CENTRE FOR APPLIED MACROECONOMIC ANALYSIS

The Australian National University

THE AUSTRALIAN NATIONAL UNIVERSITY

CAMA Working Paper Series September, 2010

# APPLYING SHAPE AND PHASE RESTRICTIONS IN GENERALIZED DYNAMIC CATEGORICAL MODELS OF THE BUSINESS CYCLE

**Don Harding**
School of Economics and Finance
La Trobe University

CAMA Working Paper 25/2010
http://cama.anu.edu.au

# Applying shape and phase restrictions in generalized dynamic categorical models of the business cycle

Don Harding[1,2]

July 26, 2010

[1]School of Economics and Finance, La Trobe University, Melbourne, Australia.
[2]Support from a La Trobe University Faculty of Law and Management research grant is gratefully acknowledged. Helpful comments were provided by Adrian Pagan, John Geweke and participants at the $15^{th}$ Australasian macroeconomics workshop. Jack Mullumby provided useful research assistance. All errors are my responsibility.

**Abstract**

To match the NBER business cycle features it is necessary to employ Generalised dynamic categorical (GDC) models that impose certain phase restrictions and permit multiple indexes. Theory suggests additional shape restrictions in the form of monotonicity and boundedness of certain transition probabilities. Maximum likelihood and constraint weighted bootstrap estimators are developed to impose these restrictions. In the application these estimators generate improved estimates of how the probability of recession varies with the yield spread.

Key Words: Generalized dynamic categorical model, Business cycle; binary variable, Markov process, probit model, yield curve

JEL Code C22, C53, E32, E37

# 1 Introduction

The business cycle is one of many cases in macroeconomics and finance where single index dynamic discrete choice (DDC) models are invalid and multiple index DDC models are required. Harding and Pagan (2009) provide a detailed discussion of this issue. They introduce a second order generalized dynamic categorical (GDC) model as a parsimonious framework to approximate the DGP of NBER business cycle states. The GDC model is set out in section 2 and the form of the transition probabilities derived in section 2.1 where it is shown that boundedness and monotonicity, in the forcing variables, is exhibited by this class of models.

The non parametric estimation method developed by Harding and Pagan (2009) imposes phase restrictions necessary to approximate the NBER business cycles, respects boundedness of transition probabilities but does not impose monotonicity.[1] This paper makes four contributions that develop and extend the framework they develop.

The first contribution is to develop parametric models that simultaneously impose phase restrictions and shape features. Two particular parametric versions are developed using the Normal and Logistic distributions. These are named GDC Probit and GDC Logit respectively. Maximum likelihood estimation (MLE) of these models is discussed in section 3.

The second contribution of this paper, in section 4, is to develop non parametric methods that simultaneously impose the GDC phase and shape restrictions. The particular method applied is the constraint weighted boot-

[1] Henceforth I refer to monotonicity in the forcing variables and boundedness of the transition probabilities as shape features of GDC models.

strap of Hall and Huang (2001). The intuition of this method is that it "involves tilting the empirical distribution to the least possible amount, subject to the constraint being enforced".

The non-parametric method requires choice of the window width in the kernel estimator and a parameter in the distance function. Procedures are developed to determine both of these parameters as functions of the data.

The third contribution is made in section 5 where I evaluate the parametric and non parametric procedures in an application that uses the yield spread to predict the state of the business cycle. Here it is shown that the GDC Probit produces a large improvement in fit over the static Probit developed by Estrella and Mishkin (1998). It is also shown that the non parametric estimator with phase and shape restrictions delivers further improvement in fit over the GDC Probit. The shape restricted non parametric estimator also reveals an important feature of United States business cycles that is relevant for the conduct of monetary policy.

The fourth contribution of the paper, made in section 6, is to show that one needs to be very careful in translating statements about improvements in fit to statements about policy relevant improvements in prediction about peaks and troughs of the business cycle.

Conclusions are in section 7.

# 2 A second order Generalized Dynamic Choice model

The variable of interest is $S_t$, a binary variable constructed using NBER procedures, it has the properties that,

- $S_t = 1$ when the economy is in expansion;
- $S_t = 0$ when the economy is in recession.
- completed phases have a minimum duration of at least two periods.

The GDC model of order 2, introduced in Harding and Pagan (2009), to model such a variable is[2]

$$\begin{aligned} Pr\left(S_t = 1 | S_{t-1}, S_{t-1}, \mathbf{x}_t\right) &= \gamma_{00}\left(\mathbf{x}_t\right) S_{t-1}^{00} + \gamma_{01}\left(\mathbf{x}_t\right) S_{t-1}^{01} \\ &\quad + \gamma_{10}\left(\mathbf{x}_t\right) S_{t-1}^{10} + \gamma_{11}\left(\mathbf{x}_t\right) S_{t-1}^{11} \end{aligned} \tag{1}$$

where $\mathbf{x}_t$ is a vector of conditioning variables. As discussed in Harding and Pagan (2009) the NBER method of constructing the $S_t$ imposes the restrictions that

$$\gamma_{10}\left(\mathbf{x}_t\right) = 1 \qquad \text{and} \qquad \gamma_{01}\left(\mathbf{x}_t\right) = 0 \tag{2}$$

[2] For later use in estimating these models it is useful to define the following variables and sets where $i, j = \{0, 1\}$:

- $S_{t-1}^{ij} = 1$ if $S_{t-1} = i$ and $S_{t-2} = j$. Otherwise $S_{t-1}^{ij} = 0$.
- $n_{ij}$ is number of cases where $S_{t-1} = i$ and $S_{t-2} = j$.
- $I_{ij}$ is subset of $\{1, ..., T\}$ where $S_{t-1} = i$ and $S_{t-2} = j$

The $\gamma_{ij}(\mathbf{x}_t)$ also have the property of boundedness, ie $0 \leq \gamma_{ij}(\mathbf{x}_t) \leq 1$.

For later use it is convenient to write the static Probit model as

$$\Pr\left(S_{t=1} = 1|\mathbf{x}_t\right) = \int_{-\mathbf{x}_t\boldsymbol{\beta}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv \equiv 1 - \Phi\left(-\mathbf{x}_t\boldsymbol{\beta}\right) \tag{3}$$

where $\Phi(.)$ is the standard normal cumulative distribution function.

## 2.1 Obtaining the form of the transition probabilities from first principles

It is useful to obtain the 2nd order GDC model from first principles so as to establish where the properties of that model come from.

The data generating process for $S_t$ is assumed to be such that the economy stays in an expansion that has lasted at least two quarters provided the latent variable $\zeta_{11t}$ is positive $(\zeta_{11t} > 0)$. The economy shifts from a recession that has lasted at least two periods to an expansion if the latent variable $\zeta_{00t}$ is positive $(\zeta_{00t} > 0)$. Reflecting the NBER method of constructing $S_t$, the economy stays in recession if the recession has lasted only one period and stays in expansion if the expansion has lasted only one period. Thus the binary variable $S_t$ representing the state of the business cycle evolves according to.

$$S_t = S_{t-1}^{11}\mathbf{1}\left(\zeta_{11t} > 0\right) + S_{t-1}^{00}\mathbf{1}\left(\zeta_{00t} > 0\right) + S_{t-1}^{10} \tag{4}$$

The two latent variables $(\zeta_{1t}, \zeta_{1t})$ have the following data generating processes (DGPs)

$$\zeta_{11t} = \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_{11t} \tag{5}$$

$$\zeta_{00t} = \mathbf{x}_t \boldsymbol{\theta} + \varepsilon_{00t} \tag{6}$$

The shocks $\varepsilon_{11t}$ and $\varepsilon_{00t}$ are mutually independent, and also are independent of $\mathbf{x}_t$, have unit variance and densities $f(\varepsilon_{1t})$ and $g(\varepsilon_{2t})$ respectively. Moreover, the effects of the forcing variables on the latent variables are monotonic since $\frac{\partial \zeta_{11t}}{\partial x_{jt}} = \beta_j$ and $\frac{\partial \zeta_{00t}}{\partial x_{jt}} = \theta_j$. So if, for example, the forcing variable is the yield spread then an increase in the yield spread does not make it less likely that the economy will be in expansion next period.

Then, the probability of remaining in an expansion that has lasted at least two periods is

$$\begin{aligned} \Pr(S_t = 1 | S_{t-1} = 1, S_{t-2} = 1, \mathbf{x}_t) &= E(S_t | S_{t-1} = 1, S_{t-2} = 1, \mathbf{x}_t) \\ &= E\{\mathbf{1}(\zeta_{11t} > 0 | \mathbf{x}_t)\} \\ &= \Pr(\zeta_{11t} > 0 | \mathbf{x}_t) \\ &= \int_{-\mathbf{x}_t\boldsymbol{\beta}}^{\infty} f(v)\, dv \equiv 1 - F(-\mathbf{x}_t\boldsymbol{\beta}) \end{aligned} \tag{7}$$

Similarly the probability of exiting a contraction that has lasted at least two periods is

$$\begin{aligned}
\Pr\left(S_t = 1 | S_{t-1} = 0, S_{t-2} = 0, \mathbf{x}_t\right) &= E\left(S_t | S_{t-1} = 0, S_{t-2} = 0, \mathbf{x}_t\right) \qquad (8) \\
&= E\left\{\mathbf{1}\left(\zeta_{00t} > 0 | \mathbf{x}_t\right)\right\} \\
&= \Pr\left(\zeta_{00t} > 0 | \mathbf{x}_t\right) \\
&= \int_{-\mathbf{x}_t \boldsymbol{\theta}}^{\infty} g\left(v\right) dv \equiv 1 - G\left(-\mathbf{x}_t \boldsymbol{\theta}\right)
\end{aligned}$$

where $F(.)$ and $G(.)$ are cumulative distribution functions corresponding to the densities $f(.)$ and $g(.)$ respectively.

The transition probabilities are monotonic in the elements of $\mathbf{x}_t'$. This can be seen by noting that since $\gamma_{11}(\mathbf{x}_t) = 1 - F(-\mathbf{x}_t\boldsymbol{\beta})$ and $\gamma_{00}(\mathbf{x}_t) = 1 - G(-\mathbf{x}_t\boldsymbol{\theta})$ it follows that

$$\frac{\partial \gamma_{11}(\mathbf{x}_t)}{\partial x_{jt}} = f(-\mathbf{x}_t\boldsymbol{\beta})\,\beta_j \qquad (9)$$

and

$$\frac{\partial \gamma_{00}(\mathbf{x}_t)}{\partial x_{jt}} = g(-\mathbf{x}_t\boldsymbol{\theta})\,\theta_j \qquad (10)$$

Now since $f(.)$ and $g(.)$ are densities they are non negative and thus $\gamma_{11}(\mathbf{x}_t)$ and $\gamma_{00}(\mathbf{x}_t)$ are weakly monotonic. Moreover, the direction(s) of the monotonicity are determined by the signs of the coefficients $\beta_j$ and $\theta_j$ respectively. The issue of whether the monotonicity is weak or strict depends firstly, on whether the densities $f(.)$ and $g(.)$ are strictly positive on their support and, secondly, on whether $\beta_j \neq 0$ and $\theta_j \neq 0$.

## 2.2 GDC Probit and Logit

The double index GDC-Probit model, used in the application, is the special case of (7) and (8) where $\varepsilon_{11t}$ and $\varepsilon_{00t}$ have independent standard normal distributions so that, $f(v) = g(v) = \phi(v) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$ and $F(z) = G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(v)\, dv$ the cumulative distribution function of the standard normal distribution. The GDC Probit model is written as

$$\begin{aligned} Pr\left(S_t = 1 | S_{t-1}, S_{t-1}, \mathbf{x}_t\right) &= S_{t-1}^{10} + \left[1 - \Phi\left(-\mathbf{x}_t' \boldsymbol{\theta}\right)\right] S_{t-1}^{00} \\ &\quad + \left[1 - \Phi\left(-\mathbf{x}_t' \boldsymbol{\beta}\right)\right] S_{t-1}^{11} \end{aligned} \tag{11}$$

The single index GDC Probit model involves the restriction that $\boldsymbol{\theta} = \boldsymbol{\beta}$ in (11). Importantly, the single index GDC Probit differs from the static Probit because the former does not include the term $S_{t-1}^{10}$ which is required if the model is to be consistent with the NBER method for constructing the $S_t$. As discussed, in Harding and Pagan (2009) this difference means that the static Probit model can never match the features of the NBER business cycle.

Other distribution functions such a the logit might be considered so that

$$F(z) = G(z) = \left[1 + \exp\left(-\frac{\pi}{\sqrt{3}} z\right)\right]^{-1} \tag{12}$$

yielding a model that might be designated as GDC-Logit.

# 3 Maximum likelihood estimation

The conditional likelihood functions are

$$L_{11} = \prod_{t \in I_{11}} \left[1 - F\left(-\mathbf{x}_t \boldsymbol{\beta}\right)\right]^{S_t^{11}} \left[F\left(-\mathbf{x}_t \boldsymbol{\beta}\right)\right]^{1 - S_t^{11}} \tag{13}$$

$$L_{00} = \prod_{t \in I_{00}} \left[1 - G\left(-\mathbf{x}_t \boldsymbol{\theta}\right)\right]^{S_t^{00}} \left[G\left(-\mathbf{x}_t \boldsymbol{\theta}\right)\right]^{1 - S_t^{00}} \tag{14}$$

To implement the maximum likelihood procedures one selects parametric densities $f(.)$ and $g(.)$. Estimation proceeds by choosing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to maximize the log of the likelihoods. In the case of the GDC Probit (11) the ML estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are:

$$\widehat{\boldsymbol{\beta}} = \arg\max \sum_{t \in I_{11}} \left\{ S_t^{11} \ln\left[1 - \Phi\left(-\mathbf{x}_t' \boldsymbol{\beta}\right)\right] + \left(1 - S_t^{11}\right) \ln \Phi\left(-\mathbf{x}_t' \boldsymbol{\beta}\right) \right\} \tag{15}$$

$$\widehat{\boldsymbol{\theta}} = \arg\max \sum_{t \in I_{00}} \left\{ S_t^{00} \ln\left[1 - \Phi\left(-\mathbf{x}_t' \boldsymbol{\theta}\right)\right] + \left(1 - S_t^{00}\right) \ln \Phi\left(-\mathbf{x}_t' \boldsymbol{\theta}\right) \right\} \tag{16}$$

# 4 Non parametric estimation

Non parametric methods do not automatically impose the monotonicity that was shown above to be central to the GDC class of models.

Henderson and Parmenter (2009) survey a range of methods for imposing constraints on non parametric estimators of conditional means. Some of these methods introduce unattractive features such as reduced smoothness of the estimator. For example, isotonic regression (Friedman and Tibshirani (1984)) is unattractive because they introduce jump discontinuities in the estimator.

Hall and Huang's (2001) method, in contrast, is designed to

- produce a curve that satisfies the constraints but exhibits the same smoothness (ie the same number of derivatives exist and are continuous) as for its unconstrained counterpart;
- be applicable to general kernel methods;
- mainly modify the unconstrained estimator in the regions where the constraints are binding; and
- require little additional computational effort over the unconstrained estimator.

## 4.1 Hall and Huang's method

To implement Hall and Huang's (2001) method for the problem at hand observe that (1) implies that[3]

$$\gamma_{11}(x_t) = E(S_t | S_{t-1} = 1, S_{t-2} = 1, x_t) \tag{17}$$

and

[3] Here $x_t$ is assumed to be a scalar. The approach readily generalizes to the case where there is a vector of forcing variables.

$$\gamma_{00}(x_t) = E(S_t | S_{t-1} = 0, S_{t-2} = 0, x) \tag{18}$$

For a wide range of non parametric methods, estimators of $\gamma_{11}(x)$ and $\gamma_{00}(x)$ can be written as

$$\widehat{\gamma_{00}}(x) = \frac{1}{n_{00}} \sum_{i \in I_{00}} A_i^{00}(x) S_i^{00} \tag{19}$$

$$\widehat{\gamma_{11}}(x) = \frac{1}{n_{11}} \sum_{i \in I_{11}} A_i^{11}(x) S_i^{11} \tag{20}$$

where the $A_i^{11}$ and $A_i^{00}$ are weighting functions. For example, if a local constant kernel method is used then let $\psi_i^{jj}(x) = \frac{x - x_i^{jj}}{h_j}$ and $A_i^{jj}(x)$ is defined as[4]

$$A_i^{jj}(x) = \frac{K\left(\psi_i^{jj}(x)\right)}{\sum K\left(\psi_i^{jj}(x)\right)} \tag{21}$$

Hall and Huang (2001) suggest a very convenient method for imposing a wide range of constraints on the estimators of $\gamma_{00}(x)$ and $\gamma_{11}(x)$. For imposing monotonicity and boundedness this involves the following steps.

First estimate $\widehat{\gamma_{00}}(x)$, $\widehat{\gamma_{11}}(x)$ saving $A_i^{11}(x)$ and $A_i^{00}(x)$. Second, check whether the boundedness and monotonicity constraints are violated. If they are not violated then $\widehat{\gamma_{00}}(x)$, $\widehat{\gamma_{11}}(x)$ are the final estimators.

Otherwise introduce observation specific weights $p_i^{11}$ and $p_i^{00}$ with the properties that $\sum_{i \in I_0} p_i^{00} = \sum_{i \in I_1} p_i^{11} = 1$. Define new estimators $\widetilde{\gamma_{00}}(x)$ and

[4]The window width used to compute $E(S_t | x_t, S_{t-1} = j, S_{t-2} = j)$ is $c_j \sigma_x n_{jj}^{\frac{1}{5}}$ where $n_{jj}$ is the number of cases where $(S_{t-1} = j, S_{t-2} = j)$, $\sigma_x$ is the standard deviation of the $x's$ and $c_j$ is a constant.

$\widetilde{\gamma_{11}}(x)$ as follows

$$\widetilde{\gamma_{00}}(x) = \sum_{i \in I_{00}} p_i^{00} A_i^{00}(x) S_i^{00} \tag{22}$$

$$\widetilde{\gamma_{11}}(x) = \sum_{i \in I_{11}} p_i^{11} A_i^{11}(x) S_i^{11} \tag{23}$$

The idea in the Hall and Huang (2001) method, as applied to the problem here, is to chose the weights so as to minimize the distances between $p_i^{00}$ and $\frac{1}{n_{00}}$ and $p_i^{11}$ and $\frac{1}{n_{11}}$ respectively. The distance metric $D_\rho(p)$ introduced by Cressie and Read (1984) proves to be useful for imposing the monotonicity constraint[5]

$$\begin{aligned} D_\rho\left(\boldsymbol{p}^{jj}\right) &= \frac{1}{\rho(1-\rho)}\left[n - \sum_{i=1}^{n}\left(np_i^{jj}\right)^\rho\right], \ |\rho| < \infty, \ \rho \neq 0, \ \rho \neq 1 \quad (24) \\ &= -\sum_{i=1}^{n} \ln\left(np_i^{jj}\right), \qquad \rho = 0 \\ &= \sum_{i=1}^{n} p_i^{jj} \ln\left(np_i^{jj}\right), \qquad \rho = 1 \end{aligned}$$

If we sought to minimize $D_\rho(\mathbf{p}^{00})$ and $D_\rho(\mathbf{p}^{11})$ with respect to $\mathbf{p}^{00}$ and $\mathbf{p}^{11}$ without imposing any constraints then the solution would be $p_i^{00} = \frac{1}{n_{00}}$ and $p_i^{11} = \frac{1}{n_{11}}$ respectively. The relevant constraints when estimating $\widetilde{\gamma_{jj}}(\mathbf{x}_t)$ are:

- Weights sum to one

$$\sum_{i \in I_{jj}} p_i^{jj} = 1 \tag{25}$$

[5]Montotonicity can be imposed with weights that are positive thereby making Cressie and Read's (1984) distance metric appropriate. See Racine, Parmenter and Du (2009) for a discussion of the distance metric that is appropriate with more general shape restrictions that may require negative weights.

- Monotonicity[6]

$$\frac{\partial \widetilde{\gamma_{jj}}(x)}{\partial x_k} = \sum_{i \in I_{00}} p_i^{00} \frac{\partial A_i^{jj}(x)}{\partial x_k} S_i^{jj} > 0 \qquad \forall\, k \tag{26}$$

- Boundedness

$$0 \leq \widetilde{\gamma_{jj}}(x) \tag{27}$$

$$\widetilde{\gamma_{jj}}(x) \leq 1 \tag{28}$$

The constrained estimator is obtained by choosing $\boldsymbol{p}^{jj}$ to minimize (24) subject to constraints (25) to (28). This constrained minimization is easily achieved using a procedure such as `fmincon` in Matlab.

Using the same arguments as in Harding and Pagan (2009) the following central limit holds

$$\sqrt{n_{jj}h_j}\left(\widetilde{\gamma_{jj}}(\mathbf{x}) - \gamma_{jj}(\mathbf{x})\right) \rightarrow \mathcal{N}\left(0, \frac{\int K(\psi)^2\, d\psi}{\pi_{jj}(x)} \gamma_{jj}(\mathbf{x})\left[1 - \gamma_{jj}(\mathbf{x})\right]\right) \tag{29}$$

Where $\pi_{jj}(x)$ is the density of $x$ on the relevant sub sample.

## 4.2 Choosing $h$ and $\rho$

There are two data driven techniques available to choose the smoothing parameter $h$ and the parameter $\rho$ in the distance function.

The first of these is least squares cross validation. It involves choosing $h$

[6] Here monotonicity is imposed as a local property via the slope one could also require that $\widetilde{\gamma_{jj}}(x) \geq \widetilde{\gamma_{jj}}(x^*)$ for $x \geq x^*$.

and $\rho$ to minimise the objective function

$$CV(h,\rho) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{m}_{-i}(h,\rho)\right)^2 \tag{30}$$

Where $\widehat{m}_{-i}(h,\rho)$ is the leave one out non parametric estimate of $E(S_i|S_{i-1}, S_{i-2}, sp_{i-2})$. Least squares cross validation is expensive in terms of computer time because the leave one out estimator must be computed $n$ times. This approach is even more expensive here because the optimal observation specific weights must also be calculated at each iteration.

An attractive alternative discussed in Li and Racine (2009, p72) is provided by the improved Aiaike information criterion $AIC_c$ criterion developed by Hurvich, Simonoff and Tsai (1998). Here the objective functions are defined as

$$AIC_c(h,\rho) = \ln\widehat{\sigma}^2 + \frac{1 + tr(H^r)/n}{1 - \{tr(H^r) + 2\}/n} \qquad r = 0,1 \tag{31}$$

Where $H^r$ is an $n \times n$ weighting matrix and[7]

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{m}_{-i}(h,\rho)\right)^2 = Y'(I - H^r)'(I - H^r)Y/n \tag{32}$$

Li and Racine (2009) report studies that show that choosing $h$ that minimizes $AIC_c$ yields a window width that performs well compared to plug-in methods, and generalized cross validation methods. Li and Racine (2004) show using

[7]Letting $H^r_{i,j}$ denote the $(i,j)^{th}$ element of $H^r$, the weighting matrices $H^r$ are related to those used in estimation as follows:

$$H^r_{i,j} = p_i^{rr}A_i^{rr}(x_j).$$

simulation methods that $AIC_c$ outperforms least squares cross validation in small samples and that in large samples the two methods yield similar window widths.

An alternative version of the AIC criterion that is suitable for working with discrete choice models involves replacing $\ln \widehat{\sigma}^2$ in (31) with minus the log probability score so that the criterion to be minimized is

$$AIC_c^* (h, \rho) = -LPS (h, \rho) + \frac{1 + tr (H^r) / n}{1 - \{tr (H^r) + 2\} / n} \tag{33}$$

Where

$$LPS (h, \rho) = \sum_{i=1}^{n} \ln \{S_t \times \widehat{m}_t (h, \rho) + (1 - S_t) \times (1 - \widehat{m}_t (h, \rho))\}$$

and $\widehat{m}_t (h, \rho)$ is the non parametric estimator of $E (S_t | S_{t-1}, S_{t-2}, sp_{t-2})$.

# 5 Application

This application explores the value of the parametric and non parametric methods discussed above in studying the value of the yield spread ($sp_t$) in predicting the state of the business cycle. To achieve comparability with the literature, these questions are addressed using the same sample of data as in Estrella and Mishkin (1998) and Harding and Pagan (2009).[8] To maintain comparability with these earlier papers I maintain Estrella and Mishkin's (1998) specification that $x_t = sp_{t-2}$.

[8] Estrella and Mishkin (1998) use the spread between the 10 year bond and 3 month Treasury Bill.

Table 1: Summary statistics: United States NBER business cycle states, 1959.Q3 to 1995.Q1

| | $S_t$ | $S_t\|S^{11}_{t-1}=1$ | $S_t\|S^{00}_{t-1}=1$ | $S_t\|S^{10}_{t-1}=1$ | $S_t\|S^{01}_{t-1}=1$ |
|---|---|---|---|---|---|
| Mean | 0.85 | 0.95 | 0.40 | 1 | 0 |
| Standard Deviation | 0.36 | 0.22 | 0.51 | 0 | 0 |

Summary statistics for the NBER business cycle date are presented in Table 1. The unconditional mean of $S_t$ in the second column and the two conditional means in the third and fourth columns will be used later in the paper as a reference point against which the various models are evaluated. The fifth and sixth columns of Table 1 confirm that the data does have the NBER feature identified in Harding and Pagan (2009) that the second quarter of each phase can be predicted with certainty. They show that failure to allow for this feature in dynamic discrete choice models results in a serious violation of the assumptions required for maximum likelihood estimation to be valid.

Inspection of the means in Table 1 makes it apparent that the static Probit model (3) cannot be a good fit to this data since that model implies that all of the means in Table 1 should be equal.

## 5.1 Maximum likelihood

The static Probit (3) estimated by Estrella and Mishkin (1998) is reported in column 1 of table 2. It can be compared with the GDC model in column 3 where the parameters are constrained so that $\beta_0 = \theta_0$ and $\beta_1 = \theta_1$. This latter model produces a substantial improvement over the static Probit in terms of the log likelihood and the standard error of the regression.[9] Since the two models have the same number of parameters we would always prefer the one with the better fit. In this case the better fit comes solely from recognizing that, as discussed above and in Harding and Pagan (2009), the method of construction employed by the NBER imposes a particular structure on the transition probabilities for the NBER states.

The results for double index GDC Probit model is shown in column 5 of table 2. It produces a large improvement in the log likelihood. The likelihood ratio statistic for the null hypothesis that there is a single index is 26.12. Since this is distributed $\chi^2(2)$ the 1% critical value is 9.21 and we can reject the null hypothesis in favour of the alternative that there is a double index.

Also shown in columns 2, 4 and 6 of table 2 are the results for the logit model which fits a little worse than the Probit in each case.

The GDC single index Probit conditions on the economy being in the same state in periods $t-1$ and $t-2$ ($S_{t-1} = S_{t-2}$). This is informative because the NBER method of constructing $S_t$ imposes the restrictions that $E(S_t|S_{t-1} = 1, S_{t-2} = 0) = 1$ and $E(S_t|S_{t-1} = 0, S_{t-2} = 1) = 0$. Panel A of Figure 1 compares the probability being in expansion for the static Probit model and the single index GDC model, this Figure clearly demonstrates the effect of conditioning on ($S_{t-1} = S_{t-2}$). The dashed lines are confidence bands with 95% coverage probability.[10]

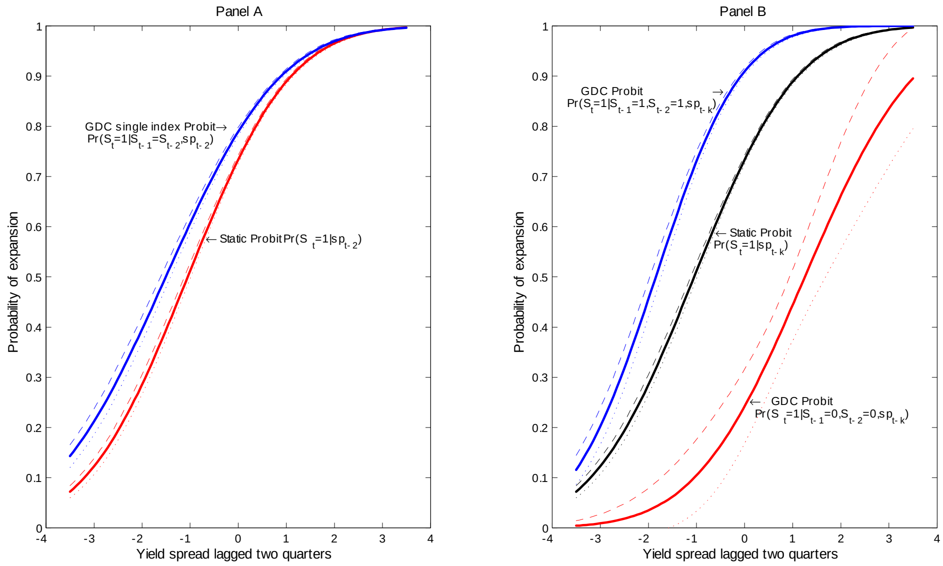
[9] In table 2 the various loglikehods on the subsample and full sample are defined as $l^{00} = \log L^{00}$ $l^{11} = \log L^{11}$ and $l = l^{00} + l^{11}$.

[10] The standard errors used in constructing the confidence intervals are obtained using the delta method.

Table 2: Results from static and GDC-Probit models

| | **Static Probit** | **Static Logit** | **GDC: Single index Probit** | **GDC: Single index Logit** | **GDC: Double index Probit** | **GDC: Double index Logit** |
|---|---|---|---|---|---|---|
| | **Col 1** | **Col 2** | **Col 3** | **Col 4** | **Col 5** | **Col 6** |
| $\beta_0$ | 0.624 (0.013) | 0.588 (0.013) | 0.804 (0.015) | 0.760 (0.015) | 1.338 (0.023) | 1.348 (0.026) |
| $\beta_1$ | 0.596 (0.011) | 0.588 (0.011) | 0.534 (0.012) | 0.540 (0.013) | 0.724 (0.021) | 0.763 (0.022) |
| $\theta_0$ | $na$ | $na$ | $=\beta_0$ | $na$ | $-0.699$ (0.121) | $-0.613$ (0.112) |
| $\theta_1$ | $na$ | $na$ | $=\beta_1$ | $na$ | 0.558 (0.099) | 0.484 (0.091) |
| $l_{00}$ | $na$ | $na$ | $na$ | $na$ | $-8.79$ | $-8.84$ |
| $l_{11}$ | $na$ | $na$ | $na$ | $na$ | $-15.71$ | $-15.82$ |
| $l$ | $-45.89$ | $-46.05$ | $-37.57$ | $-37.70$ | $-24.51$ | $-24.66$ |
| $SE_{00}$ | $na$ | $na$ | $na$ | $na$ | 0.455 | 0.458 |
| $SE_{11}$ | $na$ | $na$ | $na$ | $na$ | 0.193 | 0.208 |
| $SE$ | 0.313 | 0.330 | 0.282 | 0.296 | 0.228 | 0.239 |

Figure 1: Probability in expansion conditional on yield spead, static Probit and various GDC Probit models




Panel B of Figure 1 shows the probability of being in an expansion conditional on the yield spread lagged two periods. The static Probit conditions only on this variable. The GDC Probits also condition on the state of the business cycle at $t-1$ and $t-2$. The dashed lines are confidence bands with 95% coverage probability. The clear point made by figure 1 is that conditioning on whether the economy is in recession or expansion at $t-1$ and $t-1$ matters a great deal for the probability that the economy will be in expansion at period $t$.

## 5.2 Non parametric

The non parametric procedure described in section 4 above was implemented using a Nadaraya-Watson kernel estimator with Gaussian kernel. Four procedures were considered to obtain $h$ and $\rho$ viz 'plug-in', least squares cross validation[11], $AIC_c$ and $AIC_c^*$.

The plug-in window width, involves choosing the constants $c_0 = c_1 = 1$. When using plug-in values the coefficient $\rho$ in the distance function (24) was set equal to $\frac{1}{2}$. The results for the non parametric estimator with phase restrictions, plug-in $h$ but no shape restrictions are reported in table 3.

Table 3: Results: plug in $h$, monotonicity not imposed

| $AIC_c$ | $AIC_c^*$ | $l_{00}$ | $l_{11}$ | $l$ |
|---|---|---|---|---|
| -2.07 | 25.93 | -8.14 | -14.83 | -22.97 |

The $AIC_c$ and $AIC_c^*$ criteria proved to be computationally simple to implement and used little computer time. The results for $AIC_c$ are reported in table 4 and those for $AIC_c^*$ are in table 5. The results in both tables show that for any practical purpose $AIC_c$ is flat in $\rho$ something that is reassuring as it would be undesirable if the way in which distance is measured materially affected the estimates obtained when monotonicity is imposed. For this reason the discussion below focuses on the choice of $c_0$ and $c_1$ which affect the window width $h$.

The $AIC_c$ criterion selects $c_0 = 4.57$ for recessions which has the effect

[11] The least squares cross validation method was tried but proved to have two main disadvantages first it consumed a large amount of computer time. Second, it produced estimates of $c_0$ and $c_1$ that were implausibly low being one fifth of the plug in values cited above. Thus, least squares cross validation was not used in any of the empirical results reported below.

Table 4: $AIC_c$ results

| Method | | Value | | | | $AIC_c$ | $l_{00}$ | $l_{11}$ | l |
|---|---|---|---|---|---|---|---|---|---|
| $h$ | $\rho$ | $c_0$ | $c_1$ | $\rho_0$ | $\rho_1$ | | | | |
| Plug-in | Plug-in | 1 | 1 | 0.5 | 0.5 | -2.02 | -8.13 | -15.45 | -23.58 |
| opt | Plug-in | 4.57 | 1.00 | 0.5 | 0.5 | -2.32 | -9.85 | -15.26 | -27.58 |
| Plug-in | Opt | 1 | 1 | 0.34 | 0.56 | -2.02 | -8.13 | -15.39 | -23.51 |
| Opt | Opt | 4.57 | 1.11 | 0.28 | 0.47 | -2.32 | -9.85 | -15.59 | -25.44 |

of smoothing the yield spread out of the probability of expansion. However, the log probability score is -27.58 which is substantially worse than the -23.58 achieved when plug-in values of the smoothing parameters are used.[12] The $AIC_c$ criterion also selects $c_1 = 1$ (it also selects $c_1$ =1.11) which gives the yield spread a role in predicting the onset of recessions.

Table 5: $AIC_c^*$ results

| Method | | Value | | | | $AIC_c^*$ | $l_{00}$ | $l_{11}$ | l |
|---|---|---|---|---|---|---|---|---|---|
| $h$ | $\rho$ | $c_0$ | $c_1$ | $\rho_0$ | $\rho_1$ | | | | |
| Plug-in | Plug-in | 1 | 1 | 0.5 | 0.5 | 26.54 | -8.13 | -15.45 | -23.58 |
| opt | Plug-in | 0.88 | 0.50 | 0.5 | 0.5 | 25.74 | -8.13 | -14.67 | -22.60 |
| Plug-in | Opt | 1 | 1 | 0.45 | 0.57 | 26.40 | -8.13 | -15.31 | -23.43 |
| Opt | Opt | 0.84 | 0.77 | 0.24 | 0.19 | 25.93 | -7.94 | -14.88 | -22.82 |

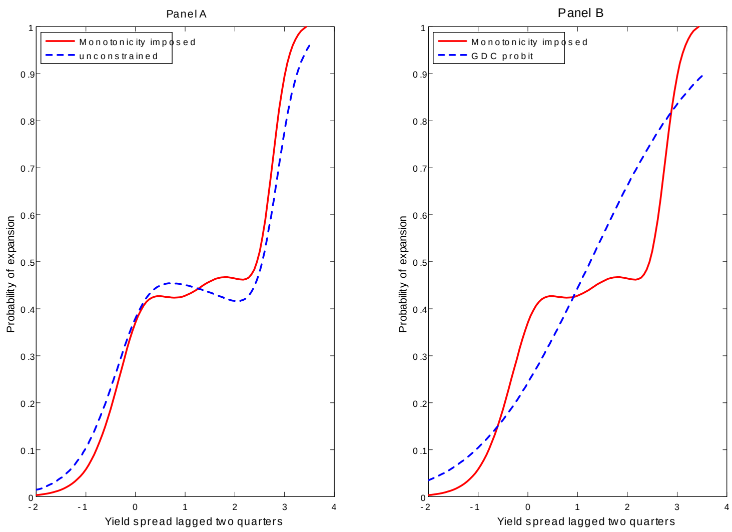
The $AIC_c^*$ criterion selects $c_0 = 0.88$ and $c_1 = 0.5$ which gives the yield spread a role in determining the probability of being in expansion. Overall the the log probability score is -22.60 which is slightly better than the model without monotonicity imposed but with plug-in values of the smoothing parameters. This result suggests that the additional smoothness achieved via imposition of monotonicity allows an improved fit to be achieved through a smaller window width.

[12]This seemingly strange result arises because the $AICc$ is not directly related to the log probability score.

### 5.2.1 Probability exit a recession that has lasted at least two quarters

Both panels of Figure 2 plot, against the yield spread lagged two quarters, non parametric estimates of the probability of exiting a recession that has lasted at least two quarters. The heavy line in both panels are estimated with phase and shape restrictions imposed. The dashed line in panel A is the probability obtained using the method in Harding and Pagan (2009) which does not impose monotonicity. Clearly, imposing monotonicity matters for the estimate of the probability of exiting a recession.

Figure 2: Probability exit a recession that has lasted for at least two quarters




Panel B of Figure 2 compares the shape constrained non parametric estimator with that from the GDC dual index Probit model. The important differences here are that the GDC dual index Probit model misses the 'flat spot' that occurs when the slope of the yield curve is between 0 and 2.5 per cent. A consequence of this is that the GDC dual index Probit is a poor approximation to the non parametric estimate of the probability of exiting a recession. The "flat spot" cited above is of considerable policy relevance because it says that unless they can get the slope of the yield curve above 2.5 per cent, policy makers in the United States have little chance of raising the probability that the economy exits a recession above 50 per cent.

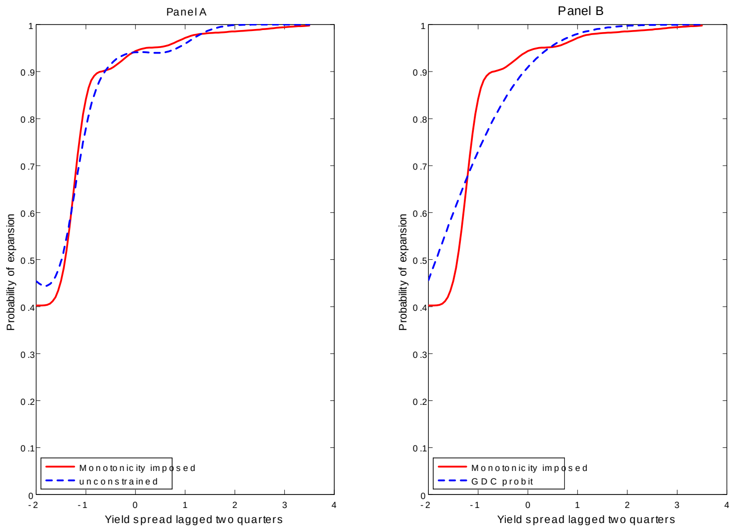
### 5.2.2 Probability of continuing in an expansion that has lasted at least two quarters

Figure 3 plots the probability of continuing in an expansion that has lasted at least two quarters conditional on the yield spread lagged two quarters. Panel A compares the estimated constrained to be monotonic with the unconstrained estimate. There are relatively minor differences between these two methods. The most notable difference is that the unconstrained method overestimates the probability of remaining in an expansion when the slope of the yield curve is below $-1$ per cent.

Panel B of Figure 3 compares the constrained non parametric estimate with that from the GDC dual index Probit model it is clear that the latter model is too restrictive. Over the range of yield spreads experienced in the United States the GDC Probit

- Over predicts the probability of remaining in an expansion for yield spreads of less than -1 per cent;
- Under predicts the probability of remaining in an expansion for yield spreads of between -1 and 0.5 per cent; .

Figure 3: Probability continue in an expansion that has lasted at least two quarters



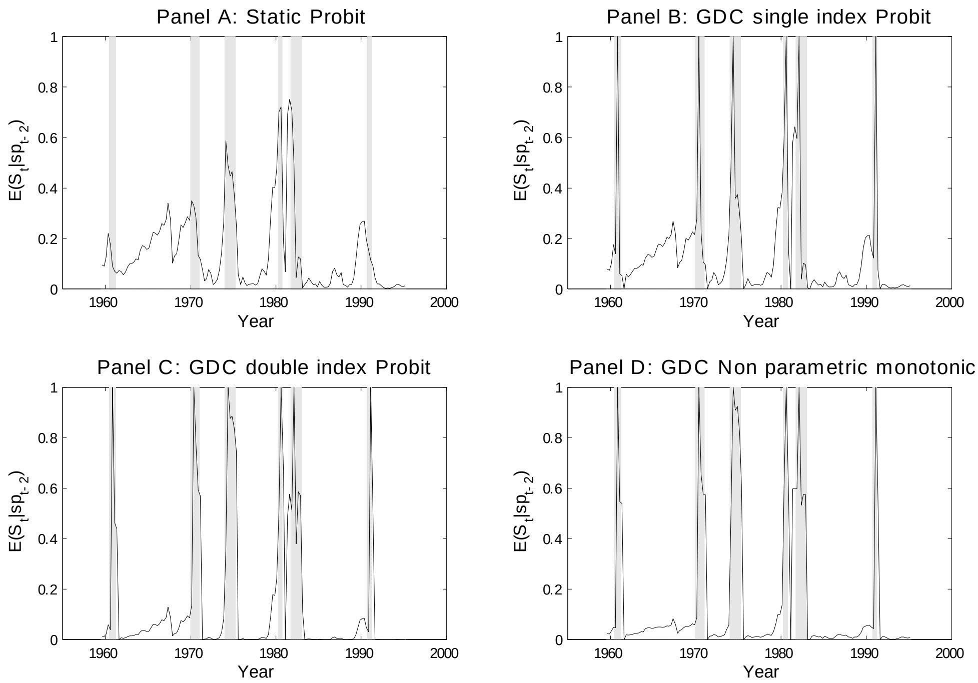

# 6 Predicted probability of recession

Analysts use dynamic categorical models to predict the probability that the economy will go into recession. Figure 4 is typical of the graphical devices used in the literature to compare the predictions of several competing models. Panel A shows that the static Probit model misses the 1960, 1970 and 1990 recessions, provided a weak signal of the mid 1970s recession and provided a somewhat stronger signal for the two recessions of the early 1980s.

The single index GDC Probit shown in panel B fits the data a little better but tends to provide a definitive recession signal after the recession has ended.

The double index GDC Probit shown in panel C does better than the two other Probits at fitting the business cycle states but this mainly comes about through better fitting the business cycle state away from turning points. The

Figure 4: Predicted probability of recession various models, 1959.3 to 1995.1



double index GDC model estimated non parametrically with shape restrictions is shown in panel D. It fits the business cycle states, away from turning points, a little better than the double index GDC Probit.

The reason for the qualification "away from turning points" above is shown in table 6 which shows $\Pr(S_t = 0|sp_{t-2})$ for the static Probit model and $\Pr(S_t = 0|S_{t-1} = 1, S_{t-2} = 1, sp_{t-2})$ for various GDC models in the first quarter of each recession . Except for the two recessions of the 1980s all models fail to achieve a probability of recession greater than one-half in the first quarter of the recession.

Table 6: $Pr(S_t = 0|S_{t-1} = 1, S_{t-2} = 1, sp_{t-2})$: First quarter of recession

| Date | Probits | | | Non parametric |
|---|---|---|---|---|
| | Static | GDC | | |
| | | Single index | Double index | |
| 1960.Q3 | 0.18 | 0.14 | 0.04 | 0.05 |
| 1970.Q1 | 0.35 | 0.28 | 0.13 | 0.09 |
| 1974.Q1 | 0.59 | 0.48 | 0.36 | 0.5 |
| 1980.Q2 | 0.70 | 0.59 | 0.50 | 0.60 |
| 1981.Q4 | 0.71 | 0.60 | 0.51 | 0.60 |
| 1990.Q4 | 0.15 | 0.12 | 0.03 | 0.04 |
| Mean | 0.45 | 0.37 | 0.26 | 0.31 |
| Std dev | 0.25 | 0.22 | 0.22 | 0.28 |

The average predicted probability of recession varies from 0.22 for the GDC Probits to 0.28 for the GDC estimated non parametrically with phase and shape restrictions. Although these probabilities are low they are substantially higher than $\Pr(S_t = 0|S_{t-1} = 1, S_{t-2} = 1)$ of 0.05 in Table 1 demonstrating that the yield spread plays a useful role in predicting recessions.

The predicted probability of $S_t = 0$ in the second quarter of recession is shown in Table 7. On average, the static Probit yields a probability of recession of 0.37 per cent even though the economy is known to be in the second quarter of a recession. All of the GDC models yield a predicted probability of being in recession of 1 — this is because they are designed to have this feature.

Table 7: $Pr(S_t = 0|S_{t-1} = 1, S_{t-2} = 0, sp_{t-2})$

| Date | Probits | | | Non parametric |
|---|---|---|---|---|
| | Static | GDC | | |
| | | Single index | Double index | |
| 1960.Q4 | 0.09 | 1 | 1 | 1 |
| 1970.Q2 | 0.33 | 1 | 1 | 1 |
| 1974.Q2 | 0.49 | 1 | 1 | 1 |
| 1980.Q3 | 0.72 | 1 | 1 | 1 |
| 1982.Q1 | 0.50 | 1 | 1 | 1 |
| 1991.Q1 | 0.11 | 1 | 1 | 1 |
| Mean | 0.37 | 1 | 1 | 1 |
| Std dev | 0.25 | 0 | 0 | 0 |

The four recessions with durations of three quarters or more provide useful information on the various models studied here. As is shown in Table 8 the static Probit yields an average predicted probability of recession of 0.21 for these cases which is a little better than the single index GDC Probit. The double index GDC Probit and non parametric estimate yield predicted probabilities of recession of 0.62 and 0.66 respectively. These should be compared with the $Pr\left(S_t = 0|S_{t-1} = 0, S_{t-2} = 0\right) = 0.6$ in Table 1. This needs to be interpreted with some caution and sophistication as it does not mean that the GDC model is useless at predicting or understanding how the yield spread influences the probability of exit from recession. Inspection of

Figure 2 shows that there is a “flat spot” in the exit probability curve which coincides with the range of yield spreads observed during recessions. It is only when policy makers push the yield spread is above this range that the economy has a high probability of exiting from recession.

Table 8: $Pr(S_t = 0|S_{t-1} = 0, S_{t-2} = 0, sp_{t-2})$: Third quarter of recession

| Date | Probits | | | Non parametric |
|---|---|---|---|---|
| | Static | GDC | | |
| | | Single index | Double index | |
| 1961.Q1 | 0.07 | 0.06 | 0.46 | 0.55 |
| 1970.Q3 | 0.28 | 0.22 | 0.77 | 0.65 |
| 1974.Q3 | 0.45 | 0.36 | 0.88 | 0.91 |
| 1982.Q2 | 0.04 | 0.04 | 0.38 | 0.53 |
| Mean | 0.21 | 0.17 | 0.62 | 0.66 |
| Std dev | 0.19 | 0.15 | 0.24 | 0.17 |

In summary, once a recession is underway the GDC models do very well at getting the second quarter of the recession but this is because the models have been constructed to capture that feature of the data. The double index GDC models do well in the third and subsequent quarters of the recession and this is where much of the improvement in fit comes from. The lesson that I take away from this is an old one that improved fit of a model does not necessarily produce improvement in economically relevant or valuable forecasts.

The economically relevant events are the turning points rather than the states $S_t$ and thus a lesson from the analysis above is that to achieve improved forecasts of these events it may be worth focusing attention on predicting quantities such as $S_{t-1}(1 - S_t)$ and $(1 - S_{t-1})S_t$ which take the values 1 if there is a turning point at $t$ and zero otherwise.

It is also instructive to compare the predictions of the various models in the first quarter of an expansion as is done in Table 9. The static probit model and the single index GDC probit yield predictions of the probability of expansion that average 0.88 and 0.90 respectively. This does not mean that such models are useful as they also yielded these predictions in the earlier stages of the recession — indeed these predictions are not far away from the unconditional probability of expansion of 0.85 shown in Table 1. The parametric and non parametric GDC models yield average probabilities that the economy is in expansion of 0.49 and 0.53 which should be compared with the 0.40 in Table 1. These results suggest that the yield spread makes only a modest contribution to the predicting the on set of an expansion.

Table 9: $Pr(S_t = 1|S_{t-1} = 0, S_{t-2} = 0, sp_{t-2})$:First quarter of expansion

| Date | Probits | | | Non parametric |
|---|---|---|---|---|
| | Static | GDC | | |
| | | Single index | Double index | |
| 1961.Q2 | 0.94 | 0.95 | 0.56 | 0.46 |
| 1971.Q1 | 0.88 | 0.90 | 0.43 | 0.43 |
| 1975.Q2 | 0.75 | 0.80 | 0.25 | 0.39 |
| 1980.Q4 | 0.82 | 0.85 | 0.33 | 0.43 |
| 1983.Q1 | 1.00 | 1.00 | 0.89 | 1.00 |
| 1991.Q4 | 0.91 | 0.92 | 0.48 | 0.44 |
| Mean | 0.88 | 0.90 | 0.49 | 0.53 |
| Std dev | 0.09 | 0.07 | 0.22 | 0.23 |

# 7 Conclusion

Allowing for the phase restrictions imposed by the method in which the data is constructed can improve the fit of DDC models of the business cycle. Taking into account the fact that double index models are required to match NBER data also improves fit. Shape restrictions implied by theory are automatically imposed in parametric models because they are inherited from the properties of probability distributions. Non parametric methods do not automatically impose these shape restrictions. Procedures were developed to impose these shape restrictions on non parametric estimators. In the application it was shown that imposing these shape restrictions did not materially worsen the fit of the model. The double index non parametric GDC model was shown to have the best fit to the business cycle states and was superior to the various Probit models. However, it was shown that this improved fit was achieved, primarily, through better fit away from turning points and did not bring as large an improvement in the prediction of turning points.

# References


Cressie, N.A.C., T.R.C. Read. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society*, Series B 46, 440-464.

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

Friedman, J. H. and R. J. Tibshirani. (1984). The monotone smoothing of scatterplots. *Technometrics* 26 243–250.

Hall, P., and L.,S. Huang. (2001) "Nonparametric Kernel Regression Subject to Monotonicity Constraints", *The Annals of Statistics*, Vol 29 No. 3 pp 624-647.

Harding, D., and A.,R. Pagan (2009). "An Econometric Analysis of Some Models for Constructed Binary Time Series", NCER working paper No 39, July.

Henderson, D.J., and C.,F., Parmenter (2009). "Imposing Economic Constraints in Nonparametric Regression: Survey, Implementation and Extension, IZA Discussion Paper No 4103, March.

Hurvich, C.M., J.S. Simonoff and C.L. Tsai. (1998). Smoothing parameter selection in non parametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B* **60**, 271-293.

Li, Q. and J.S. Racine. (2004). "Cross-validated local linear non parametric regression", *Statistica Sinica*, 14(2), 458-512.

Li, Q. and J.S. Racine. (2009). *Nonparametric Econometrics*, Princeton University Press, Princeton.

Racine, J.S., Parmeter, C.F., and P. Du. (2009). "Constrained Nonparametric Kernel Regression: Estimation and Inference, mimeo, Department of Economics, University of Victoria, Canada.