Imposing and testing monotonicity in generalized dynamic categorical models of the business cycle *

Don Harding[†]

March 31, 2010

Abstract

The generalised dynamic categorical (GDC) model introduced by Harding and Pagan (2009) allows for the wide variety of macroeconomic and financial applications where dynamic probit models are known to be inappropriate. Specifically, GDC models permit multiple indexes whereas dynamic probit is restricted to a single index. This later feature means that GDC models can represent the NBER business cycle something that is not true of dynamic probit models.

The contribution of this paper is to evaluate estimation procedures that impose monotonicity. The estimators considered are maximum likelihood (MLE) and those non parametric estimators surveyed in Henderson and Parmenter (2009). Particular attention is paid to Hall and Huang's (2001) constraint weighted bootstrapping estimator which introduces observation specific weights to the standard Nadaraya-Watson estimator.

The estimators are then applied to the NBER business cycle data to obtain 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

*Prepared for 15^{th} Australasian macroeconomics workshop, 8-9 April 2010, Wellington New Zealand. Preliminary and incomplete.

†School of Economics and Finance, La Trobe University

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 and of the second order generalized dynamic categorical (GDC) model they introduced as a parsimonious framework with the capacity 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 monotonicity, in the forcing variables, together with boundedness is an important feature exhibited by this class of models.

The non parametric estimation method developed by Harding and Pagan (2009) yields bounded transition probabilities but does not impose monotonicity. The contribution of this paper is to evaluate parametric and non parametric estimation procedures that impose monotonicity. Parametric maximum likelihood estimators are considered together with non parametric estimators surveyed in Henderson and Parmenter (2009). Particular attention is paid to Hall and Huang's (2001) estimator which introduces observation specific weights to the standard Nadaraya-Watson estimator.

The transition probabilities obtain their remaining properties from the distribution(s) of the latent variables that drive transitions between states.

Maximum likelihood estimation (MLE) and non parametric estimation of GDC models are discussed in section 3. MLE, discussed in section 3.1, requires knowledge of the distributions that drive the latent variables but has the advantage of automatically imposing boundedness and monotonicity.

Non parametric methods require weaker assumptions about the distributions that drive the latent variables but do not automatically impose boundedness or monotonicity.

Hall and Huang's (2001) methods for imposing constraints on non parametric models is particularly attractive as it is designed to have several valuable features for practitioners and is easy to implement. Their method which "involves tilting the empirical distribution to the least possible amount, subject to the constraint being enforced" is discussed in section 3.2.

The practical significance of the issues discussed here is investigated through application that is discussed in section 4.

Conclusions are in section 5.

2 A second order Generalized Dynamic Choice model

Focusing on the case where S_t, the binary variable of interest, is constructed, using NBER procedures, so that 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.

Then, the GDC model of order 2, introduced in Harding and Pagan (2009), is

$$\begin{aligned} Pr\left(S_t = 1|S_{t-1} = s_1, S_{t-2} = s_2, \mathbf{x}_t\right) &= \gamma_{00}\left(\mathbf{x}_t\right)\left(1-s_1\right)\left(1-s_2\right) \\ &\quad +\gamma_{01}\left(\mathbf{x}_t\right)\left(1-s_1\right)s_2 \\ &\quad +\gamma_{10}\left(\mathbf{x}_t\right)s_1\left(1-s_2\right)+\gamma_{11}\left(\mathbf{x}_t\right)s_1 s_2. \end{aligned} \tag{1}$$

where $\mathbf{x}_t$ is a vector of conditioning variables. As discussed in Harding and Pagan (2010) the NBER method of constructing the S_t imposes the restrictions that

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

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

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 ζ_{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 ζ_{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.

$$\begin{aligned} S_t &= S_{t-1}S_{t-2}\mathbf{1}\left(\zeta_{11t}>0\right)+\left(1-S_{t-1}\right)\left(1-S_{t-2}\right)\mathbf{1}\left(\zeta_{00t}>0\right) \\ &\quad +S_{t-1}\left(1-S_{t-2}\right)+\left(1-S_{t-1}\right)S_{t-2} \end{aligned} \tag{3}$$

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

$$\zeta_{11t}=\varphi\left(\mathbf{x}_{t,}\boldsymbol{\beta}\right)+\varepsilon_{11t} \tag{4}$$

$$\zeta_{00t}=\chi\left(\mathbf{x}_t,\boldsymbol{\theta}\right)+\varepsilon_{00t} \tag{5}$$

The shocks ε_{11t} and ε_{00t} are mutually independent, and also are independent of $\mathbf{x}_t$, have unit variance and densities $f\left(\varepsilon_{1t}\right)$ and $g\left(\varepsilon_{2t}\right)$ respectively. Moreover, we assume that the effects of the forcing variables on the latent variables are monotonic so that $\frac{\partial\zeta_{11t}}{\partial x_{jt}}$ $(\frac{\partial\zeta_{00t}}{\partial x_{jt}})$ do not change sign as x_{jt} varies. 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

$$\begin{aligned} \Pr\left(S_t=1|S_{t-1}=1,S_{t-2}=1,\mathbf{x}_t\right) &= E\left(S_t|S_{t-1}=1,S_{t-2}=1,\mathbf{x}_t\right) \\ &= E\mathbf{1}\left(\zeta_{11t}>0|\mathbf{x}_t\right) \\ &= \Pr\left(\zeta_{11t}>0|\mathbf{x}_t\right) \\ &= \int_{-\varphi(\mathbf{x}_{t,}\boldsymbol{\beta})}^{\infty} f\left(v\right)dv\equiv 1-F\left(-\varphi\left(\mathbf{x}_{t,}\boldsymbol{\beta}\right)\right) \end{aligned} \tag{6}$$

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

$$\begin{aligned} \Pr\left(S_t=1|S_{t-1}=0,S_{t-2}=0,\mathbf{x}_t\right) &= E\left(S_t|S_{t-1}=0,S_{t-2}=0,\mathbf{x}_t\right) \\ &= E\mathbf{1}\left(\zeta_{00t}>0|\mathbf{x}_t\right) \\ &= \Pr\left(\zeta_{00t}>0|\mathbf{x}_t\right) \\ &= \int_{-\chi(\mathbf{x}_t,\boldsymbol{\theta})}^{\infty} g\left(v\right)dv\equiv 1-G\left(-\chi\left(\mathbf{x}_t,\boldsymbol{\theta}\right)\right) \end{aligned} \tag{7}$$

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 $\gamma_{11}(\mathbf{x}_t) = 1 - F(-\varphi(\mathbf{x}_{t,}\boldsymbol{\beta}))$ and $\gamma_{00}(\mathbf{x}_t) = 1 - G(-\chi(\mathbf{x}_t, \boldsymbol{\theta}))$ it follows that

$$\frac{\partial \gamma_{11}(\mathbf{x}_t)}{\partial x_{jt}} = f(-\varphi(\mathbf{x}_{t,}\boldsymbol{\beta})) \frac{\partial \varphi(\mathbf{x}_{t,}\boldsymbol{\beta})}{\partial x_{jt}} \tag{8}$$

and

$$\frac{\partial \gamma_{00}(\mathbf{x}_t)}{\partial x_{jt}} = g(-\chi(\mathbf{x}_t, \boldsymbol{\theta})) \frac{\partial \chi(\mathbf{x}_{t,}\boldsymbol{\theta})}{\partial x_{jt}} \tag{9}$$

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, and they obtain the directions of the monotonicity (ie increasing or decreasing) from the direction of the monotonicity of $\varphi(\mathbf{x}_{t,}\boldsymbol{\beta})$ and $\chi(\mathbf{x}_{t,}\boldsymbol{\theta})$ respectively. In the special case where the latent variables are linear in the forcing variables $\varphi(\mathbf{x}_{t,}\boldsymbol{\beta}) = \mathbf{x}_{t,}'\boldsymbol{\beta}$ and $\chi(\mathbf{x}_{t,}\boldsymbol{\theta}) = \mathbf{x}_{t,}'\boldsymbol{\theta}$ the direction(s) of the monotonicity are determined by the signs of the coefficients β_j and θ_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 $\varphi(\mathbf{x}_{t,}\boldsymbol{\beta})$ and $\chi(\mathbf{x}_{t,}\boldsymbol{\theta})$ are strictly monotonic.

3 Estimation of GDC model

In estimating these models it is useful to define the following variables

- S_t^{11} subset of $\{S_1, ..., S_T\}$ such that $S_{t-1} = 1$ and $S_{t-2} = 1$.
- n_{11} is number of cases where $S_{t-1} = 1$ and $S_{t-2} = 1$.
- I_{11} is subset of $\{1, ..., T\}$ where $S_{t-1} = 1$ and $S_{t-2} = 1$
- S_t^{00} subset of $\{S_1, ..., S_T\}$ such that $S_{t-1} = 0$ and $S_{t-2} = 0$.
- n_{00} is number of cases where $S_{t-1} = 0$ and $S_{t-2} = 0$.
- I_{00} is subset of $\{1, ..., T\}$ where $S_{t-1} = 0$ and $S_{t-2} = 0$

- x_t^{11} subset of $\{x_1, ..., x_T\}$ such that $S_{t-1} = 1$ and $S_{t-2} = 1$.
- x_t^{00} subset of $\{x_1, ..., x_T\}$ such that $S_{t-1} = 0$ and $S_{t-2} = 0$.

With these definitions in place we can proceed to discuss maximum likelihood estimation (MLE) and non parametric estimation respectively.

3.1 Maximum likelihood estimation

The conditional likelihood functions are

$$L_{11} = \prod_{t \in I_{11}} \left[1 - F\left(-\varphi\left(\mathbf{x}_{t,}\boldsymbol{\beta}\right)\right)\right]^{S_t^{11}} \left[F\left(-\varphi\left(\mathbf{x}_{t,}\boldsymbol{\beta}\right)\right)\right]^{1-S_t^{11}} \tag{10}$$

$$L_{00} = \prod_{t \in I_{00}} \left[1 - G\left(-\chi\left(\mathbf{x}_{t,}\boldsymbol{\theta}\right)\right)\right]^{S_t^{00}} \left[G\left(-\chi\left(\mathbf{x}_{t,}\boldsymbol{\theta}\right)\right)\right]^{1-S_t^{00}} \tag{11}$$

To implement the maximum likelihood procedures one selects functional forms for $\varphi\left(\mathbf{x}_{t,}\boldsymbol{\beta}\right)$, $\chi\left(\mathbf{x}_{t,}\boldsymbol{\theta}\right)$ and parametric densities $f\left(.\right)$ and $g\left(.\right)$. For example, it is common to choose the linear specification $\varphi\left(\mathbf{x}_{t,}\boldsymbol{\beta}\right) = \mathbf{x}_{t,}'\boldsymbol{\beta}$ and $\chi\left(\mathbf{x}_{t,}\boldsymbol{\theta}\right) = \mathbf{x}_{t,}'\boldsymbol{\theta}$ and to assume that ε_{11t} and ε_{00t} have independent standard normal distributions. In this case, which might be referred to as GDC-Probit, $f\left(v\right) = g\left(v\right) = \phi\left(v\right) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$ and $F\left(z\right) = G\left(z\right) = \Phi\left(z\right) \equiv \int_{-\infty}^{z} \phi\left(v\right) dv$. In this case estimation proceeds by choosing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to maximize the log of the likelihoods viz.

$$\widehat{\boldsymbol{\beta}} = \arg\max \sum_{t \in I_{11}} \left\{S_t^{11} \ln\left[1 - \Phi\left(-\mathbf{x}_t'\boldsymbol{\beta}\right)\right] + \left(1 - S_t^{11}\right) \ln \Phi\left(-\mathbf{x}_t'\boldsymbol{\beta}\right)\right\} \tag{12}$$

$$\widehat{\boldsymbol{\theta}} = \arg\max \sum_{t \in I_{00}} \left\{S_t^{00} \ln\left[1 - \Phi\left(-\mathbf{x}_t'\boldsymbol{\theta}\right)\right] + \left(1 - S_t^{00}\right) \ln \Phi\left(-\mathbf{x}_t'\boldsymbol{\theta}\right)\right\} \tag{13}$$

Other distribution functions such a the logit might be considered so that $F\left(z\right) = G\left(z\right) = \Phi\left(z\right) = \int_{-\infty}^{z} \phi\left(v\right) dv$ yielding a model that might be designated as GDC-Logit.

The GDC-Probit and GDC-Logit models have the feature that the same density governs the entry to recession as governs the exit from recession. Investigators may wish to test this proposition. The most straight forward way to do this is to specify flexible functional forms. A useful functional form is the generalized exponential distribution discussed in Lye and Martin (1993). We restrict attention to a particular version — the generalized t distribution — defined as

$$f(v) = e^{\alpha_1 \arctan\left(\frac{v}{\delta s}\right)+\alpha_2 \ln\left(\delta^2+\left(\frac{v}{s}\right)^2\right)+\sum_{i=3}^{6} \alpha_i \left(\frac{v}{s}\right)^{i-2}-\eta} \tag{14}$$

Where

$$\eta = \ln \int_{-\infty}^{\infty} e^{\alpha_1 \arctan\left(\frac{v}{\delta s}\right)+\alpha_2 \ln\left(\delta^2+\left(\frac{v}{s}\right)^2\right)+\sum_{i=3}^{6} \alpha_i \left(\frac{v}{s}\right)^{i-2}} dv \tag{15}$$

The restrictions that the mean is zero and variance is one require that

$$0 = \int_{-\infty}^{\infty} v e^{\alpha_1 \arctan\left(\frac{v}{\delta s}\right)+\alpha_2 \ln\left(\delta^2+\left(\frac{v}{s}\right)^2\right)+\sum_{i=3}^{6} \alpha_i \left(\frac{v}{s}\right)^{i-2}-\eta} dv \tag{16}$$

and

$$1 = \int_{-\infty}^{\infty} v^2 e^{\alpha_1 \arctan\left(\frac{v}{\delta s}\right)+\alpha_2 \ln\left(\delta^2+\left(\frac{v}{s}\right)^2\right)+\sum_{i=3}^{6} \alpha_i \left(\frac{v}{s}\right)^{i-2}-\eta} dv \tag{17}$$

These restrictions tie down two parameters leaving up to seven parameters to be estimated. The unimodal students t distribution with mean zero, variance one and δ^2 degrees of freedom arises when $\alpha_2 = -0.5\left(1+\delta^2\right)$, $s = \sqrt{\frac{\delta^2}{\delta^2-2}}$ and $\alpha_1 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$. The unimodal normal distribution with mean zero and variance 1 arises where, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_5 = \alpha_6 = 0$, $s = 1$ and $\alpha_4 = -\frac{1}{2}$. These special cases provide useful starting values for estimation.

In working with this generalized distribution it is essential to note that in some circumstances parameters are unidentified. For example, δ is unidentified if $\alpha_1 = \alpha_2 = 0$.

3.2 Hall and Huang's non parametric method

Non parametric methods do not automatically impose monotonicity as is evident from Figure 3 in Harding and Pagan (2010).

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.

To implement Hall and Huang's (2001) method for the problem at hand taking expectation of 3 conditional on $(S_{t-1}=1, S_{t-2}=1, x_t)$ to obtain

$$\gamma_{11}(\mathbf{x}_t) = E(S_t | S_{t-1}=1, S_{t-2}=1, \mathbf{x}_t) \tag{18}$$

while taking the expectation conditional on $(S_{t-1}=0, S_{t-2}=0, x_t)$ yields

$$\gamma_{00}(\mathbf{x}_t) = E(S_t | S_{t-1}=0, S_{t-2}=0, \mathbf{x}_t) \tag{19}$$

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

$$\widehat{\gamma_{00}}(\mathbf{x}_t) = \frac{1}{n_{00}} \sum_{i \in I_{00}} A_i^{00}(x) S_i^{00} \tag{20}$$

$$\widehat{\gamma_{11}}(\mathbf{x}_t) = \frac{1}{n_{11}} \sum_{i \in I_{11}} A_i^{11}(x) S_i^{11} \tag{21}$$

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[1]

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

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

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

First estimate $\widehat{\gamma_{00}}(\mathbf{x}_t), \widehat{\gamma_{11}}(\mathbf{x}_t)$ 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 $\widehat{\gamma_{00}}(\mathbf{x}_t), \widehat{\gamma_{11}}(\mathbf{x}_t)$ 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 $\widetilde{\gamma_{00}}(\mathbf{x}_t)$ and $\widetilde{\gamma_{11}}(\mathbf{x}_t)$ as follows

$$\widetilde{\gamma_{00}}(\mathbf{x}) = \sum_{i\in I_{00}} p_i^{00} A_i^{11}(\mathbf{x}) S_i^{00}$$

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

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

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

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} and $\frac{1}{n_{00}}$ and p_i^{11} and $\frac{1}{n_{11}}$ respectively. The relevant constraints when estimating $\widetilde{\gamma_{jj}}(\mathbf{x}_t)$ are:

- Weights sum to one

$$\sum_{i\in I_{jj}} p_i^{jj} = 1 \tag{22}$$

- Monotonicity

$$\frac{\partial \widetilde{\gamma_{jj}}(\mathbf{x})}{\partial x_k} = \sum_{i \in I_{00}} p_i^{00} \frac{\partial A_i^{jj}(\mathbf{x})}{\partial x_k} S_i^{jj} > 0 \qquad \forall k \tag{23}$$

- Boundedness

$$0 \le \widetilde{\gamma_{jj}}(\mathbf{x}) \tag{24}$$

$$\widetilde{\gamma_{jj}(\mathbf{x})} \le 1 \tag{25}$$

The constrained estimator is obtained by choosing $\boldsymbol{p}^{jj}$ to minimize (**??**) subject to constraints (22) to (25). This constrained minimization is easily achieved using a procedure such as `fmincon` in Matlab.

Using the same arguments as in Harding and Pagan (2010) the following central limit holds

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

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

4 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 peaks and troughs in the business cycle. These questions are addressed using the same sample of data as is used by Estrella and Mishkin (1998) and Harding and Pagan (2009) this allows comparability with the literature. To maintain comparability with these earlier papers we main Estrella and Mishkin's (1998) specification that $x_t = sp_{t-2}$.

4.1 Maximum likelihood

Here we compare a variety of models.

The static probit estimated by Estrella and Mishkin (1998) is reported in column 1 of table 1 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. Since the two models have the same number of parameters we would always prefer the one with the better fit.

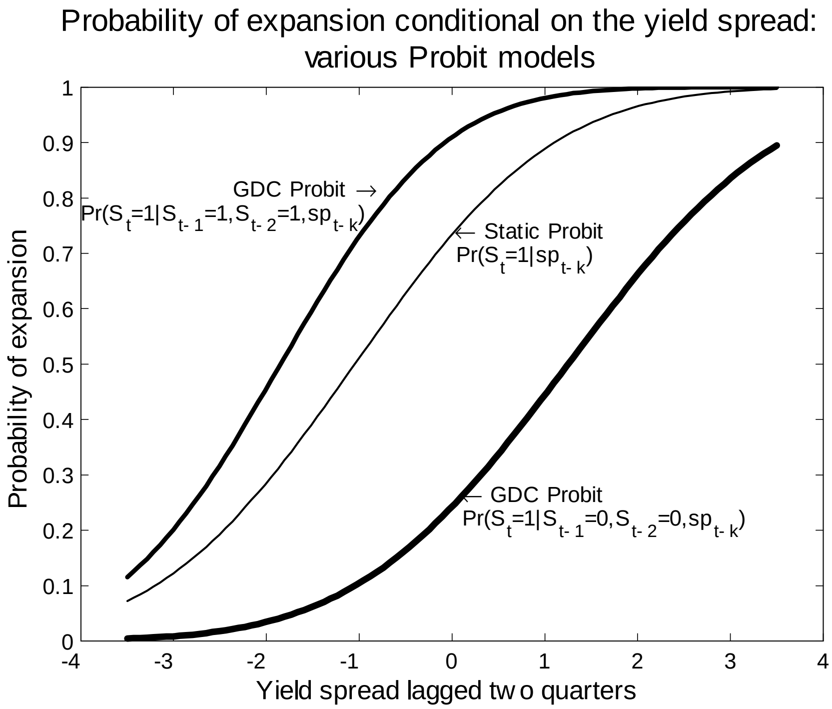
The results for double index GDC Probit model is shown in column 5 of table 1. It produces a large improvement in the log likelihood. The likelihood ratio statistic for the null hypothesis that there is a single index is 26.12. Since this is distributed $\chi^2(2)$ the 1% critical value is 9.21 and we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

Table 1: 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**
β_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)
β_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)
θ_0	*na*	*na*	$=\beta_0$	*na*	-0.699 (0.121)	-0.613 (0.112)
θ_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

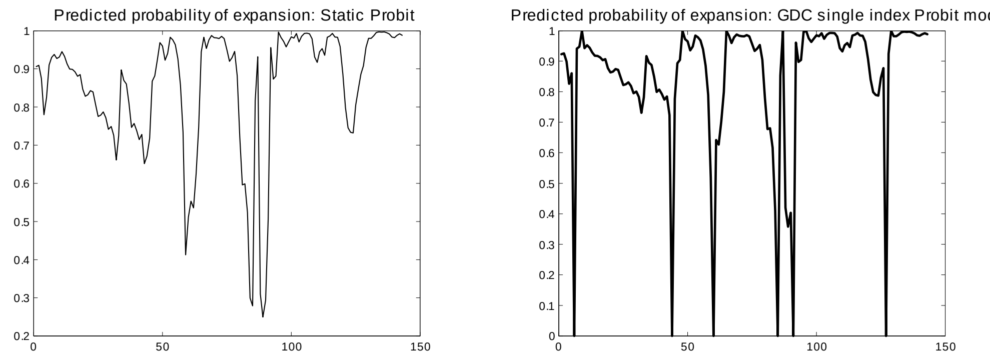
Also shown in columns 2, 4 and 6 of table 1 are the results for the logit model which fits a little worse than the Probit in each case.

Figure 4.1 show 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 clear point made by figure 4.1 is that this conditioning matters a great deal.



The GDC Probit model even with the constraint that ($\boldsymbol{\beta}_0 = \boldsymbol{\theta}_0, \boldsymbol{\beta}_1 = \boldsymbol{\theta}_1$) so there is a single index generates significan improvements in forecasting the probability of recession. This can be seen by comparing the Static Probit predictions in Panel a of Figure 2 with those for the single index GDC model in Panel B of that figure.

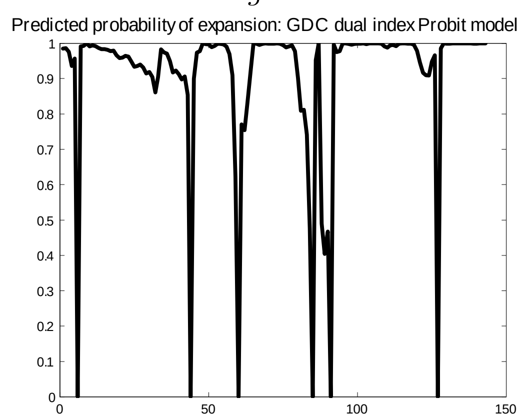
Figure 2



Panel a *Panel b*

The improvement in the predictions achieved by the dual index GDC probit are shown in Figure 3 below.

Figure 3

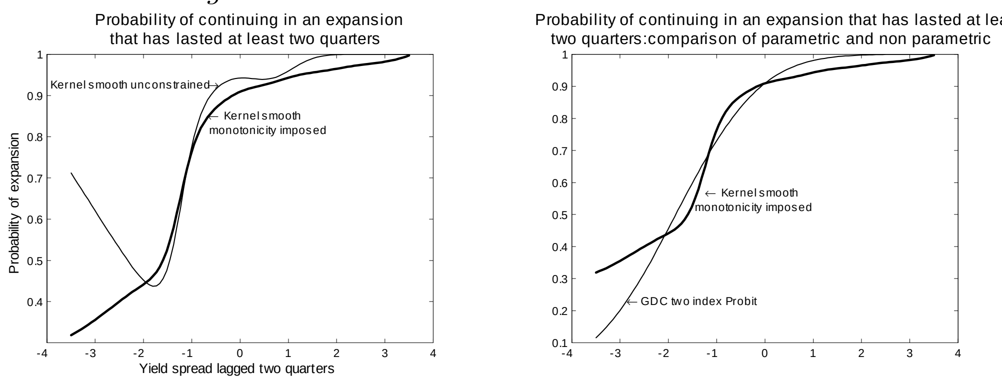


4.2 Non parametric

4.2.1 Probability of remaining in an expansion that has lasted at least two quarters

Figure 3 plots the non parametric probability of remaining in an expansion that has lasted at least two quarters against the yield spread lagged two quarters. The heavy line in both panels is from the method of Huang and Hall (HH) (2001). The method imposes monotonicity and boundedness. The lighter line is from is the probability obtained using the method in Harding and Pagan (2009) which imposes neither monotonicity nor boundedness. Clearly, imposing monotonicity matters for the estimate of the probability of remaining in expansion.

Figure 3



Panel A *Panel B*

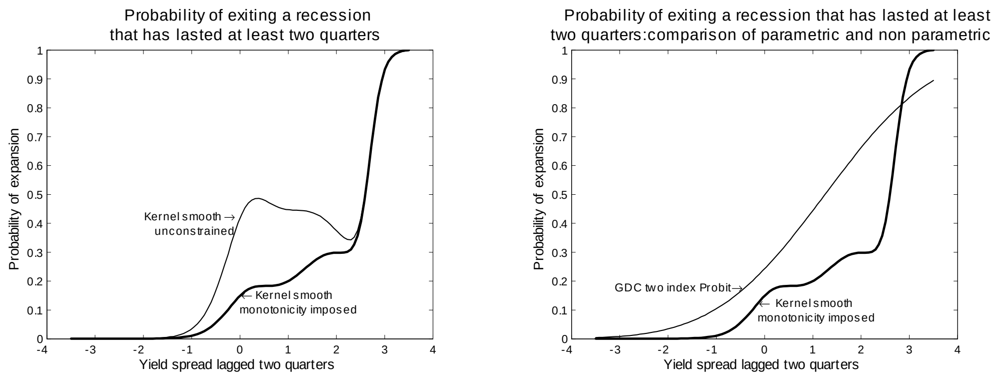
Panel B of Figure 3 compares the HH estimator with that from the GDC dual index Probit model. The important differences here are

- that the GDC dual index Probit model underestimates the rate at which the probability of continuing in expansion declines after the yield curve inverts.
- that the GDC dual index Probit model overestimates the probability of remaining in expansion when the slope of the yield curve is positive.
- That that the GDC dual index Probit model underestimates the probability of continuing in expansion when the yield curve is negative.

4.2.2 Probability of exiting a recession that has lasted at least two quarters

Figure 4 plots the probability of exiting a recession that has lasted at least two quarters conditional on the yield spread lagged two quarters. Panel A compares the HH method against the unconstrained method of Harding and Pagan (2009). It is clear that the unconstrained method over estimates the probability of exiting a recession for the yield spread in the range -1 to 2 per cent.

Figure 4



Panel A *Panel B*

Panel B compares the HH estimator with that from the GDC dual index Probit model it is clear that the latter model is too restrictive and over most of the range it over predicts the probability of exiting a recession.

5 Conclusion

We have examined the imposition of monotonicity restrictions in GDC models. Our main finding is that the method of Hall and Huang is easily applied and when applied it generates results that differ significantly from those for the unconstrained non parametric estimator.

Parametric MLE methods impose monotonicity automatically but they do so at the cost of reduced flexibility and poor fit. The generalized T distribution is a potential method for increasing flexibility within a parametric model but we have not yet applied that model in practice.

6 References

Cressie, N.A.C., Read, T.R.C., 1984. Multinomial goodnes-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 Tibshirani, R. J. (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.

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

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.