# ESSAYS IN FORECASTING

**BY NII AYI CHRISTIAN ARMAH**

**A dissertation submitted to the**

**Graduate School—New Brunswick**

**Rutgers, The State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Graduate Program in Economics**

**Written under the direction of**

**Norman Rasmus Swanson**

**and approved by**

_____

_____

_____

_____

**New Brunswick, New Jersey**

**May, 2009**

## ABSTRACT OF THE DISSERTATION

# Essays in Forecasting

**by Nii Ayi Christian Armah**

**Dissertation Director: Norman Rasmus Swanson**

This dissertation comprises three essays in macroeconomic forecasting. The first essay discusses model selection and predictive accuracy tests in the context of parameter and model uncertainty under recursive and rolling estimation schemes. Particular emphasis is placed on the construction of valid bootstrap procedures for calculating the impact of parameter estimation error on the class of test statistics with limiting distributions that are functionals of Gaussian processes. Results of an empirical investigation of the marginal predictive content of money for income are also presented.

The second essay outlines a number of approaches to the selection of factor proxies (observed variables that proxy unobserved estimated factors) using statistics based on large sample datasets. This approach to factor proxy selection is examined via a small Monte Carlo experiment and a set of prediction experiments, where evidence supporting our proposed methodology is presented.

The third essay compares the predictive content of a set of macroeconomic indicators with that of various other observable variables that act as proxies to factors constructed using diffusion index methodology. The analysis suggests that certain spreads constructed as the difference between short or long term debt instruments and the federal funds rate are found to be useful indicators. Surprisingly, traditional spreads, such as the yield curve slope and the reverse yield gap are not found to provide additional predictive power.

# Acknowledgements

I would like to express my sincere gratitude to my advisor, Norman Swanson, for his dedicated guidance and financial support throughout my research career. Norm created an enabling environment for me to complete my essays in a timely manner. He was extremely helpful in the successful completion of my graduate career.

Many thanks are owed to Roger Klein for his patience and for assiduously explaining to me the basic concepts in Econometrics. At a crucial point in my graduate career, Roger made me believe in myself again.

I also would like to thank Serena Ng, Michael McCracken, Valentina Corradi, Mark Wohar, Todd Clark, Greg Tkacz, John Landon-Lane and Bruce Mizrach for helpful comments as well as insightful discussions at various stages of my dissertation. A great many thanks to Mark Watson for making the data available for public consumption. Special thanks to Dorothy Rinaldi for actively managing my graduate career especially with respect to deadlines and requirements. She really saw to my welfare.

For their help in various capacities, I would like to thank Geetesh Bhardwaj, Yuri Zaderman, Colin Campbell, Solomon Dzakuma and Dina Essien.

Finally, I would like to thank my parents for their immense sacrifices and for believing in me. To my brother and sister, thank you for your unwavering support.

# Dedication

To Mummy and Daddy

# Table of Contents

**Abstract** . . . . . ii

**Acknowledgements** . . . . . iii

**Dedication** . . . . . iv

**List of Tables** . . . . . viii

**List of Figures** . . . . . x

**1. Introduction** . . . . . 1

**2. Predictive Inference Under Model Misspecification with an Application to Assessing the Marginal Predictive Content of Money for Output** . . . . 4

2.1. Introduction . . . . . 4

2.2. Block Bootstraps for Recursive and Rolling $m-$Estimators . . . . . 9

2.2.1. Recursive Estimation Window: . . . . . 9

2.2.2. Rolling Estimation Window: . . . . . 13

2.3. The CS Test . . . . . 14

2.4. Monte Carlo Experiments . . . . . 19

2.5. Empirical Illustration: The Marginal Predictive Content of Money for Output 25

2.6. Concluding Remarks . . . . . 31

**3. Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Largescale Macroeconomic Time Series Environments** . . . . . 40

3.1. Introduction . . . . . 40

3.2. Review: Diffusion Index Models and the Principle Components Approach to Estimation . . . 43
3.2.1. The diffusion index model . . . 43
3.2.2. Common factor estimation using principal components . . . 46
3.3. Using Proxies In Place of Factors for Prediction . . . 48
3.3.1. Prediction using factors . . . 48
3.3.2. Using the $A(j)$ and $M(j)$ tests of Bai and Ng (2006b) to uncover factor proxies . . . 50
3.3.3. Smoothed $A(j)$ and $M(j)$ tests for selecting factor proxies . . . 53
3.4. Empirical Methodology . . . 54
3.5. Data . . . 56
3.6. Monte Carlo Experiment . . . 57
3.7. Empirical Findings . . . 60
3.8. Recent Advances in the Construction of Diffusion Indices . . . 64
3.9. Concluding Remarks . . . 65

**4. Which Variables Should the Federal Reserve Monitor? New Diffusion Index Evidence** . . . 78
4.1. Introduction . . . 78
4.2. Using Macroeconomic Indicators, Factors, and Factor Proxies for Prediction 82
4.2.1. Prediction with Macroeconomic Indicators . . . 82
4.2.2. Prediction with Factors . . . 83
4.2.3. Prediction With Factor Proxies . . . 86
4.3. Predictive Content of Spreads . . . 89
4.4. Data . . . 93
4.5. Empirical Methodology . . . 95
4.6. Empirical Results . . . 98
4.7. Concluding Remarks . . . 103
**References** . . . 115

**Vita** . . . . . 123

# List of Tables


2.1. **Test Statistics, Sampling Scheme, and Data Generating Processes Used in Monte Carlo Experiments** . . . . . 33

2.2. **Recursive Estimation Scheme - Rejection Frequencies of CS Test with $T = 540$, $P = 0.5T$** . . . . . 34

2.3. **Rolling Estimation Scheme - Rejection Frequencies of CS Test with $T = 540$, $P = 0.5T$** . . . . . 35

2.4. **Recursive Estimation Scheme - Rejection Frequencies of Various Tests with $T = 540$, $P = 0.5T$** . . . . . 36

2.5. **Rolling Estimation Scheme - Rejection Frequencies of Various Tests with $T = 540$, $P = 0.5T$** . . . . . 37

2.6. **Mean Square Forecast Errors and the Marginal Predictive Content of $M2$ for Output** . . . . . 38

2.7. **Tests for the Marginal Predictive Content of $M2$ for Output** . . . 39

3.1. **Prediction Models Used in Empirical Experiments** . . . . . 67

3.2. **Monte Carlo Experiment Results** . . . . . 68

3.3. **Monte Carlo Experiment Descriptive Statistics** . . . . . 69

3.4. **Frequency of Selected Factor Proxies** . . . . . 70

3.5. **Predictive Performance of Various Models for Price Variables** . . . 71

3.6. **Predictive Performance of Various Models for Output, Employment and Sales Variables** . . . . . 73

4.1. **Predictors Used in Empirical Experiments** . . . . . 104

4.2. **Prediction Models** . . . . . 106

4.3. **Description of Prediction Models Used in Empirical Experiments** 107

4.4. **Forecast of CPI Inflation** . . . . . 109

4.5. **Forecast of Output Growth** . . . . . 111

# List of Figures

3.1. **Estimated Factors and Most Frequently Selected Factor Proxies** . 75

4.1. **Forecast of CPI Inflation by Benchmark Model 1 at the 12 Step Horizon** . . . . . 113

4.2. **Forecast of Output Growth by Benchmark Model 1 at the 1 Step Horizon** . . . . . 114

# Chapter 1

# Introduction

This dissertation considers the forecasting performance of various macroeconomic time series models. The second chapter discusses model selection and predictive accuracy tests in the context of parameter and model uncertainty under recursive and rolling estimation schemes. The chapter begins by summarizing some recent theoretical findings, with particular emphasis on the construction of valid bootstrap procedures for calculating the impact of parameter estimation error on the class of test statistics with limiting distributions that are functionals of Gaussian processes with covarinace kernels that are dependent upon parameter and model uncertainty. An example of a particular test which falls in this class is given by outlining the so-called Corradi and Swanson (CS: 2002) test of (non)linear out-of-sample Granger causality. Thereafter, a series of Monte Carlo experiments examining the properties of the CS and a variety of other related predictive accuracy and model selection type tests is given. Finally, the results of an empirical investigation of the marginal predictive content of money for income, in the spirit of Stock and Watson (1989), Swanson (1998), Amato and Swanson (2001), and the references cited therein are presented. There is evidence of predictive causation when in-sample estimation periods are ended any time during the 1980s, but less evidence during the 1970s. Furthermore, recursive estimation windows yield better prediction models when prediction periods begin in the 1980s, while rolling estimation windows yield better models when prediction periods begin during the 1970s and 1990s. Interestingly, these two results can be combined into a coherent picture of what is driving the empirical results. Namely, when recursive estimation windows yield lower overall predictive MSEs, then bigger prediction models that include money are preferred, while smaller models without money are preferred when rolling window models yield the lowest MSE predictors.

In the third chapter, common factors are often assumed to underlie the co-movements of a set of macroeconomic variables. For this reason, many authors have used estimated factors in the construction of prediction models. This chapter outlines a number of approaches to the selection of factor proxies (observed variables that proxy unobserved estimated factors) using statistics developed in Stock and Watson (2002a,b) and Bai and Ng (2006a,b). This approach to factor proxy selection is examined via a small Monte Carlo experiment, where evidence supporting the proposed methodology is presented, and via a large set of prediction experiments using the panel dataset of Stock and Watson (2005). One of the main empirical findings is that the "smoothed" approaches to factor proxy selection appear to yield predictions that are often superior not only to a benchmark factor model, but also to simple linear time series models which are generally difficult to beat in forecasting competitions. In some sense, by using the proposed approach to predictive factor proxy selection, one is able to open up the "black box" often associated with factor analysis, and to identify actual variables that can serve as primitive building blocks for (prediction) models of macroeconomic variables, and can also serve as policy instruments, for example. With regard to forecasting price variables such as CPI and PPI, autoregressive models with exogenous variables (ARX), where the exogenous variables are based on smoothed versions of the A(j) and M(j) tests, often outperform all other considered models at the 1, 3 and 12 month ahead horizons. However, the basic factor model outperforms all other alternative models at the 24-month ahead horizon. These findings suggest that important observable variables include various SP500 stock price indices and dividend series; a 1-year Treasury bond rate; housing activity variables; industrial production; and exchange rates.

In the fourth chapter, the Federal Reserve regularly monitors select financial and macroeconomic variables in order to obtain early indication of the impact of current monetary policy. This practice is discussed on the Federal Reserve Bank of New York website, where one particular set of macroeconomic indicators is given. As a measure of the "adequacy" of these particular "macroeconomic indicators", the chapter compares their predictive content with that of various other observable variables that act as proxies to factors constructed using the Stock and Watson (2002a,b) diffusion index methodology. More specifically, observable proxies for factors obtained from "diffusion index" analysis of a large scale macroeconomic

and financial dataset are constructed via application of the methodology recently introduced in Bai and Ng (2006a,b) and further developed by Armah and Swanson. Interestingly, the "macroeconomic indicators" are very similar to the observable variables that proxy the factors. The findings, thus, lend credence to the macroeconomic indicators used to monitor monetary policy. In addition, the analysis suggests that certain "spreads" are useful indicators. The particular spreads found to be useful are the difference between short or long term debt instruments and the federal funds rate. Surprisingly, traditional spreads, such as the yield curve slope and the reverse yield gap are not found to provide additional predictive power. More specifically, "spread augmented" models, which are by construction less parsimonious than those not containing spreads, yield improved inflation and output growth predictions for a variety of models and forecast horizons, based on mean square forecast error comparisons; and in particular, the macroeconomic indicators perform best when forecasting inflation in non-volatile time periods. On the contrary, the forecast performance of the indicators can be improved by including spreads when forecasting inflation in times of high volatility.

# Chapter 2

# Predictive Inference Under Model Misspecification with an Application to Assessing the Marginal Predictive Content of Money for Output

## 2.1 Introduction

In a series of recent papers, Chao *et al* (2001) and Corradi and Swanson (2002, 2004, 2006a, 2007) discuss model selection and predictive accuracy tests in the context of parameter and model uncertainty under recursive and rolling estimation schemes. In this chapter, we begin by summarizing some of the theoretical findings of these papers, with particular emphasis on the construction of valid bootstrap procedures for calculating the impact of parameter estimation error on the class of test statistics with limiting distributions that are functionals of Gaussian processes with covariance kernels that are dependent upon parameter and model uncertainty. We then provide an example of a particular test which falls in this class. Namely, we outline the so-called Corradi and Swanson (CS: 2002) test of (non)linear out-of-sample Granger causality. Thereafter, we carry out a series of Monte Carlo experiments examining the properties of the CS and a variety of other related predictive accuracy and model selection type tests, including the Deibold and Mariano (DM: 1995) and West (1996) predictive accuracy test as well as the encompassing test of Clark and McCracken (CM: 2004). This is done for both recursive and rolling window estimators, hence shedding light on the finite sample impact of using shorter rolling windows rather than recursive windows. Finally, we present the results of an empirical investigation of the marginal predictive content of money for income, in the spirit of Stock and Watson (1989), Swanson (1998), Amato and Swanson (2001), and the references cited therein. The empirical results shed new light on the importance of sample periods and estimation schemes when carrying out empirical investigations.

The main link between this chapter and the overall theme of the book is that we address the issue of model uncertainty. In particular, the tests discussed herein *do not* assume correct specification under either the null or the alternative hypothesis being tested. This is a crucial assumption to have if one believes that all models are approximations of some underlying *true* DGP. Of course, if one does not believe that all models should be viewed as approximations, then there is perhaps really no obvious need to carry out ex ante inference using forecasts (assuming no structural breaks). After all, under the assumption of correct specification under the null, why not simply carry out in-sample inference, for the sake of efficiency? Our approach differs from approaches used in many (perhaps most) currently popular prediction tests, where correct specification is assumed under the null. As a case in point, consider the predictive density testing framework discussed by the important paper of Diebold, Gunther and Tay (DGT: 1998) and in Corradi and Swanson (2006a,b,c). In their paper, DGT use the probability integral transform (see e.g. Rosenblatt (1952)) to show that $F_t(y_t|\Im_{t-1}, \theta_0)$, is identically and independently distributed as a uniform random variable on $[0,1]$, where $F_t(\cdot|\Im_{t-1}, \theta_0)$ is a parametric distribution with underlying parameter $\theta_0$, $y_t$ is the random variable of interest, and $\Im_{t-1}$ is the information set containing all "relevant" past information (see below for further discussion). They thus suggest using the difference between the empirical distribution of $F_t(y_t|\Im_{t-1}, \widehat{\theta}_T)$ and the $45^\circ-$degree line as a measure of "goodness of fit", where $\widehat{\theta}_T$ is some estimator of $\theta_0$. This approach has been shown to be very useful for financial risk management (see e.g. Diebold, Hahn and Tay (1998)), as well as for macroeconomic forecasting (see e.g. Diebold, Tay and Wallis (1998) and Clements and Smith (2000, 2002)). Likewise, Bai (2003) proposes a Kolmogorov type test of $F_t(u|\Im_{t-1}, \theta_0)$ based on the comparison of $F_t(y_t|\Im_{t-1}, \widehat{\theta}_T)$ with the CDF of a uniform on $[0,1]$. As a consequence of using estimated parameters, the limiting distribution of his test reflects the contribution of parameter estimation error and is not nuisance parameter free. To overcome this problem, Bai (2003) uses a novel approach based on a martingalization argument to construct a modified Kolmogorov test which has a nuisance parameter free limiting distribution. This test has power against violations of uniformity but not against violations of independence. Now, Corradi and Swanson (2006b), allow for (dynamic) misspecification under the null hypothesis, while the others mentioned above

do not. This feature allows them to obtain asymptotically valid critical values even when the conditioning information set does not contain all of the relevant past history. More precisely, if one is interested in testing for correct specification, given a particular information set which may or may not contain all of the relevant past information, then the Corradi-Swanson approach is preferrable. This is relevant when a Kolmogorov test is constructed, for example, as one is generally faced with the problem of defining $\Im_{t-1}$. If enough history is not included, then there may be dynamic misspecification. Additionally, finding out how much information (e.g. how many lags) to include may involve pre-testing, hence leading to a form of sequential test bias. By allowing for dynamic misspecification, one does not require such pre-testing. Another key feature of the Corradi-Swanson approach concerns the fact that the limiting distribution of Kolmogorov type tests is affected by dynamic misspecification. Critical values derived under correct specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$. Consider the following example. Assume that we are interested in testing whether the conditional distribution of $y_t|y_{t-1}$ is $N(\alpha_1^\dagger y_{t-1}, \sigma_1)$. Suppose also that in actual fact the "relevant" information set has $\Im_{t-1}$ including both $y_{t-1}$ and $y_{t-2}$, so that the true conditional model is $y_t|\Im_{t-1} = y_t|y_{t-1}, y_{t-2} = N(\alpha_1 y_{t-1} + \alpha_2 y_{t-2}, \sigma_2)$, where $\alpha_1^\dagger$ differs from $\alpha_1$. In this case, we have correct specification with respect to the information contained in $y_{t-1}$; but we have dynamic misspecification with respect to $y_{t-1}$, $y_{t-2}$. Even without taking account of parameter estimation error, the critical values obtained assuming correct dynamic specification are invalid, thus leading to invalid inference. Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under $H_0$) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is *not* false. In summary, if one is interested in the particular problem of testing for correct specification for a given information set, then the Corradi-Swanson approach is appropriate. In general, these sorts of arguments apply to all varieties of prediction based testing, such as that discussed in this chapter.[1]

Parameter estimation error is a crucial component of model selection and predictive

[1]Note that we do not address structural breaks directly, although lack of knowledge of structural breaks when specifying a model can clearly lead to misspecification under both hypotheses. This is one reason why rolling windows are sometimes used in predictive contexts.

accuracy tests that is often overlooked, or more precisely is often assumed away by making the assumption that the in-sample estimation period grows more quickly than the out-of-sample predictive evaluation period. However, in some circumstances, such as when constructing DM tests for equal (pointwise) predictive accuracy of two models, limiting distributions are normal random variables, and parameter estimation error can be accounted for using the framework of West (1996). In other circumstances, such as when constructing tests which have power against generic alternatives (e.g. the CS test), statistics have limiting distributions that can be shown to be functionals of Gaussian processes with covariance kernels that reflect both (dynamic) misspecification as well as the contribution of parameter estimation error. Such limiting distributions are not nuisance parameter free, and critical values cannot be tabulated. Nevertheless, valid asymptotic critical values can be obtained via use of a bootstrap procedure that allows for the formulation of statistics which properly mimic the contribution of parameter estimation error. In the first part of the chapter we summarize block bootstrap procedures which are valid for recursive and rolling $m$-estimators (see e.g. Corradi and Swanson (2006a, 2007)).

In the second part of the chapter we review the so-called CS test, which is an out-of-sample version of the integrated conditional moment (ICM) test of Bierens (1982, 1990) and Bierens and Ploberger (1997), and which yields out-of-sample tests that are consistent against generic (nonlinear) alternatives (see Corradi and Swanson (2002, 2007) and Swanson and White (1997)). The CS test can alternatively be viewed as a consistent specification test, in the spirit of Bierens, or as a nonlinear Granger causality test, as discussed in Chao *et al.* (2001). Note, however, that the CS test differs from the ICM test developed by Bierens (1982, 1990) and Bierens and Ploberger (1997) because parameters are estimated in either recursive or rolling fashion, the test is of the out-of-sample variety, and the null hypothesis is that the reference model delivers the best "loss function specific" predictor, for a given information set. Furthermore, the CS test allows for model misspecification under both hypotheses (see Corradi and Swanson (2006b)).

In order to provide evidence on the usefulness of the bootstrap methods discussed above, and in particular in order to compare bootstraps based on recursive and rolling estimators, we carry out a Monte Carlo investigation that compares the finite sample properties of

our block bootstrap procedures with two alternative naive block bootstraps, all within the context of the CS test and a simpler non-generic version of the CS test due to Chao, Corradi and Swanson (CCS: 2001). In addition, various other related tests, including the standard F-test, the DM test and the CM test are included in the experiments. Results support the finding of Corradi and Swanson (2007) that the recursive block bootstrap outperforms alternative naive nonparametric block bootstraps. Additionally, we find that the rolling version of the bootstrap also outperforms the naive alternatives, Finally, we find that the finite sample properties of the other tests vary to some degree. Of note is that the Kilian (1999) bootstrap is a viable alternative to ours, although theoretical assessment thereof remains to be done (see Corradi and Swanson (2007) for further discussion).

In the last part of the chapter, an empirical illustration is presented, in which it is found that results concerning the (non)linear marginal predictive content for money for output are not only sample dependent, but also vary to some limited degree depending upon whether recursive or rolling estimation windows are used. In particular, there is evidence of predictive causation when in-sample estimation periods are ended any time during the 1980s, but little evidence of causality otherwise. Furthermore, recursive estimation windows yield better models when prediction periods begin in the 1980s, while rolling estimation windows yield better models when prediction periods begin during the 1970s and 1990s. Interestingly, these two results can be combined into a coherent picture of what is driving our empirical results. Namely, when recursive estimation windows yield lower overall predictive MSEs, then bigger prediction models that include money are preferred, while smaller models without money are preferred when rolling models yield the lowest MSE predictors.

Hereafter, $P^*$ denotes the probability law governing the resampled series, conditional on the sample, $E^*$ and $Var^*$ are the mean and variance operators associated with $P^*$, $o_P^*(1)$ $\Pr -P$ denotes a term converging to zero in $P^*-$probability, conditional on the sample, and for all samples except a subset with probability measure approaching zero, and $O_P^*(1)$ $\Pr -P$ denotes a term which is bounded in $P^*-$probability, conditional on the sample, and for all samples except a subset with probability measure approaching zero. Analogously, $O_{a.s.^*}(1)$ and $o_{a.s.^*}(1)$ denote terms that are almost surely bounded and terms that approach zero almost surely, according the the probability law $P^*$ and conditional on the sample. Note

that $P$ is also used to denote the length of the prediction period, and unless otherwise obvious from the context in which it is used, clarification of the meaning is given.

## 2.2 Block Bootstraps for Recursive and Rolling $m-$Estimators

In this section, we draw largely from Corradi and Swanson (2006a, 2007).

### 2.2.1 Recursive Estimation Window:

Define the block bootstrap estimator that captures the effect of parameter estimation error in the context of *recursive* $m$-estimators, as follows. Let $Z^t = (y_t, ..., y_{t-s_1+1}, X_t, ..., X_{t-s_2+1})$, $t = 1, ..., T$, and let $s = \max\{s_1, s_2\}$. Additionally, assume that $i = 1, ..., n$ models are estimated (thus allowing us to establish notation that will be useful in the applications presented in subsequent sections). Now, define the *recursive* $m$-estimator for the parameter vector associated with model $i$ as:[2]

$$\widehat{\theta}_{i,t} = \arg\min_{\theta_i \in \Theta_i} \frac{1}{t} \sum_{j=s}^{t} q_i(y_j, Z^{j-1}, \theta_i), \quad R \leq t \leq T-1, i = 1, ..., n \tag{2.1}$$

Further, define

$$\theta_i^{\dagger} = \arg\min_{\theta_i \in \Theta_i} E(q_i(y_j, Z^{j-1}, \theta_i)), \tag{2.2}$$

where $q_i$ denotes the objective function for model $i$. As the discussion below does not depend on any specific model, we drop the subscript $i$. Following standard practice (such as in the real-time forecasting literature), this estimator is first computed using $R$ observations. In our applications we focus on 1-step ahead prediction (although results can be extended quite easily to multiple step ahead prediction), so that recursive estimators are thus subsequently computed using $R+1$ observations, and then $R+2$ observations, and so on, until the last estimator is constructed using $T-1$ observations. This results in a sequence of $P = T - R$ estimators. These estimators can then be used to construct sequences of $P$ 1-step ahead forecasts and associated forecast errors, for example.

The overlapping block resampling scheme of Künsch (1989) involves drawing $b$ blocks (with replacement) of length $l$ from the sample $W_t = (y_t, Z^{t-1})$, where $bl = T - s$, at each

[2]Within the context of full sample estimation, the first order validity of the block bootstrap for $m-$estimators has been shown by Goncalves and White (2004) for dependent and heterogeneous series.

replication. Thus, the first block is equal to $W_{i+1}, ..., W_{i+l}$, for some $i = s-1, ..., T-l+1$, with probability $1/(T-s-l+1)$, the second block is equal to $W_{i+1}, ..., W_{i+l}$, again for some $i = s-1, ..., T-l+1$, with probability $1/(T-s-l+1)$, and so on, for all blocks, where the block length grows with the sample size at an appropriate rate. More formally, let $I_k$, $k = 1, ..., b$ be *iid* discrete uniform random variables on $[s-1, s, ..., T-l+1]$. Then, the resampled series, $W_t^* = (y_t^*, Z^{*,t-1})$, is such that $W_1^*, W_2^*, ..., W_l^*, W_{l+1}^*, ..., W_T^* = W_{I_1+1}, W_{I_1+2}, ..., W_{I_1+l}, W_{I_2+1}, ..., W_{I_b+l}$, and so a resampled series consists of $b$ blocks that are discrete *iid* uniform random variables, conditional on the sample.

Suppose we define the bootstrap estimator, $\widehat{\theta}_t^*$, to be the direct analog of $\widehat{\theta}_t$. Namely,

$$\widehat{\theta}_t^* = \arg\min_{\theta\in\Theta} \frac{1}{t}\sum_{j=s}^{t} q(y_j^*, Z^{*,j-1}, \theta), R \leq t \leq T-1. \tag{2.3}$$

By first order conditions, $\frac{1}{t}\sum_{j=s}^{t} \nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_t^*) = 0$, where $\nabla_\theta$ denotes the denotes the derivative with respect to $\theta$. Via a mean value expansion of $\frac{1}{t}\sum_{j=s}^{t} \nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_t^*)$ around $\widehat{\theta}_t$, after a few simple manipulations, we have that

$$\begin{aligned} & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t^* - \widehat{\theta}_t\right) \\ = \; & B^\dagger \frac{a_{R,0}}{\sqrt{P}}\sum_{j=s}^{R} \nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_R) + B^\dagger \frac{1}{\sqrt{P}}\sum_{j=1}^{P-1} a_{R,j}\nabla_\theta q(y_{R+j}^*, Z^{*,R+j-1}, \widehat{\theta}_{R+j}) \\ & + o_{P^*}(1) \Pr -P, \end{aligned} \tag{2.4}$$

where $B^\dagger = E\left(-\nabla_\theta^2 q(y_j, Z^{j-1}, \theta^\dagger)\right)^{-1}$, $a_{R,j} = \frac{1}{R+j} + \frac{1}{R+j+1} + ... + \frac{1}{R+P-1}$, $j = 0, 1, ..., P-1$, and where the last equality on the right hand side of (2.4) follows immediately, using the same arguments as those used in Lemma A5 of West (1996). Analogously,

$$\begin{aligned} & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^\dagger\right) \\ = \; & B^\dagger \frac{a_{R,0}}{\sqrt{P}}\sum_{j=s}^{R} \nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger) + B^\dagger \frac{1}{\sqrt{P}}\sum_{j=1}^{P-1} a_{R,j}\nabla_\theta q(y_{R+j}, Z^{R+j-1}, \theta^\dagger) + o_P(1) \end{aligned} \tag{2.5}$$

Now, given (2.2), $E\left(\nabla_\theta q(y_j, Z^{j-1}, \theta^\dagger)\right) = 0$ for all $j$, and $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^\dagger\right)$ has a zero mean normal limiting distribution (see Theorem 4.1 in West (1996)). On the other hand, as any block of observations has the same chance of being drawn,

$$E^*\left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_t)\right) = \frac{1}{T-s}\sum_{k=s}^{T-1} \nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t) + O\left(\frac{l}{T}\right) \Pr -P, \tag{2.6}$$

where the $O\left(\frac{l}{T}\right)$ term arises because the first and last $l$ observations have a lesser chance of being drawn (see e.g. Fitzenberger (1997)). Now, $\frac{1}{T-s}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t) \neq 0$, and is instead of order $O_P\left(T^{-1/2}\right)$. Thus, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{T-s}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t) = O_P(1)$, and does not vanish in probability. This clearly contrasts with the full sample case, in which $\frac{1}{T-s}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_T) = 0$, because of the first order conditions. Thus, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t^* - \widehat{\theta}_t\right)$ cannot have a zero mean normal limiting distribution, but is instead characterized by a location bias that can be either positive or negative depending on the sample.

Given (2.6), our objective is thus to have the bootstrap score centered around $\frac{1}{T-s}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)$. Hence, define a new bootstrap estimator, $\widetilde{\theta}_t^*$, as:

$$\widetilde{\theta}_t^* = \arg\min_{\theta\in\Theta}\frac{1}{t}\sum_{j=s}^{t}\left(q(y_j^*, Z^{*,j-1}, \theta) - \theta'\left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)\right)\right), \tag{2.7}$$

$R \leq t \leq T-1$.[3]

Now, note that first order conditions are $\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \widetilde{\theta}_t^*) - \left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)\right)\right) = 0$; and via a mean value expansion of $\frac{1}{t}\sum_{j=s}^{t}\nabla_\theta q(y_j^*, Z^{*,j-1}, \widetilde{\theta}_t^*)$ around $\widehat{\theta}_t$, after a few simple manipulations, we have that:

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widetilde{\theta}_t^* - \widehat{\theta}_t\right) \\
= \;& B^\dagger\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\frac{1}{t}\sum_{j=s}^{t}\left(\nabla_\theta q(y_j^*, Z^{*,j-1}, \widehat{\theta}_t) - \left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)\right)\right)\right) \\
& + o_{P^*}(1), \Pr - P.
\end{aligned}$$

Thus, given (2.6), it is immediate to see that the bias associated with $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widetilde{\theta}_t^* - \widehat{\theta}_t\right)$ is of order $O\left(lT^{-1/2}\right)$, conditional on the sample, and so it is negligible for first order asymptotics, as $l = o(T^{1/2})$.

Theorem 1, which summarizes these results, requires the following assumptions.

**Assumption A1:** $(y_t, X_t)$, with $y_t$ scalar and $X_t$ an $\Re^\zeta-$valued $(0 < \zeta < \infty)$ vector, is a strictly stationary and absolutely regular $\beta-$mixing process with size $-4(4+\psi)/\psi$, $\psi > 0$.

[3] More precisely, we should use $\frac{1}{t-s}$ and $\frac{1}{T-s}$ to scale the summand in (7). For notational simplicity, $\frac{1}{t-s}$ and $\frac{1}{T-s}$ are approximated with $\frac{1}{t}$ and $\frac{1}{T}$.

**Assumption A2:** (i) $\theta^{\dagger}$ is uniquely identified (i.e. $E(q(y_t, Z^{t-1}, \theta)) > E(q(y_t, Z^{t-1}, \theta^{\dagger}))$ for any $\theta \neq \theta^{\dagger}$); (ii) $q$ is twice continuously differentiable on the interior of $\Theta$, and for $\Theta$ a compact subset of $\Re^{\varrho}$; (iii) the elements of $\nabla_{\theta} q$ and $\nabla^2_{\theta} q$ are $p-$dominated on $\Theta$, with $p > 2(2+\psi)$, where $\psi$ is the same positive constant as defined in Assumption A1; and (iv) $E\left(-\nabla^2_{\theta} q(\theta)\right)$ is negative definite uniformly on $\Theta$.[4]

**Assumption A3:** $T = R + P$, and as $T \to \infty$, $P/R \to \pi$, with $0 < \pi < \infty$.

Assumptions A1 and A2 are standard memory, moment, smoothness and identifiability conditions. A1 requires $(y_t, X_t)$ to be strictly stationary and absolutely regular. The memory condition is stronger than $\alpha-$mixing, but weaker than (uniform) $\phi-$mixing. Assumption A3 requires that $R$ and $P$ grow at the same rate. In fact, if $P$ grows at a slower rate than $R$, i.e. $P/R \to 0$, then $\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widehat{\theta}_t - \theta^{\dagger}\right) = o_P(1)$ and so there were no need to capture the contribution of parameter estimation error.

**Theorem 1 (Corradi and Swanson (2007)):** Under recursive estimation, let A1-A3 hold. Also, assume that as $T \to \infty$, $l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as $T, P$ and $R \to \infty$,

$$P\left(\omega : \sup_{v \in \Re^{\varrho}} \left| P^*_T \left( \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widetilde{\theta}^*_t - \widehat{\theta}_t\right) \leq v \right) - P\left( \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widehat{\theta}_t - \theta^{\dagger}\right) \leq v \right) \right| > \varepsilon \right) \to 0,$$

where $P^*_T$ denotes the probability law of the resampled series, conditional on the (entire) sample.

Theorem 1 states that $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widetilde{\theta}^*_t - \widehat{\theta}_t\right)$ has the same limiting distribution as $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_t - \theta^{\dagger}\right)$, conditional on sample, and for all samples except a set with probability measure approaching zero. Of note is that if Assumption 3 is violated and $P/R \to 0$, then the statement in the theorem above is trivially satisfied, in the sense that both $\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widetilde{\theta}^*_t - \widehat{\theta}_t\right)$ and $\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widehat{\theta}_t - \theta^{\dagger}\right)$ have a limiting distribution degenerate on zero. Hence, the crucial impact of allowing for non-vanishing parameter estimation error is quite apparent.

---

[4] We say that $\nabla_{\theta} q(y_t, Z^{t-1}, \theta)$ is $2r-$dominated on $\Theta$ if its $j-th$ element, $j = 1, ..., \varrho$, is such that $\left|\nabla_{\theta} q(y_t, Z^{t-1}, \theta)\right|_j \leq D_t$, and $E(|D_t|^{2r}) < \infty$. For more details on domination conditions, see Gallant and White (1988, pp. 33).

### 2.2.2 Rolling Estimation Window:

In the rolling estimation scheme, one constructs a sequence of $P$ estimators using a rolling window of $R$ observations. The first estimator is constructed using the first $R$ observations, the second using observations from 2 to $R+1$, and so on, with the last estimator being constructed using observations from $T-R$ to $T-1$, so that we have a sequence of $P$ estimators, $(\widehat{\theta}_{R,R}, \widehat{\theta}_{R+1,R}, \ldots, \widehat{\theta}_{R+P-1,R})$. In general, it is common to assume that $P$ and $R$ grow as $T$ grows. Giacomini and White (2003) propose using a rolling scheme with a fixed window that does not increase with the sample size, so that estimated parameters are treated as mixing variables. Pesaran and Timmerman (2004a,b) suggest rules for choosing the window of observations in order to take into account possible structure breaks.

Using the same notation as in the recursive case, but noting that we are now constructing a rolling estimator, define

$$\widehat{\theta}_{i,t} = \arg\min_{\theta_i \in \Theta_i} \frac{1}{R} \sum_{j=t-R+1}^{t} q_i(y_j, Z^{j-1}, \theta_i), \quad R \leq t \leq T-1, i = 1, ..., n$$

In the case of in-sample model evaluation, the contribution of parameter estimation error is summarized by the limiting distribution of $\sqrt{T}(\widehat{\theta}_T - \theta^\dagger)$, where $\theta^\dagger$ is the probability limit of $\widehat{\theta}_T$. In the case of rolling estimation schemes, the contribution of parameter estimation error is summarized by the limiting distribution of $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_t - \theta^\dagger\right)$. Under mild conditions, because of the central limit theorem, $\left(\widehat{\theta}_t - \theta^\dagger\right)$is $O_P(R^{-1/2})$. Thus, if $P$ grows at a slower rate than $R$ (i.e. if $P/R \to 0$, as $T \to \infty$), then $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\widehat{\theta}_t - \theta^\dagger\right)$ is asymptotically negligible. In other words, if the in-sample portion of the data used for estimation is "much larger" than the out-of-sample portion of the data to be used for predictive accuracy testing and generally for model evaluation, then the contribution of parameter estimation error is asymptotically negligible.

In the rolling estimation scheme, observations in the middle are used more frequently than observations at either the beginning or the end of the sample. As in the recursive case, this introduces a location bias to the usual block bootstrap, as under standard resampling with replacement, any block from the original sample has the same probability of being selected. Also, the bias term varies across samples and can be either positive or negative, depending on the specific sample. Our objective is thus to properly recenter the objective

function in order to obtain a bootstrap rolling estimator, say $\widetilde{\theta}_t^*$, such that $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\widetilde{\theta}_t^* - \widehat{\theta}_t)$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\theta}_t - \theta^\dagger\right)$, conditionally on the sample. The approach and result are largely as outlined above. Namely, resample $b$ overlapping blocks of length $l$ from $W_t = (y_t, Z^{t-1})$, and form a bootstrap sample, as in the recursive case. Then, define the rolling bootstrap estimator as

$$\widetilde{\theta}_t^* = \arg\min_{\theta\in\Theta}\frac{1}{R}\sum_{j=t-R+1}^{t}\left(q(y_j^*, Z^{*,j-1}, \theta) - \theta'\left(\frac{1}{T}\sum_{k=s}^{T-1}\nabla_\theta q(y_k, Z^{k-1}, \widehat{\theta}_t)\right)\right).$$

As in the recursive case, the following theorem can be stated.

**Theorem 2 (Corradi and Swanson (2006a)):** Under rolling estimation, let Assumptions A1-A3 and A5 hold. Also, assume that as $T \to \infty$, $l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as $T, P$ and $R \to \infty$,

$$P\left(\omega : \sup_{v\in\Re^\varrho}\left|P_T^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widetilde{\theta}_t^* - \widehat{\theta}_{t,rol}\right)\le v\right) - P\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\theta}_t - \theta^\dagger\right)\le v\right)\right| > \varepsilon\right) \to 0.$$

## 2.3 The CS Test

As an example of the implementation of the recursive and rolling bootstrap discussed above, we summarize the CS test discussed in different forms in Chao *et al* (2001) as well as in Corradi and Swanson (2002, 2007). The test is presented in a framework that is directly applicable to the empirical investigation discussed in a subsequent section of the chapter.

As discussed in the introduction, the test draws on both the consistent specification and predictive ability testing literatures in order to propose a test for predictive accuracy which is consistent against generic nonlinear alternatives, which is designed for comparing nested models, and which allows for dynamic misspecification of all models being evaluated. The CS test is an out-of-sample version of the ICM test, as discussed in the introduction of this paper. Alternative (non DM) tests for comparing the predictive ability of a fixed number of nested models have previously also been suggested. For example, Clark and McCracken (2001, 2004) propose encompassing tests for comparing two nested models for one-step and multi-step ahead prediction, respectively. Giacomini and White (2003) introduce a test for conditional predictive ability that is valid for both nested and nonnested models. The key ingredient of their test is the fact that parameters are estimated using a fixed

rolling window. Finally, Inoue and Rossi (2004) suggest a recursive test, where not only the parameters, but the statistic itself, are computed in a recursive manner. One of the main differences between these tests and the CS test is that the CS test is consistent against generic (non)linear alternatives and not only against a fixed alternative.

The CS testing approach that will be used in the Monte Carlo and empirical sections of the chapter, assumes that the objective is to test whether there exists any unknown alternative model that has better predictive accuracy than a given benchmark model, for a given loss function. The benchmark model is:

$$y_t = \theta^\dagger_{1,1} + \theta^\dagger_{1,2} y_{t-1} + \theta^\dagger_{1,3} z_{t-1} + u_{1,t}, \tag{2.8}$$

where $\theta^\dagger_1 = (\theta^\dagger_{1,1}, \theta^\dagger_{1,2}, \theta^\dagger_{1,3})' = \arg\min_{\theta_1 \in \Theta_1} E(q_1(y_t - \theta_{1,1} - \theta_{1,2} y_{t-1} - \theta_{1,3} z_{t-1}))$, $\theta_1 = (\theta_{1,1}, \theta_{1,2}, \theta_{1,3})'$, $y_t$ is a scalar, and $q_1 = g$, as the same loss function is used both for in-sample estimation and out-of-sample predictive evaluation.[5] The generic alternative model is:

$$y_t = \theta^\dagger_{2,1}(\gamma) + \theta^\dagger_{2,2}(\gamma) y_{t-1} + \theta^\dagger_{2,3}(\gamma) z_{t-1} + \theta^\dagger_{2,4}(\gamma) w(Z^{t-1}, \gamma) + u_{2,t}(\gamma), \tag{2.9}$$

where $\theta^\dagger_2(\gamma) = (\theta^\dagger_{2,1}(\gamma), \theta^\dagger_{2,2}(\gamma), \theta^\dagger_{2,3}(\gamma), \theta^\dagger_{2,4}(\gamma))' = \arg\min_{\theta_2 \in \Theta_2} E(q_1(y_t - \theta_{2,1} - \theta_{2,2} y_{t-1} - \theta_{2,3} z_{t-1} - \theta_{2,4} w(Z^{t-1}, \gamma)))$, $\theta_2(\gamma) = (\theta_{2,1}(\gamma), \theta_{2,2}(\gamma), \theta_{2,3}(\gamma), \theta_{2,4}(\gamma))'$, $\theta_2 \in \Theta_2$, $\Gamma$ is a compact subset of $\Re^d$, for some finite $d$. The alternative model is called "generic" because of the presence of $w(Z^{t-1}, \gamma)$, which is a generically comprehensive function, such as Bierens' exponential, a logistic, or a cumulative distribution function (see e.g. Stinchcombe and White (1998) for a detailed explanation of generic comprehensiveness). One example has $w(Z^{t-1}, \gamma) = \exp(\sum_{i=1}^{s_2} \gamma_i \Phi(X_{t-i}))$, where $\Phi$ is a measurable one to one mapping from $\Re$ to a bounded subset of $\Re$, so that here $Z^t = (X_t, ..., X_{t-s_2+1})$, and we are thus testing for nonlinear Granger causality. In fact, the above setup can be described within the context of our empirical example in Section 2.5. Namely, in Section 2.5 we set $X_t$ is equal to a vector of two variables including money supply growth and a cointegration term connecting output, money and prices; $y_t$ is set equal to output growth; and $z_t$ is an interest rate spread.

[5]Note that $z_{t-1}$ as used in (2.8) differs from $Z^{t-1}$ used elsewhere in the chapter (see Section 5 for an empirical illustration where $z_{t-1}$ is defined.

Turning back to our current discussion, note that the hypotheses of interest are:

$$H_0 : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) = 0 versus H_A : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) > 0. \quad (2.10)$$

Clearly, the reference model is nested within the alternative model, and given the definitions of $\theta_1^\dagger$ and $\theta_2^\dagger(\gamma)$, the null model can never outperform the alternative.[6] For this reason, $H_0$ corresponds to equal predictive accuracy, while $H_A$ corresponds to the case where the alternative model outperforms the reference model, as long as the errors above are loss function specific forecast errors. As discussed in Corradi and Swanson (2002), we can restate $H_0$ and $H_A$ as:

$$H_0 : E(g'(u_{1,t+1})w(Z^t, \gamma)) = 0 versus H_A : E(g'(u_{1,t+1})w(Z^t, \gamma)) \neq 0, \quad (2.11)$$

for $\forall \gamma \in \Gamma$, except for a subset with zero Lebesgue measure. Finally, define the forecast error as $\widehat{u}_{1,t+1} = y_{t+1} - \begin{pmatrix} 1 & y_t & z_t \end{pmatrix} \widehat{\theta}_{1,t}$. The relevant test statistic is:

$$M_P = \int_\Gamma m_P(\gamma)^2 \phi(\gamma) d\gamma, \quad (2.12)$$

where

$$m_P(\gamma) = \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} g'(\widehat{u}_{1,t+1}) w(Z^t, \gamma), \quad (2.13)$$

and where $\int_\Gamma \phi(\gamma) d\gamma = 1$, $\phi(\gamma) \geq 0$, with $\phi(\gamma)$ absolutely continuous with respect to Lebesgue measure. Note also that " $'$ " denotes derivative with respect to the argument of the function. Elsewhere, we use "$\nabla_x$" to denote derivative with respect to $x$. In the sequel, we require the following assumptions.

**Assumption A4:** (i) $w$ is a bounded, twice continuously differentiable function on the interior of $\Gamma$ and $\nabla_\gamma w(Z^t, \gamma)$ is bounded uniformly in $\Gamma$; and (ii) $\nabla_\gamma \nabla_{\theta_1} q'_{1,t}(\theta_1) w(Z^{t-1}, \gamma)$ is continuous on $\Theta_1 \times \Gamma$, where $q'_{1,t}(\theta_1) = q'_1(y_t - \theta_{1,1} - \theta_{1,2} y_{t-1} - \theta_{1,3} z_{t-1})$, $\Gamma$ a compact subset of $\Re^d$, and is $2r-$dominated uniformly in $\Theta_1 \times \Gamma$, with $r \geq 2(2 + \psi)$, where $\psi$ is the same positive constant as that defined in Assumption A1.

Assumption A5 requires the function $w$ to be bounded and twice continuously differentiable; such a requirement is satisfied by logistic or exponential functions, for example.

---

[6] Needless to say, in finite samples the forecasting mean square prediction error from the small model can be lower than that associated with the larger model.

**Theorem 3 (Corradi and Swanson (2007)):** Under either recursive or rolling estimation, let Assumptions A1-A4 hold. Then, the following results hold: (i) Under $H_0$,

$$M_P = \int_\Gamma m_P(\gamma)^2 \phi(\gamma) d\gamma d \rightarrow \int_\Gamma Z(\gamma)^2 \phi(\gamma) d\gamma,$$

where $m_P(\gamma)$ is defined in equation (2.13) and $Z$ is a Gaussian process with covariance kernel given by:

$$\begin{aligned} K(\gamma_1, \gamma_2) &= S_{gg}(\gamma_1, \gamma_2) + 2\Pi \mu'_{\gamma_1} B^\dagger S_{hh} B^\dagger \mu_{\gamma_2} + \Pi \mu'_{\gamma_1} B^\dagger S_{gh}(\gamma_2) \\ &\quad + \Pi \mu'_{\gamma_2} B^\dagger S_{gh}(\gamma_1), \end{aligned}$$

with $\mu_{\gamma_1} = E(\nabla_{\theta_1}(g'_{t+1}(u_{1,t+1}) w(Z^t, \gamma_1)))$, $B^\dagger = (E(\nabla^2_{\theta_1} q_1(u_{1,t})))^{-1}$,
$S_{gg}(\gamma_1, \gamma_2) = \sum_{j=-\infty}^{\infty} E(g'(u_{1,s+1}) w(Z^s, \gamma_1) g'(u_{1,s+j+1}) w(Z^{s+j}, \gamma_2))$,
$S_{hh} = \sum_{j=-\infty}^{\infty} E(\nabla_{\theta_1} q_1(u_{1,s}) \nabla_{\theta_1} q_1(u_{1,s+j})')$,
$S_{gh}(\gamma_1) = \sum_{j=-\infty}^{\infty} E(g'(u_{1,s+1}) w(Z^s, \gamma_1) \nabla_{\theta_1} q_1(u_{1,s+j})')$, and $\gamma$, $\gamma_1$, and $\gamma_2$ are generic elements of $\Gamma$.
(ii) Under $H_A$, for $\varepsilon > 0$, $\lim_{P \rightarrow \infty} \Pr\left( \frac{1}{P} \int_\Gamma m_P(\gamma)^2 \phi(\gamma) d\gamma > \varepsilon \right) = 1$.

Clearly, the form of the covariance kernel depends upon whether recursive or rolling estimation is used (for further detailed discussion of these covariance kernels, the reader is referred to the appendices in Corradi and Swanson (2006a, 2007)). It is also clear that the limiting distribution under $H_0$ is a Gaussian process with a covariance kernel that reflects both the dependence structure of the data and the effect of parameter estimation error. Hence, critical values are data dependent and cannot be tabulated.

In order to implement this statistic using the block bootstrap for recursive or rolling $m$-estimators discussed above, we define:

$$\begin{aligned} \widetilde{\theta}^*_{1,t} &= (\widetilde{\theta}^*_{1,1,t}, \widetilde{\theta}^*_{1,2,t}, \widetilde{\theta}^*_{1,3,t})' = \arg \min_{\theta_1 \in \Theta_1} \frac{1}{t} \sum_{j=2}^{t} [q_1(y^*_j - \theta_{1,1} - \theta_{1,2} y^*_{j-1} - \theta_{1,3} z^*_{j-1}) \\ &\quad - \theta'_1 \frac{1}{T} \sum_{i=2}^{T-1} \nabla_\theta q_1(y_i - \widehat{\theta}_{1,1,t} - \widehat{\theta}_{1,2,t} y_{i-1} - \widehat{\theta}_{1,3,t} z_{i-1})] \end{aligned} \tag{2.14}$$

Also, define $\widetilde{u}^*_{1,t+1} = y^*_{t+1} - \begin{pmatrix} 1 & y^*_t & z^*_t \end{pmatrix} \widetilde{\theta}^*_{1,t}$. The bootstrap test statistic is:

$$M^*_P = \int_\Gamma m^*_P(\gamma)^2 \phi(\gamma) d\gamma,$$

where, recalling that $g = q_1$,

$$\begin{aligned} & m_P^*(\gamma) \\ = \; & \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left( g' \left( y_{t+1}^* - \begin{pmatrix} 1 & y_t^* & z_t^* \end{pmatrix} \widetilde{\theta}_{1,t}^* \right) w(Z^{*,t}, \gamma) \right) \\ & - \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left( \frac{1}{T} \sum_{i=2}^{T-1} g' \left( y_i - \begin{pmatrix} 1 & y_{i-1} & z_{i-1} \end{pmatrix} \widehat{\theta}_{1,t} \right) w(Z^{i-1}, \gamma) \right) \end{aligned} \tag{2.15}$$

The bootstrap statistic in (2.15) is characterized by the fact that the bootstrap (re-sampled) component is constructed only over the last $P$ observations, while the sample component is constructed over all $T$ observations. This differs from the usual approach that would involve calculating:

$$\begin{aligned} m_P^{**}(\gamma) = & \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left( g' \left( y_{t+1}^* - \begin{pmatrix} 1 & y_t^* & z_t^* \end{pmatrix} \widetilde{\theta}_{1,t}^* \right) w(Z^{*,t}, \gamma) \right) \\ & - \frac{1}{P^{1/2}} \sum_{t=R}^{T-1} \left( g' \left( y_{t+1} - \begin{pmatrix} 1 & y_t & z_t \end{pmatrix} \widehat{\theta}_{1,t} \right) w(Z^t, \gamma) \right) \end{aligned} \tag{2.16}$$

However, the recursive (rolling) feature of the parameter estimation error in the CS test in the current context ensures that for all samples except a set with probability measure approaching zero, $m_P^{**}(\gamma)$ does not have the same limiting distribution as $m_P(\gamma)$ (see Corradi and Swanson (2007) for further details).

**Theorem 4 (Corradi and Swanson (2007)):** Under either recursive or rolling estimation, let Assumptions A1-A3 and A5 hold. Also, assume that as $T \to \infty$, $l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as $T, P$ and $R \to \infty$,

$$P \left( \omega : \sup_{v \in \Re} \left| P_T^* \left( \int_\Gamma m_P^*(\gamma)^2 \phi(\gamma) d\gamma \leq v \right) - P \left( \int_\Gamma m_P^\mu(\gamma)^2 \phi(\gamma) d\gamma \leq v \right) \right| > \varepsilon \right) \to 0,$$

where $m_P^\mu(\gamma) = m_P(\gamma) - \sqrt{P} E \left( g'(u_{1,t+1}) w(Z^t, \gamma) \right)$.

The above result suggests proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic, $m_P^*(\gamma)$. Perform $B$ bootstrap replications ($B$ large) and compute the quantiles of the empirical distribution of the $B$ bootstrap statistics. Reject $H_0$, if $m_P(\gamma)$ is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject.

## 2.4 Monte Carlo Experiments

In this section we carry out a series of Monte Carlo experiments comparing the recursive and rolling block bootstrap with a variety of other bootstraps, and comparing the finite sample performance of the test discussed above with a variety of other tests. In addition to the fact that rolling as well as recursive estimators are used, the experiments in this section differ from those discussed in Corradi and Swanson because they estimate an AR(1) model as their benchmark model (i.e. the model used in size experiments), while our benchmark model includes an additional explanatory variable, $z_t$, which corresponds to the interest rate spread in our empirical implementation. Furthermore, they include in all models an omitted variable, which we do not use in our specifications. As shall be discussed below, it is in fact this omitted variable that drives much of the size distortion in Corradi and Swanson (2007) when comparing the F-test with various other tests.

With regard to the bootstrap, we consider 4 alternatives. Namely: (i) the "Recur Block Bootstrap", which is the block bootstrap for recursive $m$-estimators discussed above; (ii) the "Roll Block Bootstrap", which is also discussed above, (iii) the "Block Bootstrap, no PEE, no adjust", which is a strawman block bootstrap used for comparison purposes, where it is assumed that there is no parameter estimation error (PEE), so that $\widehat{\theta}_{1,t}$ is used in place of $\widetilde{\theta}^*_{1,t}$ in the construction of $M^*_P$, and the term $\frac{1}{T}\sum_{i=1}^{T-1} g'\left(y_{i+1} - \begin{pmatrix} 1 & y_i & z_i \end{pmatrix}\widehat{\theta}_{1,t}\right) w(Z^i,\gamma)$ in $m^*_P$ is replaced with $g'\left(y_{t+1} - \begin{pmatrix} 1 & y_t & z_t \end{pmatrix}\widehat{\theta}_{1,t}\right) w(Z^t,\gamma)$ (i.e. there is no bootstrap statistic adjustment, thus conforming with the usual case when the standard block bootstrap is used) and (iv) the "Standard Block Bootstrap", which is the standard block bootstrap (i.e. this bootstrap is the same as that outlined in (iii), except that $\widehat{\theta}_{1,t}$ is replaced with $\widehat{\theta}^*_{1,t}$.

As discussed in Section 2.3, the hypotheses of interest are:

$$H_0 : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) = 0 versus H_A : E(g(u_{1,t+1}) - g(u_{2,t+1}(\gamma))) > 0. \quad (2.17)$$

where $u_{1,t}$ and $u_{2,t}$ are out-of-sample 1-step ahead prediction errors of the following models:

$$y_t = \theta^\dagger_{1,1} + \theta^\dagger_{1,2} y_{t-1} + \theta^\dagger_{1,3} z_{t-1} + u_{1,t}, \quad (2.18)$$

$$y_t = \theta^\dagger_{2,1}(\gamma) + \theta^\dagger_{2,2}(\gamma) y_{t-1} + \theta^\dagger_{2,3}(\gamma) z_{t-1} + \theta^\dagger_{2,4}(\gamma) w(Z^{t-1},\gamma) + u_{2,t}(\gamma), \quad (2.19)$$

where $\theta_1^\dagger = (\theta_{1,1}^\dagger, \theta_{1,2}^\dagger, \theta_{1,3}^\dagger)'$, and $\theta_2^\dagger = (\theta_{2,1}^\dagger, \theta_{2,2}^\dagger, \theta_{3,3}^\dagger, \theta_{3,4}^\dagger)'$ are parameter vectors, where $z_{t-1}$is an additional explanatory variable in the "small" model, and where $Z^{t-1}$ in the "big" model includes the variable which is being tested for inclusion in the small model (denoted $x_t$ in Table 2.1).

The test statistics examined in our experiments include: (i) the standard in-sample F-test; (ii) the encompassing test due to Clark and McCracken (CM: 2004) and Harvey *et al* (1997); (iii) the Diebold and Mariano (DM: 1995) test; (iv) the CS test; and (v) the CCS test.[7] Of note in this context is that in the CS test we are implicitly testing whether any (non)linear function of $Z^{t-1}$ would be useful for constructing a better prediction model of $y_t$. Alternatively, the other tests only consider inclusion of a linear function of $Z^{t-1}$, so that they are essentially setting $w$ to be an affine function.

To be more specific, note that the CM test is an out-of-sample encompassing test, and is defined as follows:

$$CM = (P-h+1)^{1/2} \frac{\frac{1}{P-h+1}\sum_{t=R}^{T-h} \widehat{c}_{t+h}}{\sqrt{\frac{1}{P-h+1}\sum_{j=-\overline{j}}^{\overline{j}} \sum_{t=R+j}^{T-h} K\left(\frac{j}{M}\right)\left(\widehat{c}_{t+h}-\overline{c}\right)\left(\widehat{c}_{t+h-j}-\overline{c}\right)}},$$

where $\widehat{c}_{t+h} = \widehat{u}_{1,t+h}\left(\widehat{u}_{1,t+h} - \widehat{u}_{2,t+h}\right)$, $\overline{c} = \frac{1}{P-h+1}\sum_{t=R}^{T-h}\widehat{c}_{t+h}$, $K(\cdot)$ is a kernel (such as the Bartlett kernel), and $0 \leq K\left(\frac{j}{M}\right) \leq 1$, with $K(0)=1$, and $M = o(P^{1/2})$. Additionally, $h$ is the forecast horizon (set equal to unity in our experiments), $P$ is as defined above, and $\widehat{u}_{1,t+1}$ and $\widehat{u}_{2,t+1}$ are the out-of-sample forecast errors associated with least squares estimation of "smaller" and "bigger" linear models, respectively (see below for further details). Note that $\overline{j}$ does not grow with the sample size. Therefore, the denominator in $CM$ is a consistent estimator of the long-run variance only when $E\left(c_t c_{t+|k|}\right) = 0$ for all $|k| > h$ (see Assumption A3 in Clark and McCracken (2004)). Thus, the statistic takes into account the moving average structure of the multistep prediction errors, but still does not allow for dynamic

[7]The CCS statistic is essentially the same as the CS test, but uses $Z^t$ instead of a generically comprehensive function thereof (recall that $Z^t$ contains the additional variables included in the "big" model defined below). Thus, this test can be seen as a special case of the CS test that is designed to have power against linear alternatives, and it is not explicitely designed to have power against generic nonlinear alternatives as is the CS test. The theory in Section 3 of this paper thus applies to both the CS and CCS tests. Additionally, the CM test is included in our study because it is an encompassing test which is designed to have power against linear alternatives, and so it is directly comparable with the CCS test. Finally, the F and DM tests are included in our analysis because they are the most commonly applied and examined in- and out-of-sample tests used for model selection. They thus serve as a kind of benchmark against which the performance of the other tests can be measured.

misspecification under the null. This is one of the main differences between the CM and CS (CCS) tests.

Note also that the DM test is the mean square error version of the Diebold and Mariano (1995) test for predictive accuracy, and is defined as follows:

$$DM = (P-h+1)^{1/2} \frac{\frac{1}{P-h+1}\sum_{t=R}^{T-h}\widehat{d}_{t+h}}{\sqrt{\frac{1}{P-h+1}\sum_{j=-\bar{j}}^{\bar{j}}\sum_{t=R+j}^{T-h}K\left(\frac{j}{M}\right)\left(\widehat{d}_{t+h}-\overline{d}\right)\left(\widehat{d}_{t+h-j}-\overline{d}\right)}},$$

where $\widehat{d}_{t+h} = \widehat{u}^2_{1,t+h} - \widehat{u}^2_{2,t+h}$, and $\overline{d} = \frac{1}{P-h+1}\sum_{t=R}^{T-h}\widehat{d}_{t+h}$. The limiting distributions of the CM and DM statistics are given in Theorems 3.1 and 3.2 in Clark and McCracken (2004), and for $h > 1$ contain nuisance parameters so that critical values cannot be directly tabulated, and hence Clark and McCracken (2004) use the Kilian parametric bootstrap to obtain critical values. In this case, as discussed above, it is not clear that the parametric bootstrap is asymptotically valid. However, again as alluded to above, the parametric bootstrap approach taken by Clark and McCracken is clearly a good approximation, at least for the DGPs and horizon considered in our experiments, given that these tests have very good finite sample properties (see discussion of results below).

Data are generated according to the DGPs summarized in Table 2.1 as : *Size1-Size2* and *Power1-Power12.*

In our setup, the benchmark model (denoted by *Size1 in* Table 2.1) is an ARX(1). (The benchmark model is also called the "small" model.) The null hypothesis is that no competing model outperforms the benchmark model. Twelve of our DGPs (denoted by *Power1-Power12*) include (non)linear functions of $x_{t-1}$. In this sense, our focus is on (non)linear out-of-sample Granger causality testing. Some regression models estimated in these experiments are misspecified not just because of neglected nonlinearity, but also because fitted regression functions ignore the MA error component that appears in some DGPs. Recall also, as discussed above, that CS and CCS tests only require estimation of the benchmark models. The CM, F, and DM tests require estimation of the benchmark models as well as the alternative models. In our context, the alternative model estimated is simply the benchmark model with $x_{t-1}$ added as an additional regressor, regardless of which DGP is used to generate the data. The alternative is also sometimes called the "big" model.

The functional forms that are specified under the alternative include: (i) exponential (*Power1, Power7*); (ii) linear (*Power2*); (iii) self exciting threshold (*Power3*), squared (*Power8*) and absolute value (*Power9*). In addition, *Power4-Power6* and *Power10-Power12* are the same as the others, except that an MA(1) term is added. Notice that *Power1* includes a nonlinear term that is similar in form to the test function, $w(\cdot)$, which is defined below. Also, *Power2* serves as a linear causality benchmark. Test statistics are constructed by fitting what is referred to in the next section as a "small model" in order to construct the CS and CCS test statistics. Note that the "big model" (which is a linear ARX(1) model in $y_{t-1}$, and $z_{t-1}$ with $x_{t-1}$ added as an additional regressor) is only fitted in order to construct the F, CM, and DM test statistics. It is not necessary to fit this model when constructing the CS and CCS statistics. All test statistics are formed using one-step ahead predictions (and corresponding prediction errors) from recursive and rolling window estimated models.

In all experiments, we set $w(z^{t-1},\gamma) = \exp(\sum_{i=1}^{3}(\gamma_i \tan^{-1}((z_{i,t-1} - \overline{z}_i)/2\widehat{\sigma}_{z_i})))$, with $z_{1,t-1} = x_{t-1}$, $z_{2,t-1} = y_{t-1}, z_{3,t-1} = w_{t-1}$ and $\gamma_1, \gamma_2, \gamma_3$ scalars. Additionally, define $\Gamma = [0.0, 5.0] \times [0.0, 5.0] \times [0.0, 5.0]$. We consider a grid that is delineated by increments of size 0.5. All results are based on 500 Monte Carlo replications, and a sample of $T$=540 is used. All tests are empirical rejection frequencies. The following parameterizations are used: $a_1 = 1.0$, $a_2 = \{0.3, 0.6, 0.9\}$, and $a_3 = 0.3$. Additionally, bootstrap critical values are constructed using 100 simulated statistics, the block length, $l$, is set equal to $\{2, 5, 10\}$, $\{4, 10, 20\}$, or $\{10, 20, 50\}$, depending upon the degree of DGP persistence, as given by the value of $a_2$. Finally, all results are based on $P = (1/2)T$ recursive and rolling window formed predictions.

We summarize our findings from the Monte Carlo simulations in Tables 2.2-2.3 for the CS test and Tables 2.4-2.5 for the F, DM, CM and CCS tests. In addition, Tables 2.2 and 2.4 consider results under recursive estimation, while Tables 2.3 and 2.5 consider results under rolling window estimation. The first column in the mentioned tables states the DGP used to generate the data. The names are further defined in Table 2.1. *Size1-Size2* refer to empirical size experiments and *Power1-Power12* refer to empirical power experiments. All numerical entries are test rejection frequencies. Details of the mnemonics used to describe the columns in the tables and the different approaches used for critical value construction

are contained in the footnotes to Table 2.2 and 2.4.

In the following discussion, we consider two broad issues. First, is the recursive/rolling bootstrap useful, or could one simply use more naive bootstraps such as the standard block bootstrap? Second, what can we say about the use of recursive as opposed to rolling window estimation schemes for estimating model parameters and in particular with respect to inference. As an ancilliary issue, we also consider the issue of in-sample versus out-of-sample testing since we include the in-sample F-test as an alternative test.

A first look at Tables 2.2 and 2.3, where the CS test is examined under the "Recur/Rolling Block Bootstrap" indicates that in general, empirical levels are larger and closer to the 10% nominal level under recursive estimation (Table 2.2) than under rolling window estimation (Table 2.3). For example, in Panel A of Tables 2.2 and 2.3, empirical rejection levels for $l = 2, 5, 10$ are $0.07, 0.07, 0.08$ (Table 2.2) and $0.05, 0.06, 0.07$ (Table 2.3) for *Size1*. However, empirical power is in general closer to 1 under rolling window estimation (Table 2.3) than under recursive estimation (Table 2.2). For example, in Panel A of Tables 2.2 and 2.3 empirical power for $l = 2, 5, 10$ is $0.53, 0.73$,and 0.80 (Table 2.2) and $0.62, 0.87$, and 0.90 (Table 2.3) for *Power1*. This observation about empirical power also holds for the other bootstrap techniques considered. These findings are not surprising, given the that the rolling windows are fixed in length, while the recursive windows increase in length. Furthermore, it is worth stressing that both window types appear to be yield quite reasonable finite sample properties, overall, when the nonparametric bootstrap is used. Finally, notice also that in all panels of Tables 2.2 and 2.3, CS tests constructed using data generated according to *Size2* yield poorer empirical level performance than under *Size1*. This is as expected, given that *Size2* DGPs include unmodelled serial error dependence.

A closer look at Table 2.2 reveals that regardless of the level of dependence in the lagged endogenous variable as determined by the value of $a_2$, the nonparametric block bootstrap developed in this paper consistently has the empirical level closest to the nominal level. For example in Table 2.2, the closest empirical level to the nominal level is 0.08 and it occurs in Panel A when under "Recur Block Bootstrap" and *Size1* when $l = 10$. This same observation can be made in Table 2.3. However, such a blanket conclusion cannot be drawn when comparing empirical power. In Panels A and B of Table 2.2, for the smallest

block lengths of 2 and 4 respectively, the "Block Bootstrap" in general has the highest power levels. For the medium block lengths of 5 and 10 of Panels A and B respectively, the "BB, no PEE, no adj" nonparametric bootstrap has higher power. Finally, for the highest block length, "Recur Block Bootstrap" has the highest empirical power. When there is too much persistence in the model as in Panel C, these conclusions no longer hold. The same conclusions can generally be drawn under the rolling window estimation in Table 2.3.

We now turn to a discussion of Tables 2.4 and 2.5, where results for the rest of the test statistics examined in the Monte Carlo experiments are reported. Relative to the Monte Carlo results in CS (2007), the F-test is not nearly as severely oversized. Indeed, judging from its empirical level and power figures, the F-test seems to have good size and power properties. The main reason for this is that the F-test is in-sample, and is carried out with a correctly specified model in the current analysis. Of course, an in-sample analysis of a correctly specified model for any test will generally yield superior performance. However, as shown in CS (2007), where there is model misspecification in the form of an omitted variable, the in-sample F-test is highly oversized. It is in such cases (i.e. model misspecification) that the argument can be made for considering alternative tests of model performance, even under an assumption of linearity, and particularly when nonlinearities may be present in the true underlying DGPs.

In addition to this, in both Tables 2.4 and 2.5, there is a dramatic improvement in the empirical size of the DM test depending upon which critical values are used (i.e. whether we assume that $\pi = 0$ or $\pi > 0$ - see footnote to Table 2.4 for further explanation of $\pi$). The empirical size under $\pi > 0$ is much closer to the nominal size of 10%. This suggests that parameter estimation error is relevant in our setup, as standard normal critical values (under $\pi = 0$) are simply too big. For the CM test in both Tables 2.4 and 2.5, the assumption that $\pi > 0$ still generates some improvement in empirical size values albeit marginal. Empirical power is very high for the F, CM and DM tests under either assumption on $\pi$; and unlike the CS test in Tables 2.2 and 2.3, power is not compromised by high persistence levels. This is however not the case for the CCS test. In Panel A and B of Table 2.4, the CCS test is grossly oversized regardless of block length. However, for both Tables 2.4 and 2.5, as the model becomes more persistent, there is an improvement in size and a reduction in power.

The fact that this sort of result arises for the CS and CCS tests and not for the F, DM or CM tests indicates that the power loss is due to the use of a block length dependent bootstrap for calculating critical values. Indeed, it is worth noting that the power reduction is also characteristic of the other naive bootstrap techniques in Tables 2.2 and 2.3. Furthermore, it is worth noting that under model misspecification of the variety looked at in CS (2007), the F, CM and DM tests are no longer dominant in the above respect. In the next section, we estimate models that are clearly approximations to the true underlying DGP and hence are probably misspecified. We use the CS test which is robust to model misspecification under both hypotheses, as well as the other tests examined above, to assess the models.

## 2.5 Empirical Illustration: The Marginal Predictive Content of Money for Output

In this section we implement the F, CM, DM, CS and CCS tests that are described in Table 2.1, and examined in the previous Monte Carlo section. In particular, we use these tests together with recursive and rolling window estimation schemes to assess the marginal predictive content of money for real income. Recent contributions to this important literature include the papers of Swanson (1998), Amato and Swanson (2001), and the papers cited therein.

The variables used are the same as those examined by Christiano and Ljungqvist (1988), Stock and Watson (1989), Friedman and Kuttner (1993) and Thoma (1994). In particular, the variables used are monthly observations of industrial production ($IP$), the wholesale price index ($P$), the secondary market rate on 90-day U.S. Treasury bills ($R$), the interest rate on three-month prime commercial paper ($C$) and Divisia monetary aggregates of money supply ($M2$). The sample period is 1959:01 to 2003:12. Seasonally adjusted nominal measures of $M2$ exhibit erratic behavior after 1985, which can be accounted for by documented shifts in the public's demand for money balances. This might explain why the relationship between nominal $M2$, $IP$ and $P$ has been unstable in recent years. Our approach in dealing with shifting money demand is to consider the Divisia monetary aggregates of $M2$. Other approaches, such as including structural breaks and explicit nonlinearities in the models are

left to future research. All data with the exception of the three-month prime commercial paper ($C$) were obtained from the St. Louis Federal Reserve Bank. The data on $C$ were obtained from Stock and Watson (2005).

We define the small model as a vector error correction model with:

$$y_t = \theta^{\dagger}_{1,1} + \theta^{\dagger}_{1,2} y_{t-1} + \theta^{\dagger}_{1,3} z_{1,t-1} + u_{1,t}$$

where

$$\theta^{\dagger}_1 = (\theta^{\dagger}_{1,1}, \theta^{\dagger}_{1,2}, \theta^{\dagger}_{1,3})' = \arg\min_{\theta_1 \in \Theta_1} E(q_1(y_t - \theta_{1,1} - \theta_{1,2} y_{t-1} - \theta_{1,3} z_{1,t-1})) is defined conformably,$$

$$y_t = (\Delta \log IP_t, \Delta \log P_t, \Delta R_t)'$$

and

$$z_{1,t-1} = C_{t-1} - R_{t-1}.$$

We further define the generic alternative (big) model as:

$$y_t = \theta^{\dagger}_{2,1}(\gamma) + \theta^{\dagger}_{2,2}(\gamma) y_{t-1} + \theta^{\dagger}_{2,3}(\gamma) z_{1,t-1} + \theta^{\dagger}_{2,4}(\gamma) w(Z^{t-1}, \gamma) + u_{2,t}(\gamma)$$

where

$$\begin{aligned} \theta^{\dagger}_2(\gamma) &= (\theta^{\dagger}_{2,1}(\gamma), \theta^{\dagger}_{2,2}(\gamma), \theta^{\dagger}_{2,3}(\gamma), \theta^{\dagger}_{2,4}(\gamma))' \\ &= \arg\min_{\theta_2 \in \Theta_2} E(q_1(y_t - \theta_{2,1} - \theta_{2,2} y_{t-1} - \theta_{2,3} z_{1,t-1} - \theta_{2,4} w(Z^{t-1}, \gamma))) \end{aligned}$$

and

$$\begin{aligned} y_t &= (\Delta \log IP_t, \Delta \log P_t, \Delta \log M2_t, \Delta R_t) \\ z_{1,t-1} &= C_{t-1} - R_{t-1} \\ z_{2,t-1} &= \log M2_{t-1} - \log IP_{t-1} - \log P_{t-1}. \end{aligned}$$

Finally, $Z^{t-1} = (z_{2,t-1},\ \Delta \log M2_{t-1})$. Notice that $z_{1,t-1}$ and $z_{2,t-1}$ can be interpreted as vector error correction terms, and are consistent with evidence presented in Swanson (1998) and Amato and Swanson (2001). Since we are interested in examining the (non)linear marginal predictive content of money for income, our forecasting analysis and test statistics are constructed based on estimates of the first equation in the vector error correction model specified above (i.e. the equation with $\Delta \log IP_t$ as dependent variable).

Of note is that standard F-tests or Wald-tests for Granger causality are prone to severe upward size distortions when vector error correction (VEC) models are estimated using only differenced data, without accounting for cointegrating restrictions (see e.g. Swanson (1998) and Swanson, Ozyildirim and Pisu (2003)). One of the reasons why this problem arises is that the moving average representation for a model with cointegrated regressors will not yield a finite order VAR representation. In Swanson (1998) it is noted that at a 1% significance level, trace test statistics support the presence of one cointegrating (CI) vector when the data are linearly detrended, and when an intercept or an intercept and a trend are included in the cointegrating relation. One of the two cointegrating vectors is $z_{1,t-1}$, based on a likelihood ratio test (see Johansen (1988,1991)). Of further note is that the null hypothesis that the other CI vector is $z_{2,t-1}$ almost always fails to reject, although confidence intervals are quite wide relative to those for the interest rate spread CI vector. Finally, it should be recalled (see the discussion in Section 2.4) that in the DM, CM, CCS, and F tests, unlike the CS test, the alternative model is explicitly estimated. In such cases, linearity is assumed, so that the bigger model includes linear functions of $z_{2,t-1}$ and $\Delta \log M2_{t-1}$. This is one of the main reasons why it should not be expected that the results of the different empirical tests "agree". Indeed, if the CS test rejects while all others fail to reject, we have direct evidence of nonlinear Granger causality coupled with evidence of an absence of linear causality, for example.[8]

We construct tests statistics using 1-steap ahead forecasts formed via recursive and rolling window estimated models. Thus, models are re-estimated (using least squares) at each point in time, before each new prediction is constructed. The beginning date for the in-sample period is 1959:1 when constructing the CS, CCS, DM, CM, and F tests, the prediction periods reported on are 1978:1-2003:12 ($\pi = 1.4$), 1981:1-2003:12 ($\pi = 1.0$) and 1987:1-2003:12 ($\pi = 0.6$), so that initial estimation samples for both the recursive and rolling window schemes include data for the periods 1959:1-1977:12, 1959:1-1980:12 and 1959:1-1986:12, respectively. The block length is set equal to 6 in application of the

[8]Here, we are using the notion of "causality" interchangeably with the notion of prediction, in the spirit of what Granger originally had in mind when he introduced causality to the time series profesion (see the discussion in Chao, Corradi and Swanson (2001) for further details).

recursive block bootstrap.[9] In all cases, the dependent variable in regressions and the target variable in forecasts is the first log difference of industrial production (output). As discussed above, all estimated models are linear, and explanatory variables include lags of the first log difference of industrial production, prices, lag first difference of interest rates as well as the CI term $C_{t-1} - R_{t-1}$ (in the benchmark or "small" model). Lags of the first log difference of $M2$ and the CI term $z_{2,t-1}$ are added for the alternative ("big") model. Lags are selected via use of the Schwarz information criterion. Again as discussed above, and given this setup, our tests can be viewed as tests of (non) linear Granger causality.[10]

Results are gathered in Tables 2.6-2.7. In Table 2.6, point mean square forecast errors (MSEs) are tabulated for the "small model" and the "big model" under rolling window and recursive window estimation schemes respectively. Results are given not only for the three prediction periods outlined above, but also for all prediction periods beginning with 1974:1, 1975:1, ..., 1993:1. In Tables 2.7, CS, CCS, F, DM and CM test results for the three prediction periods outlined above are reported.

Turning first to the MSE results in Table 2.6, note that in the case of recursive estimation, the "big" model consistently outperforms the "small" model, for every prediction period. However, in many instances the MSEs are very close in absolute and relative magnitude, with differences often less than 1%. Interestingly, this pattern does not emerge when viewing MSEs associated with models estimated using rolling windows. In particular, the bigger model that includes money only "wins" for prediction periods beginning in 1984, 1988, 1989, 1990, and 1991. This puzzle is further confounded by noting that the lowest MSE model across both estimation window types is sometimes associated with the recursive modelling strategy, and sometimes with the rolling estimation strategy (note that the bold figures denote the lowest MSE across all estimation strategies and model types for a given start year). Thus, it appears that choice of recursive versus rolling estimation in our exercise is quite dependent upon sample prediction period start date.

---

[9]It should be noted that we do not use real-time data in this empirical illustration, even though both variables considered are subject to periodic revision. Extension of our results to incorporate real-time data is left to future research. Additionally, note that various other block lengths were tried and the empirical findings where qualitatively similar regardless of block length.

[10]It should be stressed that the results presented in this section are meant primarily to illustrate the uses of the different tests, and to underscore potentially important differences between the tests.

As mentioned above, the bigger model is always preferred for recursive estimation, while the results are mixed for rolling estimation. In particular, for rolling estimation, the bigger model is preferred for only 5 start years. If the recursively estimated models always yielded the lowest overall MSE across both estimation strategies, our results would be quite straightforward. However, when one looks across estimation strategies, the rolling window approach "wins" when prediction periods begin in the 1990s or from 1974-1982. The recursive window approach "wins" for prediction periods beginning from 1983-1989. This corresponds to our ranking of the models when one looks across *both* estimation strategies. Namely, the lowest MSE model is essentially the bigger model during much of the 1980s (i.e. from 1983 through 1991), while the smaller model "wins" during the rest of the years. Thus, for prediction periods that include the more turbulent 1970s, the smaller model wins, while for prediction periods beginning after 1983, the bigger model with money "wins". This corresponds loosely with the money targeting experiment of the early 1980s. Namely, after this targeting experiment ended, one might argue that a sufficiently "stable" environment ensued for money to become a predictor for output. This is rather interesting, given that the stated goal of the Federal Reserve Board has indeed been stabilization at low levels of inflation.

A further point of interest is that the rolling 10 year estimator that we used in our analysis is indeed dominant with regard to point MSE for 13 or the 20 start years (i.e. 13 of the 20 different prediction periods). Thus, we have some evidence that there may indeed be instabilities resulting in the relatively poorer performance of recursive estimation strategies. As might be expected, this points to model misspecification in the form of structural breaks, missing variables, and omitted nonlinearity, for example.

Finally, it is worth stressing that predictions of income have clearly gotten substantially more accurate over our sample period, as evidenced by the fact that MSFEs are much bigger for early subsamples, and are much smaller for the later sub-samples. This result is clearly due in part to the smooth nature of recent data relative to more distant data, although one might also argue that the more accurate results are associated in large part with instances where models that include money yield superior point predictions, hence pointing to further evidence in favor of using money in output prediction models. It should

be stressed, however, that thus far we have only compared MSEs, and hence have focused our attention upon the comparison of purely linear models. In order to assess the potential impact of generic nonlinearity, for example, we need to either fit a variety of nonlinear models (which may be a large undertaking, given the plethora of available models), or we need to carry out tests such as the generically comprehensive nonlinear out-of-sample Granger causality CS test. We turn to this issue next.

As mentioned above, Table 2.7 contains CS, CCS, F, DM and CM test results for prediction periods beginning in 1978, 1981, and 1987. Three conclusions emerge based upon inspection of the results. First, the CS test fails to reject the null of no (non)linear predictive causation, regardless of prediction period, and regardless of whether recursive or rolling estimation is used. On the other hand, there are many rejections of the null hypothesis when the "linear" tests are used, particularly at the 10% level. Furthermore, these rejections, in the case of recursive estimation, correspond to the big model winning (as the MSE associated with the big model is always lower than that associated with the small model). Thus, based on our recursive results, there is clearly predictive causation from money to output, However, this causation appears to be "moderate" in magnitude, given the fact that the non rejection using the CS test coupled with rejections using the CCS test may be a result of low power associated with the CS test (i.e. the CS test is an omnibus test, and hence has lower power in any given specific alternative than a test designed with that alternative specifically in mind). Second, the number of rejections is close to twice as many when moving from the rolling to the recursive estimation schemes, suggesting that parameter estimation error is playing a significant role in our testing procedures. This finding is also indicative of further evidence in favor of predictive causation, given that in the rolling case, small model MSEs based on prediction periods beginning in 1978, 1981, and 1987 are always lower than corresponding big model MSEs. In other words, in the rolling cases, rejection would imply that the big model is significantly "better" than the small model; and hence fewer rejections supports the finding based on the recursive estimation scheme that there is predictive causation. Third, when changing the significance level from 10% to 5%, some rejections in the CCS, DM and CM tests become non-rejections, which again substantiates the claim that although there is predictive causation, it is somewhat

"weak" in the sense that predictions do not change to a great extent when money is added to the output equation.

In summary, power considerations are relevant, as should be expected, when using the CS test, as evidenced by the fact that in our illustration the CS test may be good at detecting non-linear Granger causality, but it is clearly not good at detecting moderate levels of linear predictive causation. Additionally, our evidence is clearly leaning toward a finding of predictive causation from money to output. However, much empirical work is needed before a complete picture emerges concerning the prevalence of nonlinear Granger causality in the income/money relationship. This is left to future research. It is clear, though, that much can be learned by using *all* of the different tests in consort with one another.

## 2.6 Concluding Remarks

We have discussed bootstrap procedures valid for construction of critical values in the case of test statistics based on recursive and/or rolling estimation schemes that have limiting distributions which are functionals of Gaussian processes, and which have covariance kernels that reflect parameter uncertainty. In these cases, limiting distributions are thus not nuisance parameter free, and valid critical values are often obtained via bootstrap methods. In this paper, we first developed a bootstrap procedure that properly captures the contribution of parameter estimation error in recursive estimation schemes using dependent data. Intuitively, when parameters are estimated recursively, as is done in our framework, earlier observations in the sample enter into test statistics more frequently than later observations. This induces a location bias in the bootstrap distribution, which can be either positive or negative across different samples, and hence the bootstrap modification that we discuss is required in order to obtain first order validity of the bootstrap. Within this framework, we discussed the Corradi and Swanson (2002: CS) model selection type test and carried out a series of experiments evaluating the CS as well as a variety of other tests including ones due to Diebold and Mariano (1995) and Clark and McCracken (2004). Finally, we carried out an empirical investigation using all of the tests examined in the Monte Carlo experiments. the investigation focused on predictive money-income causation. We found that sample size,

prediction period, and estimator type (i.e. recursive versus rolling) play an important role in our empirical findings, although concrete evidence supporting the existence of predictive causation was found, particularly for prediction periods beginning during the 1980s.

Table 2.1: **Test Statistics, Sampling Scheme, and Data Generating Processes Used in Monte Carlo Experiments**

Panel A: Test Statistic Mnemonics and Definitions

$F$ – The standard Wald version of the in-sample F-test is calculated using the entire sample of $T$ observations. In particular, we use: $F = T\left(\sum_{t=1}^{T} \widehat{u}_{1,t}^2 - \sum_{t=1}^{T} \widehat{u}_{2,t}^2)/\sum_{t=1}^{T} \widehat{u}_{2,t}^2\right)$, where $\widehat{u}_{1,t}$ and $\widehat{u}_{2,t}$ are the in-sample residuals associated with least squares estimation of the small and big models, respectively, and where $T$ denotes the sample size.

$CM$ – The Clark and McCracken (2004) test outlined in Section 2.4.

$DM$ – The Diebold and Mariano (1995) test outlined in Section 2.4.

$CS$ – The Corradi and Swanson (2002,2007) test outlined in Section 2.3.

$CCS$ – The Chao, Corradi and Swanson (2001) test discussed in Section 2.4.

Panel B: Data Generating Processes Used in Monte Carlo Experiments

$x_t = a_1 + a_2 x_{t-1} + u_{1,t},\ u_{1,t} \sim iidN(0,1)$

$z_t = a_1 + a_3 z_{t-1} + u_{2,t},\ u_{2,t} \sim iidN(0,1)$

Size1: $y_t = a_1 + a_2 y_{t-1} + a_4 z_{t-1} + u_{3,t},\ u_{3,t} \sim iidN(0,1)$

Size2: $y_t = a_1 + a_2 y_{t-1} + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$

Power1 : $y_t = a_1 + a_2 y_{t-1} + 2\exp(\tan^{-1}(x_{t-1}/2)) + a_4 z_{t-1} + u_{3,t}$

Power2 : $y_t = a_1 + a_2 y_{t-1} + 2x_{t-1} + a_4 w_{t-1} + u_{3,t}$

Power3 : $y_t = a_1 + a_2 y_{t-1} + 2x_{t-1}1\{x_{t-1} > a_1/(1-a_2)\} + a_4 z_{t-1} + u_{3,t}$

Power4 : $y_t = a_1 + a_2 y_{t-1} + 2\exp(\tan^{-1}(x_{t-1}/2)) + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$

Power5: $y_t = a_1 + a_2 y_{t-1} + 2x_{t-1} + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$

Power6: $y_t = a_1 + a_2 y_{t-1} + 2x_{t-1}1\{x_{t-1} > a_1/(1-a_2)\} + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$.

Power7 : $y_t = a_1 + a_2 y_{t-1} + 2\exp(x_{t-1}) + a_4 z_{t-1} + u_{3,t}$

Power8 : $y_t = a_1 + a_2 y_{t-1} + 2x_{t-1}^2 + a_4 z_{t-1} + u_{3,t}$

Power9 : $y_t = a_1 + a_2 y_{t-1} + 2|x_{t-1}| + a_4 z_{t-1} + u_{3,t}$

Power10 : $y_t = a_1 + a_2 y_{t-1} + 2\exp(x_{t-1}) + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$

Power11: $y_t = a_1 + a_2 y_{t-1} + 2x_{t-1}^2 + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$

Power12: $y_t = a_1 + a_2 y_{t-1} + 2|x_{t-1}| + a_4 z_{t-1} + a_3 u_{3,t-1} + u_{3,t}$.

Note that the benchmark or "small" model in our test statistic calculations is always $y_t = \alpha_1 + \alpha_2 y_{t-1} + \alpha_3 z_{t-1} + \epsilon_t$; and the "big" model is the same, but with $x_{t-1}$ or generic functions of $x_{t-1}$ added as an additional regressor.

Table 2.2: **Recursive Estimation Scheme - Rejection Frequencies of CS Test with $T = 540$, $P = 0.5T$**

| Model | Recur Block Bootstrap | | | BB, no PEE, no adj | | | Block Bootstrap | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Panel A:* $a_2 = 0.3$ | | | | | | | | |
| | $l = 2$ | $l = 5$ | $l = 10$ | $l = 2$ | $l = 5$ | $l = 10$ | $l = 2$ | $l = 5$ | $l = 10$ |
| Size1 | 0.07 | 0.07 | 0.08 | 0.01 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 |
| Size2 | 0.04 | 0.05 | 0.07 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 |
| Power1 | 0.53 | 0.73 | 0.80 | 0.00 | 0.59 | 0.83 | 0.54 | 0.77 | 0.74 |
| Power2 | 0.68 | 0.90 | 0.93 | 0.00 | 0.94 | 0.92 | 0.91 | 0.86 | 0.81 |
| Power3 | 0.68 | 0.90 | 0.92 | 0.01 | 0.95 | 0.93 | 0.94 | 0.87 | 0.82 |
| Power4 | 0.53 | 0.76 | 0.81 | 0.00 | 0.54 | 0.84 | 0.42 | 0.76 | 0.76 |
| Power5 | 0.69 | 0.88 | 0.93 | 0.00 | 0.94 | 0.93 | 0.91 | 0.84 | 0.83 |
| Power6 | 0.68 | 0.88 | 0.92 | 0.01 | 0.96 | 0.91 | 0.94 | 0.86 | 0.83 |
| Power7 | 0.57 | 0.75 | 0.77 | 0.02 | 0.76 | 0.77 | 0.77 | 0.73 | 0.70 |
| Power8 | 0.66 | 0.88 | 0.90 | 0.03 | 0.91 | 0.85 | 0.93 | 0.82 | 0.81 |
| Power9 | 0.68 | 0.93 | 0.96 | 0.00 | 0.97 | 0.94 | 0.97 | 0.90 | 0.86 |
| Power10 | 0.57 | 0.73 | 0.77 | 0.02 | 0.76 | 0.76 | 0.79 | 0.73 | 0.71 |
| Power11 | 0.68 | 0.88 | 0.89 | 0.01 | 0.92 | 0.88 | 0.92 | 0.85 | 0.80 |
| Power12 | 0.71 | 0.91 | 0.95 | 0.00 | 0.97 | 0.94 | 0.97 | 0.90 | 0.90 |
| | *Panel B:* $a_2 = 0.6$ | | | | | | | | |
| | $l = 4$ | $l = 10$ | $l = 20$ | $l = 4$ | $l = 10$ | $l = 20$ | $l = 4$ | $l = 10$ | $l = 20$ |
| Size1 | 0.05 | 0.07 | 0.08 | 0.01 | 0.01 | 0.03 | 0.01 | 0.01 | 0.01 |
| Size2 | 0.03 | 0.07 | 0.07 | 0.00 | 0.02 | 0.01 | 0.00 | 0.01 | 0.01 |
| Power1 | 0.59 | 0.69 | 0.75 | 0.00 | 0.65 | 0.80 | 0.56 | 0.64 | 0.68 |
| Power2 | 0.71 | 0.84 | 0.86 | 0.01 | 0.91 | 0.84 | 0.79 | 0.75 | 0.76 |
| Power3 | 0.78 | 0.86 | 0.89 | 0.07 | 0.92 | 0.86 | 0.80 | 0.78 | 0.78 |
| Power4 | 0.57 | 0.69 | 0.78 | 0.00 | 0.61 | 0.82 | 0.56 | 0.66 | 0.69 |
| Power5 | 0.73 | 0.85 | 0.86 | 0.01 | 0.92 | 0.86 | 0.80 | 0.78 | 0.77 |
| Power6 | 0.77 | 0.87 | 0.91 | 0.05 | 0.92 | 0.88 | 0.81 | 0.78 | 0.77 |
| Power7 | 0.56 | 0.64 | 0.68 | 0.05 | 0.64 | 0.67 | 0.53 | 0.60 | 0.63 |
| Power8 | 0.72 | 0.83 | 0.86 | 0.14 | 0.87 | 0.82 | 0.77 | 0.75 | 0.73 |
| Power9 | 0.82 | 0.92 | 0.93 | 0.04 | 0.95 | 0.88 | 0.83 | 0.80 | 0.80 |
| Power10 | 0.57 | 0.62 | 0.67 | 0.07 | 0.67 | 0.67 | 0.62 | 0.61 | 0.63 |
| Power11 | 0.76 | 0.83 | 0.87 | 0.15 | 0.86 | 0.82 | 0.79 | 0.72 | 0.73 |
| Power12 | 0.80 | 0.90 | 0.93 | 0.04 | 0.94 | 0.88 | 0.86 | 0.81 | 0.79 |
| | *Panel C:* $a_2 = 0.9$ | | | | | | | | |
| | $l = 10$ | $l = 20$ | $l = 50$ | $l = 10$ | $l = 20$ | $l = 50$ | $l = 10$ | $l = 20$ | $l = 50$ |
| Size1 | 0.01 | 0.03 | 0.07 | 0.00 | 0.01 | 0.03 | 0.00 | 0.00 | 0.01 |
| Size2 | 0.01 | 0.03 | 0.06 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 |
| Power1 | 0.41 | 0.56 | 0.64 | 0.00 | 0.47 | 0.75 | 0.29 | 0.53 | 0.61 |
| Power2 | 0.59 | 0.71 | 0.77 | 0.06 | 0.78 | 0.79 | 0.58 | 0.65 | 0.72 |
| Power3 | 0.61 | 0.72 | 0.79 | 0.09 | 0.82 | 0.77 | 0.61 | 0.68 | 0.71 |
| Power4 | 0.42 | 0.53 | 0.64 | 0.00 | 0.43 | 0.73 | 0.27 | 0.50 | 0.61 |
| Power5 | 0.61 | 0.69 | 0.77 | 0.03 | 0.81 | 0.78 | 0.57 | 0.67 | 0.70 |
| Power6 | 0.62 | 0.72 | 0.80 | 0.12 | 0.81 | 0.79 | 0.59 | 0.68 | 0.72 |
| Power7 | 0.41 | 0.47 | 0.54 | 0.07 | 0.47 | 0.54 | 0.34 | 0.46 | 0.53 |
| Power8 | 0.57 | 0.67 | 0.75 | 0.23 | 0.76 | 0.69 | 0.56 | 0.62 | 0.66 |
| Power9 | 0.61 | 0.72 | 0.82 | 0.13 | 0.83 | 0.81 | 0.62 | 0.67 | 0.72 |
| Power10 | 0.42 | 0.47 | 0.53 | 0.06 | 0.50 | 0.54 | 0.35 | 0.47 | 0.52 |
| Power11 | 0.60 | 0.66 | 0.75 | 0.27 | 0.76 | 0.73 | 0.59 | 0.63 | 0.66 |
| Power12 | 0.64 | 0.76 | 0.83 | 0.17 | 0.85 | 0.81 | 0.59 | 0.66 | 0.71 |

Notes: All entries are rejection frequencies of the null hypothesis of equal predictive accuracy based on 10% nominal size critical values constructed using the bootstrap approaches discussed above, where $l$ denotes the block length, and empirical bootstrap distributions are constructed using 100 bootstrap statistics. In particular, "Recur Block Bootstrap" is the bootstrap developed in this paper, "BB, no PEE, no adj" is a naive block bootstrap where no parameter estimation error is assumed, and no recentering (i.e. adjustment) is done in parameter estimation or bootstrap statistic construction, "Block Bootstrap" is the usual block bootstrap that allows for parameter estimation error, but does not recenter parameter estimates or bootstrap statistics. For all models denoted Power$i$, $i = 1, ..., 12$, data are generated with (non) linear Granger causality (see above for further discussion of DGPs. In all experiments, the ex ante forecast period is of length $P$, which is set equal to $(1/2)T$, where $T$ is the sample size. All models are estimated recursively, so that parameter estimates are updated before each new prediction is constructed. All reported results are based on 500 Monte Carlo simulations. See Table 2.1 and Section 2.4 for further details.

Table 2.3: **Rolling Estimation Scheme - Rejection Frequencies of CS Test with $T = 540$, $P = 0.5T$**

| Model | Rolling Block Bootstrap | | | BB, no PEE, no adj | | | Block Bootstrap | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Panel A:* $a_2 = 0.3$ | | | | | | | | |
| | $l = 2$ | $l = 5$ | $l = 10$ | $l = 2$ | $l = 5$ | $l = 10$ | $l = 2$ | $l = 5$ | $l = 10$ |
| Size1 | 0.05 | 0.06 | 0.07 | 0.01 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 |
| Size2 | 0.03 | 0.06 | 0.07 | 0.01 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 |
| Power1 | 0.62 | 0.87 | 0.90 | 0.00 | 0.67 | 0.88 | 0.80 | 0.86 | 0.83 |
| Power2 | 0.74 | 0.95 | 0.96 | 0.01 | 0.97 | 0.94 | 0.98 | 0.94 | 0.91 |
| Power3 | 0.77 | 0.95 | 0.98 | 0.01 | 0.97 | 0.94 | 0.98 | 0.94 | 0.92 |
| Power4 | 0.59 | 0.89 | 0.92 | 0.00 | 0.55 | 0.89 | 0.71 | 0.87 | 0.82 |
| Power5 | 0.75 | 0.96 | 0.97 | 0.00 | 0.97 | 0.95 | 0.98 | 0.94 | 0.92 |
| Power6 | 0.76 | 0.95 | 0.97 | 0.01 | 0.98 | 0.94 | 0.98 | 0.95 | 0.92 |
| Power7 | 0.67 | 0.83 | 0.84 | 0.03 | 0.82 | 0.82 | 0.86 | 0.82 | 0.80 |
| Power8 | 0.78 | 0.90 | 0.91 | 0.04 | 0.92 | 0.88 | 0.96 | 0.90 | 0.86 |
| Power9 | 0.79 | 0.96 | 0.96 | 0.01 | 0.97 | 0.94 | 0.98 | 0.94 | 0.92 |
| Power10 | 0.68 | 0.82 | 0.84 | 0.03 | 0.81 | 0.81 | 0.86 | 0.82 | 0.82 |
| Power11 | 0.80 | 0.90 | 0.92 | 0.04 | 0.91 | 0.89 | 0.95 | 0.89 | 0.86 |
| Power12 | 0.78 | 0.94 | 0.96 | 0.00 | 0.97 | 0.94 | 0.98 | 0.93 | 0.92 |
| | *Panel B:* $a_2 = 0.6$ | | | | | | | | |
| | $l = 4$ | $l = 10$ | $l = 20$ | $l = 4$ | $l = 10$ | $l = 20$ | $l = 4$ | $l = 10$ | $l = 20$ |
| Size1 | 0.03 | 0.04 | 0.05 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 |
| Size2 | 0.03 | 0.03 | 0.06 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 |
| Power1 | 0.68 | 0.81 | 0.85 | 0.01 | 0.69 | 0.85 | 0.74 | 0.75 | 0.75 |
| Power2 | 0.82 | 0.91 | 0.94 | 0.04 | 0.94 | 0.90 | 0.90 | 0.84 | 0.84 |
| Power3 | 0.88 | 0.96 | 0.95 | 0.08 | 0.97 | 0.91 | 0.90 | 0.84 | 0.85 |
| Power4 | 0.65 | 0.84 | 0.88 | 0.01 | 0.67 | 0.85 | 0.69 | 0.77 | 0.78 |
| Power5 | 0.84 | 0.93 | 0.94 | 0.02 | 0.95 | 0.91 | 0.88 | 0.84 | 0.83 |
| Power6 | 0.89 | 0.96 | 0.95 | 0.09 | 0.96 | 0.92 | 0.93 | 0.88 | 0.86 |
| Power7 | 0.65 | 0.70 | 0.72 | 0.07 | 0.67 | 0.68 | 0.68 | 0.67 | 0.68 |
| Power8 | 0.83 | 0.89 | 0.89 | 0.17 | 0.89 | 0.85 | 0.85 | 0.82 | 0.81 |
| Power9 | 0.90 | 0.94 | 0.94 | 0.12 | 0.94 | 0.92 | 0.92 | 0.89 | 0.88 |
| Power10 | 0.63 | 0.70 | 0.73 | 0.05 | 0.69 | 0.69 | 0.65 | 0.68 | 0.67 |
| Power11 | 0.82 | 0.88 | 0.90 | 0.14 | 0.89 | 0.86 | 0.85 | 0.85 | 0.82 |
| Power12 | 0.90 | 0.93 | 0.94 | 0.08 | 0.94 | 0.92 | 0.92 | 0.89 | 0.89 |
| | *Panel C:* $a_2 = 0.9$ | | | | | | | | |
| | $l = 10$ | $l = 20$ | $l = 50$ | $l = 10$ | $l = 20$ | $l = 50$ | $l = 10$ | $l = 20$ | $l = 50$ |
| Size1 | 0.01 | 0.03 | 0.06 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| Size2 | 0.01 | 0.02 | 0.03 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 |
| Power1 | 0.37 | 0.54 | 0.72 | 0.00 | 0.53 | 0.75 | 0.36 | 0.56 | 0.63 |
| Power2 | 0.64 | 0.75 | 0.83 | 0.13 | 0.82 | 0.82 | 0.60 | 0.69 | 0.75 |
| Power3 | 0.68 | 0.78 | 0.88 | 0.16 | 0.86 | 0.84 | 0.67 | 0.71 | 0.77 |
| Power4 | 0.35 | 0.54 | 0.74 | 0.00 | 0.43 | 0.76 | 0.32 | 0.52 | 0.67 |
| Power5 | 0.61 | 0.75 | 0.85 | 0.07 | 0.80 | 0.82 | 0.60 | 0.67 | 0.77 |
| Power6 | 0.71 | 0.80 | 0.87 | 0.16 | 0.86 | 0.84 | 0.67 | 0.72 | 0.78 |
| Power7 | 0.48 | 0.54 | 0.60 | 0.06 | 0.49 | 0.55 | 0.44 | 0.54 | 0.58 |
| Power8 | 0.67 | 0.77 | 0.81 | 0.24 | 0.81 | 0.75 | 0.65 | 0.68 | 0.72 |
| Power9 | 0.76 | 0.83 | 0.89 | 0.14 | 0.89 | 0.86 | 0.70 | 0.76 | 0.79 |
| Power10 | 0.47 | 0.56 | 0.60 | 0.06 | 0.50 | 0.56 | 0.42 | 0.54 | 0.58 |
| Power11 | 0.67 | 0.75 | 0.81 | 0.21 | 0.82 | 0.75 | 0.66 | 0.70 | 0.73 |
| Power12 | 0.75 | 0.84 | 0.89 | 0.18 | 0.89 | 0.85 | 0.71 | 0.77 | 0.79 |

Notes: See notes to Table 2.2.

Table 2.4: **Recursive Estimation Scheme - Rejection Frequencies of Various Tests with $T = 540$, $P = 0.5T$**

| Model | Assume $\pi = 0$ | | | Assume $\pi > 0$ | | Recur Block Bootstrap | | |
|---|---|---|---|---|---|---|---|---|
| | F | DM | CM | DM | CM | CCS-$l1$ | CCS-$l2$ | CCS-$l3$ |
| *Panel A:* $a_2 = 0.3$ | | | | | | | | |
| Size1 | 0.11 | 0.01 | 0.06 | 0.10 | 0.10 | 0.20 | 0.21 | 0.20 |
| Size2 | 0.11 | 0.01 | 0.07 | 0.11 | 0.11 | 0.17 | 0.17 | 0.17 |
| Power1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.94 |
| Power2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 |
| Power3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 |
| Power4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.96 |
| Power5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 |
| Power6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 |
| Power7 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.89 | 0.78 |
| Power8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 | 0.92 |
| Power9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| Power10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.88 | 0.77 |
| Power11 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.91 |
| Power12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| *Panel B:* $a_2 = 0.6$ | | | | | | | | |
| Size1 | 0.09 | 0.02 | 0.04 | 0.10 | 0.09 | 0.19 | 0.22 | 0.20 |
| Size2 | 0.11 | 0.01 | 0.06 | 0.10 | 0.09 | 0.14 | 0.16 | 0.19 |
| Power1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 0.91 | 0.89 |
| Power2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 | 0.95 |
| Power3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.94 | 0.93 |
| Power4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 | 0.91 | 0.86 |
| Power5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.96 | 0.94 |
| Power6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.94 | 0.92 |
| Power7 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.69 | 0.64 |
| Power8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.89 | 0.84 |
| Power9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 0.91 |
| Power10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.81 | 0.67 | 0.61 |
| Power11 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.86 | 0.84 |
| Power12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 0.92 |
| *Panel C:* $a_2 = 0.9$ | | | | | | | | |
| Size1 | 0.10 | 0.01 | 0.06 | 0.11 | 0.11 | 0.10 | 0.11 | 0.14 |
| Size2 | 0.13 | 0.01 | 0.08 | 0.11 | 0.14 | 0.09 | 0.11 | 0.16 |
| Power1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.68 | 0.74 | 0.76 |
| Power2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.81 | 0.83 | 0.86 |
| Power3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.80 | 0.82 |
| Power4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.63 | 0.67 | 0.73 |
| Power5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.82 | 0.85 |
| Power6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.76 | 0.79 | 0.84 |
| Power7 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 0.47 | 0.43 | 0.49 |
| Power8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.70 | 0.73 | 0.78 |
| Power9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.75 | 0.78 | 0.82 |
| Power10 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 0.42 | 0.45 | 0.48 |
| Power11 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.69 | 0.75 | 0.78 |
| Power12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.77 | 0.77 | 0.81 |

Notes: See notes to Table 2.2. Test statistics, denoted by F, DM, CM, CS, and CCS are summarized in Table 2.1. Block lengths are denoted by $l1$, $l2$, and $l3$, so that $CCS - l3$ is the CCS test with block length $l3$. Block lengths corrspond to those used in Table 2.2 and 2.3, so that for $a_2 = 0.3$, $l1, l2, l3$=2,5,10. The block lengths for $a_2 = 0.6$ and $a_2 = 0.9$ are $l1, l2, l3$=4,10,20 and $l1, l2, l3$=10,20,50, respectively. $\pi = 0$ corresponds to the case where standard critical values based upon the assumption that parameter estimation error vanishes asymptotically are used (i.e. $\pi = lim_{T->\infty} P/R = 0$). $\pi > 0$ corresponds to the case where nonstandard critical values (see McCracken (2004)) based upon the assumption that parameter estimation error does not vanish asymptotically are used (i.e. $\pi = lim_{T->\infty} P/R > 0$). In this case, we assume that $\pi = 1$.

Table 2.5: **Rolling Estimation Scheme - Rejection Frequencies of Various Tests with $T = 540$, $P = 0.5T$**

| Model | Assume $\pi = 0$ | | | Assume $\pi > 0$ | | Recur Block Bootstrap | | |
|---|---|---|---|---|---|---|---|---|
| | F | DM | CM | DM | CM | CCS-$l1$ | CCS-$l2$ | CCS-$l3$ |
| *Panel A:* $a_2 = 0.3$ | | | | | | | | |
| Size1 | 0.11 | 0.00 | 0.06 | 0.07 | 0.09 | 0.17 | 0.17 | 0.16 |
| Size2 | 0.10 | 0.00 | 0.06 | 0.10 | 0.09 | 0.14 | 0.14 | 0.13 |
| Power1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| Power2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Power3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| Power4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| Power5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Power6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Power7 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.87 | 0.81 |
| Power8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.94 |
| Power9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 |
| Power10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.87 | 0.80 |
| Power11 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 |
| Power12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 |
| *Panel B:* $a_2 = 0.6$ | | | | | | | | |
| Size1 | 0.10 | 0.01 | 0.05 | 0.08 | 0.08 | 0.13 | 0.14 | 0.16 |
| Size2 | 0.13 | 0.01 | 0.06 | 0.10 | 0.10 | 0.11 | 0.13 | 0.15 |
| Power1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.94 | 0.93 |
| Power2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.96 |
| Power3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.96 |
| Power4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.94 | 0.91 |
| Power5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.96 |
| Power6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.96 |
| Power7 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.86 | 0.71 | 0.67 |
| Power8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.94 | 0.90 |
| Power9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.95 |
| Power10 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.84 | 0.72 | 0.68 |
| Power11 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 0.91 |
| Power12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.94 |
| *Panel C:* $a_2 = 0.9$ | | | | | | | | |
| Size1 | 0.11 | 0.02 | 0.04 | 0.08 | 0.10 | 0.07 | 0.12 | 0.13 |
| Size2 | 0.13 | 0.02 | 0.06 | 0.12 | 0.13 | 0.06 | 0.07 | 0.13 |
| Power1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.72 | 0.72 | 0.80 |
| Power2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.85 | 0.83 | 0.86 |
| Power3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.82 | 0.82 |
| Power4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.70 | 0.71 | 0.78 |
| Power5 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.83 | 0.82 | 0.87 |
| Power6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.81 | 0.80 | 0.85 |
| Power7 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 0.46 | 0.44 | 0.49 |
| Power8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.78 | 0.82 |
| Power9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.87 | 0.83 | 0.87 |
| Power10 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 0.48 | 0.48 | 0.53 |
| Power11 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.81 | 0.82 |
| Power12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.83 | 0.84 | 0.87 |

Notes: See notes to Table 2.4.

Table 2.6: **Mean Square Forecast Errors and the Marginal Predictive Content of $M2$ for Output**

| | Recursive | | Rolling | |
|---|---|---|---|---|
| Start Year | *small model* | *big model* | *small model* | *big model* |
| 1974 | 0.0000398 | 0.0000395 | **0.0000393** | 0.0000410 |
| 1975 | 0.0000359 | 0.0000355 | **0.0000345** | 0.0000363 |
| 1976 | 0.0000352 | 0.0000349 | **0.0000338** | 0.0000352 |
| 1977 | 0.0000353 | 0.0000350 | **0.0000340** | 0.0000345 |
| 1978 | 0.0000350 | 0.0000348 | **0.0000337** | 0.0000344 |
| 1979 | 0.0000345 | 0.0000341 | **0.0000334** | 0.0000339 |
| 1980 | 0.0000343 | 0.0000338 | **0.0000333** | 0.0000336 |
| 1981 | 0.0000328 | 0.0000325 | **0.0000323** | 0.0000330 |
| 1982 | 0.0000321 | **0.0000317** | **0.0000317** | 0.0000323 |
| 1983 | 0.0000285 | **0.0000280** | 0.0000284 | 0.0000287 |
| 1984 | 0.0000271 | **0.0000261** | 0.0000274 | 0.0000271 |
| 1985 | 0.0000281 | **0.0000271** | 0.0000280 | 0.0000281 |
| 1986 | 0.0000284 | **0.0000274** | 0.0000281 | 0.0000283 |
| 1987 | 0.0000281 | **0.0000272** | 0.0000278 | 0.0000279 |
| 1988 | 0.0000279 | **0.0000266** | 0.0000272 | 0.0000270 |
| 1989 | 0.0000292 | **0.0000279** | 0.0000285 | 0.0000280 |
| 1990 | 0.0000281 | **0.0000269** | 0.0000275 | **0.0000269** |
| 1991 | 0.0000278 | 0.0000269 | 0.0000269 | **0.0000267** |
| 1992 | 0.0000278 | 0.0000271 | **0.0000267** | 0.0000270 |
| 1993 | 0.0000288 | 0.0000280 | **0.0000275** | 0.0000279 |

Notes: For the empirical work, the variables used are monthly observations of industrial production ($IP$), the wholesale price index ($P$), the secondary market rate on 90-day U.S. Treasury bills ($R$), the interest rate on three-month prime commercial paper ($C$) and Divisia monetary aggregates of money supply ($M2$). The sample period is 1959-01 to 2003-12.
We define the small model as:

$$y_t = \theta^\dagger_{1,1} + \theta^\dagger_{1,2} y_{t-1} + \theta^\dagger_{1,3} z_{1,t-1} + u_{1,t}$$

where

$$\theta^\dagger_1 = (\theta^\dagger_{1,1}, \theta^\dagger_{1,2}, \theta^\dagger_{1,3})' = \arg \min_{\theta_1 \in \Theta_1} E(q_1(y_t - \theta_{1,1} - \theta_{1,2} y_{t-1} - \theta_{1,3} z_{1,t-1})) is defined conformably,$$

$$y_t = (\Delta \log IP_t, \Delta \log P_t, \Delta R_t)'$$

and

$$z_{1,t-1} = C_{t-1} - R_{t-1}.$$

We further define the generic alternative (big) model as:

$$y_t = \theta^\dagger_{2,1}(\gamma) + \theta^\dagger_{2,2}(\gamma) y_{t-1} + \theta^\dagger_{2,3}(\gamma) z_{1,t-1} + \theta^\dagger_{2,4}(\gamma) w(Z^{t-1}, \gamma) + u_{2,t}(\gamma)$$

where

$$\theta^\dagger_2(\gamma) = (\theta^\dagger_{2,1}(\gamma), \theta^\dagger_{2,2}(\gamma), \theta^\dagger_{2,3}(\gamma), \theta^\dagger_{2,4}(\gamma))' = \arg \min_{\theta_2 \in \Theta_2} E(q_1(y_t - \theta_{2,1} - \theta_{2,2} y_{t-1} - \theta_{2,3} z_{1,t-1} - \theta_{2,4} w(Z^{t-1}, \gamma)))$$

and

$$\begin{aligned} y_t &= (\Delta \log IP_t, \Delta \log P_t, \Delta \log M2_t, \Delta R_t) \\ z_{1,t-1} &= C_{t-1} - R_{t-1} \\ z_{2,t-1} &= \log M2_{t-1} - \log IP_{t-1} - \log P_{t-1}. \end{aligned}$$

$z_{1,t-1}$ and $z_{2,t-1}$ can be interpreted as vector error correction terms. Mean square forecast errors are reported for the small and big models as defined above. Since we are interested in examining the (non)linear marginal predictive content of money for income, our forecasting analysis and test statistics are constructed based on estimates of the first equation in the vector error correction model specified above. All predictions are 1-step ahead output and predictive periods begin in the year given in the first column of entries in the table. Entries in bold represent the lowest MSFE for the corresponding year in which prediction started.

Table 2.7: **Tests for the Marginal Predictive Content of $M2$ for Output**

| | Prediction Period Begins in | | |
|---|---|---|---|
| Test Statistic | 1987($\pi = 0.6$) | 1981($\pi = 1.0$) | 1978($\pi = 1.4$) |
| Panel A: Sig Level = 5%; Recursive | | | |
| CS (Recur Block Bootstrap) | no reject | no reject | no reject |
| CCS (Recur Block Bootstrap) | no reject | reject | no reject |
| F | reject | reject | reject |
| DM (Tabulated CVs) | reject | no reject | no reject |
| CM (Tabulated CVs) | reject | reject | reject |
| Panel B: Sig Level = 10%; Recursive | | | |
| CS (Recur Block Bootstrap) | no reject | no reject | no reject |
| CCS (Recur Block Bootstrap) | reject | reject | no reject |
| F | reject | reject | reject |
| DM (Tabulated CVs) | reject | reject | reject |
| CM (Tabulated CVs) | reject | reject | reject |
| Panel C: Sig Level = 5%; Rolling | | | |
| CS (Recur Block Bootstrap) | no reject | no reject | no reject |
| CCS (Recur Block Bootstrap) | no reject | no reject | no reject |
| F | reject | reject | reject |
| DM (Tabulated CVs) | no reject | no reject | no reject |
| CM (Tabulated CVs) | reject | no reject | no reject |
| Panel D: Sig Level = 10%; Rolling | | | |
| CS (Recur Block Bootstrap) | no reject | no reject | no reject |
| CCS (Recur Block Bootstrap) | reject | reject | no reject |
| F | reject | reject | reject |
| DM (Tabulated CVs) | no reject | no reject | no reject |
| CM (Tabulated CVs) | reject | no reject | reject |

Notes: Entries denote either rejection (reject) or failure to reject (no reject) the null hypothesis that $M2$ has no marginal predictive content for output. Entries denote nominal 5% and 10% level test rejection based on critical values constructed using the approach signified in brackets in the first column of the table. The models are as described in the notes to Table 2.6. All models use monthly data and all predictions are based on 1-step ahead recursive and rolling schemes.

# Chapter 3

# Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Largescale Macroeconomic Time Series Environments

## 3.1 Introduction

The idea that individual economic variables can be forecast with some precision by refining the information from a large panel of data into a small set of estimated factors (predictors) is intriguing. It suggests that there is a small set of crucial latent factors which generate the co-movements in a large set of macroeconomic variables. This idea is consistent, for example, with the notion that a small set of underlying shocks are responsible for the dynamic behavior implicit in dynamic stochastic general equilibrium models. The practice of using observable economic variables to proxy the latent factors is espoused on the Federal Reserve Bank of New York's website: *"In formulating the nation's monetary policy, the Federal Reserve considers a number of factors, including the economic and financial indicators which follow, as well as the anecdotal reports compiled in the Beige Book. Real Gross Domestic Product (GDP); Consumer Price Index (CPI); Nonfarm Payroll Employment Housing Starts; Industrial Production/Capacity Utilization; Retail Sales; Business Sales and Inventories; Advance Durable Goods Shipments, New Orders and Unfilled Orders; Lightweight Vehicle Sales; Yield on 10-year Treasury Bond; S&P 500 Stock Index; M2"* (see *http://www.newyorkfed.org/education/bythe.html).* The recent literature on factor (diffusion index) models is rich and diverse. A very few of the most important papers include: Bai (2003), Bai and Ng (2002, 2005, 2006a,b,c,d), Boivin and Ng (2005), Connor and Korajczyk (1993), Ding and Hwang (1999), Forni, Hallin, Lippi, and Reichlin (2000, 2005), Forni and Reichlin (1996, 1998), Geweke (1977), Rapach and Strauss (2007), and Stock and Watson (1996, 1998, 1999, 2002a,b, 2004a,b, 2005).

In this paper, our purpose is twofold: first, we provide a review of the extant literature, with careful emphasis on the implementation of factor estimation and prediction using the methods of Bai and Ng as well as Stock and Watson. We then outline a simple methodology for the construction of factor proxies for use in prediction models, where our proxies are observable economic variables. In this sense, we attempt to look inside the "black box", in the sense that our proxy factors are observable and hence have clear economic meaning, while factors in general are often hard to interpret economically (see below for further discussion). As a case in point, policy makers use individual observable variables as policy instruments, for example. Our factor proxies might thus be used for policy, while estimated unobserved factors are not as obviously used in policy applications. In this sense, our main contribution is to add to the broad literature on prediction using factor models. The methodology that we outline is very straightforward, and is based upon application of the $A(j)$ and $M(j)$ statistics developed in Bai and Ng (2006a,b). An ancillary purpose in this paper is to note that in some cases factor proxies defined as observable variables may actually perform as well as estimates of unobserved factors based on standard factor analysis. This is rather an interesting finding, suggesting for example that factor analysis should be applied with caution, particularly in cases where parameter estimation error implicit to factor construction may be great.

Following the approach of Stock and Watson (2002a,b), diffusion index forecasts involve a two-step procedure. First, the method of principal components is used to estimate the factors from a large panel of possible predictors, $X$. Second, the estimated factors are used to forecast the variable of interest, $y_{t+1}$. Stock and Watson (2002a) demonstrate that diffusion index forecasts yield encouraging results. Bai and Ng (2006a), however, point out that the regressors (factors) in the diffusion index model are estimated, hence substantially increasing the forecast error variance. In a related paper, Bai and Ng (2006b) examine whether observable economic variables can serve as proxies for the underlying unobserved factors. In particular, they use the $A(j)$ and $M(j)$ statistics to determine whether a group of observed variables yields precisely the same information as that contained in the latent factors. Stock and Watson (2002a) have also attempted to link the factors to observed variables. Thus, in some sense, Bai and Ng, as well as Stock and Watson, have already looked

inside the "black box". Our approach is to take their argument one step further, and to argue that if observable economic variables are indeed good proxies of the unobserved factors, then these proxies can be used in place of the factors in the diffusion index model for prediction. Once the set of factor proxies is fixed, we effectively eliminate the incremental increase in forecast error variance (i.e., uncertainty) associated with the use of estimated factors. Along these lines, we consider "smoothed" versions of the $A(j)$ and $M(j)$ statistics that pre-select a set of factor proxies prior to the ex-ante construction of a sequence of predictions. It is worth noting that by replacing the estimated factors with observed variables, we are trading off the above variety of uncertainty with "variable selection uncertainty". Our empirical results suggest that there are cases in macro forecasting where the trade-off is worthwhile.

In a Monte Carlo experiment, we show that the $A(j)$ and $M(j)$ statistics can be used to construct prediction models that compare perform quite favorably when compared against standard factor model predictions. We additionally carry out a large variety of prediction experiments using the macroeconomic dataset of Stock and Watson (2005). In these experiments, we predict a number of price and income variables, including industrial production, real personal income less transfers, real manufacturing and trade sales, the number of employees on non-agricultural payrolls, the consumer price index, the personal consumption expenditure implicit price deflator, and the producer price index for finished goods. Using recursively estimated models, we construct $h = 1, 3, 12$, and 24 step ahead forecasts. We show that the $A(j)$ and $M(j)$ statistics appear to offer an interesting means by which factor proxies for later use in prediction models can be chosen. Indeed, our "smoothed" approaches to factor proxy selection appear to yield predictions that are often mean square forecast error "superior" not only relative to a benchmark factor model, but also to simple linear time series models which are often difficult to beat in forecasting competitions. Furthermore, our methods based on the use of the $A(j)$ statistic appear to perform better than those based on the $M(j)$ statistic. Finally, we provide evidence that: (i) versions of our factor proxy selection method that use only a single factor proxy are preferred to those based on the use of $\widehat{k}$ proxies, where $\widehat{k}$ is a consistent estimate of the true number of factors; and (ii) while our "smoothed" proxy selection method is clearly superior for $h = 1, 3$, and 12, the method breaks down at the longest forecast horizon that we consider (i.e., $h = 24$).

For the longest horizons, the estimated factor approach to prediction (e.g., that used by Stock and Watson (2002a,b)) dominates.

By using our approach to predictive factor proxy selection, we believe that we are able to "open up" the "black box" often associated with factor analysis, at least to a certain extent, and to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables. Our empirical analysis suggests that important underlying observable variables, in the sense that they are good proxies for latent factors, include: the S&P500 price index and dividend series; the 1-year Treasury bond rate; various housing activity variables; industrial production; and an exchange rate.

The rest of the paper is organized as follows. In Section 3.2 we review the diffusion index literature, with some focus on the methods that are used in our Monte Carlo and empirical experiments. In Section 3.3 we discuss the use of factor proxies, including a discussion of the Bai and Ng (2006a,b) tests, and a discussion of the methodological approach to the construction and use of factor proxies for prediction. Section 3.4 contains a summary of the empirical methodology used in the paper, and Section 3.5 summarizes the data used. In Section 3.6, the results of a small Monte Carlo experiment studying the finite sample properties of the Bai and Ng (2006a,b) tests are presented, and in Section 3.7 we summarize our empirical findings. Finally, in Section 3.8 we briefly discuss the most recent advances in the diffusion index methodology; and concluding remarks are gathered in Section 3.9.

## 3.2 Review: Diffusion Index Models and the Principle Components Approach to Estimation

### 3.2.1 The diffusion index model

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 $t = 1, \ldots, T$. Assume that $(y_{t+1}, X_t)$ has a dynamic factor model representation with $\overline{r}$ common dynamic factors, $f_t$. Hence, $f_t$ is an $\overline{r} \times 1$ vector. The dynamic factor model is written as:

$$y_{t+h} = \alpha(L) f_t + \beta' W_t + \varepsilon_{t+h} \tag{3.1}$$

and

$$x_{it} = \lambda_i(L) f_t + e_{it}, \tag{3.2}$$

for $i = 1, 2, \ldots, N$, where $W_t$ is an $l \times 1$ vector of other observable variables with $l << N$, such as contemporaneous and lagged values of $y_t$; $h > 0$ is the lead time between information available and the dependent variable; $x_{it}$ 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 $L$. 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 (3.1) and (3.2) as:

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h} \tag{3.3}$$

and

$$x_{it} = \Lambda_i' F_t + e_{it}, \tag{3.4}$$

where $F_t = (f_t', \ldots, f_{t-q}')'$ is an $r \times 1$ vector, with $r = (q+1)\overline{r}$ and $\alpha$ is an $r \times 1$ vector. Here, $r$ is the number of static factors (i.e., the number of elements in $F_t$). Additionally, $\Lambda_i = (\lambda_{i0}', \ldots, \lambda_{iq}')'$ is a vector of factor loadings on the $r$ static factors, where $\lambda_{ij}$ is an $\overline{r} \times 1$ vector for $j = 0, \ldots, q$ and $\beta = (\beta_1, \ldots, \beta_l)'$. Alternatively, from (3.2), the dynamic factor model can be represented as:

$$\begin{aligned} x_{it} &= \lambda_{i0}' f_t + \lambda_{i1}' f_{t-1} + \ldots + \lambda_{iq}' f_{t-q} + e_{it} && (3.5) \\ &= \lambda_i'(L) f_t + e_{it} && (3.6) \end{aligned}$$

and:

$$\lambda_i(L) = \lambda_{i0} + \lambda_{i1} L^1 + \ldots + \lambda_{iq} L^q.$$

For complete details, see Bai and Ng (2005). Now, (3.6) can be written in the static form (3.4) where $F_t$ and $\Lambda_i$ are defined as above. The static factor model refers to the contemporaneous relationship between $x_{it}$ and $F_t$. One major advantage of the static representation of the dynamic factor model is it enables us to use principal components to estimate the factors. This involves estimating $F_t$ using an eigenvalue-eigenvector decomposition of the

sample covariance matrix of the data. It is worth noting that the use of principal components to estimate the factors cannot be done with infinitely distributed lags of the factors (see Stock and Watson (2002a)). Ding and Hwang (1999), Forni et al. (2000), 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. However, many economic analyses hinge on the ability to isolate the primitive shocks or the number of dynamic factors (see Bai and Ng (2007)). Boivin and Ng (2005) also 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. If the idiosyncratic errors, $e_t = (e_{1t}, \ldots, e_{Nt})'$, are cross-sectionally independent and i.i.d. over time, then (3.4) is the classical factor analysis model. It is important at this juncture to note that the factor model does not generally require the idiosyncratic errors to be cross-sectionally independent (see e.g., Bai and Ng (2002)). This is a crucial departure, as it ensures that we can assume the existence of an "approximate" rather than "strict" factor model. (Moreover, the idiosyncratic errors are restricted to be "weakly" correlated, roughly speaking, as the basic structure of the factor model requires the factors to account for the "bulk" of the co-movement across variables). Of final note, it should be mentioned that Geweke (1977) and Sargent and Sims (1977) were among the first to extend the classical factor analysis model to dynamic models.

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 $(T \times 1)\underline{X}_i = (T \times r)F^0(r \times 1)\Lambda_i + (T \times 1)\underline{e}_i$ where $\underline{X}_i = (X_{i1}, \ldots, X_{iT})'$, $F^0 = (F_1, \ldots, F_T)'$ and $\underline{e}_i = (e_{i1}, \ldots, e_{iT})'$. The whole panel of data $X = (\underline{X}_1, \ldots, \underline{X}_N)$ can consequently be represented as $(T \times N)X = (T \times r)F^0(r \times N)\Lambda' + (T \times N)e$, where $\Lambda = (\Lambda_1, \ldots, \Lambda_N)'$ and $e = (\underline{e}_1, \ldots, \underline{e}_N)$. Connor and Korajczyk (1986, 1988, 1993) (1996, 1998) and Forni, Hallin, Lippi and Reichlin (2000) Stock and Watson (2002b) We will also assume $\{F_t\}$ and $\{e_{it}\}$ are two groups of mutually independent stochastic variables. 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 (3.3) and (3.4) 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, $\{\widetilde{F}_t\}_{t=1}^T$ by means of principal components. With the estimated factors in hand, we obtain the estimators $\widehat{\alpha}$ and $\widehat{\beta}$ by regressing $y_{t+h}$ onto $\widetilde{F}_t$ 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 $\widehat{\alpha}$ and $\widehat{\beta}$ are $\sqrt{T}$ consistent and asymptotically normal (see Bai and Ng (2005)).

### 3.2.2 Common factor estimation using principal components

The problem of obtaining the necessary estimates in (3.4) would be simplified if we knew $F^0$. Then $\Lambda_i$ could be estimated via least squares by setting $\{x_{it}\}_{t=1}^T$ to be the dependent variable and $\{F_t\}_{t=1}^T$ to be the explanatory variable. On the other hand, if $\Lambda$ were known, $F_t$ could be estimated by regressing $\{x_{it}\}_{i=1}^N$ on $\{\Lambda_i\}_{i=1}^N$. Since the common factors are not observed, in the regression analysis of (3.4), we replace $F_t$ by $\widetilde{F}_t$, 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, 2004a, 2004b) and Bai and Ng (2002) for a detailed explanation of this procedure, and to Connor and Korajczyk (1986, 1988, 1993), Forni and Reichlin (1996, 1998) and Forni, Hallin, Lippi and Reichlin (2000) for further detailed discussion of diffusion models, in general.

As noted earlier $F_t$ and $\lambda_i$ are not separately identified, but rather identifiable only up to a square matrix. Stock and Watson (1998) further demonstrate that when principal components is used, the factors remain consistent even when there is some time variation in $\Lambda$ and small amounts of data contamination, so long as the number of variables in the panel data set or the number of predictors is very large (i.e., $N >> T$). In this paper, we only give an outline of how principal component analysis is carried out, with particular emphasis on those features of the analysis that allow us to carry out our prediction experiments using

the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006b).

Let $k$ ($k < \min\{N, T\}$) be an arbitrary number of factors, $\Lambda^k$ be the $N \times k$ matrix of factor loadings, $(\Lambda_1^k, \ldots, \Lambda_N^k)'$, and $F^k$ be a $T \times k$ matrix of factors $(F_1^k, \ldots, F_T^k)'$. From (3.4), estimates of $\Lambda_i^k$ and $F_t^k$ are obtained by solving the optimization problem:

$$V(k) = \Lambda^k, F^k \min(NT)^{-1} i = 1N \sum Tt = 1 \sum (x_{it} - \Lambda_i^{k\prime} F_t^k)^2 \tag{3.7}$$

Let $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$ be the minimizers of equation (3.7). Since $\Lambda^k$ and $F^k$ are not separately identifiable, if $N > T$, a computationally expedient approach would be to concentrate out $\widetilde{\Lambda}^k$ and minimize (3.7) subject to the normalization $F^{k\prime} F^k / T = I_k$. Minimizing (3.7) is equivalent to maximizing $tr[F^{k\prime}(XX')F^k]$. This optimization is solved by setting $\widetilde{F}^k$ to be the matrix of the $k$ eigenvectors of $XX'$ that correspond to the $k$ largest eigenvalues of $XX'$. Note that $tr[\cdot]$ represents the matrix trace. The superscript in $\Lambda^k$ and $F^k$ signifies the use of $k$ factors in the estimation and the fact that the estimates will depend on $k$. Let $\widetilde{D}$ be a $k \times k$ diagonal matrix consisting of the $k$ largest eigenvalues of $XX'$. The estimated factor matrix, denoted by $\widetilde{F}^k$, is $\sqrt{T}$ times the eigenvectors corresponding to the $k$ largest eigenvalues of the $T \times T$ matrix $XX'$. Given $\widetilde{F}^k$ and the normalization $F^{k\prime} F^k / T = I_k$, $\widetilde{\Lambda}^{k\prime} = (\widetilde{F}^{k\prime} \widetilde{F}^k)^{-1} \widetilde{F}^{k\prime} X = \widetilde{F}^{k\prime} X / T$ is the corresponding factor loadings matrix.

The solution to the optimization problem in (3.7) is not unique. If $N < T$, it becomes computationally advantageous to concentrate out $\overline{F}^k$ and minimize (3.7) subject to $\overline{\Lambda}^{k\prime} \overline{\Lambda}^k / N = I_k$. This minimization is the same as maximizing $tr[\Lambda^{k\prime} X' X \Lambda^k]$, the solution of which is to set $\overline{\Lambda}^k$ 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 $\overline{F}^k = X' \overline{\Lambda}^k / N$. $\widetilde{F}^k$ and $\overline{F}^k$ span the same column spaces, hence for forecasting purposes, they can be used interchangeably depending on which one is more computationally efficient. Given $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$, let $\widehat{V}(k) = (NT)^{-1} i = 1N \sum Tt = 1 \sum (x_{it} - \widetilde{\Lambda}_i^{k\prime} \widetilde{F}_t^k)^2$ be the sum of squared residuals from regressions of $X_i$ on the $k$ factors, $\forall i$. A penalty function for over fitting, $g(N, T)$, is chosen such that the loss function

$$IC(k) = \log(\widehat{V}(k)) + kg(N, T) \tag{3.8}$$

can consistently estimate $r$. Let $kmax$ be a bounded integer such that $r \leq k \max$. Bai and Ng (2002) propose three versions of the penalty function $g(N, T)$. Namely $g_1(N, T) =$

$\left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right)$, $g_2(N,T) = \left(\frac{N+T}{NT}\right)\log C_{NT}^2$, and $g_3(N,T) = \left(\frac{\log(C_{NT}^2)}{C_{NT}^2}\right)$, all of which lead to consistent estimation of $r$. In our empirical and Monte Carlo experiments, we use $g_2(N,T)$. Of note is that we tried the other penalty functions above, and our results were qualitatively the same. However, Bai and Ng (2002), as well as others, have shown that in certain contexts, results are sensitive to the choice of penalty function. Hence, (3.8) becomes:

$$IC(k) = \log(\widehat{V}(k)) + k(\frac{N+T}{NT})\log C_{NT}^2$$

where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. The consistent estimate of the true number of factors is then:

$$\widehat{k} = \arg\min_{0\leq k\leq k\max} IC(k), \tag{3.9}$$

and $\lim_{N,T\to\infty}\text{Prob}[\widehat{k} = r] = 1$ if $g(N,T) \to 0$ and $C_{NT}^2 \cdot g(N,T) \to \infty$ as $N,T \to \infty$, as shown in Bai and Ng (2002).

## 3.3 Using Proxies In Place of Factors for Prediction

### 3.3.1 Prediction using factors

Reconsider the general equation (3.3), $y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}$. As mentioned above, and shown in Stock and Watson (2002b) and Bai and Ng (2005), under a set of moment conditions on $(\varepsilon, e, F^0)$ and an asymptotic rank condition on $\Lambda$, if the space spanned by $F_t$ can be consistently estimated, then $\sqrt{T}$ consistent estimates of $\alpha$ and $\beta$ are obtainable. Under a similar set of conditions, it is also possible to obtain $\min[\sqrt{N}, \sqrt{T}]$ consistent forecasts if $\sqrt{T/N} \to 0$ as $N,T \to \infty$. Let $z_t = (F_t', W_t')'$; $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$, for any $h > 0$; and let $z_t$ and $\varepsilon_t$ be independent of the idiosyncratic errors $e_{is}$, $\forall i, s$. If $F_t$ is observable and $\alpha$ and $\beta$ are known, based on the above assumption that the mean of $\varepsilon_{t+h}$ conditional on past information is zero, the conditional mean and minimum mean square error forecast of $y_{T+h}$ is given by:

$$y_{T+h|T} = E(y_{T+h}|z_T, z_{T-1}, \ldots) = \alpha' F_T + \beta' W_T \equiv \delta' z_T$$

Such a prediction is not feasible, however, since $\alpha$, $\beta$ and $F_t$ are all unobserved. The feasible prediction that replaces the unknown objects by their estimates is:

$$\widehat{y}_{T+h|T} = \widehat{\alpha}'\widetilde{F}_T + \widehat{\beta}' W_T = \widehat{\delta}'\widehat{z}_T, \tag{3.10}$$

where $\widehat{z}_t = (\widetilde{F}_t', W_t')'$. Here, $\widehat{\alpha}$ and $\widehat{\beta}$ are the least squares estimates obtained from regressing $y_{t+h}$ on $\widetilde{F}_t$ and $W_t$, $t = 1, \ldots, T-h$. We suppress the $k$ superscript on $\widetilde{F}_t^k$ because we assume we have consistently estimated the number of factors underlying the dataset. The factors, $F_t$, are estimated from $x_{it}$ by the method of principal components, as discussed above. As the objective is to forecast $y_{T+h}$, a crucial aspect of our analysis is the distribution of the forecast error. As explained in detail in Bai and Ng (2006a), since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, it follows that the forecast error is:

$$\widehat{\varepsilon}_{T+h} \equiv \widehat{y}_{T+h|T} - y_{T+h} = (\widehat{y}_{T+h|T} - y_{T+h|T}) - \varepsilon_{T+h}$$

If $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, then:

$$\widehat{\varepsilon}_{T+h} \sim N(0, \sigma_\varepsilon^2 + var(\widehat{y}_{T+h|T})) \tag{3.11}$$

where

$$var(\widehat{y}_{T+h|T}) = \frac{1}{T}\widehat{z}_T' Avar(\widehat{\delta})\widehat{z}_T + \frac{1}{N}\widehat{\alpha}' Avar(\widetilde{F}_T)\widehat{\alpha}. \tag{3.12}$$

Here, $var(\widehat{y}_{T+h|T})$ reflects both parameter uncertainty and regressor uncertainty. In large samples, $var(\widehat{\varepsilon}_{T+h})$ is dominated by $\sigma_\varepsilon^2$. If we ignore $var(\widehat{y}_{T+h|T})$, $\sigma_\varepsilon^2$ alone will underestimate the true forecast uncertainty for finite $T$ and $N$. Let us now assume for a moment that $F_t$ is observable. The feasible prediction of $y_{T+h}$ would then be $\overline{y}_{T+h|T} = \overline{\alpha}' F_T + \overline{\beta}' W_T = \overline{\delta}' z_T$, where $\overline{\alpha}$ and $\overline{\beta}$ are the least squares estimates obtained from regressing $y_{t+h}$ on $F_t$ and $W_t$. Once again, since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, the forecast error is:

$$\overline{\varepsilon}_{T+h} = \overline{y}_{T+h|T} - y_{T+h} = (\overline{y}_{T+h|T} - y_{T+h|T}) - \varepsilon_{T+h}$$

If $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, then

$$\overline{\varepsilon}_{T+h} \sim N(0, \sigma_\varepsilon^2 + var(\overline{y}_{T+h|T})), \tag{3.13}$$

where

$$var(\overline{y}_{T+h|T}) = \frac{1}{T} z_T' Avar(\overline{\delta}) z_T. \tag{3.14}$$

Thus, and as discussed by Bai and Ng (2006a), when comparing $var(\overline{y}_{T+h|T})$ with $var(\widehat{y}_{T+h|T})$, it is clear that estimating the factors increases the forecast error variance, $var(\widehat{y}_{T+h|T})$, by $\frac{1}{N}\widehat{\alpha}' Avar(\widetilde{F}_T)\widehat{\alpha}$. Of course, if we could observe the factors instead of estimating them, we would reduce the forecast error variance from (3.11) to (3.13). In finite samples, this may

yield important prediction error variance reduction. It is for this reason that we consider replacing the factors in (3.10) with observable variables that closely proxy the factors. The approach taken in order to do this involves implementing a "first stage" factor analysis in which proxies are formed using the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006b). In a "second stage", the observable proxies are used in the construction of a prediction model. In this way, all estimation error associated with the factor analysis and proxy selection is essentially "hidden" in the first stage, and does not directly manifest itself in the "second stage" prediction models and prediction errors. Put another way, we are trading-off "estimated factor uncertainty" for "variable selection uncertainty" (see introduction for further discussion). Of course, issues related to "pre-testing" and sequential testing bias still arise. Nevertheless, in our prediction experiments we attempt to quantify through finite sample experiments the potential gains to using the "proxy" approach.

### 3.3.2 Using the $A(j)$ and $M(j)$ tests of Bai and Ng (2006b) to uncover factor proxies

For a detailed theoretical discussion of the results presented in this subsection, the reader is referred to Bai and Ng (2006b). Here, we draw heavily on aspects of that paper that are relevant to our empirical implementation. Note that while Bai and Ng (2006b) suggest using the $A(j)$ and $M(j)$ statistics to assess whether key business cycle indicators approximate the latent factors, we use the $A(j)$ and $M(j)$ statistics to select factor proxies for subsequent use in prediction models. The $A(j)$ statistic depends on the actual size of a t-test. The $M(j)$ test is based on a measure of the distance between observed variables and factor estimates thereof.

Suppose we observe $G'$, a $(T \times m)$ matrix of observable economic variables that could potentially proxy the latent factors (i.e., $G$ is an $m \times T$ matrix). 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 $G_{jt} = \gamma_j' \widetilde{F}_t + \rho_t$. Let $\widehat{\gamma}_j$ be the least squares estimate of $\gamma_j$ and let $\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$. The test is carried out by constructing the

following t-statistic:

$$\tau_t(j) = \frac{(\widehat{G}_{jt} - G_{jt})}{(\widehat{var}(\widehat{G}_{jt}))^{1/2}} \tag{3.15}$$

where

$$\begin{aligned} \widehat{var}(\widehat{G}_{jt}) &= \frac{1}{N}\widehat{\gamma}_j'\widetilde{D}^{-1}\left(\frac{\widetilde{F}'\widetilde{F}}{T}\right)\widetilde{\Gamma}_t\left(\frac{\widetilde{F}'\widetilde{F}}{T}\right)\widetilde{D}^{-1}\widehat{\gamma}_j \\ &= \frac{1}{N}\widehat{\gamma}_j'\widetilde{D}^{-1}\widetilde{\Gamma}_t\widetilde{D}^{-1}\widehat{\gamma}_j, \end{aligned} \tag{3.16}$$

and $\widetilde{\Gamma}_t$ is defined below. The last step above is due to the normalization that $\widetilde{F}'\widetilde{F}/T = I_{\widehat{k}}$. Once again, $\widetilde{D}$ 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 $\widehat{G}_{jt}$ converges to $G_{jt}$ at rate $\sqrt{N}$, Bai and Ng (2006b) show that the limiting distribution of $\sqrt{N}(\widehat{G}_{jt} - G_{jt})$ is asymptotically normal and hence $\tau_t(j)$ has a standard normal limiting distribution. Consistent choices for the the $\widehat{k} \times \widehat{k}$ matrix $\widetilde{\Gamma}_t$ include the following:

$$\widetilde{\Gamma}_t^1 = \frac{1}{n}i = 1n\sum nj = 1\sum\widetilde{\Lambda}_i\widetilde{\Lambda}_j'\frac{1}{T}t = 1T\sum\widetilde{e}_{it}\widetilde{e}_{jt}\ ,\ \ \forall t, \tag{3.17}$$

$$\widetilde{\Gamma}_t^2 = \frac{1}{N}Ni = 1\sum\widetilde{e}_{it}^2\widetilde{\Lambda}_i\widetilde{\Lambda}_i', \tag{3.18}$$

and

$$\widetilde{\Gamma}^3 = \widehat{\sigma}_e^2\frac{\widetilde{\Lambda}'\widetilde{\Lambda}}{N}, \tag{3.19}$$

where $\widehat{\sigma}_e^2 = \frac{1}{NT}i = 1N\sum Tt = 1\sum\widetilde{e}_{it}^2$, $\widetilde{e}_{it} = x_{it} - \widetilde{\Lambda}_i'\widetilde{F}_t$ and $\frac{n}{\min[N,T]} \to 0$ as $N, T \to \infty$. In our Monte Carlo simulation and our empirical analysis, we choose $n = \min\{\sqrt{N}, \sqrt{T}\}$. Equation (3.17) allows cross-section correlation but assumes time-series stationarity of $e_{it}$. This covariance estimator is a HAC type estimator because it is robust to cross-correlation (see Bai and Ng (2006a) for complete details). Equation (3.18) allows for time-series heteroskedasticity, but assumes no cross-sectional correlation of $e_{it}$. Equation (3.19) assumes no cross-sectional correlation and constant variance, $\forall i$ and $\forall t$. 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 (3.18) and (3.19) are useful even if residual cross-correlation is genuinely present.

As mentioned earlier, $\tau_t(j)$ in (3.15) has a standard normal limiting distribution. Let $\Phi_\xi^\tau$ be the $\xi$ percentage point of the limiting distribution of $\tau_t(j)$. The hypothesis test based on

the t-statistic in (3.15) 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, our first proxy selection method is based upon the following statistic:

$$A(j) = \frac{1}{T}t = 1T\sum 1(|\tau_t(j)| > \Phi_\xi^\tau). \tag{3.20}$$

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 $2\xi$) as $T \to \infty$. 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 $2\xi$ 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 $2\xi$. 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. Given the choice of the significance level $2\xi$, the $A(j)$ statistic incorporates some degree of robustness by allowing $G_{jt}$ to deviate from $\widehat{G}_{jt}$ for a specified number of time points.

The second method for selecting the proxies considers the statistic:

$$M(j) = 1 \leq t \leq T\max|\tau_t(j)|, \tag{3.21}$$

which is based on a measure of how far the $\widehat{G}_{jt}$ curve is from $G_{jt}$. If $e_{it}$ is serially uncorrelated, then:

$$P(M(j) \leq x) \approx [2\Phi(x) - 1]^T, \tag{3.22}$$

where $\Phi(x)$ is the cdf of a standard normal random variable. From (3.21) and (3.22), we can perform a test to determine whether the time series of a candidate variable is a good proxy for the latent factors. For instance, suppose we are given a significance level $2\xi$ and a sample

of size $T$ from a particular candidate variable, $G_j$. From the right hand side of (3.22), we can calculate the corresponding critical value, $x$, for the test. For the same sample, we calculate $M(j)$ from (3.21) and conclude that $G_j$ is a good proxy if $M(j) \leq x$, and a bad proxy otherwise. The test based on the $M(j)$ statistic is thus stronger than the selection method based on the $A(j)$ statistic, as the $M(j)$ test gives a sharp decision rule. However, the $M(j)$ test has at least one disadvantage. It requires $e_{it}$ to be serially uncorrelated. We ignore this requirement in our experimental analysis. It should be noted that $x$ increases with the sample size, $T$. Depending on the nature of the observed sample, this fact could either preserve or reduce the power of the $M(j)$ test.[1]

The proxies selected depend on the structure of the $\widehat{k} \times \widehat{k}$ matrix $\widetilde{\Gamma}_t$ that we use in (4.8). For a given proxy selection method, if the choice of $\widetilde{\Gamma}_t^1, \widetilde{\Gamma}_t^2, \widetilde{\Gamma}^3$ used in calculating (4.8) all produce the same proxies, it could mean that the respective assumptions associated with the use of $\widetilde{\Gamma}_t^1, \widetilde{\Gamma}_t^2, \widetilde{\Gamma}^3$ might not be very relevant, empirically. We found no gains, in our experimental set-up, to using $\widetilde{\Gamma}_t^1$ and $\widetilde{\Gamma}^3$, and hence all reported results are for the case where we use $\widetilde{\Gamma}_t^2$

Finally, Shanken (1992) points out that it is theoretically crucial for the observed selected proxies to span the same space as the $r$ latent factors, as discussed above. We nevertheless consider versions of the above methods where the number of factors is greater than the number of proxies, given the principle of parsimony in forecasting.

### 3.3.3 Smoothed $A(j)$ and $M(j)$ tests for selecting factor proxies

The $A(j)$ and $M(j)$ statistics discussed above may yield a different set of proxies at each point in time when used to construct a sequence of recursive forecasts. Namely, if the information set used in the parameterization of a prediction model is updated prior to the construction of each new forecast for some sequence of $E$ ex ante predictions, then the "first stage" factor analysis discussed above may yield a sequence of $E$ different vectors of factor proxies. Thus, in addition to the $A(j)$ and $M(j)$ proxy selection methods, we also consider a version of these methods where the sample period in our empirical analysis is broken

[1]Note that we also considered the confidence interval approach of Bai and Ng (2006); but it did not perform better than the above methods.

into three subsamples ($R_1, R_2$, and $E$, such that $T = R_1 + R_2 + E$). The first subsample is used to estimate 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. As some of our models (such as the autoregressive model) select the number of lags and re-estimate all parameters prior to the formation of each new prediction, this smoothed approach is at a disadvantage, in the sense that it is static (i.e., the set of proxies is fixed throughout the forecast experiment). However, loading parameters for the proxies are still re-estimated prior to the formation of each new recursive prediction. Of course, the potential advantage to this approach is that noise across the proxy selection process is suppressed.

## 3.4 Empirical Methodology

In order to assess the performance of factor proxy based prediction models, we focus our attention on direct multistep-ahead predictions. Forecasts are generated as $h$-step ahead predictions of $y_t$, say. Namely, we predict $y_{t+h} = \log\left(\frac{Y_{t+h}}{Y_{t+h-1}}\right)$, where $Y_t$ is the variable of interest.[2] Our approach is to compare the performance of factor based predictions with a host of proxy based predictions as well as various "strawman" predictions. For the "strawman" forecast models, we use an autoregressive ($AR(p)$) model (with lags selected using the Schwarz Information Criterion (SIC)) and a random walk model. The "strawman" models are included because they serve as parsimonious benchmarks that are often difficult to outperform. In Table 3.1, we provide the specifications and brief descriptions of all of the forecast models examined.

We consider two classes of proxy forecasts models. The first class of models, which we

[2]While cummulative changes are very useful in prediction contexts, we predict the growth rate from one period to the next, $y_{t+h} = log(Y_{t+h}/Y_{t+h-1})$ instead of the cumulative change, $y_{t+h} = log(Y_{t+h}/Y_t)$. Our approach is in accord with the Federal Reserve Economic Database (FRED), where the same period on period growth rates are reported. We have experimented also with cummulative growth rates, with similar empirical findings to those reported here.

call "ordinary" proxy forecast models, include Model *4* - Model *7*. With these models, proxies are re-selected recursively, prior to the construction of each $h$-step ahead prediction. Let $\{A(j)\}_{j=1}^{m}$, be a set of $A(j)$ statistics calculated for each candidate proxy variable $j$. As suggested above, in this particular paper, we set $m = N$; but this need not always be the case. Define:

$$S^A = \{G_{j_1}^A, \ldots, G_{j_{\widehat{k}}}^A\} \tag{3.23}$$

where $\widehat{k} \leq m$ and $|A(j_1) - 2\xi| \leq |A(j_2) - 2\xi| \leq \ldots \leq |A\left(j_{\widehat{k}}\right) - 2\xi|$.Here, $S^A$ is the set of $\widehat{k}$ proxy variables selected via implementation of the $A(j)$ test. Further, define $G_{j_1}^A$ as the "best" possible proxy as determined by the $A(j)$ while $G_{j_2}^A$ is the next "best" proxy, and so on. Recall that $G_j$ is an observable time series variable, such as the CPI or the Federal Funds Rate. Turning next to proxies selected via implementation of the $M(j)$ test, define:

$$S^M = \{G_j \in G \mid M(j) \leq x\}, \quad j = 1, \ldots, m.$$

Here, $S^M$ is a set of proxies selected by the $M(j)$ test. The number of proxy variables selected at each recursive stage is indeterminate. Furthermore, the selected proxies are not ranked. For Model *6*, where the $M(j)$ test is used to select a single proxy, our approach is to select the proxy in the set $S^M$ that is associated with the smallest value of $M(j)$.

The second class of models, which we call "smoothed" proxy forecast models, are discussed in Section 3.3, and include Model *8* - Model *15*. The proxies used in these models are still based on implementing the $A(j)$ and $M(j)$ statistics as discussed above. The factors and the proxies are estimated recursively, just as in Models *1, 4-7,* but this is done starting with $R_1$ observations and ending with $R_1 + R_2$ observations. The "smoothed" proxies are selected as the $\widehat{k}$ proxies that are "most frequently" picked by the $A(j)$ and $M(j)$ tests. Thereafter, all proxies are fixed, although their "weights" in the prediction models are still re-estimated recursively, prior to the construction of each of the $E$ ex-ante forecasts. To differentiate between proxies picked using the "ordinary" and "smoothed" versions of the tests, we define $S^{A*}$ and $S^{M*}$ to be the "smoothed" versions of $S^A$ and $S^M$. The ex-ante prediction period, $E$, is the same for all models in our empirical experiments.

In order to evaluate forecast performance, we compare mean squared forecast errors (MSFEs) defined as $\frac{1}{E}\sum_{t=R-h+1}^{T-h}\left(\widehat{y}_{t+h} - y_{t+h}\right)^2$, where $R = R_1 + R_2$. We also carry out Diebold

and Mariano (DM: 1995) predictive accuracy tests. Let $\{\widehat{y}_{1,t}\}_{t=R-h+1}^{T-h}$ and $\{\widehat{y}_{2,t}\}_{t=R-h+1}^{T-h}$ be two forecasts of the time series $\{y_t\}_{t=R-h+1}^{T-h}$. The "benchmark" is Model *1* (i.e., the factor model), and is used to generate $\{\widehat{y}_{1,t}\}_{t=R-h+1}^{T-h}$, while Models *2-15* are used to generate $\{\widehat{y}_{2,t}\}_{t=R-h+1}^{T-h}$. Since the "benchmark" contains estimated factors and the alternative models contain no estimated factors, the "benchmark" and alternative models are non-nested. The corresponding out-of-sample forecast errors are $\{\widehat{\varepsilon}_{1,t}\}_{t=R-h+1}^{T-h}$ and $\{\widehat{\varepsilon}_{2,t}\}_{t=R-h+1}^{T-h}$. The null hypothesis of equal forecast accuracy for two forecasts is given by $H_0 : E[\widehat{\varepsilon}_{1,t}^2] = E[\widehat{\varepsilon}_{2,t}^2]$ or $H_0 : E[\widehat{d}_t] = 0$,where $\widehat{d}_t = \widehat{\varepsilon}_{1,t}^2 - \widehat{\varepsilon}_{2,t}^2$ is the loss differential series. The DM test statistic is $DM = E^{-1/2}\frac{\overline{d}}{\sqrt{\widehat{\sigma}_d^2}}$, where $\overline{d} = \frac{1}{E}\sum_{t=R-h+1}^{T-h}\widehat{d}_t$, and $\widehat{\sigma}_d^2$ is a HAC standard error for $\widehat{d}_t$. Since the forecast models are non-nested, and assuming that parameter estimation error vanishes, the $DM$ test statistic has a $N(0,1)$ limiting distribution. Finally, given this setup, a negative $DM$ t-stat indicates that the factor model yields a lower point MSFE. For further discussion of parameter estimation error and nestedness issues in the context of predictive accuracy tests, the reader is referred to Corradi and Swanson (2002, 2006a, 2006b).

### 3.5 Data

The dataset used to estimate the factors is the same as that used in Stock and Watson (2005), which can be obtained at *http://www.princeton.edu/˜mwatson*. This dataset contains 132 monthly time series for the United States for the entire period from 1960:1 to 2003:12, hence $N = 132$ and $T = 528$ observations. The series were selected to represent 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. Most of the series were taken from the Global Insights Basic Economic Database or The Conference Board's Indicators Database (TCB). Others were calculated by Stock and Watson with information from either Global Insights or TCB or both. The theory outlined assumes that the panel dataset used to estimate the factors is $I(0)$. To achieve this, some of the 132 series were subjected to transformations by taking logarithms and/or first differencing. In general, logarithms were taken for all nonnegative series that were not already in rates or

percentage units (see Stock and Watson (2002a,2005) for complete details). After these transformations were carried out, all series were further standardized to have sample mean zero and unit sample variance. Using the transformed data set, denoted above by $X$, the factors are estimated by the method of principal components. As mentioned earlier, in our implementation, the set of candidate proxies for the factors $G'$, will be the same as $X$. Although this need not be the case, it is done mainly because $X$ represents the biggest set of (standardized and stationary) observable time series variables available to us. We perform real-time forecasts of 7 of the 8 major monthly macroeconomic time series studied in Stock and Watson (2002a). The four real variables we concentrate on are total industrial production (IP), real personal income less transfers, real manufacturing and trade sales and the number of employees on nonagricultural payrolls. The three price series considered are the consumer price index (CPI), the personal consumption expenditure implicit price deflator (PCED) and the producer price index for finished goods (PPI). All of these variables are expressed in log-differences (i.e., monthly growth rates).[3]

### 3.6 Monte Carlo Experiment

Table 3.2 contains the results from a small Monte Carlo experiment used to assess the finite sample forecast performance of the $A(j)$ and $M(j)$ tests. In the empirical panel dataset spanning 1960:1 to 2003:12 discussed in Section 3.5 above, $\widehat{k} = 13$ factors are consistently estimated using the methodology of Bai and Ng (2002). For this reason, we assume there are 13 factors underlying our simulated dataset and set $r = 13$. The simulated dataset also has the same dimensions as the empirical dataset discussed in the next section. Hence, we set $N = 132$ and $T = 528$. Each of the thirteen latent factors is generated as

$$F_{kt} = \nu_k F_{kt-1} + u_{kt}, \tag{3.24}$$

where $0.6 \leq \nu_k \leq 0.8$, $u_{kt} \sim N(0,1)$, and $u_{kt}$ is uncorrelated with $u_{jt}$, for $k \neq j$, $k, j = 1, \ldots, r$. $F_t = (F_{1t}, \ldots, F_{rt})'$, $\Lambda_{ik} \sim N(0,1)$, and $e_{it}$ is uncorrelated with $e_{jt}$, for $i \neq j$,

[3]Note that Stock and Watson (1999, 2002a) model some of our price variables as $I(2)$ in logarithms. However, they find little discrepancy in performance under $I(1)$ and $I(2)$ assumptions for factor forecasts of our three target price variables. For this reason, we limit our analysis by assuming that our price variables as well as other variables in $X$ are $I(1)$ in logarithms (see Section 10 for further details). In all other respects, our dataset is the same as that used by Stock and Watson (2005).

$i, j = 1, \ldots, N$, $t = 1, \ldots, T$. Following Bai and Ng (2002), the simulated panel dataset is generated as

$$x_{it} = \Lambda_i' F_t + \sqrt{\eta} e_{it}, \tag{3.25}$$

for $i = 1, \ldots, 119$, where $\eta$ is a measure of the variance of the idiosyncratic errors, $e_{it}$, relative to the common component, $\Lambda_i' F_t$. More specifically, for $i = 1, \ldots, 39$, we make the idiosyncratic errors homoskedastic and set $e_{it} \sim N(0, 1)$. We introduce heteroskedasticity into the variables for which $i = 40, \ldots, 79$ and let

$$e_{it} = \begin{cases} e_{it}^1 & if t is even \\ e_{it}^1 + e_{it}^2 & if t is odd \end{cases} \tag{3.26}$$

where $e_{it}^1$ and $e_{it}^2$ are independent $N(0, 1)$ (see Bai and Ng (2002)). For $i = 80, \ldots, 119$, the idiosyncratic component of (3.25) is generated as an $MA(1)$ process such that

$$e_{it} = 0.6 e_{it-1} + e_{it}^3 \tag{3.27}$$

and $e_{it}^3 \sim N(0, 1)$. Define a variable to be a good proxy if it is a linear combination of the underlying latent factors (see Bai and Ng (2006b) for complete details). Thus, for $i = 120, \ldots, 132$, the proxy variables are generated as

$$x_{it} = \Lambda_i' F_t \tag{3.28}$$

Since the generated factors in (3.24) are assumed to be latent, they are not wholly included in the simulated panel dataset. The above setup ensures that four separate DGPs generate a total of 132 simulated variables. Ex ante forecasts are constructed for four variables. In Table 3.2, the target variables labelled "Homoskedastic", "Heteroskedastic" and "$MA(1)$" are all generated from (3.25). However, the corresponding idiosyncratic errors are specified by i.i.d. $N(0, 1)$, (3.26) and (3.27) respectively. Let $x_t^p = (x_{120t}, \ldots, x_{132t})$, $e_{it}^4 \sim N(0, 1)$ and $\Omega_{il} \sim N(0, 1)$ for $l = 1, \ldots, 13$, then the target variable labelled "Proxy (Homoskd.)" is generated by

$$y_p = \Omega_i' x_t^p + e_{it}^4 \tag{3.29}$$

The difference between (3.29) and (3.25) is that in (3.29), $x_t^p$ are observed and can be selected by the $A(j)$ or $M(j)$ tests as regressors in a forecast model for the respective target

variable. Of course, there is still no guarantee that they will be selected; rather this is the only case where the true regressor variables are actually in the panel dataset and can be selected. This is an important case, and defines the case we are most interested in. On the contrary in (3.25), $F_t$ are not observed and can consequently not be selected as predictors in a forecasting exercise. For each of the four target variables in Table 3.2 and Table 3.3, the last third of simulated values are recursively forecasted. Since there are 528 data points across time in our setup, we effectively use observations from $t = 352$ to $t = 528$ to evaluate forecast performance via examination of the mean squared forecast error (MSFE). Prior observations are used to estimate the forecast models. In strict recursive fashion, all models, factors, number of factors, $k$ and proxies are re-estimated and re-selected for each constructed forecast. Forecasting at time $t + 1$, the panel dataset from $1, \ldots, t$ is also standardized to have mean zero and unit variance before the factors are recursively estimated and proxies selected. In order to make the experiment credible, the model used for the factor forecasts is Model *1* and those used for the "$A(j)$" and "$M(j)$"proxy forecasts are Model *5* and Model *7,* respectively (see Table 3.1 for the specification of these models). From prior work, Model *5* and Model *7* performed worst among all the alternative proxy forecast models specified in Table 3.1, and hence our setup is as "tough as possible" on our approach. It is left to future research to establish whether other model specifications discussed in this paper that perform better in our empirical experiments also perform better in Monte Carlo simulation experiments.

We perform the same forecast evaluation exercise for a subsets of $N = 40$ and $N =$132. The 40 variable subsets are randomly selected from the original 132 simulated variables under the constraint that at least 2 and at most 7 proxies as defined by (3.28) are selected. Forecast horizons of $h = 1, 12$ are considered. The entire monte carlo experiment is conducted for 250 iterations and at each iteration, for $N = 40$ and $N = 132$, we calculate the MSFE from 176 recursive forecasts for $t = 352, \ldots, 528$.

The numerical entries in Table 3.2 represent the fraction of times (out of 250 Monte Carlo iterations) that the proxy forecasts have a lower MSFE than the factor forecasts. Regardless of the number of variables in the panel dataset or the forecast horizon, the proxy forecasts outperformed the factor forecasts about 50% of the time in almost all cases.

This is significant as it demonstrates that the worst performing proxy forecast models equally match the factor forecasts. Under $h = 1$, entries of 0.720, 0.795, 0.880 and 0.895 for "Proxy (Homoskd.)" indicate that the proxy forecasts strongly outperform the factor forecasts. This particular outcome is as might be expected, given that this is the case where the $A(j)$ and $M(j)$ test statistics are afforded the possibility of selecting the truly correct elements of $x_t^p$ used to generate $y_p$ in (3.29) and suggests that our approach is working as desired. However, under $h = 12$ for the same target variable "Proxy (Homoskd.)", the proxy forecasts perform just as well as the factor forecasts. One explanation for this result might be that as the forecast horizon gets longer, the informational content in the proxies deteriorates faster relative to that of the factors.

The entries in Table 3.3 not in parenthesis represent the mean of the various MSFEs across Monte Carlo iterations. The standard deviations of the MSFEs are reported in parentheses. From Table 3.3, proxy forecasts constructed from the $A(j)$ or $M(j)$ statistic marginally outperform the factor forecasts most of the time in terms of the mean of the MSFEs. However, the equal performance of the factor and proxy forecasts in Table 3.2 is demonstrated in Table 3.3 by the fact that the mean of the proxy MSFEs is generally only slightly less than the mean of the factor MSFEs.

Overall, these results are interesting, and suggest that our prediction/proxy approach outperforms a standard factor approach in favorable cases, and perform equally as well in non-favorable cases.

## 3.7 Empirical Findings

In this section, we discuss the results of a series of prediction experiments using the dataset discussed above, and applying the models outlined in Table 3.1 to construct sequences of recursive ex-ante $h$-step ahead predictions. The dataset consists of 132 variables (see Section 3.5), and data are available for the period 1960:1-2003:12. Furthermore, predictions are constructed for the period 1989:5-2003:12. Please see Section 3.4 for complete details concerning the strategy used to specify and estimate the prediction models prior to forecast construction. For models in which proxies were selected using the $M(j)$ and $A(j)$ tests,

we set $2\xi = 0.05$. Hence we carry out the tests at a 5% significance level. We include 1 autoregressive lag in most of the models because the importance of autoregressive lags in prediction is well established. Furthermore, adding autoregressive terms of the target variable to the basic factor model is a good way to give the factor model a fair chance to "win" our forecasting competition.

Results of our empirical experiments are gathered in Table 3.4 (frequency of selected factor proxies), Table 3.5 (CPI, PCED, and PPI forecasting competition results), and Table 3.6 (Industrial Production, Personal Income; Nonagricultural Employment, Manufacturing and Trade Sales). In Table 3.4, selection frequencies are reported, while in Tables 3.5 - 3.6 MSFEs and $DM$ test statistics are reported. The MSFE values reported for CPI, PCED and Nonagricultural Employment are multiplied by 100,000 and those reported for Producer Price Index, Industrial Production, Manufacturing and Trade Sales and Personal Income are multiplied by 10,000. For the benchmark Model *1* (i.e., the factor model), the only tabular entry for all forecast horizons is the MSFE. With all of the other models (i.e., our alternative models), there are two entries: The top entry is the MSFE and the bottom entry in parenthesis is the DM t-statistic. As mentioned earlier, a positive DM statistic value indicates that the alternative model has a MSFE that is lower than the benchmark, while a negative statistic value indicates the reverse. Entries in bold signify instances where the alternative model outperforms the factor model as determined by a point MSFE comparison. Boxed MSFE entries represent the lowest MSFE value among all the models for a particular forecast horizon. DM statistic entries with a $*$ indicate instances where the respective alternative model significantly outperforms the factor model at a 10% significance level, whereas for entries with a † sign, the factor model significantly outperforms the alternative model at a 10% significance level. We now provide a number of conclusions based on the tables.

Upon inspection of Table 3.5, it is clear that the benchmark factor model (i.e., Model *1)* significantly outperforms most of the alternative models in the forecast of CPI and PCED. This point is supported by the overwhelming number of DM test rejections in Panels A and B of Table 3.5. While the benchmark still yields the lower MSFE in many pairwise comparisons when examining PPI results (see Panel C of the table), the DM test null of

equal predictive accuracy is not frequently rejected.

A key exception to the above conclusion that the benchmark model yields superior predictions is in the case of Models 12-15. From Table 3.1, recall that these are autoregressive models with exogenous variables (ARX). The lags of the ARX models are selected by the SIC and the exogenous variables are based on smoothed versions of the $A(j)$ and $M(j)$ tests. For $h = 1, 3, 12$, these models not only frequently yield lower point MSFEs than the benchmark, but the difference in performance is often significant. Across all 3 panels and 3 forecast horizons (i.e., 9 variable/horizon combinations), it is interesting to note that one or many of Models 12-15 are "MSFE-best" 7 times. Furthermore, of these 7 "wins" it is Model 12 that yields the lowest MSFE in 4 instances. Thus, we have direct evidence that the parsimonious single proxy smoothed $A(j)$ model fares very well when compared not only to the benchmark, but also to other models which yield lower MSFEs than the benchmark. This suggests that while the factor approach is very useful, often beating the pure autoregressive and other linear models when used for predicting price variables, a parsimonious version of the smoothed $A(j)$ factor proxy approach performs the best, overall. Thus, as pointed out by Bai and Ng (2006c), parsimony is still important. This is even true in the context of ordinary proxy models (Models 4-7), as choosing one proxy rather than $\widehat{k}$ proxies often yields the lowest MSFE model.

Interestingly, in Table 3.5, the above conclusions hold for $h = 1, 3, 12$ and not for $h = 24$. Indeed, it appears that all models perform quite poorly for $h = 24$, with the notable exception of the benchmark, which clearly outperforms virtually all competitors in all price variable cases when $h = 24$. Thus, at the longest forecast horizons, we have evidence that our simple factor proxy approaches are not faring well at all.

Turning now to Table 3.6, the above conclusions still hold, with the exception that many other alternative models, and not just Models 12-15, are point MSFE "better" than the benchmark. Summarizing the results in Table 3.6, the benchmark model does yield the lowest MSFE for 3 of the 4 variables when $h = 1$ and for 1 variable when $h = 3$, although the DM test null is not rejected in any of these cases. Furthermore, for all remaining horizon/variable combinations, the benchmark does not yield the lowest MSFE. Indeed, in all but one of these other cases, factor proxy approaches yield the lowest MSFE (the sole

exception is a random walk "win" for Manufacturing and Trade Sales when $h = 3$).

Given the above results, it is of interest to tabulate which factor proxies were used in our prediction experiments. This is done in Table 3.4, where factor proxies that are (most frequently) selected using the $A(j)$ and $M(j)$ test and the frequencies with which they are selected are reported. The second column under "Trans" indicates the data transformation that was performed to induce data stationarity. As is evident, S&P's Common Stock Price Index, Industrials; S&P's Common Stock Price Index, Composite; Dividend Yields, a 1-Year Bond Rate; and Housing Starts are the five most common proxies selected by both $A(j)$ and $M(j)$. Structural change could account for some of the proxies being selected less frequently than the five above proxies. Clearly, the importance of proxies may in some cases depend on the period in history represented by the data. However, it is interesting that a variety of factor proxies are "picked" across our entire ex-ante prediction period.

The diagrams in Panel 1- 3 of Figure 3.1 are time series plots of the first three estimated factors (i.e., the most important factors for explaining the variability in our panel dataset). Panels 4-6 are time series plots of the three most frequently selected proxies based on use of the $A(j)$ and $M(j)$ test statistics. The S&P Common Stock Price Index does proxy the estimated factors to some extent, although the relatively high level of noise in the S&P variable does appear to obscure this fact to a certain degree. The Housing Starts, Nonfarm variable (which has less noise - see Panel 6) better illustrates the close relationship between the estimated factors and selected proxies. Results in Table 3.4 indicate that almost all three proxies in Figure 3.1 are selected 100% of the time by both the $A(j)$ and $M(j)$ statistics although the $M(j)$ test has more power than the $A(j)$ test. The lone exception to this is the Housing Starts, Nonfarm variable which is selected 95% of the time by the $M(j)$ test. This suggests how strongly the three variables proxy the underlying factors. In addition, one gets a "sense" of the robustness of the $A(j)$ and $M(j)$ test statistics in consistently selecting good proxies, since the uderlying factors are re-estimated at each recursive iteration.

In closing, we note that factor proxies appear useful for prediction. Additionally, since factors are unobserved, analyzing and studying them on their own can be quite difficult. For instance, in our context is not clear how relevant it is to study the evolution of the individual factors over time because prior to each new prediction, the factors are re-estimated.

Creating a clearly defined historical path for a factor is consequently complicated. The ability to proxy the unobserved factors with observed variables enables us to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables.

## 3.8 Recent Advances in the Construction of Diffusion Indices

In this section, we briefly highlight some of the most recent work relating to diffusion index (factor) models. Many of these ideas could potentially be applied to the issues discussed in this paper, although we leave that to future research. Some of the concerns raised in this paper such as the use of the same factors and consequently the same proxies to forecast *any* variable are addressed in a number of the papers. For example, Bai and Ng (2006c) offer two refinements to the method of factor forecasting. The current framework is confined to a linear relation between the predictors and the forecasted series. Bai and Ng (2006c) propose a more flexible structure. Their so-called squared principal components approach allows the relationship between the predictors and the factors to be non-linear. They use a non-linear "link" function that involves expanding the set of predictors to include non-linear functions of the observed variables. In this regard, (3.4) can be modified as follows:

$$h(x_{it}) = \vartheta_i' J_t + e_{it},$$

where $h(\cdot)$ is a non-linear link function, $J_t$ are the common factors, and $\vartheta_i$ is the vector of factor loadings. The second order factor model is consequently:

$$x_t^* = \Omega J_t + e_t \tag{3.30}$$

where $x_t^* = \{x_{it}, x_{it}^2\}\forall i$ is an $N^* \times 1$ vector and $N^* = 2N$. Estimation of $J_t$ from (3.30) is done using the usual method of principal components. The forecasting equation in (3.10) remains linear regardless of the form of $h(\cdot)$. The second refinement proposed by Bai and Ng (2006c) takes explicit account of the fact that the ultimate aim is to forecast a specific time series variable, say $y_t$. The authors propose using principal components analysis with a "targeted" subset of the predictors in $X$, which have been tested to have predictive power for $y$. This implies that the set of predictors used to extract the factors change with $y$, the

targeted forecast variable. "Hard" and so-called "soft" thresholding is used to determine which subset of $X$ the factors are to be extracted from. Under "hard" thresholding, a test with a sharp decision rule determines which variables are "in" or "out". With "soft" thresholding, the top variables are kept in the subset of predictors used to extract the factors. The ordering of the predictors is based on the particular soft-thresholding rule. The "soft" thresholding approach is thus related to our "smoothed" test statistic approach to factor proxy selection.

As a reminder, the use of factor models (diffusion indices) involves a two-step approach in which the factors are first estimated from a large panel dataset. The estimated factors are then used as predictors in the forecast models. Although the estimated factors in the first stage are capable of parsimoniously capturing almost all the information in a large dataset, standard tools for specifying the forecast model in the second stage remain unsatisfactory in certain contexts. The specified prediction models are still susceptible to overfitting or underfitting, for example. In this light, Bai and Ng (2006d) suggest a stopping rule for "boosting" that prevents a model from being overfitted with estimated factors or other predictors. Boosting is a procedure that estimates the conditional mean using $M$ stagewise regressions (Bai and Ng (2006d)). The authors also propose two ways to handle lagged predictors: a component-wise approach that treats each lag as a separate variable, and a block-wise approach that treats lags of the same variable jointly. Some important papers on boosting include Schapire (1990), Freund (1995), Friedman (2001) and Buhlmann and Hothorn (2006).

## 3.9 Concluding Remarks

Using Monte Carlo and empirical analysis, we have shown that the $A(j)$ and $M(j)$ statistics of Bai and Ng (2006b) appear to offer an interesting means by which factor proxies for later use in prediction models can be chosen. Indeed, our "smoothed" approaches to factor proxy selection appear to yield predictions that are often superior not only to a benchmark factor model, but also to simple linear time series models which have in many practical applications hitherto been found to be difficult to beat in forecasting competitions. More specifically, we find that our factor proxy models (e.g., see Model 5 and Model 7 in Table

3.1) perform (slightly) better than a standard factor model (Model 1) in our Monte Carlo experiments. The implication is that a policymaker will be better served by using the proxy model. At the very least, the methodology suggested in this paper should perhaps be added to the practitioners "tool-box", and one should examine on a case-by-case basis whether or not proxy observable factors are more effective that standard factors. This is particularly relevant since, unlike the factor model which has estimated regressors, the proxy model uses observed regressors that can act as policy instruments, for example. By using our approach to predictive factor proxy selection, one is able to open up the "black box" often associated with factor analysis, to some extent. This is because one can identify actual variables that can serve as "primitive" building blocks for (prediction) models of a host of other macroeconomic variables. This approach in some cases leads to improved prediction, and may also possibly lead to improved policy analysis if used in policy related prediction modelling.

Table 3.1: **Prediction Models Used in Empirical Experiments**

| |
|---|
| Model 1 (Factor Model): This is the standard factor forecast model: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\alpha}'\widetilde{F}_T + \widehat{\beta}y_T$ |
| Model 2 (Autoregressive Model): This is an $AR(p)$ forecast model, with lags selected by the SIC: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + j = 1p\sum\widehat{\alpha}_j y_{T-j+1}$ |
| Model 3 (Random Walk Model): This is a random walk forecast model: $\widehat{y}_{T+h\|T} = y_T$ |
| Model 4 (Ordinary $A(j)$ - 1 Proxy Model): In this forecast model, the single "best" proxy selected by the $A(j)$ test (i.e., the proxy associated with the $A(j)$ statistic value closest to $2\xi$ in absolute value) is used as the only proxy regressor in the forecast model: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\alpha}G^A_{j_1T} + \widehat{\beta}y_T$ |
| Model 5 (Ordinary $A(j)$ - $\widehat{k}$ Proxies Model): The "best" $\widehat{k}$ factor proxies selected by the $A(j)$ test are used: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\alpha}'S^A_T + \widehat{\beta}y_T$, where $S^A_T = \{G^A_{j_1T}, \ldots, G^A_{\widehat{j_k}T}\}$. |
| Model 6 (Ordinary $M(j)$ - 1 Proxy Model): In this forecast model, the single "best" factor proxy selected by the $M(j)$ test (i.e., the proxy associated with the lowest $M(j)$-statistic) is used as the only proxy regressor in the forecast model: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\alpha}G^M_{jT} + \widehat{\beta}y_T$. Since it is possible for the $M(j)$ test to select no proxies at all, should that scenario occur, the model degenerates to: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\beta}y_T$ . |
| Model 7 (Ordinary $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 6, but $\widehat{k}$ factor proxies selected by the $M(j)$ test are used: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\alpha}'S^M_T + \widehat{\beta}y_T$. |
| Model 8 (Smoothed $A(j)$ - 1 Proxy Model): This forecast model is the same as Model 4, except that the smoothed version of the $A(j)$ test is used (see Section 3.3.3 for further discussion). |
| Model 9 (Smoothed $A(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 5, except that the smoothed version of the $A(j)$ test is used (see Section 3.3.3 for further discussion). |
| Model 10 (Smoothed $M(j)$ - 1 Proxy Model): This forecast model is the same as Model 6, except that the smoothed version of the $M(j$ ) test is used (see Section 3.3.3 for further discussion). |
| Model 11 (Smoothed $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 7, except that the smoothed version of the $M(j$ ) test is used (see Section 3.3.3 for further discussion). |
| Model 12 (Autoregressive plus Smoothed $A(j)$ - 1 Proxy Model): This forecast model is the same as Model 8, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1: $\widehat{y}_{T+h\|T} = \widehat{a}_0 + \widehat{\alpha}G^{A*}_{j_1T}j = 1+p_x\sum\widehat{\beta}_j y_{T-j+1}$. |
| Model 13 (Autoregressive plus Smoothed $A(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 9, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1. |
| Model 14 (Autoregressive plus Smoothed $M(j)$ - 1 Proxy Model) : This forecast model is the same as Model 10, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1. |
| Model 15 (Autoregressive plus Smoothed $M(j)$ - $\widehat{k}$ Proxies Model): This forecast model is the same as Model 8, except that the lag of the autoregressive component is selected by the SIC rather than restricted to 1. |

Note: See Sections 3.3.3 and 3.4 for further discussion of the factor proxy selection methodology used in the construction of the above models.

Table 3.2: **Monte Carlo Experiment Results**

| | | $h = 1$ | | $h = 12$ | |
|---|---|---|---|---|---|
| N | Error Structure | A(j) | M(j) | A(j) | M(j) |
| 40 | Homoskedastic | 0.425 | 0.330 | 0.460 | 0.560 |
| 40 | Heteroskedastic | 0.425 | 0.435 | 0.530 | 0.685 |
| 40 | MA(1) | 0.520 | 0.620 | 0.575 | 0.675 |
| 40 | Proxy (Homoskd.) | 0.720 | 0.795 | 0.390 | 0.450 |
| 132 | Homoskedastic | 0.585 | 0.430 | 0.545 | 0.595 |
| 132 | Heteroskedastic | 0.680 | 0.475 | 0.545 | 0.615 |
| 132 | MA(1) | 0.460 | 0.600 | 0.585 | 0.605 |
| 132 | Proxy (Homoskd.) | 0.880 | 0.895 | 0.585 | 0.620 |

Notes: The numeric entries under "N" indicate the number of variables in the simulated panel dataset. Entries under "A(j)" and "M(j)" indicate the fraction of times that the alternative model (*Model 5* or *Model 7*, repsectively) has a lower MSFE than the benchmark (*Model 1*), in 250 Monte Carlo iterations. Under "Error Structure", we state the forecast "target" variable. "Homoskedastic", "Heteroskedastic" and "$MA(1)$" represent target variables for which the idiosyncratic error, $e_{it}$, in the DGP is an i.i.d. $N(0, 1)$, a heteroskedastic, or a moving average process, respectively. For all three of these cases, the independent variables in the DGP are the latent factors. For "Proxy (Homoskd.)", the idiosyncratic error, $e_{it}$, is i.i.d. $N(0, 1)$; and the independent variables in the DGP are potential "proxy" variables, so that the "A(j)" and "M(j)" in this case might select the "true" proxy, if they perform as desired. See Section 3.6 for complete details.

Table 3.3: **Monte Carlo Experiment Descriptive Statistics**

| | | | $h = 1$ | | | $h = 12$ | |
|---|---|---|---|---|---|---|---|
| N | Error Structure | Factor | A(j) | M(j) | Factor | A(j) | M(j) |
| 40 | Homoskedastic | 52.809 | 53.023 | 53.472 | 60.043 | 60.118 | 60.008 |
| | | (8.882) | (8.933) | (9.219) | (12.770) | (12.826) | (12.822) |
| 40 | Heteroskedastic | 49.111 | 49.481 | 49.725 | 56.925 | 56.928 | 56.685 |
| | | (8.950) | (8.761) | (8.783) | (14.240) | (14.226) | (14.174) |
| 40 | MA(1) | 38.829 | 38.812 | 38.736 | 66.851 | 66.649 | 66.468 |
| | | (6.669) | (6.665) | (6.619) | (16.385) | (16.333) | (16.178) |
| 40 | Proxy (Homoskd.) | 25.751 | 25.569 | 25.466 | 56.548 | 56.953 | 56.681 |
| | | (26.382) | (26.695) | (26.704) | (59.791) | (59.849) | (59.543) |
| 132 | Homoskedastic | 49.371 | 48.955 | 50.024 | 62.079 | 61.987 | 61.814 |
| | | (8.567) | (8.627) | (8.680) | (13.959) | (13.609) | (13.578) |
| 132 | Heteroskedastic | 44.948 | 44.305 | 45.265 | 57.560 | 57.444 | 57.298 |
| | | (7.369) | (7.501) | (7.834) | (12.337) | (12.245) | (12.384) |
| 132 | MA(1) | 39.227 | 39.234 | 39.054 | 69.809 | 69.586 | 69.236 |
| | | (6.225) | (6.266) | (6.254) | (15.916) | (16.042) | (16.556) |
| 132 | Proxy (Homoskd.) | 28.249 | 27.579 | 27.422 | 60.651 | 60.446 | 60.065 |
| | | (31.545) | (31.141) | (30.988) | (71.739) | (72.218) | (71.744) |

Notes: See notes to Table 3.2 above. The numerical entries not in parentheses under "Factor", "A(j)" or "M(j)" are the means of the various MSFEs calculated under the respective models, across 250 Monte Carlo iterations. The corresponding entries in parentheses are MSFE standard deviations, again calculated across all Monte Carlo iterations.

Table 3.4: **Frequency of Selected Factor Proxies**

| Selected Factor Proxy | Trans | A(j) | M(j) |
|---|---|---|---|
| fspin: S&P's Common Stock Price Index, Industrials | $\Delta \log$ | 1.000 | 1.000 |
| fspcom: S&P's Common Stock Price Index, Composite | $\Delta \log$ | 1.000 | 1.000 |
| fsdxp: S&P's Composite Common Stock: Dividend Yield | $\Delta$lv | 1.000 | |
| fygt1: Interest Rate: U.S. Treasury Const Maturities, 1-Yr | $\Delta$lv | 1.000 | |
| hsfr: Housing Starts, Nonfarm | log | 1.000 | 0.949 |
| hsbr: Housing Authorized, Total New Private Housing Units | log | 0.989 | 0.455 |
| ips10: Industrial Production Index, Total Index | $\Delta \log$ | 0.909 | |
| exrus: United States, Effective Exchange Rate | $\Delta \log$ | 0.835 | 0.370 |
| sfygm6: 6 month Treasury Bills - Federal Funds, spread | lv | 0.813 | |
| sfygt5: 5 yr Treasury Bond Const. Maturities - Federal Funds, spread | lv | 0.750 | |
| sfygt10: 10 yr Treasury Bond Const. Maturities - Federal Funds, spread | lv | 0.659 | 0.420 |
| fygm6: Interest Rate, U.S. Treasury Bills, Sec Mkt, 6-Mo. | $\Delta$lv | 0.460 | |
| a0m077: Ratio, Mfg. and Trade Inventories to Sales | $\Delta$lv | 0.341 | 0.261 |

Notes: In this table we report proxies that were frequently selected using the $A(j)$ and $M(j)$ tests, and the frequencies with which they were selected, in our recursive forecasting experiments. The second column under "Trans" indicates the data transformation that was performed to induce stationarity, lv means no transformation; the series was left at level. $\Delta$lv means first difference of the level. log means the natural log function was applied to the data. $\Delta \log$ means the series was first differenced after the natural log function was applied. Empty entries in the fourth column under $M(j)$ indicate that the respective variables were not selected at all by the $M(j)$ test.

Table 3.5: **Predictive Performance of Various Models for Price Variables**

| Forecast Horizon (h) | 1 | 3 | 12 | 24 |
|---|---|---|---|---|
| Panel A: CPI | | | | |
| Model 1 | 3.496 | 3.464 | 4.299 | 4.089 |
| Model 2 | **3.457** | **3.330** | 4.357 | 5.069 |
| | **(0.136)** | **(0.375)** | (-0.155) | (-2.270)† |
| Model 3 | 4.785 | 5.270 | 6.347 | 6.129 |
| | (-3.788)† | (-3.795)† | (-3.768)† | (-3.087)† |
| Model 4 | 3.809 | 4.075 | 4.792 | 5.305 |
| | (-1.164) | (-1.873)† | (-1.336) | (-2.737)† |
| Model 5 | 4.079 | 4.592 | 5.255 | 5.337 |
| | (-1.125) | (-1.775)† | (-1.650)† | (-1.878)† |
| Model 6 | 3.802 | 4.107 | 4.757 | 4.891 |
| | (-1.139) | (-2.011)† | (-1.347) | (-1.770)† |
| Model 7 | 4.516 | 4.747 | 5.095 | 5.103 |
| | (-1.479) | (-2.223)† | (-1.480) | (-1.600) |
| Model 8 | 3.810 | 4.111 | 4.759 | 4.960 |
| | (-1.169) | (-2.048)† | (-1.382) | (-2.014)† |
| Model 9 | 3.677 | 3.921 | 4.472 | 4.665 |
| | (-0.775) | (-1.798)† | (-0.618) | (-1.645)† |
| Model 10 | 3.819 | 4.101 | 4.769 | 5.208 |
| | (-1.212) | (-2.040)† | (-1.304) | (-2.576)† |
| Model 11 | 3.720 | 4.050 | 4.563 | 4.740 |
| | (-0.935) | (-2.022)† | (-0.881) | (-1.659)† |
| Model 12 | **3.340** | **3.158** | **4.020** | 4.448 |
| | **(0.549)** | **(0.995)** | **(0.921)** | (-0.981) |
| Model 13 | 3.519 | **3.296** | **4.097** | 4.259 |
| | (-0.086) | **(0.539)** | **(0.606)** | (-0.537) |
| Model 14 | **3.486** | **3.381** | 4.351 | 5.124 |
| | **(0.035)** | **(0.232)** | (-0.145) | (-2.379)† |
| Model 15 | **3.351** | **3.331** | **3.999** | 4.297 |
| | **(0.527)** | **(0.411)** | **(0.938)** | (-0.634) |
| Panel B: Consumption Deflator (PCE) | | | | |
| Model 1 | 2.689 | 2.882 | 3.162 | 2.902 |
| Model 2 | **2.613** | **2.540** | **3.097** | 3.918 |
| | **(0.245)** | **(1.598)** | **(0.275)** | (-2.985)† |
| Model 3 | 4.318 | 3.956 | 4.521 | 4.823 |
| | (-2.312)† | (-3.275)† | (-3.082)† | (-3.373)† |
| Model 4 | 3.561 | 3.214 | 3.608 | 4.114 |
| | (-1.911)† | (-1.525) | (-1.983)† | (-3.754)† |
| Model 5 | 2.900 | 3.488 | 3.557 | 3.663 |
| | (-1.106) | (-2.348)† | (-1.990)† | (-2.308)† |
| Model 6 | 3.542 | 3.220 | 3.587 | 3.835 |
| | (-1.871)† | (-1.593) | (-2.118)† | (-2.933)† |
| Model 7 | 3.123 | 3.386 | 3.501 | 3.648 |
| | (-1.865)† | (-2.486)† | (-1.834)† | (-2.349)† |
| Model 8 | 3.562 | 3.283 | 3.921 | 4.412 |
| | (-1.910)† | (-1.847)† | (-3.021)† | (-4.066)† |
| Model 9 | 3.375 | 3.233 | 3.491 | 3.826 |
| | (-1.687)† | (-1.948)† | (-1.729)† | (-2.957)† |
| Model 10 | 3.593 | 3.227 | 3.673 | 4.207 |
| | (-1.887)† | (-1.614) | (-1.969)† | (-3.925)† |
| Model 11 | 3.548 | 3.196 | 3.496 | 3.781 |
| | (-1.717)† | (-1.504) | (-1.769)† | (-2.905)† |
| Model 12 | **2.619** | **2.485** | **3.118** | 3.846 |
| | **(0.237)** | **(2.005)*** | **(0.191)** | (-2.904)† |
| Model 13 | **2.669** | **2.554** | **2.874** | 3.294 |
| | **(0.066)** | **(1.669)*** | **(1.360)** | (-1.562) |
| Model 14 | **2.637** | **2.558** | **3.123** | 3.978 |
| | **(0.163)** | **(1.544)** | **(0.160)** | (-3.229)† |
| Model 15 | **2.633** | **2.525** | **2.817** | 3.271 |
| | **(0.175)** | **(1.870)*** | **(1.617)** | (-1.542) |

Table 3.5 (continued)

| Forecast Horizon (h) | 1 | 3 | 12 | 24 |
|---|---|---|---|---|
| Panel C: Producer Price Index (PPI) | | | | |
| Model 1 | 2.142 | 2.152 | 2.351 | 2.198 |
| Model 2 | 2.445 | 2.360 | 2.433 | 2.385 |
| | (-1.813)† | (-1.349) | (-0.660) | (-1.232) |
| Model 3 | 3.140 | 4.070 | 3.625 | 3.737 |
| | (-3.026)† | (-3.407)† | (-3.214)† | (-3.404)† |
| Model 4 | 2.201 | 2.413 | **2.300** | 2.421 |
| | (-0.387) | (-1.424) | **(0.370)** | (-1.599) |
| Model 5 | 2.282 | 2.391 | 2.370 | 2.536 |
| | (-1.143) | (-1.339) | (-0.152) | (-1.576) |
| Model 6 | 2.203 | 2.392 | **2.256** | 2.303 |
| | (-0.402) | (-1.320) | **(0.729)** | (-0.743) |
| Model 7 | 2.332 | 2.480 | **2.273** | 2.420 |
| | (-1.205) | (-1.828)† | **(0.632)** | (-1.110) |
| Model 8 | 2.206 | 2.397 | **2.257** | 2.332 |
| | (-0.420) | (-1.351) | **(0.730)** | (-1.021) |
| Model 9 | **2.115** | 2.192 | **2.245** | 2.238 |
| | **(0.394)** | (-0.769) | **(1.369)** | (-0.352) |
| Model 10 | 2.217 | 2.474 | **2.345** | 2.407 |
| | (-0.465) | (-1.806)† | **(0.043)** | (-1.350) |
| Model 11 | 2.199 | 2.409 | **2.200** | 2.313 |
| | (-0.385) | (-1.569) | **(1.449)** | (-0.938) |
| Model 12 | 2.396 | 2.299 | 2.356 | 2.332 |
| | (-1.654)† | (-0.888) | (-0.054) | (-1.021) |
| Model 13 | **2.115** | 2.344 | **2.245** | 2.238 |
| | **(0.394)** | (-1.512) | **(1.369)** | (-0.352) |
| Model 14 | 2.447 | 2.401 | 2.465 | 2.407 |
| | (-1.784)† | (-1.558) | (-0.912) | (-1.350) |
| Model 15 | 2.406 | 2.387 | 2.383 | 2.313 |
| | (-1.650)† | (-1.337) | (-0.327) | (-0.938) |

Notes: Primary entries in this table are mean square forecast errors (MSFEs) based upon recursively constructed ex ante predictions for the period 1960:01-2003:12, using Models 1-15 (see Table 3.1 for an explanation of the different models). Bracketed entries are MSFE type Diebold and Mariano (DM: 1995) predictive accuracy test statistics, where Model 1 is compared with each of the other models). Entries in bold indicate instances where the alternative model (i.e. each of Models 2-15) outperforms the factor model (i.e. Model 1), as indicated by both a lower MSFE and a positive DM test stistic. Boxed MSFE entries represent the lowest MSFE value amongst all models, for a particular forecast horizon, $h$. DM statistic entries with a $*$ sign indicate instances where the respective alternative model significantly outperforms the factor model at a 10% significance level, whereas for entries with a † sign, the factor model significantly outperforms the alternative model at a 10% significance level, under the assumption that the DM test statistic has a standard normal limiting distribution (see above for further discussion).

Table 3.6: **Predictive Performance of Various Models for Output, Employment and Sales Variables**

| Forecast Horizon (h) | 1 | 3 | 12 | 24 |
|---|---|---|---|---|
| Panel A: Industrial Production | | | | |
| Model 1 | 2.226 | 2.459 | 3.114 | 2.871 |
| Model 2 | 2.471 | 2.490 | **2.811** | **2.797** |
| | (-1.529) | (-0.192) | **(1.343)** | **(0.673)** |
| Model 3 | 4.267 | 3.931 | 4.541 | 5.528 |
| | (-4.910)† | (-3.142)† | (-3.165)† | (-4.884)† |
| Model 4 | 2.804 | 2.655 | **2.785** | **2.708** |
| | (-3.270)† | (-1.093) | **(1.436)** | **(1.417)** |
| Model 5 | 2.284 | 2.478 | **3.100** | **2.747** |
| | (-0.419) | (-0.147) | **(0.081)** | **(0.560)** |
| Model 6 | 2.682 | 2.623 | **2.795** | **2.688** |
| | (-2.613)† | (-1.039) | **(1.383)** | **(1.584)** |
| Model 7 | 2.678 | **2.352** | **2.708** | **2.620** |
| | (-2.563)† | **(0.948)** | **(1.752)*** | **(1.853)*** |
| Model 8 | 2.719 | 2.652 | **2.737** | **2.584** |
| | (-2.542)† | (-1.210) | **(1.598)** | **(2.195)*** |
| Model 9 | 2.445 | **2.406** | **2.912** | **2.681** |
| | (-1.542) | **(0.447)** | **(0.803)** | **(1.504)** |
| Model 10 | 2.666 | **2.164** | **2.758** | **2.846** |
| | (-2.474)† | **(2.155)*** | **(1.565)** | **(0.232)** |
| Model 11 | 2.512 | **2.291** | **2.654** | **2.609** |
| | (-1.911)† | **(1.268)** | **(1.784)*** | **(1.852)*** |
| Model 12 | 2.594 | 2.615 | **2.737** | **2.584** |
| | (-2.009)† | (-0.976) | **(1.598)** | **(2.195)*** |
| Model 13 | 2.445 | **2.402** | **2.912** | **2.681** |
| | (-1.542) | **(0.490)** | **(0.803)** | **(1.504)** |
| Model 14 | 2.453 | **2.123** | **2.758** | **2.846** |
| | (-1.445) | **(2.445)*** | **(1.565)** | **(0.232)** |
| Model 15 | 2.502 | **2.240** | **2.654** | **2.609** |
| | (-1.840)† | **(1.608)** | **(1.784)*** | **(1.852)*** |
| Panel B: Personal Income Less Transfers | | | | |
| Model 1 | 5.919 | 5.841 | 5.660 | 6.235 |
| Model 2 | 7.167 | 6.811 | **5.576** | **5.994** |
| | (-1.444) | (-1.522) | **(0.293)** | **(1.841)*** |
| Model 3 | 15.316 | 12.858 | 6.533 | 10.327 |
| | (-2.046)† | (-1.697)† | (-0.534) | (-1.459) |
| Model 4 | 6.408 | 6.028 | **5.225** | **6.083** |
| | (-0.725) | (-0.927) | **(1.627)** | **(1.587)** |
| Model 5 | 6.030 | 6.028 | **5.642** | **6.148** |
| | (-0.292) | (-1.125) | **(0.118)** | **(1.117)** |
| Model 6 | 6.373 | 5.996 | **5.298** | **6.071** |
| | (-0.674) | (-0.790) | **(1.513)** | **(1.889)*** |
| Model 7 | 6.570 | 6.249 | **5.518** | **6.027** |
| | (-0.941) | (-1.328) | **(0.418)** | **(2.272)*** |
| Model 8 | 6.368 | 5.991 | **5.300** | **6.075** |
| | (-0.666) | (-0.764) | **(1.505)** | **(1.840)*** |
| Model 9 | 6.334 | 6.147 | 5.690 | **6.132** |
| | (-0.741) | (-2.102)† | (-0.074) | **(0.969)** |
| Model 10 | 6.569 | 6.077 | **5.363** | **6.026** |
| | (-0.734) | (-0.834) | **(0.940)** | **(1.581)** |
| Model 11 | 6.336 | 6.057 | **5.358** | **6.042** |
| | (-0.610) | (-0.782) | **(1.347)** | **(1.887)*** |
| Model 12 | 6.766 | 6.674 | **5.490** | **6.075** |
| | (-1.268) | (-1.327) | **(0.767)** | **(1.840)*** |
| Model 13 | 6.659 | 6.791 | 5.920 | **6.150** |
| | (-1.220) | (-1.589) | (-0.676) | **(1.004)** |
| Model 14 | 7.164 | 6.809 | **5.587** | **6.007** |
| | (-1.440) | (-1.491) | **(0.269)** | **(1.548)** |
| Model 15 | 6.649 | 6.796 | **5.482** | **6.042** |
| | (-1.022) | (-1.417) | **(0.936)** | **(1.887)*** |

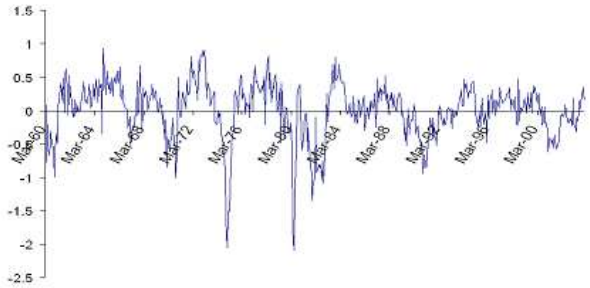
Table 3.6 (continued)

| Forecast Horizon (h) | 1 | 3 | 12 | 24 |
|---|---|---|---|---|
| Panel C: Nonagricultural Employment | | | | |
| Model 1 | 1.893 | 1.693 | 3.587 | 3.279 |
| Model 2 | **1.135** | **1.471** | **3.446** | 3.626 |
| | **(4.013)*** | **(1.323)** | **(0.561)** | (-1.836)† |
| Model 3 | **1.655** | **1.571** | 3.685 | 6.021 |
| | **(0.991)** | **(0.542)** | (-0.239) | (-5.224)† |
| Model 4 | 2.203 | 2.134 | 3.607 | 3.424 |
| | (-1.460) | (-2.614)† | (-0.079) | (-0.970) |
| Model 5 | 2.360 | 2.441 | **3.345** | **2.726** |
| | (-2.191)† | (-3.580)† | **(0.977)** | **(3.068)*** |
| Model 6 | 2.102 | 2.032 | **3.566** | 3.408 |
| | (-0.982) | (-2.115)† | **(0.090)** | (-0.866) |
| Model 7 | 2.235 | 2.102 | **3.177** | **2.992** |
| | (-1.323) | (-2.570)† | **(1.569)** | **(2.170)*** |
| Model 8 | 2.090 | 2.024 | **3.547** | 3.426 |
| | (-0.929) | (-2.073)† | **(0.170)** | (-0.986) |
| Model 9 | 2.223 | 2.219 | **3.385** | **2.772** |
| | (-1.635) | (-3.206)† | **(0.786)** | **(2.767)*** |
| Model 10 | **1.772** | **1.632** | **3.311** | 3.657 |
| | **(0.574)** | **(0.333)** | **(1.066)** | (-2.064)† |
| Model 11 | 2.084 | 2.009 | **3.029** | **2.784** |
| | (-0.935) | (-2.081)† | **(2.256)*** | **(3.210)*** |
| Model 12 | **1.275** | 1.719 | **3.547** | 3.426 |
| | **(3.526)*** | (-0.187) | **(0.170)** | (-0.986) |
| Model 13 | **1.327** | 1.744 | **3.385** | **2.772** |
| | **(3.691)*** | (-0.428) | **(0.786)** | **(2.767)*** |
| Model 14 | **1.128** | **1.406** | **3.311** | 3.657 |
| | **(4.087)*** | **(1.546)** | **(1.066)** | (-2.064)† |
| Model 15 | **1.257** | 1.695 | **3.029** | **2.784** |
| | **(3.825)*** | (-0.015) | **(2.256)*** | **(3.210)*** |
| Panel D: Manufacturing and Trade Sales | | | | |
| Model 1 | 7.001 | 8.243 | 8.603 | 8.187 |
| Model 2 | 7.294 | **7.729** | **8.075** | **7.920** |
| | (-0.639) | **(1.802)*** | **(1.494)** | **(0.912)** |
| Model 3 | 21.172 | 12.915 | 15.844 | 18.207 |
| | (-5.572)† | (-3.449)† | (-4.636)† | (-5.484)† |
| Model 4 | 7.811 | **8.132** | **8.076** | **7.881** |
| | (-1.696)† | **(0.447)** | **(1.461)** | **(1.073)** |
| Model 5 | 7.885 | **7.787** | **8.292** | 8.425 |
| | (-1.239) | **(2.022)*** | **(0.734)** | (-0.914) |
| Model 6 | 7.541 | **7.808** | **8.074** | **7.925** |
| | (-1.197) | **(1.895)*** | **(1.451)** | **(0.915)** |
| Model 7 | 7.706 | **7.890** | **8.183** | 8.420 |
| | (-1.359) | **(1.643)** | **(1.083)** | (-0.907) |
| Model 8 | 7.429 | **7.795** | **8.079** | **7.926** |
| | (-0.959) | **(1.955)*** | **(1.447)** | **(0.910)** |
| Model 9 | 7.199 | **7.836** | **8.148** | **8.033** |
| | (-0.458) | **(1.589)** | **(1.128)** | **(0.602)** |
| Model 10 | 7.571 | **7.895** | **8.091** | **7.964** |
| | (-1.109) | **(1.546)** | **(1.424)** | **(0.763)** |
| Model 11 | 7.465 | **7.917** | **8.092** | **7.984** |
| | (-1.019) | **(1.585)** | **(1.237)** | **(0.687)** |
| Model 12 | 7.429 | **7.795** | **8.079** | **7.926** |
| | (-0.959) | **(1.955)*** | **(1.447)** | **(0.910)** |
| Model 13 | 7.199 | **7.836** | **8.013** | **8.033** |
| | (-0.458) | **(1.589)** | **(1.422)** | **(0.602)** |
| Model 14 | 7.195 | **7.895** | **8.091** | **7.964** |
| | (-0.398) | **(1.546)** | **(1.424)** | **(0.763)** |
| Model 15 | 7.465 | **7.917** | **8.092** | **7.984** |
| | (-1.019) | **(1.585)** | **(1.237)** | **(0.687)** |

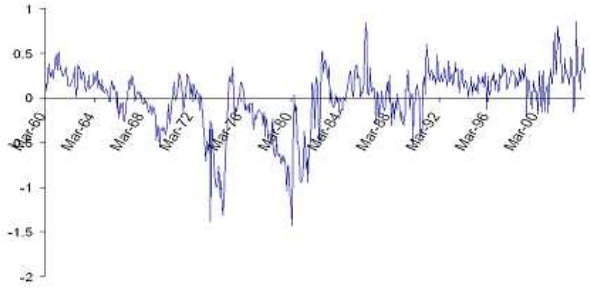
Notes: See notes to Table 3.4.

Figure 3.1: **Estimated Factors and Most Frequently Selected Factor Proxies**

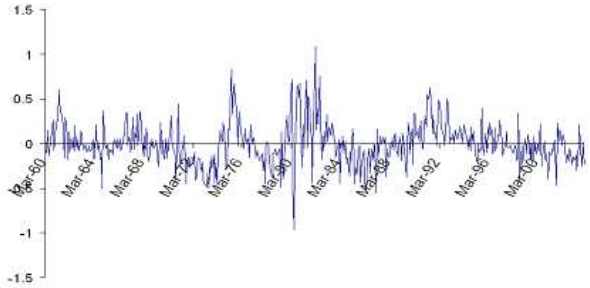
Panel 1: Estimated Factor 1



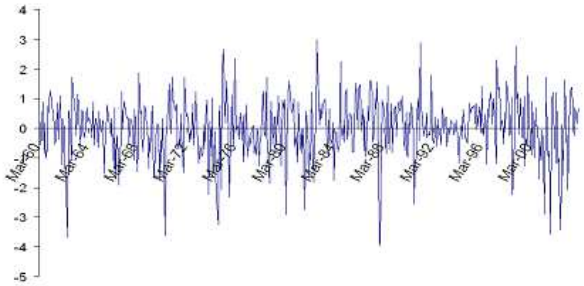

Panel 2: Estimated Factor 2



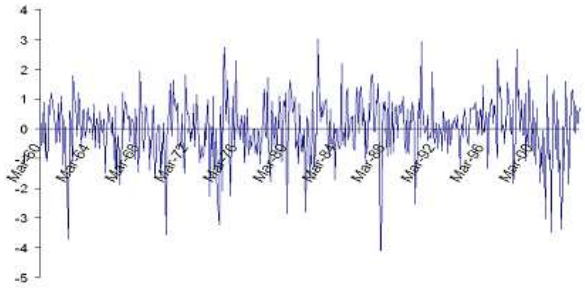
Panel 3: Estimated Factor 3



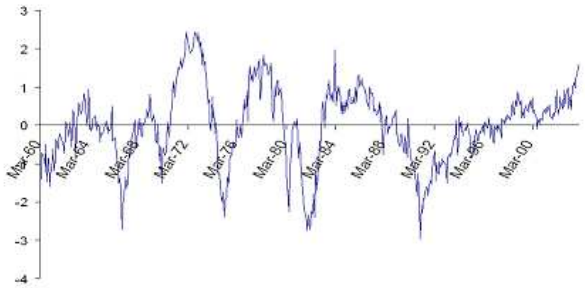

Panel 4: SP's Common Stock Price Index, Composite



Panel 5: SP's CommonStock Price Index, Industrials




Panel 6: Housing Starts, Nonfarm



# Chapter 4

# Which Variables Should the Federal Reserve Monitor? New Diffusion Index Evidence

## 4.1 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 (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) enabling one to "control" the target variable, but might be characterized by poor predictive accuracy. In this sense, an ideal model might be one that has satisfactory out-of-sample forecast performance and includes relevant control variables. In light of this, we assume in this paper that both policy analysis and forecast performance are relevant, and we assess, via the use of diffusion index methodology, the importance of a particular set of variables that are monitored by the Federal Reserve.

In order to construct a "reasonable approximation" to the set of variables that the Federal Reserve (Fed) Banks might typically be expected to monitor, note that the different sectors of the economy can be expected to respond to changes in interest rates and other monetary policy instruments with varying reaction times. For this reason, in formulating the nation's monetary policy, monetary regulators like the Fed monitor the behavior of a diverse group of macroeconomic and financial indicators. In particular, it is stated on

the Federal Reserve Bank of New York's website that: *"In formulating the nation's monetary policy, the Federal Reserve considers a number of factors, including the economic and financial indicators which follow, as well as the anecdotal reports compiled in the Beige Book. Real Gross Domestic Product (GDP); Consumer Price Index (CPI); Nonfarm Payroll Employment Housing Starts; Industrial Production/Capacity Utilization; Retail Sales; Business Sales and Inventories; Advance Durable Goods Shipments, New Orders and Unfilled Orders; Lightweight Vehicle Sales; Yield on 10-year Treasury Bond; S&P 500 Stock Index; M2" (see http://www.newyorkfed.org/education/bythe.html).* Thus, some of the financial and macroeconomic indicators the Federal Reserve relies on to create the country's monetary policy 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 economic and financial indicators will be referred to as the "macroeconomic indicators". The aim of this paper is to determine the "adequacy" of these particular macroeconomic indicators, in a controlled experimental environment. Moreover, we argue that the methodology used herein will be useful also for examining other macroeconomic indicators of interest to policy makers and forecasters. The main tool that we use in our assessment is ex ante prediction. In order to retain "policy relevance", all of our experiments involve models that include various "control" variables. Our methodology centers on the construction of out-of-sample forecasts of CPI inflation and output growth using a variety of models including, among others: ones that take the macroeconomic indicators as the relevant set of explanatory variables; ones that include factors constructed from a large scale macroeconomic dataset (that includes the macroeconomic indicators and many other variables) using the diffusion index methodology of Stock and Watson (1998,1999,2002a,b); ones that are based on factor proxies constructed using the methodology discussed in Bai and Ng (2006a,b) and further developed in Armah and Swanson (2008).

The reason why we use the diffusion index methodology of Stock and Watson as a starting point in our analysis is that their methodology has been shown to consistently

estimate the relevant common factors that underlie the co-movements of a given set of macroeconomic variables. We take it as given that two variables that the Fed wishes to track are output growth and inflation, in which case our comparison of their macroeconomic indicators with our proxies is an appropriate assessment device. Put differently, if we discover that the factor proxies most useful for inclusion in prediction models of output and inflation in fact correspond to the macroeconomic indicators, then we have direct evidence of the appropriateness of the approach used by the Fed in tracking monetary policy induced economic behavior. This is a main goal of our paper. Another goal is to uncover any additional variables that might be of use in monitoring economic activity. Finally, our third goal is to add to the methodological literature on the use of diffusion indices in macroeconometric applications. This is done by positing a framework for implementation of the tools for factor proxy selection developed in Bai and Ng (2006a,b) and Armah and Swanson (2008).

Interestingly, examination of the macroeconomic indicators and the observable variables selected via our methodology indicates that the two groups are largely the same; and at the very least, individual members of the two groups belong to the same "classes" of variables. Examples of "classes" of variables include *CPI variables* (i.e. CPI all items, CPI transportation, CPI apparel and upkeep, etc.) and *IP variables*. There is one important exception, however. Our factor proxy analysis additionally finds evidence in favor of the use of spreads for monitoring the effects of monetary policy. Moreover, and perhaps more importantly, the particular spreads found to be useful are constructed as the difference between short and long term debt instruments and the federal funds rate. Surprisingly, traditional spreads, such as the yield curve slope and the reverse yield gap are not found to be useful. Moreover, and with regard to our findings of the usefulness of spreads, we find that the Fed's macroeconomic indicators perform best when forecasting inflation in non-volatile time periods. On the contrary, the forecast performance of the indicators can be improved by including spreads when forecasting inflation in times of high volatility.

In a series of ex ante forecasting experiments carried out to assess the adequacy of the variables monitored by the Federal Reserve, we consider models that do and do not include the three spread variables. Interestingly, "spread augmented" models, which are by

construction less parsimonious than those 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 one half the time, and at 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 spreads are also significantly better, based on application of Diebold and Mariano (1995) tests (see Clark and McCracken (2005) and McCracken (2007) for a discussion of Diebold and Mariano tests in various time series contexts). 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 suggest that certain "non-traditional" spreads (i.e. none of the traditional spreads proposed by economic theory in the existing literature) may be useful for monitoring economic activity. Of final note is these spreads are picked from our huge set of observable macroeconomic and financial variables in a completely agnostic manner, in the sense that we let the data "do the talking". It remains to examine the implications, for macroeconomic theory, of the relevance of these variables.

The rest of the paper is organized as follows. In Section 4.2, we summarize some useful features of the diffusion index literature and briefly discuss the methodology associated with forming predictions using factor proxies and using the Federal Reserve's macroeconomic indicators. In Section 4.3, we discuss the predictive content of spreads and review some of the existing literature on the subject. Section 4.4 contains a brief overview of the data used in our prediction experiments, and empirical methodology is further outlined in Section 4.5. Section 4.6 summarizes our empirical findings, and Section 4.7 offers concluding remarks.

## 4.2 Using Macroeconomic Indicators, Factors, and Factor Proxies for Prediction

### 4.2.1 Prediction with Macroeconomic Indicators

To craft the nation's monetary policy, the Federal Reserve (Fed) monitors a number of macroeconomic and financial variables. Monetary regulators and researchers have traditionally relied on indicators such as monetary aggregates (particularly for large countries) and the exchange rate (usually for smaller countries) for the stance of monetary policy (see Davis and Fagan (1997)). Davis and Fagan (1997), however, point out that the distortion of monetary aggregates in a number of countries following financial innovation, and the increasing volatility of exchange rates have caused researchers and central banks to consider alternative indicators like the ones used by the Fed. Although the macroeconomic indicators as listed in this paper are known, functional forms or transformations of these variables used in the specification of Fed's models remain unknown. In the absence of this knowledge, we begin by specifying simple prediction models for inflation and output growth. Even though simple, the sorts of parsimonious models used in our analysis are prototypical "strawman" models that have been found to perform well for many variables (see e.g. Swanson and White (1995,1997)). Moreover, predictions from these models can be easily analyzed using the forecast evaluation methodology of Clark and McCracken (2005). Our simple model is called Model A, and is given by:

$$y_{t+h} = \alpha_a' W_{at} + \sum_{j=1}^{p} \beta_{aj} y_{t-j+1} + \varepsilon_{at+h}, \tag{4.1}$$

where $W_{at}$ is an $l_a \times 1$ vector of observable variables (e.g. the macroeconomic indicators). 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 also evaluated, 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 $W_{at}$, is a benchmark against we compare other models. Another benchmark that we consider is the set of "Green Book" forecasts generated by the Federal Reserve Board of Governors' Research and Statistics Division. (Results based upon the latter benchmark

are not yet available.)

### 4.2.2 Prediction with Factors

The factor model of Stock and Watson (2002a,b) is included as one of our benchmark prediction models 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. We view this as a sensible approach, given that by design, the principal component factors of Stock and Watson would be very good candidate variables for assessing how the economy responds to a particular policy stance.

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 $t = 1, \ldots, T$. Assume that $(y_{t+1}, X_t)$ has a dynamic factor model representation with $\overline{r}$ common dynamic factors, $f_t$. Hence, $f_t$ is an $\overline{r} \times 1$ vector. The dynamic factor model is written as:

$$y_{t+h} = \alpha(L) f_t + \sum_{j=1}^{p} \beta_j y_{T-j+1} + \varepsilon_{t+h} \tag{4.2}$$

and

$$x_{it} = \lambda_i(L) f_t + e_{it}, \tag{4.3}$$

for $i = 1, 2, \ldots, N$, 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 $\overline{r}$. This implies that data generated under (4.3) would have $\overline{r}$ 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 (4.2) and (4.3) as:

$$y_{t+h} = \alpha' F_t + \sum_{j=1}^{p} \beta_j y_{T-j+1} + \varepsilon_{t+h} \tag{4.4}$$

and

$$x_{it} = \Lambda_i' F_t + e_{it}, \tag{4.5}$$

where $F_t = (f_t', \ldots, f_{t-q}')'$ is an $r \times 1$ vector, with $r = (q+1)\overline{r}$ and $\alpha$ is an $r \times 1$ vector. Here, $r$ is the number of static factors (i.e. the number of elements in $F_t$). Additionally, $\Lambda_i = (\lambda_{i0}', \ldots, \lambda_{iq}')'$ is a vector of factor loadings on the $r$ static factors, where $\lambda_{ij}$ is an $\overline{r} \times 1$ vector for $j = 0, \ldots, q$. The $N$ dimensional population covariance matrix of $x_{it}$ generated under (4.5) 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 $\frac{1}{N}\sum_{i=1}^{N} \lambda_i \lambda_i' p \longrightarrow \Lambda$ as $N \to \infty$, and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' p \longrightarrow {}_F$ as $T \to \infty$, where ${}_\Lambda$ and ${}_F$ 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) 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. For this reason, the diffusion index model proposed by Stock and Watson is made one of the benchmark models.

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:

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

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

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

where $\Lambda = (\Lambda_1, \ldots, \Lambda_N)'$ and $e = (\underline{e}_1, \ldots, \underline{e}_N)$. $X$ 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 $e_t = (e_{1t}, \ldots, e_{Nt})$ 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 $(\Lambda'\Lambda/N) \to I_r$, 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 (4.4) and (4.5) 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, $\{\widetilde{F}_t\}_{t=1}^T$ by means of principal components. With the estimated factors in hand, we obtain the estimators $\widehat{\alpha}$ and $\widehat{\beta}$ by regressing $y_{t+1}$ onto $\widetilde{F}_t$ and the observable variables in $W_t$. Of note is that if $\sqrt{T}/N \to 0$, then the generated regressor problem does not arise, in the sense that least squares estimates of $\widehat{\alpha}$ and $\widehat{\beta}$ are $\sqrt{T}$ consistent and asymptotically normal (see Bai and Ng (2006a)).

Since the common factors are not observed, in the regression analysis of (4.5), we replace $F_t$ by $\widetilde{F}_t$, estimates that span the same space as $F_t$ when $N, T \to \infty$. 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, 2004a, 2004b) and Bai and Ng (2002, 2007) for a detailed explanation of this procedure.

From (4.5), estimates of $\Lambda_i^k$ and $F_t^k$ are obtained by solving the optimization problem:

$$V(k) = \Lambda^k, F^k \min (NT)^{-1} i = 1N \sum Tt = 1 \sum (x_{it} - \Lambda_i^{k\prime} F_t^k)^2 \tag{4.6}$$

Let $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$ be the minimizers of equation (4.6). Since $\Lambda^k$ and $F^k$ are not separately identifiable, if $N > T$, a computationally expedient approach would be to concentrate out $\widetilde{\Lambda}^k$ and minimize (4.6) subject to the normalization $F^{k\prime} F^k / T = I_k$. Minimizing (4.6) is equivalent to maximizing $tr[F^{k\prime}(XX')F^k]$. This optimization is solved by setting $\widetilde{F}^k$ to be the matrix of the $k$ eigenvectors of $XX'$ that correspond to the $k$ largest eigenvalues of $XX'$. Note that $tr[\cdot]$ represents the matrix trace. The superscript in $\Lambda^k$ and $F^k$ signifies the use of $k$ factors in the estimation and the fact that the estimates will depend on $k$. Let $\widetilde{D}$ be a $k \times k$ diagonal matrix consisting of the $k$ largest eigenvalues of $XX'$. The estimated

factor matrix, denoted by $\widetilde{F}^k$, is $\sqrt{T}$ times the eigenvectors corresponding to the $k$ largest eigenvalues of the $T \times T$ matrix $XX'$. Given $\widetilde{F}^k$ and the normalization $F^{k\prime}F^k/T = I_k$, $\widetilde{\Lambda}^{k\prime} = (\widetilde{F}^{k\prime}\widetilde{F}^k)^{-1}\widetilde{F}^{k\prime}X = \widetilde{F}^{k\prime}X/T$ is the corresponding factor loadings matrix.

The solution to the optimization problem in (4.6) is not unique. If $N < T$, it becomes computationally advantageous to concentrate out $\overline{F}^k$ and minimize (4.6) subject to $\overline{\Lambda}^{k\prime}\overline{\Lambda}^k/N = I_k$. This minimization is the same as maximizing $tr[\Lambda^{k\prime}X'X\Lambda]$, the solution of which is to set $\overline{\Lambda}^k$ 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 $\overline{F}^k = X'\overline{\Lambda}^k/N$. $\widetilde{F}^k$ and $\overline{F}^k$ 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 $\widetilde{F}^k$ and $\widetilde{\Lambda}^k$, let $\widehat{V}(k) = (NT)^{-1}i = 1N\sum Tt = 1\sum(x_{it} - \widetilde{\Lambda}_i^{k\prime}\widetilde{F}_t^k)^2$ be the sum of squared residuals from regressions of $X_i$ on the $k$ factors, $\forall i$ and $IC(k) = \log(\widehat{V}(k)) + k(\frac{N+T}{NT})\log C_{NT}^2$ 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 $\widehat{k} = \arg\min_{0 \leq k \leq k\max} IC(k)$.

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 $N, T \rightarrow \infty$.

### 4.2.3 Prediction With 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 (2008) 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. The factor proxies represent a set of methodically selected observable variables that can possibly indicate an economy's response to current monetary policy. We compare the macroeconomic indicators with the selected factor proxies in order to determine whether the indicators contain all "crucial information" that is useful for assessing the economy's response to a change in monetary policy.

Recall the general equation (4.4):

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

As mentioned above, and shown in Stock and Watson (2002b) and Bai and Ng (2006b), under a set of moment conditions on $(\varepsilon, e, F^0)$ and an asymptotic rank condition on $\Lambda$, if the space spanned by $F_t$ can be consistently estimated, then $\sqrt{T}$ consistent estimates of $\alpha$ and $\beta$ are obtainable. Under a similar set of conditions, it is also possible to obtain $\min[\sqrt{N}, \sqrt{T}]$ consistent forecasts if $\sqrt{T/N} \to 0$ as $N, T \to \infty$.

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 $G_{jt} = \gamma_j' \widetilde{F}_t + \rho$. Let $\widehat{\gamma}_j$ be the least squares estimate of $\gamma_j$ and let $\widehat{G}_{jt} = \widehat{\gamma}_j' \widetilde{F}_t$. The test is carried out by constructing the following t-statistic:

$$\tau_t(j) = \frac{(\widehat{G}_{jt} - G_{jt})}{(\widehat{var}(\widehat{G}_{jt}))^{1/2}} \tag{4.7}$$

where

$$\begin{aligned} \widehat{var}(\widehat{G}_{jt}) &= \frac{1}{N} \widehat{\gamma}_j' \widetilde{D}^{-1} \left( \frac{\widetilde{F}' \widetilde{F}}{T} \right) \widetilde{\Gamma}_t \left( \frac{\widetilde{F}' \widetilde{F}}{T} \right) \widetilde{D}^{-1} \widehat{\gamma}_j \\ &= \frac{1}{N} \widehat{\gamma}_j' \widetilde{D}^{-1} \widetilde{\Gamma}_t \widetilde{D}^{-1} \widehat{\gamma}_j, \end{aligned} \tag{4.8}$$

and $\widetilde{\Gamma}_t$ is defined below. The last step above is due to the normalization that $\widetilde{F}' \widetilde{F}/T = I_{\widehat{k}}$. Once again, $\widetilde{D}$ 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 $\widehat{G}_{jt}$ converges to $G_{jt}$ at rate $\sqrt{N}$, Bai and Ng (2006b) show that the limiting distribution of $\sqrt{N}(\widehat{G}_{jt} - G_{jt})$ is asymptotically normal and hence $\tau_t(j)$ has a standard normal limiting distribution. The $\widehat{k} \times \widehat{k}$ matrix $\widetilde{\Gamma}_t$ is consistently estimated as

$$\widetilde{\Gamma}_t = \frac{1}{N} Ni = 1 \sum \widetilde{e}_{it}^2 \widetilde{\Lambda}_i \widetilde{\Lambda}_i', \tag{4.9}$$

and where $\widetilde{e}_{it} = x_{it} - \widetilde{\Lambda}_i' \widetilde{F}_t$. Equation (4.9) 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 (4.9) is useful even if residual cross-correlation is genuinely present.

As mentioned earlier, $\tau_t(j)$ in (4.7) has a standard normal limiting distribution. Let $\Phi_\xi^\tau$ be the $\xi$ percentage point of the limiting distribution of $\tau_t(j)$. The hypothesis test based on the t-statistic in (4.7) 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:

$$A(j) = \frac{1}{T}t = 1T\sum 1(|\tau_t(j)| > \Phi_\xi^\tau). \tag{4.10}$$

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 $2\xi$) as $T \to \infty$. 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 $2\xi$ 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 $2\xi$. 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 (2008) 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 potential advantage to this approach is that noise across the proxy selection process is potentially suppressed.

In our empirical implementation, the factor proxies selected by the smoothed $A(j)$ statistic from the panel dataset are listed in Panel C and Panel I of Table 4.1. Our so-called macroeconomic indicators are listed in Panel A of Table 4.1. A close 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 4.1, 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 is a strong testament to the usefulness of the macroeconomic indicators. It remains to see, however, whether the three spread variables actually improve the forecast performance of the macroeconomic indicators in out-of-sample forecasts of CPI inflation and output growth.

## 4.3 Predictive Content of Spreads

The yield on a debt instrument like a government bond is the annual rate of return that would be earned by a lender who holds the bond to maturity. For assuming more risk, lenders (investors) would generally demand a higher annual rate of return (yield) on debt instruments with longer maturities than those with shorter maturities. The yield curve describes the relationship between the yields (interest rates) and maturities of a particular debt instrument. Spreads may be defined as the difference between yields on two financial instruments. The bigger the spread is between a long-term and short-term debt instrument, the steeper the slope of the corresponding yield curve will be. Spreads exist because assets and more specifically, debt instruments are imperfect substitutes of each other. Reasons for this imperfect substitutability include differences in liquidity, currency, maturity, risk and

levels and covariances of yields on alternative assets (see Davis and Fagan (1997)). There is a vast literature on the predictive content of spreads for inflation and output growth and the fundamental idea to this body of literature is that financial market participants are forward-looking. Asset prices consequently embody useful information such as expectations of future economic activity. Another advantageous indicator property of spreads is that they are readily observable usually some periods ahead of currently available macroeconomic data. Spreads are also reliable since they are not subject to revisions. The usefulness of spreads in the forecast of output growth and inflation in the US have been studied in Laurent (1988, 1989); Harvey (1988, 1989); Stock and Watson (1989); Mishkin (1990, 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 UK, 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).

Estrella and Hardouvelis (1991) find that spreads have predictive content for output growth not contained in other variables like the current level of real interest rates and advocate the sole use of spreads in predicting output growth. They argue that the spread between the yield on the ten-year Treasury bond and the three-month Treasury bill is useful for predicting both cumulative and marginal output growth. Estrella and Hardouvelis (1991) further find the spread useful for predicting the likelihood of a recession. Plosser and Rouwenhorst (1994) use discount equivalent yields and match the maturity structure of the interest rate spread with the forecast horizon being studied to examine the predictive content of spreads for different countries between 1973 to 1988. They also consider spreads of long-term bonds and short-term bills and discover that the term spread's ability to predict economic activity at short horizons of up to two years provides some significant in-sample predictive content for cumulative changes in output. Estrella and Mishkin (1998) argue that the spread between ten-year Treasury bond yields and three-month Treasury bill yields is the best out-of-sample predictor of the likelihood of a recession within a four quarter horizon. By adding an equity price indexes, Estrella and Mishkin (1998) improve

forecast accuracy at shorter horizons. Mishkin (1990a) and Mishkin and Jorion (1991) employ simple bivariate analysis to determine whether the correlation between spreads and future inflation or output growth is significant. Bernanke (1990) and Friedman and Kuttner (1991) use bivariate Granger Causality analysis on US data to assess whether spreads contain predictive information content beyond that contained in the lags of the dependent variables. Finally, Garnett, Hall and Henry (1992), Davis and Henry (1992, 1993) and Davis, Henry and Pesaran (1994) among others use vector autoregressive (VAR) analysis to examine the predictive content of spreads for output growth and inflation.

One traditional spread 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)). Violation of any of these strong assumptions would make the slope of the yield curve spread less accurate in forecasting inflation. Empirical evidence from work by Mishkin (1990b); Jorion and Mishkin (1991); Browne and Manasse (1989) does suggest that although the link between the slope of the yield curve and future inflation is not perfect, the spread has significant predictive content for inflation. In addition to its role for forecasting future inflation, the slope of the yield curve is also considered a useful indicator of future cyclical output movements. Laurent (1988) further explains that a positive shift in the yield curve may induce banks to purchase long-term securities and make long-term loans, which could from a monetarist perspective boost money and hence economic activity.

Another variety of spread considered 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 whiles 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)).

Conclusions of empirical studies aiming to assess the predictive content of spreads for inflation and output growth have been mixed. This is because in general, the predictive content of a regressor for a target variable is often difficult to assess definitively, and conclusions drawn are only valid with respect to the model specification and the data sample used in the analysis. Davis and Fagan (1997) point out that in bivariate analysis, the slope of the yield curve may be found to be useful in forecasting inflation. However, if some other variables such as short-term interest rates, money stock, other leading indicators or even past values of inflation are included in the model specification, the 'marginal forecasting power' of the yield curve slope may be crowded out although the overall forecasting performance of the new model may increase (Davis and Fagan (1997)). In consequence, one could argue that while a spread variable may contain some useful information about the future of inflation, it may contain no information beyond that contained in other monetary variables or lags of the target variable. Davis and Fagan (1997) further emphasize that this line of argument leads to the conclusion that in assessing the 'marginal predictive content' of one variable for another, there is no definitive 'correct answer', rather conclusions are heavily subject to the information set considered in the model specification and may differ significantly as the model specification is varied. The above discussion is one reason the lag length $p$ in our models was determined by the SIC (we found values of $p = 3$ for inflation and $p = 1$ for output growth). Moreover, in our empirical forecasting exercises, we examine the marginal predictive content of spreads using various different models. The objective is to assess whether spreads have predictive information content over and above different information sets used in specifying the relevant models. Our model that incorporates spreads as predictors, and which corresponds to Model B is:

$$y_{t+h} = \alpha_b' W_{bt} + \sum_{j=1}^{p} \beta_{bj} y_{t-j+1} + \varepsilon_{bt+h} \tag{4.11}$$

$W_{bt} \equiv (W_{at}', S_t')'$, $W_{at}$ is an $l_a \times 1$ vector described in (4.1), $S_t$ is an $l \times 1$ vector of spreads and hence $W_{bt}$ is an $l_b \times 1$ vector where $l_b = l_a + l$.

In past work, spreads have been included in forecasting models mainly because of economic theory or more specifically, asset pricing theory. In this paper, however, we do not rely on any theory to explicitly include spreads as predictors when specifying models. Instead, spreads are part of our huge set of observable macroeconomic and financial variables that can potentially proxy the underlying factors in the economy. From these candidate variables, the smoothed $A(j)$ statistic is used to select the variables that actually proxy the underlying factors. In this regard and independent of economic theory, we allow the massive dataset that represents the economy to "talk" to us and to determine what possible variables should be included in the forecasting models. Of the many possibilities, some spread variables were indeed selected by the smoothed $A(j)$ statistic as close proxies of the underlying factors. The interesting outcome however was that, none of the traditional spreads proposed by economic theory in the existing literature cited here were selected. Rather, three spreads composed of the difference in yields of long-term and short-term debt instruments and the federal funds rate were selected. These spreads appear to have some predictive content at least for inflation in the long run and output growth in the short run.

## 4.4 Data

The dataset used in this paper to represent the economy and from which the economy's underlying factors are estimated is derived from that used in Stock and Watson (2005). The original Stock and Watson (2005) dataset can be found at *http://www.princeton.edu/˜mwatson* and contains 132 monthly time series for the United States from 1960:1 to 2003:12 to give $N = 132$ and $T = 528$ observations.[1] Time series of the various variables 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 in the original Stock and Watson dataset were selected from the following categories of macroeconomic time series: real output and income; employment, manufacturing and trade sales; consumption; housing starts

[1] An updated version of the Stock and Watson dataset is available, although that dataset contains only quarterly data, whereas our dataset consists of monthly data. Neverthless, in order to check the robustness of our finding, all experiments were also carried out using the updated quarterly data, and results were found to be largely unchanged from those reported here.

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 make up the macroeconomic indicators can be found at the Federal Reserve Bank of New York's website at *http://www.ny.frb.org/education/bythe.html* and are listed as Real Gross Domestic Product (GDP); Consumer Price Index (CPI); Non-farm 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; Yield on 10-year Treasury Bond; S&P 500 Stock Index; M2. For Business Sales and Inventories as well as Advanced Durable Goods Shipments, New Orders and Unfilled Orders, the available series begin in 2002. In order to maintain the integrity of the balanced panel, we use Manufacturing and Trade Sales in addition to Manufacturing and Trade Inventories as close substitutes for Business Sales and Inventories. Industrial Production: Durable Goods Materials is further used as a close substitute for Advanced Durable Goods Shipments, New Orders and Unfilled Orders. The data on Lightweight Vehicle Sales starts from 1976 and since there were not that many good substitutes for that variable, the initial 14 years of data from the original Stock and Watson dataset were redacted to make the actual dataset used in this paper run from 1976 to 2003. Obviously, empirical results are heavily dependent on the panel dataset used to represent the economy. Evidence from prior work seems to suggest that the information content of the estimated factors and selected factor proxies improves with a higher variety of variables as well as a higher number of subaggregates contained in the panel dataset.

The original Stock and Watson dataset includes spreads constructed as the difference between the yields on the following debt instruments and the Federal Funds rate: commercial paper; 3-month, 6-month, 1-year, 5-year, 10-year Treasury bills/bonds; moody's AAA; moody's BAA corporate bonds. For the purposes of this paper, specific yield curve spreads considered and added to the dataset include the difference between the following variables and the 3-month Treasury bill rate: 6-month, 1-year, 5-year, 10-year Treasury bills/bonds; moody's AAA; moody's BAA corporate bonds. Also, the specific reverse yield gap spread considered is the difference between the yield on the 10-year Treasury bond and the S&P 500 common stock dividend yield. With all these modifications, the final dataset used in

this paper runs from 1976:1 to 2003:12 with $N = 139$ and $T = 336$.

The theory underlying the factor forecast and the methodology for selecting the factor proxies assumes that all the variables in the panel dataset are $I(0)$. Some of the variables in the panel dataset are made stationary by applying transformations that include taking logarithms and/or first differencing. At each recursive iteration, all the variables in the panel dataset are further standardized to have mean zero and unit variance before the factors were estimated by principal components and proxies selected by the Bai and Ng (2006b) methodology. The stationary panel dataset with variables standardized to have zero mean and unit variance is represented in this paper by $X$. Real time forecasts are performed of the growth rate in industrial production index: total index and the growth rate in CPI: all items.

## 4.5 Empirical Methodology

Forecasts are generated as $h$-step ahead recursive predictions of $y_t$. That is, we 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. The available data is split into three subsamples such that $T = R_1 + R_2 + E$. For the factor models, 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 actual factors are consistently 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 set to 1 for output growth and 3 for inflation throughout the ex-ante forecast period, based on an initial lag selection procedure carried out using the Schwarz information criteria, using the first subsample of data. Unlike the factors in the factor forecast model, the observable variables in the alternative forecast models (Model 1 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. As a robustness check, permutations of the selected factor proxies and the macroeconomic indicators are used as predictors in the alternative models. All alternative models are also re-estimated by OLS at each recursive iteration before the

$h$-step ahead forecast is constructed.

Various classes of variables in the Stock and Watson dataset have aggregates as well as subaggregates. It is consequently not a trivial task to pick a representative variable from a particular class. For instance, included in the Fed's list are CPI, industrial production and Housing Starts. However, there are at least 9 CPI, 12 industrial production, 9 housing, and 8 interest rate variables in the panel dataset. Picking a representative CPI or industrial production variable can consequently be tricky if not ad hoc. In specifying the benchmark model with the list of variables in $W_{1t}$, we subjectively pick the aggregate variable to be the default representative of the class. So for example, CPI: all items is made the representative CPI variable. The rationale behind this is that by construction, the aggregate variable contains some of the information from the other subaggregates. An alternative approach could be to use the smoothed $A(j)$ statistic on a specific class of variables. In this regard, let $\widehat{X} \subset X$ where $\widehat{X}$ contains all the relevant aggregate and subaggregate variables in a particular class like Housing Starts or CPI. Therefore, for CPI, $\widehat{X}$ will contain CPI: all items; CPI: apparel and upkeep; CPI: transportation; CPI: medical care; etc. The first two principal components are estimated from $\widehat{X}$ and the smoothed $A(j)$ statistic is used to select variables from $X$ that proxy these two principal components. This way, we methodically select a representative variable from a class. The principal components factors estimated from $\widehat{X}$ and the corresponding 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 $\widehat{X}$. Some of the models have focused factor proxies as regressors and we select focused variables only from classes with enough subaggregates. In this regard, the classes considered are CPI, Industrial Production, Housing, Employment and Yields. The Yield class includes the various interest rates and spreads.

Model 1 represents the linear benchmark model, where $W_{1t}$ contains the macroeconomic indicators as listed in panel A of Table 4.1. 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. In all even numbered models, $W_{nt}$ $(n = 2, 4, 6, 8)$ contains the variables in the corresponding previous models with the relevant ones replaced by variables selected under focused principal components. In Model $nS$ $(n = 1, \ldots, 8)$, the corresponding variables in $W_{nt}$ have been augmented with the spreads in Panel I of Table 4.1.

In evaluating forecast performance to determine the predictive content of spreads, a parsimonious benchmark model (without spreads) in (4.1)

$$y_{t+h} = \alpha_a' W_{a,t} + \sum_{j=1}^{p} \beta_{a,j} y_{t-j+1} + \varepsilon_{a,t+h}$$

is effectively compared to a larger alternative model (with spreads) in (4.11)

$$y_{t+h} = \alpha_b' W_{b,t} + \sum_{j=1}^{p} \beta_{b,j} y_{t-j+1} + \varepsilon_{b,t+h}$$

$W_{bt} \equiv (W_{at}', S_t')'$, $W_{at}$ is an $l_a \times 1$ vector described in (4.1), $S_t$ is an $l \times 1$ vector of spreads and hence $W_{bt}$ is an $l_b \times 1$ vector, where $l_b = l_a + l$. The two models being compared are thus nested in the sense that in this setup, the alternative model uses a set of predictors $W_{bt}$ to predict the target variable $y_{t+h}$ whereas the null or parsimonious model uses a set of predictors $W_{at}$ that is a strict subset of $W_{bt}$. The alternative model consequently contains $l$ excess parameters. In applications where the forecast performance of models is compared and the benchmark model is a restricted version of the alternative model, the asymptotic and finite-sample properties of equal forecast accuracy test statistics based on non-nested models as described in West (1996, 2001) may not apply. West (2005) argues that the nesting of models violates a rank condition required in the asymptotic normality results of West (1996). However, the exercise of comparing the forecast performance of all the alternative models to the factor model involves comparing two non-nested models. For the non-nested cases, the asymptotic normality results of West (1996) are indeed applicable.

The test statistic used for all forecast evaluations in this paper is the $t$-statistic for equal forecastibility developed by Diebold and Mariano (1995) and West (1996) under quadratic loss. The total 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 model 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 number of $E-h$ predictions/observations. Forecasts for both null and alternative linear models are made recursively using least squares estimated parameters. In the context of Diebold and Mariano (1995), let $\widehat{\varepsilon}_{a,t+h} = y_{t+h} - \widehat{\alpha}_a' W_{a,t} + \sum_{j=1}^{p} \widehat{\beta}_{a,j} y_{t-j+1}$ and $\widehat{\varepsilon}_{b,t+h} = y_{t+h} - \widehat{\alpha}_b' W_{b,t} + \sum_{j=1}^{p} \widehat{\beta}_{b,j} y_{t-j+1}$, then the sample $MSFE = \frac{1}{E-h} \sum_{t=R+1}^{T-h} \widehat{\varepsilon}_t^2$. The null hypothesis of equal forecast accuracy from the two models is given by $H_0 : E[\widehat{d}_t] = 0$, where $\widehat{d}_t = \widehat{\varepsilon}_{a,t}^2 - \widehat{\varepsilon}_{b,t}^2$ is the loss differential and $\overline{d} = \frac{1}{E-h} \sum_{t=R+1}^{T-h} \widehat{d}_t$. The DM test statistic is then

$$DM = \sqrt{E-h} \frac{\frac{1}{E-h} \sum_{t=R+1}^{T-h} \widehat{d}_t}{\sqrt{\frac{1}{E-h} \sum_{j=-\overline{j}}^{\overline{j}} \sum_{t=R+1+j}^{T-h} K(\frac{j}{M})(\widehat{d}_t - \overline{d})(\widehat{d}_{t-j} - \overline{d})}}$$

where $K(\frac{j}{M})$ is the kernel with bandwidth $M$.

As already stated, for non-nested 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 $d_{t+h} = 0$, $\forall t$, in population. This means that the population variance of $d_{t+h}$ is equal to 0 and hence standard inference a la Diebold and Mariano (1995) or West (1996) does not apply, and one must use the results of Clark and McCracken (2005). Indeed, McCracken (2007) proves that for nested models, the DM $t$-statistic converges in distribution to a functional of stochastic integrals of quadratics of Brownian motion, with a limiting distribution that depends on the sample split, $\pi = \lim_{R,E\to\infty} \frac{E}{R}$, and the number of exclusion restrictions, $l$, but does not depend on any other nuisance parameters. Appropriate limiting distributions together with the critical values can be found in Clark and McCracken (2005) and McCracken (2007). Of final note is that we carry out our forecasting experiments for 1, 3 and 12 step ahead forecast horizons.

## 4.6 Empirical Results

Tables 4.4 and 4.5 contain the results of various out-of-sample prediction experiments where the target variables are both inflation and output growth. The first column in both tables

contains the names of the models considered. For a complete explanation of the models considered, refer to Table 4.3 and the above discussion. Numerical entries in the second column are mean squared forecast errors (MSFEs). Those in bold correspond to models with lower MSFEs, relative to Model 1. Boxed MSFEs represent the lowest MSFE amongst all models considered.

Our other "benchmark" model against which the out-of-sample forecast performance of all alternative models is compared is the Stock and Watson (SW) diffusion index model. The SW model is a useful benchmark because of its well established superior out-of-sample forecast performance as well as its flexibility and efficiency at distilling useful information from large macroeconomic datasets. Numerical entries in the third column of the tables report the DM test statistic where the benchmark model is the Stock and Watson factor model. Since the factor model and all the alternative models are non-nested, these DM test statistics have a standard normal limiting distribution (see Corradi and Swanson (2006b)). Negative entries show instances where the benchmark SW factor model has a lower point MSFE than the respective alternative model and positive entries show the converse. DM statistics with $^{*,**,***}$ signs denote models where the null of equal predictive accuracy is rejected at 20%(*), 10%(**) and 5%(***) significance levels, respectively. Numerical entries in the remaining two columns of the tables report DM test results for different benchmark models.

In Table 4.4, 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. 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 4.5, 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 are "MSFE-better" 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 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. Namely, models that use actual observable variables corresponding to those the Fed is interested in (as well as our additional spread variables), and that are hence useful for policy monitoring, are also the models yielding the best predictions. Thus, the fusion of the latest diffusion index methodology with the monitoring objectives of monetary regulators yields models that use variables relevant for both prediction and policy analysis.

Note that DM test statistics calculated under "No Spread" report on predictive accuracy tests, where the benchmark model is simply the model listed in the row in which the statistic is reported, minus spread variables. Negative entries thus indicate instances where the benchmark no spread model outperforms the alternative model, whereas positive entries indicate the converse. Note that the models in these comparisons are nested, and hence the DM test limiting distribution is no longer standard normal, as discussed in Clark and McCracken (2005) and McCracken (2007), and hence we use critical values tabulated by these authors. Moreover, Clark and West (2006, 2007) observed that the parsimonious model in such cases has a smaller MSFE because it gains efficiency by setting to zero parameters that are zero in the population. The alternative model on the other hand, in finite samples, inflates its MSFE by virtue of parameter estimation error introduced into the forecasting process. Thus, under the null hypothesis, the extra variables in the alternative model impair prediction, and the MSFE of the parsimonious model should effectively be smaller than that of the alternative model in finite samples, under equality (see Clark and

West (2007)). Interestingly, our larger models that include spreads actually yield smaller MSFEs in many cases, and all cases where there is a significant difference in MSFEs favor the spread-augmented models, yielding further evidence as to the usefulness of the spread variables. However, it should 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 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) to suggest that spreads have marginal predictive content for inflation at longer horizons. It is consequently conclusive from Table 4.4 that spreads generally have strong marginal predictive content for CPI inflation at the 12 step horizon but are not too helpful at the shorter horizons. At the 12 step horizon, models with spreads significantly outperform the corresponding restricted models without spreads for almost all model specifications. 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.

Recall our earlier discussion of the work of Clark and West, where we commented that parameter estimation error must be carefully assessed when comparing forecast performance. 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. Evidence in support of this point is provided in Table 4.4, 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 4.5, 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 4.4 and 4.5 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 predictions 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. One of the main driving arguments behind the principle of parsimony certainly involves efficiency gains associated with parameter estimation error reduction. This principle is valid so long as the extra variables are redundant and do not contain too "much predictive" content for the target variable. The above analysis suggests that the trade-off between predictive content and parameter estimation error is tipped in favor of predictive content when considering the spreads discussed in this paper.

In Panel 1 of Figure 4.1, 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 4.1, 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 the macroeconomic indicators perform best when forecasting inflation in non-volatile time periods. On the contrary, 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 in this case, there is not much of a discrepancy between the MSFEs of Models 1 and 1S when forecasting output growth (see Panel 2 of Figure 4.2).

Finally, notice that for both inflation and output, the lowest MSFE occurs for one of our even numbered models, at all forecast horizons. As specified in Tables 4.2 and 4.3, 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.

## 4.7 Concluding Remarks

In order to obtain early indication of the impact of current monetary policy, the Federal Reserve monitors select financial and macroeconomic variables. This practice is discussed on the Federal Reserve Bank of New York website. We lend credence to the set of "macroeconomic indicators" by establishing that they are largely the same as those variables that proxy diffusion indices (factors) constructed via analysis of a largescale macroeconomic dataset. Out-of-sample forecast exercises further suggest that augmenting the macroeconomic indicators with certain spreads is in some cases useful when forecasting inflation or output growth. The particular spreads found to be valuable 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 were not found to provide additional predictive content. Moreover, we find that the macroeconomic indicators perform 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.

Table 4.1: **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 | $\Delta$levels |
| 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 | $\Delta$levels |
| 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 | $\Delta$levels |
| 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 | $\Delta$levels |
| 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 | $\Delta$levels |
| Yield on 6-month Treasury Bill | $\Delta$levels |
| Yield on 1-year Treasury Bond | $\Delta$levels |
| 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 | $\Delta$levels |

Table 4.1 (continued)

| Regressors | Stationarity Transformtion |
|---|---|
| Panel D: Model 4 ($W_{4t}$) | |
| Housing Starts: northeast | log |
| Housing Authorized by Building Permits: northeast | log |
| Industrial Production Index: manufacturing | $\Delta \log$ |
| Industrial Production Index: nondurable consumer goods | $\Delta \log$ |
| Capacity Utilization | $\Delta$levels |
| Yield on 3-month Treasury Bill | $\Delta$levels |
| Yield on 6-month Treasury Bill | $\Delta$levels |
| Yield on 10-year Treasury Bond | $\Delta$levels |
| S&P 500 Stock Price Index: Composite | $\Delta \log$ |
| S&P 500 Stock Price Index: Industrials | $\Delta \log$ |
| Panel E: Model 5 ($W_{5t}$) | |
| Housing Starts: total farm and nonfarm | log |
| Industrial Production Index: total index | $\Delta \log$ |
| Capacity Utilization | $\Delta$levels |
| S&P 500 Stock Price Index: Composite | $\Delta \log$ |
| Panel F: Model 6 ($W_{6t}$) | |
| Housing Starts: northeast | log |
| Industrial Production Index: manufacturing | $\Delta \log$ |
| Capacity Utilization | $\Delta$levels |
| S&P 500 Stock Price Index: Composite | $\Delta \log$ |
| Panel G: Model 7 ($W_{7t}$) | |
| Housing Starts: total farm and nonfarm | log |
| Industrial Production Index: total index | $\Delta \log$ |
| Capacity Utilization | $\Delta$levels |
| Yield on 6-month Treasury Bill | $\Delta$levels |
| S&P 500 Stock Price Index: Composite | $\Delta \log$ |
| Money Supply - M2 | $\Delta \log$ |
| Panel H: Model 8 ($W_{8t}$) | |
| Housing Starts: northeast | log |
| Industrial Production Index: manufacturing | $\Delta \log$ |
| Capacity Utilization | $\Delta$levels |
| Yield on 6-month Treasury Bill | $\Delta$levels |
| S&P 500 Stock Price Index: Composite | $\Delta \log$ |
| Money Supply - M2 | $\Delta \log$ |
| 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 Bonds − 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) | $\Delta \log$ |
| Industrial Production: Manufacturing (Industrial Production Class) | $\Delta \log$ |
| Yield on 6-month Treasury Bill (Yield Class) | $\Delta$levels |
| Nonfarm Payroll Employment: goods producing (Employment Class) | $\Delta \log$ |

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

Table 4.2: **Prediction Models**

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

Table 4.3: **Description of Prediction Models Used in Empirical Experiments**

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

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

Model 2: $\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}'W_{2t} + \sum_{j=1}^{p} \widehat{\beta}_j y_{t-j+1}$: 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 $\widehat{X}_1$: CPI; $\widehat{X}_2$: Industrial Production; $\widehat{X}_3$: Housing; $\widehat{X}_4$: Employment; $\widehat{X}_5$: 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 $\widehat{F}_i$ $\forall i$. $W_{2t}$ 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 selelcted 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.

Table 4.3 (continued)

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 Fed's "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.

For Model $nS$ ( $n = 1, \ldots, 8$), the forecast model is specified as

$$\widehat{y}_{t+h} = \widehat{c} + \widehat{\alpha}' W_{nt}^{S} + \sum_{j=1}^{p} \widehat{\beta}_j y_{t-j+1}$$

where $W_{nt}^{S} = (W_{nt}', S_t')'$ and $S_t$ is a vector of spreads listed in Panel I of Table 1.

Notes: In Model $n$ ($n = 2, 4, 6, 8$), variables selected by the methodology discussed in Table 4.2 above were used where possible to represent some classes. The specific variables and their corresponding classes are listed in Panel J of Table 4.1. With Model $n$ ($n = 1, 3, 5, 7$), aggregate variables are used by default as class representatives. All variables contained in $W_{nt}$ are listed in Table 4.1. The SIC is used to select a value of $p = 1$ for ouput growth forecasting and $p = 3$ for inflation forecasting.

Table 4.4: **Forecast of 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 4.4 (continued)

| 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: Numerical entries in the second column represent the mean squared forecast errors (MSFEs) of recursively constructed ex ante predictions for the period 1994:09-2003:12, using the models listed in the first column (see Table 4.2 for an explanation of the different models). Entries in bold font represent lower MSFEs relative to the assumed Fed model (Model 1). Boxed MSFE entries represent the lowest MSFE value amongst all considered models. All numerical entries in the last three columns are MSFE Diebold and Mariano (DM) predictive accuracy test statistics. Negative DM statistics represent instances where the benchmark model (noted at the top of the table) has a lower MSFE relative to the respective model. Positive entries indicate the opposite scenario. Models with DM statistics that have $^{*,**,***}$ signs significantly outperform the respective benchmark model at the 20%($^{*}$), 10%($^{**}$) and 5%($^{***}$) significance levels. In calculating the DM statistic under "No Spread", the benchmark model is the respective model restricted to exclude spreads. This DM statistic consequently reflects the marginal predictive content of spreads for that particular model specification.

Table 4.5: **Forecast of 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 |

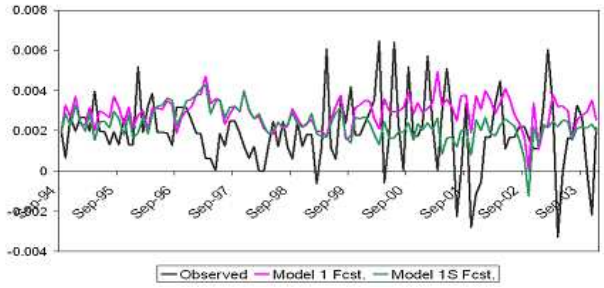
Table 4.5 (continued)

| 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 4.4.

Figure 4.1: **Forecast of CPI Inflation by Benchmark Model 1 at the 12 Step Horizon**

Panel 1: Observed and Forecast Values of CPI Inflation




Panel 2: Forecast Squared Errors of CPI Inflation

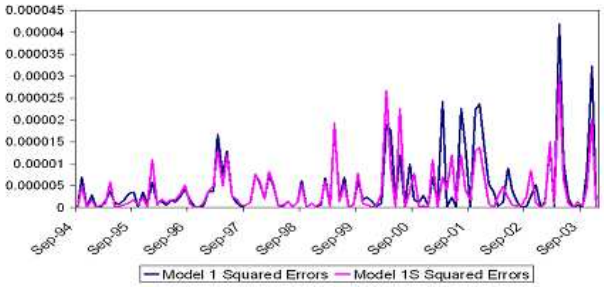
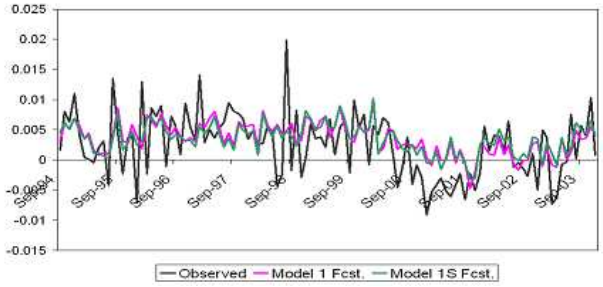
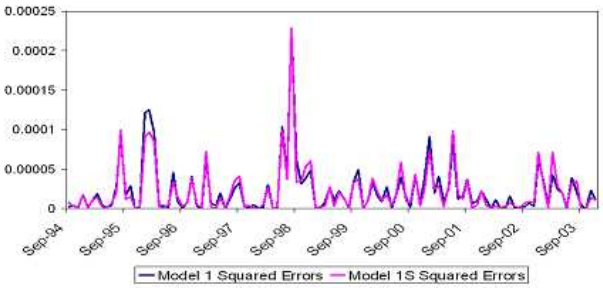


Figure 4.2: **Forecast of Output Growth by Benchmark Model 1 at the 1 Step Horizon**

Panel 1: Observed and Forecast Values of Output Growth




Panel 2: Forecast Squared Errors of Output Growth



# References


[1] Amato, J.D. and Swanson, N.R. (2001). The Real Time Predictive Content of Money for Output. Journal of Monetary Economics 48, 3-24.

[2] Armah, N. and Swanson, N. (2008). Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Large Scale Macroeconomic Time Series Environments. Working paper, Rutgers University.

[3] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. Econometrica 71, 135-172.

[4] Bai, J. (2003). Testing Parametric Conditional Distributions of Dynamic Models. Review of Economics and Statistics 85, 531-549.

[5] Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. Econometrica 70, 191-221.

[6] Bai, J. and Ng, S. (2006a). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. Econometrica 74, 1133-1150.

[7] Bai, J. and Ng, S. (2006b). Evaluating Latent and Observed Factors in Macroeconomics and Finance. Journal of Econometrics 113, 507-537.

[8] Bai, J. and Ng, S. (2006c). Forecasting Economic Time Series Using Targeted Predictors. Working Paper, University of Michigan, Ann-Arbor.

[9] Bai, J. and Ng, S. (2006d). Boosting Diffusion Indices. Working Paper, University of Michigan, Ann-Arbor.

[10] Bai, J. and Ng, S. (2007). Determining the Number of Primitive Shocks in Factor Models. Journal of Business and Economic Statistics 25, 52-60.

[11] Battini, J. and Haldane, A. (1999). Forward-looking Rules for Monetary Policy. In Taylor, J., (ed.), Monetary Policy Rules. University of Chicago Press for NBER, Chicago.

[12] Bernanke, B. (1990). On the Predictive Power of Interest Rates and Interest Rates Spreads. New England Economic Review, 51-68.

[13] Bernanke, B. and Boivin, J. (2003). Monetary Policy in a Data-Rich Environment. Journal of Monetary Economics 50, 525-546.

[14] Bierens, H.B. (1982). Consistent Model Specification Tests. Journal of Econometrics 20, 105-134.

[15] Bierens, H.B. (1990). A Conditional Moment Test of Functional Form. Econometrica 58, 1443-1458.

[16] Bierens, H.J. and Ploberger, W. (1997). Asymptotic Theory of Integrated Conditional Moment Tests. Econometrica 65, 1129-1152.

[17] Boivin, J. and Ng, S. (2005). Understanding and Comparing Factor Based Macroeconomic Forecasts. International Journal of Central Banking 1, 117-152.

[18] 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, 37-53.

[19] Breeden, D., Gibbons, M. and Litzenberger, R. (1989). Empirical Tests of the Consumption-Oriented CAPM. Journal of Finance XLIV, 231-262.

[20] Browne, F. and Manasse, P. (1989). Information Content of the Term Structure of Interest Rates: Theory and Practice. OECD Working Paper No. 69.

[21] Buhlmann, P. and Hothorn, T. (2006). Boosting Algorithms: Regularization, Prediction and Model Fitting. Working Paper, ETH Zürich.

[22] Chao, J., Corradi, V. and Swanson, N. (2001). An Out of Sample Test for Granger Causality. Macroeconomic Dynamics 5, 598-620.

[23] Christiano, L.J. and Ljungqvist, L. (1988). Money Does Granger-Cause Output in the Bivariate Money-Output Relation. Journal of Monetary Economics 22, 217-235.

[24] Clarida, R, Gali, G. and Gertler, M. (1999). The Science of Monetary Policy: A new Keynesian Perspective. Journal of Economic Literature 37, 1661-1707.

[25] Clarida, R, Gali, G. and Gertler, M. (2000). Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. Quarterly Journal of Economics 115, 147-180.

[26] Clark, T.E., and McCracken, M.W. (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. Journal of Econometrics 105, 85-110.

[27] Clark, T. and McCracken, M. (2005). Evaluating Direct Multi-Step Forecasts, Econometric Reviews, 24, 369-404.

[28] Clark, T. and West, K. (2006). Usinig Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis. Journal of Econometrics 135, 155-186.

[29] Clark, T. and West, K. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. Journal of Econometrics 138, 291-311.

[30] Clements, M.P. and Smith, J. (2000). Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment. Journal of Forecasting 19, 255-276.

[31] Clements, M.P. and Smith, J. (2002). Evaluating Multivariate Forecast Densities: A Comparison of Two Approaches. International Journal of Forecasting 18, 397-407.

[32] Connor, G., and Korajczyk, R. (1988). Risk and Return in an Equilibrium APT: Application of a New Test Methodology. Journal of Financial Economics 21, 255-289.

[33] Connor, G., and Korajczyk, R. (1993). A Test for the Number of Factors in an Approximate Factor Model. Journal of Finance 48, 1263-1291.

[34] Corradi, V., and Swanson, N.R. (2002). A Consistent Test for Out of Sample Nonlinear Predictive Ability. Journal of Econometrics 110, 353-381.

[35] Corradi, V. and Swanson, N.R. (2004). Some Recent Developments in Predictive Accuracy Testing with Nested Models and (Generic) Nonlinear Alternatives. International Journal of Forecasting 20, 185-199.

[36] Corradi, V. and Swanson, N.R. (2005). Predictive Density and Confidence Intervals Accuracy Tests. Journal of Econometrics forthcoming.

[37] Corradi, V. and Swanson, N.R. (2006a). Predictive Density Evaluation. In C. Granger, G.Elliot and A. Timmerman (Eds.) Handbook of Economic Forecasting (pp. 197-284). Amsterdam: Elsevier.

[38] Corradi, V. and Swanson, N.R. (2006b). Bootstrap Conditional Distribution Tests In the Presence of Dynamic Misspecification. Journal of Econometrics 133, 779-806.

[39] Corradi, V. and Swanson, N.R. (2006c). Predictive Density and Conditional Confidence Intervals Accuracy Tests. Journal of Econometrics 135, 187-228.

[40] Corradi, V. and N.R. Swanson, N.R. (2007). Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes. International Economic Review forthcoming.

[41] Davis, E. (1993). VAR Modeling of the German Economy with Financial Spreads as Key Indicator Variables. London School of Economics, Discussion Paper 59.

[42] 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.

[43] Davis, E. and Henry, S. (1992). An Aggregate VAR Model with Financial Spreads. Working Paper. Bank of England.

[44] Davis, E. and Henry, S. (1993). The Use of Financial Spreads as Indicators of Real Activity. In Arestis, P. (ed.), Money and Banking: Issues for the Twenty-First Century, London, Macmillan.

[45] Davis, E. and Henry, S. (1994). The Use of Financial Spreads as Indicator Variables: Evidence for the UK and Germany. IMF, Working Paper 31.

[46] 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, 374-394.

[47] Diebold, F.X., T. Gunther and Tay, A.S. (1998). Evaluating Density Forecasts with Applications to Finance and Management. International Economic Review 39, 863-883.

[48] Diebold, F.X., Hahn, J. and Tay, A.S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High Frequency Returns on Foreign Exchange. Review of Economics and Statistics 81, 661-673.

[49] Diebold, F. and Mariano, R. (1995). Comparing Predictive Accuracy. Journal of Business and Economic Statistics. 13, 253-263.

[50] Diebold, F.X., A.S. Tay and Wallis, K.D. (1998). Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters, in Festschrift in Honor of C.W.J. Granger, eds. R.F. Engle and H. White, Oxford University Press, Oxford.

[51] 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.

[52] Estrella, A. and Hardouvelis, G. (1991). The Term Structure as a Predictor of Real Economic Activity. Journal of Finance 46, 555-576.

[53] 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, 1375-1401.

[54] Estrella, A. and Mishkin, F. (1998). Predicting U.S. Recessions: Financial Variables as Leading Indicators. Review of Economics and Statistics 80, 45-61.

[55] Fitzenberger, B. (1997). The Moving Block Bootstrap and Robust Inference for Linear Least Square and Quantile Regressions. Journal of Econometrics 82, 235-287.

[56] 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, 540-552.

[57] 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, 830-840.

[58] Forni, M. and Reichlin, L. (1996). Dynamic Common Factors in Large Cross-Sections. Empirical Economics 21, 27-42.

[59] 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.

[60] Freund, Y. (1995). Boosting a Weak Learning Algorithm by Majority. Information and Computation 121, 256-285.

[61] Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29, 1189-1232.

[62] Friedman, B. and Kuttner, K. (1991). Why Does the Paper-Bill Spread Predict Real Economic Activity? NBER Working Paper, No. 3879.

[63] Friedman, B.M. and Kuttner, K.N. (1993). Another Look at the Evidence on Money-Income Causality. Journal of Econometrics 57, 189-203.

[64] Gallant, A.R. and White, H. (1988). A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Oxford: Blackwell.

[65] Garnett, T., Hall, S. and Henry, S. (1992). Measuring and Forecasting Underlying Economic Activity. London School of Economics, Discussion Paper 18.

[66] Gerlach, S. (1997). The Information Content of the Term Structure: Evidence for Germany. Empirical Economics 22, 161-179.

[67] Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series. In Aigner, D. and Goldberger, A. (eds.), Latent Variables in Socio-Economic Models. Amsterdam.

[68] Giacomini, R. and White H. (2003). Conditional Tests for Predictive Ability. Manuscript, University of California, San Diego.

[69] Goncalves, S. and White, H. (2004). Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models. Journal of Econometrics 119, 199-219.

[70] Harvey, C. (1988). The Real Term Structure and Consumption Growth. Journal of Financial Economics 22, 305-330.

[71] Harvey, C. (1989). Forecasts of Economic Growth from the Bond and Stock Markets. Financial Analysts Journal 45, 38-45.

[72] Harvey, D.I., Leybourne, S.J. and Newbold, P. (1997). Tests for Forecast Encompassing. Journal of Business and Economic Statistics 16, 254-259.

[73] Hayashi, F. (2000). Econometrics. Princeton University Press, Princeton.

[74] Inoue, A. and Rossi, B. (2004). Recursive Predictive Ability Tests for Real Time Data. Working Paper, Duke University and NC State.

[75] Johansen, S. (1988). Statistical Analysis of Cointegrating Vectors. Journal of Economic Dynamics and Control 12, 231-254.

[76] Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica 59, 1551-1580.

[77] Jorion, P. and Mishkin, F. (1991). A Multi-Country Comparison of Term Structure Forecasts at Long Horizons. Journal of Financial Economics 29, 59-80.

[78] Kilian, L. (1999). Exchange Rate and Monetary Fundamentals: What do we Learn from Long-Horizon Regressions. Journal of Applied Econometrics 14, 491-510.

[79] Kozicki, S. (1997). Predicting Real Growth and Inflation With the Yield Spread. Federal Reserve Bank of Kansas City Economic Review 82, 39-57.

[80] Künsch, H.R. (1989). The Jackknife and the Bootstrap for General Stationary Observations, Annals of Statistics 17, 1217-1241.

[81] Laurent, R. (1988). An Interest Rate-Based Indicator of Monetary Policy. Federal Reserve Bank of Chicago Economic Perspectives 12, 3-14.

[82] Laurent, R. (1989). Testing the Spread. Federal Reserve Bank of Chicago Economic Perspectives 13, 22-34.

[83] McCracken, M. (2007). Asymptotics for Out-of-Sample Tests of Granger Causality. Journal of Econometrics 140, 719-752.

[84] Mishkin, F. (1990a). The Information in the Longer-Maturity Term Structure About Future Inflation. Quarterly Journal of Economics 55, 815-828.

[85] Mishkin, F. (1990b). What Does the term Structure Tell Us About Future Inflation. Journal of Monetary Economics 25, 77-96.

[86] Miskin, 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, 2-22.

[87] Newey, W. and West, K. (1987). A Simple Positive-Definite Heteroskedasticity and Autocorelation Consistent Covariance Matrix. Econometrica 55, 703-708.

[88] Nobili, A. (2005). Forecasting Output Growth and Inflation in the Euro Area: Are Financial Spreads Useful? Banca D'Italia, No. 544.

[89] Pesaran, M.H. and Timmerman, A. (2004a). How Costly is to Ignore Breaks when Forecasting the Direction of a Time Series? International Journal of Forecasting 20, 411-425.

[90] Pesaran, M.H. and Timmerman, A. (2004b). Selection of Estimation Window for Strictly Exogenous Regressors. Working Paper, Cambridge University and University of California, San Diego.

[91] Plosser, C. and Rouwenhorst, K. (1994). International Term Structures and Real Economic Growth. Journal of Monetary Economics 33, 133-155.

[92] Rapach, D. and Strauss, J. (2007). Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Unemployment Growth. Econometric Reviews, (forthcoming).

[93] Sargent, T., and Sims, C. (1977). Business Cycle Modeling without Pretending to Have Too Much A-Priori Economic Theory. In Sims, C. et al. (eds.), New Methods in Business Cycle Research. Minneapolis: Federal Reserve Bank of Minneapolis.

[94] Schapire, R. (1990). The Strength of Weak Learnability. Machine Learning 5, 197-227.

[95] Schorfheide, F. (2004). VAR Forecasting under Misspecification. Journal of Econometrics, (forthcoming).

[96] Shanken, J. (1992). On the Estimation of Beta-Pricing Models. Review of Financial Studies 5, 1-33.

[97] Stinchcombe, M.B. and White, H. (1998). Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative. Econometric Theory 14, 295-325.

[98] Stock, J. and Watston, 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.

[99] Stock, J. and Watson, M. (1998). Diffusion Indexes. Working Paper 6702, National Bureau of Economic Research.

[100] Stock, J. and Watson, M. (1999). Forecasting Inflation. Journal of Monetary Economics 44, 293-335.

[101] Stock, J. and Watson, M. (2002a). Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics 20, 147-161.

[102] Stock, J. and Watson, M. (2002b). Forecasting Using Principal Components from a Large Number of Predictors. Journal of American Statistical Association 97, 1167-1179.

[103] Stock, J. and Watson, M. (2004a). An Empirical Comparison of Methods for Forecasting Using Many Predictors. Working Paper, Princeton University.

[104] Stock, J. and Watson, M. (2004b). Forecasting with Many Predictors. Handbook of Forecasting, (forthcoming).

[105] Stock, J. and Watson, M. (2005). Implications of Dynamic Factor Models for VAR Analysis. Working Paper 11467, National Bureau of Economic Research.

[106] Swanson, N.R. (1998). Money and Output Viewed Through a Rolling Window. Journal of Monetary Economics 41, 455-474.

[107] Swanson, N.R., Ozyildirim, A. and Pisu, M.(2003). A Comparison of Alternative Causality and Predictive Ability Tests in the Presence of Integrated and Cointegrated Economic Variables. In D. Giles (Ed.), Computer Aided Econometrics 91-148. New York: Marcel Dekker.

[108] 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.

[109] 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.

[110] Taylor, J. (1993). Discretion Versus Policy Rules in Practice. Carnegie Rochester Conference Series on Public Policy 39, 195-214.

[111] Thoma, M.A. (1994). Subsample Instability and Asymmetries in Money-Income Causality. Journal of Econometrics 64, 279-306.

[112] West, K. (1996). Asymptotic Inference About Predictive Ability. Econometrica 64, 1067-1084.

[113] West, K. (2001). Tests for Forecast Encompassing When Forecasts Depend on Estimated Regression Parameters. Journal of Business and Economic Statistics 19, 29-33.

[114] West, K. (2005). Forecast Evaluation. In Elliott, G., Granger, C. and Timmermann, A. (eds.), Handbook of Economic Forecasting, (forthcoming).

[115] White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica 48, 817-838.

[116] White, H. (2000). A Reality Check for Data Snooping. Econometrica 68, 1097-1126.

# Vita

## Nii Ayi Christian Armah

**2005-2009** Ph.D. in Economics, Rutgers University

**2003-2005** M.A. in Economics, Rutgers University

**1999-2002** B.A. in Economics and Mathematics, cum laude, Knox College

**2007-2009** Part Time Lecturer, Department of Economics, Rutgers University

**2004-2007** Teaching assistant, Department of Economics, Rutgers University