Improving Recession Probability Forecasts in the U.S. Economy

Munechika Katayama*

Louisiana State University

mkatayama@lsu.edu

This Draft: March 29, 2010

First Draft: July 2008

Abstract

Using out-of-sample and cross validation methods, I systematically compare the performance of recession probability forecasting models that differ in terms of variables included and functional forms used. I find substantial gains from including additional variables, such as the S&P 500 and employment growth, together with the Treasury term spread. In addition, we can further improve forecasting accuracy by utilizing a non-Normal cumulative distribution function. I also examine this possibility by using the generalized Edgeworth expansion, which enables us to explore a wider set of skewness and excess kurtosis. Results are favorable for allowing for more flexibility in the functional form.

Keywords: Recession Probability; Forecasting; Generalized Edgeworth Expansion.

JEL Classification: C25, C53, E32, E37.

*I am indebted to James Hamilton for his encouragement and helpful comments. For helpful comments and suggestions, I thank Hiroaki Kaido and participants at the 2009 Far East and South East Asia Meeting of the Econometric Society, the UC Riverside Conference on Business Cycles, and the Ninth Annual Missouri Economics Conference. This paper is based on Chapter 3 of my doctoral dissertation at UCSD.

1 Introduction

There are many attempts to forecast recessions in the U.S. economy. Information contained in forward-looking variables can be used for making predictions about the future state of the economy. Stock and Watson (1989) report that the slope of the Treasury yield curve provides useful information in constructing their leading indicator. Estrella and Mishkin (1998) examine the usefulness of financial variables in making predictions about future recession probabilities at various forecasting horizons. They find that the slope of yield curve is the best single predictor of future recession probabilities.[1] Other studies also utilize the predictive content of the term spread in order to forecast recession probabilities.[2] Although developments in the literature achieve some success in making predicting recession, there is a room to improve forecasting accuracy.

To obtain further improvements, there are two margins that we can exploit. One is the question of selecting a set of predictor variables. The other is a choice of a functional form. Typical studies on forecasting recession probabilities utilize a simple binary response model, such as a probit or logit model. In this paper, I will explore the importance of the two margins in improving forecasting accuracy.

Within the class of probit/logit models, there is favorable evidence for benefits from including additional information, other than the term spread. For example, Estrella and Mishkin (1998) find some evidence on usefulness of including an additional variable, such as a New York Stock Exchange index or GDP growth. Wright (2006) reports that including the level of the federal funds rate together with the term spread results in better forecasting performance than the term spread alone.[3] King, Levin, and Perli (2007) recently find superior predictive power of the 5-year corporate credit spread on AA-rated firms compared with the 10-year-3-month Treasury term spread over

[1] Their finding is consistent with a simple rule of thumb that the Treasury yield curve inversions or negative term spreads are likely followed by recessions in subsequent periods and the term spread has been cited as a leading economic indicator.

[2] Other studies include Dotsey (1998), Birchenhall, Jessen, Osborn, and Simpson (1999), Estrella, Rodrigues, and Schich (2003), Stock and Watson (2003), Clements and Galvão (2006), Rudebusch and Williams (2008).

[3] He also finds some evidence favorable for controlling for a term premium proxy.

the Great Moderation period (since the mid 1980's). Furthermore, they also report that once the corporate spread is augmented by the term spread, the forecasting performance dramatically improves by reducing "false positive" predictions of recession.

There are also some attempts to extend a simple probit model in the recent literature. For example, Chauvet and Potter (2005) extend a probit model in a way that it allows for business-cycle dependent coefficients (i.e., multiple breaks) and/or autocorrelated errors. Kauppi and Saikkonen (2008) develop dynamic probit models that include lagged explanatory variables and lagged recessionary dummies. Dueker (2005) presents a framework (a Qual VAR) that enables us to treat a qualitative variable as endogenous in a typical VAR framework, so that we can obtain dynamic forecasts of the qualitative variable.

Going beyond the typical univariate probit model with the term spread, such as allowing additional features in a forecasting model and/or including additional predictor variables, most likely would improve in-sample forecasting performance. However, at the same time, we also have to worry about a potential over-fitting problem and we want to make a forecasting model as parsimonious as possible. Hansen (2008) examines the pitfalls of relying on in-sample fit. He shows that too good in-sample fit (over-fitting) tends to be associated poor out-of-sample fit and that model selection based on in-sample fit is not reliable. In addition, he demonstrates that tendencies of yielding spurious results are much higher if we use in-sample evaluation and are more pronounced when we compare performance of a large number of alternative models. For example, Sephton (2001) applies non-linear and non-parametric methods (multivariate adaptive regression splines) to forecasting recession probability. While in-sample fit with this very flexible approach shows great success, he finds that out-of-sample predictions are relatively not helpful. In this paper, for these reasons, I will use out-of-sample and cross validation methods to examine the importance of the two margins that can contribute to improve forecasting performance.

First, I systematically compare the forecasting performance of six-month-ahead recession probability predictions in the U.S. economy. I look at all possible combinations of 32 variables up to trivariate models with six different functional forms. Results highlight the importance of variable selection. Finding the best combination of predictor variables greatly improve forecasting performance. Especially, among variables considered, the combination of the term spread, changes in the S&P 500 stock price index, and growth rate of non-farm employment is found to achieve the best forecasting accuracy. Furthermore, additional gains can be obtained by using non-Normal cumulative distribution functions. These additional features help improve forecasting accuracies by amplifying positive signals during recessions and also by dampening false positive alarms during expansions. However, there is some mixed evidence on what features of functional form are necessary in order to improve forecasting performance. Using known CDFs amounts to imposing restrictions on the combination of skewness and excess kurtosis. Thus, I proceed to allow for more flexibility in the functional form by utilizing the generalized Edgeworth expansion of Jarrow and Rudd (1982), which allows us to explore a wider space of skewness and excess kurtosis. By using the genenelized Edgeworth expansion, we can obtain further improvement in out-of-sample forecasting accuracy. Findings in this paper complements other developments in the literature mentioned above. We can easily include additional variables into other extensions to probit models and also the more flexible functional form can be incorporated into a dynamic model or it can include multiple breaks, in principle.

The rest of the paper is organized as follows. Section 2 lays out basic framework of the study. I present basic binary response models for forecasting recession probability, including the discussion of variables to be used and the method of evaluating forecasting performance. Section 3 presents empirical results and performs robustness check. Section 4 introduces the generalized Edgeworth expansion and applies it to making predictions of the U.S. recession probability. Finally, Section 5 concludes.

2 Basic Framework

2.1 Recession Probability Forecasting Model

Let y_t represent an NBER recession binary variable, which equals 1 when the economy is in recession in month t and equals 0 in expansion.[4] Typical models of forecasting h-period-ahead recession probabilities using the information available at time t assume that

$$\text{Prob}(y_{t+h} = 1|\boldsymbol{x}_t) = F(\boldsymbol{\beta}'\boldsymbol{x}_t), \tag{1}$$

where $F(\cdot)$ is a monotonically increasing function, whose range is the unit interval, $\boldsymbol{\beta}$ is a vector of coefficients associated with a vector of predictors $\boldsymbol{x}_t' = [1, x_{1,t}, \cdots, x_{k,t}]$, and k is the number of variables included.

It is commonly assumed that y_{t+h} is a conditionally independent Bernoulli random variable, so that the likelihood function is given by:

$$L = \prod_{t=1}^{T} \left[F(\boldsymbol{\beta}'\boldsymbol{x}_t)\right]^{y_{t+h}} \left[1 - F(\boldsymbol{\beta}'\boldsymbol{x}_t)\right]^{1-y_{t+h}}. \tag{2}$$

In an empirical analysis, I will set $h = 6$ and focus on six-month-ahead predictions.

In this formulation, predicting recession probability involves two issues that are possibly related to each other. The first one is to choose a set of predictor variables, so that we can obtain useful information from data. A workhorse predictor in the literature is the Treasury term spread between 10-year and 3-month bonds, which is due to the finding in Estrella and Mishkin (1998).

Given a choice of $F(\cdot)$, finding a better combination of predictor variables obviously helps improve forecasting accuracy if those variables contain different information and jointly provide

[4] It is important to mention that, in reality, the NBER's announcement often involves a significant time lag. This is one drawback of relying on the NBER's decision. We can employ alternative definitions of business cycles, such as more than two consecutive quarters of negative output growth as a recession. Alternatively, we can utilize an algorithm proposed by Chauvet and Hamilton (2006), which is motivated by the delay in the NBER's announcement, for "nowcasting," in order to determine near-real-time values of y_t. However, since it is the NBER's decisions that is most widely used in the academia as well as the public, I will use it to define the state of the U.S. economy.

useful signals. Since there are no *a priori* variable selection procedures available, our approach is to try all possible combinations of variables in order to find a better combination of variables that are helpful and stable in forecasting recession probabilities. Although knowing the best single predictor is helpful, it is unlikely that we can obtain some insights about a better combination of variables by just looking at forecasting performance of single predictors because those variables that show relatively good forecasting performance tend to contain similar information.

The second issue is how to translate signals into a probability measure between 0 and 1, which is related to a shape of $F(\cdot)$. In order to guarantee that $F(\cdot)$ is monotonically increasing and takes values between 0 and 1, we typically use a known cumulative distribution function (CDF). A popular choice is to use either the Standard Normal CDF (a probit model) or the Logistic CDF (a logit model), or some extensions to those (e.g., Chauvet and Potter, 2005; Kauppi and Saikkonen, 2008). However, the shape of $F(\cdot)$ is not necessarily restricted to typical ones. In principle, a CDF of any continuous random variable will suffice.

Differences in the shape of CDFs can be characterized in terms of "skewness" and "excess kurtosis."[5] It should be noted that it is not appropriate to use terms skewness and excess kurtosis here because we are not talking about characteristics of underlying statistical distributions, but just utilizing functional forms. However, for expositional simplicity, I will use those terms in describing the shape of a CDF. Higher excess kurtosis makes a CDF steeper around $F(\boldsymbol{\beta}'\boldsymbol{x}) = 0.5$. On the other hand, allowing non-zero skewness makes a CDF asymmetric, in the sense that $F(\boldsymbol{\beta}'\boldsymbol{x}) \neq 1 - F(-\boldsymbol{\beta}'\boldsymbol{x})$. With non-zero skewness, the shape of the CDF around $F(\boldsymbol{\beta}'\boldsymbol{x}) = 0$ and that around $F(\boldsymbol{\beta}'\boldsymbol{x}) = 1$ are especially different.

In order to understand what features of a CDF are helpful in improving forecasting accuracy, I will consider six different CDFs. Table 1 summarizes characteristics of the CDFs considered. In addition to the Standard Normal CDF and the Logistic CDF, I consider Student-t, Laplace,

[5] In general, changing location and scale parameters does not affect forecasting results and they are fixed to avoid an identification problem. Changing the location parameter just changes the estimate of the constant term and changing the variance just results in different scaling of coefficients $\boldsymbol{\beta}$.

Gumbel, and Type III Generalized Extreme Value (GEV3). All location parameters and scale parameters, if applicable, are set to 0 and 1, respectively. These CDFs are chosen and configured to incorporate non-zero skewness and/or higher excess kurtosis that the Standard Normal CDF does not have. The first four CDFs have zero skewness and differ in terms of degree of excess kurtosis. The Logistic CDF has excess kurtosis of 1.2. The degrees of freedom parameter for Student-t is set to be 6.5, such that its excess kurtosis equals 2.4. The excess kurtosis of Laplace is 3. The last two CDFs also have positive skewness. For the Gumbel CDF, skewness is equal to 1.1395 and excess kurtosis is 2.4. The shape parameter of the Type III GEV, s, is set such that its excess kurtosis equals 3 (i.e., $s = -0.1732$). The associated skewness becomes 0.3492. Thus, for a given set of predictor variables, I will be able to infer the importance of allowing for excess kurtosis in $F(\cdot)$ by looking at results based on the first four CDFs. Furthermore, comparing the Gumbel with the Student-t or the Type III GEV with the Laplace enables us to obtain some insight on whether allowing positive skewness is helpful in improving forecasting performance.[6]

2.2 Data

Table 2 lists all 32 monthly variables considered in this paper. The sample period starts from January 1960 and ends at December 2008. It is chosen to maximize data availability and to include as many recession episodes as possible. As a result, the sample covers eight post-war recessions. This is important in choosing a set of predictor variables that are robust. It is not so difficult to find a particular variable combination that can explain previous recessions ex post. However, there is no guarantee that a good predictor for a particular recession works well in predicting another one.

The data set contains the term spreads, the credit spreads, various interest rates, employment data, stock price indices, monetary aggregates, and other macroeconomic variables. Most of those

[6] It is also possible to introduce negative skewness by using the Type III GEV. However, preliminary estimation indicated that forecasting performance based on the Type III GEV with negative skewness is inferior to others in general. So, I do not report their results.

Table 1: List of the Cumulative Distribution Functions Considered

Type of CDF	$F(x)$	Skewness	Excess Kurtosis
Standard Normal	$\frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} \exp\left(\frac{-u^2}{2}\right) du$	0	0
Logistic	$\frac{\exp(x)}{1+\exp(x)}$	0	1.2
Student-t with $\nu = 6.5$	$\int_{-\infty}^{x} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1+\frac{u^2}{\nu}\right)^{\frac{-(\nu+1)}{2}} du$	0	2.4
Laplace	$\frac{1}{2}\left[1+\operatorname{sgn}(x)\left\{1-\exp\left(-\lvert x\rvert\right)\right\}\right]$	0	3
Gumbel	$1-\exp\left\{-\exp(-x)\right\}$	1.1395	2.4
Type III GEV with $s = -0.1732$	$\exp\left\{-\left(1+sx\right)^{-1/s}\right\}$	0.3492	3

Note: All location parameters are set to be 0 and all scale parameters are set to be 1. ν represents degrees of freedom parameter for the Student-t distribution. $\Gamma(\cdot)$ is a gamma function. Excess kurtosis of Student-t is given by $6/(\nu-4)$ for $\nu > 4$. Skewness of the Gumbel distribution is $\frac{12\sqrt{6}\zeta(3)}{\pi^3}$, where $\zeta(\cdot)$ is a zeta function. s denotes a shape parameter of the generalized extreme value distribution. For the Type III GEV, skewness is given by $\frac{-\Gamma(1-3s)+3\Gamma(1-s)\Gamma(1-2s)-2(\Gamma(1-s))^3}{\left(\Gamma(1-2s)-(\Gamma(1-s))^2\right)^{3/2}}$ and excess kurtosis is give by $\frac{\Gamma(1-4s)-4\Gamma(1-s)\Gamma(1-3s)+6\Gamma(1-2s)(\Gamma(1-s))^2-3(\Gamma(1-s))^4}{\left(\Gamma(1-2s)-(\Gamma(1-s))^2\right)^2}$.

are investigated in the earlier studies or used in constructing a composite leading indicator. It also includes variables, to which the NBER Business Cycle Dating Committee pays particular attention in deciding business cycle peaks and troughs. Since Kane (2008) documents the usefulness of employment data in predicting occurrence of recessions, the data set includes employment-related variables as well.

Recently, King, Levin, and Perli (2007) report that, in the period of the Great Moderation, the credit spread on AA-rated firms has particularly good forecasting performance. However, the corporate bond yields used in their study (maturity of 5-year and 10-year) do not cover the entire sample. Since there are not many recessionary periods, especially after the 1980's, using shorter sample periods may have considerable effects on evaluating forecasting performance. Thus, I have

Table 2: List of Variables

Predictor	Description	Info. Lag
Interest Rates		0
FF	Federal Funds rate	0
3M	3-month Treasury Bill rate	0
5Y	5-year Treasury Bond rate	0
10Y	10-year Treasury Bond rate	0
AAA	Moody's corporate bond yield, AAA 20 years or longer	0
AA	Moody's corporate bond yield, AA 20 years or longer	0
A	Moody's corporate bond yield, A 20 years or longer	0
Term Spreads		
TS10YFF	10Y-FF Treasury term spread	0
TS10Y3M	10Y-3M Treasury term spread	0
TS10Y5Y	10Y-5Y Treasury term spread	0
Credit Spreads		
CSAAA	AAA - 10Y spread	0
CSAA	AA - 10Y spread	0
CSA	A - 10Y spread	0
Employment Data		
EMP	Non-agricultural employment (log-differenced)	0
CEMP	Civilian employment (log-differenced)	0
UICLAIM	Initial unemployment insurance claims (log-differenced)	1
UNEMP	Unemployment rate	0
UNEMPD	Changes in unemployment rate	0
HOURS	Average weekly hours in manufacturing (log-differenced)	0
Stock Price Indices		
DJ30	Dow Jones 30 average (% changes over 3 months)	0
SP500	S&P 500 stock price index (% changes over 3 months)	0
Monetary Aggregates		
M0	Monetary base (log-differenced)	1
M1	M1 (log-differenced)	1
M2	M2 (log-differenced)	2
Other Macroeconomic Variables		
CLI11	Composite leading indicators (11 series, 1987=100, log-differenced)	1
CPI	CPI, all urban, all items (log-differenced)	1
EXP	Consumer expectation (1966.1 = 100)	0
EXPD	Changes in consumer expectation	0
HOUSE	New private housing units authorized by building permits (log-differenced)	1
VENDOR	Vendor performance (slower deliveries diffusion index, %)	0
INCOME	Personal income less transfer payments (log-differenced)	2
IP	Industrial production (log-differenced)	1
SALES	Manufacturing & trade sales (log-differenced)	1

Note: Information lag is measured at the end of month. Strictly speaking, those employment data with zero information lag and vendor performance are not available at the end of the month. However, they will be available at the very beginning of the next month and there are virtually no considerable lags. Thus, they are categorized in the zero information lag variable.

decided not to include the credit spreads that King, Levin, and Perli (2007) have used. Instead, as a crude proxy, the data set include spreads between Moody's AAA-, AA-, or A- rated corporate bond yield (20 years or longer) and the 10-year Treasury bond yield.[7]

When we focus on financial variables as predictors, we do not need to consider a gap between when an observation is made and when it is available for forecasting. However, as shown in Table 2, some series are not reported immediately and we need to take account of the information lag, in order to accurately assess forecasting models. In this paper, any variable z_t represents the latest data on z available at month t, instead of an observation at month t. For example, industrial production (IP) has 1 month of the information lag. So, $\text{IP}_{2000:01}$ refers to the industrial production data on December 1999.

Because the total number of models increases exponentially as the number of variables included (k) increases, I will restrict my attention to $k \leq 3$. This will result in examining a total of $\sum_{k=1}^{3} \frac{32!}{k!(32-k!)} \times 6 = 32928$ forecasting models.

2.3 Evaluating Forecasting Performance

In order to evaluate various forecasting models, I will primarily focus on out-of-sample results. Hansen (2008) shows that in-sample and out-of-sample fits are negatively correlated, which implies that good in-sample performance is not a useful indicator of out-of-sample accuracy and that relying on in-sample fit is highly misleading. This over-fitting problem is particularly important, since the likelihood of obtaining spurious results is more pronounced when we search a large number of alternative models. For this reason, I will focus on recursive out-of-sample forecasting evaluation and I will use cross validation as a robustness check.

For recursive (pseudo) out-of-sample forecasting exercises, the out-of-sample prediction starts from January 1989 and ends at the end of the full sample. The out-of-sample period covers the last

[7] Since the 20-year Treasury bond yield data has discontinuity between January 1987 and September 1993, I use the 10-year T-bond rate, instead. Thus, precisely speaking, this "credit spread" is a combination of the "true" credit spread and the term spread.

three recessions in the U.S. economy. By adding observations one by one, I estimate a forecasting model again to produce a forecast for the next month. In reality, when we forecast future recession probability, we are not certain about the true state of the economy in recent months. It is because the decision of the NBER Business Cycle Dating Committee typically involves substantial time lag. To be realistic, I assume that forecasters do not know the true state of the economy for a year and assume that $y_t = y_{t-1} = \cdots = y_{t-11} = y_{t-12}$. In other words, a forecaster assumes that the economy is in the same state as a year ago.

Following Clements and Galvão (2006), I will use three different measures for evaluating accuracy of out-of-sample recession probability predictions. The first measure is the probability-analogue of mean squared error, the quadratic probability score (QPS), which is commonly used in evaluating probability forecasts. The QPS is defined as

$$QPS = \frac{2}{T}\sum_{t=1}^{T}(\hat{p}_t - y_t)^2, \tag{3}$$

where $\hat{p}_t$ is the recession probability forecast month t. The QPS takes values between 0 and 2 and smaller value indicates more accurate forecasts.

The second measure is the log probability score (LPS), which is defined as

$$LPS = -\frac{1}{T}\sum_{t=1}^{T}\big\{y_t \log(\hat{p}_t) + (1-y_t)\log(1-\hat{p}_t)\big\}. \tag{4}$$

The LPS ranges from 0 to $+\infty$ and a smaller value corresponds to more accurate predictions and penalizes larger mistakes more heavily than the QPS.

The last measure is the Kuipers Score (KS), which is given by

$$\begin{aligned} KS &= \frac{\sum_{t=1}^{T} y_t \mathbb{1}_{[(\hat{p}_t>0.5)]}}{\sum_{t=1}^{T} y_t} - \frac{\sum_{t=1}^{T}(1-y_t)\mathbb{1}_{[(\hat{p}_t>0.5)]}}{\sum_{t=1}^{T}(1-y_t)}, \\ &= \text{hit rate} - \text{false rate}, \end{aligned} \tag{5}$$

where $\mathbb{1}_{[\cdot]}$ is an indicator function that equals 1 if its argument is true and 0 otherwise. The KS calculates the difference between the hit rate and the rate of false signals by using 50% probability of recession as a cutoff. The KS takes values between -1 and 1. A score of 1 corresponds to making perfect predictions. The KS evaluates the recession predictions from a slightly different aspect, compared with other two measures. Even when recession predictions never show "strong" indications (say, higher than 50% probability), it is possible to have seemingly good results based on the QPS and LPS. The KS discounts such "weak" predictions. In this sense, the KS captures the strength and accuracy of predictions by using the 50% probability cutoff.

There is a potential problem of just relying on the recursive out-of-sample exercises described above, especially in the context of recession probability forecasting. Since there are not many recession episodes in the out-of-sample period, it is possible to select a forecasting model that has particularly good performance for the last three recessions, but not a next recession.

In order to robustify the results, I will also carry out a cross-validation type exercise, called leaving 2-years out. The detailed procedures of the leaving 2-years out exercises are as follows. Let $S = \{(y_t, \boldsymbol{x}_{t-h}) : t = 1, \cdots, T\}$ denote a full sample and $L_\tau = \{(y_t, \boldsymbol{x}_{t-h}) : t = \tau - 12, \cdots, \tau + 12\}$ represent a set of excluding observations. For each $\tau = 13, \cdots, T - 12$,

(i) Take $E_\tau = S \setminus L_\tau$ as a training sample.

(ii) Estimate parameter values $\boldsymbol{\beta}_\tau$ based on E_τ.

(iii) Make a prediction for y_τ by using $\boldsymbol{\beta}_\tau$ and $\boldsymbol{x}_{\tau-h}$ and store it.

(iv) Repeat steps (i) – (iii).

Then calculate the QPS, LPS, and KS based on $\{\hat{y}_\tau\}_{\tau=13}^{T-12}$.

It is necessary to find a forecasting model whose forecasting performance is stable over time and is robust to different recessions. The recursive out-of-sample forecasting exercises seem to be a more realistic setup. However, at the same time, it could be vulnerable to serially correlated errors. Given the conditional independence assumption, the leaving 2-years out exercise provides a completely valid evaluation and hopefully picks up a forecasting model whose performance is robust across different recession episodes. For this reason, it is very important to draw conclusions based on two different ways of evaluating forecasting performance.

3 Results

3.1 Univariate Probit Models

First, we will start off by looking at forecasting performance within a class of univariate probit models of predicting six-month-ahead recession probabilities ($h = 6$) as a benchmark. Table 3 summarizes the out-of-sample forecasting performance of univariate probit models.

Overall, the out-of-sample forecasting performance evaluated by the QPS and LPS shows broadly similar patters for the rankings of predictor variables. While the QPS selects the composite leading indicator (CLI11) as the best predictor within univariate probit models, the LPS chooses the term spread between 10-year Treasury bond yield and the Federal Funds rate (TS10YFF) as the best predictor. Based on both the QPS and LPS, 3-month change in S&P 500 (SP500) appears to be one of best predictors. It is important to point out that these variables outperform the widely used recession probability predictor, the term spread between 10-year and 3-month Treasury yields (TS10Y3M). Given the forecasting horizon of six months, superiority of TS10YFF over the conventionally used TS10Y3M is true not only in the out-of-sample forecasts, but also in the entire sample (detailed in-sample results are not reported here).

In contrast to better performance of these term spreads the credit spreads (CSA, CSAA, and CSAAA) have poor forecasting performance. Among other variables, CPI inflation (CPI) and University of Michigan's index of consumer expectation (EXP) have relatively good performance based on both the QPS and LPS. Some of employment related variables, such as changes in unemployment rate (UNEMPD) and growth rate of the initial unemployment insurance claims (UICLAIM), might contain useful information. However, other employment related variables, those variables

Table 3: Variable Rankings with Univariate Probit Models

QPS Ranking		*LPS Ranking*		*KS Ranking*	
CLI11	0.2094	TS10YFF	0.3446	TS10YFF	0.0000
TS10YFF	0.2106	CLI11	0.3462	TS10Y3M	0.0000
SP500	0.2144	SP500	0.3562	UNEMPD	0.0000
TS10Y3M	0.2211	TS10Y3M	0.3685	EXPD	0.0000
CPI	0.2220	EXP	0.3798	EMP	0.0000
EXP	0.2236	CPI	0.3817	IP	0.0000
UNEMPD	0.2266	UICLAIM	0.3879	SALES	0.0000
3M	0.2271	UNEMPD	0.3880	UNEMP	0.0000
FF	0.2271	EXPD	0.3893	CEMP	0.0000
EXPD	0.2285	EMP	0.3914	HOURS	0.0000
UICLAIM	0.2289	IP	0.3921	M0	0.0000
IP	0.2298	SALES	0.3961	A	0.0000
EMP	0.2314	UNEMP	0.3961	FF	0.0000
UNEMP	0.2314	CEMP	0.3978	HOUSE	0.0000
A	0.2320	HOURS	0.3999	AA	0.0000
SALES	0.2323	M0	0.4008	AAA	0.0000
CEMP	0.2325	A	0.4008	VENDOR	0.0000
AA	0.2329	FF	0.4018	3M	0.0000
AAA	0.2334	HOUSE	0.4020	M1	0.0000
HOURS	0.2334	AA	0.4030	10Y	0.0000
M0	0.2336	AAA	0.4049	5Y	0.0000
HOUSE	0.2344	VENDOR	0.4063	CSA	0.0000
5Y	0.2344	3M	0.4121	CSAA	0.0000
10Y	0.2350	M1	0.4124	CSAAA	0.0000
VENDOR	0.2364	INCOME	0.4138	M2	-0.0012
CSAAA	0.2409	10Y	0.4142	CLI11	-0.0048
CSAA	0.2424	5Y	0.4200	EXP	-0.0048
CSA	0.2431	CSA	0.4318	CPI	-0.0048
M1	0.2447	CSAA	0.4616	UICLAIM	-0.0048
INCOME	0.2465	M2	0.4798	INCOME	-0.0096
TS10Y5Y	0.2528	CSAAA	0.5110	TS10Y5Y	-0.0144
M2	0.2955	TS10Y5Y	0.6059	SP500	-0.0287

that the NBER Business Cycle Dating Committee is paying attention to, and monetary aggregates have considerably poorer results.

Although the rankings of out-of-sample forecasting accuracy based on the QPS and LPS are broadly consistent each other, some discrepancy results from the fact that the LPS penalizes larger mistakes more heavily. For instance, the levels of short-term interest rate, such as the Federal Funds rate (FF) and the 3-month Treasury bill yield (3M) are somewhat informative based on the

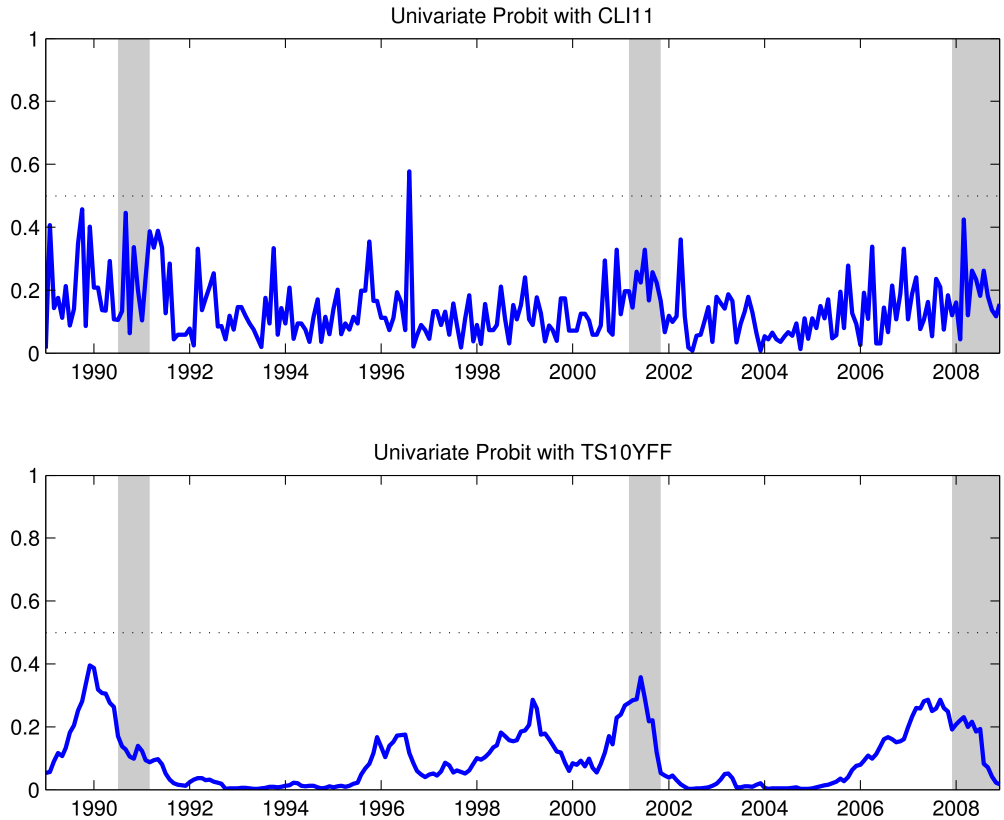

Figure 1: Out-of-Sample Predictions from Univariate Probit

Note: The shaded areas represent the NBER recessions. Horizontal axes measures recession probabilities.

QPS, the LPS suggests that they are relatively poor predictors.

Figure 1 compares six-month-ahead out-of-sample predictions that are selected as the best based on the QPS and LPS. The shaded areas indicate the NBER recession periods. Visual inspection of these picture suggests a potential pitfalls in relying on one particular measure of forecasting accuracy. Although predictions from CLI11 have the smallest errors in terms of the quadratic loss, they are not useful forecasts at all in practice, since they appear to be just noisy signals and they do not show clear contrasts between recessions and expansions. On the other hand, out-of-sample predictions with TS10YFF have better distinctions about the state of the economy.

Although overall performance is better than others, there are a couple of issues with predictions with TS10YFF. For the 1990-91 recession, the predictions made by the term spreads miss the timing of the recession and they do not have strong signals. This difficulty is also noted in other studies. See Stock and Watson (2003) for more detailed discussions on this issue. Furthermore, the magnitude of the recession signal is not strong enough, so that it is difficult to distinguish between a true signal and a false alarm. Peaks in the recession probability predictions never exceed 50%. For example, it is extremely difficult in real time to distinguish these two humps during the 1990's from those for the true signal for the 2001 recession. Overall pattern of this out-of-sample predictions is promising. However, it would be better if predictions exhibit stronger signals during true recessions. This gives us an incentive to pay attention to another measure of forecasting accuracy, the KS, which takes account of the accuracy and strength of predictions.

The out-of-sample forecasting performance evaluated by using the KS gives us a completely different picture. Most predictor variables have zeros for the KS. In other words, they usually do not give us strong predictions about the occurrence of future recessions or correct predictions are largely offset by false predictions. In the worst case, we will get more false signals, which indicate more than 50% probability of a recession during an expansion, than correct ones. Such a situation happens to many of predictors that are considered to be good based on the QPS and LPS, such as CLI11, SP500, EXP, and CPI.

3.2 Other Univariate Models

Now we will look at the performance of univariate models based on alternative CDFs, which allow positive skewness and/or excess kurtosis. Table 4 summarizes the ranking of univariate models based on out-of-sample fit.[8]

As in the probit models, the QPS and LPS suggest that those models with CLI11, TS10YFF, SP500, and TS10Y3M have relatively good forecasting performance. For CLI11, while adding skewness marginally improves forecasting accuracy in the out-of-sample forecasting exercises compared

[8]Due to space limitation, hereafter top 20 forecasting models are shown.

Table 4: Top 20 Univariate Models

QPS Ranking			*LPS Ranking*			*KS Ranking*		
CDF	x_1	QPS	CDF	x_1	LPS	CDF	x_1	KS
Gumbel	CLI11	0.2090	Normal	TS10YFF	0.3446	Laplace	TS10Y3M	0.0597
GEV3	CLI11	0.2090	GEV3	CLI11	0.3450	Laplace	CLI11	0.0549
Normal	CLI11	0.2094	Gumbel	CLI11	0.3451	Laplace	EXP	0.0083
Logistic	CLI11	0.2104	Normal	CLI11	0.3462	Gumbel	SP500	0.0035
Student-t	CLI11	0.2104	GEV3	TS10YFF	0.3465	GEV3	SP500	0.0035
Normal	TS10YFF	0.2106	Logistic	TS10YFF	0.3471	Laplace	SP500	0.0035
GEV3	TS10YFF	0.2108	Student-t	CLI11	0.3472	Normal	TS10YFF	0.0000
Logistic	TS10YFF	0.2118	Logistic	CLI11	0.3472	GEV3	TS10YFF	0.0000
Gumbel	TS10YFF	0.2119	Student-t	TS10YFF	0.3477	Logistic	TS10YFF	0.0000
Student-t	TS10YFF	0.2120	Laplace	TS10YFF	0.3493	Gumbel	TS10YFF	0.0000
Gumbel	SP500	0.2121	Laplace	CLI11	0.3507	Student-t	TS10YFF	0.0000
Laplace	CLI11	0.2127	Gumbel	TS10YFF	0.3521	Laplace	TS10YFF	0.0000
Laplace	TS10YFF	0.2127	Gumbel	SP500	0.3521	Normal	TS10Y3M	0.0000
GEV3	SP500	0.2132	GEV3	SP500	0.3538	GEV3	TS10Y3M	0.0000
Normal	SP500	0.2144	Normal	SP500	0.3562	Gumbel	TS10Y3M	0.0000
Student-t	SP500	0.2162	Logistic	SP500	0.3591	Logistic	TS10Y3M	0.0000
Logistic	SP500	0.2162	Student-t	SP500	0.3591	Student-t	TS10Y3M	0.0000
Laplace	SP500	0.2209	Laplace	SP500	0.3659	GEV3	EXP	0.0000
Normal	TS10Y3M	0.2211	Gumbel	TS10Y3M	0.3679	Gumbel	EXP	0.0000
GEV3	TS10Y3M	0.2215	Normal	TS10Y3M	0.3685	Laplace	UNEMPD	0.0000

to the probit model, allowing higher excess kurtosis alone deteriorates performance. Especially, the CLI11 with Laplace, which has the highest excess kurtosis with zero skewness is the worst among the models with CLI11. In fact, it is worse than the second-best predictor in probit models, TS10TFF. We can observe a similar pattern in those models with SP500.

Within the class of univariate models with a term spread measure, Normal works better than alternative CDFs. In fact, according to the LPS, Normal with TS10YFF has the best result.

There are some role played by using non-Normal CDFs in improving forecasting accuracy. It is not easy to generalize the role of different functional forms and the benefit of allowing skewness and/or excess kurtosis may be variables specific. However, we can see clear effects of using non-Normal CDFs if we look at the ranking based on the KS. Within the class of univariate models, most of models scores non-positive values for the KS. In other words, most of predictors do not make correct and strong predictions, in terms of the 50% probability cutoff. However, by using non-Normal CDF we can improve the KS. There are six models with positive values for the KS. Especially, using Laplace is likely to contribute to producing stronger and more accurate signals, net of false alarms. Especially, CLI11, EXP and SP500 all have negative values for the KS in the probit models. Using non-Normal CDF can alter bad performance into better predictions.

3.3 Bivariate Models

Now we turn out attention to forecasting results based on bivariate models in order to see the importance of additional information. Table 5 shows the top 20 bivariate forecasting models based on the QPS, LPS, and KS.

The combination of TS10YFF and SP500 is the best predictor variables based on the QPS and LPS. Both the QPS an LPS suggest that it is important to include a measure of term spread (either TS10YFF or TS10Y3M) and the relative importance of CLI11 decreases. Interestingly, there are some variables that have relatively poor performance in univariate models and produce good out-of-sample forecasting accuracy, *together with* the term spread measure. According to the rankings based on the univariate models, SP500 and EMP are not a useful single predictor. Especially, EMP has relatively poor out-of-sample forecasting performance. However, all three measures of forecasting accuracy suggest that they are an important companion variable to other variables. In fact, that the best bivariate model is not a combination of the two best single predictors (CLI11 and TS10YFF). This suggests that the univariate ranking is not a helpful guide for choosing multiple predictors.

The importance of SP500 together with a term spread measure is consistent with the finding of Estrella and Mishkin (1998). In our bivariate models, even without being combined with the term spread, some bivariate models that contain SP500 perform relatively well. However, it should be mentioned that King, Levin, and Perli (2007) do not find superiority of SP500 in conjunction with the term spread. Rather, they report that a combination of variables, which have better

Table 5: Top 20 Bivariate Models

QPS Ranking				*LPS Ranking*				*KS Ranking*			
CDF	x_1	x_2	QPS	CDF	x_1	x_2	LPS	CDF	x_1	x_2	KS
Gumbel	TS10YFF	SP500	0.1828	Normal	TS10YFF	SP500	0.3051	Logistic	EXP	UNEMP	0.1648
GEV3	TS10YFF	SP500	0.1833	Logistic	TS10YFF	SP500	0.3062	Student-t	EXP	UNEMP	0.1648
Normal	TS10YFF	SP500	0.1844	Student-t	TS10YFF	SP500	0.3064	Laplace	EXP	UNEMP	0.1553
Student-t	TS10YFF	SP500	0.1846	GEV3	TS10YFF	SP500	0.3064	Gumbel	TS10Y3M	SP500	0.1195
Logistic	TS10YFF	SP500	0.1847	Laplace	TS10YFF	SP500	0.3091	GEV3	TS10Y3M	SP500	0.1195
Laplace	TS10YFF	SP500	0.1866	GEV3	TS10YFF	EMP	0.3112	Normal	TS10Y3M	SP500	0.1195
Gumbel	TS10Y3M	SP500	0.1883	Gumbel	TS10YFF	SP500	0.3114	Logistic	TS10Y3M	SP500	0.1195
GEV3	TS10Y3M	SP500	0.1885	Normal	TS10YFF	EMP	0.3114	Student-t	TS10Y3M	SP500	0.1195
Normal	TS10Y3M	SP500	0.1886	Logistic	TS10YFF	EMP	0.3135	Laplace	TS10Y3M	SP500	0.1195
Logistic	TS10Y3M	SP500	0.1896	Student-t	TS10YFF	EMP	0.3139	Normal	TS10Y3M	EMP	0.1003
Student-t	TS10Y3M	SP500	0.1897	Gumbel	TS10Y3M	EXP	0.3140	GEV3	TS10Y3M	EMP	0.1003
Laplace	TS10Y3M	SP500	0.1911	Gumbel	TS10YFF	EMP	0.3150	Gumbel	TS10Y3M	EMP	0.0955
Gumbel	CLI11	SP500	0.1947	Gumbel	TS10Y3M	SP500	0.3151	Laplace	UICLAIM	SP500	0.0955
GEV3	CLI11	SP500	0.1954	GEV3	TS10Y3M	SP500	0.3160	Gumbel	EMP	SP500	0.0955
Laplace	EXP	UNEMP	0.1954	Normal	TS10Y3M	SP500	0.3165	GEV3	EMP	SP500	0.0955
Normal	TS10YFF	EMP	0.1959	Laplace	TS10YFF	EMP	0.3184	Normal	EMP	SP500	0.0955
GEV3	TS10YFF	EMP	0.1962	GEV3	TS10Y3M	EXP	0.3187	Student-t	EMP	SP500	0.0955
Normal	CLI11	SP500	0.1964	Logistic	TS10Y3M	SP500	0.3188	Logistic	EMP	SP500	0.0955
Logistic	TS10YFF	EMP	0.1966	Student-t	TS10Y3M	SP500	0.3193	Logistic	CLI11	SALES	0.0920
Student-t	TS10YFF	EMP	0.1967	Gumbel	CLI11	SP500	0.3209	Student-t	CLI11	SALES	0.0920

performance in univariate models, also perform better in bivariate models.[9]

Based on the QPS, allowing positive skewness and excess kurtosis (Gumbel and GEV3) outperform the probit counterparts. However, introducing only excess kurtosis (Logistic, Student-t, and Laplace) worsens the QPS, compared with the probit. However, for the best bivariate models with TS10YFF and SP500, utilizing non-Normal CDFs deteriorates forecasting accuracy, based on the LPS. In other words, it tends to some improvements in the sense of the quadratic loss, but at the same time, it is likely to produce larger mistakes. Especially, positive skewness makes the deterioration worse.

Although forecasting performance generally improves by adding one more predictor, probably the biggest gain appears in the KS. In univariate models, only 6 models out of 192 have positive scores. In bivariate models, 294 models out of 3168 have at least positive values for the KS. Furthermore, the magnitude of the KS improves significantly. This means that by incorporating additional variables, we are likely to produce more accurate and stronger signals. Although the variable combination of TS10YFF and SP500 is the best based on the QPS and LPS, it does not perform well in terms of the KS (KS = 0.0323 for Normal and KS = 0.0323 for Gumbel). Instead of using TS10YFF, the term spread with 3-month Treasury bill yields, together with SP500, gives us better results for the KS and provides consistently good performance across different forecasting accuracy measures. Relatively consistent performance can be also observed in the variable combination of TS10Y3M and EMP as well.

3.4 Trivariate Models and Overall Rankings

Now we move on to the overall forecasting performance, including all of univariate, bivariate, and trivariate models with six different CDFs. Tables 6 lists the top 20 forecasting models out of

[9]This could be because of a couple of reasons. First, it could be attributed to the difference in periods used for the out-of-sample forecasting exercises and forecasting horizon. Second, it might be because of the fact that my data set does not include the credit spread measures that they use and perform quite well in their univariate models. Finally, it could be due to the difference in evaluating out-of-sample forecasting performance. They look at average out-of-sample predictions over two test periods, the 2001 recession and the post-2001 expansion.

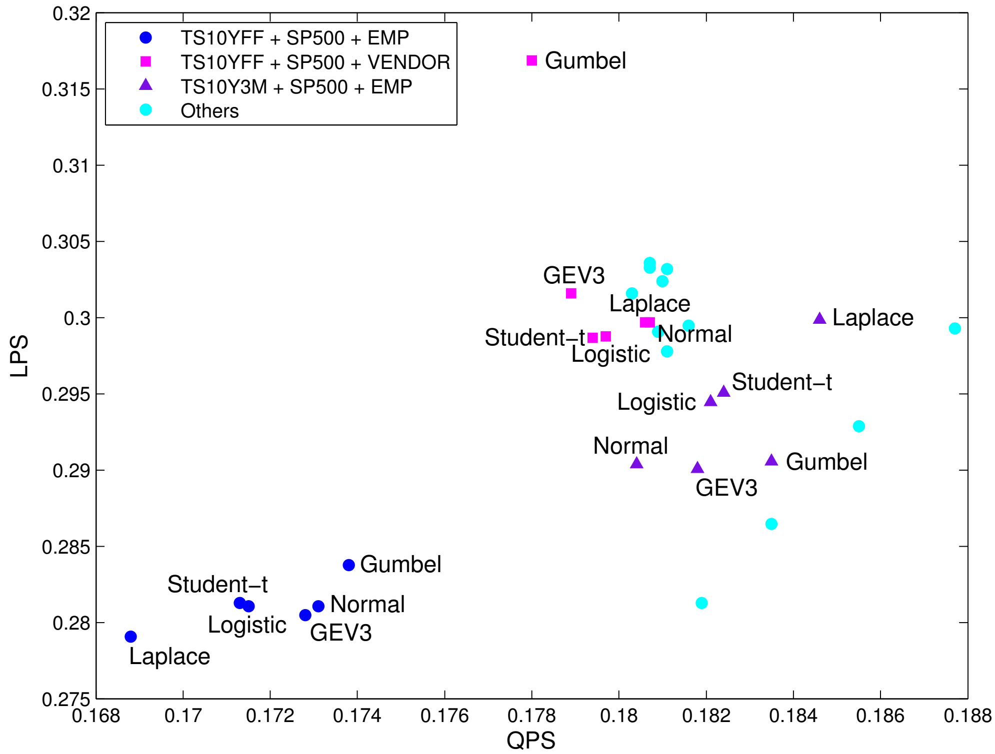

Figure 2: Overall Performance of Selected Models in Out-of-sample Forecasting Exercises

Note: Horizontal axis takes QPS and vertical axis measures the LPS. For both measures, the smaller the values are, the better the forecasting performance is. Thus, a southwest corner of the plot corresponds to the best forecasting model based on the QPS and the LPS. Top 20 models based on the QPS and LPS are plotted.

32928 forecasting models, based on different criteria and Figure 2 summarizes and compares overall performance of selected models.

According to the out-of-sample forecasting performance, based on the QPS and the LPS, the Laplace with TS10YFF, SP500, and EMP is the best forecasting model. Regardless of the functional form used, this variable combination records outstanding out-of-sample performance. This results is consistent with the fact that in bivariate models, SP500 and EMP are found to be good companion variables to TS10YFF. Other models with TS10YFF, SP500, and VENDOR (vendor performance index), and those with TS10Y3M, SP500, and EMP also show relatively good forecasting performance based on the QPS and LPS.

Table 6: Top 20 Models from Out-of-sample Forecasting Exercises

QPS Ranking					*LPS Ranking*				
CDF	x_1	x_2	x_3	QPS	CDF	x_1	x_2	x_3	LPS
Laplace	TS10YFF	SP500	EMP	0.1688	Laplace	TS10YFF	SP500	EMP	0.2791
Student-t	TS10YFF	SP500	EMP	0.1713	GEV3	TS10YFF	SP500	EMP	0.2805
Logistic	TS10YFF	SP500	EMP	0.1715	Logistic	TS10YFF	SP500	EMP	0.2811
GEV3	TS10YFF	SP500	EMP	0.1728	Normal	TS10YFF	SP500	EMP	0.2811
Normal	TS10YFF	SP500	EMP	0.1731	Student-t	TS10YFF	SP500	EMP	0.2813
Gumbel	TS10YFF	SP500	EMP	0.1738	Gumbel	TS10Y3M	SP500	EXP	0.2813
Gumbel	TS10YFF	SP500	VENDOR	0.1780	Gumbel	TS10YFF	SP500	EMP	0.2838
GEV3	TS10YFF	SP500	VENDOR	0.1789	GEV3	TS10Y3M	SP500	EXP	0.2865
Student-t	TS10YFF	SP500	VENDOR	0.1794	GEV3	TS10Y3M	SP500	EMP	0.2901
Logistic	TS10YFF	SP500	VENDOR	0.1797	Normal	TS10Y3M	SP500	EMP	0.2904
Gumbel	TS10YFF	SP500	CEMP	0.1803	Gumbel	TS10Y3M	SP500	EMP	0.2906
Normal	TS10Y3M	SP500	EMP	0.1804	Normal	TS10Y3M	SP500	EXP	0.2929
Laplace	TS10YFF	SP500	VENDOR	0.1806	Logistic	TS10Y3M	SP500	EMP	0.2945
Gumbel	TS10YFF	SP500	TS10Y3M	0.1807	Student-t	TS10Y3M	SP500	EMP	0.2951
Gumbel	FF	SP500	5Y	0.1807	Gumbel	TS10YFF	SP500	EXP	0.2978
Normal	TS10YFF	SP500	VENDOR	0.1807	Student-t	TS10YFF	SP500	VENDOR	0.2987
GEV3	TS10YFF	SP500	CEMP	0.1809	Logistic	TS10YFF	SP500	VENDOR	0.2988
Gumbel	TS10YFF	SP500	CLI11	0.1810	GEV3	TS10YFF	SP500	CEMP	0.2991
Gumbel	TS10YFF	SP500	UNEMPD	0.1811	Logistic	TS10Y3M	SP500	EXP	0.2993
Gumbel	TS10YFF	SP500	EXP	0.1811	Student-t	TS10YFF	SP500	CEMP	0.2995

KS Ranking				
CDF	x_1	x_2	x_3	KS
Gumbel	TS10Y3M	SP500	EMP	0.2795
GEV3	TS10Y3M	SP500	EMP	0.2712
Logistic	TS10Y3M	SP500	EMP	0.2712
Student-t	TS10Y3M	SP500	EMP	0.2712
Laplace	UNEMP	EXP	EMP	0.2520
Logistic	UNEMP	SP500	EMP	0.2473
Student-t	UNEMP	SP500	EMP	0.2473
Normal	TS10Y3M	SP500	EMP	0.2389
Laplace	UNEMP	SP500	EMP	0.2329
Gumbel	UNEMP	SP500	EMP	0.2150
GEV3	UNEMP	SP500	EMP	0.2150
Laplace	UNEMP	EXP	M2	0.2138
GEV3	UNEMP	EXP	EMP	0.2067
Normal	UNEMP	EXP	EMP	0.2019
Laplace	TS10Y3M	SP500	EMP	0.1971
Laplace	UNEMP	EXP	UNEMPD	0.1971
Laplace	UNEMP	EXP	IP	0.1971
Student-t	UNEMP	EXP	EMP	0.1971
Logistic	UNEMP	EXP	EMP	0.1971
Laplace	UNEMP	SP500	EXP	0.1923

For the variable combination of TS10YFF, SP500, and EMP, there are some roles played by positive skewness and/or excess kurtosis in improving the forecasting accuracy. In terms of the QPS, allowing higher excess kurtosis without positive skewness is likely to outperform the probit counterpart. With the LPS, it is hard to see some clear patterns but there is non-trivial difference between the Laplace model and the probit model.

However, it the term spread measure is replaced with TS10Y3M, the role of non-Normal CDFs is completely flipped. For these models with TS10Y3M, SP500, and EMP, allowing additional features in the functional form deteriorates forecasting accuracy. In other words, the differences in the measures of the term spread (and interactions with other variables) affects the usefulness of the different functional form. Although there is favorable evidence that allowing excess kurtosis and skewness help improve forecasting accuracy, such a possibility can be variable specific.

Out-of-sample forecasting accuracy evaluated by the KS suggests different ranking of models. The best models according to the KS utilize information contained in TS10Y3M, SP500, and EMP. Especially, the one with Gumbel achieves the best result. For this combination of variables, positive skewness and excess kurtosis appear to be very helpful to improve the out-of-sample forecasting accuracy in terms of the KS. Good forecasting performance based on the QPS and LPS usually is not supported by the KS performance and vice versa. Except for the combination of TS10Y3M, SP500, and EMP, the variable combinations that are ranked relatively high in the KS ranking do not have good out-of-sample forecasting performance in terms of the QPS or LPS. In other words, making strong predictions during recessions involves risks of making positive false signals during expansions, which worsens the QPS and LPS. In this sense, the variable combination TS10Y3M, SP500, and EMP might be preferred since it has relatively stable performance across different measures of forecasting accuracy.

Table 7: Importance of the Two Margins

Models	QPS			LPS			KS		
	Normal	Laplace	GEV3	Normal	Laplace	GEV3	Normal	Laplace	GEV3
Univariate	0.2106	0.2127	0.2108	0.3446*	0.3493	0.3465	0.0000	0.0000	0.0000
Bivariate	0.1844	0.1866	0.1833	0.3051*	0.3091	0.3064	0.0323	0.0275	0.0275
Trivariate	0.1731	0.1688*	0.1728	0.2811	0.2791*	0.2805	0.0920	0.1517	0.1565

Note: The univariate model includes TS10YFF. The bivariate model refers to models with TS10YFF and SP500 and the trivariate models add EMP to the bivariate models. Asterisks indicate the best model in each of univariate, bivariate, and trivariate models, based on different criteria.

3.5 Discussion

Forecasting performance based on all models indicates that the term spread measure (TS10YFF) is one of the most important variables, as widely documented in other studies. However, there are two margins to improve forecasting performance. One is to include additional predictors. By combining different information, we can increase accuracy of recession probability predictions. As demonstrated, a particular combination of variables greatly outperforms a typical univariate probit model with TS10YFF. Bivariate results suggest the importance of the stock price index (SP500) as a companion variable to the term spread measure. In addition, among the all combinations of variables considered in this paper, augmenting the term spread measure by SP500 and EMP results in the best out-of-sample forecasting performance. Another margin is to change a functional form of $F(\cdot)$, in order to improve how to translate signals to a recession probability measure.

It is important to understand how these two margins contribute to better forecasting performance. Table 7 summarizes how forecasting performance improves by changing the two margins and illustrates different effects of the margins. We compare probit models with the Laplace and GEV3 counterparts. Provided that a forecaster knows which one is the best companion variable, including additional predictors has greater impact on improving all three measures of forecasting accuracy. However, it is possible to obtain non-negligible gains from changing the functional form. Although up to the bivariate models, probit specification works better than any other CDFs, changing the functional form has have bigger impacts, once EMP is included in addition to TS10YFF

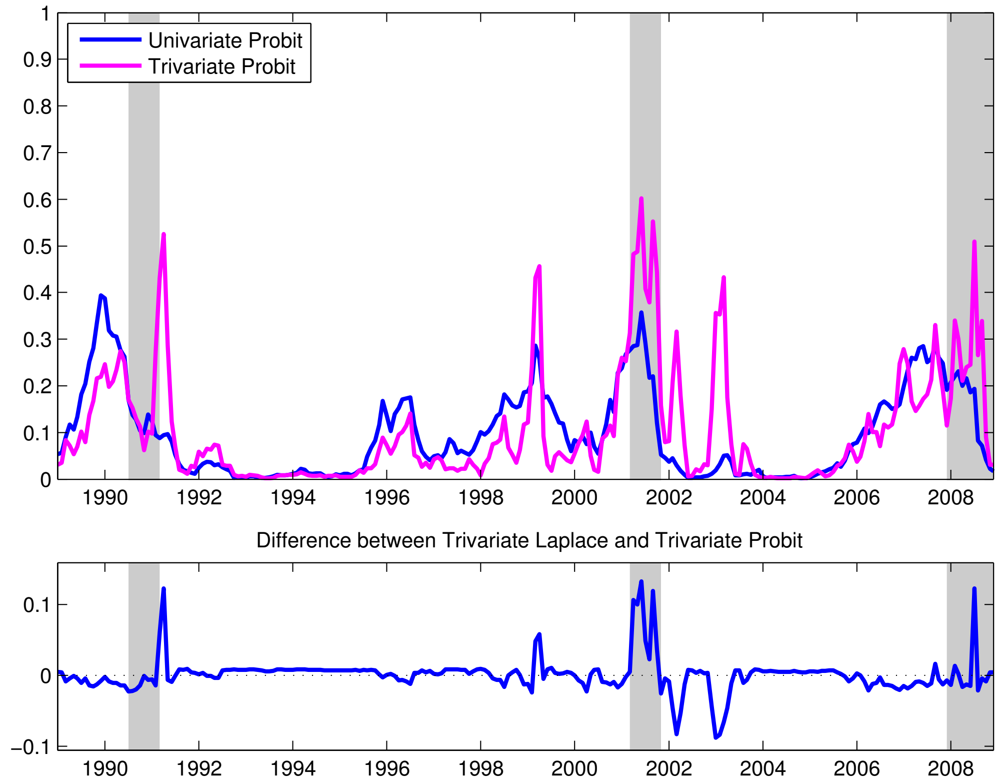

Figure 3: Comparison of Out-of-Sample Predictions

Note: Shaded areas indicate the NBER recession episodes. The top panel compares predictions from the univariate probit model with TS10YFF with those from the trivariate model with TS10YFF, SP500, and EMP. The bottom panel shows differences in predictions by changing the functional form. The bottom panel plots the difference between the trivariate Laplace and the trivariate probit.

and SP500.

The top panel of Figure 3 compares out-of-sample predictions from the univariate probit model with the term spread with the trivariate probit model (TS10YFF, SP500, and EMP). As can be seen, including additional variables reduces forecasting errors during expansions. Especially, predictions from the univariate model suggest moderate possibilities of recessions during the last half of the 1990's, when the economy was in fact in the longest expansion. Those positive false predictions are dampened by including additional variables. At the same time, it contributes to making stronger predictions during recessions, especially for the 2001 recession. However, in exchange for such improvements, the trivariate model also contains a couple of relatively strong false alarms during expansions, which are not present in the univariate model.

The bottom panel of Figure 3 plots differences in predictions between the trivariate Laplace model and the trivariate probit model. This suggests some benefits from changing functional form. Using the Laplace, instead of the Normal, will amplify positive signals during the 2001 recession and also will dampen strong false positive signals that happen right after the 1990-91 recession and before and after the 2001 recession. In general, using the Laplace will contribute to amplifying strong signals and to dampening weak signals, so that on average there will be less forecasting errors. The magnitude of such effects can be ± 10 percentage points, which make significant difference between the probit predictions and the Laplace predictions. However, such a feature works in an unfavorable way during the 1990-91 recession, since the underlying predictions are somewhat weak and miss the correct timing.

In order to further understand the role of additional variables, the middle and bottom panels of Figure 4 show contributions of SP500 and EMP, respectively. Contributions are calculated by taking difference between out-of-sample predictions from the trivariate model and predictions obtained by shutting down the corresponding coefficient to be zero. The top panel of Figure 4 compares out-of-sample predictions from the univariate Laplace with TS10YFF and the trivariate Laplace with TS10YFF, SP500, and EMP, indicating overall impact of including additional variables.

As we can see from the top panel, predictions solely based on the term spread usually fail to indicate the duration of recessions. Even when the timing of recession is more or less closer to the actual recession, the forecasted recession probability starts to decline before the recession ends. Plots on the middle and bottom panels suggest that including additional variables, especially SP500, help strengthen positive signals at the later stage of recessions. At the same time, additional information is useful in correcting false alarms during expansions. Especially, the contribution of EMP is coming from reducing false positive signals, rather than strengthening positive signals during recession. In this sense, combining information from different sources is helpful in amplifying

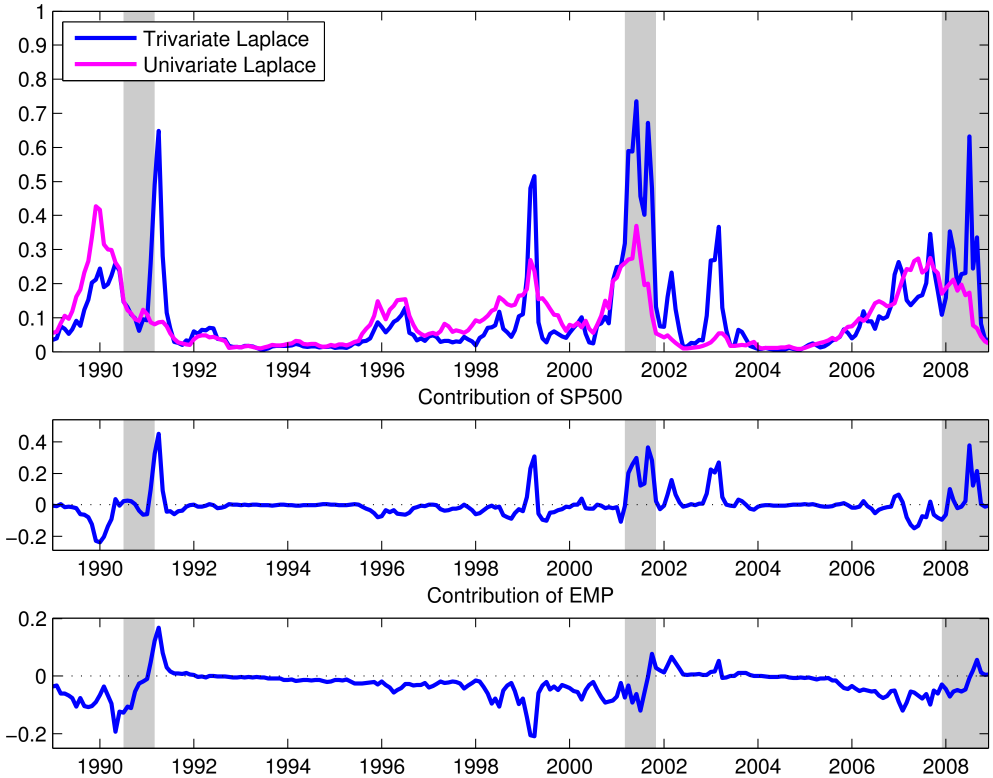

Figure 4: Understanding Contributions of Additional Variables

Note: Shaded areas indicate the NBER recession episodes. The top panel compares predictions from the univariate Laplace model with TS10YFF with those from the trivariate model with TS10YFF, SP500, and EMP. The middle and bottom panels plot contributions of the SP500 and EMP, respectively. Contributions are calculated by taking difference between out-of-sample predictions from the trivariate model and predictions obtained by setting the corresponding coefficient to be zero.

correct signals and discounting false signals.

Although there are some trade-offs in using different functional forms and in including additional variables, overall such modifications improve out-of-sample forecasts.

3.6 Robustness Check

As a robustness check, I also evaluate forecasting performance of all models by using an alternative scheme. It is possible that the selected forecasting models based on the recursive out-of-sample forecasting exercises do not outperform other competing models consistently over time. Such a problem could be sever when we search for a large number of alternative models. One of possibilities is the choice of periods for the out-of-sample forecasting exercises. Some predictors may have particularly good predictive power for a particular episode, but there is no guarantee that it will work quite well again for the next one in the future. Thus, it is important to carry out a robustness check. Table 8 reports the top 20 models from the leaving 2-years out exercises described in Section 2.3.

It turns out that the best variable combination based on the QPS and LPS in the out-of-sample forecasting exercises (TS10YFF, SP500, and EMP) also has the best performance in the all three different measures of forecasting accuracy in the leaving 2-years out exercises. Furthermore, among other models that have relatively good forecasting performance in the recursive out-of-sample forecasting exercises, only this variable combination shows consistently superior forecasting accuracy. Although those models with TS10YFF, SP500, and EMP do not have good results in terms of the KS in the out-of-sample forecasting exercises, this variable combination show strong performance even in terms of the KS in the leaving 2-years out exercises. Other models, such as those with TS10YFF, SP500, and VENDOR, or those with TS10Y3M, SP500, and EMP, do not survive in the leaving 2-years out exercises, implying that their performance is somewhat fragile against different recession episodes.

There is more tendency favorable for non-Normal CDFs, compared to the out-of-sample forecasting exercises. Especially, for the model with TS10YFF, SP500, and EMP, non-Normal CDFs outperform the Normal counterpart. Unlike the case in the out-of-sample case, allowing both excess kurtosis and positive skewness further improve the LPS.

Another thing we can observe from the leaving 2-years out exercises is that there are not so many trivariate models that outperform the best bivariate model. The bivariate Laplace model with TS10YFF and SP500 beat most of trivariate models in terms of the QPS and KS. Furthermore, the bivariate Gumbel model with TS10YFF and EMP shows better results than most of the trivariate

Table 8: Top 20 Models from the Leaving 2-years Out

QPS Ranking					*LPS Ranking*				
CDF	x_1	x_2	x_3	QPS	CDF	x_1	x_2	x_3	LPS
Laplace	TS10YFF	SP500	EMP	0.1149	Gumbel	TS10YFF	SP500	EMP	0.2064
Student-t	TS10YFF	SP500	EMP	0.1195	GEV3	TS10YFF	SP500	EMP	0.2182
Logistic	TS10YFF	SP500	EMP	0.1198	Laplace	TS10YFF	SP500	EMP	0.2194
Gumbel	TS10YFF	SP500	EMP	0.1210	Gumbel	TS10YFF	CLI11	EMP	0.2198
GEV3	TS10YFF	SP500	EMP	0.1216	Gumbel	TS10YFF	UICLAIM	EMP	0.2236
Normal	TS10YFF	SP500	EMP	0.1232	Student-t	TS10YFF	SP500	EMP	0.2236
Laplace	TS10YFF	SP500	IP	0.1256	Gumbel	TS10YFF	EXPD	EMP	0.2236
Logistic	TS10YFF	SP500	IP	0.1285	Gumbel	TS10YFF	TS10Y3M	EMP	0.2244
Student-t	TS10YFF	SP500	IP	0.1285	Gumbel	TS10YFF	EMP		0.2246
Laplace	TS10YFF	SP500	UICLAIM	0.1289	Gumbel	TS10YFF	CEMP	EMP	0.2247
Laplace	TS10YFF	EXPD	EMP	0.1290	Gumbel	TS10YFF	SALES	EMP	0.2249
Laplace	TS10YFF	SP500	HOUSE	0.1297	Logistic	TS10YFF	SP500	EMP	0.2249
Laplace	TS10YFF	SP500		0.1300	Gumbel	TS10YFF	IP	EMP	0.2250
Laplace	TS10YFF	SP500	CSA	0.1301	Gumbel	TS10YFF	HOURS	EMP	0.2251
Laplace	TS10YFF	SP500	SALES	0.1302	Gumbel	TS10YFF	HOUSE	EMP	0.2251
Laplace	TS10YFF	EMP	CLI11	0.1305	Gumbel	TS10YFF	INCOME	EMP	0.2253
Laplace	TS10YFF	SP500	HOURS	0.1306	Gumbel	TS10YFF	UNEMPD	EMP	0.2255
Gumbel	TS10YFF	TS10Y	SP500	0.1306	Gumbel	TS10YFF	TS10Y3M	SP500	0.2259
Normal	TS10YFF	SP500	IP	0.1306	Gumbel	TS10YFF	M0	EMP	0.2262
Student-t	TS10YFF	SP500	UICLAIM	0.1306	Gumbel	TS10YFF	SP500	CEMP	0.2263

KS Ranking				
CDF	x_1	x_2	x_3	KS
Laplace	TS10YFF	SP500	EMP	0.6856
Gumbel	TS10YFF	SP500	EMP	0.6606
Student-t	TS10YFF	SP500	EMP	0.6315
GEV+	TS10YFF	SP500	EMP	0.6221
Logistic	TS10YFF	SP500	EMP	0.6221
Laplace	TS10YFF	SP500	HOUSE	0.5964
Laplace	TS10YFF	SP500	UICLAIM	0.5910
Laplace	TS10YFF	TS10Y	SP500	0.5869
Laplace	TS10YFF	SP500	IP	0.5849
Normal	TS10YFF	SP500	EMP	0.5836
Laplace	TS10YFF	EMP	CLI11	0.5775
Gumbel	TS10YFF	SP500	IP	0.5755
Gumbel	TS10YFF	TS10Y3M	SP500	0.5734
Logistic	TS10YFF	SP500	IP	0.5734
Laplace	TS10YFF	SP500	HOURS	0.5734
Laplace	TS10YFF	SP500	CEMP	0.5734
Laplace	TS10YFF	SP500	CPI	0.5734
Student-t	TS10YFF	SP500	IP	0.5714
Gumbel	TS10YFF	SP500	HOUSE	0.5714
Laplace	TS10YFF	SP500		0.5714

Note: Results reported here is based on the leaving 2-years out exercises described in Section 2.3.

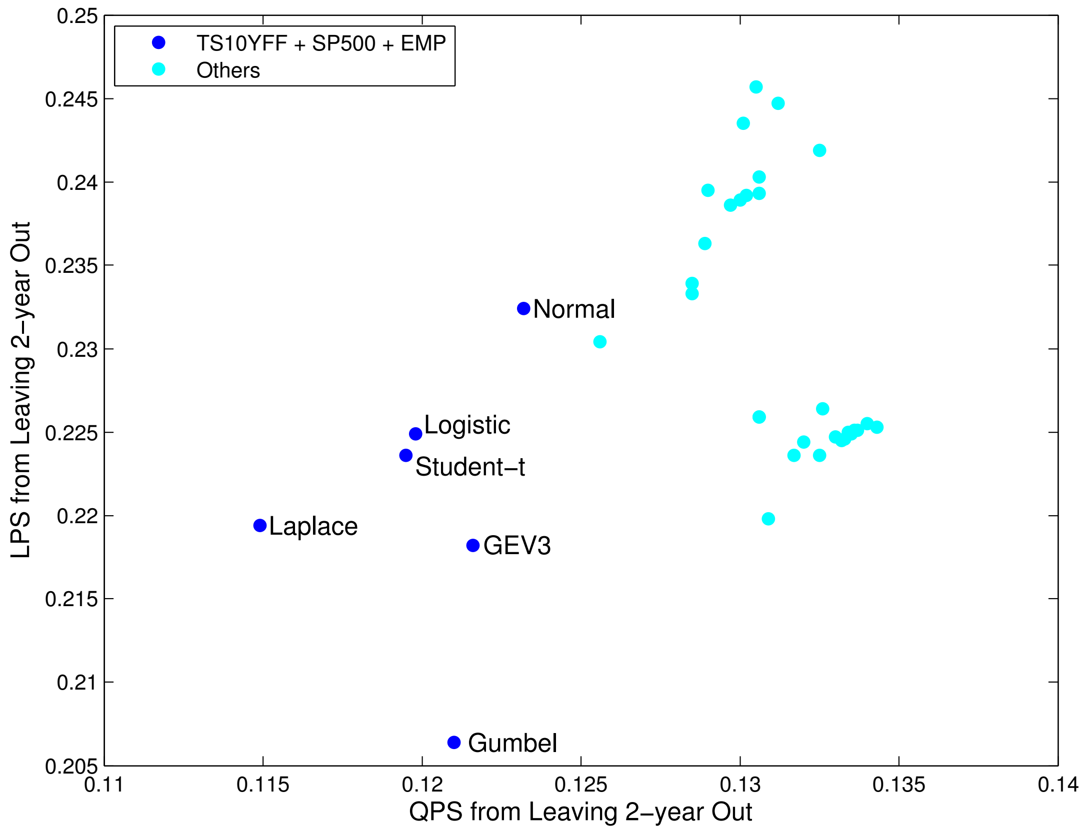

Figure 5: Robustness Check by the Leaving 2-years Out

models based on the the LPS.

4 Allowing More Flexible Skewness and Excess Kurtosis

4.1 Generalized Edgeworth Expansion

Results in the previous section suggest some benefits of allowing for positive skewness or excess kurtosis, in order to improve recession probability forecasts. However, there is some mixed evidence on what features in the functional form are important, compared with a typical probit model. It is possibly because of the fact that I just examine a relatively restrictive set of combinations of skewness and excess kurtosis. In this section, I will examine the effect of allowing for more flexibility in the functional form.

In particular, I will allow for more flexibility in skewness and excess kurtosis. In order to do so, I will utilize the generalized Edgeworth series expansion, which is introduced by Jarrow and Rudd (1982).[10] The generalized Edgeworth expansion (GEE) will approximate the "true" distribution $F(s)$ by using a series expansion involving higher moments of the approximating distribution $A(s)$. I assume that both $F(s)$ and $A(s)$ are continuous, that is,

$$\frac{dF(s)}{ds} = f(s) \quad \text{and} \quad \frac{dA(s)}{ds} = a(s) \tag{6}$$

exist. Furthermore, I assume that the moments of $F(s)$, $\alpha_j(F)$, exist for $j \leq n$. Given $\alpha_n(F)$ exists, the first $n-1$ cumulants $\kappa_j(F)$ from $j = 1, \cdots, n-1$ also exist. Jarrow and Rudd (1982) show that

$$f(s) \approx a(s) + \sum_{j=1}^{N} E_j \frac{(-1)^j}{j!} \frac{d^j a(s)}{ds^j}, \tag{7}$$

where E_j is a function of the cumulants of $F(s)$ and $A(s)$ up to the jth cumulant, which will be given below. By taking integral of (7), the "true" $F(x)$ can be approximated by:

$$F(x) \approx \int_{\infty}^{x} a(s)ds + \int_{\infty}^{x} \left[\sum_{j=1}^{N} E_j \frac{(-1)^j}{j!} \frac{d^j a(s)}{ds^j} \right] ds. \tag{8}$$

Since I am interested in having more flexibility in skewness and excess kurtosis, it is sufficient to set $N = 4$. The third and fourth cumulants are related to skewness and excess kurtosis, respectively, as

$$\kappa_3 = \gamma_1 (\kappa_2)^{3/2} \qquad \text{and} \qquad \kappa_4 = \gamma_2 (\kappa_2)^2, \tag{9}$$

where γ_1 and γ_2 denote skewness and excess kurtosis, respectively, and κ_2 denotes the 2nd cumulant

[10] They use the generalized Edgeworth expansion to study the problem of option valuation, where the underlying security distribution is unknown.

which corresponds to the variance. Then, E_j for $j = 1, \cdots, 4$ are given by:

$$E_1 = \kappa_1(F) - \kappa_1(A), \tag{10}$$

$$E_2 = \kappa_2(F) - \kappa_2(A) + E_1^2, \tag{11}$$

$$E_3 = \kappa_3(F) - \kappa_3(A) + 3E_1\left(\kappa_2(F) - \kappa_2(A)\right) + E_1^3, \tag{12}$$

$$\begin{aligned} E_4 = {} & \kappa_4(F) - \kappa_4(A) + 4\left(\kappa_3(F) - \kappa_3(A)\right)E_1 \\ & + 3\left(\kappa_2(F) - \kappa_2(A)\right)^2 + 6E_1^2\left(\kappa_2(F) - \kappa_2(A)\right) + E_1^4. \end{aligned} \tag{13}$$

Given a choice of the approximating distribution $A(x)$ (i.e., $\kappa_j(A)$ for $j = 1, \cdots, 4$), E_j is a function of the cumulants of $F(x)$ from $\kappa_1(F)$ to $\kappa_j(F)$. For my purpose of using this approximation in the binary response model, location and scale parameters are needed to be fixed for normalization. Otherwise, the coefficient vector $\boldsymbol{\beta}$ cannot be uniquely estimated. Without loss of generality, I assume that $F(x)$ have zero mean and unit variance and also assume that $A(x)$ has zero mean, so that $\kappa_1(F) = \kappa_1(A)$ and $E_1 = 0$. Then the approximation equation becomes:

$$\widetilde{F}(x) \equiv A(x) + \frac{E_2}{2!}\int_{\infty}^{x} \frac{d^2a(s)}{ds^2}ds - \frac{E_3}{3!}\int_{\infty}^{x} \frac{d^3a(s)}{ds^3}ds + \frac{E_4}{4!}\int_{\infty}^{x} \frac{d^4a(s)}{ds^4}ds. \tag{14}$$

Although it is easily shown that $\lim_{x\to\infty} \widetilde{F}(x) = 1$ and $\lim_{x\to-\infty} \widetilde{F}(x) = 0$, there is no guarantee that $\widetilde{F}(x)$ is monotonically increasing. In order for (14) being monotonically increasing, I need to impose the following condition:

$$a(x) + \sum_{j=2}^{4} E_j \frac{(-1)^j}{j!} \frac{d^j a(s)}{ds^j} \geq 0 \qquad \text{for all } x \in \mathbb{R}. \tag{15}$$

This condition (15) determines admissible combinations of skewness and excess kurtosis of $\widetilde{F}(x)$. This admissible set depends on a choice of $A(x)$. In order to cover all CDFs considered in the previous section, I will use the Logistic distribution whose mean equals zero and variance is 1.2 as the approximating distribution $A(x)$. Figure 6 illustrates the admissible set of (γ_1, γ_2) with the

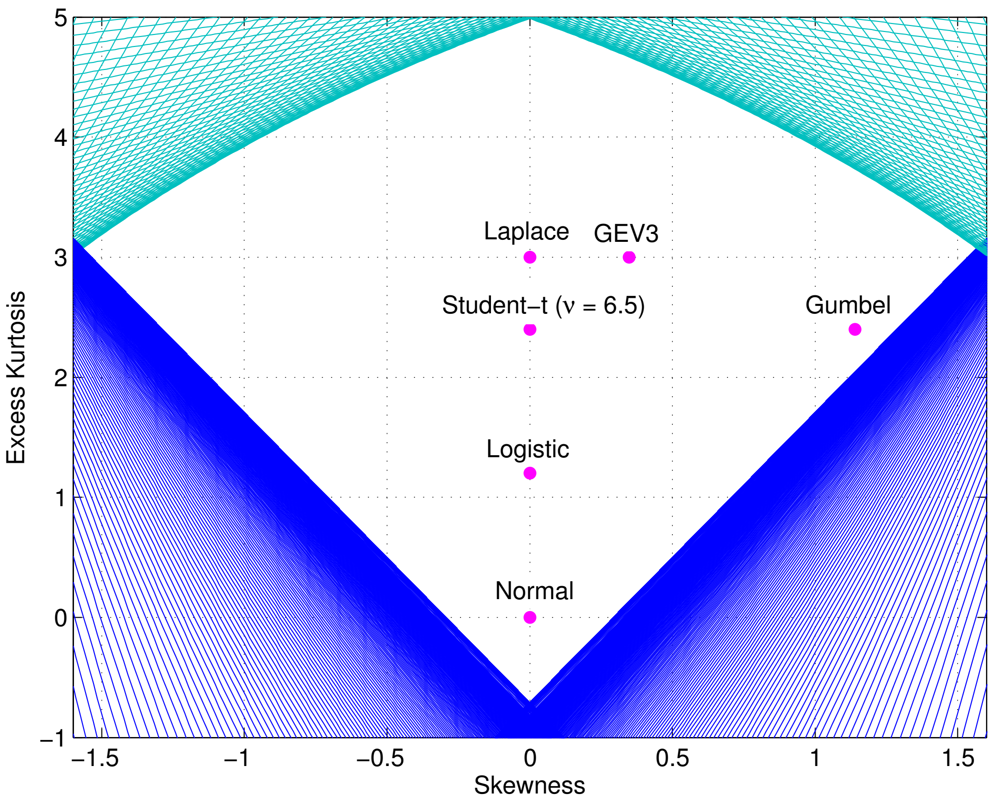

Figure 6: Admissible Set of Skewness and Excess Kurtosis with the Logistic ($\mu = 0, \sigma^2 = 1.2$)

Note: The white area bounded by lines approximates the admissible combinations of skewness and excess kurtosis that satisfy (15) when the Logistic with mean = 0 and $\sigma^2 = 1.2$ is used as the approximating distribution. This is drawn by discretizing x.

Logistic distribution, together with skewness and excess kurtosis of other CDFs.[11] Since I cannot estimate the variance of $A(x)$ due to an identification problem, I will restrict my search of skewness and excess kurtosis to combinations within this admissible set in Figure 6.

[11] Basically, the bottom corner of the admissible set depicted in Figure 6 moves up and the angle of the bottom corner becomes smaller as the variance increases. For example, with unit variance the area is slightly below the skewness and excess kurtosis of the Gumbel and with larger variance the admissible set passes by the Gumbel's combination.

4.2 The GEE Results

I will use (14) for $F(\cdot)$. Since E_3 and E_4 are functions of γ_1 and γ_2, respectively, there are two additional parameters to be estimated. This is a parsimonious way to cover a relatively large space of skewness and excess kurtosis.[12]

By using the GEE, I will estimate the best trivariate model, which includes TS10YFF, SP500, and EMP. For this variable combination, the Laplace model works better than others. It is of our interest to see whether we can obtain further improvement in the forecasting accuracy by exploring much wider space of the excess kurtosis and positive skewness.

Estimated values of γ_1 and γ_2 based on the variable combination of TS10YFF, SP500, and EMP are 1.4144 and 3.3264, respectively. The GEE estimates from the both trivariate models suggest to include both positive skewness and excess kurtosis. Furthermore, the degree of positive skewness and excess kurtosis is much higher than those incorporated in the GEV3 used in this paper. Figure 7 graphically compares a functional form estimated by the GEE with the Normal CDF. The GEE makes smaller predictions around $x = -1$. While the GEE predictions will be more aggressive than the probit ones at somewhere between $x = -0.5$ and 1, its prediction becomes modest around $x > 1$, given the same information.

Table 9 reports performance of predictions made by using the GEE. Regarding the out-of-sample forecasting performance, the GEE models always outperform the probit counterparts. Furthermore, the GEE models almost always have better out-of-sample forecasting performance than other non-Normal CDFs, except for the LPS, where the GEE and the Laplace have the same score. The largest gain appears in the KS. In other words, given the same information, the GEE produces more accurate and stronger predictions than other known CDFs. Although the GEE does not outperform the best trivariate model based on the KS (KS = 0.2795 for the Gumbel with TS10Y3M, SP500,

[12] Instead of using the GEE, we could use more complicated CDFs that can cover a much wider set of skewness and excess kurtosis, such as the one based on the skewed generalized t distribution of Theodossiou (1998). See Hansen, McDonald, and Theodossiou (2007) for other flexible distributions. However, there are basically two problems. First, those functions have more parameters and moments are complicated functions of those parameters, so that it is relatively difficult to control the identification issue in the binary response model. Second, it is often the case that there are no closed form CDFs available, which could increase computational burden.

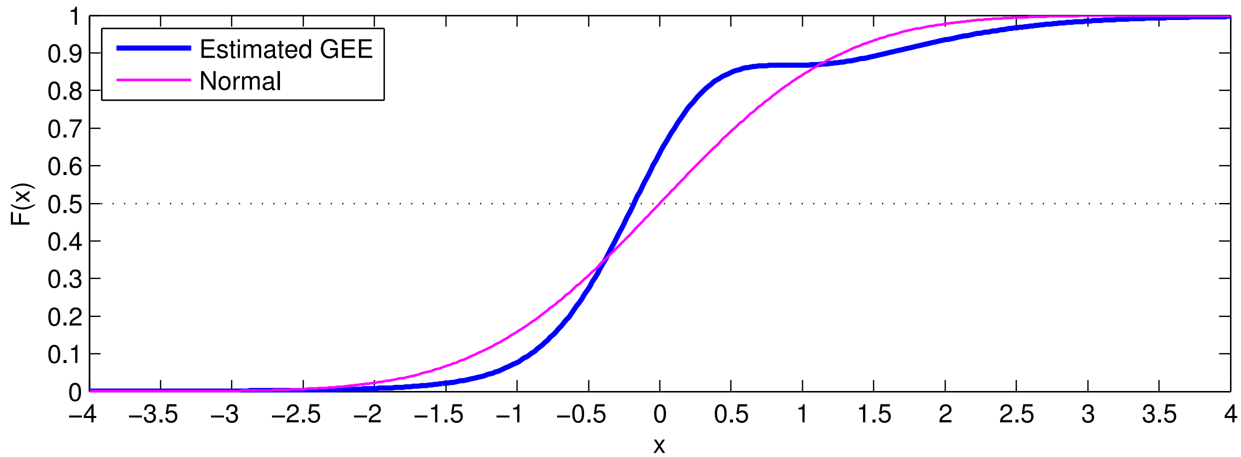

Figure 7: Comparison of Functional Forms

Table 9: Forecasting Accuracy with the Generalized Edgeworth Expansion

CDF	QPS	LPS	KS
GEE	0.1683 [0.1181]	0.2791 [0.2044]	0.2389 [0.6910]
Normal	0.1731 [0.1232]	0.2811 [0.2324]	0.0920 [0.5836]
Logistic	0.1715 [0.1198]	0.2811 [0.2249]	0.1565 [0.6221]
Student-t	0.1713 [0.1195]	0.2813 [0.2236]	0.1565 [0.6315]
Laplace	0.1688 [0.1149]	0.2791 [0.2194]	0.1517 [0.6856]
Gumbel	0.1738 [0.1210]	0.2838 [0.2064]	0.1469 [0.6606]
GEV3	0.1728 [0.1216]	0.2805 [0.2182]	0.1565 [0.6221]

Note: Results based on the know CDFs from the previous section are reproduced for comparison. Values reported in brackets are the corresponding values obtained from the leaving 2-years out exercises.

and EMP), using the GEE helps greatly overcome the weakness of the relatively small KS.

In addition to the recursive out-of-sample forecasting exercises, I also perform the leaving 2-years out exercises as a robustness check. Table 9 also reports results from the leaving 2-years out exercises in brackets. The results are robust. In the leaving 2-years exercises, the Lapalce and Gumbel models are two leading models. While the GEE has better predictions than the Gumbel model in all three measures, the GEE beats the Laplace counterpart in terms of the LPS and the KS, but not in the QPS. With only one exception, we can confirm the superiority of the GEE model over other models.

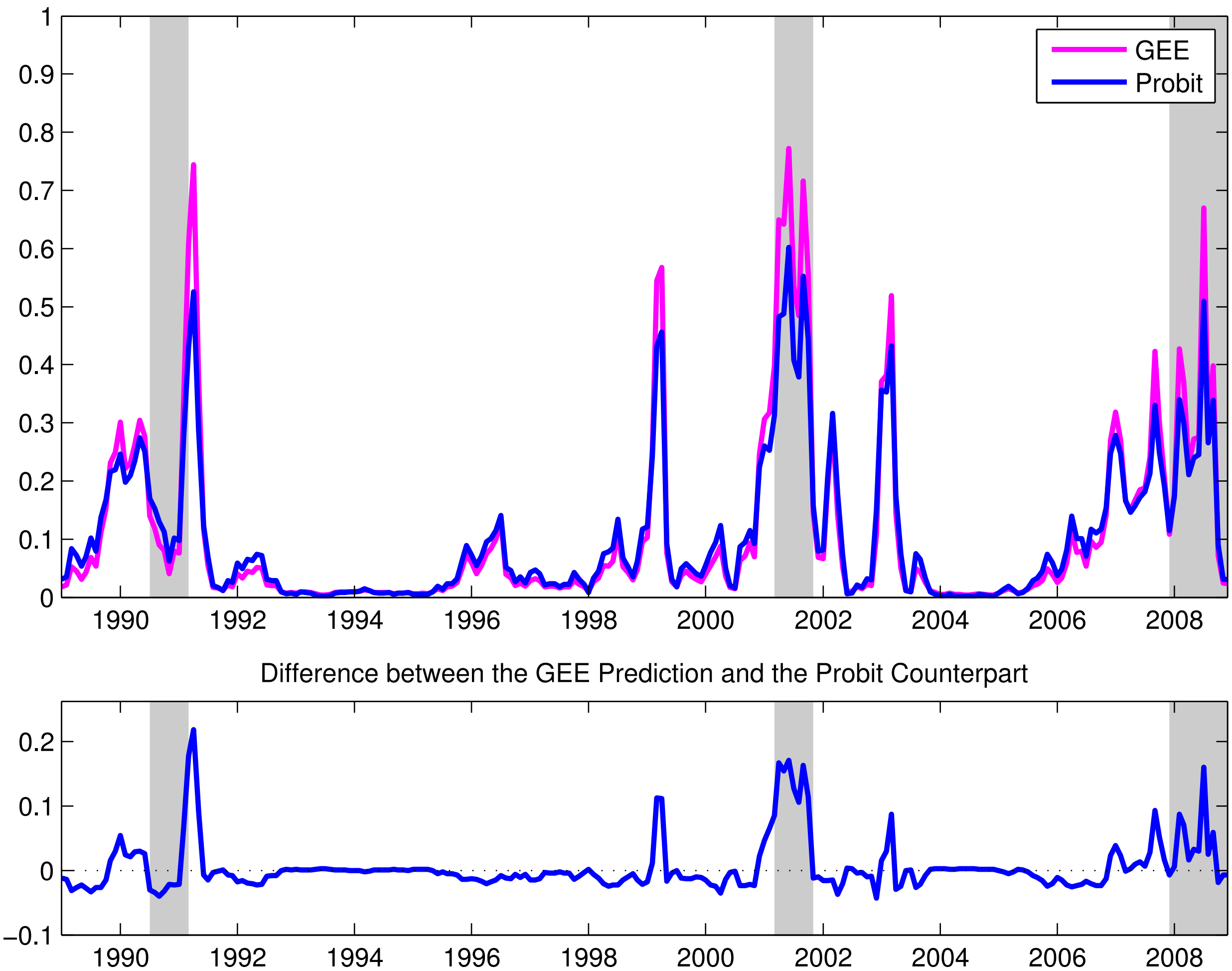

Figure 8: Out-of-Sample Predictions by the GEE

Figure 8 compares predictions made by the GEE model with those from the probit model. Predictions from the GEE model amplify the positive signals during recessions, especially for the 2001 recession and the current recession. This amplification is stronger than the one we can obtain from using the Laplace, instead of the Normal, reaching nearly 20 percentage points. For the 1990-91 recession, as in other models, the GEE model fails to predict the correct timing of the recession. The GEE also reduces false positive signals that arise from the trivariate probit model. However, the size of the correction is somewhat modest. Overall, this comparison suggests that most gains are obtained from making stronger and more accurate predictions, rather than effectively dampening false positive signals.

As we have seen, allowing flexibility in the functional form will effectively improve recession probability forecasts. Although there is a possibility to make false predictions worse, there are net gains in forecasting accuracy from using the GEE. Using more flexible functional form can be combined with other extensions to probit models, such as multiple breaks and dynamic probit models with lagged dependent variables, in order to achieve further improvements in forecasting accuracy.

5 Conclusion

Systematically comparing forecasting models of recession probability in the U.S. economy, I have found substantial gains in out-of-sample forecasting accuracy from including additional variables in a typical univariate probit model with the Treasury term spread. Furthermore, there is a room to further improve forecasting performance by utilizing non-Normal CDFs, which allow for positive skewness and/or excess kurtosis.

Among 32928 forecasting models considered, the combination of the term spread between 10-year Treasury yield and the Federal Funds rate, changes in the S&P 500, and non-farm employment growth outperforms other variable combinations in the out-of-sample forecasting exercise. Although its out-of-sample performance based on the KS is somewhat weak, the combination achieves the best results in both the QPS and LPS. Its superior forecasting performance is robust and the combination achieves the best results in the alternative leaving 2-years out forecasting exercise, in terms of the all three measures of forecasting accuracy.

In terms of the functional form, there is favorable evidence for using non-Normal CDFs. In the case of the best variable combination, the Laplace model seems to work better than other functional forms. Allowing excess kurtosis magnifies positive signals during recessions and dampens false positive signals during expansion. In general, however, it is not so clear what additional features may help improve forecasting accuracy and it is highly likely that such conclusion can be variable specific.

In order to allow for more flexibility in the functional form, I have applied the generalized Edgeworth expansion, which enables us to explore a wider set of skewness and excess kurtosis. By using the GEE, we can further improve forecasting accuracy by making stronger predictions. In the out-of-sample forecasting exercises, the GEE model outperforms the best model among those examined in the paper.

References

Birchenhall, C. R., H. Jessen, D. R. Osborn, and P. Simpson (1999): "Predicting U.S. Business-Cycle Regimes," *Journal of Business Economic Statistics*, 17(3), 313–323.

Chauvet, M., and J. D. Hamilton (2006): "Dating Business Cycle Turning Points," in *Nonlinear Time Series Analysis of Business Cycles*, ed. by C. Milas, P. Rothman, and D. van Dijk, vol. 276 of *Contributions to Economic Analysis*, chap. 1, pp. 1–54. Elsevier.

Chauvet, M., and S. Potter (2005): "Forecasting Recessions Using the Yield Curve," *Journal of Forecasting*, 24(2), 77–103.

Clements, M. P., and A. B. Galvão (2006): "Combining Predictors & Combining Information in Modelling: Forecasting US Recession Probabilities and Output Growth," in *Nonlinear Time Series Analysis of Business Cycles*, ed. by C. Milas, P. Rothman, and D. van Dijk, vol. 276 of *Contributions to Economic Analysis*, chap. 2, pp. 55 – 74. Elsevier.

Dotsey, M. (1998): "The Predictive Content of the Interest Rate Term Spread for Future Economic Growth," *Federal Reserve Bank of Richmond Economic Quarterly*, 84(3), 31–51.

Dueker, M. (2005): "Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of U.S. Recessions," *Journal of Business & Economic Statistics*, 23(1), 96–104.

Estrella, A., and F. S. Mishkin (1998): "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics*, 80(1), 45–61.

Estrella, A., A. P. Rodrigues, and S. Schich (2003): "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States," *Review of Economics and Statistics*, 85(3), 629–644.

Hansen, C. B., J. B. McDonald, and P. Theodossiou (2007): "Some Flexible Parametric

Models for Partially Adaptive Estimators of Econometric Models," Economics Discussion Papers 2007-13, Kiel Institute for the World Economy.

Hansen, P. R. (2008): "In-Sample and Out-of-Sample Fit: Their Joint Distribution and its Implications for Model Selection," Working paper, Stanford University.

Jarrow, R., and A. Rudd (1982): "Approximate Option Valuation for Arbitrary Stochastic Processes," *Journal of Financial Economics*, 10(3), 347–369.

Kane, T. (2008): "Employment Numbers as Recession Indicators," Discussion paper, Joint Economic Committee, U.S. Congress.

Kauppi, H., and P. Saikkonen (2008): "Predicting U.S. Recessions with Dynamic Binary Response Models," *Review of Economics and Statistics*, 90(4), 777–791.

King, T. B., A. T. Levin, and R. Perli (2007): "Financial Market Perceptions of Recession Risk," Finance and economics discussion series, Board of Governors of the Federal Reserve System.

Rudebusch, G. D., and J. C. Williams (2008): "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve," Federal Reserve Bank of San Francisco Working Paper Series.

Sephton, P. (2001): "Forecasting Recessions: Can We Do Better on MARS?," *Review, Federal Reserve Bank of St. Louis*, (Mar), 39–49.

Stock, J. H., and M. W. Watson (1989): "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 4, 351–394.

——— (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829.

Theodossiou, P. (1998): “Financial Data and the Skewed Generalized t Distribution,” *Management Science*, 44(12), 1650–1661.

Wright, J. H. (2006): “The Yield Curve and Predicting Recessions,” Finance and Economics Discussion Series.