose of comparison with the combination we started with, it is useful to focus upon the implications for an MP(1). Note that the situation is like approximating an AR of higher order with an AR(1).

the form of MPs. These have a simple additive structure, which is evident in the MP(2) below

$$S_t = \phi_0 + \phi_1 S_{t-1} + \phi_2 S_{t-2} + \phi_4 S_{t-1} S_{t-2} + \eta_t, \tag{10}$$

where $E_{t-1}(\eta_t) = 0$. Higher order MPs involve all products of $S_{t-j}$ taken two at a time, then three at a time etc. To illustrate this an MP(3) would be

$$\begin{aligned} S_t &= \phi_0 + \phi_1 S_{t-1} + \phi_2 S_{t-2} + \phi_3 S_{t-1} S_{t-2} + \phi_4 S_{t-1} S_{t-3} \\ &\quad + \phi_5 S_{t-2} S_{t-3} + \phi_6 S_{t-1} S_{t-2} S_{t-3} + \eta_t. \end{aligned}$$

For simplicity we will generally work with the second order case above.

Estimation of (10) is easy. OLS will provide consistent estimators, although one might improve on its efficiency by making an assumption about the nature of the heteroskedasticity in $\eta_t$. The main difficulty is that, as the order of the MP increases, we may find that there are not enough observations in a given sample to ensure that the regressors are independent of one another. In work with MPs various simplifications have been proposed that involve restricting the parameters in some way e.g. Raftery (1985), but we do not use these in what follows, although they may be useful in empirical work.

Most of the extant literature working with the $S_t$ has not adopted the MP framework. Instead a static Probit model has been proposed and the assumption used in constructing the log likelihood was that the $S_t$ had no dependence. Since this is unlikely to be correct the likelihood has therefore been mis-specified. In an attempt to deal with the dependence it has been suggested that a dynamic Probit (DP) model might be used instead - see Deuker(1997). Now one has to exercise some care here in deciding on what is meant by a DP model. In micro-econometrics the Probit model has the following structure

$$\begin{aligned} \zeta_t &= x_t'\beta + \varepsilon_t, \ \varepsilon_t \tilde{} n.i.d.(0,1) \\ \\ z_t &= 1(\zeta_t > 0), \end{aligned} \tag{11}$$

and this implies that $E(z_t|x_t) = \Phi(x_t'\beta)$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal. A natural correspondence to our case would seem to be to set $S_t = z_t$. But what are the $x_t$? If $x_t = \zeta_{t-j}$ then the estimation task becomes very complex indeed. Possibly for this reason a popular version of the DP

model used in working with the $S_t$ has been to set $x_t$ to lags of $S_t$ rather than $\zeta_t$. Such a model has been studied by de Jong and Woutersen (2007) and used by Dueker (1997) *inter alia*. A likelihood is derived for this model under the assumption that $E(S_{t-1}\varepsilon_t) = 0$. However the latter is generally not true of constructed $S_{t-1}$ since these are often formed using contemporaneous and future values of the underlying data they are constructed from For example, with the two quarters rule for business cycle turning points, $S_t$ depends on the outcomes of $\Delta y_t, \Delta y_{t+1}$ and $\Delta y_{t+2}$, and so there must be a correlation between $S_{t-1}$ and $\varepsilon_t$. Studies that just add $S_{t-1}$ to the single index and proceed with a Probit model will therefore produce inconsistent estimators of $\beta$. This is also true of the $S_t$ set out by the NBER. Modelling the dependence between $S_{t-1}$ and $\varepsilon_t$ would be a very complex task, particularly if we do not know exactly what rule is used and what $y_t$ the $S_t$ are based on.

There is another approach. As noted above the Probit model implied a specific functional form connecting $z_t$ and $x_t$ i.e. $\Phi(x_t'\beta)$. Although there is nothing to suggest that this will be true when $x_t = S_{t-1}$, we might nevertheless *assume* that it is correct. It is this assumption that we will take to be a DP model. To investigate its adequacy in capturing the MP process that is the DGP of $S_t$ we assume that the orders of the MP and DP processes are equal. Then a DP(2) process would be

$$\begin{align}
\Pr(S_t &= 1|S_{t-1}, S_{t-2}) = E(S_t|S_{t-1}, S_{t-2}) \tag{12}\\
&= \Phi(c_0 + c_1 S_{t-1} + c_2 S_{t-2} + c_3 S_{t-1} S_{t-2}). \tag{13}
\end{align}$$

To compare this to an MP(2) suppose we perform an expansion of $\Phi(c_0 + z_t)$ around $z_t = 0$, where $z_t = c_1 S_{t-1} + c_2 S_{t-2} + c_3 S_{t-1} S_{t-2}$. Then we would get $\Phi(c_0 + z_t) = c_0 + \rho_1 z_t + \rho_2 z_t^2 + ...$ where $\rho_j = \frac{\partial^j \Phi(c_0+z)}{\partial z^j}|_{z=0}$. Now the terms $z_t^j$ can be written as $\gamma_{1j} S_{t-1} + \gamma_{2j} S_{t-1} + \gamma_{3j} S_{t-1} S_{t-2}$ by utilizing the fact that $S_{t-k}^j = S_{t-k}$. The coefficients $\gamma_{kj}$ are functions of the four coefficients $c_j$. Hence

$$E(S_t|S_{t-1}, S_{t-2}) = c_0 + \psi_1 S_{t-1} + \psi_2 S_{t-2} + \psi_3 S_{t-1} S_{t-2}, \tag{14}$$

and so the DP(2) has the form of an MP(2). Consequently, in this case, provided we make the number of parameters in the DP model the same as in the MP process, the former is capable of replicating the latter.

## 3.3 Effects of Censoring Rules on the DGP of the States

Now, as we observed earlier, censoring is a pervasive feature of situations where the $S_t$ are constructed. To investigate how this impacts upon the models needed to capture the dependence in $S_t$ we will begin with an MP(2) process for $S_t$. For illustrative purposes, the NBER censoring constraint that recessions and expansions have a minimum duration of two quarters is chosen. One reason for this is that the data employed in Section 5 has been generated in such a way. Thus we begin with

$$\Pr(S_t = 1 | S_{t-1}, S_{t-2}) = \phi_0 + \phi_1 S_{t-1} + \phi_2 S_{t-2} + \phi_3 S_{t-1} S_{t-2}.$$

Consider any 3-tuple of $\{S_t, S_{t-1}, S_{t-2}\}$. In the event that $\{S_{t-1} = 0, S_{t-2} = 1\}$ it must be the case that $S_t = 0$, since recessions have to be of two-period duration. Similarly $\{S_{t-1} = 1, S_{t-2} = 0\}$ means that $S_t = 1$. Thus the censoring restriction implies that

$$\Pr(S_t \;=\; 1 | S_{t-1} = 0, S_{t-2} = 1) = 0 \tag{15}$$
$$\Pr(S_t \;=\; 1 | S_{t-1} = 1, S_{t-2} = 0) = 1. \tag{16}$$

There are no restrictions for the 3-tuples $\{S_t = (0,1), S_{t-1} = 1, S_{t-1} = 1\}$ and $\{S_t = (0,1), S_{t-1} = 0, S_{t-1} = 0\}$. Translating (15) and (16) into parametric restrictions on the MP(2) we get

$$\phi_0 + \phi_2 \;=\; 0 \tag{17}$$
$$\phi_0 + \phi_1 \;=\; 1. \tag{18}$$

Thus the presence of censoring induces restrictions upon the nature of the MP(2). Indeed, in the regression (10) we would have a zero residual for any sequence $(1,1,0)$ and $(0,0,1)$ i.e. the observations at the beginning of an expansion and a contraction.

The parametric restrictions above does show up in regressions with NBER-defined states representing the business cycle. Using the $S_t$ from their web page over 1953/2 to 2001/4 we get

$$S_t = 0.429 + 0.571 S_{t-1} - 0.429 S_{t-2} + 0.37 S_{t-1} S_{t-2}. \tag{19}$$

Imposing the constraints on (10) involves estimating a regression of the form

$$\Delta S_t = \phi_0 (1 - S_{t-1} - S_{t-2}) + \phi_4 S_{t-1} S_{t-2},$$

12

and this gives identical results to those in (19) i.e. the regression model (10) *automatically imposes* any censoring constraints even if they are unknown to us, and this is a decided advantage of it.

To see what happens if the MP is of higher order, suppose we are dealing with an MP(3). Then this will be

$$
\begin{aligned}
S_t &= \phi_0 + \phi_1 S_{t-1} + \phi_2 S_{t-2} + \phi_3 S_{t-1} S_{t-2} + \phi_4 S_{t-1} S_{t-3} \\
&\quad + \phi_5 S_{t-2} S_{t-3} + \phi_6 S_{t-1} S_{t-2} S_{t-3} + \eta_t,
\end{aligned}
$$

and the restrictions become

$$
\begin{aligned}
1 &= \phi_0 + \phi_1 + \phi_4 S_{t-3} \\
0 &= \phi_0 + \phi_2 + \phi_5 S_{t-3}
\end{aligned}
$$

Since $S_{t-3}$ could be either one or zero, unless $\phi_4 = 0$ there would be two incompatible restrictions for the first set above, these corresponding to $S_{t-3} = 1$ and $S_{t-3} = 0$. A similar argument for the second restriction means that $\phi_5 = 0$. Consequently, we are left with just the term $S_{t-1} S_{t-2} S_{t-3}$ as an addition to the MP(2) model[3].

Are such restrictions compatible with a DP model? It is clear that a DP model would need to satisfy

$$
\begin{aligned}
1 &= \Phi(c_0 + c_1) \\
0 &= \Phi(c_0 + c_2),
\end{aligned}
$$

and this would mean that $c_0 + c_1 = \infty$ and $c_0 + c_2 = -\infty$. Thus, although we saw in the previous sub-section that the DP model is a restricted MP, once one imposes censoring constraints on the states this correspondence breaks down. Indeed, in the few studies where $S_{t-j}$ have been added to the single index in a DP model, an example being Dueker (1997), it is noticeable that only $S_{t-1}$ has been added. If $S_{t-2}$ had been included $\Phi(\cdot)$ would take values that cannot be defined in a log likelihood.

---

[3]In the previous version of this paper we described how one can test for the order of an MP, starting with some pre-defined higher order and testing downwards, just as is done in any standard autoregressive set up. To compare an MP(2) and MP(3) with NBER censoring restrictions we therefore need to test if the coefficient of $S_{t-1} S_{t-2} S_{t-3}$ is zero. The t ratio for this is -0.63 when working with the NBER data, so an MP(2) seems the preferred model for that data.

Is there a formulation of the DP model that will adapt to censoring? Technically a variant that ensures this is to *assume* that

$$E(S_t|S_{t-1}, S_{t-2}) = c_0 S_{t-1} S_{t-2} + c_1(1 - S_{t-1})(1 - S_{t-2})$$
$$+ S_{t-1}(1 - S_{t-2}) - (1 - S_{t-1})S_{t-2},$$

as this makes $\Pr(S_t = 1|S_{t-1} = 0, S_{t-2} = 1) = 0$ and $\Pr(S_t = 1|S_{t-1} = 1, S_{t-2} = 0) = 1$. There are certain technical difficulties with this formulation and we leave its development for later research. It should be said however that the correct formulation is to treat $S_t$ as an MP, so that the DP model is just an approximation. Consequently, it is not clear why there are any advantages in using a DP estimator over the simple regression based MP estimator given above. Any difficulties encountered in estimating the MP process due to high orders will be shared by the DP process as well.

# 4 The Impact of Covariates upon the DGP of the States

## 4.1 MPs with Covariates and no Censoring Restrictions

We now wish to introduce covariates into the Markov process.[4] Again it is simplest to discuss the issues when there is a single covariate. From a theoretical perspective the extension to the multiple variable case will be obvious, although in practice it may be difficult numerically. Moreover, in many applications there is just a single covariate such as the yield spread. To gain some appreciation of how this complicates the analysis we provide a treatment in the appendix of $\Pr(S_t = 1|x_t)$ when the two quarters dating rule is applied to a series whose growth rate is driven by some variable $x_t$. It emerges that this conditional probability is not a function of just the contemporaneous value of $x_t$ ( as in the static Probit model) but involves all past values of $x_t$. This points to a modification of the Markov process in which $x_t$ influences the transition probabilities; we will call such processes MPC. For the second order MP an appropriate generalization to an MPC(2) format might be

$$S_t = \alpha(x_t) + \beta(x_t)S_{t-1} + \gamma(x_t)S_{t-2} + \delta(x_t)S_{t-1}S_{t-2} + \eta_t, \qquad (20)$$

---

[4]We are referring to extra determinants of $S_t$ over and above its past history.

where $\alpha(x_t), \beta(x_t), \gamma(x_t)$ and $\delta(x_t)$are some non-linear functions of $x_t$. Using the same argument as earlier it should be clear that a dynamic Probit model would have the same form as (20), with the polynomials effectively being of infinite order but combined together by functions of a small number of parameters.

## 4.2 MPs with Covariates and Censoring Restrictions

In the MPC(2) case the censoring restrictions now imply that

$$\begin{aligned}
\alpha(x_t) + \beta(x_t) &= 1 \\
\alpha(x_t) + \gamma(x_t) &= 0
\end{aligned}$$

so that the Markov process becomes

$$\Delta S_t = \alpha(x_t)(1 - S_{t-1} - S_{t-2}) + \delta(x_t)S_{t-1}S_{t-2} + \eta_t \qquad (21)$$

If the polynomials were linear we would be adding on to the regression of the previous section the regressors $x_t(1 - S_{t-1} - S_{t-2})$ and $x_t S_{t-1}S_{t-2}$. In that instance we can test the significance of the extra regressors to determine if there is any contribution from $x_t$. If the order of the polynomials is unknown we could proceed with various non-parametric approaches such as splies. In the next section we describe how to use kernel methods to perform non-parametric analysis that exploits the additivity of the terms in the MPC.[5]

---

[5]It may be possible to extend the DP model to handle censoring in the covariate case as described in the case when there are no covariates. It is hard to know whether it is worth developing such an extension. The proposed MPC estimator is non-parametric versus a parametric form for the DP, but we could have made the procedure parametric simply by assuming some known forms for the polynomials. Basically the Probit model assumes a specific parametric functional form for the conditional mean and then some extra assumptions are needed for it to handle censoring. Because the non-parametric MPC estimator is more general, we have not further developed the modified DP estimator described above.

## 4.3 Non-parametric Estimation of the MPC Model

To estimate the MPC(2) model we observe from (20) that

$$
\begin{aligned}
E(S_t|S_{t-1} &= 1, S_{t-2} = 1, x_t) = \alpha(x_t) + \beta(x_t) + \gamma(x_t) + \delta(x_t) \quad (22) \\
E(S_t|S_{t-1} &= 1, S_{t-2} = 0, x_t) = \alpha(x_t) + \beta(x_t) \quad (23) \\
E(S_t|S_{t-1} &= 0, S_{t-2} = 1, x_t) = \alpha(x_t) + \gamma(x_t) \quad (24) \\
E(S_t|S_{t-1} &= 0, S_{t-2} = 0, x_t) = \alpha(x_t), \quad (25)
\end{aligned}
$$

so that the polynomials in the MPC can be identified from the conditional expectations. Moreover

$$
E(S_t|x_t) = \sum_{j=0}^{1} \sum_{k=0}^{1} E(S_t|S_{t-1} = j, S_{t-2} = k, x_t) \Pr(S_{t-1} = j, S_{t-2} = k|x_t).
$$

$$(26)$$

Consequently, it is necessary to evaluate a number of conditional expectations if $E(S_t|x_t)$ is to be computed.

Let $S_{1t} = S_{t-1}$ and $S_{2t} = S_{t-2}$. Then we have

$$
\begin{aligned}
E(S_t|S_{t-1} &= s_1, S_{t-2} = s_2, x_t) = \int s f(s|s_1, s_2, x) ds \\
&= [\frac{\int s f(s, s_1, s_2, x) ds}{f(s_1, s_2, x)}]
\end{aligned}
$$

Replacing the densities by a kernel estimator produces

$$
\hat{E}(S|S_1, S_2, x) = \left[ \frac{\int s \frac{1}{Th^4} \sum_{t=1}^{T} K_z(\frac{S_t-s}{h}, \frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h}) ds}{\frac{1}{Th^3} \sum_{t=1}^{T} K_{z^-}(\frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h})} \right],
$$

where $K_z$ and $K_{z^-}$ are multivariate kernels for the variables $z_t = [\, S_t \quad S_{1t} \quad S_{2t} \quad x_t \,]'$ and $z_t^- = [\, S_{1t} \quad S_{2t} \quad x_t \,]'$. After making a change of variable to $\psi_t = \frac{S_t-s}{h}$ ( so that $s = S_t - h\,\psi_t$),

$$
\hat{E}(S|S_1, S_2, x) = \left[ \frac{\frac{1}{Th^3} \sum_{t=1}^{T} \int (S_t - h\psi_t) K_z(\psi_t, \frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h}) d\psi}{\frac{1}{Th^3} \sum_{t=1}^{T} K_{z^-}(\frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h})} \right].
$$

16

If the kernel is symmetric then $\int \psi_t K_z(\psi_t, \frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h}) d\psi = 0$, leaving

$$\hat{E}(S|S_1 = s_1, S_2 = s_2, x) = \left[ \frac{\sum_{t=1}^{T} S_t K_z(\frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h})}{\sum_{t=1}^{T} K_{z^-}(\frac{S_{1t}-s_1}{h}, \frac{S_{2t}-s_2}{h}, \frac{x_t-x}{h})} \right],$$

where $s_1$ and $s_2$ take the values of zero and unity. Because the variables $S_{1t}$ and $S_{2t}$ are binary, $\frac{S_{jt}-s_j}{h}$ is either zero, when $S_{jt} = s_j$, or will be a large number, because $h \to 0$ with $T$. Thus, with a product kernel of the form $K_1(\frac{S_{1t}-s_1}{h})K_2(\frac{S_{2t}-s_2}{h})K(\frac{x_t-x}{h})$, we get

$$\hat{E}(S_t|S_{t-1} = j, S_{t-2} = k, x) = \frac{\sum_{t \in I_{jk}} S_t K(\frac{x_t-x}{h})}{\sum_{t \in I_{jk}} K(\frac{x_t-x}{h})} \tag{27}$$

where $I_{jk} = \{t \text{ s.t. } S_{t-1} = j, S_{t-2} = k\}$.

In the same way

$$\Pr(S_{t-1} = j, S_{t-2} = k|x_t) = \frac{\sum_{t \in I_{jk}}^{T} K(\frac{x_t-x}{h})}{\sum_{t=1}^{T} K(\frac{x_t-x}{h})} \tag{28}$$

(27) and (28) are then used to compute (26).

To determine the asymptotic distribution of the estimator of $m(x) = E(S_t|x_t = x)$ we have $S_t = E(S_t|S_{t-1}, S_{t-2}, x_t) + \eta_t$ and $S_t = m(x_t) + v_t$, giving $v_t = \sum_{j=0}^{1} \sum_{k=0}^{1} K_t \eta_t \mathbf{1} (t \in I_{jk})$. Then, from Theorem 3.5 of Pagan and Ullah (1999, p110),

$$(Th)^{\frac{1}{2}} (\widehat{m}(x) - m(x)) = \hat{f}^{-1} \left[ (Th)^{1/2} \sum_{t=1}^{T} K_t(m(x_t) - m(x)) \right] + \hat{f}^{-1} \left[ (Th)^{-\frac{1}{2}} \sum_{t=1}^{T} v_t \right], \tag{29}$$

Since under general conditions $\widehat{f}^{-1}$ converges in probability to $f^{-1}$ the terms that we need to focus on are those in square brackets. Because $\eta_t$ is a martingale difference with respect to $\{S_{t-j}\}_{j=1}^{\infty}$, $K_t$ is bounded by a suitable choice of kernel, and $\mathbf{1}\,(t \in I_{jk})$ is also bounded, $v_t$ will be a martigale difference with bounded moments provided those of $\eta_t$ are.[6] Under these conditions a central limit theorem will apply and the last term in brackets will converge to $\mathcal{N}\left(0, \lim\limits_{T\to\infty} E\frac{1}{h}v_t\right)$. Turning to the first term in brackets Theorem 3.6 of Pagan and Ullah (1999) showed that, when $x_t$ was $i.i.d.$ and $m(\cdot)$ was well behaved, it was $o_p(1)$. Following the structure of that proof it will also be true that the result holds when $x_t$ follows a dependent process that has a suitable mixing condition imposed on it, and so we will assume that the $x_t$ (and $m(\cdot)$) are such as to asymptotically eliminate the first term.

To evaluate the variance of the asymptotic distribution we use the law of iterated expectations to get

$$
\begin{aligned}
E\left(\frac{1}{h}v_t\right) &= E\left(E\left\{\frac{1}{h}\sum_{j=0}^{1}\sum_{k=0}^{1}K_t^2\eta_t^2\mathbf{1}\,(t \in I_{jk})\,|x_t\right\}\right) \\
&= E\left(\left\{\frac{1}{h}\sum_{j=0}^{1}\sum_{k=0}^{1}K_t^2\sigma^2\,(x_t, t \in I_{jk})\Pr\,(t \in I_{jk}|x_t)\right\}\right), \quad (30)
\end{aligned}
$$

where $\sigma^2\,(x_t, t \in I_{jk})$ is the variance of $S_t$ conditional on $x$ and $t \in I_{jk}$. Because $S_t$ is binary we have

$$
\begin{aligned}
\sigma^2\,(x_t, t \in I_{jk}) &= E\left[\eta_t^2|x_t, t \in I_{jk}\right] \\
&= E(S_t|S_{t-1}=j, S_{t-2}=k, x)\left[1 - E(S_t|S_{t-1}=j, S_{t-2}=k, x)\right].
\end{aligned}
$$

Evaluating the expectation in (30) yields

$$
E\left(\frac{1}{h_t}v_t\right) = \sum_{j=0}^{1}\sum_{k=0}^{1}\int\frac{1}{h}K_t^2\sigma^2\,(x_t, t \in I_{jk})\Pr\,(t \in I_{jk}|x_t)\,f\,(x_t)\,dx_t. \quad (31)
$$

Making a change of variable to $\psi = \frac{x_t-x}{h}$, the right hand side of (31) becomes

$$
\sum_{j=0}^{1}\sum_{k=0}^{1}\int K\,(\psi)^2\,\sigma^2\,(x + h\psi, t \in I_{jk})\Pr\,(t \in I_{jk}|x + h\psi)\,f\,(x + h\psi)\,d\psi
$$

$$
(32)
$$

---

[6]Because $S_t$ is binary the variance of $\eta_t$ will be $\Pr(S_t = 1|S_{t-1}, S_{t-2}, x_t)(1 - \Pr(S_t = 1|S_{t-1}, S_{t-2}, x_t))$ and so will be bounded.

Applying Corollary 1 of Lemma 1 of Pagan and Ullah (1999,p 364) to (32),
$\int K(\psi)^2 \sigma^2 (x + h\psi, t \in I_{jk}|) \Pr(t \in I_{jk}|x + h\psi) f(x + h\psi) d\psi$ converges to

$\sigma^2 (x, t \in I_{jk}) \Pr(t \in I_{jk}|x) f(x) \int K(\psi)^2 d\psi$ as $T \to \infty$. Thus $(Th)^{\frac{-1}{2}} \sum_{t=1}^{T} v_t$

converges in distribution to

$$\mathcal{N}\left(0, f(x) \int K(\psi)^2 d\psi \sum_{j=0}^{1} \sum_{k=0}^{1} \sigma^2 (x, t \in I_{jk}) \Pr(t \in I_{jk}|x)\right) \qquad (33)$$

so that of $(Th)^{\frac{1}{2}} (\widehat{m}(x) - m(x))$ becomes

$$\mathcal{N}\left(0, \frac{\int K(\psi)^2 d\psi}{f(x)} \sum_{j=0}^{1} \sum_{k=0}^{1} \sigma^2 (x, t \in I_{jk}) \Pr(t \in I_{jk}|x)\right). \qquad (34)$$

We use (34) to establish asymptotic confidence intervals for the non-parametric estimator of the MPC process.

# 5 An Application to the Probability of Recessions Given the Yield Spread

We apply the methods developed above to assess the extent to which the yield spread ($sp_t$) affects the probability of a recession occurring. Estrella and Mishkin (1998) did this by applying a static Probit model to the NBER states i.e. they ignored the dependence and censoring in the binary variable $S_t$. In this application we use the methods described earlier to take account of the fact that the NBER states $S_t$ are neither independent, identically distributed nor uncensored. The conditional mean is calculated using (26) with a kernel that is a product of Gaussian densities.

Estrella and Mishkin (1998) find that the best fit occurs with the yield spread being lagged two quarters, and we continue with that assumption here, so that $x_t = sp_{t-2}$ in this application. Figure 1 plots the probability of a recession given the spread i.e. $E(1 - S_t|sp_{t-2})$, against $sp_{t-2}$. It is assumed that the NBER $S_t$ follow an MPC(2) process and the conditional expectation is computed using (26).[7] Also shown on the figure are the estimate of

---

[7]As reported earlier the NBER states were tested for MP(3) against MP(2) and the latter was favoured.

19

$Pr\left(S_t = 0|sp_{t-2}\right)$ obtained from a static Probit model and the 95% confidence bands obtained using the asymptotic results for the estimator of $1 - E(S_t|x_t)$ given in (34). It is clear that there is a difference between the probability of recession obtained from the static Probit and MPC(2) models at a number of values for the spread. Most notably this occurs for spreads in the range -0.55% to 0.5%, although there is close to being a significant difference for a spread around $-1\%$ - at that point the static Probit model yields a predicted probability of recession that is much lower than the MPC(2) model.

Figure 1: Probability of recession from MP(2) and Probit models conditional on the yield spread lagged two quarters
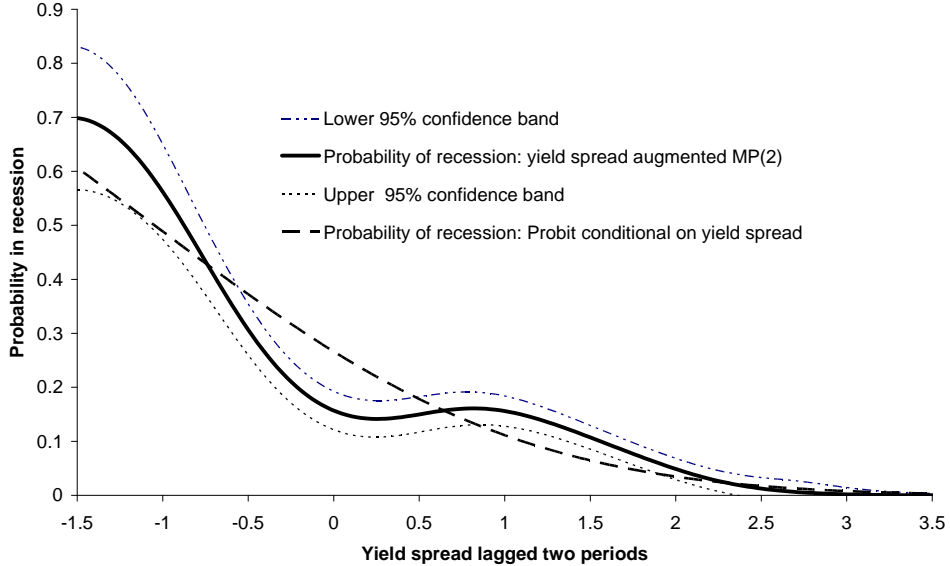


Having established that making an allowance for the nature of $S_t$ is both theoretically and empirically important it is of interest to evaluate the extent to which the yield spread is useful for looking at the probability of moving from an established phase to the opposite one. To assess this we focus on either the probability that an expansion which has lasted for two or more periods will be terminated or the probability of continuing in a contraction that has lasted for two or more periods. The former is the quantity $E(S_t = 0|x_t, S_{t-1} = 1, S_{t-2} = 1)$ while the latter is $E(S_t = 1|x_t, S_{t-1} = 0, S_{t-2} = 0)$.

The probability of leaving an expansion that has lasted for two or more

20

quarters is shown in Figure 2. There is a substantial difference between the estimates obtained from the non-parametric estimates of the MPC(2) model and those from a *dynamic* Probit model that uses $sp_{t-2}$ and $S_{t-1}$ as covariates (the variant used in some of the cycle literature). The dynamic Probit model over predicts the probability of leaving an expansion for yield spreads in the range -1.1% to 0.3%, and under predicts the probability of leaving an expansion for yield spreads below $-1.1\%$. These differences are statistically significant at the 5% level for spreads in the interval $-0.75\%$ to 0%. The main finding that the probability of terminating an expansion is low for spreads above $-0.75$ should be of interest to policy makers.

Figure 2: Probability of terminating an expansion that has lasted for two quarters
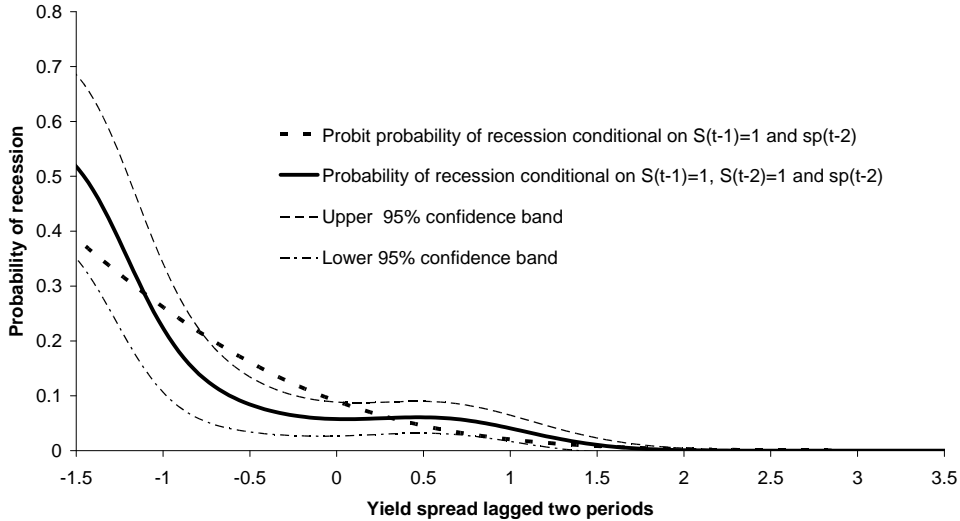


The probability of continuing in a recession that has lasted for two quarters is plotted in Figure 3. Again the probabilities are from the MPC(2) model estimated non parametrically and the dynamic Probit model. There is a substantial difference between the predicted probabilities from the two models, and this difference is both economically and statistically significant. The most important difference between the probabilities from the two methods is that the MPC(2) model suggests that there is no decrease in the

21

Figure 3: Probability of continuing in a recession
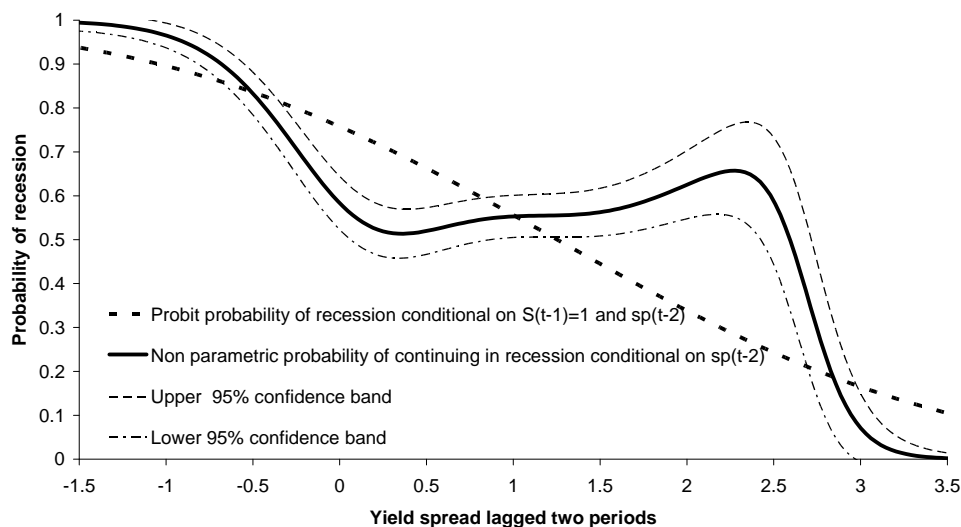


probability of staying in a recession with a rise in the yield spread from zero to 2.5 per cent. In contrast the DP(2) model suggests that the probability of remaining in recession declines monotonically as the yield spread increases.

Of course, one may question the accuracy of the asymptotic confidence intervals as there are only 10 per cent of cases where the economy is in contraction for two or more periods. But, even allowing for this caveat, the results presented above are likely to be of considerable practical interest.

# 6   Conclusion

We have made the argument that constructed states $S_t$ require careful treatment if they are to be used in econometric work, since they are very different in their nature to the binary states often modelled in micro-econometrics. When engaging in a broad range of estimation and inference methods one has to allow for the fact that they are essentially Markov processes. But, to date, the nature of the $S_t$ has mostly been ignored, with the potential for misleading estimates and inferences. We have suggested some methods to deal with this fact. In the application these methods produce results that

differ from those obtained by a standard Probit procedure that does not allow for the Markov process nature of the binary states and which forces a particular functional form upon the data. We have shown that these difference are economically and statistically significant.

# Appendix A

The determination of these transition probabilities becomes much more complex with the "two quarters rule" as the conditioning event $S_{t-1} = 1$ will place some restrictions upon the past sample paths for $\{\Delta y_t\}$ that are associated with an $ETS$. For example the sequence

$$\{\Delta y_{t+1}, \Delta y_t, \Delta y_{t-1}, \Delta y_{t-2}, ......\} = \{-, -, -., +, ....\} \tag{35}$$

would be incompatible with $S_{t-1} = 1$ since the negative growth at $t-1$ would match with the negative growth at $t$ and so the expansion would have been terminated at $t-1$. It is clear that the sample paths $\{\Delta y_{t-1}, \Delta y_{t-2}, ...\}$ that are compatible with $S_{t-1} = 1$ and $\{\Delta y_{t+1} < 0, \Delta y_t < 0\}$ must have the form $\{+, ...\}$ and in such paths we must encounter a $\{+, +\}$ before we encounter a $\{-, -\}$. If this did not happen e.g. we had for $\{\Delta y_{t-1}, \Delta y_{t-2}, ...\}$the path $\{+, -, +, -, -, ...\}$, then the recession would have begun at $t-5$ and would still be running when we reach $t-5$

Now let us consider an enumeration of the paths that are consistent with $S_{t-1} = 1$. This is done in the matrix below where the first column represents time and subsequent columns represent paths along which we are assured that $S_{t-1} = 1$. The notation used is as follows:

- "$+$" indicates $\Delta y_t > 0$;

- "$-$" indicates $\Delta y_t < 0$;

- "$*$" before a "$-$" indicates that any pattern for the observations can occur along the path up to and including that point;

- "$*$" following a "$+$" indicates that any pattern for the observations can occur along the path from that point forward.

Thus looking at the second column the "$+, +$" at $t$ and $t-1$ assures us that $S_{t-1} = 1$ along all paths that exhibit this pattern at $t$ and $t-1$.

23

Similarly, the "$-$" at $t$ and the "$+, +$" at $t-1$ and $t-2$ assures us that all paths with this pattern are consistent with $S_{t-1} = 1$. Similar logic can be applied to all the subsequent paths.

$$
\begin{bmatrix}
t+1 & * & * & * & * & * & * & \cdots \\
t & + & - & + & - & + & - & \cdots \\
t-1 & + & + & - & + & - & + & \cdots \\
t-2 & * & + & + & - & + & - & \cdots \\
t-3 & & * & + & + & - & + & \cdots \\
t-4 & & & * & + & + & - & \cdots \\
t-5 & & & & * & + & + & \cdots \\
t-6 & & & & & * & + & \cdots \\
\vdots & & & & & & * & \ddots
\end{bmatrix}
\tag{36}
$$

To understand the derivation of these paths suppose we start with the four possible outcomes for $(\Delta y_t, \Delta y_{t-1}\}$, namely $\{+,+\}, \{-,+\}, \{+,-\}$ and $\{-,-\}$. The last would give $S_{t-1} = 0$ and the first $S_{t-1} = 1$; thus the first becomes the second column of the table. The other two outcomes do not enable us to decide what the state for $S_{t=1}$ is and so we proceed to observation $t-2$ and consider what happens to each of them as we add on a $-$ or a $+$. Thus $\{-,+,+\}$ will give $S_{t-1} = 1$ and that becomes the third column. But $\{-,+,-\}$ produces no resolution and one needs to proceed to $t-3$. Augmenting $\{+,-\}$ with a $+$ also fails to resolve the indeterminacy while adding on a $-$ result in $S_{t-1} = 0$. Consequently that path has to be continued on to $t-3$ as well. The process continues in this way and all columns of the matrix will eventually be enumerated by such a strategy.

To formalize the discussion it is helpful to separate the set of paths that are consistent with $S_{t-1} = 1$ into two subsets. Let $E_t$ be the set of paths such that $\{\Delta y_t > 0 \ and \ S_{t-1} = 1\}$ and $F_t$ be the set of paths such that $\{\Delta y_t < 0 \ and \ S_{t-1} = 1\}$. If we introduce the notation that

- $[+-]_t^j$ represents the fragment of the path along which there are j repetitions of the pattern in the $[+-]$.with the leading term in the pattern being located at time t,

- $[++]_t$ represents the fragment of path where the pattern " $++$" occurs with the first " $+$ " being at $t$ and the second at $t-1$

24

- $[-]_t$ represents the case where $\Delta y_t < 0$,

the sets $E_t$ and $F_t$ can be enumerated as

$$E_t = \left\{ [++]_t \,;\, [+-]_t\,[++]_{t-2} \,;\, [+-]_t^2\,[++]_{t-4} \,;\, ...;\, [+-]_t^j\,[++]_{t-2j} \,;\, .... \right\} \quad (37)$$

$$F_t = \left\{ \begin{array}{c} [-]_t\,[++]_{t-1} \,;\, [-]_t\,[+-]_{t-1}\,[++]_{t-3} \,;\, \\ [-]_t\,[+-]_{t-1}^2\,[++]_{t-5} \,;\, ...;\, [-]_t\,[+-]_{t-1}^j\,[++]_{t-2j-1} \,;\, .... \end{array} \right\}. \quad (38)$$

Thus, using the notation that $\Pr(E_t)$ represents the probability that the path is drawn from the set $E_t$, and recognizing that the paths are mutually exclusive, we have ( to simplify notation we have omitted the conditioning on $\Im_{t+1}$ in equations (39), (40), (41) and (43).[8])

$$\Pr(E_t) = \sum_{j=0}^{\infty} \Pr\left( [+-]_t^j\,[++]_{t-2j} \right) \quad (39)$$

and

$$\Pr(F_t) = \sum_{j=0}^{\infty} \Pr\left( [-]_t\,[+-]_{t-1}^j\,[++]_{t-2j-1} \right). \quad (40)$$

By definition

$$\Pr(S_{t-1} = 1) = \Pr(E_t) + \Pr(F_t). \quad (41)$$

Interest also centres on the joint event $Pr\{S_t = 0, S_{t-1} = 1\}$; this will involve the set $G_{t+1}$ defined as

$$G_{t+1} = \left\{ \begin{array}{c} [--]_{t+1}\,[++]_{t-1} \,;\, [--]_{t+1}\,[+-]_{t-1}\,[++]_{t-3} \,;\, [--]_{t+1}\,[+-]_{t-1}^2\,[++]_{t-5} \,;\, \cdots \\ ...;\, [--]_{t+1}\,[+-]_{t-1}^j\,[++]_{t-2j-1} \,;\, .... \end{array} \right\} \quad (42)$$

Then, since $S_t$ is a stationary process,

---

[8]To simplify notation we have omitted the conditioning on $\Im_{t+1}$ in equations (39), (40), (41) and (43).

$$p_{10} = \frac{\Pr(S_t = 0, S_{t-1} = 1)}{\Pr(S_{t-1} = 1)} = \frac{\Pr(G_{t+1})}{\Pr(E_t) + \Pr(F_t)}. \tag{43}$$

If $\Pr(S_t = 1, S_{t-1} = 0)$ is constant, which essentially requires $\Delta y_t$ to be a random walk with time invariant drift and variance, then $\Pr(S_t = 1, S_{t-1} = 0) = \Pr(S_t = 0, S_{t-1} = 1)$ (as the number of peaks and troughs must be the same). Using this in conjunction with $\Pr(S_t = 0) = 1 - \Pr(S_t = 1)$ we can directly derive $p_{01}$ from the same information as used to construct $p_{10}$. If $\Pr(S_t = 1, S_{t-1} = 0)$ is time varying (as would be the case where $\mu_t$ depends on some exogenous variable) then one also needs to enumerate the various paths where $S_{t-1} = 0$.

Considering the limits of $E_t$ etc we get

$$
\begin{aligned}
\Pr(E) &= \sum_{j=0}^{\infty} (1 - \psi)^2 \left[ \psi(1 - \psi) \right]^j \\
&= \frac{(1 - \psi)^2}{1 - \psi(1 - \psi)} \tag{44}
\end{aligned}
$$

$$
\begin{aligned}
\Pr(F) &= \sum_{j=0}^{\infty} \psi (1 - \psi)^2 \left[ \psi (1 - \psi) \right]^j \\
&= \frac{\psi(1 - \psi)^2}{1 - \psi(1 - \psi)} \tag{45}
\end{aligned}
$$

$$
\begin{aligned}
\Pr(G) &= \sum_{j=0}^{\infty} \psi^2 (1 - \psi)^2 \left[ \psi (1 - \psi) \right]^j \\
&= \frac{\psi^2 (1 - \psi)^2}{1 - \psi (1 - \psi)} \tag{46}
\end{aligned}
$$

and so

$$
p_{10} = \frac{\psi^2}{(1 + \psi)} \tag{47}
$$

$$
p_{11} = \frac{1 + \psi - \psi^2}{(1 + \psi)} \tag{48}
$$

26

$$p_{01} = \frac{(1-\psi)^2}{2-\psi} \tag{49}$$

$$p_{00} = \frac{1+\psi-\psi^2}{2-\psi} \tag{50}$$

Now in some of the literature we deal with it is assumed that the process for $\Delta y_t$ depends linearly upon some other variable $x_t$ in the following way:

$$\Delta y_t = a + bx_t + u_t \tag{51}$$

where the $x_t$ are taken to be strictly exogenous (and so can be conditioned upon) and $u_t$ is $n.i.d.(0,1)$. If $\psi_t = \Phi(-a - bx_t)$, applying the enumeration method results in

$$
\begin{aligned}
\Pr\left(E_{t+1}|\Im_{t+1}\right) &= \sum_{j=0}^{\infty} \Pr\left([+-]_{t+1}^{j}[++]_{t+1-2j}\right) \\
&= \left(1-\psi_{t+1}\right)\left(1-\psi_t\right) \tag{52} \\
&\quad + \sum_{j=1}^{\infty}\left\{\left[\prod_{i=0}^{j-1}\left(1-\psi_{t+1-i}\right)\psi_{t-i}\right]\left(1-\psi_{t-2j+1}\right)\left(1-\psi_{t-2j}\right)\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\Pr\left(F_{t+1}|\Im_{t+1}\right) &= \sum_{j=0}^{\infty} \Pr\left([-]_{t+1}[+-]_{t}^{j}[++]_{t-2j}\right) \\
&= \psi_{t+1}\left(1-\psi_t\right)\left(1-\psi_{t-1}\right) + \tag{53} \\
&\quad \psi_{t+1}\sum_{j=1}^{\infty}\left\{\left[\prod_{i=0}^{j-1}\left(1-\psi_{t-i}\right)\psi_{t-i-1}\right]\left(1-\psi_{t-2j}\right)\left(1-\psi_{t-2j-1}\right)\right\}.
\end{aligned}
$$

Letting $P_t = Pr\left(S_t = 1|\Im_{t+1}\right)$ under the two quarters rule gives

$$P_t = \Pr\left(E_{t+1}|\Im_{t+1}\right) + \Pr\left(F_{t+1}|\Im_{t+1}\right) \tag{54}$$

It is clear from this expression that the use of the two quarters dating rule means that $P_t$ is a function not only of $x_t$ but also of $x_{t+1}$ and the entire past history of $x_t$. Moreover it does not have a single index form i.e. does not

27

depend upon $\alpha + x_t\beta$ alone. Only if the dating rule had been the "calculus" one would $\Pr(S_t = 1|\mathfrak{S}_{t+1}) = (1 - \psi_t)$ be a function of $x_t$ only. Clearly the lesson of this analysis is that one cannot just assume that $\Pr(S_t = 1)$ is a function of a contemporaneous variable only; it is necessary that one know how the $S_t$ were generated in order to be able to write down the correct likelihood.

# 7    References

Abiad, A. (2003), "Early Warning Systems: A Survey and a Regime-Switching Approach", *IMF Working Paper No. 3/32*.

Bae, H-H, G.A. Karolyi and R.M. Stulz (2003), "A New Approach to Measuring Financial Contagion", *Review of Financial Studies*, 16, 717-763.

Baur, D. and N. Schulze (2005), "Coexceedances in Financial Markets-a Quantile Regression Analysis of Contagion", *Emerging Markets Review*, 6, 21-43.

Birchenall, C.R., H. Jessen, D.R. Osborne and P. Simpson (1999), "Predicting U.S. Business-Cycle Regimes", *Journal of Business and Economic Statistics*, 17, 313-23.

Bordo, M, B. Eichengreen, D. Klingbiel and M. S. Martinez-Peria (2001), "Financial Crises", *Economic Policy*, 54-82.

Bordo, M.D. and D.C. Wheelock (2006), "When Do Stock Market Booms Occur? The Macroeconomic and Policy Environments of 20th Century Booms", *Federal Reserve Bank of St Louis Working Paper 2006-051A*

Brailsford, T., Heaney, R., A. and Shi, J., 2001, "The Cyclical Behaviour of the IPO market in Australia", *Accounting Research Journal*, 14, 1,17-34.

Bry, G., Boschan, C., (1971), *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, New York, NBER.

Candelon, B., J. Piplach and S. Straetmans (2008), " On Measuring Synchronization of Bulls and Bears" The Case of East Asia", *Journal of Banking and Finance*, 32, 1022-1035.

Cashin, P., C.J. McDermott and A. Scott (2002), "Booms and Slumps in World Commodity Prices", *Journal of Development Economics*, 69, 277-96.

Cashin, P. and C.J. McDermott (2002), "Riding on the Sheep's Back: Examining Australia's Dependence on Wool Exports", *Economic Record*, 78, 249-263.

Chin, D., Geweke, J., and P., Miller, (2000), "Predicting Turning Points", *Federal Reserve Bank of Minneapolis Research Department Staff Report No. 267*.

Claessens, S,. M. A. Kose and M. E. Terrones (2008), "What Happens During Recessions: Crunches and Busts", International Monetary Fund, Working Paper 08/274, December.

de Jong, R and T. Woutersen (2007), "Dynamic Time Series Binary Choice", *Econometric Theory* ( forthcoming)

Dueker, M. (1997), "Strengthening the Case for the Yield Curve as a Predictor of U.S. Recessions", *Federal Reserve Bank of St Louis Review*, 79, 41-51.

Eichengreen, B., A.K. Rose and C. Wyplosz (1995), "Exchange Rate Mayhem: The Antecedents and Aftermath of Speculative Attacks", *Economic Policy*, 21, 251-312.

Estrella, A. and F.S. Mishkin (1998), "Predicting US Recessions: Financial Variables as Leading Indicators", *Review of Economics and Statistics*, LXXX, 28-61.

Hamilton, J.D., (1989), "A New Approach to the Economic Analysis of Non-Stationary Times Series and the Business Cycle", *Econometrica*, 57, 357-84.

Ibbotson, R.G., and J.J. Jaffee (1975), " 'Hot Issue' Markets", *Journal of Finance*, 30, 1027-1042.

Kaminsky, G.I. and C.M. Reinhart (1999), "The Twin Crises: The Causes of Banking and Balance-of- Payments Problems", *American Economic Review*, 89, 473-500.

Kedem, B., (1980), *Binary Time Series*, Marcel Dekker, New York.

Lunde, A. and A.Timmermann (2004), "Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets", *Journal of Economic and Business Statistics*, 22, 253-73.

Maheu, J.M and T.H., McCurdy (2000), "Identifying Bull and Bear Markets in Stock Returns," *Journal of Business & Economic Statistics*, American Statistical Association, vol. 18(1), pages 100-112, January

Neftci, S.N. (1984), "Are Economic Times Series Asymmetric over the Business Cycle", *Journal of Political Economy*, 92, 307-28.

Pagan, A.R. and K. Sossounov (2003), "A Simple Framework for Analysing Bull and Bear Markets", *Journal of Applied Econometrics*, 18, 23-46.

Pagan, A.R. and A., Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press.

Raftery, A.E (1985), " A Model for High-Order Markov Chains", *Journal of the Royal Statistical Society*, Series B, 47, 528-539.