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

Don Harding[†] and Adrian Pagan[‡]

December 11, 2008

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 later are rarely i.d. We show how popular rules for constructing the binary states interact with the stochastic process of the variables they are constructed from to make these variables Markov processes. Consequently, one needs to recognize the Markov nature of the binary variables when performing analyses with them, and it is not valid to adopt a model like static Probit which fails to recognize this 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 Markov process which means that it is generally not correct to utilize a dynamic Probit model with them. Given this we describe methods for modelling with these variables that respects their Markov process nature and recognizes any censoring constraints. An application is provided that investigates the relation between the business cycle and the yield spread.

*We would like to thank a referee and an Associate Editor for constructive comments on an earlier version of the paper.

[†]*La Trobe University

[‡]Queensland University of Technology and University of New South Wales. Research supported by ESRC Grant 000 23-0244.

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 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, such 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). A failure to make an allowance for this fact when using S_t 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 take up in section 3. A crucial element in this discussion will be the fact that often the rules adopted to construct S_t from y_t are designed to ensure that certain patterns are seen in a history of S_t. We demonstrate that such censoring imposes restrictions upon the Markov process generating S_t which simple dynamic Probit models cannot capture. Section 4 then considers how one would model and/or use the S_t when there are other covariates determining it. 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) 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 Δy_t. Perhaps the best known of these arises by assuming that Δy_t is a function of a latent binary variable ξ_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 Δy_t or perhaps the complete sample of observations - see Hamilton (1989). Of course one could use other parametric models of Δy_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 classification into binary outcomes S_t is produced which will effectively involve checking if movements in some function of the Δy_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* ξ_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]

3 The DGP of The Binary States

3.1 Interactions of Rules and the DGP of Δy_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 Δy_t (and perhaps

[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.

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. 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 Δy_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 η_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 Δy_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 Δy_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 Δy_t will mean an MP(k) for the S_t process. It is probably not surprising that the DGP of Δy_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 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 a MP(3) would be

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

[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).

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 their efficiency by making an assumption about the nature of the heteroskedasticity in η_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 was 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 misspecified. 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 it will be necessary to know exactly what y_t is and, as we have pointed out, often this information is not available e.g. the NBER Dating Committee is rather vague about the precise derivation of S_t. If we were to treat ζ_t as a latent variable then the estimation task becomes very complex indeed since the single index depends on a latent variable. One version of the DP model that seeks to avoid such complexity, and which has sometimes been used in micro-econometric studies with S_t that are dependent, involves assuming that x_t include lags of S_t rather than ζ_t. Such a model has been studied by de Jong and Woutersen (2007). A likelihood is then 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 i.e. ζ_t. 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 Δy_{t+2}, and so there must be a correlation between S_{t-1} and ε_t. Studies that just add S_{t-1} to the single index and proceed with a Probit model will therefore produce inconsistent estimators of β. Modelling the dependence between S_{t-1} and ε_t would be a very complex task, particularly if we do not know exactly what rule is used and what y_t is.

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 fix the order of the MP and DP processes to be the same. Then a DP(2) process would be

$$\Pr(S_t = 1|S_{t-1}, S_{t-2}) = E(S_t|S_{t-1}, S_{t-2}) \quad (12)$$

$$= \Phi(c_0 + c_1 S_{t-1} + c_2 S_{t-2} + c_3 S_{t-1} S_{t-2}). \quad (13)$$

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 γ_{kj} are functions of the four coefficients c_j. Hence

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

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 4 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

$$\begin{aligned} \Pr(S_t &= 1|S_{t-1} = 0, S_{t-2} = 1) = 0 && (15) \\ \Pr(S_t &= 1|S_{t-1} = 1, S_{t-2} = 0) = 1. && (16) \end{aligned}$$

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

$$\begin{aligned} \phi_0 + \phi_2 &= 0 && (17) \\ \phi_0 + \phi_1 &= 1. && (18) \end{aligned}$$

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

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

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

$$\Delta S_1 = \phi_0(1 - S_{t-1} - S_{t-2}) + \phi_4 S_{t-1} S_{t-2}.$$

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

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

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 there would be two incompatible restrictions for the first set above corresponding to $S_{t-3} = 1$ and $S_{t-3} = 0$, unless $\phi_4 = 0$. 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. Thus there seems no reason to adopt a DP model in preference to the simpler MP approach.

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. Again it is simplest to discuss the issues when we have a single covariate. From a

[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.

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

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

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 clear that a dynamic Probit model would have the same form as (20), with the polynomials effectively being of infinite order.

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

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

If the polynomials were first order 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 $x_tS_{t-1}S_{t-2}$. In that instance we can test the significance of the extra regressors to determine if there is any contribution. One can do this non-parametrically by using a series expansion for the unknown α and δ.

4.3 Estimating MPs with Covariates

The problem now is to estimate the covariate-augmented MP model when there are censoring restrictions. To decide on how to proceed it is necessary to first ask what it is envisaged that the model will be used for. If it is to estimate $E(S_t|x_t)$ then we do not need to estimate $\alpha(x_t)$ and $\delta(x_t)$ separately. An alternative quantity of interest would be $E(S_t|x_t, S_{t-1} = 1, S_{t-2} = 1)$ i.e. what is the probability that we will continue in an expansion for various values of x_t (and a similar question can be asked about recessions)? This can also be computed without knowing the separate functional forms. Indeed it is hard to think of any issue that requires us to estimate $\alpha(x_t)$ and $\delta(x_t)$ separately and we will proceed under this presumption. In the event that they were of interest the additive nature of the model suggests that the polynomials be estimated as spline functions.

Let us therefore look at the estimation of $E(S_t|x_t)$. By the law of iterated expectations this is $E_{S(-1)S(-2)}E(S_t|x_t, S_{t-1}, S_{t-2})$ and, given the discrete nature of the S_t, it will become

$$E(S_t|x_t) = \sum_{k=0}^{1}\sum_{j=0}^{1} E(S_t|x_t, S_{t-1} = k, S_{t-2} = j) \Pr(S_{t-1} = k, S_{t-2} = j). \quad (21)$$

Now it is possible to estimate $E(S_t|x_t, S_{t-1} = k, S_{t-1} = j)$ by kernel regression. Letting $I_{kj,t} = \{t \text{ s.t. } S_{t-1} = k, S_{t-2} = j\}$, in large samples these expectations can be found by applying the kernel estimator to each of the four sets of observations distinguished by I_{kj}. The probabilities $\Pr(S_{t-1} = k, S_{t-2} = j)$ can be estimated by the fraction of observations in each set of outcomes. Of course if we were looking at $E(S_t|x_t, S_{t-1} = 1, S_{t-2} = 1)$ we would just use those observations in I_{11}.

To determine the asymptotic distribution of the estimator of $m(x) = E(S_t|x_t)$ we note that from Theorem 3.5 p 110 of Pagan and Ullah (1999) we have

$$(nh)^{1/2}(\hat{m}(x^*) - E_X m(x)) = \hat{f}^{-1}[(\frac{1}{Th})^{1/2}\sum_{i=1}^{T} K_t\eta_t],$$

Standard central limit theorems for independently distributed random variables mean that the term in brackets will be normal with zero mean and variance $lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{n}E_X(\frac{1}{h}K_t^2\eta_t)$, where $K_t = K(\frac{x_t - x^*}{h})$. This is

$$lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T}E(\frac{1}{h}K_t\sigma^2(x_t)).$$

Making the change in variable to $\psi = h^{-1}(x_t - x^*)$ we get

$$E(\frac{1}{h}K_t\sigma^2(x_t)). = \int K^2(\psi)\sigma^2(x^* + h\psi)d\psi$$

and, as $T \to \infty, h \to 0$,

$$lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{n}E(\frac{1}{h}K_t^2\sigma^2(x_t)) = \sigma^2(x^*)\int K^2(\psi)d\psi.$$

To compute $\sigma^2(x^*)$ we note that the conditional variance $\sigma^2(x_t) = E(S_t^2|x_t) - [E(S_t|x_t)]^2$. But because $S_t^2 = S_t$ this is just $(1 - E(S_t|x_t))E(S_t|x_t)$. Hence we compute the conditional variance by weighting the different sets of observations as for the mean in (21).

4.4 Modified DP with covariates and censoring restrictions

The DP model can be modified to deal with censoring restrictions by extending standard probit model as follows

$$\begin{aligned}
\zeta_t^c &= S_{t-1}S_{t-2}\left[x_t'\beta_{11} + \varepsilon_t\right] + \left(1 - S_{t-1}\right)\left(1 - S_{t-2}\right)\left[x_t'\beta_{00} + \varepsilon_t\right] \\
&\quad + S_{t-1}\left(1 - S_{t-2}\right) - \left(1 - S_{t-1}\right)S_{t-2} \qquad (22) \\
&\quad \varepsilon_t \tilde{\ } n.i.d.(0,1) \\
z_t^c &= 1(\zeta_t^c > 0),
\end{aligned}$$

The DP model allowing for censoring can be estimated by running separate Probit models on the two sub samples one where $S_{t-1} = 1$ and $S_{t-2} = 1$ and the other sub sample where $S_{t-1} = 0$ and $S_{t-2} = 0$. Thus $E(S_t|x_t, S_{t-1} = k, S_{t-2} = j) = \Phi\left(x_t'\beta_{kj}\right)$ and using the law of iterated expectations we obtain $E(S_t|x_t)$ using (21)

$$\begin{aligned}
E(S_t|x_t) &= \Phi\left(x_t'\beta_{11}\right)\Pr(S_{t-1} = 1, S_{t-2} = 1) \qquad (23) \\
+\Phi\left(x_t'\beta_{00}\right)\Pr(S_{t-1} &= 0, S_{t-2} = 0) + \Pr(S_{t-1} = 1, S_{t-2} = 0)
\end{aligned}$$

Where the last term in (23) arises because cycles are censored so that phases last at least two periods. The probabilities $\Pr(S_{t-1} = k, S_{t-2} = j)$ can be estimated by the fraction of observations in each set of outcomes.

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

We apply the method developed above to the extent to which the yield spread (sp_t) affects the probability of an expansion occurring. Estrella and Mishkin (1998) did this using a Probit model assuming that the NBER S_t were *i.d.* and uncensored. In this application we use the method described above 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 (21) with the kernel regression methods discussed in chapter 3 of Pagan and Ullah (1999).[4]

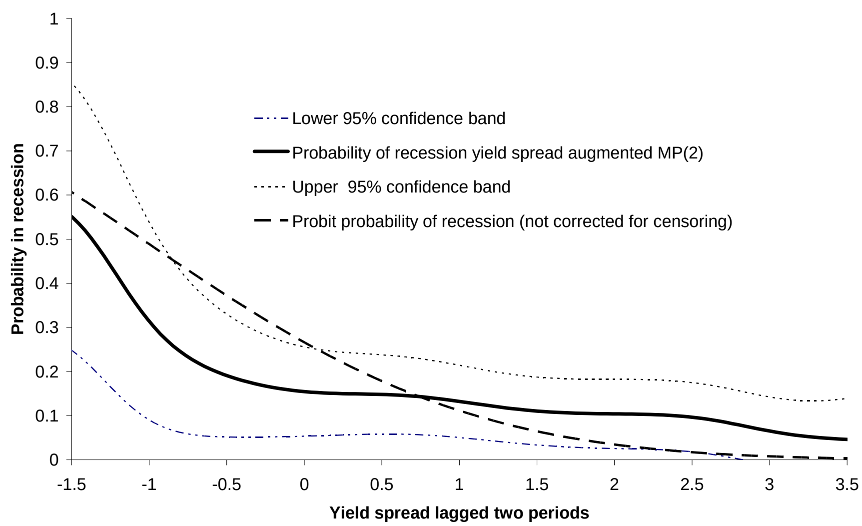
Estrella and Mishkin (1998) find the best fit using the yield spread lagged two quarters and we continue with that assumption here, so $x_t = sp_{t-2}$ in this application. Figure 1 plots the probability of recession given the spread, $E(1 - S_t|sp_{t-2})$, against sp_{t-2}. assuming that S_t follows an MPC(2) process obtained via (21).[5] Also shown on Figure 1 are 95% confidence bands obtained using the asymptotic results for $E(S_t|x_t)$ given earlier. Figure 1 also shows the estimate of $Pr\,(S_t = 0|sp_{t-2})$ obtained from a probit model. It is clear that there is a difference between the probability of recession obtained from a Probit model and that obtained via (21) from the MPC(2) model with sp_{t-2} as the covariate. Indeed the estimate of $Pr\,(S_t = 0|sp_{t-2})$ obtained from the Probit model lies outside of the two standard deviation confidence bands for spreads between 0.125 and -0.75 i.e. the Probit model over predicts the probability of recession when the yield curve is just inverted and this over-prediction is statistically significant. The MPC(2) model with sp_{t-2} as the covariate suggests that large negative yield spreads are needed to produce a high probability of recession. Another difference is that for yield spreads higher than 2.0 the Probit model yields a predicted probability of re-

[4]The window width used to compute $E(S_t|x_t, S_{t-1} = k, S_{t-2} = j)$ is $n_{kj}^{\frac{1}{5}}$ where n_{kj} is the number of cases where $(S_{t-1} = k, S_{t-2} = j)$.

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

cession that is statistically significantly below that obtained from the MP(2) model with sp_{t-2} as the covariate. This is explained by the fact that, as the yield spread becomes strongly positive, the MPC(2) model implies that the probability of a recession tends to its unconditional probability, which makes sense. However, the Probit model puts this probability to zero since its functional form does not allow such behaviour. That is, unlike the Probit model the MP(2) model does not predict that economies are assured to avoid a recession if the yield spread is sufficiently large and positive.

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



Earlier we cited papers that estimate DP models with lagged values of S_t as the explanatory variables. Such a DP model would take the form of (11). We have shown above that the model cannot be estimated when x_t contains both S_{t-1} and S_{t-2} as fitting the censoring of the data requires certain parameters to be infinite. We can, however, estimate models where $x_t = \{1, S_{t-1}, sp_{t-2}\}$. Such a model would be invalid because it does not allow for censoring but it is instructive to estimate this DP model and compare the probabilities of recession from this model with those obtained from the DP model (22) which does allow for censoring.

Figure 2 plots the probability of going into a recession next period conditional on the yield spread and being in an expansion last period. Clearly, there is a small difference between the DP model that allows for censoring and the one that does not make that allowance.

Figure 2: Probability of moving from expansion to recession in DP models that vary in whether they allow for censoring

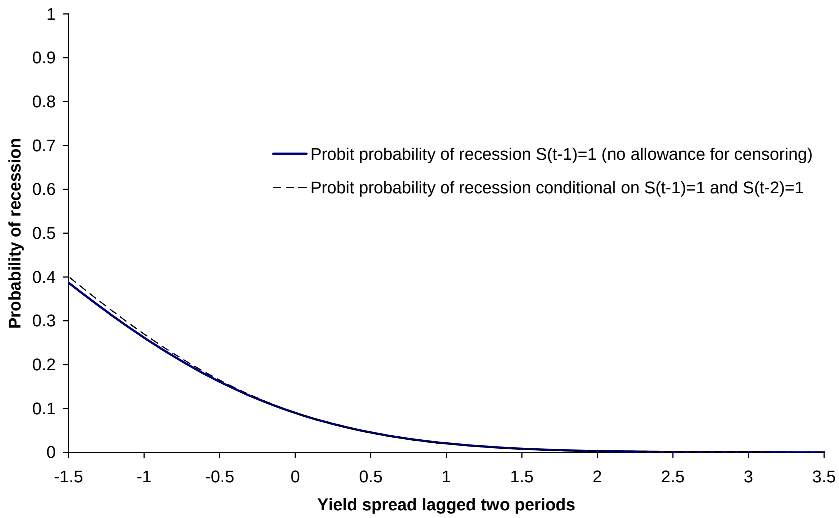


Figure 3 plots the probability of continuing in a recession. Here there is a substantial difference between the estimated probabilities. The DP model that makes no allowance for censoring over predicts the probability of staying in a recession when the yield spread is below 2.3 and under precicts the probability of a recession when the yield spread is above 2.3 per cent.

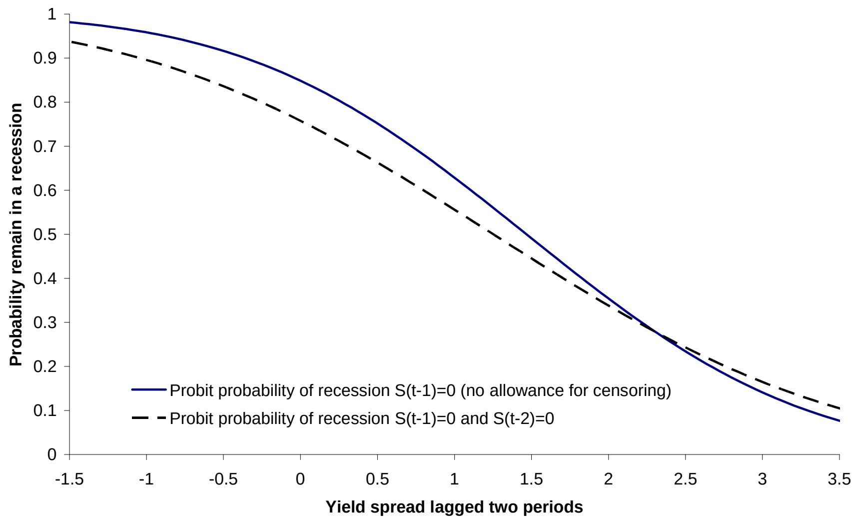
$E(S_t|x_t)$ can be calculated from the DP model where there is no allowance of censoring as follows[6]

$$E(S_t|x_t) = \Phi\left(\alpha_0 + \alpha_1 + \alpha_2 sp_{t-2}\right)\Pr\left(S_{t-1} = 1\right) \quad (24)$$
$$+\Phi\left(\alpha_0 + \alpha_2 sp_{t-2}\right)\Pr\left(S_{t-1} = 0\right) \quad (25)$$

Figure 4 plots the probability of recession $[1 - E(S_t|x_t)]$ estimated from the DP models that allow for censoring (equation 23) and do not allow for

[6]Here α_0 is the constant and α_1 the coefficient on S_{t-1} in the Probit model.

Figure 3: Probability of continuing in recession, DP models that vary in whether they allow for censoring



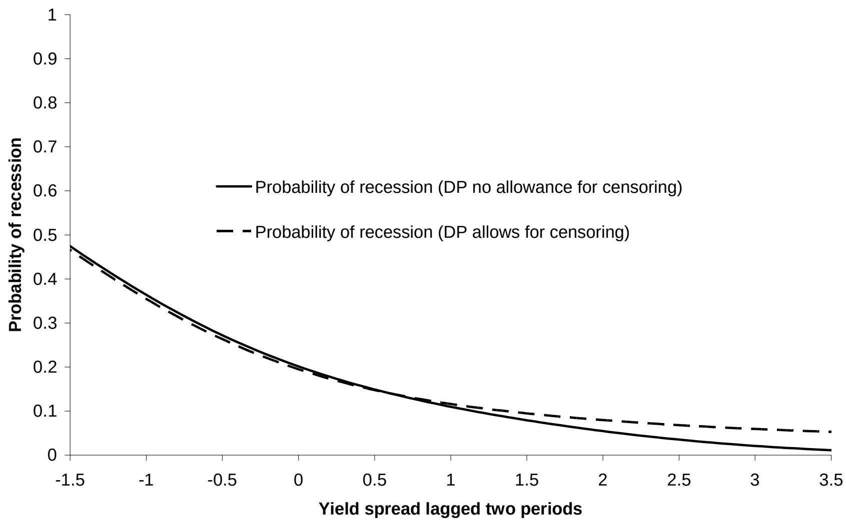
censoring (equation 24). The DP model that doesn't allow for censoring over predicts the probability of recession when the yield spread is below 0.53 and under precicts the probability of recession otherwise. The extent of the under prediction of recession is substantial. The evidence in these three figures suggests that failure to allow for censoring results in subtantial bias in estimates of the probability of recession.

Having established that making allowance for censoring is both theoretically and empirically important it is of interest to evaluate the extent to which the yield spread is useful in predicting the probability that an expansion which has lasted for two or more periods will be terminated and the probability of continuing in a contraction that has lasted for two or more periods. The former is the quantity $E(S_t|x_t, S_{t-1} = 1, S_{t-2} = 1)$ while the latter is $E(S_t|x_t, S_{t-1} = 0, S_{t-2} = 0)$.

The probability of leaving an expansion that has lasted for two or more quarters is shown in Figure 5. There is a substantial difference between the estimates obtained from the MP(2) model using non parametric methods (described above) and those from the DP (2) model that allows for censoring.[7]

[7]The confidence intervals for the non parametric methods were obtained using the

Figure 4: Probability of recession from DP models that vary in whether they allow for censoring



The DP (2) 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. While these differences are not statistically significant at the 5% level they are sufficiently large to be of economic interest to policy makers.

The probability of leaving a recession that has lasted for two quarters is plotted in Figure 6. Again the probabilities are from the MP(2) model estimated non parametrically and the DP(2) model allowing for censoring. There is a substantial, difference between the 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 MP(2) model suggests that there is no decrease in the probability of staying in recession from when the yield spread returns to positive until is reaches 2.5 per cent. The DP(2) model, in contrast 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

methods described above.

Figure 5: Probability of terminating an expansion that has lasted for two quarters

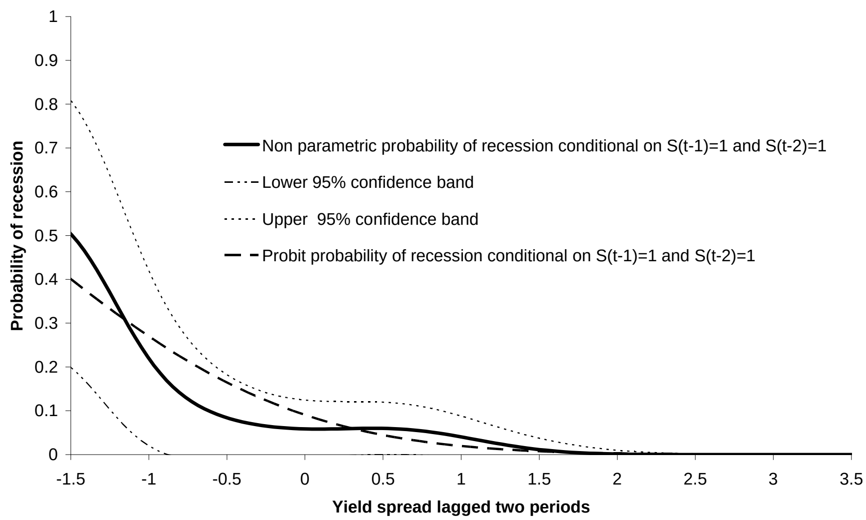
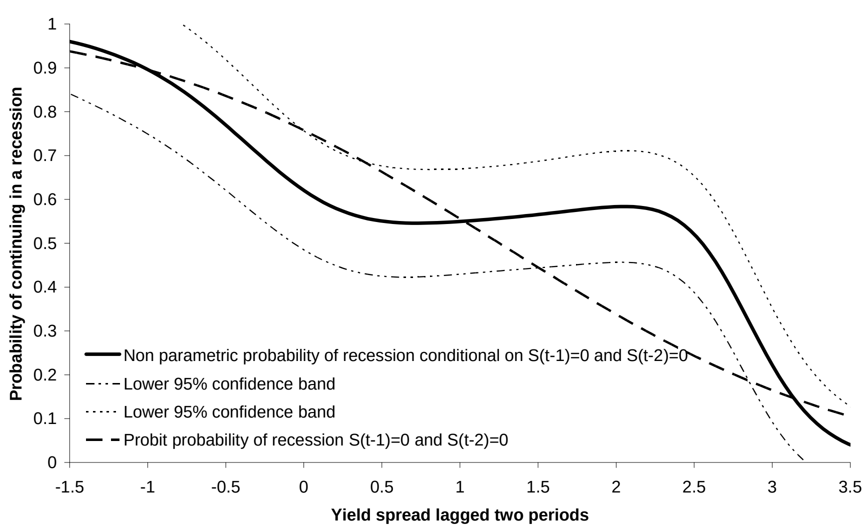


Figure 6: Probability of continuing in a recession



intervals as there are only 10 per cent of cases where the economy is in contraction for two or more periods. But even after 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. One has to allow for the fact that they are essentially Markov processes when engaging in a broad range of estimation and inference methods. 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 procedure that does not allow for the Markov process nature of the binary states and forces a functional form upon the data that is not appropriate. We have shown that these difference are economically and statistically significant.

Appendix

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

$$\{\Delta y_{t+1}, \Delta y_t, \Delta y_{t-1}, \Delta y_{t-2}, ...\} = \{-, -, -., +, ...\} \tag{26}$$

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 $t-1$. 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.

$$\begin{bmatrix} t+1 & * & * & * & * & * & * & \cdots \\ t & + & - & + & - & + & - & \cdots \\ t-1 & + & + & - & + & - & + & \cdots \\ t-2 & * & + & + & - & + & - & \cdots \\ t-3 & & * & + & + & - & + & \cdots \\ t-4 & & & * & + & + & - & \cdots \\ t-5 & & & & * & + & + & \cdots \\ t-6 & & & & & * & + & \cdots \\ \vdots & & & & & & * & \ddots \end{bmatrix} \tag{27}$$

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 $t-1$
- $[-]_t$ represents the case where $\Delta y_t < 0$,

the sets E_t and F_t can be enumerated as

$$E_t = \left\{ [++]_t \,; [+-]_t \, [++]_{t-2} \,; [+-]_t^2 \, [++]_{t-4} \,; ...; [+-]_t^j \, [++]_{t-2j} \,; .... \right\} \quad (28)$$

$$F_t = \left\{ \begin{array}{c} [-]_t \, [++]_{t-1} \,; [-]_t \, [+-]_{t-1} \, [++]_{t-3} \,; \\ [-]_t \, [+-]_{t-1}^2 \, [++]_{t-5} \,; ...; [-]_t \, [+-]_{t-1}^j \, [++]_{t-2j-1} \,; .... \end{array} \right\}. \quad (29)$$

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 (30), (31), (32) and (34).[8])

$$\Pr(E_t) = \sum_{j=0}^{\infty} \Pr\left([+-]_t^j \, [++]_{t-2j} \right) \quad (30)$$

and

[8] To simplify notation we have omitted the conditioning on $\Im_{t+1}$ in equations (30), (31), (32) and (34).

$$\Pr(F_t) = \sum_{j=0}^{\infty} \Pr\left([-]_t [+-]_{t-1}^{j} [++]_{t-2j-1} \right). \tag{31}$$

By definition

$$\Pr(S_{t-1} = 1) = \Pr(E_t) + \Pr(F_t). \tag{32}$$

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

$$G_{t+1} = \left\{ \begin{array}{c} [--]_{t+1} [++]_{t-1} ; [--]_{t+1} [+-]_{t-1} [++]_{t-3} ; [--]_{t+1} [+-]_{t-1}^{2} [++]_{t-5} ; \ldots \\ \ldots ; [--]_{t+1} [+-]_{t-1}^{j} [++]_{t-2j-1} ; \ldots . \end{array} \right\} \tag{33}$$

Then, since S_t is a stationary process,

$$p_{10} = \frac{\Pr(S_t = 0, S_{t-1} = 1)}{\Pr(S_{t-1} = 1)} = \frac{\Pr(G_{t+1})}{\Pr(E_t) + \Pr(F_t)}. \tag{34}$$

If $\Pr(S_t = 1, S_{t-1} = 0)$ is constant, which essentially requires Δy_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 μ_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

$$\begin{aligned} \Pr(E) &= \sum_{j=0}^{\infty} (1-\psi)^2 \left[\psi(1-\psi)\right]^j \\ &= \frac{(1-\psi)^2}{1-\psi(1-\psi)} \end{aligned} \tag{35}$$

$$\begin{aligned} \Pr(F) &= \sum_{j=0}^{\infty} \psi (1-\psi)^2 \left[\psi (1-\psi)\right]^j \\ &= \frac{\psi(1-\psi)^2}{1-\psi(1-\psi)} \end{aligned} \tag{36}$$

$$\begin{aligned}
\Pr(G) &= \sum_{j=0}^{\infty} \psi^2 (1-\psi)^2 [\psi (1-\psi)]^j \\
&= \frac{\psi^2 (1-\psi)^2}{1-\psi(1-\psi)} \qquad (37)
\end{aligned}$$

and so

$$\begin{aligned}
p_{10} &= \frac{\psi^2}{(1+\psi)} \qquad (38) \\
p_{11} &= \frac{1+\psi-\psi^2}{(1+\psi)} \qquad (39)
\end{aligned}$$

$$\begin{aligned}
p_{01} &= \frac{(1-\psi)^2}{2-\psi} \qquad (40) \\
p_{00} &= \frac{1+\psi-\psi^2}{2-\psi} \qquad (41)
\end{aligned}$$

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

$$\Delta y_t = a + b x_t + u_t \qquad (42)$$

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 $\psi_t = \Phi(-a - b x_t)$, applying the enumeration method results in

$$\begin{aligned}
\Pr(E_{t+1} | \Im_{t+1}) &= \sum_{j=0}^{\infty} \Pr\left([+-]_{t+1}^{j} [++]_{t+1-2j} \right) \\
&= \left(1-\psi_{t+1}\right)(1-\psi_t) \qquad (43) \\
&\quad + \sum_{j=1}^{\infty} \left\{ \left[\prod_{i=0}^{j-1} \left(1-\psi_{t+1-i}\right) \psi_{t-i} \right] \left(1-\psi_{t-2j+1}\right) \left(1-\psi_{t-2j}\right) \right\}
\end{aligned}$$

and

$$
\begin{aligned}
\Pr\left(F_{t+1}|\Im_{t+1}\right) &= \sum_{j=0}^{\infty}\Pr\left([-]_{t+1}[+-]_t^j[++]_{t-2j}\right) \\
&= \psi_{t+1}\left(1-\psi_t\right)\left(1-\psi_{t-1}\right)+ \qquad (44)\\
&\quad \psi_{t+1}\sum_{j=1}^{\infty}\left\{\left[\prod_{i=0}^{j-1}\left(1-\psi_{t-i}\right)\psi_{t-i-1}\right]\left(1-\psi_{t-2j}\right)\left(1-\psi_{t-2j-1}\right)\right\}.
\end{aligned}
$$

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

$$
P_t = \Pr\left(E_{t+1}|\Im_{t+1}\right) + \Pr\left(F_{t+1}|\Im_{t+1}\right) \qquad (45)
$$

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 depend upon $\alpha + x_t\beta$ alone. Only if the dating rule had been the “calculus” one would $\Pr(S_t = 1|\Im_{t+1}) = (1-\psi_t)$ 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.