Forecasting Recessions Using the Mixed Frequency Probit

Michael T. Owyang
Research Division
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
owyang@stls.frb.org

keywords: term spread, business cycle, MIDAS

December 31, 2009

Abstract

Previous studies have found that the term spread has significant information content for forecasting recessions. These studies typically use probit or logit models to assess the probability of recession at various horizons. Most of these models use monthly averages of interest rates, potentially discarding important information about the timing of changes in the yield curve. In this paper, we exploit this timing information by implementing a mixed sampling probit, a binary variable extension of the MIDAS model suggested by Ghysels, Santa Clara, and Valkanov (2004). [JEL: C31, E32]
1 Introduction

Macroeconomics has a long tradition of characterizing the economy as moving through distinct phases – expansion and recession – dubbed the business cycle [see Burns and Mitchell (1946 book) and, more recently, Hamilton (1989 Econometrica), for example]. Further, many argue that these phases are asymmetric, resulting in very different behavior from consumers, firms, and policymakers.\(^1\) Agents, therefore, have incentives to forecast the timing of economic downturns and their subsequent upswings. A substantial literature has blossomed around the problem of estimating and forecasting the (discrete) transitions between business cycle phases.\(^2\)

The general framework for these problems is to construct the probability that the economy will experience a recession some number of months (or quarters) ahead. Many of the papers in this literature have utilized binary dependent variable models such as probit or logit. The covariates used to assess the probability of a business cycle turning point are typically the standard toolkit of macroeconomic variables: GDP, employment, inflation, etc [Chauvet and Hamilton (); Chauvet and Piger (JBES)]. Of course, some of these variables tend to lead the business cycle, while others tend to trail [Stock and Watson; Paap, Van Dyck, and co-author (JBES 2009)]. Whether the variable is a leading or lagging indicator of the business cycle can determine its general utility as an indicator in the binary forecasting regression.

In addition to the macroeconomic variables discussed above, financial market variables have become popular for forecasting turning points. For example, Estrella and Mishkin (1998 ReSTAT) and many others have found that the term spread – the difference between long- and short-term interest rates (typically on government securities) – can be informative about future turning points. The information content in the term spread for forecasting recessions is particularly apparent at longer horizons.\(^3\)

---

\(^1\)See Morley and Piger (2008 wp) for a detailed discussion and further evidence of business cycle asymmetries.

\(^2\)The literature is too vast to survey here but, as a general reference, see Estrella and Mishkin (1998 ReSTAT); Wright (2006 FRBoG working paper); Chauvet and Potter; and Rudebusch and Williams (this just got published). For a recent survey of the literature, see section XX of Wheelock and Wohar (2008 Review).

\(^3\)A number of other papers have investigated the ability of the term spread for forecasting macro variables such as
The models discussed above, however, may not be exploiting all the available information. While the business cycle is assessed at a monthly or quarterly basis, financial variables are often sampled at a higher frequency, e.g., daily or even higher. Models using macroeconomic variables alone do not suffer much from the difference in sampling frequency; however, the inclusion of daily financial variables may be problematic. The problem is typically resolved by converting the financial variables to the frequency of that the business cycle is sampled by simple time-averaging. Dueker (1996? Review), in a different context, warns that this procedure may discard valuable information about the timing of innovations to the high frequency variables.\textsuperscript{4} The goal of this paper is to incorporate some of this discarded timing information by weighting the daily financial data by adopting a technique introduced by Ghysels, Santa Clara, and Valkanov (2004, GSV). This weighting technique – known as Mixed Data Sampling (MIDAS) – utilizes a parsimoniously parameterized lag polynomial to weight the high frequency observations so they may be entered into the regression with data of a lower frequency. The specification allows us to estimate the weights on the high frequency data as functions of their proximity to sampling the macro variables.\textsuperscript{5}

The model proposed here is the mixed frequency probit or MIDAS probit. Its features are similar to those of a standard probit with the addition that the estimated probability depends on terms which aggregate the high frequency data via the nonlinear weighting polynomials proposed in GSV. In principle, the MIDAS probit can be estimated using either classical or Bayesian techniques. Because of its additional complexity introduced by the nonlinear weighting functions, we estimate the model using the Gibbs sampler.

The balance of the paper is constructed as follows: Section 2 introduces the baseline MIDAS probit specification. Section 3 outlines the estimation procedure, including details about the priors GDP. Although similar in motivation, these papers typically do not account for the nonlinearity that characterizes the Burns-and-Mitchell-type business cycle phases.

\textsuperscript{4}In particular, suppose the Fed lowers the fed funds rate by 25 basis points on the 21st of the month. The resulting monthly fed funds rate series would exhibit a 6-basis-point reduction in the current month and a 19-basis-point reduction in the subsequent month.

\textsuperscript{5}a couple of citations here about the ability of the framework to forecast macro variables...Andreou, Ghysels and co-author...Clements and Galvao.
and data. Section 4 discusses the empirical results for nowcasting conditional on data up through the current release. Section 5 incorporates daily leads of the high frequency variable to account for all available information. Section 6 summarizes and offers some conclusions.

2 The MIDAS-Probit Model

One of the standard models used for forecasting recessions is the probit.\(^6\) The variable of interest is the period–t latent business cycle indicator, \(S_t\), a binary variable sampled at the monthly frequency. We wish to use information available at time \(t\) to forecast \(S_{t+h}\), the value of the business cycle indicator at a (monthly) horizon \(h\), where \(h\) may be zero.\(^7\) Available at time \(t\) is a set of macroeconomic and financial covariates summarized by the forecasters information set, \(\Omega_t\). For the purposes of discussion, suppose that \(\Omega_t\) includes \(n_x\) variables \(X_t\) sampled at a monthly frequency and \(n_z\) variables \(Z_t = \left(z_{t-\frac{1}{m}}^{[m]}, z_{t-\frac{2}{m}}^{[m]}, \ldots, z_{t-\frac{m}{m}}^{[m]}\right)\) sampled \(m\) times more often than monthly reflected in the superscript. The subcripted notation \(t - \frac{i}{m}\) reflects the \(i\)th daily sample prior to the period–t sample of either \(X_t\) or \(S_t\).

In the absence of \(Z_t\), we can construct the canonical probit model in which the probability that \(S_{t+h}\) takes on a value of 1, conditional on the information at time \(t\), is determined by the relation

\[
p \left[S_{t+h | t} = 1\right] = \Phi \left(X_t' \beta_h\right), \tag{1}
\]

where \(\beta_h\) is a vector of coefficients particular to horizon \(h\) and \(\Phi(.)\) is the probit link function, the normal CDF. The value \(X_t' \beta_h\) is often termed the horizon–\(h\) score. We can normalize \(S_t = 1\) to reflect recessions; then, an increase in the score increases the probability of a recession at horizon \(h\).

Previous studies have shown that the standard probit specification (1) has some predictive power

\(^6\)Recently, some have studied the consequences of altering the binary choice model. For example, Katayama (2008) investigates a number of models including the logit. XXXX considers a dynamic probit. For simplicity, we consider only the standard probit here; extension of the model to other link functions should be transparent.

\(^7\)In the literature, setting \(h = 0\) is termed "nowcasting".
for calling recessions at short forecast horizons. These studies, however, have not taken advantage of the availability of high frequency financial data. In order to match the sampling frequency of the available recession data, most models use time averaged financial data. Recently, a series of papers has developed a method of parsimoniously incorporating high frequency information while preserving some of its timing characteristics. This so-called Mixed Data Sampling or MIDAS time aggregates the high frequency data using a polynomial weighting function. The slope coefficients and the polynomial hyperparameters are jointly estimated, typically using nonlinear least squares.

We can use the MIDAS framework to augment the standard probit (1) to account for the high frequency sampling of the variables $Z_t$. For brevity, we consider only a single high frequency variable, $z_t$. This is accomplished by adding a new term similar to those found in the standard MIDAS regressions and yields

$$p [S_{t+h|t} = 1] = \Phi \left( X_t' \beta_h + \gamma_h w \left( L_{\frac{h}{m}} ; \theta_h \right) z_t^{(m)}_{t-k/m} \right), \tag{2}$$

where $w \left( L_{\frac{h}{m}} ; \theta_h \right)$ is an order $m$ polynomial in the lag operator with an exogenously chosen form. The hyperparameters $\theta_h$ determine the shape of the polynomial. Consistent with the literature on MIDAS regressions, we choose a beta polynomial with $\theta_h = \{ \theta_{1h}, \theta_{2h} \}$:

$$w \left( L_{\frac{h}{m}} ; \theta_1, \theta_2 \right) = \frac{ f \left( \frac{k}{m}, \theta_1, \theta_2 \right) L_{\frac{h}{m}} }{ \sum_{j=1}^{m} f \left( \frac{k}{m}, \theta_1, \theta_2 \right) }, \tag{3}$$

where we have suppressed the dependence on the forecast horizon $h$ for convenience. The function $f \left( \frac{k}{m}, \theta_1, \theta_2 \right)$ is defined as follows:

$$f \left( j, \theta_1, \theta_2 \right) = \frac{ j^{\theta_1-1} (1 - j)^{\theta_2-1} \Gamma (\theta_1 + \theta_2) }{ \Gamma (\theta_1) \Gamma (\theta_2) }, \tag{4}$$

where $\theta_1$ and $\theta_2$ are hyperparameters governing the shape of the weighting function and
\[ \Gamma(\theta_p) = \int_0^\infty e^{-j} j^{\theta_i - 1} dj \]  

is the standard Gamma function.

The beta specification is a flexible polynomial form. Variation in the hyperparameters can produce a number of weighting polynomials. For example, the weighting function described above nests the simple time averaged specification used in the previous literature. In particular, if both of the MIDAS polynomial hyperparameters \( \theta_h \) are identically 1, the weighting function places equal weight on each of the high frequency variables. Other parameterizations of \( \theta \) can generate differently shaped weighing functions (e.g., strictly decaying, hump shaped, increasing, u-shaped).  

### 3 Empirical Approach

The MIDAS-probit model specified above can be estimated using standard Bayesian techniques [Albert and Chib (1993)]. We employ the Gibbs sampler [see Gelfand and Smith (1990), Cassella and George (1992), Carter and Kohn (1994)] with an additional Metropolis-in-Gibbs [Chib and Greenberg (1995)] step to estimate the polynomial hyperparameters. Estimation with the Gibbs sampler is accomplished by data-augmentation [Tanner and Wong (1987)], which calls for sampling an additional variable, \( y_{t+h|t} \), defined as

\[ y_{t+h|t} = \beta_h X_t + \gamma_h w \left( L^{\frac{1}{m}}; \theta_h \right) z_{t-k/m}^{(m)} + u_{t+h|t}, \]  

subject to the restriction

\[ y_{t+h|t} \geq 0 \quad \text{if} \quad S_{t+h|t} = 1, \]

\[^8\text{Other weighting polynomial forms used in the MIDAS literature include the Almon and exponential Almon. Substituting alternative polynomial specifications in our framework is straightforward but requires a change in the proposal density for the MH step discussed below.}\]
where \( u_t \) is iid standard normal. The sampler can then be broken down into blocks: the slope coefficients, \( \{ \beta_h, \gamma_h \} \); the polynomial hyperparameters, \( \{ \theta_h \} \); and the augmented data, \( \{ y^h_T = y_{1+h|1}; y_{2+h|2}, \ldots, y_{T+h|T} \} \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( [\beta, \gamma]^T )</td>
<td>( N(\mathbf{m}_0, \sigma^2 \mathbf{M}_0) )</td>
<td>( \mathbf{m}<em>0 = 0</em>{n_x+n_z}; \mathbf{M}<em>0 = \mathbf{I}</em>{n_x+n_z} )</td>
</tr>
<tr>
<td>( \Theta )</td>
<td>( \Gamma(\mathbf{d}_0, \mathbf{D}_0) )</td>
<td>( \mathbf{d}_0 = 1_2; \mathbf{D}_0 = 1_2 )</td>
</tr>
</tbody>
</table>

As is standard in the MIDAS literature, we specify a polynomial weighting function with two hyperparameters. We adopt the relatively standard normal prior for the slope coefficients. Similarly, the AR component has a normally-distributed prior with zero mean. The priors for the polynomial weight parameters are gamma distributed and constructed to centered around the belief that the high frequency data is equally weighted.\(^9\) Hyperparameters for the priors are given in Table 1.

### 3.1 Drawing \([\beta, \gamma]^T\) conditional on \(\theta, y, S\)

Conditional on \(\theta, y, S\), (6) is a linear regression. Let \( \mathbf{X}_t^* = [1, \mathbf{X}_t, \mathbf{w} \left( L^\frac{1}{z} ; \theta_h \right) z_{t-k/m}^{(m)}]' \) and \( \mathbf{X}^* \) represent the \( T \times (2 + n_x) \) vector of stacked variables. Then, given the prior \( N(\mathbf{m}_0, \mathbf{M}_0) \), a draw of \( \rho = [\beta, \gamma]^T \) can be made from \( \rho|\theta, y, S \sim N(\mathbf{m}, \mathbf{M}) \), where

\[
\mathbf{M} = \left( \mathbf{M}_0^{-1} + \mathbf{X}^* \mathbf{X}^* \right)^{-1}
\]

and

\[
\mathbf{m} = \mathbf{M} \left( \mathbf{M}_0^{-1} \mathbf{m}_0 + \mathbf{X}^* \mathbf{y} \right),
\]

where \( \mathbf{y} \) is the stacked vector of latent variables.

\(^9\)We could adopt a diffuse prior over the \( \theta \) hyperparameters. This would be an improper prior and be invalid for computation of marginal likelihoods.
3.2 Drawing $\theta$ conditional on $\beta, \gamma, y, S$

Obtaining a draw of $\theta$ can be accomplished using Metropolis in Gibbs step [Chib and Greenberg (2005)] to sample from the nontractable posterior distribution. The Metropolis step requires a candidate draw from a proposal density which is accepted with a probability that depends on both the likelihood and parameters’ prior distribution.

We utilize a Gamma proposal density, whose hyperparameters depend on the previous accepted draw. In other words, for the $(i+1)$ iteration, we draw a candidate $\theta^* = (\theta_1^*, \theta_2^*)'$ from

$$\theta_j^* \sim \Gamma\left(\sqrt{c}\theta_j^{[i]},\left(c\theta_j^{[i]}\right)^2\right),$$

where the superscript $i$ represents the draw from the previous iteration. The parameter $c$ is a scaling factor that can be tuned to achieve a reasonable acceptance rate. The candidate draw is then accepted with probability $a = \min\{\alpha, 1\}$, where

$$\alpha = \frac{f(y_{t}\mid \theta^*) \, dG(\theta^*\mid \mathbf{d}_0, \mathbf{D}_0)}{f(y_{t}\mid \theta^{[i]}) \, dG(\theta^{[i]}\mid \mathbf{d}_0, \mathbf{D}_0) \, dG(\theta^*\mid \sqrt{c}\theta^{[i]},\left(c\theta^{[i]}\right)^2)},$$

where $f(\cdot\mid\cdot)$ reflects the conditional likelihood whose log is

$$\ln f(y_{t}\mid \theta^*) = \sum_t y_{t+h|t} \ln \Phi \left( \frac{1}{\sqrt{m}} \left( X_t \beta_h + \gamma_h w \left( L^{\frac{1}{m}}; \theta^* \right) z_{t-k/m}^{(m)} \right) \right)$$

$$+ y_{t+h|t} \ln \left[ 1 - \Phi \left( \frac{1}{\sqrt{m}} \left( X_t \beta_h + \gamma_h w \left( L^{\frac{1}{m}}; \theta^* \right) z_{t-k/m}^{(m)} \right) \right) \right]$$

and $dG(\cdot\mid\cdot)$ is the gamma pdf.
3.3 Drawing $y$ conditional on $\beta, \theta, \gamma, S$

Given (6) and the covariates ($X$ and $z$), and conditional on the parameters ($\beta$, $\gamma$, and $\theta$), we can draw the latent variable from a normal distribution with mean $\delta_t$ and unit variance, where

$$
\delta_t = X_t' \beta + \gamma_h w \left( \frac{L_m}{\theta} ; \theta \right) z_{t-k/m}^{(m)}.
$$

However, we also know that the sign of $y$ must match the realization of $S$. Therefore, $y$ can be drawn from the truncated normal

$$
y_{t+h|t} \sim \begin{cases} 
N(\delta_t, 1) I[y_{t+h|t} \geq 0] & \text{if } S_{t+h|t} = 1 \\
N(\delta_t, 1) I[y_{t+h|t} < 0] & \text{if } S_{t+h|t} = 0
\end{cases},
$$

where the indicator $I_{[\cdot]}$ determines the direction of the truncation.

4 Results

[Results to follow]