

Armah, Nii Ayi; Swanson, Norman

Working Paper

Some variables are more worthy than others: New diffusion index evidence on the monitoring of key economic indicators

Working Paper, No. 2011-15

Provided in Cooperation with:

Department of Economics, Rutgers University

Suggested Citation: Armah, Nii Ayi; Swanson, Norman (2011) : Some variables are more worthy than others: New diffusion index evidence on the monitoring of key economic indicators, Working Paper, No. 2011-15, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/59477>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Some Variables are More Worthy Than Others: New Diffusion Index Evidence on the Monitoring of Key Economic Indicators*

Nii Ayi Armah¹ and Norman R. Swanson²

¹Bank of Canada and ²Rutgers University

April 2010

Abstract

Central banks regularly monitor select financial and macroeconomic variables in order to obtain early indication of the impact of monetary policies. This practice is discussed on the Federal Reserve Bank of New York website, for example, where one particular set of macroeconomic “indicators” is given. In this paper, we define a particular set of “indicators” that is chosen to be representative of the typical sort of variable used in practice by both policy-setters and economic forecasters. As a measure of the “adequacy” of the “indicators”, we compare their predictive content with that of a group of observable factor proxies selected from amongst 132 macroeconomic and financial time series, using the diffusion index methodology of Stock and Watson (2002a,b) and the factor proxy methodology of Bai and Ng (2006a,b) and Armah and Swanson (2010). The variables that we predict are output growth and inflation, two representative variables from our set of indicators that are often discussed when assessing the impact of monetary policy. Interestingly, we find that the indicators are all contained within the set the observable variables that proxy our factors. Our findings, thus, support the notion that a judiciously chosen set of macroeconomic indicators can effectively provide the same macroeconomic policy-relevant information as that contained in a largescale time series dataset. Of course, the large-scale datasets are still required in order to select the key indicator variables or confirm one’s prior choice of key variables. Our findings also suggest that certain yield “spreads” are also useful indicators. The particular spreads that we find to be useful are the difference between Treasury or corporate yields and the federal funds rate. After conditioning on these variables, traditional spreads, such as the yield curve slope and the reverse yield gap are found to contain no additional marginal predictive content. We also find that the macroeconomic indicators (not including spreads) perform best when forecasting inflation in non-volatile time periods, while inclusion of our spread variables improves predictive accuracy in times of high volatility.

JEL classification: C22, C33, C51.

Keywords: diffusion index, factor, federal reserve bank, forecast, macroeconometrics, monetary policy, parameter estimation error, proxy.

* Nii Ayi Armah, Bank of Canada, 234 Wellington Street, Ottawa, ON K1A 0G9, Canada. narmah@bankofcanada.ca. Norman R. Swanson: Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA. nswanson@econ.rutgers.edu. This paper has been prepared for a memorial issue of the Applied Economics Journal celebrating the accomplishments of Sir Clive W. J. Granger, and many thanks are due to Mark Taylor for organizing the special issue. Useful comments on earlier versions of this paper were also provided by Roberto Chang, Valentina Corradi, Roger Klein, Fuchun Li, Esfandia Maasoumi, Marcelo Medeiros, Serena Ng, Greg Tkacz, as well as to participants of the conference on “Real-Time Data” at the Philadelphia Federal Reserve Bank in 2009 and seminar participants at the Bank of Canada. Finally, a great many thanks are also owed to Mark Watson for making the data used in this paper available for public consumption. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada. Swanson wishes to thank the Research Council at Rutgers University for financial support.

1 Introduction

In assessing the usefulness of macroeconomic models, two criteria are often used. The first is the model's performance against well established alternative models in out-of-sample forecasting exercises. The second is the assessment of the (perceived) relevance of the model in policy formulation and analysis. These two criteria may be exclusive, in the sense that a model might offer nothing from a policy perspective, but be extremely useful for forecasting purposes, if it offers superior out-of-sample forecast performance. For instance, a univariate autoregressive (AR) model might exhibit superior forecasting performance, but might not contain relevant regressors that can act as policy instruments for controlling a given "target" variable. On the other hand, a model formulated solely based on economic theory could by design provide a number of instruments (regressors) or policy "control" variables, but might be characterized by poor predictive accuracy. In this paper, we evaluate the argument that a judiciously chosen set of macroeconomic "indicators" that can be easily extracted from a largescale dataset may be useful from both a policy and a forecasting perspective. We further argue that such a group of "key" indicators can take the place of all other variables in the largescale dataset, both when forecasting and when carrying out analysis of the impact of monetary policies.

Central banks regularly monitor select financial and macroeconomic variables in order to obtain early indication of the impact of monetary policies. This practice is discussed on the Federal Reserve Bank of New York website, for example, where one particular set of macroeconomic "indicators" is given. Our approach is to define a particular set of "indicators" chosen to be representative of the type of variables used in practice. As a measure of the "adequacy" of our "indicators", we compare their predictive content with that of a group of observable factor proxies selected from amongst 132 macroeconomic and financial time series, using the diffusion index methodology of Stock and Watson (2002a,b) and the factor proxy methodology of Bai and Ng (2006a,b) and Armah and Swanson (2010). The variables that we predict are output growth and inflation, two representative variables from our set of indicators. One of our primary objectives is to ascertain whether a carefully chosen set of indicators can effectively provide the same macroeconomic policy-relevant information as that contained in a largescale time series dataset. Ancillary questions that we attempt to provide answers for include: Does the relevant set of indicators remain constant across different phases of the business cycle? Are there additional indicators that may hitherto

have failed to be recognized in the prediction literature, and that can be obtained via analysis of large scale datasets commonly used in diffusion index construction?

Some of the financial and macroeconomic indicators that central banks often rely **on** for monitoring monetary policy, and that we examine in this paper include: real gross domestic product (GDP), the consumer price index (CPI), nonfarm payroll employment, housing starts, industrial production/capacity utilization, retail sales, business sales and inventories, advanced durable goods shipments, new orders and unfilled orders, lightweight vehicle sales, the yield on 10-year Treasury bonds, the S&P 500 stock price index, and the money supply (M2). Collectively, these are referred to as our macroeconomic “indicators”. In order to examine these indicators, we carry out an extensive set of real-time prediction experiments in which benchmark linear models that directly include the above indicators are compared against a variety of other prediction models. One of the alternative groups of models that we examine involves constructing predictions using common factors of the Stock and Watson (1998,1999,2002a,b) variety. The reason why we use the diffusion index methodology of Stock and Watson as a starting point in our factor analysis is that their methodology has been shown to quite robustly estimate the relevant common factors that underlie the co-movements of a given set of macroeconomic variables. A second alternative group of models that we examine replaces common factors with observable factor proxies. The idea behind this latter group of models is that a properly chosen set of factor proxies may in some cases contain the same marginal predictive content as the factors that they replace. The theory for this approach is discussed in Bai and Ng (2006a,b) and further developed in Armah and Swanson (2010).

Interestingly, we find that our macroeconomic indicators are all contained within the set of observable variables that proxy our factors. Our findings, thus, support the notion that a judiciously chosen set of macroeconomic indicators can effectively provide the same macroeconomic policy-relevant *and* forecast relevant information as that contained in a largescale time series dataset, inasmuch as the indicators can be used to construct predictions as accurate or more accurate than those constructed using the entire dataset. In addition, our analysis suggests that certain yield “spreads” are also useful indicators. The particular spreads that we find to be useful are the difference between Treasury or corporate yields and the federal funds rate. When predicting inflation, we find that models with these spreads outperform models without these spreads at the 1-month ahead prediction horizon around one half the time, and at the 3- and 12-month ahead horizons in virtually all cases, based on point mean square forecast error (MSFE) comparison. Moreover, at

the 12-month horizon, most of the prediction models including these spreads are also significantly better, based on application of Diebold and Mariano (1995) tests (see also Clark and McCracken (2005) and McCracken (2007)). For output growth, models with these spreads outperform those without spreads around 90% of the time, at the 1-month ahead horizon; but add little when considering longer horizons, based on point MSFE comparison. In summary, these results suggest that certain “non-traditional” spreads may be useful for monitoring economic activity. Moreover, after conditioning on the “non-traditional” variables, traditional spreads, such as the yield curve slope and the reverse yield gap are found to contain no additional marginal predictive content for output growth and inflation. We also find that the macroeconomic indicators (not including our spread variables) perform best when forecasting inflation in non-volatile time periods, while inclusion of our spread variables improves predictive accuracy in time of high volatility. This suggests that there are relevant business-cycle induced informational asymmetries that should not be overlooked when using our spread indicators to predict macroeconomic variables that are in turn used to evaluate monetary policy.

The rest of the paper is organized as follows. In Section 2, we summarize some useful features of the diffusion index literature and briefly discuss the methodology associated with forming predictions using factor proxies and the macroeconomic indicators. Section 3 gives an outline of the empirical methodology used in the sequel, and Section 4 contains a brief overview of the data employed in our prediction experiments. Empirical findings are summarized in Section 5 and Section 6 offers concluding remarks.

2 Diffusion Index Methodology

2.1 Prediction Using Diffusion Index Models

Following Stock and Watson (2002a,b), let y_{t+1} be the series we wish to forecast and X_t be an N -dimensional vector of predictor variables, for $t = 1, \dots, T$. Assume that (y_{t+1}, X_t) has a dynamic factor model representation with \bar{r} common dynamic factors, f_t . Hence, f_t is an $\bar{r} \times 1$ vector. The dynamic factor model is written as:

$$y_{t+h} = \alpha(L)f_t + \sum_{j=1}^p \beta_j y_{T-j+1} + \varepsilon_{t+h} \quad (1)$$

and

$$x_{it} = \lambda_i(L)f_t + e_{it}, \quad (2)$$

for $i = 1, 2, \dots, N$, where $h > 0$ is the forecast horizon; x_{it} is a single datum for a particular predictor variable; e_{it} is the idiosyncratic shock component of x_{it} ; and $\alpha(L)$ and $\lambda_i(L)$ are lag polynomials in nonnegative powers of L . The N dimensional spectral density of x_{it} has rank \bar{r} . This implies that data generated under (2) would have \bar{r} dynamic factors (see Boivin and Ng (2005)). In general, dynamic factor models can be transformed into static factor models. In Stock and Watson (2002a), the lag polynomials $\alpha(L)$ and $\lambda_i(L)$ are modeled as $\alpha(L) = \sum_{j=0}^q \alpha_j L^j$ and $\lambda_i(L) = \sum_{j=0}^q \lambda_{ij} L^j$. The finite order of the lag polynomials allows us to rewrite (1) and (2) as:

$$y_{t+h} = \alpha' F_t + \sum_{j=1}^p \beta_j y_{T-j+1} + \varepsilon_{t+h} \quad (3)$$

and

$$x_{it} = \Lambda_i' F_t + e_{it}, \quad (4)$$

where $F_t = (f_t', \dots, f_{t-q}')'$ is an $r \times 1$ vector, with $r = (q+1)\bar{r}$ and α is an $r \times 1$ vector. As discussed in the next section, the model depicted as (3) is the factor model that we use to predict output growth and inflation. Here, r is the number of static factors (i.e. the number of elements in F_t). Additionally, $\Lambda_i = (\lambda_{i0}', \dots, \lambda_{iq}')'$ is a vector of factor loadings on the r static factors, where λ_{ij} is an $\bar{r} \times 1$ vector for $j = 0, \dots, q$. The N dimensional population covariance matrix of x_{it} generated under (4) has r nonzero eigenvalues that diverge with N . Thus the model is said to have r static factors (see Boivin and Ng (2005)). Some technical assumptions we make are $\frac{1}{N} \sum_{i=1}^N \lambda_i \lambda_i' \xrightarrow{p} \Sigma_\Lambda$ as $N \rightarrow \infty$, and $\frac{1}{T} \sum_{t=1}^T F_t F_t' \xrightarrow{p} \Sigma_F$ as $T \rightarrow \infty$, where Σ_Λ and Σ_F are $r \times r$ positive definite matrices.

Ding and Hwang (1999), Forni and Reichlin (1996,1998), Forni et al. (2000, 2005), Stock and Watson (2002b), Bai and Ng (2002) and Bai (2003) showed that the space spanned by both the static and dynamic factors can be consistently estimated when N and T are both large. For forecasting purposes, little is gained from a clear distinction between the static and the dynamic factors. Boivin and Ng (2005), and Rapach and Strauss (2007) compare alternative factor based forecast methodologies, and conclude that when the dynamic structure is unknown and the model is characterized by complex dynamics, the approach of Stock and Watson performs favorably.

2.2 Estimating the Number of Factors

Following Bai and Ng (2002), let \underline{X}_i be a $T \times 1$ vector of observations for the i th variable. For a given cross-section i , we have:

$$\underset{(T \times 1)}{\underline{X}_i} = \underset{(T \times r)}{F^0} \underset{(r \times 1)}{\Lambda_i} + \underset{(T \times 1)}{\underline{e}_i}$$

where $\underline{X}_i = (X_{i1}, \dots, X_{iT})'$, $F^0 = (F_1, \dots, F_T)'$ and $\underline{e}_i = (e_{i1}, \dots, e_{iT})'$. The whole panel of data $X = (\underline{X}_1, \dots, \underline{X}_N)$ can consequently be represented as:

$$\underset{(T \times N)}{X} = \underset{(T \times r)}{F^0} \underset{(r \times N)}{\Lambda'} + \underset{(T \times N)}{e},$$

where $\Lambda = (\Lambda_1, \dots, \Lambda_N)'$ and $e = (\underline{e}_1, \dots, \underline{e}_N)$. X can be viewed as a representative set of variables that characterize the whole economy. The set of macroeconomic indicators is a subset of X .

We work with high-dimensional factor models that allow both N and T to tend to infinity, and in which e_{it} may be serially and cross-sectionally correlated so that the covariance matrix of $e_t = (e_{1t}, \dots, e_{Nt})$ does not have to be a diagonal matrix. Furthermore, it is well known that for $\Lambda F_t = \Lambda Q Q^{-1} F_t$, a normalization is needed in order to uniquely define the factors, where Q is a nonsingular matrix. Now, assuming that $(\Lambda' \Lambda / N) \rightarrow I_r$, we restrict Q to be orthonormal, for example. This assumption, together with others noted in Stock and Watson (2002b), enables us to identify the factors up to a change of sign and consistently estimate them up to an orthonormal transformation. Forecasts of y_{T+h} based on (3) and (4) involve a two step procedure because both the regressors and coefficients in the forecasting equations are unknown. The data sample $\{X_t\}_{t=1}^T$ are first used to estimate the factors, $\{\tilde{F}_t\}_{t=1}^T$ by means of principal components. With the estimated factors in hand, we obtain the estimators $\hat{\alpha}$ and $\hat{\beta}$ by regressing y_{t+1} onto \tilde{F}_t and the observable variables in W_t . Of note is that if $\sqrt{T}/N \rightarrow 0$, then the generated regressor problem does not arise, in the sense that least squares estimates of α and β are \sqrt{T} consistent and asymptotically normal (see Bai and Ng (2006a)).

Since the common factors are not observed, in the regression analysis of (4), we replace F_t by \tilde{F}_t , estimates that span the same space as F_t when $N, T \rightarrow \infty$. Estimation of these common factors from large panel data sets of macroeconomic variables can be carried out using principal component analysis. We refer the reader to Stock and Watson (1998, 2002a, 2002b, 2004, 2006) and Bai and Ng (2002, 2007) for a detailed explanation of this procedure.

From (4), estimates of Λ_i^k and F_t^k are obtained by solving the optimization problem:

$$V(k) = \min_{\Lambda^k, F^k} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \Lambda_i^{k'} F_t^k)^2 \quad (5)$$

Let \tilde{F}^k and $\tilde{\Lambda}^k$ be the minimizers of equation (5). Since Λ^k and F^k are not separately identifiable, if $N > T$, a computationally expedient approach would be to concentrate out $\tilde{\Lambda}^k$ and minimize (5) subject to the normalization $F^{k'} F^k / T = I_k$. Minimizing (5) is equivalent to maximizing $\text{tr}[F^{k'} (X X') F^k]$. This optimization is solved by setting \tilde{F}^k to be the matrix of the k eigenvectors of $X X'$ that correspond to the k largest eigenvalues of $X X'$. Note that $\text{tr}[\cdot]$ represents the matrix trace. The superscript in Λ^k and F^k signifies the use of k factors in the estimation and the fact that the estimates will depend on k . Let \tilde{D} be a $k \times k$ diagonal matrix consisting of the k largest eigenvalues of $X X'$. The estimated factor matrix, denoted by \tilde{F}^k , is $\sqrt{\tilde{D}}$ times the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix $X X'$. Given \tilde{F}^k and the normalization $F^{k'} F^k / T = I_k$, $\tilde{\Lambda}^{k'} = (\tilde{F}^{k'} \tilde{F}^k)^{-1} \tilde{F}^{k'} X = \tilde{F}^{k'} X / T$ is the corresponding factor loadings matrix.

The solution to the optimization problem in (5) is not unique. If $N < T$, it becomes computationally advantageous to concentrate out \tilde{F}^k and minimize (5) subject to $\tilde{\Lambda}^{k'} \tilde{\Lambda}^k / N = I_k$. This minimization is the same as maximizing $\text{tr}[\tilde{\Lambda}^{k'} X' X \tilde{\Lambda}]$, the solution of which is to set $\tilde{\Lambda}^k$ equal to the eigenvectors of the $N \times N$ matrix $X' X$ that correspond to its k largest eigenvalues. One can consequently estimate the factors as $\tilde{F}^k = X' \tilde{\Lambda}^k / N$. \tilde{F}^k and \tilde{F}^k span the same column spaces, hence for forecasting purposes, they can be used interchangeably depending on which one is more computationally efficient. We employ the methodology of Bai and Ng (2002) to consistently estimate the true number of a factors, r . Given \tilde{F}^k and $\tilde{\Lambda}^k$, let $\hat{V}(k) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \tilde{\Lambda}_i^{k'} \tilde{F}_t^k)^2$ be the sum of squared residuals from regressions of X_i on the k factors, $\forall i$ and $IC(k) = \log(\hat{V}(k)) + k(\frac{N+T}{NT}) \log C_{NT}^2$ be the Bai and Ng (2002) information criterion where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. The consistent estimate of the true number of factors is then $\hat{k} = \arg \min_{0 \leq k \leq k_{\max}} IC(k)$.

Stock and Watson (2002b) show that the difference between feasible (estimated model) and unfeasible (true model) factor based forecasts converge in probability to zero as $N, T \rightarrow \infty$.

2.3 Observable Factor Proxies

As the Stock and Watson principal components factors are not observable, the methodology of Bai and Ng (2006b) is used to select observable macroeconomic and financial variables that closely proxy the constructed diffusion indices. Armah and Swanson (2010) demonstrate that the factor proxies perform just as well as and at times significantly better than the Stock and Watson estimated factors in a number of forecasting exercises. We compare the macroeconomic indicators with the observable factor proxies in order to determine whether the indicators contain the same prediction-relevant information as that contained in the factors constructed using the above methodology.

Recall that $y_{t+h} = \alpha' F_t + \sum_{j=1}^p \beta_j y_{T-j+1} + \varepsilon_{t+h}$. As mentioned above, and shown in Stock and Watson (2002b) and Bai and Ng (2006b), under a set of moment conditions on (ε, e, F^0) and an asymptotic rank condition on Λ , if the space spanned by F_t can be consistently estimated, then \sqrt{T} consistent estimates of α and β are obtainable. Under a similar set of conditions, it is also possible to obtain $\min[\sqrt{N}, \sqrt{T}]$ consistent forecasts if $\sqrt{T/N} \rightarrow 0$ as $N, T \rightarrow \infty$.

Suppose that we observe G' , a $(T \times m)$ matrix of observable economic variables that could potentially proxy the latent factors. At any given time t , any of the m elements of G_t ($m \times 1$) will be a good proxy if it is a linear combination of the $r \times 1$ latent factors, F_t . Let G_{jt} be an element of the m vector G_t . The null hypothesis is that G_{jt} is an exact proxy, or more precisely, $\exists \theta_j$ ($r \times 1$) such that $G_{jt} = \theta_j' F_t$. In order to implement all of the methods, consider the regression $G_{jt} = \gamma_j' \tilde{F}_t + \rho$. Let $\hat{\gamma}_j$ be the least squares estimate of γ_j and let $\hat{G}_{jt} = \hat{\gamma}_j' \tilde{F}_t$. The test is carried out by constructing the following t-statistic:

$$\tau_t(j) = \frac{(\hat{G}_{jt} - G_{jt})}{(\widehat{\text{var}}(\hat{G}_{jt}))^{1/2}} \quad (6)$$

where

$$\begin{aligned} \widehat{\text{var}}(\hat{G}_{jt}) &= \frac{1}{N} \hat{\gamma}_j' \tilde{D}^{-1} \left(\frac{\tilde{F}' \tilde{F}}{T} \right) \tilde{\Gamma}_t \left(\frac{\tilde{F}' \tilde{F}}{T} \right) \tilde{D}^{-1} \hat{\gamma}_j \\ &= \frac{1}{N} \hat{\gamma}_j' \tilde{D}^{-1} \tilde{\Gamma}_t \tilde{D}^{-1} \hat{\gamma}_j, \end{aligned} \quad (7)$$

and $\tilde{\Gamma}_t$ is defined below. The last step above is due to the normalization that $\tilde{F}' \tilde{F} / T = I_{\hat{k}}$. Once again, \tilde{D} is a $k \times k$ diagonal matrix consisting of the k largest eigenvalues of XX' . Given the null hypothesis that $G_{jt} = \theta_j' F_t$ and that \hat{G}_{jt} converges to G_{jt} at rate \sqrt{N} , Bai and Ng (2006b) show that the limiting distribution of $\sqrt{N}(\hat{G}_{jt} - G_{jt})$ is asymptotically normal and hence $\tau_t(j)$ has a

standard normal limiting distribution. The $\hat{k} \times \hat{k}$ matrix $\tilde{\Gamma}_t$ is consistently estimated as

$$\tilde{\Gamma}_t = \frac{1}{N} \sum_{i=1}^N \tilde{e}_{it}^2 \tilde{\Lambda}_i \tilde{\Lambda}_i', \quad (8)$$

and where $\tilde{e}_{it} = x_{it} - \tilde{\Lambda}_i' \tilde{F}_t$. Equation (8) allows for time-series heteroskedasticity, but assumes no cross-sectional correlation of e_{it} . For small cross-sectional correlation in e_{it} , Bai and Ng (2006a) found that constraining the correlations to be zero could sometimes be desirable. In this regard, they make the point that (8) is useful even if residual cross-correlation is genuinely present.

As mentioned earlier, $\tau_t(j)$ in (6) has a standard normal limiting distribution. Let Φ_ξ^τ be the ξ percentage point of the limiting distribution of $\tau_t(j)$. The hypothesis test based on the t-statistic in (6) enables us to determine whether an observed value of a candidate variable is a good proxy at a specific time t . For our purposes however, given information up to time T , whatever methods or procedures we use to select the proxies ought to select whole time series G_j , for which G_{jt} satisfies the null hypothesis, $\forall t$. In this regard, the proxy selection method is based upon the following statistic:

$$A(j) = \frac{1}{T} \sum_{t=1}^T 1(|\tau_t(j)| > \Phi_\xi^\tau). \quad (9)$$

The $A(j)$ statistic is the actual size of the test (i.e. the probability of Type I error given the sample size). Since $\tau_t(j)$ is asymptotically standard normal and the test is a two-tailed test, the actual size, $A(j)$, of the t -test should converge to the nominal size (the desired significance level is 2ξ) as $T \rightarrow \infty$. This means that if a candidate variable is a good proxy of the underlying factors of a data set, the $A(j)$ statistic calculated from its sample time series should approach 2ξ as the sample size increases. This is the basis on which we use the $A(j)$ statistic to select proxies. It should be noted that the $A(j)$ statistic does not constitute a test in the strict sense since we do not compare a test statistic to a critical value to determine whether or not to reject a null hypothesis. Rather, this procedure gives a ranking of the proxies with the best proxy having an $A(j)$ statistic value closest to 2ξ . In our implementation, the candidate set of proxies, G' , is the same as the the panel data set X from which we estimate the factors.

The $A(j)$ statistic discussed above may yield a different set of proxies at each recursive forecast iteration. This is because the $A(j)$ statistic is composed of some estimated values. In view of this, Armah and Swanson (2010) develop a version of the $A(j)$ statistic where the sample period in an empirical analysis is broken into three subsamples (R_1, R_2 , and E , such that $T = R_1 + R_2 +$

E). The first subsample, R_1 , is used to select the initial set of factor proxies. Thereafter, one observation from R_2 is added, and this new larger sample is used to recursively select a second set of factor proxies. This is continued until the second subsample is exhausted, yielding a sequence of R_2 different vectors of factor proxies. Individual proxies are then ranked according to their selection frequency, and those occurring the most frequently are selected and fixed for further use in constructing E ex ante predictions. Loading parameters for the proxies are still re-estimated prior to the formation of each new recursive prediction although the set of proxies is fixed throughout the forecast experiment. The advantage to this approach is that noise across the proxy selection process is potentially suppressed.

3 Empirical Methodology

3.1 Prediction

In our setup, the target variable to be predicted (either output growth or inflation) is defined as y_t . Namely, our objective is to predict the marginal growth rates $y_{t+h} = \log(\frac{Y_{t+h}}{Y_{t+h-1}})$, where Y_t is the level of the variable of interest. As discussed in the previous subsection, the available data is split into three subsamples such that $T = R_1 + R_2 + E$. Here, E denotes our ex-ante prediction period, and $R = R_1 + R_2$ denotes the number of observations used to estimate factors, and the sub-samples R_1 and R_2 are used in the selection of observable factor proxies.

In our prediction experiments, we begin by specifying simple linear prediction models. These so-called “strawman” models have been found to perform well when used to predict a variety of macroeconomic variables (see e.g. Swanson and White (1995,1997)), and have the following functional form:

$$y_{t+h} = \sum_{i=1}^{p1} \alpha'_i W_{t-i+1} + \sum_{j=1}^{p2} \beta_j y_{t-j+1} + \varepsilon_{t+h}, \quad (10)$$

where W_t is a vector of observable variables, $p1$ and $p2$ are lags selected using the Schwarz information criterion (SIC), and h is the prediction horizon. Note that this model nests random walk, random walk with drift, and AR models, which are other commonly used “strawman” models. These other strawman models were examined, but they are excluded from further discussion because their inclusion does not change our findings. The above linear model, where our so-called

macroeconomic indicators are contained in W_t , is one benchmark against which we compare all other prediction models. When constructing predictions, this model is estimated recursively.

A second set of models that we examine are factor models of the variety discussed in Section 2.1. Namely, predictions are formed using models of the following variety:

$$y_{t+h} = \alpha' F_t + \sum_{j=1}^p \beta_j y_{t-j+1} + \varepsilon_{t+h} \quad (11)$$

Forecasts using this model are constructed in the following manner. At each recursive iteration, the panel dataset of stationary variables is standardized to have zero mean and unit variance. The number of factors and the factors are re-estimated from this stationary panel dataset with unit variance and zero mean. The factor forecast model is then re-estimated by OLS and the h -step ahead forecast is constructed. This means that the specification for the factor forecast model can change at each recursive iteration. The number of lags of the target variable, p , is selected using the SIC, based on examination of the first subsample of data, R_1 . One of the main reasons that the factor model of Stock and Watson (2002a,b) is included as one of the prediction models is because such models are typically difficult to consistently outperform. The factor model thus serves as a sort of “credibility check” against which all other models are compared.

Our final group of prediction models are those constructed using observable factor proxies (see Section 2.3 for complete details on how observable proxies are selected). Unlike the factors in the factor forecast model, the observable variables in these models (denoted below as Model 2 to Model 8) are kept the same at each recursive iteration, although they are standardized to have zero mean and unit variance at each recursive iteration.

In summary, the prediction models that we consider include the following:

Factor: $\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' \tilde{F}_t + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$: \tilde{F}_t contains the estimated Stock and Watson diffusion indices.

Model 1: $\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{1t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$: W_{1t} contains the variables that make up the “macroeconomic indicators”.

Model 2: $\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{2t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$: Consider the model $\hat{X}_i = \hat{F}_i \Omega' + e$ where $\hat{X}_i \subset X$ is a specific class of variables. \hat{F}_i is made up of the first two principal component factors that underlie \hat{X}_i alone. The classes considered for \hat{X}_i are \hat{X}_1 : CPI; \hat{X}_2 : Industrial Production; \hat{X}_3 : Housing; \hat{X}_4 : Employment; \hat{X}_5 : Yields. Each of these five classes contains one of the “macroeconomic indicators”. The smoothed $A(j)$ statistic is then used to select one observable variable from X that proxies $\hat{F}_i \forall i$. W_{2t} contains these five

factor proxies in conjunction with the remaining macroeconomic indicators that are not included in the above classes. Only five classes are considered for \widehat{X}_i because the other macroeconomic indicators are members of classes that are too small to meaningfully apply our factor analysis.

Model 3: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{3t} + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: W_{3t} contains the variables selected by the smoothed $A(j)$ statistic without spreads

Model 4: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{4t} + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: Consider the model $\widehat{X}_i = \widehat{F}_i \Omega' + e$ where $\widehat{X}_i \subset X$ is a specific class of variables. \widehat{F}_i is made up of the first two principal component factors that underlie \widehat{X}_i alone. The classes considered for \widehat{X}_i are \widehat{X}_1 : Housing; \widehat{X}_2 : Industrial Production; \widehat{X}_3 : Yields. Each of these three classes contains one of the “macroeconomic indicators”. The smoothed $A(j)$ statistic is then used to select two observable variable from X that proxy $\widehat{F}_i \forall i$. W_{4t} contains these six factor proxies in conjunction with the remaining observable variables that are included in W_{3t} but not in the above classes. Only three classes are considered for \widehat{X}_i because the other variables are members of classes that are too small to meaningfully apply our factor analysis.

Model 5: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{5t} + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: W_{5t} contains a subset of the “macroeconomic indicators” selected by the smoothed $A(j)$ statistic

Model 6: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{6t} + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: Consider the model $\widehat{X}_i = \widehat{F}_i \Omega' + e$ where $\widehat{X}_i \subset X$ is a specific class of variables. \widehat{F}_i is made up of the first two principal component factors that underlie \widehat{X}_i alone. The classes considered for \widehat{X}_i are \widehat{X}_1 : Housing Starts; \widehat{X}_2 : Industrial Production. Each of these two classes contains one of the “macroeconomic indicators”. The smoothed $A(j)$ statistic is then used to select one observable variable from X that proxies $\widehat{F}_i \forall i$. W_{6t} contains these two factor proxies in conjunction with the remaining observable variables that are not included in W_{5t} . Only five classes are considered for \widehat{X}_i because the other variables are members of classes that are too small to meaningfully apply our factor analysis.

Model 7: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{7t} + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: W_{7t} contains a subset of the “macroeconomic indicators” selected by the smoothed $A(j)$ statistic in addition to Money Supply and the 6-month Treasury Bill Yield.

Model 8: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{8t} + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: W_{8t} contains the variables in W_{6t} plus Money Supply and the 6-month Treasury Bill Yield.

Models 1S-8S: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{nt}^S + \sum_{j=1}^p \widehat{\beta}_j y_{t-j+1}$: $W_{nt}^S = (W'_{nt}, S'_t)'$ and S_t is a vector of spreads listed in Panel I of Table 1.

Note that in the above models, explanatory variable are denoted as W_{1t} to W_{8t} . The observable contained in each of W_{1t} to W_{8t} are listed in Table 2.

Of note is that various classes of variables in the Stock and Watson dataset have aggregates as well as subaggregates. However, when picking macroeconomic “indicators” one is often given little guidance as to which variable is suitable. For example, if an indicator is CPI, one is left with a choice concerning precisely which CPI variable to use. In general, it is not a trivial task to pick a representative variable from a particular class of variables. Included in our macroeconomic indicators are CPI, industrial production and housing starts. However, there are at least 9 CPI, 12 industrial production, and 9 housing start variables in our panel dataset. Picking a representative CPI variable can consequently be tricky if not ad hoc. In specifying our benchmark model (i.e. Model 1), we subjectively pick variables as default representatives of the associated variable class. So for example, CPI (all items) is our representative CPI variable. The rationale behind this is that by construction, the aggregate variable contains information from the subaggregates. An alternative approach could be to use the smoothed $A(j)$ statistic on a specific class of variables. In this regard, let $\hat{X} \subset X$ where \hat{X} contains all the relevant aggregate and subaggregate variables in a particular class like CPI. Therefore, for CPI, \hat{X} will contain CPI: all items; CPI: apparel and upkeep; CPI: transportation; CPI: medical care; etc. The first two principal components are estimated from \hat{X} and the smoothed $A(j)$ statistic is used to select variables from X that proxies these two principal components. In this way, we methodically select a representative variable from a class. The principal components factors estimated from \hat{X} and the corresponding observable factor proxies will be called focused factors and focused factor proxies. In the empirical implementation, although X is the candidate set of proxies, the variable selected by the smoothed $A(j)$ statistic to proxy the focused principal components factors always ends up being a variable from \hat{X} . Some of the models have focused factor proxies as regressors, and in these cases we select focused variables only from classes with more than 10 subaggregates. Given this approach, the classes considered are CPI, Industrial Production, Housing, Employment and Yields. The Yield class includes the various interest rates and spreads.

From the above discussion, it follows that Model 1 represents our linear benchmark model, where W_{1t} contains the macroeconomic indicators as listed in Panel A of Table 1. Model 3 is the same model specified for Model 1 except that W_{3t} contains the factor proxies selected by the smoothed $A(j)$ statistic. In Model 5, W_{5t} contains a subset of the macroeconomic indicators selected by the smoothed $A(j)$ statistic (i.e. macroeconomic indicators that were selected as factor proxies). W_{7t} in Model 7 contains the same variables as W_{5t} with the explicit addition of the Money Supply

(M2) and the Yield on the 6-month Treasury Bill. All even numbered models, W_{nt} ($n = 2, 4, 6, 8$) contain the variables in the corresponding previous models with the relevant ones replaced by variables selected using our “focused principal components” approach which is discussed above. In Models 1S-8S, the corresponding variables in W_{nt} have been augmented with the spreads in Panel I of Table 1. Figure 1 depicts various spreads, including those listed in Panel I of Table 1 that are “selected” using our observable factor proxy approach (see Panel 1 of the figure), as well as standard yield spreads often used in the empirical macroeconomics literature (see Panel 2 of Figure 1). Of note is that the dynamic behavior of the preads differs markedly when comparing the two panels of plots.

3.2 Predictive Evaluation

The test statistic used for all forecast evaluation is the t -statistic for equal forecast accuracy discussed in Diebold and Mariano (1995) and West (1996), under quadratic loss. Note that because we use the same loss function both in- and out-of-sample, parameter estimation error vanishes asymptotically, and the simpler DM version of the test can be used. As discussed above, the sample is divided into two sub-samples R and E such that $T = R + E$ and $R = R_1 + R_2$. The sub-sample used to initially estimate the models spans 1 to R . The number of out-of-sample observations as well as the number of h -step ahead predictions span $R + 1$ to $R + E - h$, for a total of $E - h$ predictions. Forecasts for both null and alternative models are made recursively using least squares. In the context of Diebold and Mariano (DM: 1995), let $\hat{\varepsilon}_{a,t+h} = y_{t+h} - \hat{\alpha}'_a W_{a,t} + \sum_{j=1}^p \hat{\beta}_{a,j} y_{t-j+1}$ and $\hat{\varepsilon}_{b,t+h} = y_{t+h} - \hat{\alpha}'_b W_{b,t} + \sum_{j=1}^p \hat{\beta}_{b,j} y_{t-j+1}$, then the sample mean square forecast error is defined to be: $MSFE = \frac{1}{E-h} \sum_{t=R+1}^{T-h} \hat{\varepsilon}_t^2$. The null hypothesis of equal forecast accuracy from two models is given by $H_0 : E[d_t] = 0$, where $d_t = \varepsilon_{a,t}^2 - \varepsilon_{b,t}^2$ is the loss differential and $\bar{d} = \frac{1}{E-h} \sum_{t=R+1}^{T-h} \hat{d}_t$, where $\hat{d}_t = \hat{\varepsilon}_{a,t}^2 - \hat{\varepsilon}_{b,t}^2$. The DM test statistic is then:

$$DM = \frac{\frac{1}{E-h} \sum_{t=R+1}^{T-h} \hat{d}_t}{\sqrt{\frac{1}{E-h} \sum_{j=-\bar{j}}^{\bar{j}} \sum_{t=R+1+j}^{T-h} K\left(\frac{j}{M}\right) (\hat{d}_t - \bar{d})(\hat{d}_{t-j} - \bar{d})}},$$

where $K\left(\frac{j}{M}\right)$ is the kernel with bandwidth M . For non-nested models of the sort that arise when comparing our alternative predictive models, the distribution of this statistic is standard normal, but this is not the case for nested models. Under the null hypothesis, population forecast errors of

the restricted and alternative models are identical, implying $d_t = 0, \forall t$, in population. This means that the population variance of d_{t+h} is equal to 0, and standard inference does not apply. One can, however, use the results of Clark and McCracken (2005) and McCracken (2007) in order to obtain valid critical values in this case.

4 Data

The dataset used in this paper is that used in Stock and Watson (SW: 2005). The SW dataset contains 132 monthly time series for the United States from 1960:1 to 2003:12. A updated version of this dataset is examined by Stock and Watson (2006) using data aggregated to a quarterly frequency, and ending in 2006:4. In this paper, however, we only report results based on analysis of the monthly dataset, given that the reporting frequency of the variables that we use as our macroeconomic “indicators” is monthly. Moreover, by using data only up until 2004, we avoid tricky issues associated with data revision that may impact upon our empirical findings (see Diebold and Rudebusch (1991) and Croushore and Stark (2001,2003) for further discussion). Results based on our analysis of the quartely dataset are available upon request from the authors.

The time series in the dataset were obtained from the Global Insights Basic Economic Database or The Conference Board’s Indicators Database (TCB). Other series were calculated by Stock and Watson with prior information from the two databases mentioned above. The variables were selected from the following categories of macroeconomic time series: real output and income; employment, manufacturing and trade sales; consumption; housing starts and sales; real inventories and inventory-sales ratios; orders and unfilled orders; stock price indices; exchange rates; interest rate spreads; money and credit quantity aggregates; and price indexes.

The variables that we initially chose to comprise our group of our macroeconomic “indicators” were real gross domestic product (GDP); the consumer price index (CPI); nonfarm payroll employment; housing starts; industrial production/capacity utilization; retail sales, business sales and inventories; advanced durable goods shipments, new orders and unfilled orders; lightweight vehicle sales; the yield on the 10-year treasury bond; the S&P 500 stock price index; and M2. The choice of these variables was motivated by examination of a list of “policy-relevant” variables appearing on the New York Federal Reserve Bank website (<http://www.newyorkfed.org/education/bythe.html>). However, data for two of these variables (business sales and inventories as well as advanced durable

goods shipments, new orders and unfilled orders), is available only from 2002. In light of this fact, these two “indicators” are replaced with manufacturing and trade sales as well as manufacturing and trade inventories. In addition, we use industrial production: durable goods materials as a replacement for advanced durable goods shipments, new orders and unfilled orders. Finally, data for lightweight vehicle sales is only available beginning in 1976:1, and hence all other variables were truncated at this date in order to obtain a balanced panel dataset.

The SW dataset includes spreads constructed as the difference between the yields on the following bonds and the effective federal funds rate: commercial paper; 3-month, 6-month, 1-year, 5-year, and 10-year Treasury bills/bond; and Moody’s Aaa and Baa corporate bond yields. We additionally considered the difference between the following yield variables and the 3-month Treasury bill yield: 6-month, 1-year, 5-year, and 10-year Treasury bills/bond yields; and Moody’s Aaa and Baa corporate bond yields. Finally, we examined a reverse yield gap spread defined as the difference between the yield on the 10-year Treasury bond and the S&P 500 common stock dividend yield. In summary, the final dataset used in this paper includes a total of 139 variables.

Note that, given the above discussion, one of the traditional spreads considered in this paper is the slope of the yield curve. This is defined as the difference between a long-term and a short-term interest rate. Under some restrictive assumptions such as constant real interest rates over time, perfect substitutability between assets of different maturities and that the expectations theory of the term structure holds, the slope of the yield curve will provide an exact measure of the market’s expected inflation path (see Davis and Fagan (1997)). Another variety of spread that we consider is the reverse yield gap which is defined as the difference between yields on long-term or short-term debt instruments and the dividend yield on domestic equity. Debt securities issued by the government are typically regarded as risk-free assets with guaranteed coupon payments while equities are risky assets with non-guaranteed dividend payments. The reverse yield gap consequently reflects the premium that an investor is likely to demand to compensate for the extra risk (Nobili (2005)). Thus, increases in this spread will predict downturns in economic activity. Furthermore, the reverse yield gap is expected to be positively related to inflation because a rising spread will accompany a tightening of monetary policy in response to increased inflationary pressures (Nobili (2005)).

All variables that are not already stationary (based upon examination of the properties of the variables using augmented Dickey-Fuller tests), were differenced or log differenced prior to

prediction model estimation and forecast construction.

5 Empirical Results

5.1 Observable Factor Proxies

In our empirical implementation of the diffusion index methodology discussed in Section 2.3, the factor proxies selected by the smoothed $A(j)$ statistic are listed in Panel C and Panel I of Table 1. Our macroeconomic indicators are listed in Panel A of Table 1. Examination of the two groups of observable variables indicates a strong resemblance. With the exception of the three spread variables listed in Panel I of Table 1, most of the other variables from the two groups are either identical or can be viewed as belonging to similar classes of variables. This result suggests that the macroeconomic indicators are suitable proxies to the underlying factors describing the dynamics of our entire dataset. Moreover, we shall see below that various subsets of these macroeconomic indicators additionally outperform diffusion indices when used to predict output growth and inflation at various forecast horizons.

It is noteworthy that the only observable factor proxies selected that are not in our set of macroeconomic indicators are the three spread variables defined as the difference between the federal funds rate and yields on either the 5-year treasury bond, the 10-year treasury bond, or Moody's Aaa corporate bond yield. In general, yield spread variables have played an important role in empirical research in recent years, although these particular spreads are not always those that have been found to be useful, particularly when the objective has been to construct the most accurate prediction models. More specifically, note that there is a vast literature on the predictive content of spreads for inflation and output growth. The fundamental idea underlying this body of literature is that financial market participants are forward-looking. Asset prices and yields consequently embody useful information such as expectations of future economic activity. Spreads are also quite reliable as real-time indicators, as they are not subject to revision. The usefulness of spreads for forecasting output growth and inflation in the U.S. has previously been studied in Laurent (1988, 1989); Harvey (1988, 1989); Stock and Watson (1989); Mishkin (1990a,b, 1991); Estrella and Hardouvelis (1991); Jorion and Mishkin (1991); Friedman and Kuttner (1991). Other researchers who have also considered the predictive content of spreads in the U.K., some European countries and other OECD countries include Davis (1993); Davis and Henry (1994); Plosser and

Rouwenhorst (1994); Davis, Henry and Pesaran (1994); Bonser-Neal and Morley (1997); Kozicki (1997); Gerlach (1997); Davis and Fagan (1997); Estrella and Mishkin (1997). In order to observe the difference between the yield spreads that were selected as observable factor proxies and other spreads often examined in the literature, please refer to Figure 1. Recall from Section 3.1 that Figure 1 contains plots of the three yields spreads selected as proxies, as well as a yield curve slope (defined to be the difference between the 10-year treasury bond and the 3-month treasury bill yield) and the reverse yield gap.

5.2 Prediction Experiments

The variables that we predict in our experiments are output growth, which is proxied for by the growth rate of industrial production (total index), and the growth rate in CPI (all items). The prediction models used in our experiments are summarized in Table 2 and in the Section 3.1. All prediction models are estimated recursively using least squares, and the ex-ante prediction period is 1994:09-2003:12. The choice of ex-ante prediction period was set arbitrarily by partitioning the dataset into three approximately equal sub-samples, the first corresponding to *R1*, the second to *R2*, and the third to *E*, as discussed in Section 3.1.

The main results of our prediction experiments are collected in Tables 3 and 4. The first column in both tables contains the mnemonic used to denote the prediction model (see Table 2 for further details). Numerical entries in the second column are MSFEs. Those in bold correspond to models with lower MSFEs, relative to Model 1. Boxed MSFEs represent the lowest MSFE amongst all models considered, for a particular prediction horizon, $h = 1, 3, \text{ or } 12$. In addition to Model 1, the other key “benchmark” model against which the out-of-sample forecast performance of the alternative models is compared is the diffusion index model. Numerical entries in the third column of the tables report DM test statistic, where the benchmark model is the diffusion index model. Since the factor model and all the alternative models are non-nested, these DM test statistics have a standard normal limiting distribution (see Section 3.2 for further discussion). Negative entries denote cases for which the benchmark model has a lower point MSFE than the alternative model, and positive entries denote the converse. Starred entries denote models where the null of equal predictive accuracy is rejected at 20%(*), 10%(**) and 5%(***) significance levels, respectively. Numerical entries in the next column in the tables are DM test statistics for cases where the benchmark model is Model 1. The final column of entries in the tables are DM test statistics where

a model with spreads is compared against one without spreads. Thus, for example, the first entry in Panel A of Table 3 reports the DM test statistic for the case where the benchmark model is Model 1, and the alternative model is Model 1 with three spread variables added as additional regressors. As discussed below, the three spread variables that we use in these tests correspond to the three spread variables chosen as observable proxies for the factors using our diffusion index model construction methodology.

In Table 3, where the target variable is CPI inflation, the SW factor model significantly outperforms all alternative models at the 3 step horizon. On the other hand, the evidence is more mixed at other forecast horizons, as the null hypothesis of equal predictive accuracy between the factor model and the alternative models is never rejected. Moreover, at the 1 and 12 month forecast horizon, the point MSFE of SW model is higher than that associated with a variety of alternative models. Thus, most of the alternative models do indeed provide reduced MSFEs, relative to the factor model. However, this reduction is generally not enough to cause a rejection of the null hypothesis of equal predictive accuracy.

Turning to Table 4, where results from our output growth prediction experiments are summarized, note that the factor model is often “worse” than various competitor models, when comparing point MSFEs, at all horizons; and is in various cases significantly “worse”. Moreover, results increasingly favor our alternative models, as the forecast horizon is increased.

Summarizing the above results, we have evidence suggesting that many of our alternative models have lower MSFEs than our benchmark factor model, in 5 of 6 variable/horizon combinations. Moreover, the lowest-MSFE model is a model with our new spread variables in 4 of 6 combinations. Given that one of the remaining 2 combinations is one for which the factor model is MSFE-best, we have rather surprising evidence of the usefulness of our three spread variables; particularly when one considers the fact that some of the alternative models are quite parsimonious, and that the inclusion of three new spread variables to them adds substantially to the parameter estimation error associated with estimation of the models. Finally, the specification of models that provide superior predictions and the specification of models that are useful for policy monitoring are usually one and the same, in the sense that models with our macroeconomic “indicators” are generally MSFE-best.

Turning now to a discussion of our yield spread-type observable factor proxies variables, note that we have already pointed out that the lowest-MSFE model is a model with our new spread variables in 4 of 6 combinations. However, it should also be noted that evidence in favor of the

spread-augmented models is weakest at our shorter forecast horizons, when considering inflation. This could be because at the shorter horizons, the macroeconomic indicators already encapsulate much of the information contained in spreads. For longer horizons, inflation prediction is improved by including spreads, though. In particular, Model 1S strongly and significantly outperforms Model 1 at the 12-month ahead horizon (the DM test statistic is 2.02), suggesting that spreads have marginal predictive content for inflation at longer horizons. The story is somewhat similar when considering output growth - spread-augmented models are MSFE-best at a 10% significance level in many cases; and in cases where the point MSFEs associated with spread augmented models are larger, there is usually nothing to choose between the models, as the null hypothesis of equal predictive accuracy fails to reject. Thus, we have evidence that our macroeconomic indicators can usefully be augmented by including our three yield spread-type observable factor proxy variables.

Recall that parsimonious models may outperform larger models simply because parsimonious models set to zero coefficients that are truly zero (or close to zero) in the population. By setting these coefficients to zero instead of estimating them, the parsimonious models gain efficiency, and thereby benefit from increased predictive accuracy. Evidence in support of this point is provided in Table 3, where Models 5 and 7, which are restricted versions of Model 1, both have lower MSFEs, relative to Model 1, at all forecast horizons. Also, in Table 4, Models 5 and 7 outperform Model 1 half the time, with the more parsimonious models outperforming Model 1 at longer horizons. On the other hand, setting the coefficients of variables with “strong” predictive content for the target variable to zero simply to deliver parsimony can clearly be costly. Tables 3 and 4 illustrate this point. The “spread augmented” models, which are by construction less parsimonious than their counterparts not containing spreads, yield improved predictions for a variety of models and forecast horizons. In particular, for inflation, prediction models with spreads outperform models without spreads at the 1-month ahead horizon around half the time, and at 3- and 12-month ahead horizons in virtually all cases, based on point MSFE comparisons. Moreover, at the 12-month horizon, most of the prediction models including spreads are also significantly better. For output growth, models with spreads outperform those without spreads around 90% of the time at the 1-month ahead horizon, but add little when considering longer horizons, based on point-MSFE comparison. These results constitute further evidence in favor of the marginal predictive content of our spread proxies variables.

In Panel 1 of Figure 2, note that observed CPI inflation is very volatile between September

1999 and September 2002, but relatively calmer in years prior to September 1999. Interestingly, the MSFEs of Models 1 and 1S are also almost identical in the relatively calmer period prior to September 1999. However, as illustrated in Panel 2 of Figure 2, the MSFE of Model 1S is substantially lower than that for Model 1 during the high volatility period between September 1999 and September 2002. These observations suggest that our macroeconomic indicators (without spreads) perform best when forecasting inflation in non-volatile time periods. On the other hand, the forecast performance of the indicators can be improved by including spreads when forecasting inflation in times of high volatility. Although the addition of spreads significantly improves the forecast performance of the macroeconomic indicators when predicting inflation in this case, there is little difference between the MSFEs of Models 1 and 1S when forecasting output growth (see Panel 3 of Figure 2).

Finally, notice that for both inflation and output, the lowest MSFE occurs for one of our even numbered models, at all forecast horizons. As discussed in Section 3.1, all even numbered models have variables selected from their respective classes using the smoothed $A(j)$ statistic. One interpretation of this result is that although aggregate variables such as those used in our benchmark Model 1 by design contain some information from all of the members in a class, they might not necessarily be “optimally” representative of that class, at least when it comes to prediction. Rather, we find that a variety of subaggregates chosen using the smoothed $A(j)$ statistic have “better” predictive content. This in turn suggests that one direction for future research is the inclusion of multiple members of particular classes in our prediction models.

6 Concluding Remarks

In order to obtain early indication of the impact of current monetary policy, policy makers monitor select financial and macroeconomic variables. Likewise, many forecasters do the same, using revisions and fresh information on select variables to update their models and predictions of key macroeconomic variables. We lend credence to the use of such judiciously chosen macroeconomic “indicators” by establishing that a set of variables that are likely to be related closely to those used by federal and private agencies are largely the same as those variables that proxy common factors constructed via analysis of a largescale macroeconomic dataset using the diffusion index methodology of Stock and Watson (2002a,b). Out-of-sample forecast exercises further suggest that

augmenting the particular set of macroeconomic indicators that we consider with yield spreads is in some cases useful when forecasting inflation and output growth. The particular spreads that we find to be most helpful are constructed as the difference between short or long term debt instruments and the federal funds rate. Interestingly, spreads constructed as yield curve slopes and reverse yield gaps are not found to provide additional marginal predictive content. Finally, we find that our initial set of macroeconomic indicators performs best when forecasting inflation in non-volatile time periods, while the forecast performance of the indicators is most clearly improved by including spreads when forecasting inflation in times of high volatility.

7 References

- Armah, N. and Swanson, N. (2010). Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Large Scale Macroeconomic Time Series Environments. *Econometric Reviews* (forthcoming).
- Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica* 71(1), 135-172.
- Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70, 191-221.
- Bai, J. and Ng, S. (2006a). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* 74(4), 1133-1150.
- Bai, J. and Ng, S. (2006b). Evaluating Latent and Observed Factors in Macroeconomics and Finance. *Journal of Econometrics* 113, 507-537.
- Bai, J. and Ng, S. (2007). Determining the Number of Primitive Shocks in Factor Models. *Journal of Business and Economic Statistics* 25, 52-60.
- Boivin, J. and Ng, S. (2005). Understanding and Comparing Factor Based Macroeconomic Forecasts. *International Journal of Central Banking* 1(3), 117-152.
- Bonser-Neal, C and Morley, T. (1997). Does the Yield Spread Predict Real Economic Activity? A Multicountry Analysis. *Federal Reserve Bank of Kansas City Economic Review* 82(3), 37-53.
- Clark, T. and McCracken, M. (2005). Evaluating Direct Multi-Step Forecasts, *Econometric Reviews*, 24, 369-404.
- Davis, E. (1993). VAR Modeling of the German Economy with Financial Spreads as Key Indicator Variables. London School of Economics, *Discussion Paper* 59.
- Croushore, D. and T. Stark (2001). A Real-Time Dataset for Macroeconomists. *Journal of Econometrics* 105, 111-130.
- Croushore, D. and T. Stark (2003). A Real-Time Dataset for Macroeconomists: Does Data Vintage Matter? *Review of Economics and Statistics* 85, 605-617.
- Davis, E. and Fagan, G. (1997). Are Financial Spreads Useful Indicators of Future Inflation and Output Growth in EU Countries? *Journal of Applied Econometrics* 12, 701-714.
- Davis, E. and Henry, S. (1992). An Aggregate VAR Model with Financial Spreads. *Working Paper*. Bank of England.
- Davis, E. and Henry, S. (1994). The Use of Financial Spreads as Indicator Variables: Evidence for the UK and Germany. IMF, *Working Paper* 31.
- Davis, E, Henry, S. and Pesaran, B. (1994). The Role of Financial Spreads: Empirical Analysis of Spreads and Real Economic Activity. *Manchester School of Economic & Social Studies* 62(4), 374-394.
- Diebold, F. and Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*. 13, 253-263.
- Diebold, F.X. and G.D. Rudebusch (1991). Forecasting Output with the Composite Leading Index: A Real-Time Analysis. *Journal of the American Statistical Association* 86, 603—610.
- Ding, A. and Hwang, J. (1999). Prediction Intervals, Factor Analysis Models, and High-Dimensional Empirical Linear Prediction. *Journal of the American Statistical Association* 94, 446-455.
- Estrella, A. and Hardouvelis, G. (1991). The Term Structure as a Predictor of Real Economic Activity. *Journal of Finance* 46(2), 555-576.

- Estrella, A. and Mishkin, F. (1997). The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank. *European Economic Review* 41(7), 1375-1401.
- Estrella, A. and Mishkin, F. (1998). Predicting U.S. Recessions: Financial Variables as Leading Indicators. *Review of Economics and Statistics* 80, 45-61.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *The Review of Economics and Statistics* 82(4), 540-552.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association* 100(471), 830-840.
- Forni, M. and Reichlin, L. (1996). Dynamic Common Factors in Large Cross-Sections. *Empirical Economics* 21, 27-42.
- Forni, M. and Reichlin, L. (1998). Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle. *Review of Economic Studies* 65, 453-474.
- Friedman, B. and Kuttner, K. (1991). Why Does the Paper-Bill Spread Predict Real Economic Activity? *NBER Working Paper*, No. 3879.
- Gerlach, S. (1997). The Information Content of the Term Structure: Evidence for Germany. *Empirical Economics* 22(2), 161-179.
- Harvey, C. (1988). The Real Term Structure and Consumption Growth. *Journal of Financial Economics* 22(2), 305-330.
- Harvey, C. (1989). Forecasts of Economic Growth from the Bond and Stock Markets. *Financial Analysts Journal* 45(5), 38-45.
- Jorion, P. and Mishkin, F. (1991). A Multi-Country Comparison of Term Structure Forecasts at Long Horizons. *Journal of Financial Economics* 29(1), 59-80.
- Kozicki, S. (1997). Predicting Real Growth and Inflation With the Yield Spread. *Federal Reserve Bank of Kansas City Economic Review* 82(4), 39-57.
- Laurent, R. (1988). An Interest Rate-Based Indicator of Monetary Policy. *Federal Reserve Bank of Chicago Economic Perspectives* 12(1), 3-14.
- Laurent, R. (1989). Testing the Spread. *Federal Reserve Bank of Chicago Economic Perspectives* 13(4) 22-34.
- McCracken, M. (2007). Asymptotics for Out-of-Sample Tests of Granger Causality. *Journal of Econometrics*, 140, 719-752.
- Mishkin, F. (1990a). The Information in the Longer-Maturity Term Structure About Future Inflation. *Quarterly Journal of Economics* 55, 815-828.
- Mishkin, F. (1990b). What Does the term Structure Tell Us About Future Inflation. *Journal of Monetary Economics* 25(1), 77-96.
- Mishkin, F. (1991). A Multi-Country Study of the Information in the Shorter Maturity Term Structure About Future Inflation. *Journal of International Money and Finance* 10(1), 2-22.
- Nobili, A. (2005). Forecasting Output Growth and Inflation in the Euro Area: Are Financial Spreads Useful? *Banca D'Italia*, No. 544.
- Plosser, C. and Rouwenhorst, K. (1994). International Term Structures and Real Economic Growth. *Journal of Monetary Economics* 33(1), 133-155.

- Rapach, D. and Strauss, J. (2007). Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Unemployment Growth. *Econometric Reviews*, (forthcoming).
- Stock, J. and Watson, M. (1989). New Indexes of Coincident and Leading Economic Indicators. In Blanchard, O. and Fischer, S. (eds.), *NBER Macroeconomics Annual*, Cambridge, MIT Press, 351-393.
- Stock, J. and Watson, M. (1998). Diffusion Indexes. *Working Paper 6702*, National Bureau of Economic Research.
- Stock, J. and Watson, M. (1999). Forecasting Inflation. *Journal of Monetary Economics* 44, 293-335.
- Stock, J. and Watson, M. (2002a). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics* 20, 147-161.
- Stock, J. and Watson, M. (2002b). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of American Statistical Association* 97, 1167-1179.
- Stock, J. and Watson, M. (2004). An Empirical Comparison of Methods for Forecasting Using Many Predictors. *Working Paper*, Princeton University.
- Stock, J. and Watson, M. (2005). Implications of Dynamic Factor Models for VAR Analysis. *Working Paper 11467*, National Bureau of Economic Research.
- Stock, J. and Watson, M. (2006). Macroeconomic Forecasting Using Many Predictors. In Elliott, G., Granger, C. and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, North Holland.
- Swanson, N. and White, H. (1995). A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks. *Journal of Business and Economic Statistics* 13, 265-279.
- Swanson, N. and White, H. (1997). A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics* 79, 540-550.
- West, K. (1996). Asymptotic Inference About Predictive Ability. *Econometrica* 64, 1067-1084.

Table 1: Predictors Used in Empirical Experiments*

Regressors	Stationarity Transformtion
Panel A: Model 1 (W_{1t})	
Consumer Price Index: all items	$\Delta \log$
Nonfarm Payroll Employment: total private	$\Delta \log$
Housing Starts: total farm and nonfarm	\log
Industrial Production Index: total index	$\Delta \log$
Capacity Utilization	Δlevels
Retail Sales of Stores	$\Delta \log$
Manufacturing and Trade Sales	$\Delta \log$
Manufacturing and Trade Inventories	$\Delta \log$
Industrial Production Index: durable goods materials	$\Delta \log$
Lightweight Vehicle Sales	$\Delta \log$
Yield on 10-year Treasury Bond	Δlevels
S&P 500 Stock Price Index: Composite	$\Delta \log$
Money Supply - M2	$\Delta \log$
Panel B: Model 2 (W_{2t})	
Consumer Price Index: apparel and upkeep	$\Delta \log$
Nonfarm Payroll Employment: goods producing	$\Delta \log$
Housing Starts: northeast	\log
Industrial Production Index: manufacturing	$\Delta \log$
Capacity Utilization	Δlevels
Retail Sales of Stores	$\Delta \log$
Manufacturing and Trade Sales	$\Delta \log$
Manufacturing and Trade Inventories	$\Delta \log$
Industrial Production Index: durable goods materials	$\Delta \log$
Lightweight Vehicle Sales	$\Delta \log$
Yield on 6-month Treasury Bill	Δlevels
S&P 500 Stock Price Index: Composite	$\Delta \log$
Money Supply - M2	$\Delta \log$
Panel C: Model 3 (W_{3t})	
Housing Starts: total farm and nonfarm	\log
Housing Authorized: total new private housing units	\log
Industrial Production Index: total index	$\Delta \log$
Industrial Production Index: products, total	$\Delta \log$
Capacity Utilization	Δlevels
Yield on 6-month Treasury Bill	Δlevels
Yield on 1-year Treasury Bond	Δlevels
S&P 500 Stock Price Index: Composite	$\Delta \log$
S&P 500 Stock Price Index: Industrials	$\Delta \log$
S&P 500 Stock Price Index: Dividend Yield	Δlevels

Table 1 (cont.)*

Regressors	Stationarity Transformation
Panel D: Model 4 (W_{4t})	
Housing Starts: northeast	log
Housing Authorized by Building Permits: northeast	log
Industrial Production Index: manufacturing	Δ log
Industrial Production Index: nondurable consumer goods	Δ log
Capacity Utilization	Δ levels
Yield on 3-month Treasury Bill	Δ levels
Yield on 6-month Treasury Bill	Δ levels
Yield on 10-year Treasury Bond	Δ levels
S&P 500 Stock Price Index: Composite	Δ log
S&P 500 Stock Price Index: Industrials	Δ log
Panel E: Model 5 (W_{5t})	
Housing Starts: total farm and nonfarm	log
Industrial Production Index: total index	Δ log
Capacity Utilization	Δ levels
S&P 500 Stock Price Index: Composite	Δ log
Panel F: Model 6 (W_{6t})	
Housing Starts: northeast	log
Industrial Production Index: manufacturing	Δ log
Capacity Utilization	Δ levels
S&P 500 Stock Price Index: Composite	Δ log
Panel G: Model 7 (W_{7t})	
Housing Starts: total farm and nonfarm	log
Industrial Production Index: total index	Δ log
Capacity Utilization	Δ levels
Yield on 6-month Treasury Bill	Δ levels
S&P 500 Stock Price Index: Composite	Δ log
Money Supply - M2	Δ log
Panel H: Model 8 (W_{8t})	
Housing Starts: northeast	log
Industrial Production Index: manufacturing	Δ log
Capacity Utilization	Δ levels
Yield on 6-month Treasury Bill	Δ levels
S&P 500 Stock Price Index: Composite	Δ log
Money Supply - M2	Δ log
Panel I: Spreads	
Yield on 5-year Treasury Bond – Federal Funds Rate	levels
Yield on 10-year Treasury Bond – Federal Funds Rate	levels
Yield on Moody’s Aaa Corporate – Federal Funds Rate	levels
Panel J: Class Representatives Selected by the Smoothed $A(j)$ Statistic	
Housing Starts: Northeast (Housing Class)	log
CPI: Apparel and Upkeep (CPI Class)	Δ log
Industrial Production: Manufacturing (Industrial Production Class)	Δ log
Yield on 6-month Treasury Bill (Yield Class)	Δ levels
Nonfarm Payroll Employment: goods producing (Employment Class)	Δ log

* Notes: The second column under “Stationarity Transformation” indicates the data transformation that was performed to induce stationarity, levels means no transformation; Δ levels denotes first difference of the levels; log denotes the natural log function; and Δ log denotes first log differences.

Table 2: Prediction Models

Model	Specification
Factor	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' F_t + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 1	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{1t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 2	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{2t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 3	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{3t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 4	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{4t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 5	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{5t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 6	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{6t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 7	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{7t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$
Model 8	$\hat{y}_{t+h} = \hat{c} + \hat{\alpha}' W_{8t} + \sum_{j=1}^p \hat{\beta}_j y_{t-j+1}$

* Notes: See discussion in Section 3.1 for complete details on the models given in the table. The W_t variables in the models contain various regressors, as outlined in Table 1.

Table 3: Forecasting Experiment Results for CPI Inflation*

Model	MSFE	DM-Test Benchmark		
		Factor	Model 1	No Spread
Panel A: One-Month Ahead Forecast				
Factor	3.66			
Model 1	3.93	-0.98		
Model 1S	3.96	-0.81	-0.08	-0.08
Model 2	3.96	-0.96	-0.21	
Model 2S	3.88	-0.59	0.18	0.26
Model 3	3.51	0.53	2.39	
Model 3S	3.60	0.17	1.06	-0.32
Model 4	3.60	0.23	2.23	
Model 4S	3.49	0.51	1.44	0.36
Model 5	3.61	0.20	2.28	
Model 5S	3.68	-0.06	0.76	-0.20
Model 6	3.71	-0.18	1.59	
Model 6S	3.63	0.10	0.91	0.23
Model 7	3.61	0.19	2.27	
Model 7S	3.62	0.12	1.00	-0.05
Model 8	3.66	0.01	1.92	
Model 8S	3.55	0.34	1.33	0.34
Panel B: Three-Month Ahead Forecast				
Factor	3.16			
Model 1	4.15	-2.79***		
Model 1S	4.20	-3.39***	-0.21	-0.21
Model 2	4.34	-3.35***	-0.85	
Model 2S	4.30	-3.40***	-0.44	0.21
Model 3	3.81	-2.53***	1.46	
Model 3S	3.85	-3.19***	0.90	-0.21
Model 4	4.54	-4.28***	-1.21	
Model 4S	4.45	-4.15***	-0.72	0.39
Model 5	3.77	-2.30***	2.07	
Model 5S	3.76	-2.77***	1.11	0.02
Model 6	4.35	-3.95***	-0.67	
Model 6S	4.31	-3.77***	-0.36	0.16
Model 7	4.02	-2.53***	0.91	
Model 7S	3.99	-3.08***	0.57	0.14
Model 8	4.41	-3.70***	-1.17	
Model 8S	4.30	-3.54***	-0.41	0.49

Table 3 (cont.): Forecasting Experiment Results for CPI Inflation*

Model	MSFE	DM-Test Benchmark		
		Factor	Model 1	No Spread
Panel C: Twelve-Month Ahead Forecast				
Factor	4.57			
Model 1	4.71	-0.24		
Model 1S	3.93	1.18	2.02	2.02**
Model 2	4.26	0.57	1.81	
Model 2S	4.03	1.00	1.38	0.70**
Model 3	4.24	0.58	2.48	
Model 3S	3.75	1.51*	2.36	1.59**
Model 4	4.00	1.07	2.62	
Model 4S	3.78	1.49*	1.77	0.67**
Model 5	4.56	0.01	1.12	
Model 5S	3.79	1.52*	2.29	1.95**
Model 6	3.96	1.15	3.10	
Model 6S	3.73	1.55*	1.82	0.64**
Model 7	4.60	-0.06	0.82	
Model 7S	3.82	1.40	2.15	1.95**
Model 8	3.96	1.11	3.39	
Model 8S	3.69	1.56*	1.97	0.79**

* Notes: The first column in both tables contains the mnemonic used to denote the prediction model (see Table 2 for further details). Numerical entries in the second column are MSFEs. Those in bold correspond to models with lower MSFEs, relative to Model 1. Boxed MSFEs represent the lowest MSFE amongst all models considered, for a particular prediction horizon, $h=1,3$, or 12. In addition to Model 1, the other key "benchmark" model against which the out-of-sample forecast performance of the alternative models is compared is the diffusion index model. Numerical entries in the third column of the tables report DM test statistic, where the benchmark model is the diffusion index model. Since the diffusion index (i.e. the factor model) and all alternative models are non-nested, these DM test statistics have a standard normal limiting distribution (see Section 3.2 for further discussion). Negative entries denote cases for which the benchmark model has a lower point MSFE than the alternative model, and positive entries denote the converse. Starred entries denote models where the null of equal predictive accuracy is rejected at 20%(*), 10%(**) and 5%(***) significance levels, respectively. Numerical entries in the next column in the tables are DM test statistics for cases where the benchmark model is Model 1. The final column of entries in the tables are DM test statistics where a model with spreads is compared against one without spreads. Thus, for example, the first entry in Panel A of Table 3 reports the DM test statistic for the case where the benchmark model is Model 1, and the alternative model is Model 1 with three spread variables added as additional regressors.

Table 4: Forecasting Experiment Results for Output Growth*

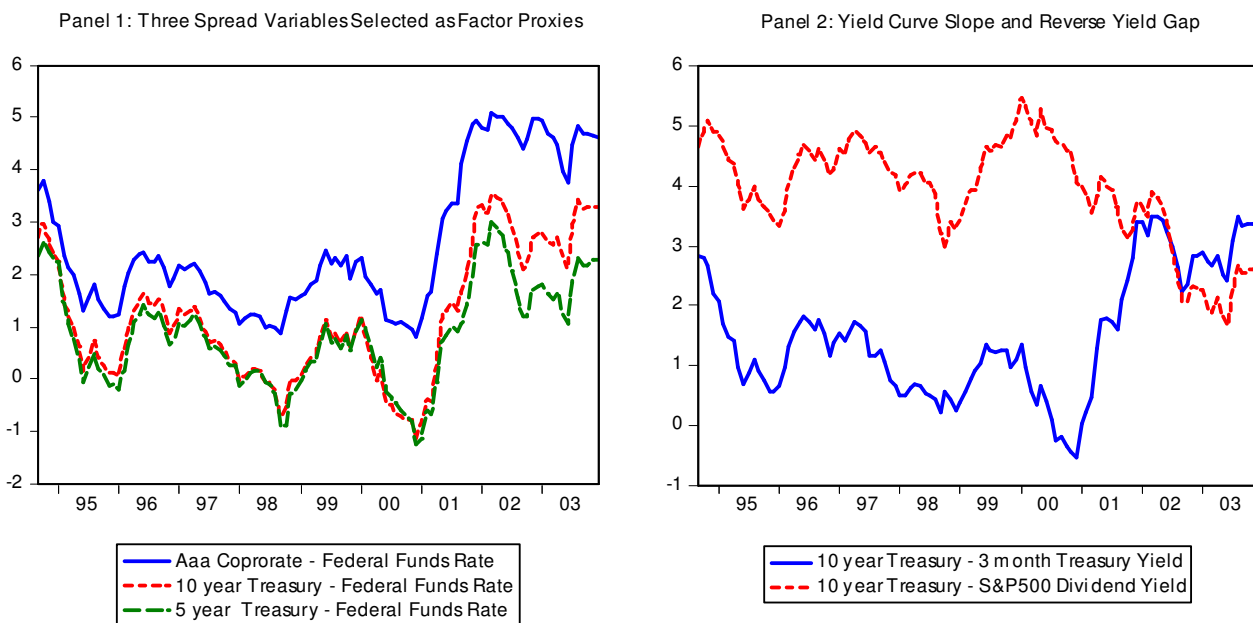
Model	MSFE	DM-Test Benchmark		
		Factor	Model 1	No Spread
Panel A: One-Month Ahead Forecast				
Factor	2.29			
Model 1	2.31	-0.07		
Model 1S	2.19	0.38	1.49	1.49**
Model 2	2.31	-0.08	0.00	
Model 2S	2.09	1.06	1.02	2.62**
Model 3	2.54	-1.21	-1.24	
Model 3S	2.41	-0.58	-0.54	1.92**
Model 4	2.63	-1.66**	-1.40	
Model 4S	2.33	-0.17	-0.08	3.04**
Model 5	2.73	-1.89**	-2.04	
Model 5S	2.50	-0.98	-1.01	3.86**
Model 6	2.74	-2.21***	-1.70	
Model 6S	2.58	-1.66**	-0.98	2.11**
Model 7	2.63	-1.55*	-1.93	
Model 7S	2.67	-1.41*	-2.16	-0.47
Model 8	2.56	-1.49*	-1.22	
Model 8S	2.41	-0.82	-0.46	1.81**
Panel B: Three-Month Ahead Forecast				
Factor	2.45			
Model 1	2.40	0.27		
Model 1S	2.43	0.16	-0.21	-0.21
Model 2	2.36	0.63	0.45	
Model 2S	2.45	-0.02	-0.36	-0.99
Model 3	2.44	0.05	-0.24	
Model 3S	2.52	-0.33	-0.58	-0.68
Model 4	2.26	1.07	0.77	
Model 4S	2.15	1.81**	1.39	0.87**
Model 5	2.37	0.90	0.24	
Model 5S	2.72	-1.43*	-1.75	-2.32
Model 6	2.31	1.49*	0.48	
Model 6S	2.50	-0.43	-0.55	-1.45
Model 7	2.56	-0.77	-1.61	
Model 7S	2.67	-1.28	-1.87	-0.94
Model 8	2.45	0.04	-0.36	
Model 8S	2.50	-0.41	-0.67	-0.58

Table 4 (cont.): Forecasting Experiment Results for Output Growth*

Model	MSFE	DM-Test Benchmark		
		Factor	Model 1	No Spread
Panel C: Twelve-Month Ahead Forecast				
Factor	3.44			
Model 1	3.07	2.36***		
Model 1S	3.02	2.52***	0.43	0.43
Model 2	3.07	1.59*	-0.01	
Model 2S	3.06	1.80**	0.06	0.10
Model 3	2.73	2.36***	1.31	
Model 3S	3.17	1.09	-0.45	-2.39
Model 4	2.94	1.58*	0.48	
Model 4S	3.02	1.70**	0.23	-0.48
Model 5	2.71	2.47***	1.53	
Model 5S	3.03	1.78**	0.22	-1.74
Model 6	2.70	2.59***	1.65	
Model 6S	3.00	1.93**	0.35	-1.73
Model 7	2.73	2.52***	1.80	
Model 7S	2.87	2.45***	1.15	-0.93
Model 8	2.72	2.71***	2.04	
Model 8S	2.87	2.53***	1.17	-1.10

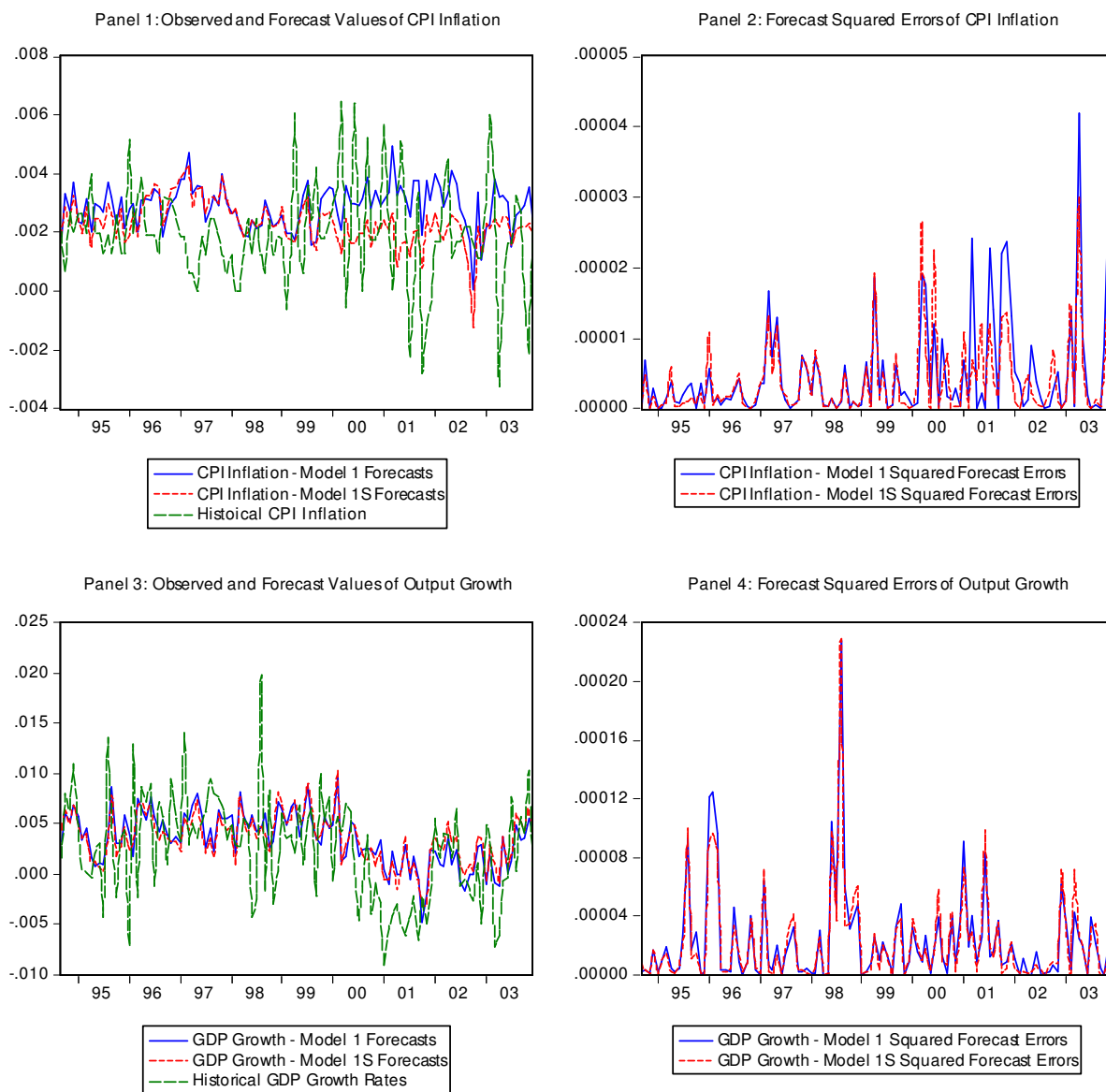
* Notes: See notes to Table 3.

Figure 1: Various Yield Spread Variables



Notes: The yield spreads in Panel 1 are the three spread variables selected as observable factor proxies, using the methodology discussed in Section 2.3, while those depicted in Panel 2 are standard measures of the slope of the yield curve and the reverse yield gap.

Figure 2: Forecasts of CPI Inflation and Output Growth at a 1-Step Ahead Horizon



Notes: Plots in Panels 1 and 3 depict actual and forecast values of CPI inflation and GDP growth for the period 1994:9-2003:12, and are all approximately “fully” revised data (see Section 4 for further details). Corresponding squared forecast errors are depicted in Panels 2 and 4. See Section 3.1 and 5.2 for further details about the prediction methodology used.