C.I.R.P.É.E

Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **13-41**

Probability and Severity of Recessions

Rachidi Kotchoni

Dalibor Stevanovic

Novembre/November 2013

Version révisée/revised : Juin/June 2014

Kotchoni : African School of Economics
rachidi.kotchoni@gmail.com
Stevanovic : Université du Qébec à Montréal, Département des sciences économiques, 315, rue Ste-Catherine Est, Montréal, QC H2X 3X2; and CIRPÉE
dstevanovic.econ@gmail.com

We thank Yvon Fauvel, Mark Watson and Jonathan Wright for useful discussions and comments.

Abstract:
This paper tackles the prediction of the probability and severity of US recessions. We employ parsimonious Probit models to estimate the probability of a recession h periods ahead, for h varying between 1 and 8 quarters. A novel goodness-of-fit measure derived from the Kullback-Leibler Information Criterion is developed and used to select the regressors to include in the Probit models. Next, an autoregression (AR) augmented with inverse Mills ratio (IMR) and diffusion indices (DI) is fitted to selected measures of real economic activity. The resulting "IMR-DI-AR" model is used to generate forecasts conditional on optimistic and pessimistic scenarios for the horizon of interest. The severity of recessions is defined as the gap between the pessimistic scenario and the recent trend of the series. For a time series of GDP growth, our measure of recession severity has the interpretation of the output loss. Our results support that U.S. recessions are predictable to a great extent, both in terms of occurrence and severity. All recessions are not alike: some are more predictable than others while some are more severe than expected.

Keywords: Forecasting, Principal Components, Probit, Real Activity, Recessions

JEL Classification: C3, C5, C35, E27, E37

1 Introduction

The world economic history consists of an endless succession of business cycles characterized by swings across peaks and troughs of real economic activity. A period running between any given peak and the next trough is called a recession while a period between a trough and the next peak is an expansion. Although quite simple, this definition raises two practical issues. The first issue concerns the precise meaning of the expression "*real economic activity*". The Business Cycle Dating Committee of the National Bureau of Economic Research (NBER) does not provide a precise definition to this expression. Rather, the NBER defines a recession as "*a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales.*" The second issue concerns the identification of the peaks and troughs of real economic activity from observed data. The Business Cycle Dating Committee does provide a precise response to the latter issue by regularly publishing recession dates with six months to one year lag[1].

The general objective of this paper is to assess the predictability of recessions in the United States and construct forecasts that are conditional on assumptions about the state of the economy at the horizon of interest. More precisely, we seek to answer two questions. First, how likely is a recession to occur at a given forecast horizon? And second, how severe (or deep) is a recession expected to be if it were to effectively occur? We conduct an "in-sample" analysis based on historical data (final releases) [2].

The exercise which consists of predicting the probability of recessions is not new in the literature. (Stock & Watson 1989) used a probabilistic framework to construct a coincident and a leading index of economic activity as well as a recession index. The latter index is

[1] The announcement dates can be found at http://www.nber.org/cycles.html.

[2] Other issues are also of interest regarding recession, e.g., prediction of the business cycle turning points [(Chauvet 1998), (Chauvet & Hamilton 2006), (Chauvet & Piger 2008), (Stock & Watson 2010) or (Stock & Watson 2012*b*)], identification of the variables that lead future economic activity [(Stock & Watson 1989), (Issler & Vahid 2006), (Ng & Wright 2013)]. A natural extension of the approach developed in this paper could be to consider a nowcasting exercise based on real-time data.

nothing but "*the probability that the economy will be in a recession six months ahead, given data available through the month of its construction.*" (Estrella & Mishkin 1998) examined the individual performance of financial variables such as interest rates, spreads, stock prices, and monetary aggregates in predicting the probability of a recession. They found that stock prices are good predictors of recessions at one to three quarters horizon while the slope of the yield curve emerges as a better predictor beyond one quarter.

The forecasting power of the yield curve is also well documented in (Rudebusch & Williams 2009) who find that professional forecasters do not incorporate the information from the yield spread. (Anderson & Vahid 2001) applied nonlinear models to predict the probability of U.S. recession using the interest-rate spread and growth in M2 as leading indicators. Using (Fair 1993) definitions of a recession, they found that "*conditional on the spread, the marginal contribution of M2 growth in predicting the probability of recessions is negligible*"[3]. Wright (2006) estimated several Probit models and found that "*models that use both the level of the federal funds rate and the term spread give better in-sample fit and better out-of-sample predictive performance than models with the term spread alone.*" Christiansen, Eriksen, and Møller (2013) found that sentiment variables have predictive power beyond the standard financial series.

Overall, the prediction of the probability of recessions has been successfully tackled by several authors to various extent. Our contribution to this question resides in a novel methodology for formally testing the significance of the predictive power of the regressors included in a discrete choice probabilistic model. Indeed, traditional information criteria like the AIC and the BIC are relative measures of the quality of a model while pseudo R-squares are more or less "ad hoc" in the context of discrete choice probabilistic models. We propose an R-square type goodness-of-fit measure deduced from the Kullback-Leibler Information Criterion (KLIC). The proposed statistic, denoted R^2_{KLIC}, always lies between 0 and 1 and

[3](Fair 1993) defines a recession as either "at least two consecutive quarters of negative growth in real GDP over the next five quarters" or "at least two quarters of negative growth in real GDP over the next five quarters." This definition is not retained by the NBER.

it provides an absolute measure of goodness-of-fit. It is equal to 0 for a model with no explanatory power (i.e., constant probability model) and it converges to 1 as the predicted conditional probabilities diverge away from the sample proportions. Conditional probabilities are more informative than sample proportions. Therefore, a value of the R^2_{KLIC} far from 0 indicates that the regressors included in the probabilistic model are relevant for predicting the binary variable of interest. We obtain parsimonious Probit models by implementing a variable selection procedure that relies on the statistical significance of the R^2_{KLIC}.

The second question about the severity of recessions has not been much addressed in the literature. This severity can be measured either in terms of its duration or in terms of its impact on economic activity. The current paper focuses on the latter aspect. Thus, let y_t denote a real economic activity variable (e.g., GDP growth or unemployment rate), $R_t \in \{0, 1\}$ the indicator of recession at time t and X_t a set of potential predictors. Three different forecasts can be produced for y_{t+h}. The first forecast is based on an average scenario that does not explicitly depend on the outlook about a recession, i.e., $E(y_{t+h}|X_t)$, for $h \geq 1$. The second forecast is pessimistic as it assumes *a priori* that there will be a recession at time $t + h$, i.e., $E(y_{t+h}|X_t, R_{t+h} = 1)$. The third forecast given by $E(y_{t+h}|X_t, R_{t+h} = 0)$ is based on the optimistic assumption that there will be an expansion at period $t + h$.

All three forecasts can be computed ex ante regardless of the actual ex post realization of R_{t+h}. Also, the parameters needed to compute $E(y_{t+h}|X_t, R_{t+h} = 1)$ and $E(y_{t+h}|X_t, R_{t+h} = 0)$ can be inferred from a model that uses only the information available at time t. The *expected severity of a recession* is defined as the difference between the pessimistic scenario and the recent trend of the series. A recession that is expected to be quite severe *ex ante* may turn out to be mild *ex post* (and *vice versa*). Hence, we define the *realized severity of a recession* as the difference between the actual realization of $y_{i,t+h}$ and its recent trend.

The optimistic and pessimistic scenario forecasts can in principle be obtained by splitting the sample according to the values of R_{t+h}, as done for example in (Hamilton 2011). Here, we follow an alternative approach that involves inverse Mills ratio (IMR) corrections. We advocate an IMR-DI-AR model, i.e. an autoregressive model augmented with diffusion indexes and IMRs . Our model is reminiscent of, but different from, the Qual VAR model of (Dueker 2005) and (Dueker & Wesche 2005). The Qual VAR is a VAR system which includes a latent variable that governs the occurrence of a binary outcome. In the IMR-DI-AR model, the IMR is treated as exogenous vis-a-vis the lagged realizations of the variables included in the AR recursion. Moreover, the IMR-DI-AR allows us to generate optimistic and pessimistic forecasts, which is not readily possible with a Qual VAR.

We design an empirical framework in three steps. First, we combine a large number of potential predictors (observed on a quarterly basis) into as many principal components (PC) as possible. Our list of variables includes the most widely used predictors in the literature (yield, credit spreads, orders, housing, employment, stock market returns, etc.) as well as new candidates (realized volatility and skewness of the SP500 and DJIA indices). The PCs are synthetic variables with no structural meaning a priori. However, they can be interpreted by examining the variables to which they are correlated the most. Second, we estimate Probit models in which selected PCs lead the probability of a recession up to two years ahead. Third, we use the IMR-DI-AR model to predict six indicators of economic activity, namely the GDP growth, industrial production growth, unemployment, employment growth, inflation and SP500 returns. For each variable, optimistic, average and pessimistic forecasts are generated, which allows us to predict the severity of recessions along six dimensions.

Our results suggest that U.S. recessions are predictable to a great extent, both in terms of occurrence and severity. NBER recession dates are reasonably well predicted up to 5 quarters ahead. Our variable selection procedure suggest that employment growth, inflation, credit spreads, yield curve and stock market realized measures (returns, volatility and skewness) are the best predictors of future recessions. Some of these predictors have also been identified in (Ng 2013). The power of the PCs at predicting recessions has changed over time, which suggests that all recessions are not alike regarding their origin. The actual GDP growth rates were above the pessimistic scenario during the recessions of 1969, 1990 and

2001 while the recessions of 1973, 1980, 1981 and 2008 have been more severe than expected. It is interesting to note that some measures of macroeconomic uncertainty that have been proposed in the recent literature (e.g. see (Jurado, Ludvigson & Ng 2013)) exhibit peaks during those recessions where the realized severity has been worse than expected.

The remainder of the paper is organized as follows. Section 2 presents our framework. Section 3 presents the empirical application and results while Section 4 concludes. An appendix contains estimation outputs that are not shown in the main text.

2 The Framework

This section presents our empirical framework in details. The first subsection presents the Probit model used to predict the probability of recessions at a given horizon. The second subsection presents the derivation of the R^2_{KLIC}, i.e. our new goodness-of-fit measure for discrete probabilistic models. The third subsection shows our variable selection procedure based on the R^2_{KLIC} for Probit models. The fourth subsection presents the IMR-DI-AR model used to forecast real economic activity variables. The fifth subsection presents the variable selection procedure used for the IMR-DI-AR models. Finally, the sixth subsection discusses the measurement of the severity of recessions.

2.1 Predicting the Probability of Recessions using a Probit

Let R_t be a variable such that $R_t = 1$ if the NBER committee designates period t as a recession time and $R_t = 0$ otherwise. Assume that we have a large number of potential predictors of recession in hand, gathered in a N-dimensional vector X_t. Ideally, X_t should contain all relevant real economic activity indicators as well as macro-financial variables. The candidate predictors may be partially redundant or highly correlated (e.g., GDP deflator versus CPI inflation, or SP500 versus DJIA), but they should all be observable at time t or a few periods ahead of $t + h$, where h is the forecast horizon. In order to reduce the dimensionality of X_t and by the same token avoid multicollinearity issues, we consider summarizing the information content of X_t into a smaller number (q) of principal components F_t. By abuse of language, we may sometimes refer to F_t as factors although we do not pretend that that data obey a structural factor model. We interpret each factor F_t, ex post, by examining the five variables to which it is the most correlated. Subsequently, F_t is augmented with a constant variable so that $F_t \in \mathbb{R}^{q+1}$. To fix ideas, we assume that the data are observed on a quarterly basis.

To model the probability of a recession, we assume that there exist a latent leading index $Z_{h,t}$ which satisfies:

$$Z_{h,t} = F_t \gamma_h + u_{h,t}, \text{ for all } t, \tag{1}$$

where $u_{h,t} \sim N(0,1)$ for all $h = 1, 2, ...$ and h is the forecast horizon. The latent index $Z_{h,t}$ predicts the state of the economy h periods ahead such that:

$$R_{t+h} = \begin{cases} 1 & \text{if } Z_{h,t} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The Probit model allows us to predict the probability of a recession in h periods as:

$$\Pr\left(R_{t+h} = 1 | X_t\right) = \Phi\left(F_t \gamma_h\right), \text{ for all } \ h, \tag{3}$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. This Probit approach has been used by (Estrella & Mishkin 1998) to investigate the indicators that lead U.S. recession at horizons ranging from 1 to 8 quarters. They estimated several Probit models by using one predictor at a time. Among other results, they found that stock prices are good predictors of recessions at one to three quarters horizon while the slope of the yield curve is a better predictor beyond one quarter horizon. As suggested by (Stock & Watson 1989), the predicted probability of recession may be interpreted as a recession index. Model (1) and (2) can be estimated based on historical data. If release lags exist, the Probit model above can still be used for forecasting purposes as long as the release lags are shorter enough than the horizon h of interest.[4]

2.2 Measuring the Goodness-of-fit of a Probit

In order to assess the goodness-of-fit of the estimated Probit models, we consider using the Kullback-Leibler Information Criterion (KLIC) as starting point. The KLIC is given by:

$$KLIC = \sum_{t=1}^{T} D\left(\widehat{p}, \widehat{p}_t\right),$$

where

$$D\left(\widehat{p}, \widehat{p}_t\right) = -\widehat{p}\log\frac{\widehat{p}_t}{\widehat{p}} - (1-\widehat{p})\log\frac{1-\widehat{p}_t}{1-\widehat{p}},$$

$\widehat{p}_t = \Phi\left(X_t\widehat{\gamma}\right)$ is the probability of a recession predicted by the Probit model at an arbitrary horizon and $\widehat{p} = \frac{1}{T}\sum_{t=1}^{T} R_t$ is the sample proportion of the quarters during which the economy experienced a recession. It is easy to verify that $D\left(\widehat{p}, \widehat{p}_t\right) \geq 0$ and $D\left(\widehat{p}, \widehat{p}_t\right) = 0 \Leftrightarrow \widehat{p} = \widehat{p}_t$. Therefore, $D\left(\widehat{p}, \widehat{p}_t\right)$ is a measure of the distance between the distributions $\left(\widehat{p}, 1-\widehat{p}\right)$ and $\left(\widehat{p}_t, 1-\widehat{p}_t\right)$.

If the explanatory variables included in the Probit model are irrelevant for predicting recession, then $\widehat{p}_t$ should be close to $\widehat{p}$ and $D\left(\widehat{p}, \widehat{p}_t\right)$ should be quite small. In the neighborhood of $\widehat{p}$, a second order expansion of $D\left(\widehat{p}, \widehat{p}_t\right)$ with respect to $\widehat{p}_t$ yields:

$$D\left(\widehat{p}, \widehat{p}_t\right) \simeq D\left(\widehat{p}, \widehat{p}\right) + \frac{\partial D\left(\widehat{p}, \widehat{p}\right)}{\partial \widehat{p}_t}\left(\widehat{p}_t - \widehat{p}\right) + \frac{1}{2}\frac{\partial^2 D\left(\widehat{p}, \widehat{p}\right)}{\partial \widehat{p}_t^2}\left(\widehat{p}_t - \widehat{p}\right)^2,$$

[4]Release lags raise no issue when training a model based on historical data. If the data are released with one lag, a forecast of Quarter $t+4$ vintage based on Quarter t vintage will be available only at $t+1$.

where

$$\frac{\partial D(\widehat{p},\widehat{p}_t)}{\partial \widehat{p}_t} = \frac{\widehat{p}_t - p}{\widehat{p}_t(1-\widehat{p}_t)}; \qquad \frac{\partial D(\widehat{p},\widehat{p})}{\partial \widehat{p}_t} = 0;$$

$$\frac{\partial^2 D(\widehat{p},\widehat{p}_t)}{\partial \widehat{p}_t^2} = \frac{(\widehat{p}_t^2 - 2p\widehat{p}_t + \widehat{p})}{\widehat{p}_t^2(\widehat{p}_t - 1)^2}; \quad \frac{\partial^2 D(\widehat{p},\widehat{p})}{\partial \widehat{p}_t^2} = \frac{1}{\widehat{p}(1-\widehat{p})}.$$

Hence, if the regressors included in the Probit have no explanatory power, $\widehat{p}_t$ will be close to $\widehat{p}$ so that:

$$KLIC \simeq \sum_{t=1}^{T} \frac{(\widehat{p}_t - \widehat{p})^2}{2\widehat{p}(1-\widehat{p})}.$$

Let us find the $O\left((\widehat{p}_t - \widehat{p})^3\right)$ and $O\left((\widehat{p}_t - \widehat{p})^4\right)$ terms of the remainder of the expansion above. The third and fourth derivatives of $D(\widehat{p},\widehat{p}_t)$ are given by:

$$\frac{\partial^3 D(\widehat{p},\widehat{p}_t)}{\partial \widehat{p}_t^3} = \frac{-2\left(-\widehat{p}_t^3 + 3p\widehat{p}_t^2 - 3p\widehat{p}_t + p\right)}{\widehat{p}_t^3(1-\widehat{p}_t)^3}; \qquad \frac{\partial^3 D(\widehat{p},\widehat{p})}{\partial \widehat{p}_t^3} = \frac{-2(1-2\widehat{p})}{\widehat{p}^2(1-\widehat{p})^2}.$$

$$\frac{\partial^4 D(\widehat{p},\widehat{p}_t)}{\partial \widehat{p}_t^4} = \frac{6\left(\widehat{p}_t^4 - 4p\widehat{p}_t^3 + 6p\widehat{p}_t^2 - 4p\widehat{p}_t + p\right)}{\widehat{p}_t^4(1-\widehat{p}_t)^4}; \quad \frac{\partial^4 D(\widehat{p},\widehat{p})}{\partial \widehat{p}_t^4} = \frac{6(3\widehat{p}^2 - 3\widehat{p} + 1)}{\widehat{p}^3(1-\widehat{p})^3}.$$

Therefore, we have:

$$D(\widehat{p},\widehat{p}_t) - \frac{(\widehat{p}_t - \widehat{p})^2}{2\widehat{p}(1-\widehat{p})} \simeq \frac{-(1-2\widehat{p})}{3\widehat{p}^2(1-\widehat{p})^2}(\widehat{p}_t - \widehat{p})^3 + \frac{(3\widehat{p}^2 - 3\widehat{p} + 1)}{4\widehat{p}^3(1-\widehat{p})^3}(\widehat{p}_t - \widehat{p})^4. \tag{4}$$

When $\widehat{p} = 1/2$, the $O\left((\widehat{p}_t - \widehat{p})^3\right)$ vanishes and the remainder reduces to the last term.

If $\widehat{p}_t$ lies far apart from $\widehat{p}$, the sum of the $O\left((\widehat{p}_t - \widehat{p})^3\right)$ and $O\left((\widehat{p}_t - \widehat{p})^4\right)$ terms of the remainder will be non-negligible relatively to the $O\left((\widehat{p}_t - \widehat{p})^2\right)$ term. Building on this intuition, we measure the goodness of fit of the Probit model by:

$$R^2_{KLIC} = \frac{1}{T}\sum_t \frac{\left|D(\widehat{p},\widehat{p}_t) - \frac{(\widehat{p}_t-\widehat{p})^2}{2\widehat{p}(1-\widehat{p})}\right|}{\frac{(\widehat{p}_t-\widehat{p})^2}{2\widehat{p}(1-\widehat{p})} + \left|D(\widehat{p},\widehat{p}_t) - \frac{(\widehat{p}_t-\widehat{p})^2}{2\widehat{p}(1-\widehat{p})}\right|}. \tag{5}$$

By construction, $0 \leq R^2_{KLIC} \leq 1$. Both the numerator and denominator of $D(\widehat{p},\widehat{p}_t)$ converge to zero as $\widehat{p}_t$ approaches $\widehat{p}$. However, the numerator converges faster than the denominator so that $D(\widehat{p}, \widehat{p}_t)$ admits a well-defined limit as $\widehat{p}_t \to \widehat{p}$. This theoretical limit must be assigned to $D(\widehat{p}, \widehat{p})$ to avoid that R^2_{KLIC} be numerically undefined when $\widehat{p}_t$ coincides with $\widehat{p}$ for a given t.

Approximating $D(\widehat{p}, \widehat{p}_t) - \frac{(\widehat{p}_t - \widehat{p})^2}{2\widehat{p}(1-\widehat{p})}$ by the RHS of (4) yields:

$$R^2_{KLIC} = \frac{1}{T} \sum_t \frac{\left| \frac{2}{3}(2\widehat{p}-1)(\widehat{p}_t - \widehat{p}) + \frac{1}{2}\left(\frac{3\widehat{p}^2 - 3\widehat{p} + 1}{\widehat{p}(1-\widehat{p})}\right)(\widehat{p}_t - \widehat{p})^2 \right|}{\widehat{p}(1-\widehat{p}) + \left| \frac{2}{3}(2\widehat{p}-1)(\widehat{p}_t - \widehat{p}) + \frac{1}{2}\left(\frac{3\widehat{p}^2 - 3\widehat{p} + 1}{\widehat{p}(1-\widehat{p})}\right)(\widehat{p}_t - \widehat{p})^2 \right|}. \tag{6}$$

This shows that the limit of R^2_{KLIC} as $(\widehat{p}_1, ..., \widehat{p}_T) \to (\widehat{p}, ..., \widehat{p})$ is zero. Indeed:

$$R^2_{KLIC} = 0 \iff (\widehat{p}_1, ..., \widehat{p}_T) = (\widehat{p}, ..., \widehat{p}).$$

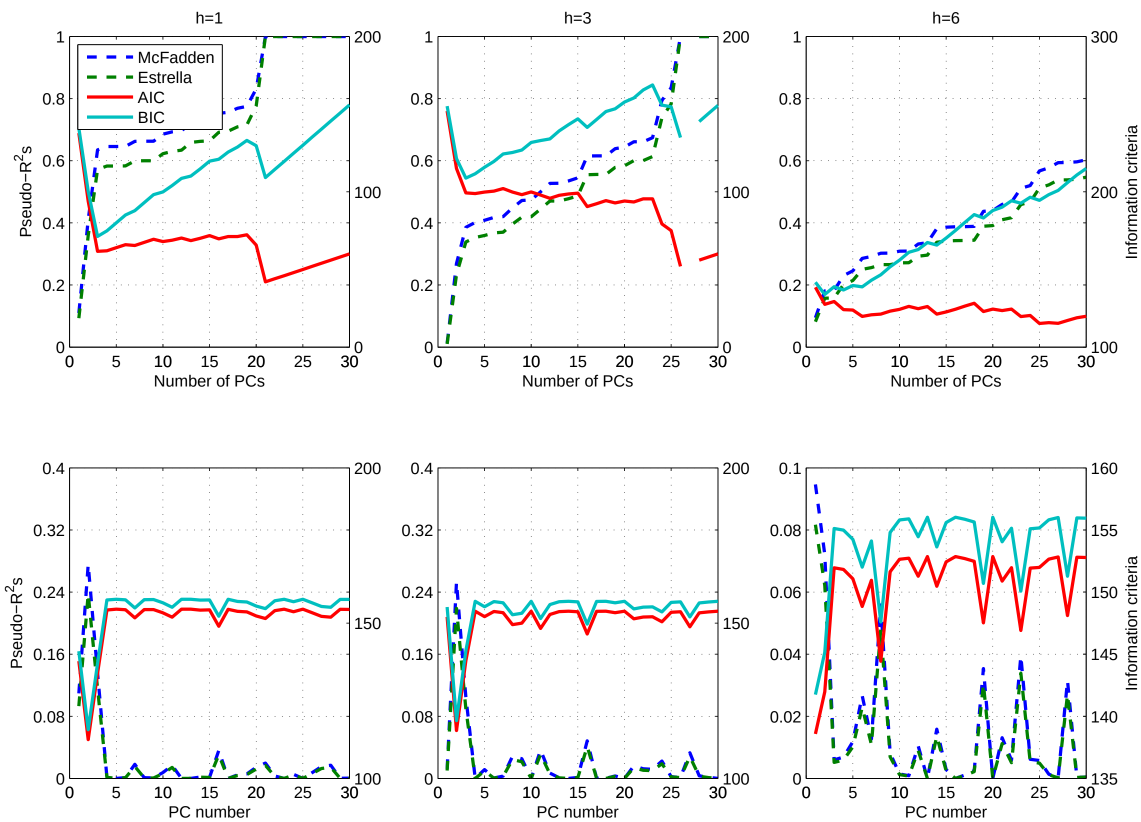
By avoiding a division by zero when $(\widehat{p}_1, ..., \widehat{p}_T) = (\widehat{p}, ..., \widehat{p})$, the expression (6) is numerically stable and is therefore preferable to (5).

2.3 Selecting the Factors to Include in the Probit

Suppose we had to use the AIC or the BIC to select the best Probit model, where the models differ only in their number of regressors. The ideal procedure consists of estimating a model for all possible combinations of regressors and computing the relevant information criterion for each model. With q regressors and a constant variable, the total number of models to be estimated is 2^q. Therefore, this approach is unfeasible with a number of regressors as small as $q = 20$.

The curse of dimensionality identified above can be avoided by ranking the regressors by order of importance. A natural ordering of the factors in our case is dictated by their importance as principal component, i.e., F_1, F_2, ... and F_q. A simplified model selection procedure under this ordering consists of comparing $q+1$ models with an increasing number of regressors, where the k^{th} model includes a constant and $(F_1,...,F_k)$ as regressors. The k^{th} model is declared the best if it has a smaller AIC or BIC than all the previous models and the next one. Unfortunately, the ranks attributed to the regressors do not necessarily reflect their power at predicting the occurrence of recessions. As a result, the simplified model selection procedure may miss a factor that is a good predictor of R_{t+h} but has low importance as a PC. More precisely, F_{10} can be a good predictor of recessions while F_2,F_3 and F_4 are not.

Figure 1: Problems with standard goodness-of-fit and information criteria



The data used to construct this figure are discussed latter in Application section. The first row panels contains standard goodness-of-fit and information criteria for cumulative number of principal components. The second row presents the same results when each principal component is considered individually.

The first row panels of Figure 1 show the AIC, the BIC and two pseudo R-squares (Estrella and McFadden) as a function of the number of factors used as predictors in the simplified model selection procedure. The BIC selects the model with 3 PCs as the best for forecasting a recession at horizons $h = 1$ and $h = 3$ quarters, and it suggests using only 2 PCs for $h = 6$ quarters. However, the second row panel suggests that some PCs lying farther also have predictive power. Such PCs will never be selected because they are preceded in the ordering by several factors that are irrelevant for predicting recessions at the horizon of interest.

We propose a variable selection procedure based on the R^2_{KLIC} that avoids the issue raised with the AIC and BIC. This procedure consists of testing the significance of the explanatory power of each factor taken individually. We design a test procedure to infer whether the theoretical counterpart of R^2_{KLIC} is significantly greater than zero or not. The distribution of R^2_{KLIC} under the null hypothesis that the Probit does not fit the data better than the sample proportion depends on the regressors included in the model. Obtaining the distribution of R^2_{KLIC} in closed form is tedious. Therefore, we consider simulating it conditional on the regressors.

To draw one realization of R^2_{KLIC} from its unknown distribution, one proceeds as follows:

Step 1: Simulate T observations according to the void model given by:

$$z_t = \overline{\gamma}_0 + \varepsilon_t, t = 1, ..., T,$$

where $\overline{\gamma}_0 = \Phi^{-1}(\widehat{p})$, $\widehat{p} = \frac{1}{T}\sum R_t$, ε_t is IID standard normal and z_t is the latent variable leading the occurrence of recessions.

Step 2: Deduce simulated recessions as:

$$R_{simul,t} = 1\left\{z_t = 0\right\}, \; t = 1, ..., T.$$

Step 3: Fit a Probit model to $R_{simul,t}$ using K regressors of interest plus a constant, contained in $\breve{X}_t$:

$$\widehat{p}_t = \widehat{p}_t = \Phi\left(\breve{X}_t \widehat{\gamma}\right),$$

where the first element of $\breve{X}_t$ is the constant and $\widehat{\gamma} = (\widehat{\gamma}_0, \widehat{\gamma}_1, ..., \widehat{\gamma}_K)$. Note that $\widehat{\gamma}$ is a consistent estimator of $\overline{\gamma} = (\overline{\gamma}_0, 0, ..., 0)$.

Step 4: Compute the realization of the R^2_{KLIC} using (6).

The finite sample distribution of R^2_{KLIC} conditional on $\left(\breve{X}_1, ..., \breve{X}_T\right)$ under the null hypothesis that the regressors are irrelevant is approximated by repeating the steps 1-4 so as to obtain a large number of replicas of R^2_{KLIC}, denoted $R^2_{KLIC,m}, m = 1, ..., M$. The critical values inferred from this simulated distribution can be used for hypothesis testing. We implement this procedure by using each PC factor as single regressor since they are orthonormal. More precisely, we let $\breve{X}_t = (1, F_{k,t})$, where $F_{k,t}$ is a PC, and simulate the corresponding critical values. Figure 2 shows the finite-sample distributions of the test statistic, under the null, and the corresponding 1%, 5% and 10% critical values, for different choices of sample size and $\widehat{p}$. For a fixed proportion of ones in the sample, $\widehat{p}$, the critical values become smaller when the sample increases. In addition, for a fixed sample size, T, the critical values decreases when $\widehat{p}$ increases. This means that in small samples and in cases where ones are rather infrequent (as is the case for the indicator of US recessions) the candidate factor must have considerable forecasting power in order to be selected.

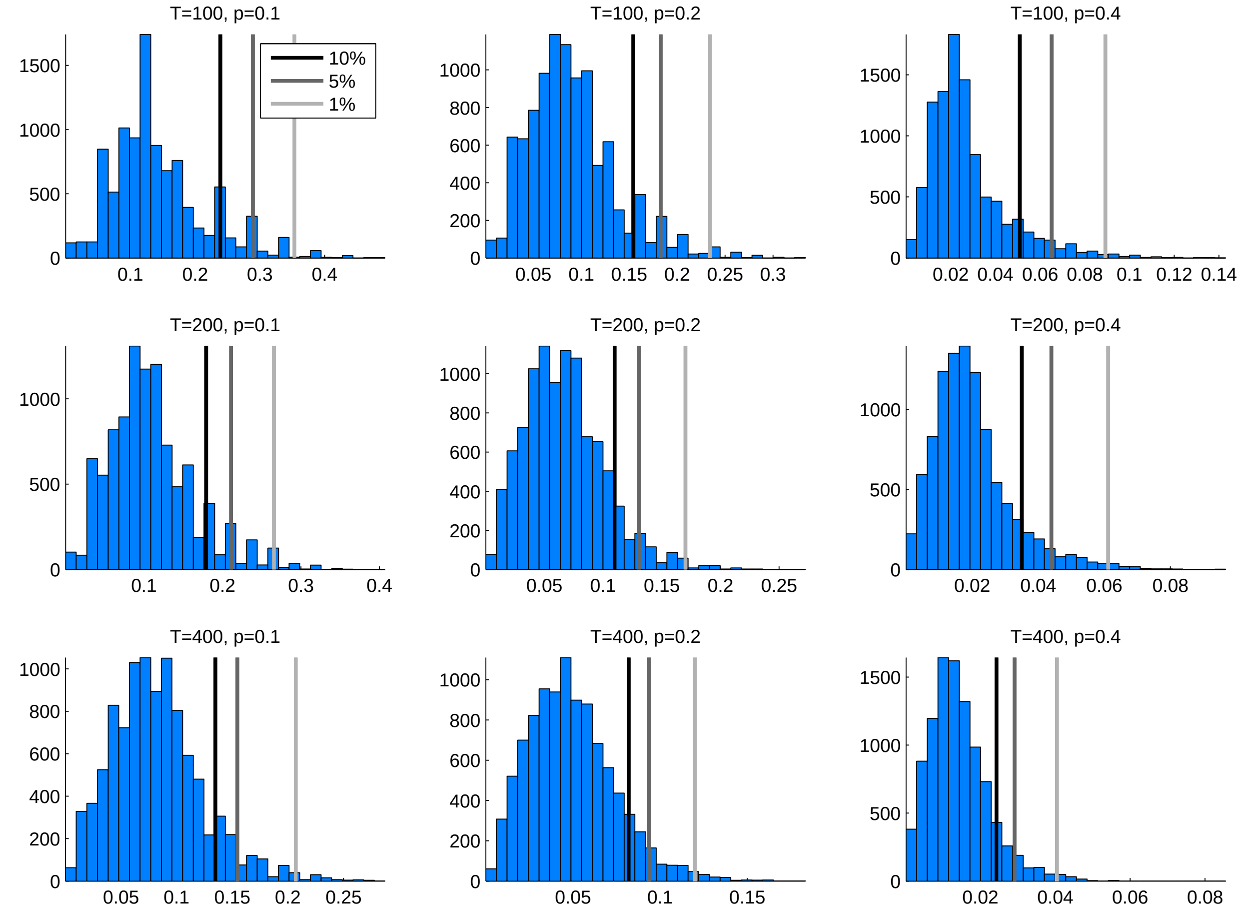
2.4 Modeling the Real Economic Activity: an IMR-DI-AR Approach

In order to assess the severity of a recession, we consider an AR model augmented with diffusion indexes as starting point:

$$y_{i,t+h} = \alpha_{i,h} y_t + F_t \beta_{i,h} + \delta_{i,h} R_{t+h} + v_{i,t+h}, t = 1, ..., T - h, \tag{7}$$

where $y_{i,t}$, $i = 1, \ldots, M$, is a measure of economic activity and $v_{i,t+h} \sim N(0, \sigma^2_{i,h})$ is assumed uncorrelated with F_t and y_t. This model may be viewed as a sparse version of the Diffusion Index AR of (Stock & Watson 2002b). We recall that F_t includes a constant variable and R_{t+h} is the indicator of recession at $t + h$. Equation (7) stipulates that the ex-post realization

Figure 2: Finite-sample distributions of R^2_{KLIC} under H_0



This figure presents the finite-sample distributions of the test statistic, under the null, and the corresponding 1%, 5% and 10% critical values, for different choices of sample size and $\widehat{p}$. The regressor is simulated from the $N(0,1)$.

of $y_{i,t+h}$ depends on whether the economy experiences a recession at period $t+h$ of not. Clearly, this equation cannot be used for forecasting as it contains a regressor that is not observed yet at period t. However, its structure can be exploited to infer a useful forecasting formula and a feasible estimating equation.

Taking the expectation of y_{t+h} conditional on the information available at time t yields:

$$E\left(y_{i,t+h}|y_t, X_t\right) = \alpha_{i,h} y_t + F_t \beta_{i,h} + \delta_{i,h} \Phi\left(F_t \gamma_h\right) \equiv \widehat{y}_{t+h}. \tag{8}$$

Based on Equation (8), we represent y_{t+h} as follows:

$$y_{i,t+h} = \alpha_{i,h} y_t + F_t \beta_{i,h} + \delta_{i,h} \Phi \left(F_t \gamma_h \right) + \widetilde{v}_{i,t+h}, \tag{9}$$

where $\widetilde{v}_{i,t+h} \equiv v_{i,t+h} + \delta_{i,h} \left(R_{t+h} - \Phi \left(F_t \gamma_h \right) \right)$. Unlike Equation (7), Equation (8) can be used for forecasting.

Assuming that the distribution of the error term depends on the future state of the economy, two other forecasts can be constructed beside the average scenario given by (8). The first forecast is based on the pessimistic assumption that the economy will experience a recession at period $t+h$, i.e.:

$$E \left(y_{i,t+h} | y_t, X_t, R_{t+h} = 1 \right) = \alpha_{i,h} y_t + F_t \beta_{i,h} + \overline{\delta}_{i,h} + \delta_{i,h,1} \frac{\phi \left(F_t \gamma_h \right)}{\Phi \left(F_t \gamma_h \right)} = \underline{y}_{i,t+h}, \tag{10}$$

where $\delta_{i,h,1} = Cov \left(u_{h,t}, v_{i,t+h} | R_{t+h} = 1 \right)$, $\overline{\delta}_{i,h}$ is the intrinsic effect of the recession and $\delta_{i,h,1} \frac{\phi(F_t \gamma_h)}{\Phi(F_t \gamma_h)}$ stems from a "break" in the structure of dependence between $y_{i,t+h}$ and F_t due to the recession. Equation (10) is obtained by assuming that $(u_{h,t}, v_{i,t+h})$ are jointly Gaussian, where $u_{h,t}$ is the error term of the relevant Probit. The second forecasting formula is based on the optimistic assumption that there will be an expansion at period $t+h$:

$$E \left(y_{i,t+h} | y_t, X_t, R_{t+h} = 0 \right) = \alpha_{i,h} y_t + F_t \beta_{i,h} + \delta_{i,h,0} \frac{-\phi \left(F_t \gamma_h \right)}{1 - \Phi \left(F_t \gamma_h \right)} = \overline{y}_{i,t+h}, \tag{11}$$

where $\delta_{i,h,0} = Cov \left(u_{h,t}, v_{i,t+h} | R_{t+h} = 0 \right)$ and $\delta_{i,h,0} \frac{\phi(F_t \gamma_h)}{1-\Phi(F_t \gamma_h)}$ is a break that marks an expansion period. The forecasting formulas (10) and (11) depend on quantities that are all known at time t.

Our optimistic and pessimistic forecasts fall in the broad family of conditional forecasts studied by (Clark & McCracken 2013). This family includes all macroeconomic forecasts that are made conditional on a particular policy path for the period [t,t+h] (e.g., announced inflation target), or conditional on a given scenario for the future path of certain macroeconomic variables (e.g., low inflation and high unemployment)[5].

The variables $\frac{\phi(F_t\gamma_h)}{\Phi(F_t\gamma_h)}$ and $\frac{\phi(F_t\gamma_h)}{1-\Phi(F_t\gamma_h)}$ are the well-known inverse Mills ratios (IMR). The parameters $\overline{\delta}_{i,h}$, $\delta_{i,h,0}$ and $\delta_{i,h,1}$ are all expected to be negative if $y_{i,t}$ is cyclical (i.e., increases during expansions and shrinks during recessions) and $\delta_{i,h,0}$ and $\delta_{i,h,1}$ should be positive otherwise. Among other things, the terms $\delta_{i,h,1}\frac{\phi(F_t\gamma_h)}{\Phi(F_t\gamma_h)}$ and $\delta_{i,h,0}\frac{-\phi(F_t\gamma_h)}{1-\Phi(F_t\gamma_h)}$ capture the combined effects of variables that are hard to measure such as policy announcements, investors sentiments, consumer confidence, agents anticipations, etc. By definition, the weighted average of the optimistic and pessimistic scenarios should coincide with the average scenario, that is:

$$\begin{aligned} E\left(y_{i,t+h}|y_t, X_t\right) \;\; \equiv \;\; & \Phi\left(F_t\gamma_h\right) E\left(y_{i,t+h}|Y_t, X_t, R_{t+h}=1\right) \\ & + \left(1-\Phi\left(F_t\gamma_h\right)\right) E\left(y_{i,t+h}|Y_t, X_t, R_{t+h}=0\right). \end{aligned}$$

Therefore, we have:

$$\delta_{i,h}\Phi\left(F_t\gamma_h\right) = \overline{\delta}_{i,h}\Phi\left(F_t\gamma_h\right) + \left(\delta_{i,h,1}-\delta_{i,h,0}\right)\phi\left(F_t\gamma_h\right).$$

Hence, $\overline{\delta}_{i,h}$ does not coincide with $\delta_{i,h}$ when the distribution of $v_{i,t+h}$ is state dependent. This is not surprising as Equation (9) then becomes a reduced form, unlike Equations (10) and (11) which contain more structure. In fact, (9) is subject to the forbidden regression problem when the distribution of $v_{i,t+h}$ is state dependent.

We build our measure of the severity of recessions within the structural model. By pooling the forecasting formulas (10) and (11) together, we obtain:

$$y_{i,t+h} = \alpha_{i,h}y_t + F_t\beta_{i,h} + \overline{\delta}_{i,h}R_{t+h} + \delta_{i,h,0}IMR_{t,h,0} + \delta_{i,h,1}IMR_{t,h,1} + \widetilde{\widetilde{v}}_{i,t+h}, \qquad (12)$$

[5] For instance, (Giannone, Lenza, Momferatou & Onorante 2010) perform an inflation forecasting exercise conditional on pre-specified paths for oil price indicators. (Schorfheide & Song 2013) produce inflation and growth forecasts conditional on forecasts obtained from judgmental sources. Other references on conditional forecasts include (Sims 1982), (Doan, Litterman & Sims 1984), (Meyer & Zaman 2013) and (Aastveit, Carriero, Clark & Marcellino 2014).

where $IMR_{t,h,0}$ and $IMR_{t,h,1}$ are given by:

$$IMR_{t,h,1} = \begin{cases} \frac{\phi(F_t\gamma_h)}{\Phi(F_t\gamma_h)} & \text{if } R_{t+h} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$IMR_{t,h,0} = \begin{cases} \frac{-\phi(F_t\gamma_h)}{1-\Phi(F_t\gamma_h)} & \text{if } R_{t+h} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

and $\widetilde{\widetilde{v}}_{i,t+h}$ is a zero mean error.

A feasible estimating equation is deduced from (12) by replacing R_{t+h} by $\Phi\left(F_t\gamma_h\right)$:

$$y_{i,t+h} = \alpha_{i,h}y_t + F_t\beta_{i,h} + \overline{\delta}_{i,h}\Phi\left(F_t\gamma_h\right) + \delta_{i,h,0}IMR_{t,h,0} + \delta_{i,h,1}IMR_{t,h,1} + \eta_{i,t+h}, \tag{13}$$

where $\eta_{i,t+h} = \widetilde{\widetilde{v}}_{i,t+h} + \overline{\delta}_{i,h}\left(R_{t+h} - \Phi\left(F_t\gamma_h\right)\right)$. Equation (13) will be estimated to identify the parameters $\left(\alpha_{i,h}, \beta_{i,h}, \overline{\delta}_{i,h}, \delta_{i,h,0}, \delta_{i,h,1}\right)$. The following equation deduced from (9) will be estimated to identify $\delta_{i,h}$:

$$\widetilde{y}_{i,t+h} = \delta_{i,h}\Phi\left(F_t\gamma_h\right) + \widetilde{v}_{i,t+h},$$

where $\widetilde{y}_{i,t+h} = y_{i,t+h} - \widehat{\alpha}_{i,h}y_t + \widehat{\beta}_{i,h}F_t$, $\widehat{\alpha}_{i,h}$ and $\widehat{\beta}_{i,h}$ are obtained from Equation (13). Finally, Equations (8), (10) and (11) will be used to generate the average, the pessimistic and the optimistic forecast scenarios.

2.5 Selecting the Factors to Include in the IMR-DI-AR Model

There is no direct relationship between the importance of a PC and its power at explaining a given real economic activity variable. For instance, GDP growth might be explained by the first and fifth principal components and not by the factors in between. Therefore, we need a procedure to identify the factors that are potentially important for explaining the particular real activity variable of interest.

Our selection procedure relies on the residuals $\widehat{v}_{i,t+h}$ of the following regression:

$$y_{i,t+h} = \widehat{\rho}_0 + \widehat{\rho}_1 y_t + \widehat{\rho}_2 \Phi\left(F_t^P \widehat{\gamma}_h\right) + \widehat{\rho}_3 \widehat{IMR}_{t,h,0} + \widehat{\rho}_4 \widehat{IMR}_{t,h,1} + \widehat{w}_{i,t+h}, \tag{14}$$

where F_t^P is the set of factors selected for the Probit, and:

$$\widehat{IMR}_{t,h,1} = \begin{cases} \frac{\phi\left(F_t^P \widehat{\gamma}_h\right)}{\Phi\left(F_t^P \widehat{\gamma}_h\right)} & \text{if } R_{t+h} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$\widehat{IMR}_{t,h,0} = \begin{cases} \frac{-\phi\left(F_t^P \widehat{\gamma}_h\right)}{1-\Phi\left(F_t^P \widehat{\gamma}_h\right)} & \text{if } R_{t+h} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

In fact, $\Phi\left(F_t^P \widehat{\gamma}_h\right)$, $\widehat{IMR}_{t,h,0}$ and $\widehat{IMR}_{t,h,1}$ are all computed using the output of the Probit model.

Note that the residuals $\widehat{v}_{i,t+h}$ have at most $T-h-5$ degree of freedoms (h observations are lost because of lagged variable, 4 regressors included plus the constant). We regress these residual on each factor individually:

$$\widehat{v}_{i,t+h} = \widetilde{\rho}_{h,k} F_{k,t} + \widetilde{w}_{i,t+h}.$$

There is no need to include an intercept in these regressions because the factors and the regressand are both centered. As the factors are mutually orthogonal, this amounts to regress $\widehat{v}_{i,t+h}$ on all the factors once and for all:

$$\widehat{v}_{i,t+h} = F_t \widetilde{\rho}_h + \widetilde{w}_{i,t+h}, \tag{15}$$

where $\widetilde{\rho}_h = (\widetilde{\rho}_{h,1}, ..., \widetilde{\rho}_{h,K})'$. As the factors are standardized, the contributions of $F_{k,t}$ to the variance of $\widehat{v}_{i,t+h}$ is $\widetilde{\rho}^2_{h,k}$.

Under the null hypothesis that the $F_{k,t}$ has no explanatory power for $\widehat{v}_{i,t+h}$, we have:

$$\frac{\widetilde{\rho}^2_{h,k}}{\left(\frac{1}{T-1}\sum\left(\widehat{v}_{i,t+h}-\overline{\overline{v}}_{i,t+h}\right)^2-\widetilde{\rho}^2_{h,k}\right)/(T-h-6)}\sim F\left(1,T-h-6\right),$$

where $\overline{\overline{v}}_{i,t+h}$ is the sample average of $\widehat{v}_{i,t+h}$, $\widetilde{\rho}^2_{h,k}$ is the portion of the variance of $\widehat{v}_{i,t+h}$ explained by $F_{k,t}$ and $\frac{1}{T-1}\sum\left(\widehat{v}_{i,t+h}-\overline{\overline{v}}_{i,t+h}\right)^2-\widetilde{\rho}^2_{h,k}$ is the sum of squared residuals of a simple regression of $\overline{\overline{v}}_{i,t+h}$ on $F_{k,t}$. We use the critical values of this distribution to assess the significance of the explanatory power of each factor.

2.6 Measuring the Severity of Recessions

We define the *expected severity of a recession* as the gap between the pessimistic forecast given by (10) and the medium run trend of the series $y_{i,t+h}$. For simplicity, we assume that this trend is indicated by the average of the eight preceding quarters [6]. Therefore, we have:

$$\Delta^e\left(y_{i,t+h}\right)=\underline{y}_{i,t+h}-\frac{1}{8}\sum_{t-7}^{t}y_{i,t}. \tag{16}$$

If $y_{i,t}$ denotes the GDP growth for instance, a recession is expected to generate an output loss and therefore, $\Delta^e\left(y_{i,t+h}\right)\leq 0$. If $y_{i,t}$ represents the unemployment rate instead, its values are expected to increase during a recession, leading to $\Delta^e\left(y_{i,t+h}\right)\geq 0$.

A recession that is expected to be quite severe *ex ante* may turn out to be mild *ex post* (and *vice versa*). Hence, we define the *realized severity of a recession* as the difference between the actual realization of $y_{i,t+h}$ and its recent trend.

$$\Delta^r\left(y_{i,t+h}\right)=y_{i,t+h}-\frac{1}{8}\sum_{t-7}^{t}y_{i,t}. \tag{17}$$

[6] There is some amount of arbitrariness in the design of the moving average. Note however that the bandwidth must not be too short in order to avoid mixing the momentum with the medium run trend. Likewise, two-sided filters that exploit future information to extract the trend should not be used.

The difference between the expected and realized severity of a recession is given by:

$$\Delta^{e}\left(y_{i,t+h}\right)-\Delta^{r}\left(y_{i,t+h}\right)=\underline{y}_{i,t+h}-y_{i,t+h}.$$

This difference accounts for the impact of policy actions aimed at fighting against the recession as well as for unexpected favorable conditions (good luck) and exogenous adverse shocks (bad luck). Our methodology cannot disentangle these two cases, but can be used to identify episodes where the realized severity was higher than expected.

3 Application to NBER Recession

For this application, we use the quarterly NBER recession indicator available in the FRED2 database. The set of regressors included in X_t is comprised of 42 variables covering the yield curve, credit spreads, the stock market, the housing market, the job market, etc. See Table 6 for details. The data cover the period running from 1967Q2 to 2012Q3. An important variable often used in the literature to predict recessions, Initial Claims (IC4WSA), is available only since 1967Q2. Our time series stop at 2012Q3 because of the availability of the Consumer Sentiment (ConsMICH) computed by the University of Michigan.

3.1 Predictability of US Recessions

The first step of our investigation strategy consists of estimating the probability of a recession occurring at horizon h quarters. For that purpose, we start by selecting the relevant PCs to use as regressors in the Probit models.

Selecting the Regressors to Include in the Probits: We estimate Probit models that use only one PC and a constant as regressors and test the significance of the R^2_{KLIC} for each model. Table 1 shows the critical values of the distribution of the R^2_{KLIC} under the null hypothesis that the included PC has no predictive power. We note that these critical values are quite stable across the PCs and horizons. Figure 3 shows the distribution of the R^2_{KLIC} under the null. The set of predictors selected for inclusion in the Probit models depends much on the forecasting horizon. For instance, the first PC is relevant for predicting recessions at horizon $h = 1$, but not at $h = 2$ and $h = 4$ quarters. The second PC is selected at several horizons and should therefore be considered an important lead indicator of recessions. At 10% significance level, our test concludes that the 19th PC is a good predictor of recessions at horizon $h = 7$. Overall, our methodology allows us to select the PCs with strongest signals without overloading the Probit model with irrelevant regressors.

Table 1: Distribution of critical values

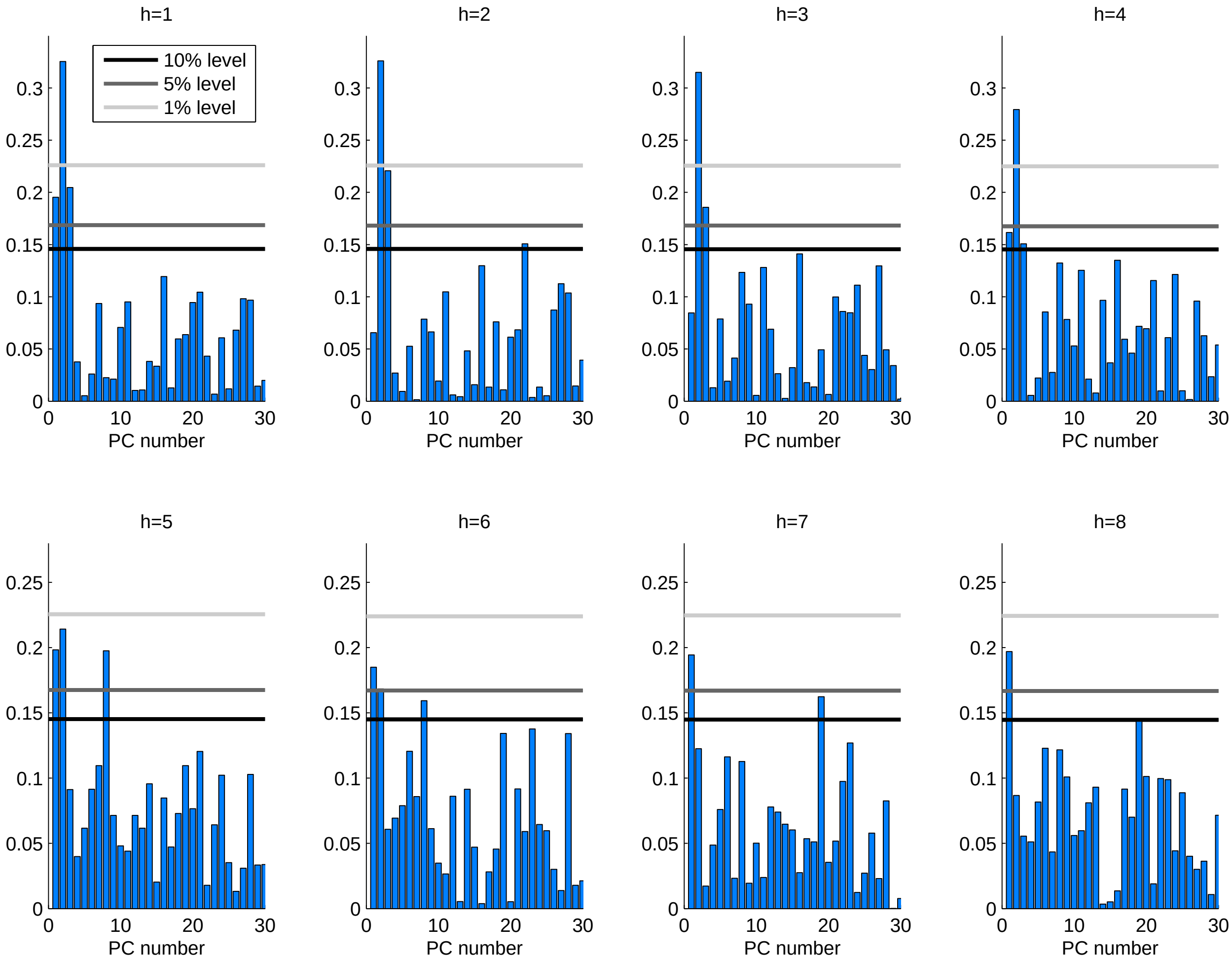
Level	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	All horizons
10%	0,1458	0,1459	0,1455	0,1454	0,1453	0,1450	0,1449	0,1447	0,1453
	(0,0010)	(0,0002)	(0,0008)	(0,0006)	(0,0004)	(0,0010)	(0,0005)	(0,0008)	(0,0007)
5%	0,1686	0,1681	0,1682	0,1676	0,1675	0,1671	0,1670	0,1668	0,1676
	(0,0012)	(0,0011)	(0,0017)	(0,0011)	(0,0010)	(0,0010)	(0,0010)	(0,0012)	(0,0012)
1%	0,2260	0,2257	0,2255	0,2249	0,2256	0,2239	0,2247	0,2243	0,2251
	(0,0027)	(0,0032)	(0,0027)	(0,0034)	(0,0018)	(0,0040)	(0,0023)	(0,0023)	(0,0028)

The columns contain critical values for each horizon averaged over all 40 principal components. The rows present levels. The values in parenthesis are average standard deviations.

The PCs selected at 10% significance level are presented in Table 2 along with the five variables to which they are correlated the most. The first PC (F_1) is highly related to employment growth, credit spread, capital utilization and PMI index[7]. The second PC (F_2) represents inflation and short term interest rates. F_3 is linked to stock market returns and realized volatility as well as to consumption growth. The fact that stock prices contribute to the formation of F_3 suggests that robust predictors of the state of the economy can be obtained by using their linear combinations with other variables. Interestingly, the PCs F_4 through F_7 do not have significant power at predicting recessions at up to eight quarters. F_8 is correlated with the skewness of stock market returns, money growth, durable consumption growth and yield. F_{19} loads onto money growth and yield spread. Finally, F_{22} is related to

[7](Ng 2013) finds similar series using boosting algorithms within more than a hundred of monthly indicators.

Figure 3: Choosing individual factors using R^2_{KLIC} testing procedure



R^2_{KLIC} across principal components and horizons, according to the testing procedure in Section 2.3. We show only 30 principal components per horizon in order to keep the dimension readable.

ISM manufacturing price index, credit spreads and stock returns skewness. These empirical findings suggests that the first three moments of stock market returns contain strong signals that lead the future states of the economy.

Probit Estimation Results: The first panel in Table 3 shows the selected predictors along with goodness-of-fit measures at different horizons. The first PC is selected at all horizons except for $h = 2$ and $h = 3$. The second PC is selected at all horizons between $h = 1$ and $h = 6$ quarters. The third PC is not selected for horizons longer than four quarters. A factor related to the stock market's realized skewness (F_8), is important for

Table 2: Most correlated variables with selected PCs in Probit models

F_1	PAYEMS 0,86	BAA-GS10 0,84	MANEMP 0,80	CUmftg 0,79	NAPM 0,78
F_2	CPILFEL 0,86	FEDFUNDS 0,84	CPIAUCSL 0,80	PCEPI 0,79	TB3MS 0,79
F_3	SP500 0,63	DJIA 0,61	RPCE 0,57	SP500-RV 0,54	DJIA-RV 0,53
F_8	DJIA-SV 0,41	SP500-SV 0,40	M2SL 0,38	RPCEDG 0,31	GS5 0,28
F_{19}	M1SL 0,18	GS1-FFR 0,17	GS10-TB3MS 0,16	CUmftg 0,13	NAPMPRI 0,13
F_{22}	NAPMPRI 0,22	BAA-AAA 0,12	GS1-FFR 0,10	DJIA-SK 0,08	NAPMOI 0,08

Notes: Five most correlated series with principal components selected by 10% *level* R^2_{KLIC} *test, for each horizon, following the steps in Section 2.3.*

predicting recessions at horizon 5 and 6 quarters. Figure 4 shows the predicted probabilities of recessions. Most of the recession dates are well predicted 1 and 2 quarters ahead while the recessions in 70s and 80s were predictable at up to 6 quarters ahead. There is a misleading peak in the predicted probability of a recession in 1987Q3 at horizons 1 to 4. This is due to an important slump of the SP500 index (i.e., a decrease of roughly 20%) which occurred between 1987Q2 and 1987Q3. The observed false positives are driven by the fact that the third PC loads heavily on stock prices while it is selected as relevant for predicting recessions at horizons $h = 1$ through $h = 4$.

The second panel of the Table 3 shows other measures of quality of fit commonly used for binary choice models. The *% of good shots* is the proportion of times where the predicted probability is higher than a given threshold while the NBER effectively called for a recession. Similarly, the *% of bad shots* is the proportion of times where the model calls for a recession (given the retained threshold) while the NBER did not. When a fixed and arbitrary threshold of 50% is used to compute these statistics, we find that 67% of recessions periods are well predicted at horizons 1 and 2. Not surprisingly, this percentage decreases as h increase and is roughly equal to zero at horizons 7 and 8 quarters. At the same time, the percentage of

Table 3: Predicting NBER recessions: in-sample goodness-of-fit

Quarter	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
Selected PCs	[1 2 3]	[2 3 22]	[2 3]	[1 2 3]	[1 2 8]	[1 2 8]	[1 19]	[1]
R^2_{KLIC}	0,4811	0,4394	0,3752	0,3553	0,3379	0,2890	0,2327	0,1970
McFadden R^2	0,6347	0,5218	0,3658	0,3328	0,3095	0,2241	0,1617	0,1112
Estrella R^2	0,5720	0,4640	0,3205	0,2914	0,2712	0,1955	0,1408	0,0967
Fixed threshold	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
% of good shots	0,6667	0,6667	0,4074	0,3333	0,2963	0,2222	0,0000	0,0370
% of bad shots	0,0390	0,0392	0,0329	0,0397	0,0467	0,0403	0,0068	0,0000
Varying threshold	0,39	0,38	0,42	0,47	0,48	0,45	0,34	0,31
% of good shots	0,7778	0,7778	0,5926	0,4074	0,2963	0,2222	0,2222	0,1481
% of bad shots	0,0455	0,0458	0,0461	0,0464	0,0467	0,0470	0,0473	0,0476

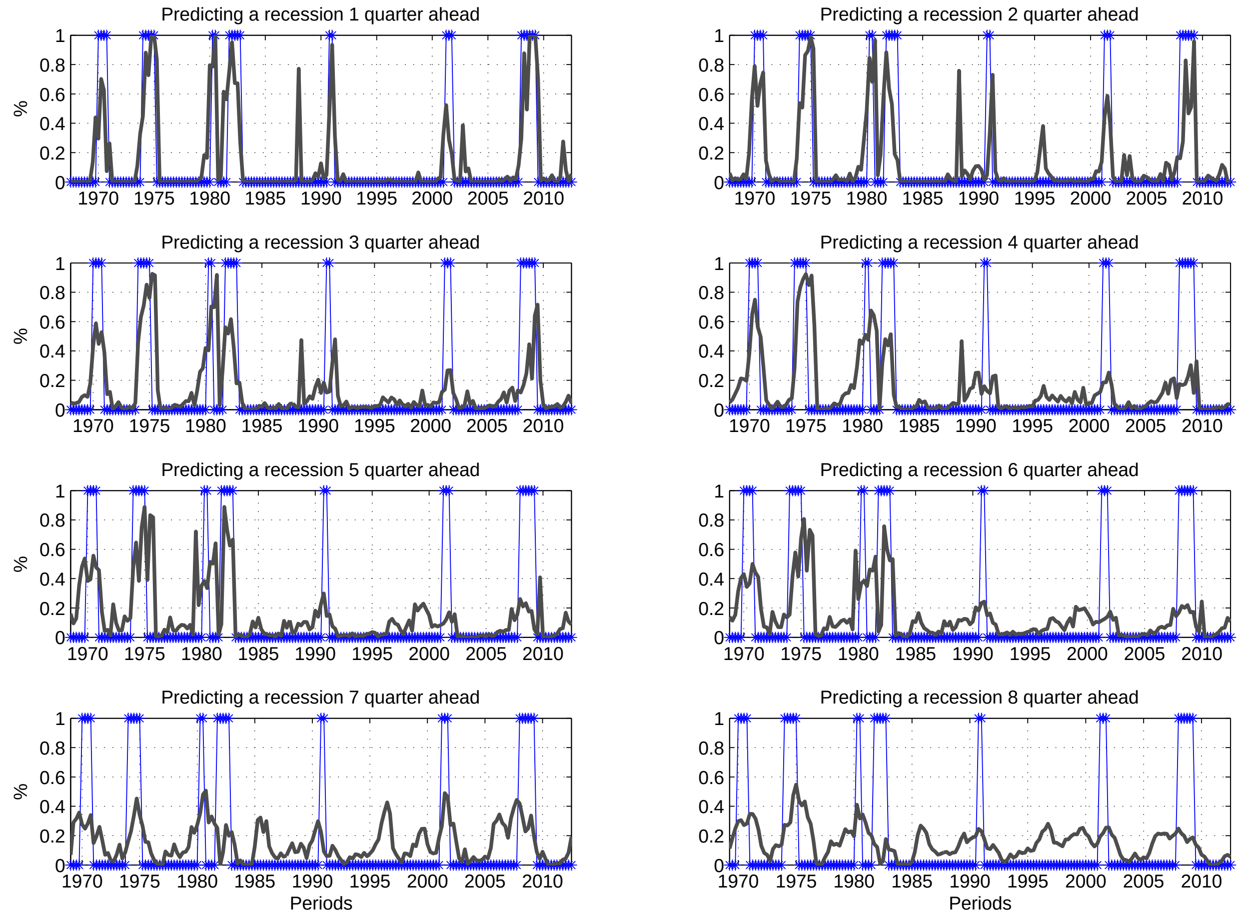
In-sample goodness-of-fit measures from the Probit model in (3). For each horizon the Probit model is specified according to testing procedure in Section 2.3. The % of good shots measures the number of successively called recession (a recession is called if $\Phi(\gamma_h F_t)$ is higher than the threshold). The % of bad shots measures the proportion of situations when a recession is called while $R_t = 0$. The fixed threshold means that $\Phi(\gamma_h F_t)$ must be larger than 0.5 *to call a recession. The varying threshold is obtained such that the percentage of bad calls is no larger than* 5 *inducing then a percentage of good shots.*

bad shots remains very low at long horizons.

To understand these results, recall that our objective is to decide whether there will be a recession or not at horizon h. In this decision process, the null hypothesis is "H0: no recession at horizon h." Our decision rule consists of rejecting H0 when the predicted probability of a recession exceeds a given threshold. The percentage of bad shot is the type I error of our decision rule while the percentage of good shots is the power (one minus the type II error). In light of this, we see that a decision process that relies on a arbitrary threshold of 50% keeps the type I error low but has no power at long horizon.

Instead of using an arbitrary threshold of 50%, it might be interesting to design a threshold that varies with the forecast horizon so as to control the percentage of bad shot at a conventional 5% level. Indeed, we do not expect the Probit model to have the same accuracy at predicting a recession at 1 quarter and at 2 years horizons. Therefore, a 30% probability of recession at horizon $h = 8$ should perhaps be taken as seriously as a 50% probability at horizon h=1. We have computed varying thresholds that keep the percentage of bad shots at approximately 5%. The results are shown on the lower panel of Table 3. The percentage

Figure 4: Predicted in-sample US recessions

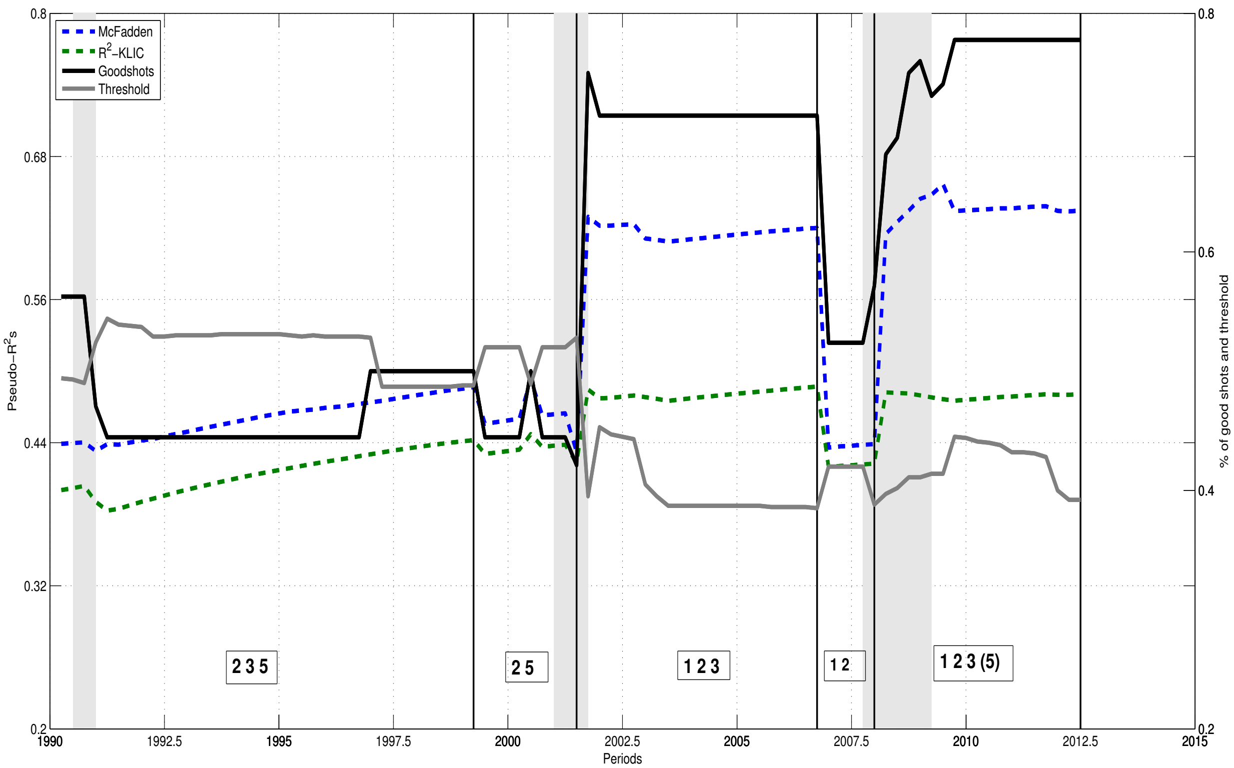


Predicted in-sample probabilities of NBER recessions from Probit model in (3). For each horizon the Probit model is specified according to testing procedure in Section 2.3.

of good shots (i.e., the power of the decision rule) increases significantly at the shortest and longest horizons. Hence, the adaptive varying threshold permits to increase the power of our decision rule. Figure 12 in Appendix shows the percentages of good and bad shots across horizons.

Predictive Power of the PCs over time: We conduct a recursive selection of the PCs to include in the Probit model that predicts recessions at horizon $h = 1$. The aim of this exercise is to assess whether all recessions that occurred in the US are led by the same factors. Figure 5 shows the selected PCs over time and the associated goodness-of-fit measures. The principal components have been calculated on the full sample prior to the recursive estimation. The set of lead indicators of recessions (listed in the boxes) is clearly not stable over time. The first PC (related to employment growth, credit spread, capital utilization and PMI index) becomes important only after 2001. On the other hand, the fifth factor (representing the stock market volatility and the housing market) was relevant between 1990 and 2001. This suggests that all recessions are not alike as far as their determinants are concerned. Regarding the quality-of-fit measures, we note that the R^2_{KLIC} is smoother than McFadden's R-square and the percentage of good shots over time.

Figure 5: Recursive in-sample selection of PCs for $h = 1$



This figure presents the recursive in-sample selection of principal components into predictive Probit model for 1-quarter forecasting horizon, as well as several goodness-of-fit measures over time. The principal components have been calculated for the full sample prior to recursive estimation, in order to keep constant their interpretation in terms of correlations with observables. The five most correlated series with the fifth principal components are : DJIA-RV, SP500-RV, M2SL, HOUS, PERMIT.

3.2 Predicting the Real Economic Activity

With the predicted probabilities of recessions in hand, we can now estimate the IMR-DI-AR model presented in Section 2.4. We also consider three restricted versions of our benchmark model: a diffusion index AR model (DI-AR) of (Stock & Watson 2002b), an IMR-augmented AR model (IMR-AR) and a standard AR model. The autoregressive part is limited to one lag for all models. The PCs included in the factor augmented models are selected as described in Section 2.5. The Probit model estimated at the previous step relies on factors that have been selected once and for all on the whole sample (and not recursively). Table 4 shows the adjusted R-squares of the models fitted to the GDP growth rate.

Adding the IMRs as regressors in an DI-AR model significantly improves its forecasting performance at all horizons. The gain in performance is even higher when the IMRs are added as regressors in a standard AR model, especially at short horizons. Also, note that different sets of factors are selected for the IMR-DI-AR and the DI-AR models. In particular, the first two PCs are never selected for the benchmark model, arguably because these factors already play an important role in the determination of the IMR.

An interesting exercise consists of comparing the variable selection procedure described in Section 2.5 to another approach that relies on the BIC and the natural ordering of the PCs[8]. The second panel of Table 4 show the adjusted R-square for this alternative approach. When $h = 1$, the BIC selects the first three PCs while our benchmark procedure selects the F_3 and F_5. Both models have the same quality-of-fit at that horizon. However, the performance of the models selected by the BIC deteriorates rapidly as the horizon increases. This simply confirms that the natural ordering of the PCs is irrelevant for the choice of predictors to include in a forecasting model. Tables 7 - 10 in Appendix show the estimation results for all six measures of economic activity (Industrial production, employment growth, Unemployment rate, GDP deflator inflation and SP500 returns). All results are qualitatively

[8] In the latter approach, the PCs are first ranked by decreasing order of importance. Models that use an increasing number of regressors are then estimated and compared using the BIC.

similar to those described here.

Table 4: Predicting US GDP growth: adjusted R^2 from predictive regressions

Quarter	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
IMR-DI-AR	0,4041	0,3264	0,3232	0,1756	0,2385	0,1727	0,1541	0,1109
	3 5	3 5 24	3 5 8 19 21	16 19 32	3 19 31 32	19 26 31 32	20 32 35	12 16 34
			27 31 32 35					
DI-AR	0,3459	0,3053	0,2903	0,1502	0,2189	0,1585	0,1526	0,0979
	2 3	1 2 3	1 2 3 8 19	1 2 3 32	1 3 8 19	1 3 19	3 6 19	12 22 34
		11 24	21 27 31 36		31 32	31 32	20 22 32	
IMR-AR	0,3266	0,2317	0,0996	0,0865	0,0750	0,0263	0,0838	0,0427
AR	0,1004	0,0538	0,0049	0,0016	-0,0040	-0,0046	0,0069	0,0196
BIC selection of DIs								
IMR-DI-AR	0,4192	0,2317	0,0996	0,0865	0,0750	0,0263	0,0838	0,0427
	3	0	0	0	0	0	0	0
DI-AR	0,4183	0,2518	0,1363	0,0643	0,0363	0,0246	0,0069	0,0196
	3	3	3	1	1	1	0	0

This table presents adjusted R^2 results from predictive regressions, (9) for each forecasting horizon. IMR-DI-AR is our benchmark model, DI-AR is the diffusion index model from Stock and Watson (2002), IMR-AR is the simple direct AR forecasting model augmented by the Probit probability of recession. In the first panel principal components have been selected using the testing procedure from Section 2.5 with 5% level. In the second panel the number of consecutive principal components has been selected according to BIC.

Cyclicality of real activity variables: Table 5 presents the point estimates and p-values for the coefficients $\bar{\delta}_h$, $\delta_{h,0}$ and $\delta_{h,1}$. See Equation (13) for a reminder. All three coefficients should be negative for cyclical series (i.e., increasing during expansions and shrinking during recessions) and positive for countercyclical series[9]. Not surprisingly, the estimated coefficients of the GDP growth equation are all negative and significant at all horizons. This result is not trivial given that the NBER business cycles are not specifically calibrated to exactly match the movements of GDP growth. The same is true for Total Industrial Production and Employment growths.

The delta parameters of the unemployment rate (UNRATE) equation are positive in all but one case. $\bar{\delta}_h$ is positive and significant at horizons $h = 1$ to 6; $\delta_{h,0}$ is negative and non

[9]Recall that $IMR_{t,h,0}$ only takes negative values so that $\delta_{h,0} < 0$ if the variable increases during expansion episodes.

significant at $h = 4$, and positive and significant at $h = 1, 7$ and 8; finally, $\delta_{h,1}$ is positive and significant at horizons $h = 1$ to 5. The outlook for future unemployment rate is clearly nonlinear in the forecast horizon and asymmetric in the (future) states: for a given forecast horizons, $\delta_{h,1}$ is often significant while $\delta_{h,0}$ is not.

The delta coefficients of the GDP Deflator equation are rarely significant. The coefficient $\hat{\delta}_{h,1}$ is negative when it is significant, which means that the inflation rate as measured by the GDP Deflator decreases during recessions. However, $\hat{\delta}_{h,0}$ is significant and positive at horizon $h = 8$[10]. Finally, in the equation estimated for the SP500 returns, $\hat{\delta}_{h,0}$ and $\hat{\delta}_{h,1}$ are significant and negative at $h = 1, 2$ and $h = 5, 6$ respectively. Overall, the results suggest that the IMR-DI-AR model can be used to formally test the cyclicality of any economic time series.

3.3 Severity of recessions

Figure 6 shows one quarter ahead average, pessimistic and optimistic scenarios for the GDP growth rate using the forecasting formulas 8, 10 and 11 respectively. The average forecast is given by the dotted blue line. Interestingly, the actual GDP growth rate usually remains above the optimistic scenario before the beginning of recession episodes and drops near to the turning point. The pessimistic scenario is closer to the actual data around the NBER recession dates. Figure 6 shows that the actual GDP growth rates were above the pessimistic scenario during the recessions of 1970, 1991 and 2001. However, the recessions of 1974, 1980, 1981 and 2008 have been worse than suggested by the model.

Figure 7 shows the 1-quarter ahead expected and realized measures of severity using equations 16 and 17. The realized severity (gray line) have been worse than expected at the beginning of the 1980 and 1981 recessions, which suggests that the initial magnitudes of

[10]We have also tried using the inflation growth rate, i.e., the second difference of the logarithm of the GDP Deflator but the results did not change.

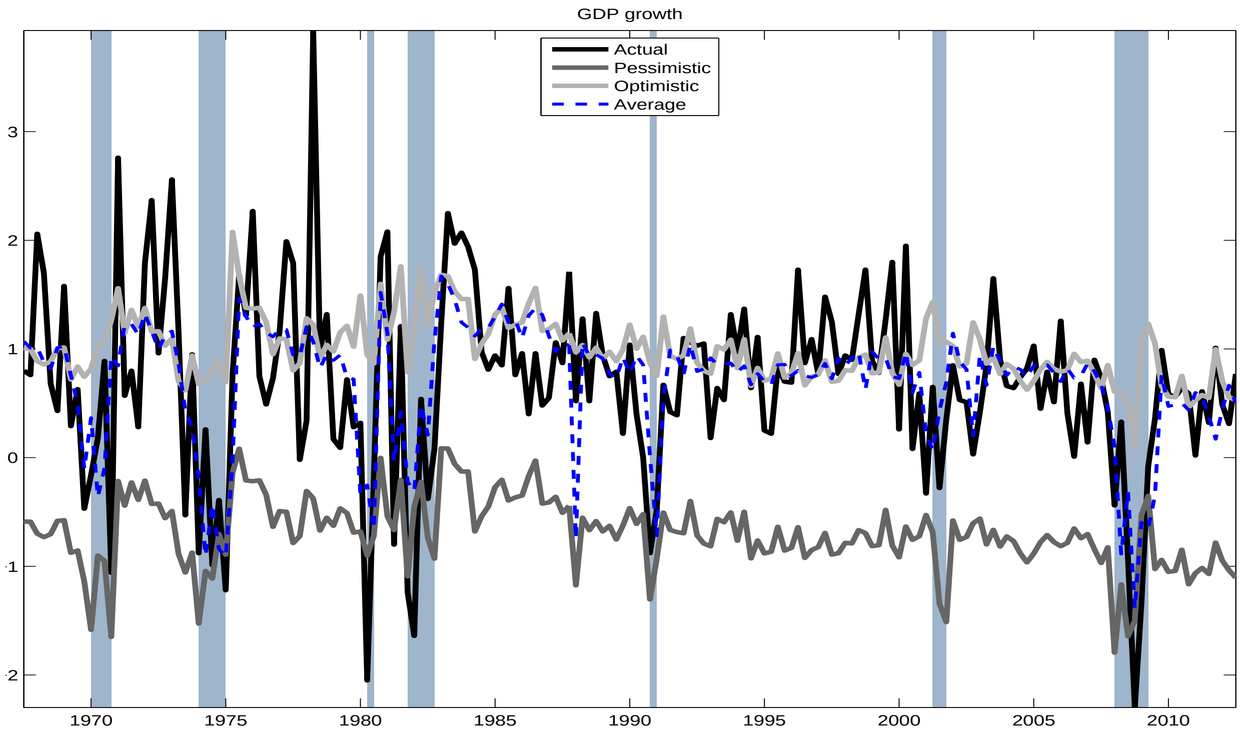
Table 5: IMR-DI-AR estimation results

		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
	$\bar{\delta}_h$	-1,5874	-1,5631	-1,7229	-1,5316	-1,7740	-1,2329	-1,9074	-2,3470
		(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,005)	(0,001)	(0,002)
GDP	$\delta_{h,0}$	-0,4198	-0,8040	-1,1442	-0,9462	-1,3016	-1,0193	-0,9562	-2,0341
		(0,055)	(0,000)	(0,000)	(0,000)	(0,000)	(0,002)	(0,014)	(0,000)
	$\delta_{h,1}$	-0,6995	-0,6124	-0,5236	-0,6781	-0,5046	-0,7140	-0,7048	-0,4682
		(0,000)	(0,000)	(0,000)	(0,000)	(0,001)	(0,000)	(0,000)	(0,003)
	$\bar{\delta}_h$	-2,4892	-3,9890	-3,6454	-3,7307	-3,4312	-3,7588	-5,0540	-7,3398
		(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,005)	(0,000)	(0,000)
INDPRO	$\delta_{h,0}$	-1,3621	-2,1747	-2,0155	-2,1809	-2,3287	-2,4923	-2,1350	-4,4852
		(0,000)	(0,000)	(0,001)	(0,002)	(0,000)	(0,008)	(0,010)	(0,000)
	$\delta_{h,1}$	-0,9690	-0,9618	-1,3370	-1,4080	-1,4744	-1,5180	-1,3380	-1,0275
		(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,001)
	$\bar{\delta}_h$	0,7399	1,5049	2,6095	2,7920	1,8278	3,2355	0,0618	1,3163
		(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,961)	(0,319)
UNRATE	$\delta_{h,0}$	0,2093	0,2913	0,3526	-0,4694	0,1095	0,4645	1,4908	1,3522
		(0,007)	(0,211)	(0,269)	(0,116)	(0,663)	(0,162)	(0,009)	(0,093)
	$\delta_{h,1}$	0,2274	0,4318	0,4688	0,7374	0,5318	0,2773	0,0250	0,1728
		(0,000)	(0,000)	(0,001)	(0,000)	(0,003)	(0,172)	(0,898)	(0,549)
	$\bar{\delta}_h$	-0,5386	-0,8472	-0,9333	-1,2245	-1,3131	-1,5859	-2,1505	-2,0683
		(0,000)	(0,000)	(0,011)	(0,000)	(0,000)	(0,000)	(0,000)	(0,001)
EMPL	$\delta_{h,0}$	-0,2822	-0,3063	-0,6040	-0,6116	-0,7314	-0,7800	-0,6834	-1,0217
		(0,000)	(0,044)	(0,004)	(0,001)	(0,000)	(0,001)	(0,019)	(0,001)
	$\delta_{h,1}$	-0,2098	-0,2529	-0,2682	-0,3660	-0,4293	-0,3400	-0,3945	-0,3717
		(0,001)	(0,000)	(0,002)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
	$\bar{\delta}_h$	0,0563	0,2130	0,5443	0,9749	0,3167	-0,1127	0,7562	2,7666
		(0,677)	(0,122)	(0,013)	(0,000)	(0,275)	(0,779)	(0,183)	(0,000)
GDPDEF	$\delta_{h,0}$	0,0892	0,1550	0,2598	0,1099	-0,0545	0,0483	0,1916	1,0000
		(0,407)	(0,435)	(0,170)	(0,527)	(0,763)	(0,811)	(0,308)	(0,001)
	$\delta_{h,1}$	0,0607	-0,0269	-0,1696	-0,1759	-0,0555	0,0100	0,0365	-0,2246
		(0,473)	(0,664)	(0,028)	(0,011)	(0,491)	(0,916)	(0,503)	(0,005)
	$\bar{\delta}_h$	-4,0782	-5,0794	-4,0531	-4,7855	-0,7150	2,4236	-9,8337	-0,1316
		(0,138)	(0,121)	(0,410)	(0,331)	(0,876)	(0,706)	(0,190)	(0,989)
SP500	$\delta_{h,0}$	-3,7451	-3,6961	-3,3019	-4,6478	-0,4984	-1,3772	-6,3313	-3,4509
		(0,036)	(0,048)	(0,255)	(0,154)	(0,869)	(0,764)	(0,200)	(0,588)
	$\delta_{h,1}$	-0,8998	-3,1836	-2,7299	-1,4891	-3,9951	-4,1480	-1,3268	-2,4387
		(0,717)	(0,152)	(0,142)	(0,530)	(0,052)	(0,007)	(0,430)	(0,193)

Point estimates of IMR coefficients $\bar{\delta}_h$, $\delta_{h,0}$ and $\delta_{h,1}$ from equation (13). P-values are in parenthesis.

these downturns were unexpected. In 1974 and 2009, the realized severity of the recessions has been worse than our pessimistic scenario towards the end of the recessions. This is

Figure 6: Predicting US GDP growth: 1-quarter ahead scenarios

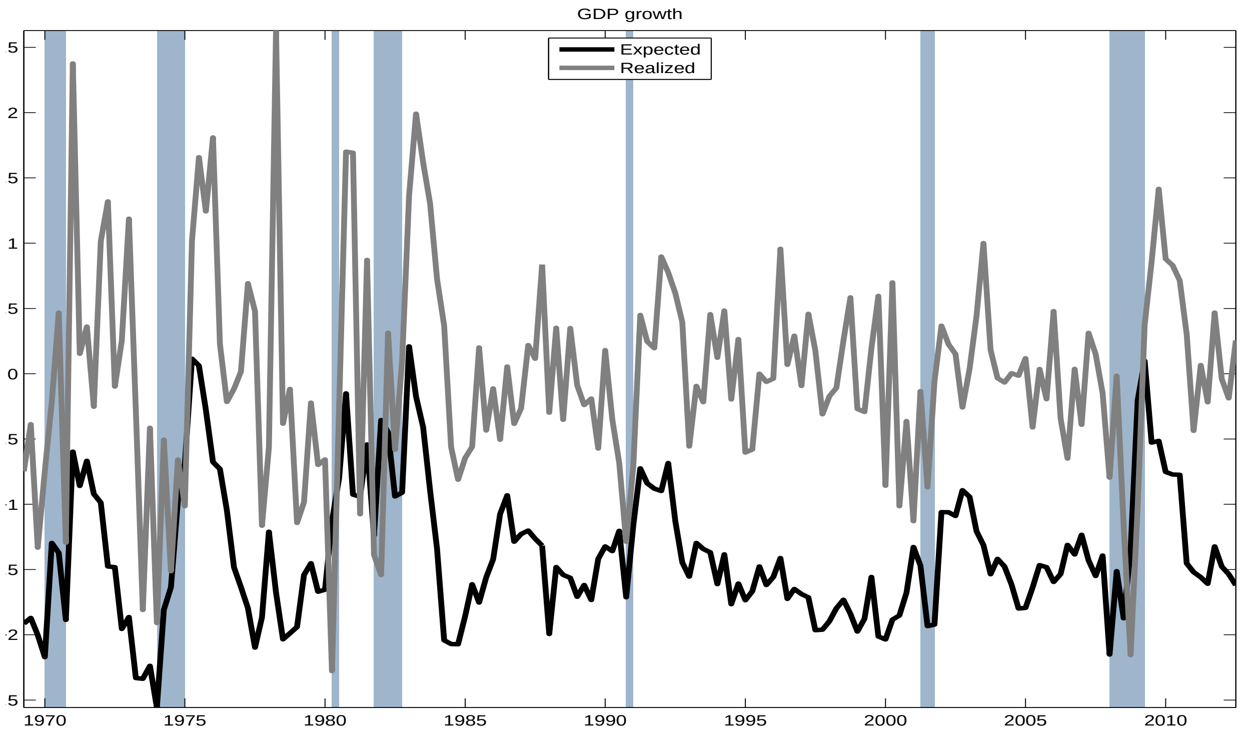


Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

particularly true for the Great Recession of 2009.

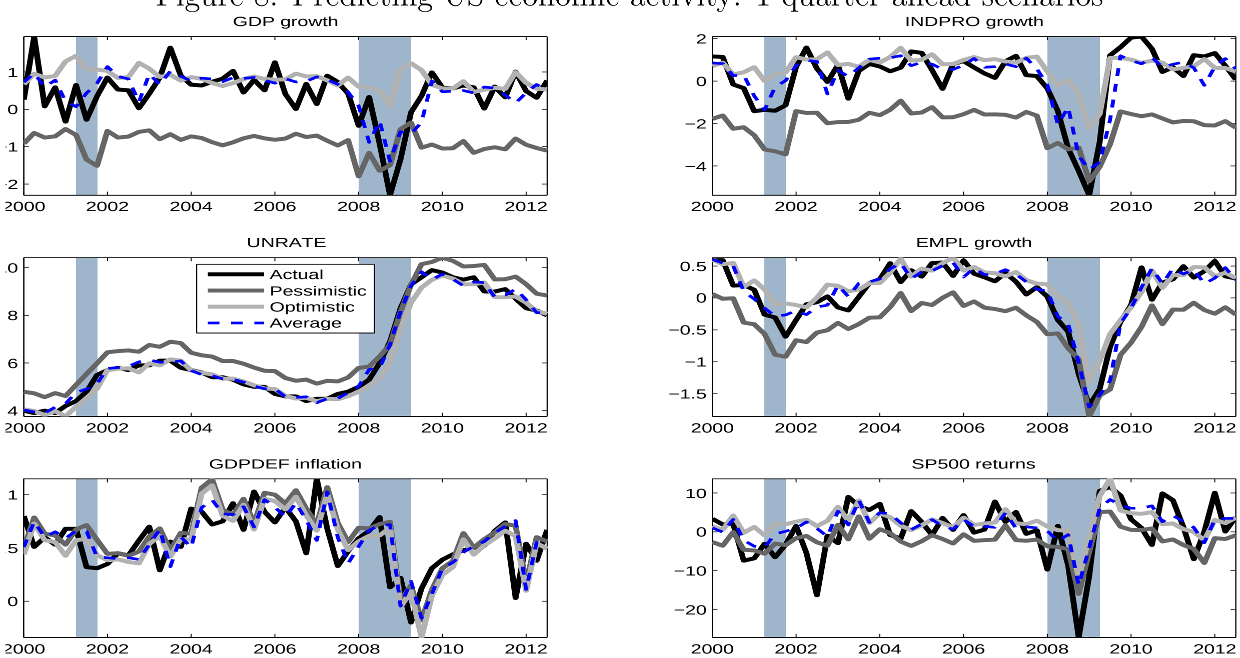
Figures 8 and 9 show the forecast scenarios and severity of recessions for all six measures of economic activity during 2000 - 2013. The severity of the last recession was less anticipated by our model compared to 2001. There have been surprises towards the end of the recession: at the end of 2008 for GDP growth, during 2009 for Industrial production, unemployment rate and SP500, etc. According to (Stock & Watson 2012*a*), the Great Recession was not different from the others with respect to its roots. Our results suggest that its severity could not be correctly anticipated. (Ng & Wright 2013) suggest that recessions are not all alike over the US business cycle. Our results suggest that recessions differ across time with respect to their degree of predictability and severity.

Figure 7: Predicting severity of recessions: 1-quarter ahead scenarios



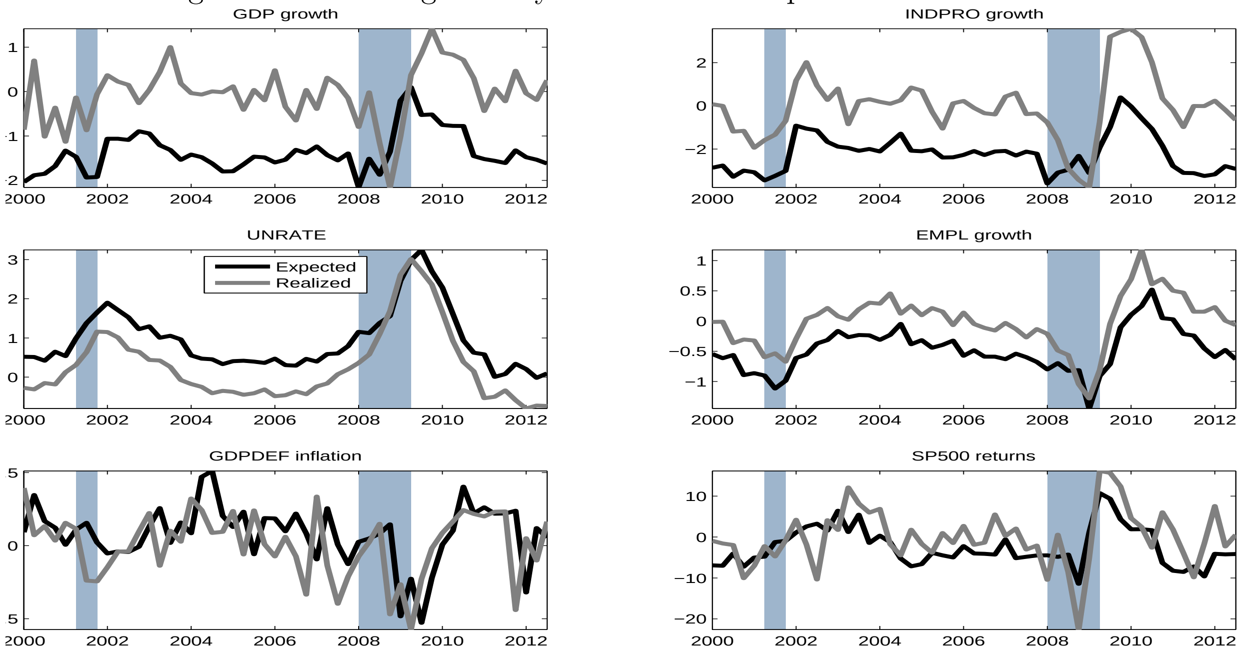
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 8: Predicting US economic activity: 1-quarter ahead scenarios



Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

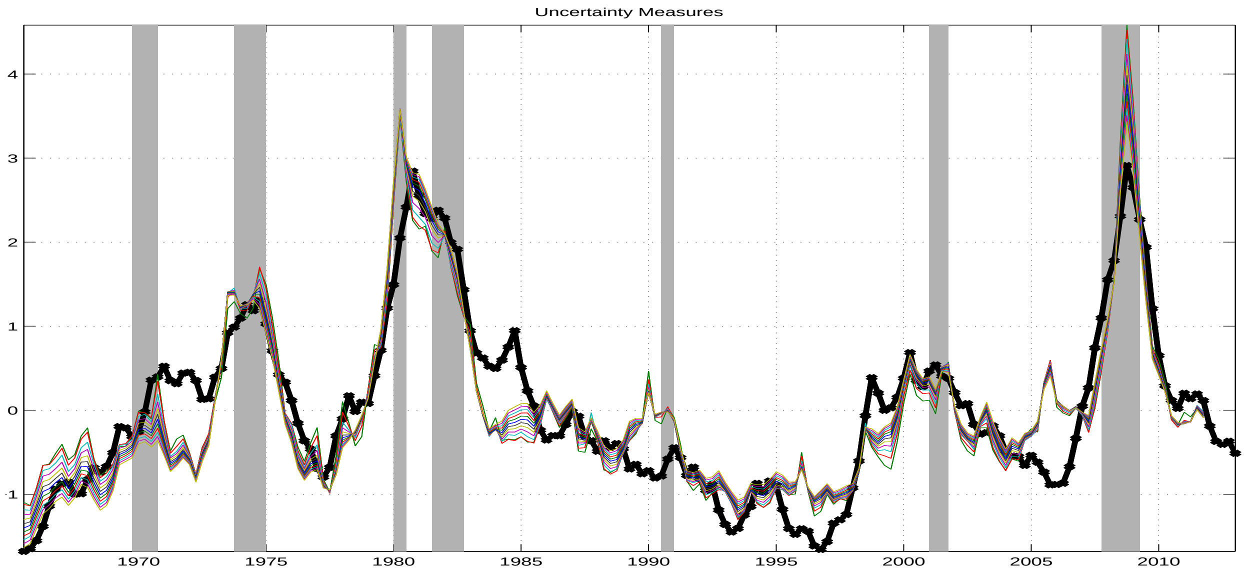
Figure 9: Predicting severity of recessions: 1-quarter ahead scenarios



Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

It is difficult to explain in the current framework why the decrease in real activity during the second half of Great Recession became more unpredictable than before. However, we note that some empirical measures of macroeconomic uncertainty have peaked during those periods. Figure 10 shows the macroeconomic uncertainty measures calculated by (Jurado et al. 2013) and (Amir-Ahmadi & Stevanovic 2014). The authors define four episodes where the uncertainty level was high. Interestingly, these episodes coincide with the recessions where the realized severity has worse than predicted by the IMR-DI-AR model. The fact that macroeconomic uncertainty rises during periods where the severity of recessions is larger than expected suggests that some important mechanisms at the roots of these recessions remains unexplained.

Figure 10: Empirical measures of macroeconomic uncertainty



The black line is the stochastic volatility factor from (Amir-Ahmadi & Stevanovic 2014) obtained from a 5-variable Factor-TVP-VAR that includes GDP growth, GDP deflator inflation, Federal funds rate, Business loans growth and Credit spread. The other lines are measures of common macroeconomic uncertainty, for different horizons and aggregated to quarterly frequency, from (Jurado et al. 2013). These are obtained as common volatiities of forecasting equations of hundreds of macroeconomic series.

4 Conclusion

This paper proposes a framework to predict the probability and severity of US recessions in a data-rich environment. We employ a principal component analysis to decompose the candidate predictors available to us into a set of uncorrelated variables. This approach allows us to account for variables that are highly, but imperfectly correlated in the analysis. Next, we design parsimonious Probit models to predict the probability of a recession h periods ahead, for h varying between 1 and 8 quarters. The quality-of-fit of the Probit models are measured using a novel metrics derived from the Kullback-Leibler Information Criterion. The same metrics serves test statistic to assess the significance of the predictive power of the principal components used as regressors in the Probit models. Finally, we utilize an autoregressive model augmented with inverse Mills ratios and diffusion indices (i.e., the principal components) to generate forecasts of real economic activity that are conditional on the future states of the economy. Indeed, the IMR-DI-AR model is able to generate an average, an optimistic and a pessimistic forecast. The optimistic forecast relies on the assumption that there will be an expansion at the forecast horizon of interest while the pessimistic forecast assumes the opposite. The severity of recessions is defined as the gap between the pessimistic scenario and the medium run trend of the series. Our results support that the occurrence and severity of U.S. recessions are predictable to a great extent. Some are more predictable than others while some are more severe than expected

References

Aastveit, K. A., Carriero, A., Clark, T. E. & Marcellino, M. (2014), Have standard vars remained stable since the crisis?, Technical report, Manuscript.

Amir-Ahmadi, P. & Stevanovic, D. (2014), Common sources of instabilities in macroeconomic dynamics, Technical report, Department of Economics, UQAM.

Anderson, H. M. & Vahid, F. (2001), 'Predicting the probability of a recession with nonlinear autoregressive leading-indicator models', *Macroeconomic Dynamics* **5**, 482–505.

Chauvet, M. (1998), 'An econometric characterization of business cycles with factor structure and markov switching', *International Economic Review* **39**, 969–996.

Chauvet, M. & Hamilton, J. (2006), Dating business cycle turning points, *in* P. R. C" Milas & D. van Dikj, eds, 'Nonlinear Time Series Analysis of Business Cycles', Vol. 77, North Holland, Amsterdam, pp. 1–54.

Chauvet, M. & Piger, J. (2008), 'A comparison of the real time performance of business cycle dating methods', *Journal of Business and Economic Statistics* **26**, 42–49.

Clark, T. E. K. & McCracken, W. (2013), Evaluating the accuracy of forecasts from vector autoregressions, *in* T. Fomby, L. Kilian & A. Murphy, eds, 'Advances in Econometrics: VAR models in Macroeconomics New Developments and Applications'.

Doan, T., Litterman, R. & Sims, C. (1984), 'Forecasting and conditional projection using realistic prior distributions (with discussion)', *Econometric Reviews* **3**, 1–144.

Dueker, M. (2005), 'Dynamic forecasts of qualitative variables: A qual var model of u.s. recessions', *Journal of Business & Economic Statistics* **23**, 96–104.

Dueker, M. & Wesche, K. (2005), 'Forecasting macro variables with a qual var business cycle turning point index', *Applied Economics* **42**(23), 2909–2920.

Estrella, A. & Mishkin, F. (1998), 'Predicting us recessions: Financial variables as leading indicators', *Review of Economics and Statistics* **80**, 45–61.

Fair, R. C. (1993), Estimating event probabilities from macroeconometric models using stochastic simulation, *in* J. H. Stock & M. W. Watson, eds, 'Business Cycles, Indicators, and Forecasting', University of Chicago Press.

Giannone, D., Lenza, M., Momferatou, D. & Onorante, L. (2010), Short-term ination projections: A bayesian vector autoregressive approach, Technical report, European Central Bank.

Hamilton, J. D. (2011), 'Calling recessions in real time', *International Journal of Forecasting* **27**, 1006–1026.

Issler, J. & Vahid, F. (2006), 'The missing link: using the nber recession indicator to construct coincident and leading indices of economic activity', *Journal of Econometrics* **132**, 281–303.

Jurado, K., Ludvigson, S. & Ng, S. (2013), Measuring uncertainty, Technical report, Columbia University.

Meyer, B. & Zaman, S. (2013), Its not just for ination: The usefulness of the median cpi in bvar forecasting, Working Paper 13-03, Federal Reserve Bank of Cleveland.

Ng, S. (2013), Boosting recessions, Technical report, Columbia University.

Ng, S. & Wright, J. (2013), 'Facts and challenges from the great recession for forecasting and macroeconomic modeling', *Journal of Economic Literature* **51**(4), 1120–1154.

Rudebusch, G. D. & Williams, J. (2009), 'Forecasting recessions: The puzzle of the enduring power of the yield curve', *Journal of Business and Economic Statistics* **27(4)**, 492–503.

Schorfheide, F. & Song, D. (2013), Real time forecasting with a mixed-frequency var, Technical report, NBER.

Sims, C. (1982), 'Policy analysis with econometric models', *Brookings Papers on Economic Activity* pp. 107–152.

Stock, J. H. & Watson, M. W. (1989), 'New indexes of coincident and leading economic indicators', *NBER Macroeconomics Annual* pp. 351–393.

Stock, J. H. & Watson, M. W. (2002b), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics* **20**(2), 147–162.

Stock, J. H. & Watson, M. W. (2010), 'Indicators for dating business cycles: Cross-history selection and comparisons', *American Economics Review: Papers and Proceedings* **100**.

Stock, J. H. & Watson, M. W. (2012*a*), 'Disentangling the channels of the 2007-2009 recession', *Brookings Papers on Economic Activity* .

Stock, J. H. & Watson, M. W. (2012*b*), 'Estimating turning points using large data sets', *Journal of Econometrics* p. forthcoming.

Appendix A: Description of the data

The transformation codes are: 1 - no transformation; 2 - first difference; 4 - logarithm; 5 - first difference of logarithm; 0 - variable not used in the estimation (only used for transforming other variables). A * indicate a series that is deflated by the Personal Consumption Expenditures: Chain-type Price Index. GDP and GDPDEF are observed quarterly and are not in X_t.

Table 6: Data used to construct the diffusion indices

INDPRO	5	Industrial Production Index
UNRATE	1	Civilian Unemployment Rate
PAYEMS	5	All Employees: Total nonfarm
MANEMP	5	All Employees: Manufacturing
USMINE	5	All Employees: Mining and logging
IC4WSA	4	4-Week Moving Average of Initial Claims
RPCE	5*	Real Personal Consumption Expenditures
RPCEDG	5*	Real Personal Consumption Expenditures: Durable Goods
AWHMAN	1	Average Weekly Hours: Manufacturing
AWOTMAN	1	Average Weekly Overtime Hours: Manufacturing
NAPM	1	ISM Manufacturing: PMI Composite Index
NAPMOI	1	ISM Manufacturing: New Orders Index
NAPMEI	1	ISM Manufacturing: Employment Index
NAPMII	1	ISM Manufacturing: Inventories Index
NAPMSDI	1	ISM Manufacturing: Supplier Deliveries Index
NAPMPRI	1	ISM Manufacturing: Prices Index
CUmftg	1	Capacity Utilization: Manufacturing
CPILFESL	5	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy
CPIAUCSL	5	Consumer Price Index for All Urban Consumers: All Items
PCEPI	5	Personal Consumption Expenditures: Chain-type Price Index
M1SL	5	M1 Money Stock
M2SL	5	M2 Money Stock
HOUST	4	Housing Starts: Total: New Privately Owned Housing Units Started
PERMIT	4	New Private Housing Units Authorized by Building Permits
ConsMICH	1	University of Michigan: Consumer Sentiment
OILPRICE	5	Spot Oil Price: West Texas Intermediate
FFR	1	Effective Federal Funds Rate
INVEST	5	Securities in Bank Credit at All Commercial Banks
TB3MS	1	3-Month Treasury Bill: Secondary Market Rate
GS1	0	1-Year Treasury Constant Maturity Rate
GS5	1	5-Year Treasury Constant Maturity Rate
GS10	0	10-Year Treasury Constant Maturity Rate
SP500	5	S&P 500 Stock Price Index
DJIA	5	Dow Jones Industrial Average
BAA	0	Moodyś Seasoned Baa Corporate Bond Yield
AAA	0	Moodyś Seasoned Aaa Corporate Bond Yield
BAA-GS10	1	
BAA-AAA	1	
BAA-FFR	1	
GS10-TB3MS	1	
GS5-FFR	1	
GS1-FFR	1	
SP500-RV	1	S&P500: realized volatility
SP500-SK	1	S&P500: realized skewness
DJIA-RV	1	DJIA: realized volatility
DJIA-SK	1	DJIA: realized skewness
GDP	5	Real Gross Domestic Product
GDPDEF	5	GDP Deflator

Appendix B: Additional estimation results

Table 7: Adjusted R^2 and selected factors from IMR-DI-AR predictive regressions

	h=1	h=2	h=3	h=4
GDP	0,4041 [3 5]	0,3264 [3 5 24]	0,3232 [3 5 8 19 21 27 31 32 35]	0,1756 [16 19 32]
INDPRO	0,5544 [3 25]	0,4816 [3 6 19 25 27 34]	0,3807 [1 3 11 16 31 32 35]	0,3369 [3 14 16 19 32]
UNRATE	0,9839 [1 4 9 27 33]	0,9526 [1 4 22 24 27 32]	0,9248 [1 3 16 20 27 32 35]	0,8935 [1 2 3 16 27 32]
EMPL	0,8279 [3 17 24 27 40]	0,7564 [3 6 19 21 24 27 32 35]	0,6630 [3 11 16 19 21 27 32 35]	0,6213 [3 10 14 16 19 21 27 32]
GDPDEF	0,7795 [1 10 26]	0,7458 [1 6 27 33]	0,7665 [1 4 6 22 24 26 27 34]	0,7995 [1 3 4 5 6 18 22 27 32 39]
SP500	0,2674 [1 8 13 27]	0,1753 [1 8 16 27]	0,1645 [1 6 14 17 32]	0,1033 [6 8 17]

	h=5	h=6	h=7	h=8
GDP	0,2385 [3 19 31 32]	0,1727 [19 26 31 32]	0,1541 [20 32 35]	0,1109 [12 16 34]
INDPRO	0,3199 [3 10 16 19 31 32]	0,2194 [19 23 28 31 32]	0,2289 [22 26]	0,1675 [19 32 35]
UNRATE	0,8753 [1 2 3 5 16 21 27 32 33]	0,8471 [1 2 3 5 16 21 27 31 32 33]	0,7535 [1 2 3 4 5 8 31 32]	0,7167 [2 3 4 8 19 31 32]
EMPL	0,5072 [3 10 19 31 32 33]	0,4910 [3 4 10 19 31 32]	0,4696 [3 4 8 10 26 31 32 37]	0,4634 [3 4 8 10 16 19 31 32 34 35 37]
GDPDEF	0,7252 [1 3 4 5 6 10 18 27 32]	0,7233 [1 2 3 4 5 6 10 18 22 32]	0,7318 [2 3 4 6 10 13 19 22 26 32 34]	0,6762 [2 3 4 6 7 10 12 32]
SP500	0,0385 [6 38]	0,0727 [6 19 40]	0,0739 [6 14 27]	0,0313 [7]

This table presents goodness-of-fit results for IMR-DI-AR predictive regressions for each forecasting horizon. The first row, e.g. in GDP equation, contains the adjusted R^2 while the second row enumerate the principal components that have been selected by a 5% F-test.

Table 8: Adjusted R^2 and retained factors from DI-VAR predictive regressions

	h=1	h=2	h=3	h=4
GDP	0,3459 [2 3]	0,3053 [1 2 3 11 24]	0,2903 [1 2 3 8 19 21 27 31 36]	0,1502 [1 2 3 32]
INDPRO	0,5834 [2 3 6 17 24 25]	0,4127 [1 2 3 6 22 27]	0,3593 [1 2 3 11 16 31]	0,3106 [1 2 3 16 32]
UNRATE	0,9838 [1 2 3]	0,9521 [1 2 3 27]	0,9113 [1 2 3 27]	0,8908 [1 2 3 16 27 32]
EMPL	0,8246 [2 3 6 17 24 27]	0,7265 [2 3 6 21 24 27 32 35]	0,6475 [1 2 3 11 16 21 27 32 35]	0,5749 [1 2 3 10 16 19 27 32]
GDPDEF	0,7848 [1 6 10 26]	0,7447 [1 6 27 33]	0,7626 [1 4 6 22 24 26 27 34]	0,7755 [1 4 5 6 18 22 27 32 39]
SP500	0,2644 [1 8 13 27]	0,1573 [1 8 16 27]	0,1421 [1 6 14 17]	0,1276 [1 6 8 17]
	h=5	h=6	h=7	h=8
GDP	0,2189 [1 3 8 19 31 32]	0,1585 [1 3 19 31 32]	0,1526 [3 6 19 20 22 32]	0,0979 [12 22 34]
INDPRO	0,3018 [1 2 3 16 19 20 31 32]	0,2140 [1 2 3 19 23 28 31 32]	0,1861 [1 19 22 26]	0,1506 [5 19 22]
UNRATE	0,8631 [1 2 3 4 8 16 21 27 33]	0,8175 [1 2 3 4 5 8 16 21 27 31]	0,7444 [1 2 3 4 5 8 31 32]	0,7183 [2 3 4 8 19 31 32]
EMPL	0,5221 [1 2 3 8 10 19 21 31 32 33]	0,4678 [1 2 3 4 8 10 19 31 32]	0,3778 [1 3 8 10 19 31 32]	0,3421 [1 3 8 10 19 31 32]
GDPDEF	0,7237 [1 4 5 6 10 18 22 27 32]	0,7238 [1 3 4 5 6 10 18 22 32]	0,7331 [1 2 3 4 5 6 10 13 22 26 32 34]	0,6900 [1 2 3 4 5 6 10 12 22 27 32]
SP500	0,0641 [6 8 38]	0,0774 [6 19 40]	0,0641 [6 14 27]	0,0369 [7]

This table presents goodness-of-fit results for DI-AR predictive regressions for each forecasting horizon. The first row, e.g. in GDP equation, contains the adjusted R^2 while the second row enumerate the principal components that have been chosen by a 5% F-test.

Table 9: Adjusted R^2 from IMR-AR and standard AR predictive regressions

IMR-AR	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
GDP	0,3266	0,2317	0,1052	0,1137	0,0750	0,0594	0,0838	0,0515
INDPRO	0,5107	0,3812	0,2185	0,2105	0,1495	0,1564	0,1698	0,1401
UNRATE	0,9794	0,9228	0,8671	0,7199	0,5672	0,4815	0,2494	0,1713
EMPL	0,7759	0,6007	0,4218	0,3087	0,1897	0,1668	0,2098	0,1999
GDPDEF	0,7611	0,6853	0,6397	0,6193	0,5286	0,4664	0,4055	0,3761
SP500	0,1003	0,0113	0,0081	0,0014	-0,0014	-0,0085	0,0169	-0,0048
AR	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
GDP	0,1004	0,0538	0,0049	0,0016	-0,0040	-0,0046	0,0069	0,0196
INDPRO	0,3412	0,0621	0,0201	-0,0056	0,0188	0,0094	0,0193	0,0623
UNRATE	0,9555	0,8561	0,7290	0,5930	0,4658	0,3526	0,2537	0,1711
EMPL	0,6905	0,3941	0,2015	0,0754	0,0187	0,0005	-0,0057	0,0092
GDPDEF	0,7618	0,6869	0,6414	0,6066	0,5072	0,4558	0,3920	0,3613
SP500	0,1047	-0,0048	-0,0043	-0,0052	-0,0028	-0,0027	0,0066	0,0005

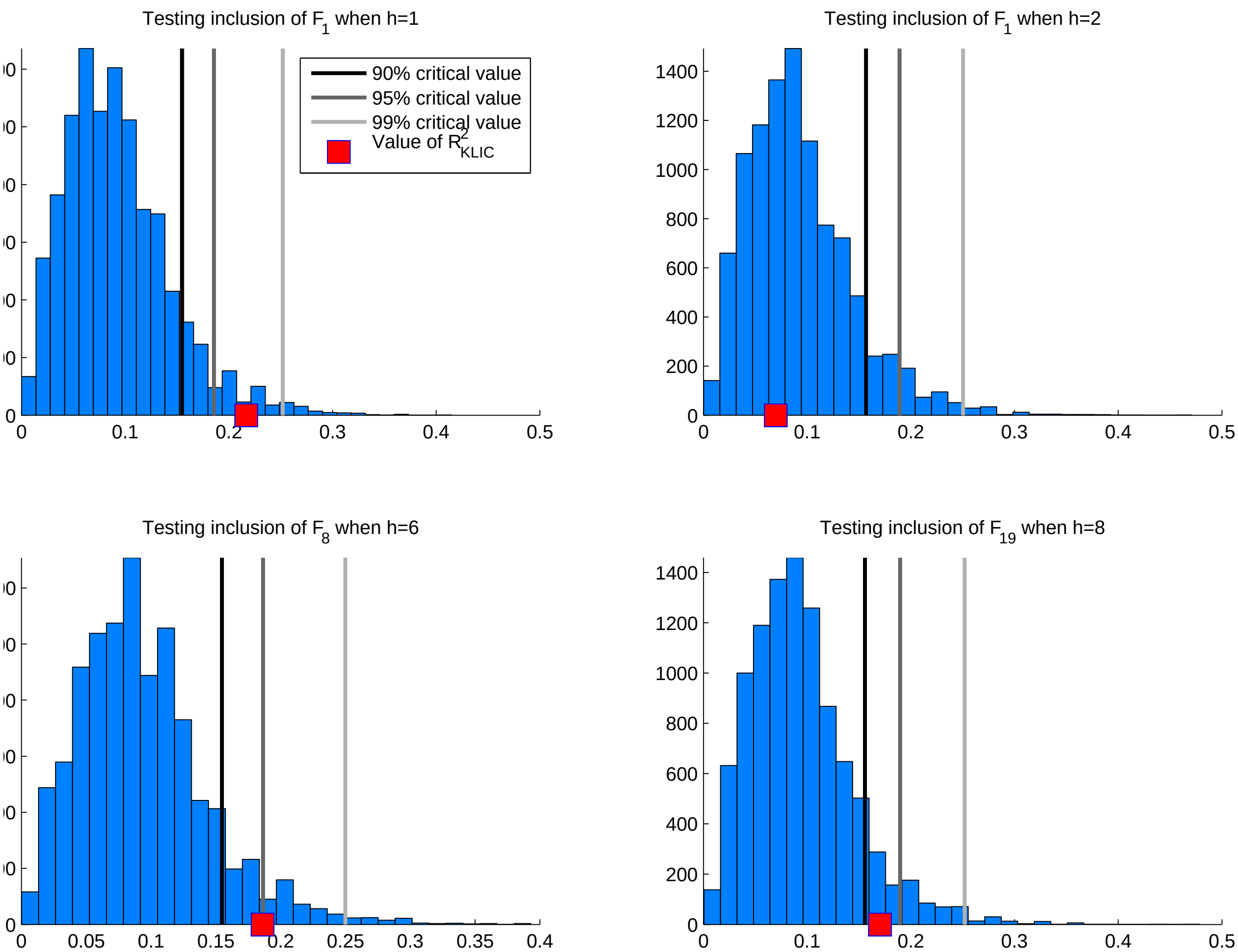
This table presents the adjusted R^2 results for IMR-VA and standard AR predictive regressions for each forecasting horizon.

Table 10: Adjusted R^2 and number of selected factors from IMR-DI-AR and DI-AR predictive regressions

	h=1		h=2		h=3		h=4	
	IMR-DI	DI	IMR-DI	DI	IMR-DI	DI	IMR-DI	DI
GDP	0,4192 3	0,4183 3	0,2317 0	0,2518 3	0,1052 0	0,1363 3	0,1137 0	0,0643 1
INDPRO	0,5601 3	0,5554 3	0,3812 0	0,3489 3	0,2664 1	0,2889 3	0,2105 0	0,2317 3
UNRATE	0,9838 2	0,9838 3	0,9572 7	0,9555 7	0,9353 13	0,9339 13	0,9143 17	0,9148 17
EMPL	0,8012 3	0,7975 3	0,6007 0	0,6054 3	0,5018 3	0,4891 3	0,3993 3	0,3923 3
GDPDEF	0,8081 6	0,8071 6	0,7733 6	0,7715 6	0,7721 6	0,7735 6	0,7366 6	0,7343 6
SP500	0,1597 1	0,1459 1	0,0674 1	0,0471 1	0,0389 1	-0,0043 0	0,0014 0	-0,0052 0
	h=5		h=6		h=7		h=8	
GDP	0,0750 0	0,0363 1	0,0594 0	0,0246 1	0,0838 0	0,0069 0	0,0515 0	0,0196 0
INDPRO	0,1495 0	0,1637 3	0,1564 0	0,0958 1	0,1698 0	0,0724 1	0,1401 0	0,0623 0
UNRATE	0,8939 17	0,8944 17	0,8492 17	0,8488 17	0,7879 17	0,7874 17	0,6725 8	0,6555 8
EMPL	0,3262 3	0,2937 3	0,3037 4	0,2387 3	0,2939 3	0,1883 3	0,1999 0	0,1024 1
GDPDEF	0,6558 6	0,6568 6	0,6237 6	0,6260 6	0,5846 6	0,5858 6	0,6159 10	0,6181 10
SP500	-0,0014 0	-0,0028 0	-0,0085 0	-0,0027 0	0,0169 0	0,0066 0	-0,0048 0	0,0005 0

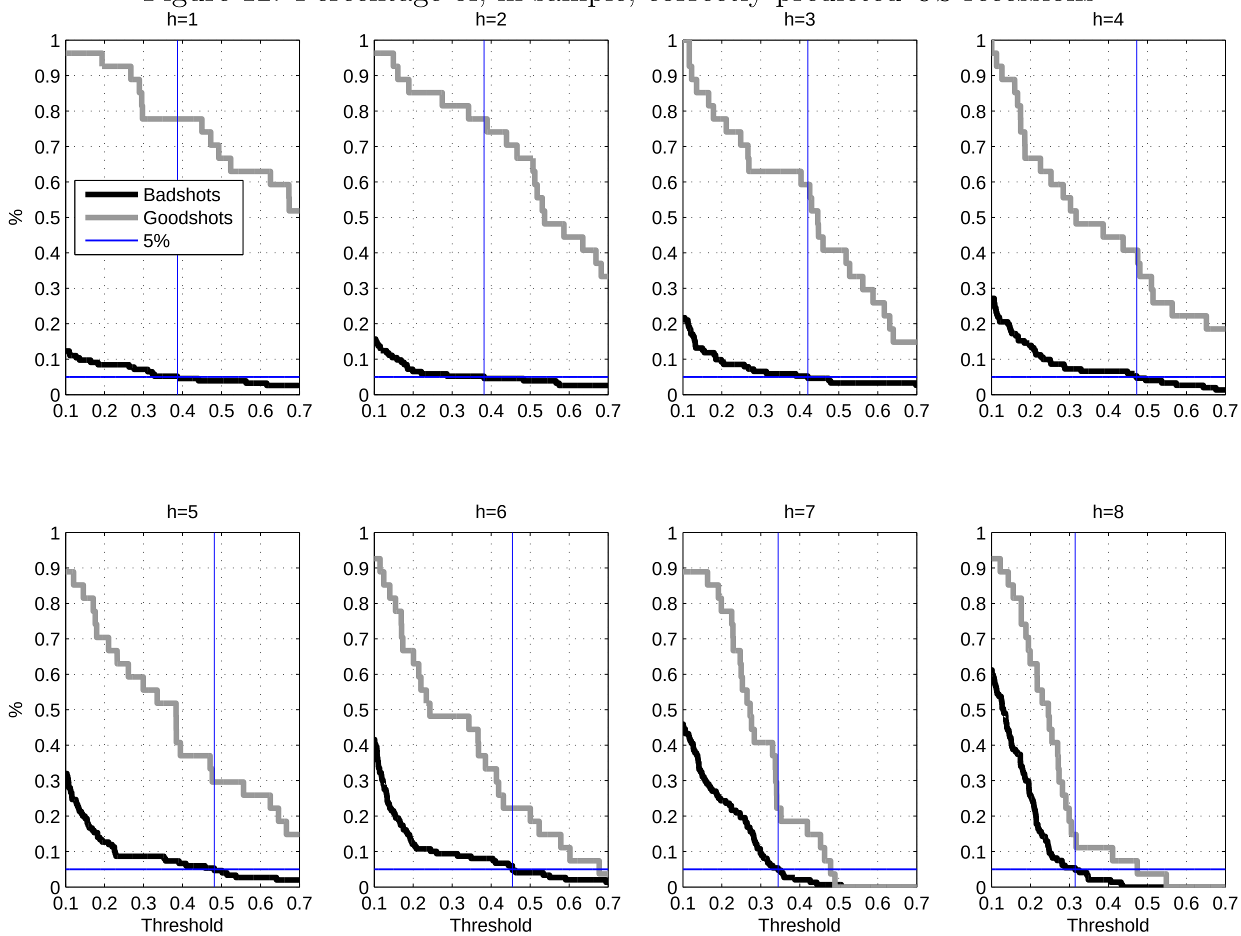
This table presents goodness-of-fit results for IMR-DI-AR and Di-AR predictive regressions for each forecasting horizon. The first row, e.g. in GDP equation, contains the adjusted R^2 for IMR-DI and DI autoregressive models, while the second row enumerate the number principal components that have been chosen by BIC for each model.

Figure 11: Examples of testing factors inclusion in predictive Probit models



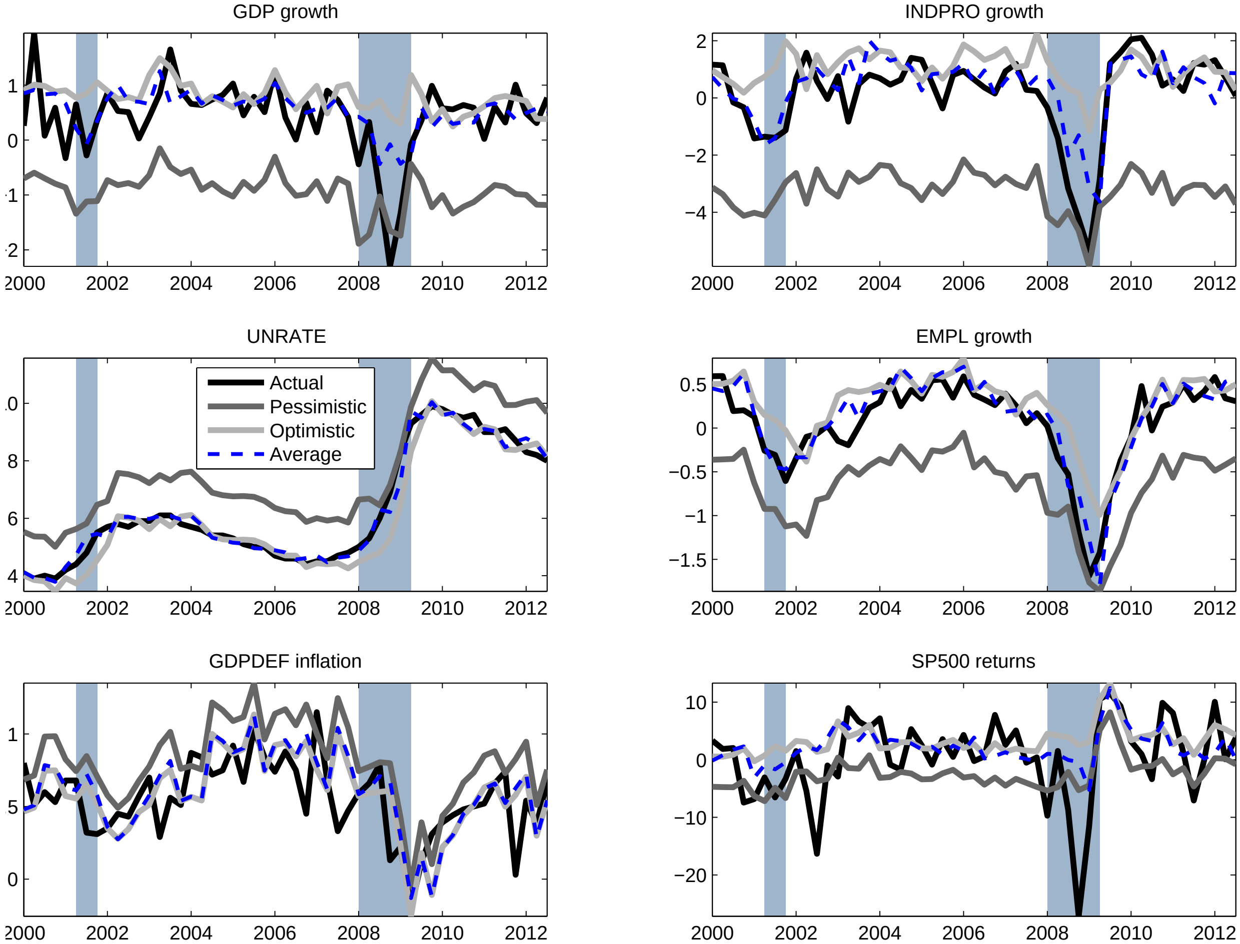
R^2_{KLIC} distribution under H_0 for different choices of principal components and horizons, according to the testing procedure in Section 2.3.

Figure 12: Percentage of, in-sample, correctly predicted US recessions



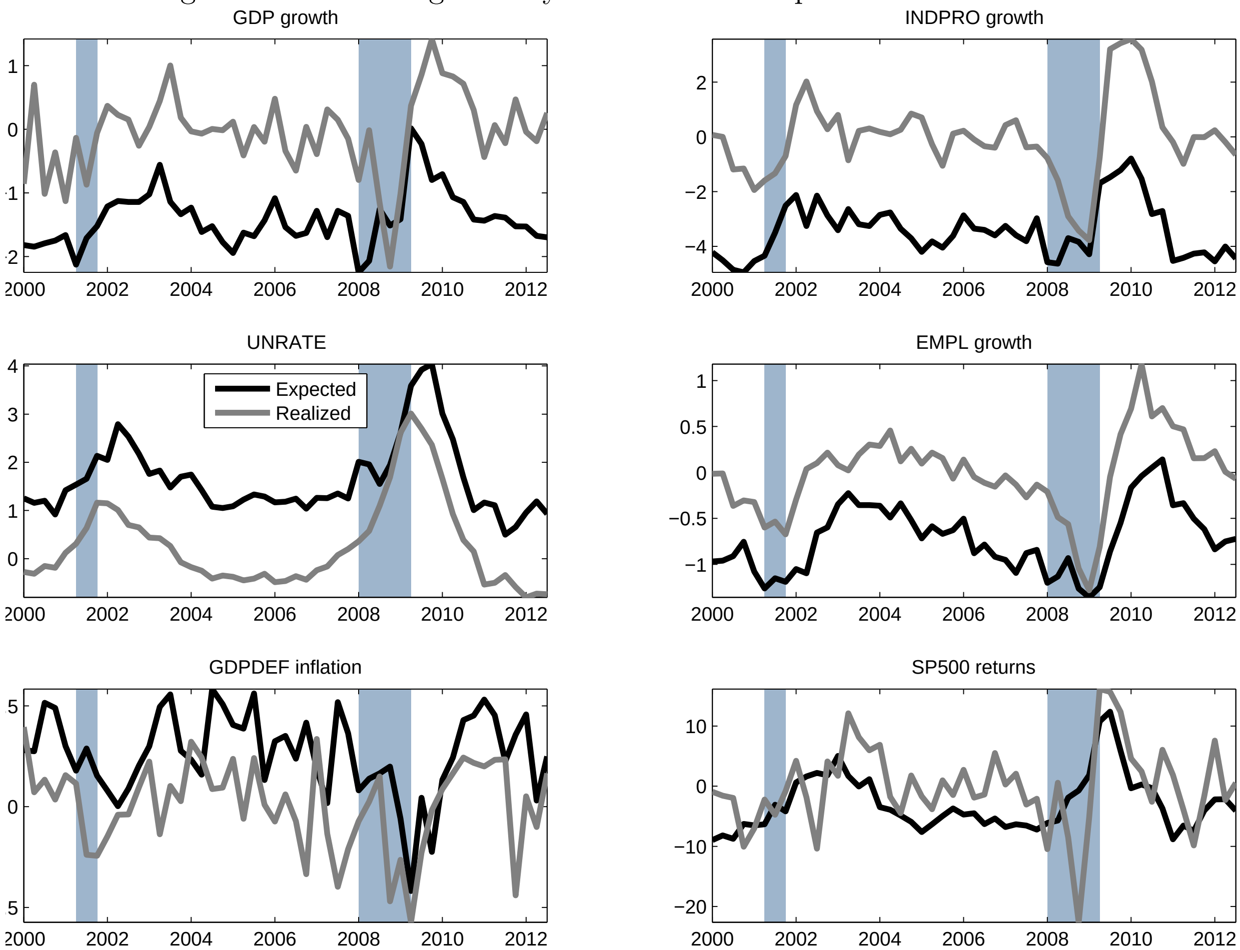
Predicted in-sample goodshots and badshots of NBER recessions using predicted probabilities from Probit model in (3). For each horizon the Probit model is specified according to testing procedure in Section 2.3.

Figure 13: Predicting US economic activity: 2-quarter ahead scenarios



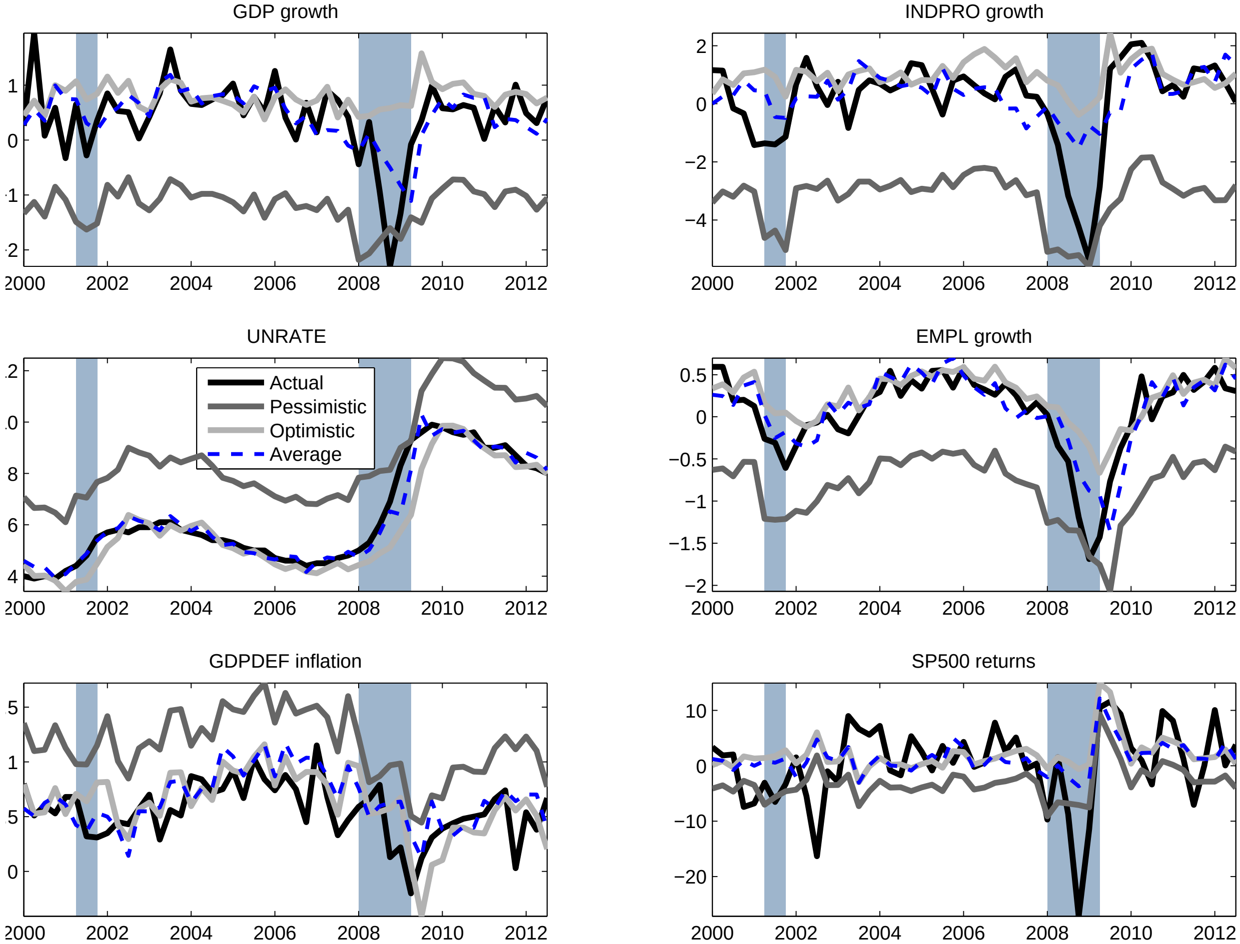
Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 14: Predicting severity of recessions: 2-quarter ahead scenarios



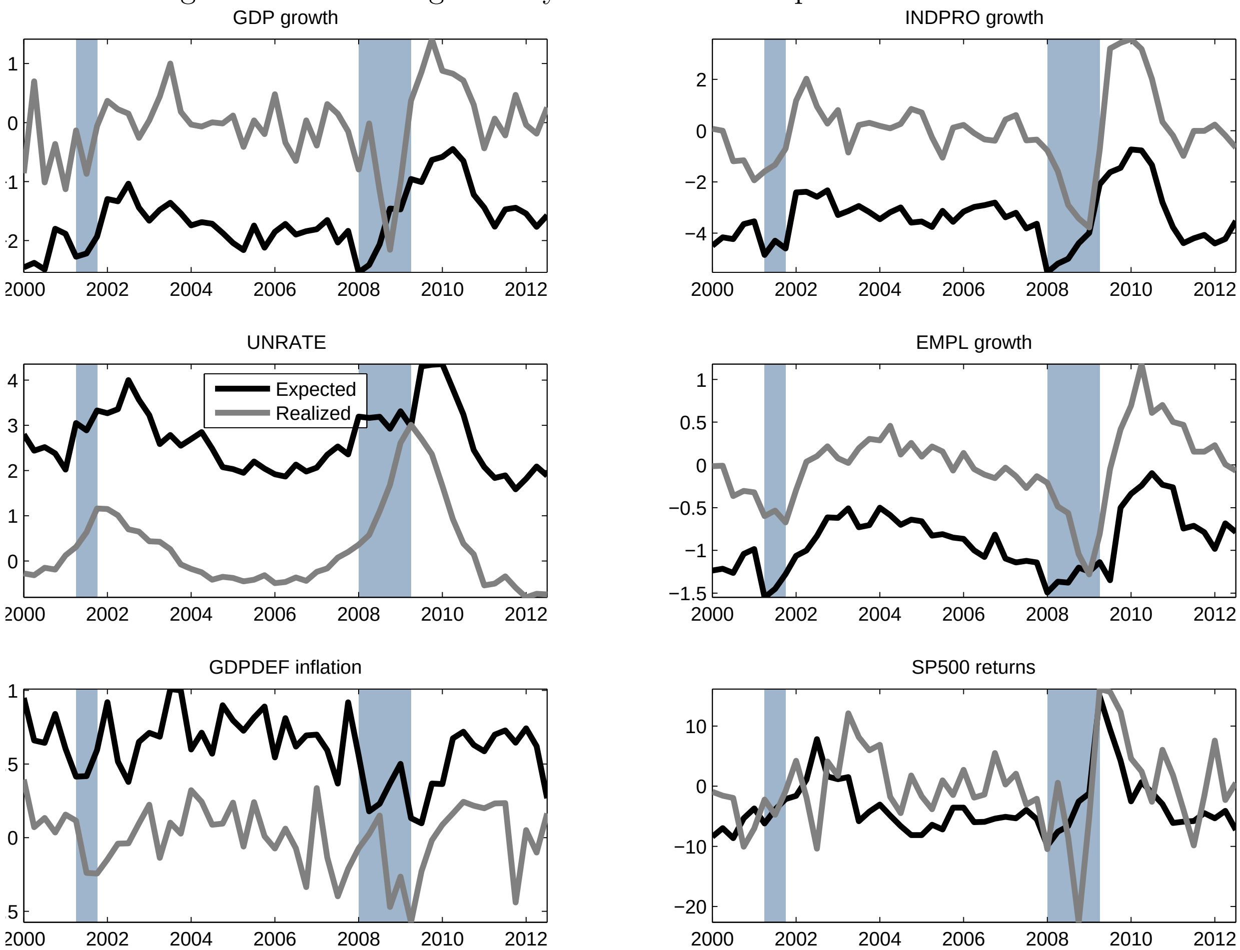
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 15: Predicting US economic activity: 3-quarter ahead scenarios



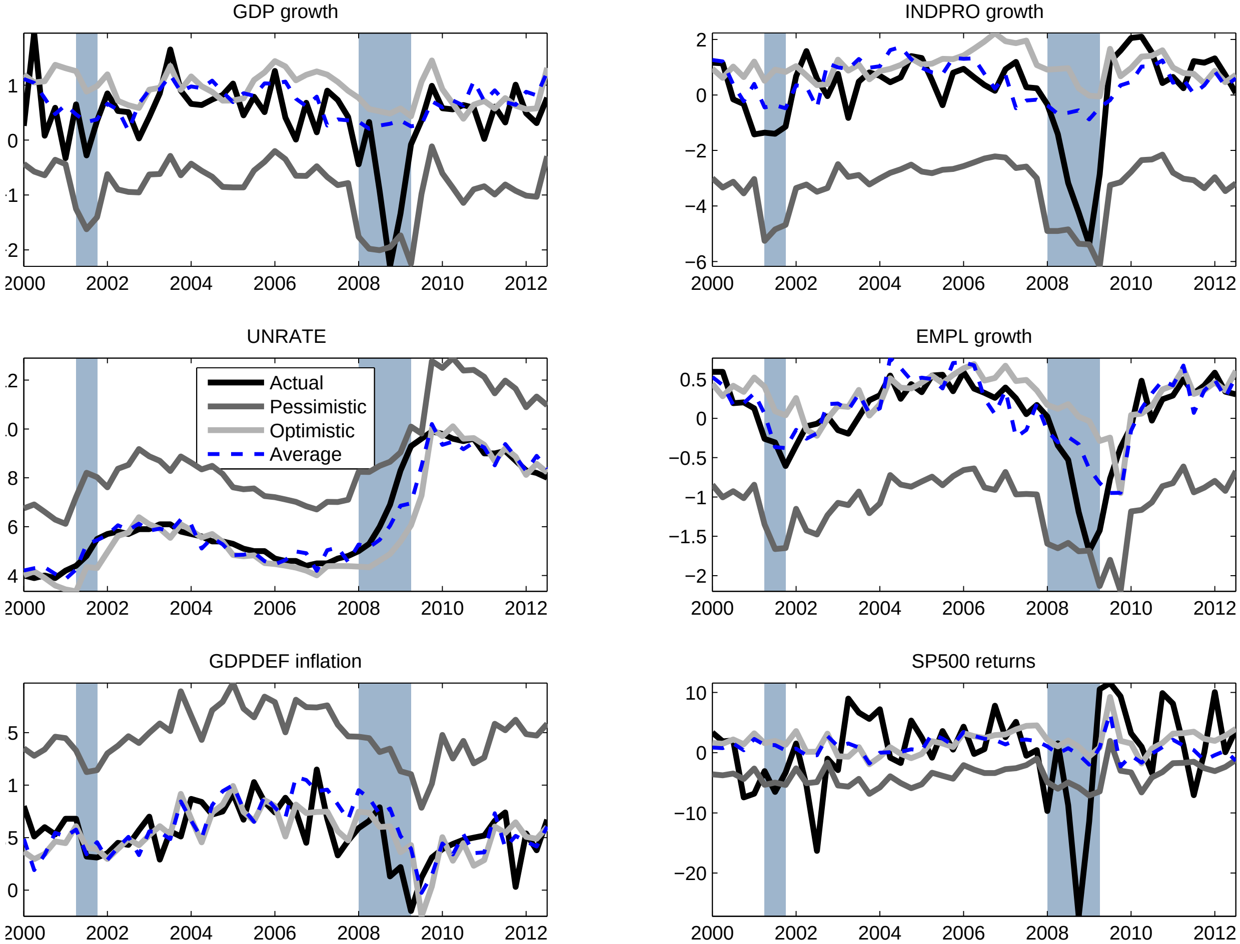
Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 16: Predicting severity of recessions: 3-quarter ahead scenarios



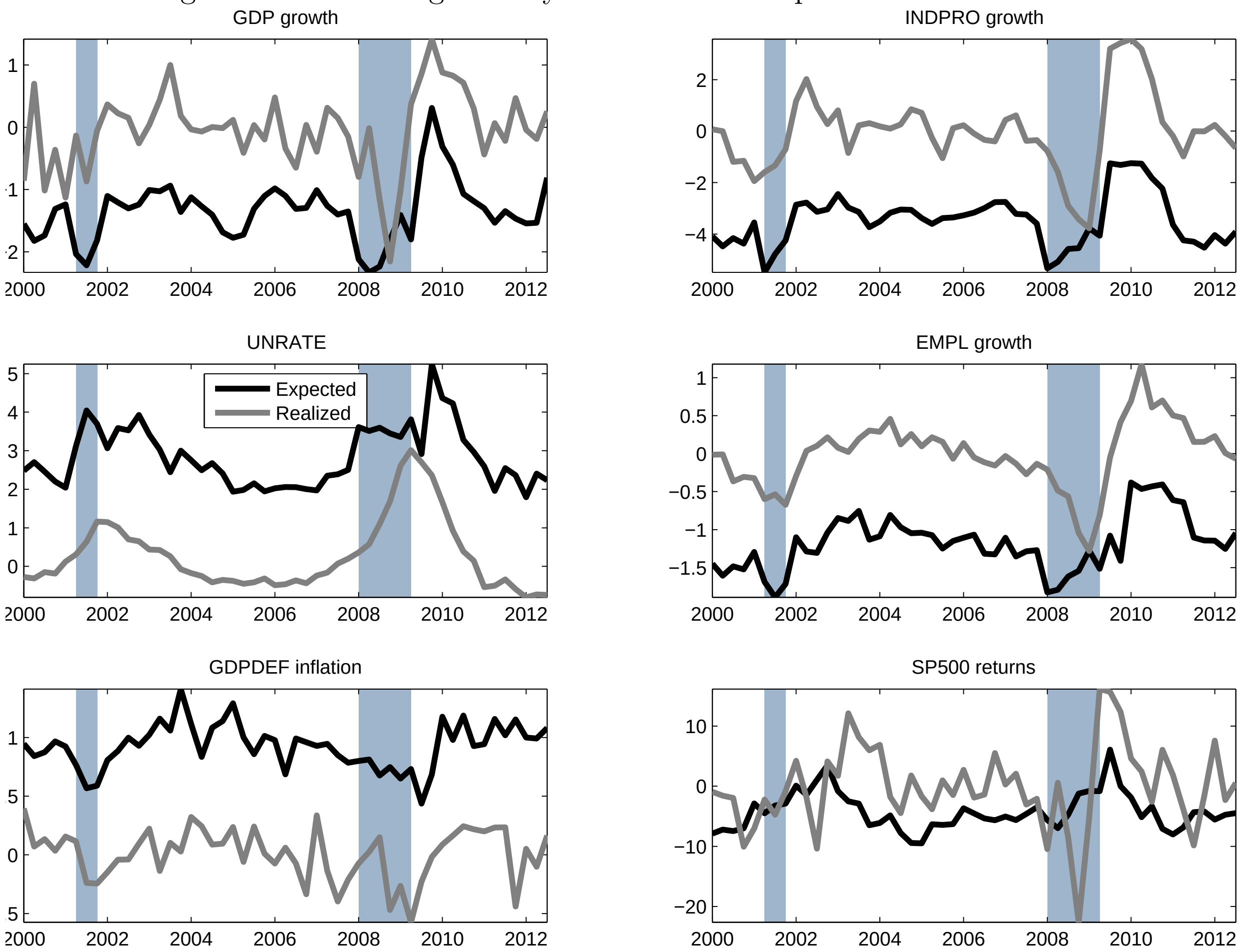
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 17: Predicting US economic activity: 4-quarter ahead scenarios



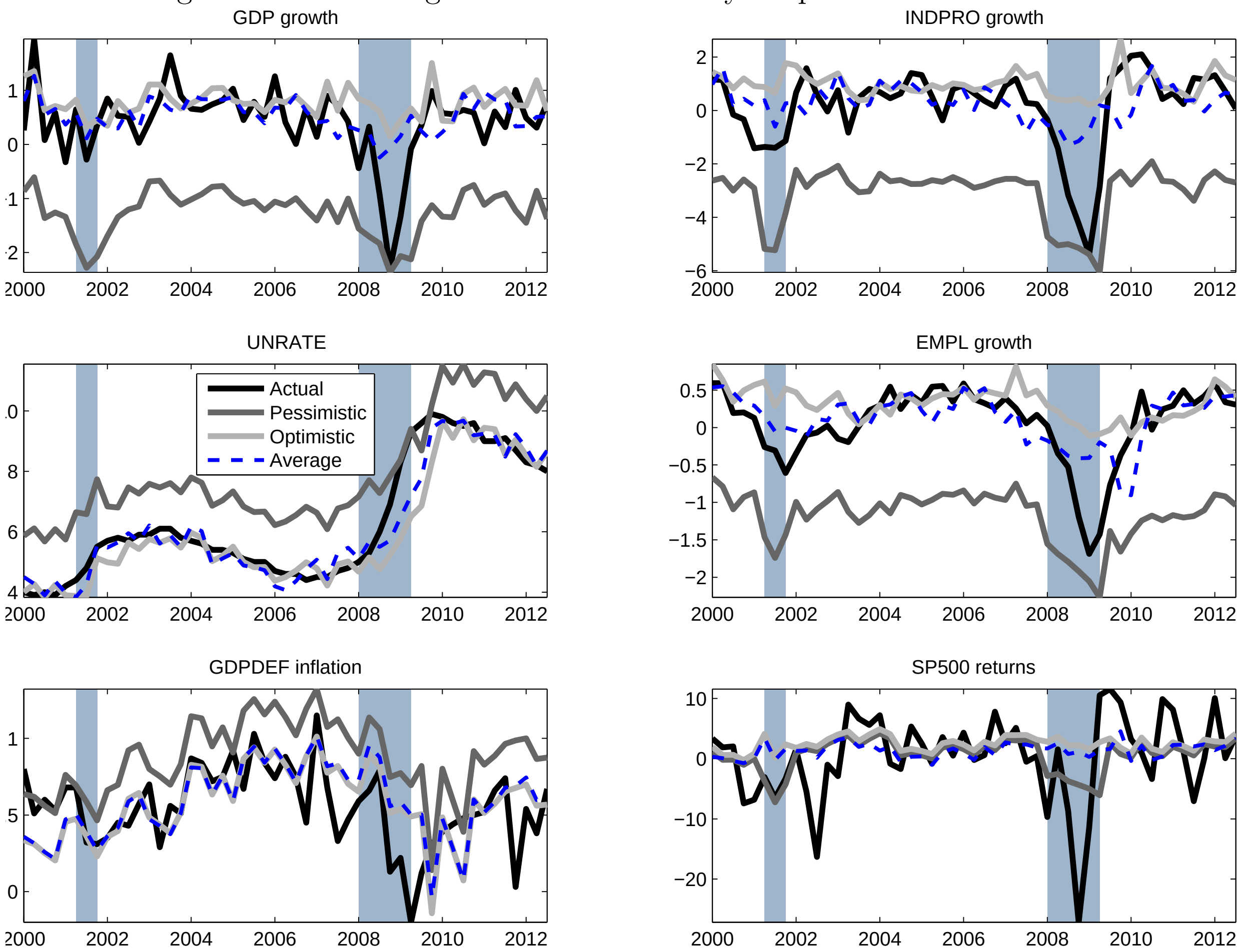
Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 18: Predicting severity of recessions: 4-quarter ahead scenarios



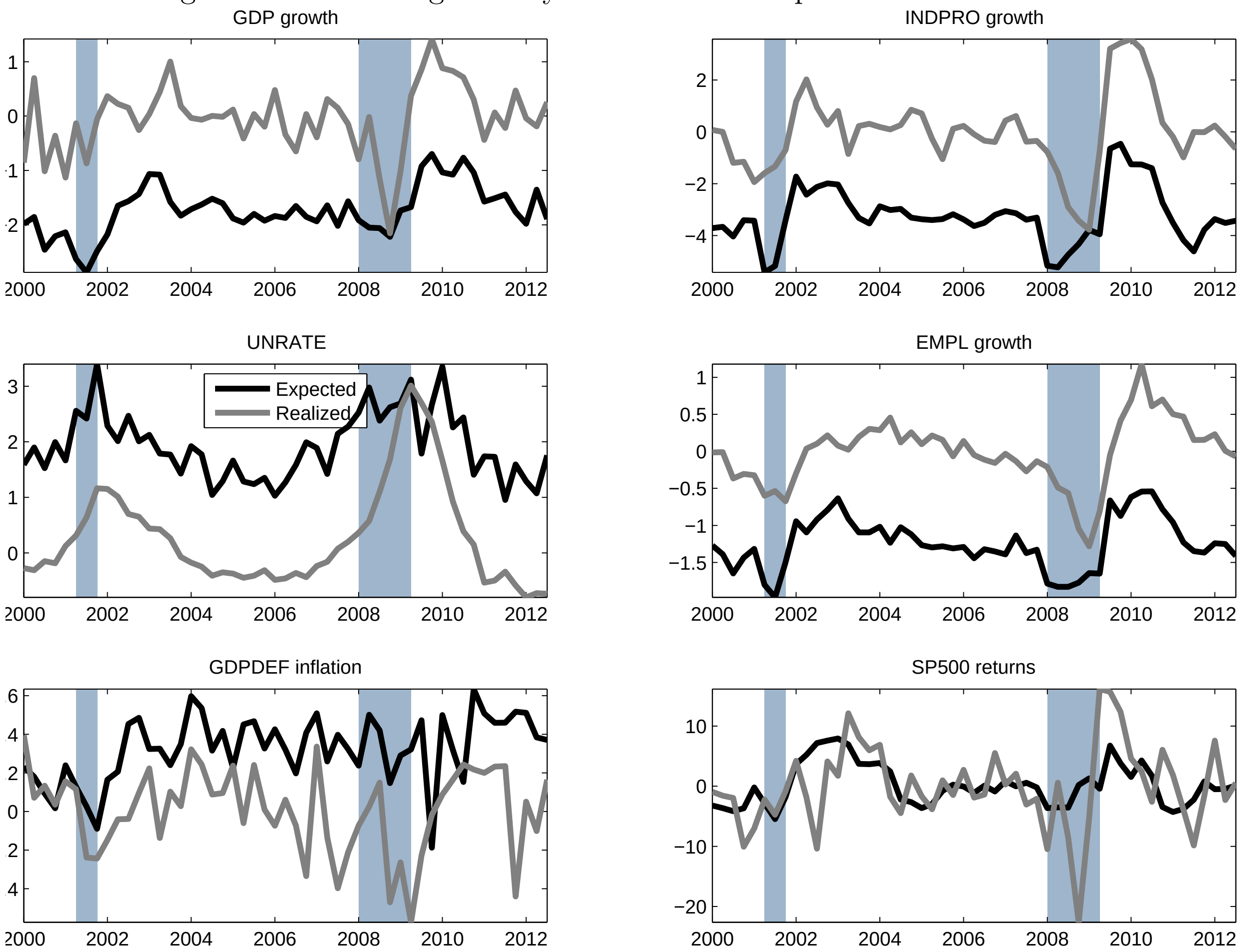
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 19: Predicting US economic activity: 5-quarter ahead scenarios



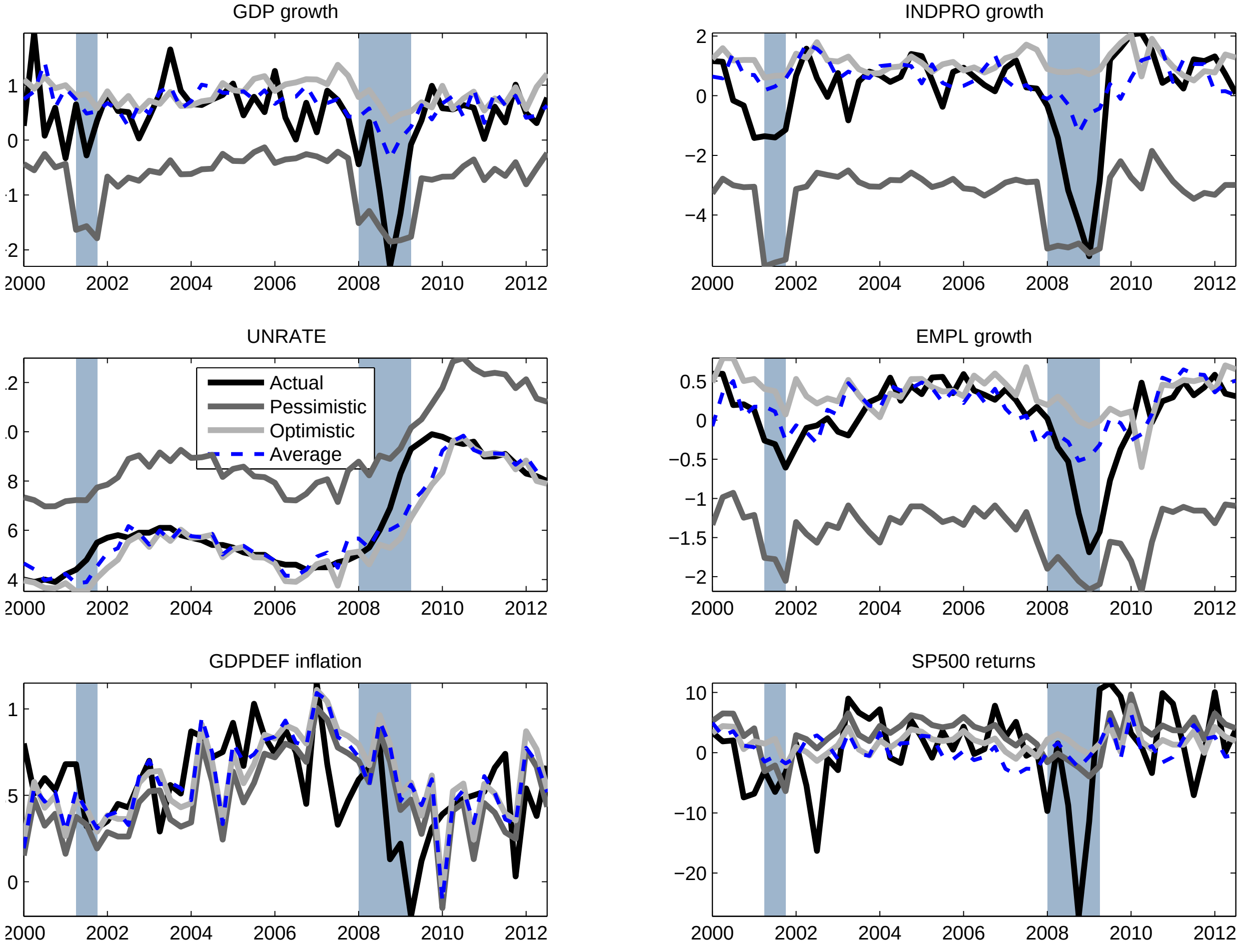
Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 20: Predicting severity of recessions: 5-quarter ahead scenarios



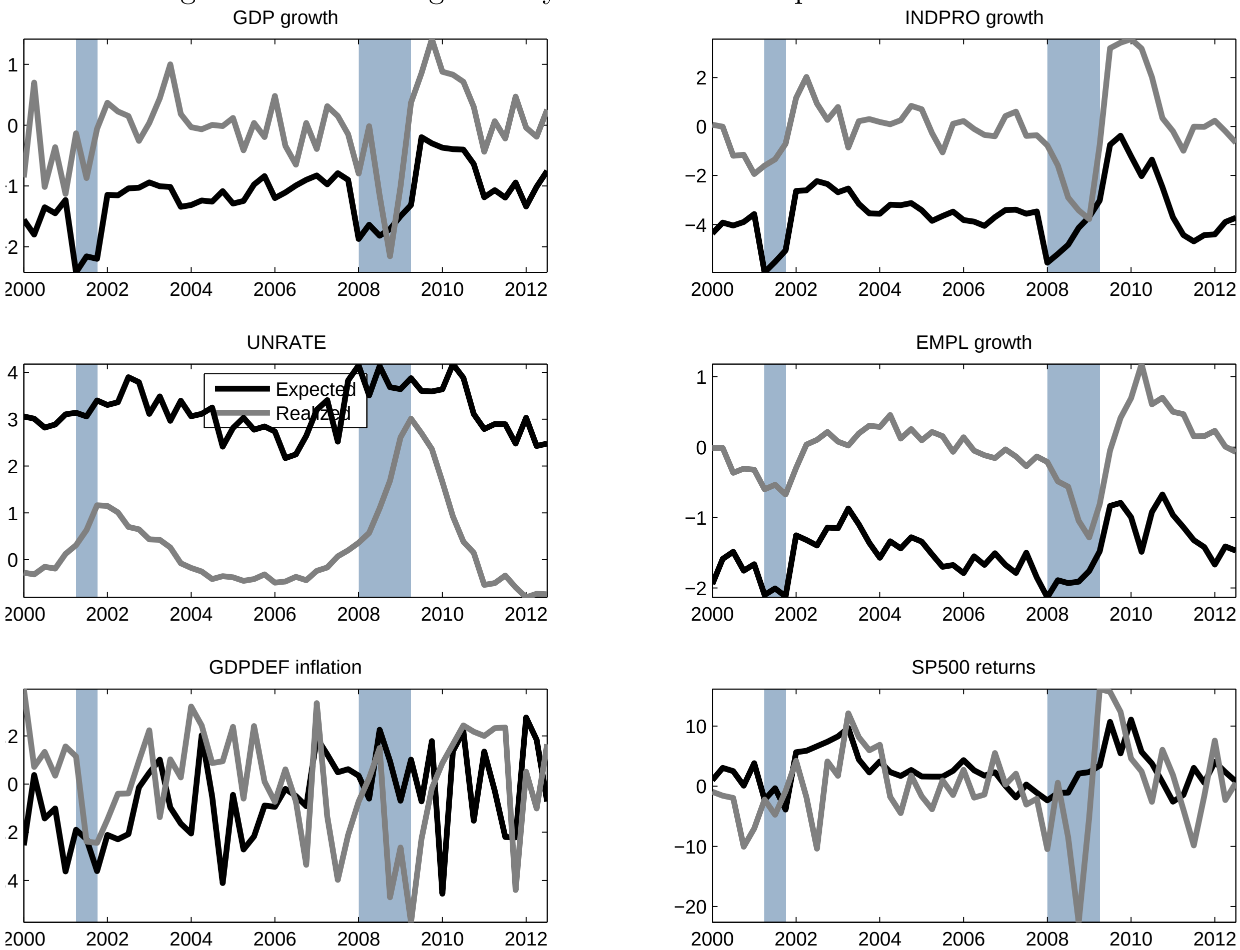
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 21: Predicting US economic activity: 6-quarter ahead scenarios



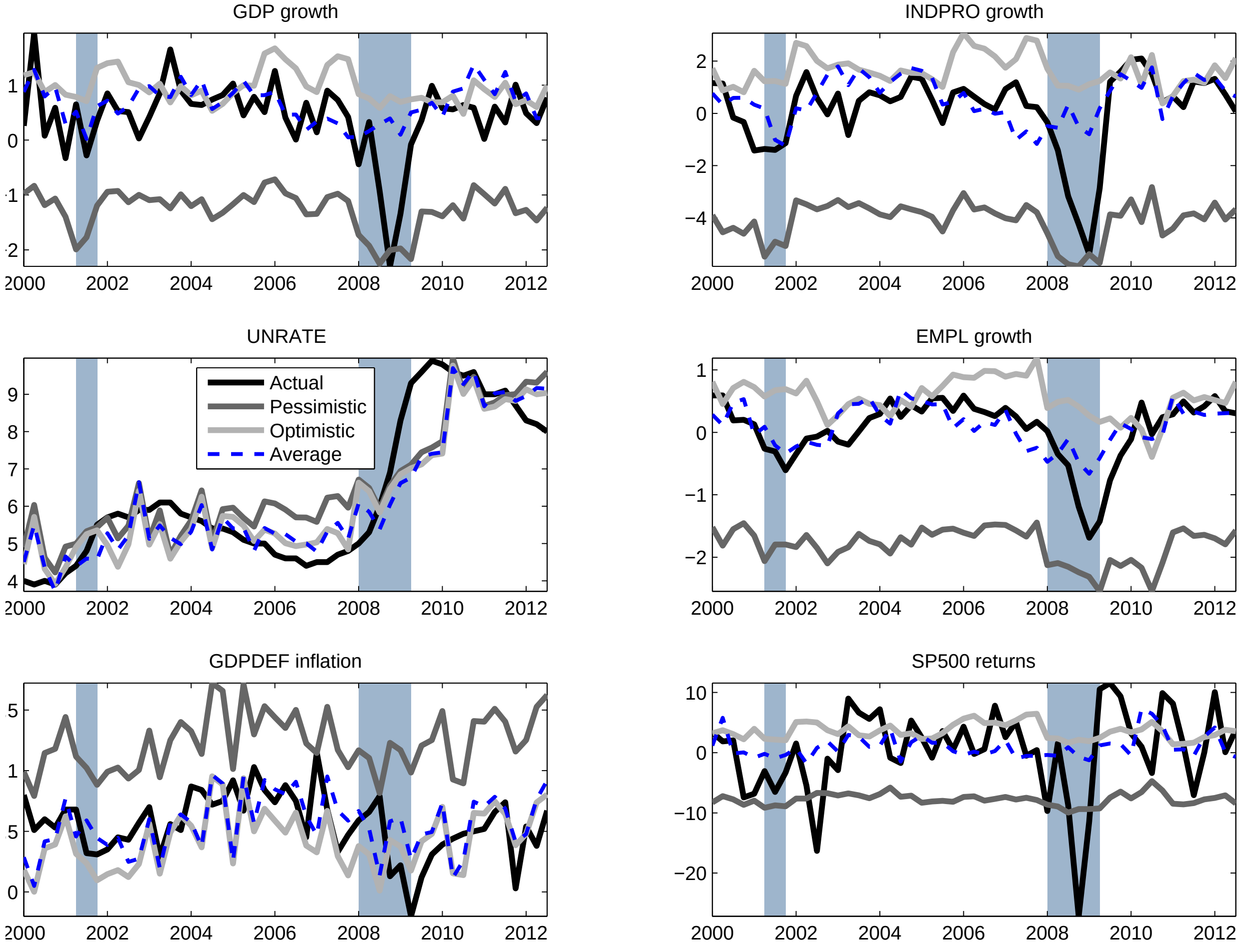
Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 22: Predicting severity of recessions: 6-quarter ahead scenarios



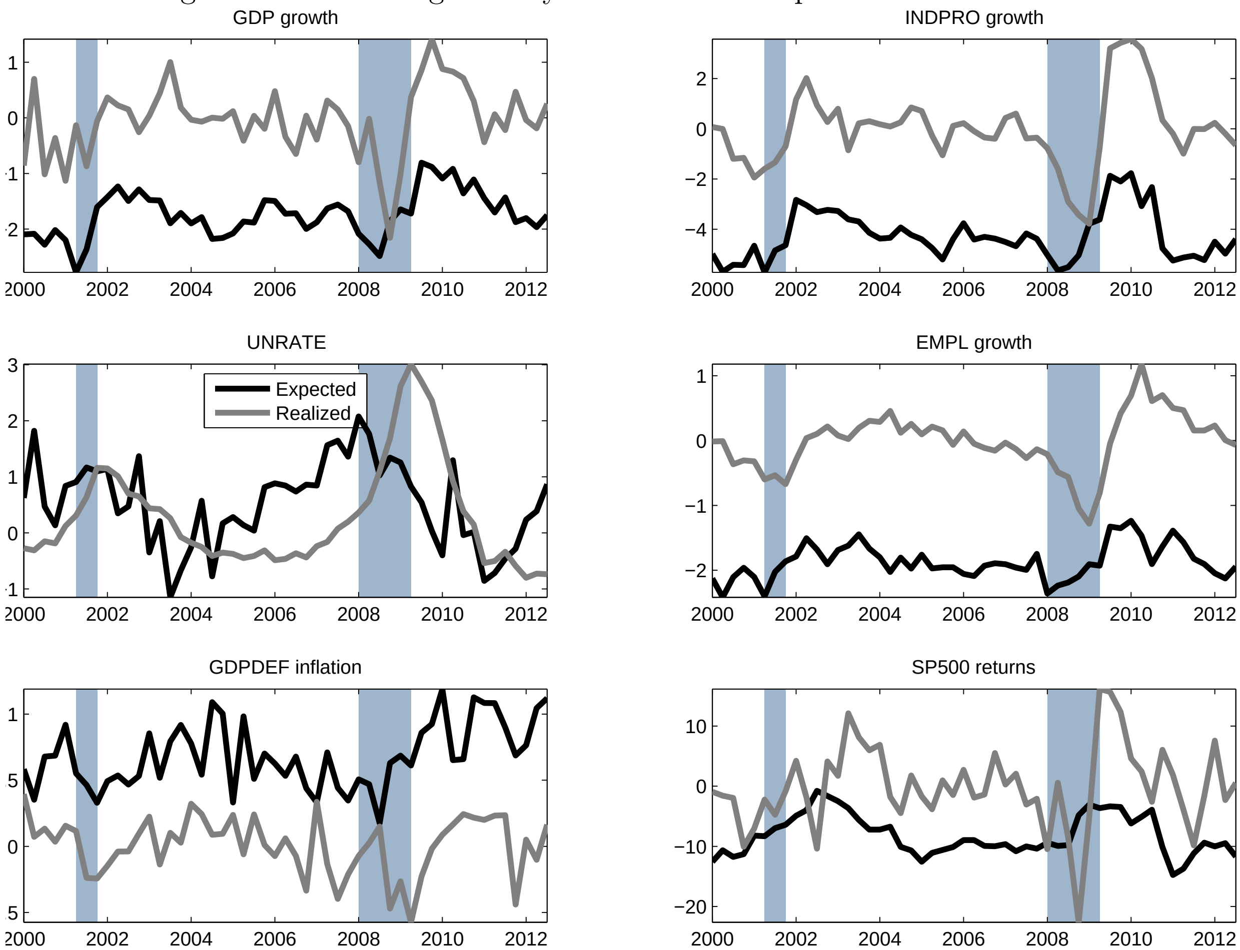
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 23: Predicting US economic activity: 7-quarter ahead scenarios



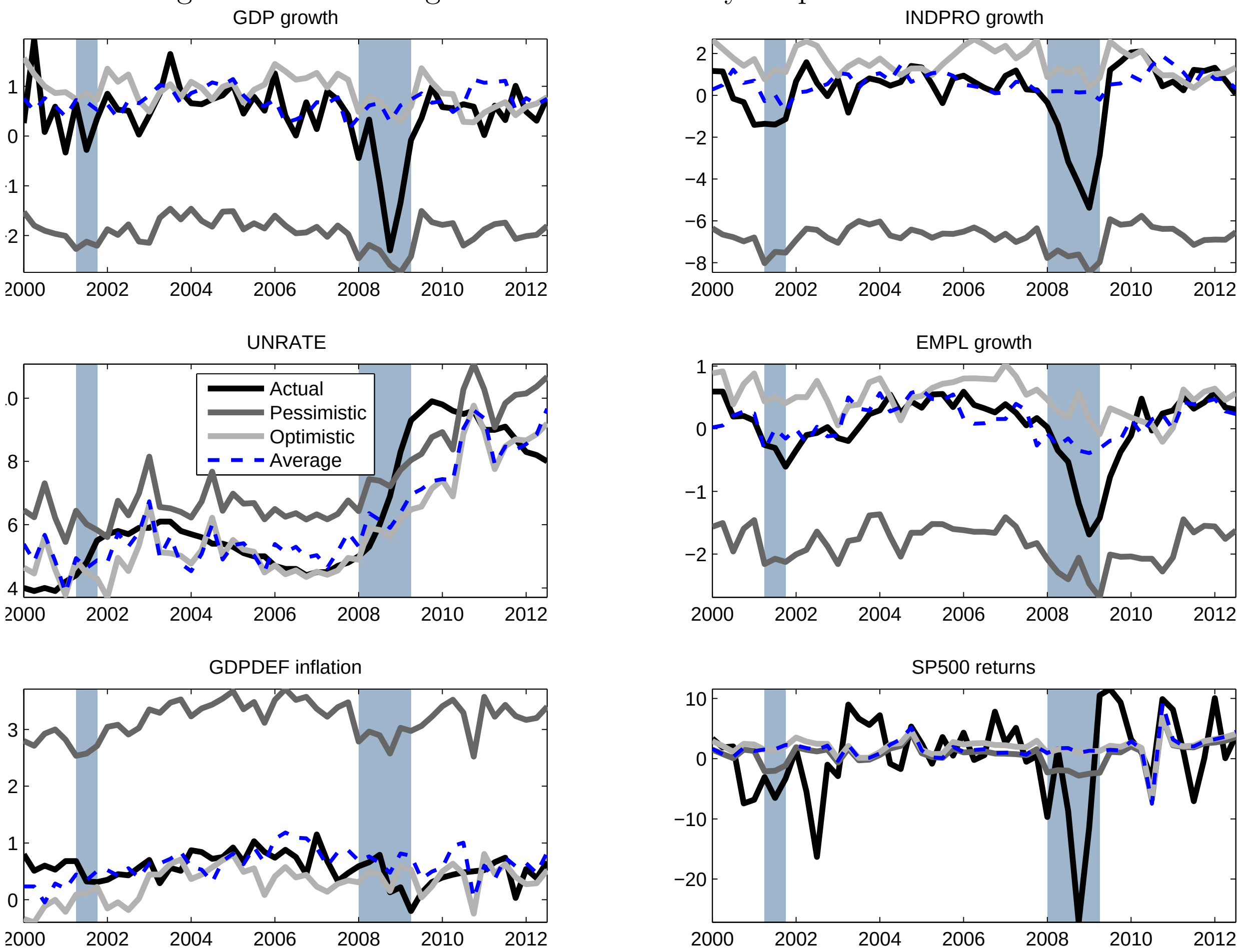
Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 24: Predicting severity of recessions: 7-quarter ahead scenarios



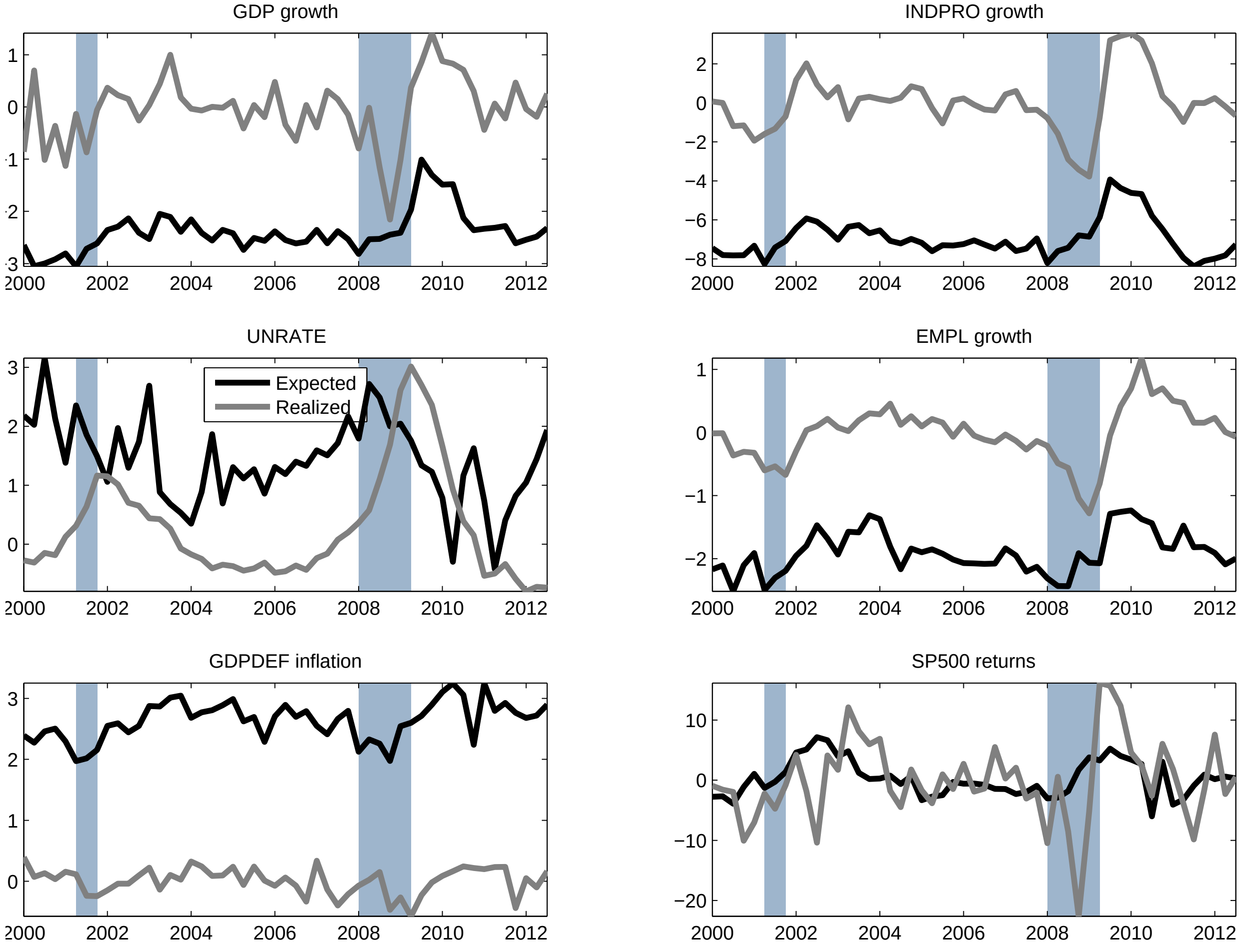
Predicted in-sample expected and realized severities from equations 16 and 17 respectively.

Figure 25: Predicting US economic activity: 8-quarter ahead scenarios



Predicted in-sample average, pessimistic and optimistic scenarios from IMR-DI-AR forecasting models 8, 10 and 11 respectively.

Figure 26: Predicting severity of recessions: 8-quarter ahead scenarios



Predicted in-sample expected and realized severities from equations 16 and 17 respectively.