NONLINEAR FORECASTING ANALYSIS USING DIFFUSION INDEXES: AN APPLICATION TO JAPAN

by

Mototsugu Shintani

Working Paper No. 03-W22

October 2003

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Nonlinear Forecasting Analysis Using Diffusion Indexes: An Application to Japan*

Mototsugu Shintani†

January 2002

This version: October 2003

Abstract

This paper extends the diffusion index (DI) forecast approach of Stock and Watson (1998, 2002) to the case of possibly nonlinear dynamic factor models. When the number of series is large, a two-step procedure based on the principal components method is useful since it allows the wide variety of the nonlinearity in the factors. The factors extracted from a large Japanese data suggest some evidence of nonlinear structure. Furthermore, both the linear and nonlinear DI forecasts in Japan outperform traditional time series forecasts, while the linear DI forecast, in most cases, performs as well as the nonlinear DI forecast.

Keywords: Diffusion Index; Dynamic Factor Model; Nonlinearity; Prediction

JEL classification: F31; F41

*A large part of this research was conducted while the author was visiting the Economic and Social Research Institute (ESRI), the Cabinet Office of the Japanese government. The author would like to thank two referees, Todd Clark, Yanqin Fan, Koichi Hamada, Yasutomo Murasawa, Serena Ng, Mark Watson, Ken West, and participants of the 2002 NBER Summer Institute, the Common Features in Rio conference, and the 2003 Midwest Econometrics Group meetings, for their helpful comments and discussions. The author also thanks the officers of ESRI for their help in obtaining the data set.

†Department of Economics, Vanderbilt University, Nashville, TN 37235, USA e-mail: mototsugu.shintani@vanderbilt.edu

1 Introduction

When a very large number of time series are available, forecasters should take advantage of all the usable information rather than restricting their attention to a small subset of the whole list of variables. Following this idea, Stock and Watson (1998, 2002) extracted common factors by applying the principal components method to 215 economic time series, and then showed the significant predictive ability of the estimated factors in the out-of-sample forecasts of several key U.S. macroeconomic variables. Since the estimated factors naturally fit the notion of diffusion indexes developed at the National Bureau of Economic Research (NBER), Stock and Watson called their method the diffusion index (DI) forecast. This DI forecast approach can be justified under the framework of the dynamic factor model originally considered by Sargent and Sims (1977) and Geweke (1977). When the common factor is generated from a linear time series model, employing a linear forecasting regression seems to be the most appropriate procedure. Alternatively, if the dynamic factor model has a nonlinear structure, we may gain from considering nonlinearity in the forecasting regression.

The first goal of this paper is to consider a simple procedure to estimate the nonlinear time series model of common factors and to test its nonlinear functional form nonparametrically. Instead of estimating the full model simultaneously, we focus on a two-step procedure, namely, the estimation of the factors by principal components, followed by the estimation of the dynamic factor structure using estimated factors. We emphasize that such a two-step method is useful and convenient in a nonlinear framework since the principal components method in the first step remains valid under the very flexible nonlinear dynamic factor structure. In particular, for both linear and nonlinear models, when the number of the series (N) increases at a sufficiently fast rate compared to the time series observations (T), the effect of the estimation error in the first step is negligible in the asymptotic property of the final estimators or the statistics of the specification tests.

The second goal of this paper is to explore the possibility of improving the performance of the DI forecast by incorporating the nonlinearity in the forecasting regression. We follow Stock and Watson and include the common factors estimated in the first step as predictors of the variable of interest in the second step. However, in addition to the linear forecasting model, we consider the nonlinear forecasting model and the combination of the two. We use several tests to evaluate the forecasting performance. By the same argument used in the estimation of the dynamic factor structure, we can expect the first step estimation error to have no effect on the criteria of forecasting performance and test statistics given sufficiently large N.

In this paper, we consider two different means of utilizing the estimated factors in the second step — estimation of the factor structure and estimation of the forecasting model. Theoretically, a nonlinear factor structure implies a nonlinear forecasting model. However, in practice, even if we detect a nonlinear factor structure, neglected nonlinearity may have only a marginal effect in forecasting. In other words, linear approximation of the model may be sufficient for forecasting purposes. The usefulness of our procedure is, therefore, more or less an empirical question. As an empirical example of our method, we apply it to a forecasting analysis of the Japanese economy.

We find that using Japanese data instead of U.S. data is well-motivated for the following reasons. First, the empirical success of DI forecast in the U.S. by Stock and Watson raised the question of whether such a procedure would also work well for other countries. Regardless of linear or nonlinear, additional evidence from Japan can be used to evaluate the general applicability of the DI forecast procedure.[1] Second, a reliable forecasting model of Japanese aggregate activity is currently being highly sought after among forecasters given the fact that major public and private research institutes failed to provide a satisfactory

[1] The usefulness of the DI forecast was recently supported by the European data (Marcellino, Stock and Watson, 2003).

forecast of business cycles and prolonged recessions during the 1990s (See Fukuda and Onodera, 2001).

Incorporating nonlinearity in a dynamic factor model is certainly not new in the literature. One of the most popular approaches in practice is to introduce Hamilton's (1989) Markov switching structure of a common factor mainly for the purpose of estimating the turning points in business cycles (Kim and Nelson, 1998, for example). This class of nonlinear model is also considered in the context of large N factor model by Diebold (2003) who suggested estimating the Markov switching model in the second step using the estimated factors by principal components in the first step. In contrast to Diebold who employed a parametric model, we use a nonparametric approach to allow flexibility in the nonlinear dynamic factor structure. Among many available nonparametric methods, we employ the artificial neural networks (ANNs). This particular estimation method has been widely used in studies on the forecasting performance of nonlinear models, including Swanson and White (1997), Chen, Racine, and Swanson (2000), Hong and Lee (2002).

The remainder of the paper is organized as follows: Section 2 explains the model. Section 3 provides the empirical results using Japanese data. Some concluding remarks are made in Section 4.

2 Model

In this section, we introduce a dynamic factor model that will be the basis of the linear and nonlinear DI forecasts. Because our purpose here is mainly the illustration of model structure rather than providing propositions and proofs for the general case, we consider only a single factor generated from an autoregressive (AR) model of order one. Nevertheless, the model can be extended to the multiple factor model and/or the AR model of higher order, which will be used in the empirical section. In what follows, we first describe a linear model that has been employed in typical applications, then introduce the model with a nonlinear structure.

Let x_{it} be an i-th component of N-dimensional multiple time series $X_t = (x_{1t}, \ldots, x_{Nt})'$ and $t = 1, ..., T$. A simple dynamic factor model associates each x_{it} with a scalar common factor f_t^0 in equations

$$x_{it} = \lambda_i^0 f_t^0 + e_{it}, \quad i = 1, ..., N, \tag{1}$$

$$f_t^0 = \phi f_{t-1}^0 + \varepsilon_t \tag{2}$$

where λ_i^0's are factor loadings with respect to i-th series, e_{it}'s are idiosyncratic shocks, $|\phi| < 1$, $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ where $\mathcal{F}_{t-1}$ is a filtration. While the factor f_t^0 is not directly observable, the model can be estimated by the maximum likelihood method combined with the Kalman filter technique if distribution of e_{it}'s and ε_t is specified (see Stock and Watson, 1989, for example). Alternatively, the model can be estimated by a two-step procedure with the factors (and factor loadings) being estimated by the principal components method in the first step [the measurement equation (1)], followed by the estimation of time series models of the factors in the second step [the transition equation (2)].[2] The latter method is particularly convenient when e_{it}'s are cross-sectionally and/or serially correlated and when both N and T are large. A recent large N asymptotic theory developed by Stock and Watson (1998), Bai and Ng (2002), and Bai (2003) shows that, under mild conditions on moments and memory, the principal components estimator $\tilde{f}_t$ is a consistent estimator of f_t^0 up to a scaling constant.[3] In addition, the $\sqrt{N}$-consistency of the first-step estimator $\tilde{f}_t$ can be used to show $\hat{\phi} - \tilde{\phi} = o_p(T^{-1/2})$ where $\hat{\phi}$ and $\tilde{\phi}$ are the ordinary least squares (OLS) estimators of ϕ in (2) based on f_t^0 and $\tilde{f}_t$, respectively. It implies that the infeasible estimator $\hat{\phi}$ can be replaced by the two-step estimator $\tilde{\phi}$ since the estimation error in the first step is negligible in

[2] For example, such a two-step procedure was considered by Kariya (1993) under the name of Multivariate Time-Series Variance-Component (MTV) model.

[3] Instead of using standard principal components, Forni, Hallin, Lippi, and Reichlin (2000) proposed an alternative estimator based on dynamic principal components and showed its consistency under $N, T \to \infty$. However, we do not use dynamic principal components because their result requires an assumption of linearity in the dynamic factor structure.

the limiting distribution and thus in the inference. One obvious sufficient condition is $N/T \to \infty$ while slower N is also possible (see Shintani, 2003, for a more formal discussion). This dynamic factor structure can also be used to construct the h-step ahead forecast of a scalar series y_t being generated by

$$y_t = \lambda_0^0 f_t^0 + e_{0t} \tag{3}$$

with $E(e_{0t}^0|\mathcal{F}_{t-1}) = 0$. A simple calculation leads to the representation

$$y_{t+h} = \beta_h f_t^0 + u_{t+h} \tag{4}$$

where $\beta_h = \lambda_0^0 \phi^h$ and $u_{t+h} = \lambda_0^0 \left[f_{t+h}^0 - E(f_{t+h}^0|\mathcal{F}_t) \right] + e_{0t+h}^0$. Therefore, the optimal h-step ahead forecast at T is $y_{T+h|T} = E(y_{T+h}|\mathcal{F}_T) = \beta_h f_T^0$. While f_t^0's are not observable, researchers have two options. One is to use the two-step method described above and roll equation (2) forward using the second step estimator $\hat{\phi}$. The other approach is to run a forecasting regression (4) with f_t^0 replaced by $\hat{f}_t$ instead of estimating (2) in the second step. The feasible forecast then is given by $\hat{y}_{T+h|T} = \hat{\beta}_h \hat{f}_T$ where $\hat{\beta}_h$ is the OLS estimator of β_h. Stock and Watson (1998) recommended using the latter approach and showed that $\hat{y}_{T+h|T}$ was asymptotically equivalent to $y_{T+h|T}$ as $N, T \to \infty$.

Let us now turn to the model with a nonlinear dynamic factor structure replacing the linear dynamics in (2). For example, the observed common asymmetricity of x_{it}'s in expansions and contractions can be a motivation of introducing nonlinearity in f_t^0. To incorporate such a nonlinearity, a Markov switching structure of common factor has been often employed in the empirical studies of business cycles. Just as in the linear case, the system of two equations can be simultaneously estimated by the maximum likelihood method (Kim and Nelson, 1998), or they can be estimated in two steps using principal components method in the first step (Diebold, 2003). It is important to note that the $\sqrt{N}$-consistency result of the principal components estimator $\hat{f}_t$ can be derived under some moment conditions of f_t^0 without using an assumption of linearity in f_t^0 (Stock and Watson, 1998, Bai and Ng, 2002, and Bai, 2003). Therefore, the principal components estimator remains valid under a very general nonlinear dynamic factor structure and is less subject to the misspecification problem.[4] In addition, the two-step method is also preferable from the computational point of view since the simultaneous estimation method becomes computationally difficult with an increasing number of parameters as $N \to \infty$.

Suppose a common factor f_t^0 is generated by the following nonlinear AR(1) model,

$$f_t^0 = m\left(f_{t-1}^0\right) + \varepsilon_t \tag{5}$$

where $m\left(f_{t-1}^0\right) = E(f_t^0|\mathcal{F}_{t-1}) = E(f_t^0|f_{t-1}^0)$ is a conditional mean function. This nonlinear AR model can be estimated by a parametric method if function m is specified. Alternatively, it can be estimated by a nonparametric method without specifying the functional form of m. Here we take the latter approach and consider a nonparametric estimator of (5) with a convergence rate T^{δ} where $0 < \delta < 1/2$. Then, the consistency result of factors, along with some conditions on the smoothness of m function, the speed of N, and the controlling parameter of the nonparametric method can be used to derive $\widehat{m}(f) - \widetilde{m}(f) = o_p(T^{-\delta})$ where $\widehat{m}(f)$ and $\widetilde{m}(f)$ are the infeasible and feasible nonparametric estimators of $m\left(f_{t-1}^0\right)$ evaluated at $f_{t-1}^0 = f$, respectively, analogous to the linear estimators $\widehat{\phi}$ and $\widetilde{\phi}$.[5] Again, the effect of the estimation error in the first step becomes negligible in the limiting distribution of the nonparametric estimator for the nonlinear factor dynamics in the second step. Finally, we consider running a nonlinear (nonparametric) forecasting regression. By combining (3) with (1) and (5), we have

$$y_{t+h} = g_h(f_t^0) + u_{t+h} \tag{6}$$

[4]For example, the validity of the procedure employed by Diebold (2003) can be considered by checking the corresponding moments of the Markov switching model recently derived by Timmermann (2000).

[5]The earlier version of the paper contained the proof of this claim for the case of kernel regression estimator. It is available from the author upon request.

where $g_h(f_t^0) = E(y_{t+h}|\mathcal{F}_t) = \lambda_0^0 m_h(f_t^0)$, $m_h(f_t^0) = E(f_{t+h}^0|\mathcal{F}_t) = E(f_{t+h}^0|f_t^0)$ and $u_{t+h} = \lambda_0^0 \left[f_{t+h}^0 - E(f_{t+h}^0|\mathcal{F}_t) \right] + e_{0t+h}$. As in the linear case, the optimal forecast, $y_{T+h|T} = g_h(f_t^0)$, is not feasible. Therefore, we employ $\hat{y}_{T+h|T} = \hat{g}_h(\hat{f}_T)$ where $\hat{g}_h(\hat{f}_T)$ is a nonparametric regression estimator of y_{t+h} on $\hat{f}_t$ evaluated at $\hat{f}_T$. The first order efficiency of $\hat{y}_{T+h|T}$ can heuristically be shown as follows. By a Taylor series expansion, the dominant term of $\hat{y}_{T+h|T} - y_{T+h|T}$ is given by

$$\nabla \hat{g}_h(f_T^0)(\hat{f}_T - f_T^0) + \left(\hat{g}_h(f_T^0) - g_h(f_T^0) \right) \tag{7}$$

where $\nabla \hat{g}_h(f)$ is the first derivative of $\hat{g}_h(f)$. The boundedness of $\nabla \hat{g}_h(f_T^0)$ and consistency of $\hat{f}_T$ and $\hat{g}_h(f_T^0)$ implies $\hat{y}_{T+h|T} - y_{T+h|T} \xrightarrow{p} 0$.

In the next section, both two-step methods of estimating a dynamic factor structure and running a forecasting regression are applied to Japanese data. In particular, we first employ nonparametric specification tests to choose between (2) and (5). Then we construct optimal forecasts using both a linear forecasting regression (4) and a nonlinear forecasting regression (6) and compare the out-of-sample forecasting performance of DI forecasts with that of conventional time series forecasts.

3 Empirical Results

3.1 Construction of Diffusion Indexes in Japan

Similarly to the NBER in the U.S., the Economic and Social Research Institute (ESRI) of the Cabinet Office (formerly the Economic Planning Agency) is in charge of releasing official diffusion indexes in Japan. Currently, twelve and eleven series are used to construct the leading index and coincident index, respectively. After each business cycle, the ESRI considers replacing the components of indexes, with the latest revision made in January, 2002 (the eighth major revision after the introduction of the official DI). However, since such revisions rely on expert judgment rather than on formal selection criteria, whether a new index would be better than the current one is always open to question (see Kanoh, 1990, for discussions regarding this issue). A DI based on the principal components of a large number of series, as proposed by Stock and Watson (1998), is certainly less subject to this problem since it automatically summarizes all the available information based on a statistical model.

Our factor DI utilizes a balanced panel of 234 monthly series from 1973:2 to 2000:12 (see Appendix for the list of variables). It should be noted that a large number of the series overlap with the candidate series considered by the ESRI in the revision of the official DI.[6] Most variables are expressed in first differences of logs of seasonally adjusted series or seasonal growth rates of unadjusted series to obtain the I(0) stationarity. In addition, all the series are standardized to have sample mean zero and unit sample variance since principal components are not scale-invariant. In a single factor case, $\left\{\widehat{f}_t\right\}_{t=1}^{T}$ is the first eigenvector of the $T \times T$ matrix XX' with normalization $T^{-1}\sum_{t=1}^{T}\widehat{f}_t^2 = 1$, where X is the $T \times N$ data matrix with t-th row given by $X_t' = (x_{1t}, \cdots, x_{Nt})$. For multiple factors, the k-th principal component estimator, $\left\{\widehat{f}_t^{(k)}\right\}_{t=1}^{T}$ for $k \geq 1$ is given by the k-th eigenvector of the same matrix.

Figure 1 plots the factor DI from the first principal component $\widehat{f}_t^{(1)}$ rescaled to have the same drift and variance as the (log of) industrial production in mining and manufacturing (hereafter referred to as IP). In the same figure, we also plot the IP series as well as the official ESRI recessionary episodes shown as the shaded area. On the whole, the factor DI and IP move together, and thus it is consistent with the U.S. finding by Stock and Watson (1998, 2002) that first factor loads primary on the series related the real output. However, there are some notable differences between the two series. First, the decline during

[6] The report by the Cabinet Office of Japan (1997) contains a list of 253 candidate series used in the seventh revision of the Japanese official business cycle index. Candidate variables employed for the eighth revision are not published but are similar to those used in the previous revision.

the recession of 1973-1975 is much larger in the factor DI than in the IP. Second, in contrast to the IP, no clear trough is observed in the factor DI series during the recession of 1985-1986. This second point has an interesting implication if we estimate the turning points using a Markov switching factor model and compare it with the ESRI reference cycle. Figure 2 shows the recession probabilities computed by fitting a Markov switching model with AR(2) dynamics to $\hat{f}_t^{(1)}$ following the two-step procedure of Diebold (2003).[7] While the extracted recession probability does not differ much from the ESRI recessionary episodes, the probability of recession in the official recession of 1985-1986 is very low. The probability is indeed lower than in 1995 despite the fact that, according to the ESRI business cycle chronology, there was no recession in 1995.

3.2 Testing for Linear Factor Dynamics

Since the neural network can be interpreted as a method of approximating nonlinear function, it can be used to estimate the nonlinear model when the functional form is not specified. The nonparametric estimator based on single hidden layer feedforward ANNs can be obtained by minimizing the least square criterion $\sum_{t=1}^{T} [Y_t - m(Z_t)]^2$ where Y_t is a single output, X_t is a vector of input, $m(Z_t)$ is the neural network approximation function given by

$$m(Z_t) = \alpha' Z_t + \sum_{j=1}^{q} \beta_j \psi\left(\gamma_j' Z_t\right) \tag{8}$$

where ψ is an activation function, q is the number of hidden units.[8] For the AR(1) case of (5), the neural network estimator $\hat{m}(f)$ is obtained by setting output $y_t = \hat{f}_t$ and input $Z_t = \hat{f}_{t-1}$ and by minimizing the criterion with respect to α, β_j's and γ_j's. Figure 3 shows the linear model of the rescaled factors

[7] In case of a small number of series ($N = 4$), Watanabe (2001) also investigated the performance of a Markov switching factor model in Japan estimated by the method proposed by Kim and Nelson (1998).

[8] Throughout this paper, we use the logistic activation function. Also, the criterion function is modified to have the weight decay identical to the one employed in Franses and van Dijk (2000).

estimated by OLS and Figure 4 shows the nonlinear model estimated by ANNs, under AR(2) specifications. While comparison of the two figures seems to suggest the presence of nonlinearity, we would like to know whether the difference is statistically significant. For this purpose, we conduct nonparametric specification tests for the null hypothesis of linear specification of (2) that are consistent against a wide range of nonlinear alternatives given by (5).[9] Since the null hypothesis can be written as a conditional moment restriction $E\left[\varepsilon_t | f'_{t-1}\right] = 0$ with $\varepsilon_t = f'_t - \phi f'_{t-1}$, it implies the unconditional moment restriction of the form $E\left[h(f'_{t-1})\varepsilon_t\right] = 0$ with any vector of measurable functions $h(f'_{t-1})$. Therefore, a number of tests for linearity (or neglected nonlinearity) can be constructed with a different choice of $h(f'_{t-1})$.

Ramsey's (1969) regression specification error test (RESET), which is one of the most well-known tests in the specification testing literature, uses a $r \times 1$ vector of polynomial functions of fitted value from linear regression, $h(f'_{t-1}) = \left((\hat{\phi} f'_{t-1})^2, \ldots, (\hat{\phi} f'_{t-1})^r\right)'$. The test statistic $RESET = T\left(\sum_{t=1}^{T} \hat{\varepsilon}_t^2 - \sum_{t=1}^{T} \hat{u}_t^2\right) / \sum_{t=1}^{T} \hat{\varepsilon}_t^2$, where $\hat{\varepsilon}_t = f'_t - \hat{\phi} f'_{t-1}$ and $\hat{u}_t$ are the residuals from the regression of $\hat{\varepsilon}_t$ on auxiliary regressors $h(f'_{t-1})$ (and f'_{t-1}), asymptotically follows χ^2 distribution with r degree of freedom.

For White's (1989) neural network test, $h(f'_{t-1}) = \left(\psi\left(\gamma_1 f'_{t-1}\right), \ldots, \psi\left(\gamma_q f'_{t-1}\right)\right)' = \Psi_t$ is a $q \times 1$ vector of logistic activation functions ψ with the coefficients γ_j's being randomly drawn independent of f'_{t-1}. The test statistic can be similarly constructed by using auxiliary regressors (NN), or by using quadratic form (NN-HAC), $T\overline{m}'\hat{\Omega}_m\overline{m}$ where $\overline{m} = T^{-1}\sum_{t=1}^{T} \Psi_t \hat{u}_t$ and $\hat{\Omega}_m$ is the heteroskedasticity and autocorrelation consistent (HAC) covariance estimator of $\overline{m}$. In either case, the limit distribution of the test statistic is χ^2 distribution with q degree of freedom.

One drawback of the White's neural network test is the unidentifiability of γ_j's under the null hypothesis. Instead of using random γ_j's, Teräsvirta, Lin, and Granger (1993) replaced the activation functions by

[9] Instead of using specification tests, Hess and Iwata (1997) evaluated the performance of nonlinear models by checking to see if they could replicate business cycle features. However, we do not use their approach since true factors are latent variables and thus we cannot define cycles unlike the one based on observed GDP series.

their Volterra expansion up to the third order under the null. This LM type neural network test (NN-LM) can be constructed by using auxiliary regressors based on quadratic and cubic terms from Volterra expansion of nonlinear AR model ($h(f_{t-1}^0) = \left((f_{t-1}^0)^2, (f_{t-1}^0)^3\right)'$ for AR(1) case). The test statistic asymptotically follows χ^2 distribution with $p(p+1)/2 + p(p+1)(p+2)/6$ degree of freedom where p is lag order of AR model.

The last test we consider is the kernel-based consistent specification test for AR models proposed by Fan and Li (1997). It utilizes the $h(f_{t-1}^0) = E\left(\varepsilon_t | f_{t-1}^0\right) f\left(f_{t-1}^0\right)$ where $f\left(f_{t-1}^0\right)$ is a density function of f_{t-1}^0. The test statistic (KERNEL) is based on the kernel estimator of $E\left[h(f_{t-1}^0)\varepsilon_t\right] = E\left[E\left(\varepsilon_t | f_{t-1}^0\right) f\left(f_{t-1}^0\right) \varepsilon_t\right]$ and follows asymptotically normal with an appropriate standardization.

We apply five different asymptotic tests to $\hat{f}_t$ estimated by principal components method since f_t^0 is not available. Following the discussion in section 2, we expect that estimation error has a negligible effect on the limiting distribution of the test statistics for linearity under certain regularity conditions. Table 1 reports the results of all five tests applied to each of the first to sixth diffusion indexes ($k = 1, ..., 6$) with autoregressive orders ranging from one to four ($p = 1, ..., 4$). For RESET, the results based on $r = 4$ are reported. For NN and NN-HAC, we use three (excluding the first) principal components of Ψ_t with $q = 10$ to avoid collinearity of f_{t-1}^0 and Ψ_t. Then, the improved Bonferroni procedure from five draws is used to construct p-values (see Lee, White, and Granger, 1993, for this procedure in detail). The p-values less than 0.10 are indicated by bold font.

It is fair to say that the results are rather mixed. The RESET, NN and NN-LM tests reject the linear hypothesis of factor-diffusion indexes for many cases at the conventional significance level. In contrast, based on the NN-HAC and KERNEL tests, the same hypothesis is not rejected for almost all cases. One possibility of this mixed outcome may be related to the power of the specification tests. Among all the tests we considered, NN-LM provides the strongest evidence against linearity. Based on a simulation experiment, Teräsvirta, Lin, and Granger (1993) argue that NN-LM is more powerful than the standard neural network tests with random draw of hidden layer parameters. In addition, Lee's (2001) simulation study compares the performance of NN and KERNEL and reports that KERNEL is less powerful than NN unless bootstrapped critical value is used. While these simulation results do not take the effect of estimation error of the common factor in the first step, we can still conclude that there are some possibilities of nonlinearity in the factor dynamic structure.

3.3 Linear and Nonlinear Diffusion Index Forecasts

In this subsection, we evaluate the out-of-sample forecasting performance of the linear and nonlinear DI forecasts in Japan. We first consider the following h-period ahead linear forecasting regression, a generalization of (4) to allow for multiple factors as well as lags of factors and y_t,

$$y_{t+h} = \alpha_h + \beta_h'(L)\hat{F}_t + \gamma_h(L)y_t + u_{t+h} \tag{9}$$

where $\hat{F}_t = (\hat{f}_t^{(1)}, \ldots, \hat{f}_t^{(K)})'$ is the $K \times 1$ vector of K estimated factors in the first step, and $\beta_h(L)$ and $\gamma_h(L)$ are the lag polynomials of finite order $s-1$ and $p-1$, respectively. As a forecasting variable y_t, we consider five measures of aggregate activity currently used as ESRI coincident indicators: the index of industrial production (IP); the index of producer's shipments (SHIP); the index of the capacity utilization ratio (CAP); the index of sales in small and medium-sized enterprises (SALE); and the index of non-scheduled worked hours (HOUR). In addition, while it is not an ESRI coincident indicator, the inflation rate based on the consumer price index (CPI) is also included as a forecasting variable.[10] Based on the assumption of I(1) in logarithm, the IP (similarly for the other series) is transformed as follows

[10] Effects of the introduction of the consumption tax in April, 1989, and the increased tax rate in April, 1997, on the CPI have been adjusted using the X12-ARIMA program. We employ the I(1) specification of the price index for Japan rather than the I(2) specification which has been used for the U.S. by Stock and Watson (2002).

$$y_{t+h} = (1200/h)\ln(IP_{t+h}/IP_t) \quad \text{and} \quad y_t = 1200\ln(IP_t/IP_{t-1}). \tag{10}$$

Following Stock and Watson (2002), we evaluate the performance of (9) based on a simulated out-of-sample forecasting methodology using the recursive scheme. First, the sample is divided into first R observations and last $P+h-1$ observations, and the factor is estimated by the principal components method using normalized x_{it}'s from period 1 to R. The estimated factor is then used in the forecasting regression to obtain the forecast of y_{R+h}. For the second forecast y_{R+h+1}, the data is again standardized and the factors and forecasting models are reestimated using the observations from 1 to $R+1$. This procedure is repeated P times to obtain P simulated out-of-sample forecasts. We compare this DI forecast with two alternative linear forecasts; the autoregressive (AR) forecast and the leading indicator (LI) forecast. The AR forecast uses only current and lagged y_t and excludes $\hat{F}_t$ from the forecasting regression (9). The LI forecast replaces $\hat{F}_t$ in (9) with a vector W_t where elements of W_t are the official leading indicators provided by the ESRI. We use the following ten leading indicators: the index of the producer's inventory ratio of finished goods (final demand goods) (L1); the index of the raw materials inventory to consumption ratio (manufacturing) (L2); new job offers (excluding new school graduates) (L3); new orders for machinery at constant prices (except for volatile orders) (L4); the total floor area of building construction started (L5); the total floor area of new housing construction started (L6); the number of new passenger car registrations and reports (L7); the Nikkei commodity price index (17 items) (L8); the money supply (M2+CD) (L9); and the index of investment climate (manufacturing) (L10).

We consider, as measures of forecasting performance, the mean squared forecast error (MSFE) defined by $P^{-1}\sum_{t=R}^{T-h}\hat{e}_{t+h}^2$ where $\hat{e}_{t+h}$ is the h-period ahead forecast error, and the mean absolute forecast error (MAFE) defined by $P^{-1}\sum_{t=R}^{T-h}|\hat{e}_{t+h}|$, and report the ratio of each criterion of the candidate model to that of the benchmark model. In addition, we use Diebold and Mariano's (DM, 1995) test of equal forecast accuracy based on the normalized loss differential given by

$$DM = \sqrt{P}\frac{\overline{d}}{\sqrt{\widehat{\omega}_d}} \tag{11}$$

where $\overline{d} = P^{-1}\sum_{t=R}^{T-h}\left(\widehat{u}_{1,t+h}^2 - \widehat{u}_{2,t+h}^2\right)$ for MSFE, $\overline{d} = P^{-1}\sum_{t=R}^{T-h}(|\widehat{u}_{1,t+h}| - |\widehat{u}_{2,t+h}|)$ for MAFE, $\widehat{u}_{1,t+h}$ and $\widehat{u}_{2,t+h}$ are h-period ahead forecast errors of two models, and $\widehat{\omega}_d$ is the HAC variance estimator of the loss differential $\overline{d}$ with lag truncation parameter $h-1$. Suppose the case of observable f_t^0 in DI forecasting regression (4) with $\widehat{\beta}_h$ being the OLS estimator of β_h. As West (1996) has shown, the effect of parameter estimation error $\widehat{\beta}_h - \beta_h$ needs to be incorporated into the limiting variance of $\overline{d}$, unless $P/R \to 0$ as $T \to \infty$. Similarly, since the true DI forecasting regression involves a latent variable f_t^0, additional uncertainty from the factor estimation error $\widehat{f}_t - f_t^0$ needs to be incorporated in general. From the argument used in Section 2, however, we expect that both factor estimation error and parameter estimation error will become negligible in the limiting distribution given $N/T \to \infty$ and $P/R \to 0$ as $T \to \infty$.

The simulated out-of-sample forecast periods are 1991:1 to 2000:12 so that the number of forecasts (P) is 120. Table 2 shows the results of various linear forecasts for five real series and one inflation series with the 6-month forecast horizon ($h = 6$). The AR lag (p) for all the models is fixed to two, and only current LI is included in the LI forecast ($s = 1$). For the DI forecast, six factor diffusion indexes are included ($K = 6$) and two cases with $s = 1$ (DI1) and $s = 2$ (DI2) are considered.[11] The DM test statistics are based on HAC estimator with the Bartlett kernel and associated p-values are based on the standard normal distribution. The p-values less than 0.10 are indicated by bold font. The results of the linear forecasts can be summarized as follows. First, the LI forecast does not provide a better forecast as compared to the AR forecast. This result is consistent with Fukuda and Onodera's (2001) finding that LI failed to forecast

[11] This paper uses Hong and Lee's (2002) approach where a forecasting model with fixed specification is reestimated to construct each forecast. In contrast, Stock and Watson (2002) allow a different choice of the lag lengths and the number of factors to construct each forecast. We do not use the latter approach since it will complicate the issue of nested models which will be discussed later in the section.

the composite indicators during the 1990s in Japan. Second, both DI1 and DI2 outperform the AR and LI forecasts for all cases except for HOUR. Among alternative DI forecasts, significant improvement is observed when lags of the DIs are included (DI2). The ratio implies that a 10 to 30 percent reduction in MSFE is obtained by DI2 compared to the AR and LI forecasts and the reduction is significant in many cases. Third, all the results seem to be robust among two choices of criteria, MSFE and MAFE. This outcome is very encouraging and suggests the usefulness of the (linear) DI forecast in Japan.[12]

We now turn to the nonlinear DI forecast. In Section 3.2, we found some evidence suggesting the possibility of nonlinearity in factor dynamics. As discussed in Section 2, this possibility implies that there may be some gain from employing a nonlinear forecasting regression. As in the case of the linear DI forecast, we consider a generalization of nonlinear DI forecast (6) to allow for multiple factors as well as lags of factors and y_t,

$$y_{t+h} = g_h(\hat{F}_t', ..., \hat{F}_{t-(s-1)}', y_t, ..., y_{t-(p-1)}) + u_{t+h}. \tag{12}$$

For the purpose of estimating g_h function, we again employ the ANNs given by (8). Here, the estimator $\hat{g}_h$ the for a nonlinear DI forecast is obtained with the output $Y_t = y_{t+h}$ and the input vector $Z_t = (\hat{F}_t', ..., \hat{F}_{t-(s-1)}', y_t, ..., y_{t-(p-1)})$. While the lag lengths ($s = 1, 2$ and $p = 2$) and the number of factors ($K = 6$) are fixed as in the case of the linear forecasts, the number of the hidden unit q is selected by minimizing BIC. The MSFE and MAFE of the forecasts from the nonlinear models with $s = 1$ (NN1) and $s = 2$ (NN2) are compared to those of the corresponding linear DI forecasts (DI1 and DI2). In addition to the result based on the single nonlinear DI forecast, we also provide the result based on the forecast combination of the linear and nonlinear DI models using the weight employed in Hong and Lee (2002). Table 3 shows the performance of the nonlinear DI forecasts compared to that of the linear DI forecasts.

[12] We also conducted forecasting with a longer horizon, $h = 12$, as well as a shorter horizon, $h = 1$. We obtained similar results with $h = 12$ case but no evidence of improvement with DI forecast was found in the case of $h = 1$.

COMB1 (COMB2) forecasts are the combination of DI1 (DI2) and NN1 (NN2). Unfortunately, for most cases, we do not find strong evidence suggesting the advantage of nonlinear forecasts over linear forecasts. The only exception is HOUR, the case in which the linear DI performs poorly in Table 2. For both MSFE and MAFE, nonlinear DI forecasts outperform linear DI forecasts, and the loss differentials are significant in the case of a forecast combination (COMB1 and COMB2). Although not reported in the table, the nonlinear DI forecasts also outperform the linear AR and LI forecasts. From this observation, we conjecture that there are some other variables with which the nonlinear DI forecast works even if the linear version fails.

Finally, we would like to discuss the issue of the nested structure of competing models. In the linear case, the DI and LI models are nonnested, but the DI and AR models are nested models. In addition, the nonlinear and linear DI models are also nested. As emphasized in Clark and McCracken (2001), the DM test of equal MSFE of two nested models may have a non-standard limiting distribution unless $P/R \to 0$ as $T \to \infty$. Their table shows that the critical values based on non-standard distribution are smaller than the standard normal critical values. Therefore, when P/R is not very small, a test based on the standard normal critical value may better be considered as a conservative test. In that case, the implications to our results are as follows. First, when the p-values of the DM test for nested models obtained in Tables 2 and 3 are small, the conclusion is still valid. Second, even if the p-values are large, such as the one for forecasting SALE using a DI model in Table 2, there are some possibilities that the loss differential is indeed significant if an appropriate test is employed. Because of this second implication, it may be worth examining the forecasting performance by using an additional test designed for the nested case. For this reason, we also compute Chao, Corradi and Swanson's (2001) test statistics for the null hypothesis of equal predictive ability of the DI and AR models,

$$CCS = P\overline{m}'\widehat{\Omega}_m\overline{m} \tag{13}$$

where $\overline{m} = P^{-1}\sum_{t=R}^{T-h}\widehat{v}_{t+h}\widehat{F}_t$, $\widehat{v}_{t+h}$ is the h-period ahead forecast error of the AR model, and $\widehat{\Omega}_m$ is the HAC covariance estimator of $\overline{m}$ with lag truncation parameter $h-1$. Under the null hypothesis of equal MSFE, the test follows χ^2 distribution with K degree of freedom. For the test of comparing the MSFE of the linear and nonlinear DI forecasts in Table 3, $\widehat{v}_{t+h}$ is the forecast error of the linear DI model and $\widehat{F}_t$ is replaced by $\Psi_t = \left(\psi\left(\gamma_1' Z_t\right), \ldots, \psi\left(\gamma_q' Z_t\right)\right)'$ where $Z_t = (\widehat{F}_t', ..., \widehat{F}_{t-(s-1)}', y_t, ..., y_{t-(p-1)})$. Implementation of the nested nonlinear prediction test is similar to that employed for the neural network test for neglected nonlinearity described in Section 3.2. We report the improved Bonferroni p-values from five draws of the test statistic based on three principal components of Ψ_t with $q = 10$ and randomly drawn γ_j's. The first column of Table 4 shows the results of nested tests for the linear case and the second column shows those of nested tests for the nonlinear case. On the whole, the results are consistent with those in Tables 2 and 3 in the sense that the evidence supports the linear DI forecast over the linear AR forecast for almost all cases, but relatively weak evidence is found regarding the advantage of the nonlinear DI forecast over the linear DI forecast. However, most p-values are found to be less than those based on a nonnested assumption. As a result, evidence of the usefulness of the DI forecast becomes stronger, including the linear DI forecast for HOUR and the nonlinear DI forecast for CPI.

4 Conclusion

This paper has considered the possibility of extending the diffusion index (DI) forecast approach proposed by Stock and Watson (1998, 2002) to the case of dynamic factor models with a possibly nonlinear dynamic factor structure. When the number of series is large, a two-step procedure based on principal components method is useful and convenient as it is robust to the wide variety of the nonlinear structures of latent factors. The DIs constructed from principal components, thus, can be used to estimate the nonlinear time series models or to conduct specification tests regarding the nonlinearity of the model. Furthermore, the DI can be included as a regressor in the nonlinear forecasting regression.

As an empirical application of this procedure, we constructed factor DIs based on 234 monthly macroeconomic series from Japan. We estimated the nonlinear time series model of DIs nonparametrically using artificial neural networks (ANNs). The results of nonparametric specification tests provided some evidence of a nonlinear dynamic factor structure. We then applied both linear and nonlinear DI forecasting regression to several measures of aggregate activity currently used as coincident indicators in Japan, as well as the CPI-based inflation series. As with Stock and Watson's (2002) finding with the U.S. data, the DI forecast approach is found to be useful in forecasting the Japanese economy. Both linear and nonlinear DI forecasts outperformed the conventional time series forecast, while only limited evidence was obtained regarding the advantage of the nonlinear DI forecast over the linear DI forecast.

In closing, we raise some issues to extend the analysis of this paper. First, this paper's approach relies on the large N asymptotics. In the linear case, Shintani's (2003) simulation results on the AR estimation of the factors show that asymptotic approximation works well with a sample size typically available for economic time series. Similar simulation design may be used to check the finite sample performance in the nonlinear case. Second, using the parametric nonlinear models as well as other nonparametric methods may provide different results from this paper that uses ANNs. Third, the performance of the nonlinear DI forecast may be improved by allowing time-varying specification regarding the lag length and the number of factors included which was not considered in this paper.

Appendix: Data Description

This appendix lists the series used to construct the diffusion index based on the factor model described in the main text. Sample period is from February 1973 to December 2000. All the data are transformed using the first difference of logs of seasonally adjusted series (or seasonal growth rate) except for the interest rates (194 to 206) where the series are based on the levels.

Series Number	
	Real Output
1 (IP)	Index of Industrial Production (Mining and Manufacturing)
2	Index of Industrial Production (Manufacturing)
3	Index of Industrial Production (Mining)
4	Index of Industrial Production (Iron and Steel)
5	Index of Industrial Production (Non-Ferrous Metals)
6	Index of Industrial Production (Fabricated Metals)
7	Index of Industrial Production (General Machinery)
8	Index of Industrial Production (Electrical Machinery)
9	Index of Industrial Production (Transport Equipment)
10	Index of Industrial Production (Precision Instruments)
11	Index of Industrial Production (Ceramics, Clay and Stone Products)
12	Index of Industrial Production (Chemicals)
13	Index of Industrial Production (Petroleum and Coal Products)
14	Index of Industrial Production (Plastic Products)
15	Index of Industrial Production (Pulp, Paper and Paper Products)
16	Index of Industrial Production (Textiles)
17	Index of Industrial Production (Foods and Tobacco)
18	Index of Industrial Production (Other Manufacturing)
19	Index of Industrial Production (Final Demand Goods)
20	Index of Industrial Production (Producer Goods)
21	Index of Industrial Production (Producer Goods for Mining and Manufacturing)
22	Index of Industrial Production (Producer Goods for Others)
23	Index of Producer's Shipments (Final Demand Goods)
24 (SHIP)	Index of Producer's Shipments (Producer Goods)
25	Index of Producer's Shipments (Producer Goods for Mining and Manufacturing)
26	Index of Producer's Shipments (Producer Goods for Others)
27	Index of Raw Materials Consumption (Manufacturing)
28	Large Consumption of Electric Energy (Total)
29 (CAP)	Index of Capacity Utilization Ratio (Manufacturing)
30	Index of Capacity Utilization Ratio (Iron and Steel)
31	Index of Capacity Utilization Ratio (Non-Ferrous Metals)
32	Index of Capacity Utilization Ratio (Fabricated Metals)
33	Index of Capacity Utilization Ratio (General Machinery)
34	Index of Capacity Utilization Ratio (Electrical Machinery)
35	Index of Capacity Utilization Ratio (Transport Equipment)
36	Index of Capacity Utilization Ratio (Precision Instruments)
37	Index of Capacity Utilization Ratio (Ceramics, Clay and Stone Products)
38	Index of Capacity Utilization Ratio (Chemicals)
39	Index of Capacity Utilization Ratio (Petroleum and Coal Products)
40	Index of Capacity Utilization Ratio (Textiles)
41	Index of Capacity Utilization Ratio (Rubber Products)
42	Index of Capacity Utilization Ratio (Machinery)
43(SALE)	Index of Sales in Small and Medium-Sized Enterprises (Manufacturing)
44	Index of Tertiary Industry Activity (Total)
45	Index of Tertiary Industry Activity (Electricity, Gas, Heat and Water Supply)
46	Index of Tertiary Industry Activity (Transport and Communication)
47	Index of Tertiary Industry Activity (Transport)
48	Index of Tertiary Industry Activity (Wholesale, Retail Trade, Eating and Drinking Places)
49	Index of Tertiary Industry Activity (Eating and Drinking Places)
50	Index of Tertiary Industry Activity (Finance and Insurance)
51	Index of Tertiary Industry Activity (Real Estate)
52	Index of Tertiary Industry Activity (Services)
53	Index of Tertiary Industry Activity (Personal Services)
54	Index of Tertiary Industry Activity (Business Services)

Inventories

55	Index of Producer's Inventory Ratio of Finished Goods (Mining and Manufacturing)
56(L1)	Index of Producer's Inventory Ratio of Finished Goods (Final Demand Goods)
57	Index of Producer's Inventory Ratio of Finished Goods (Investment Goods)
58	Index of Producer's Inventory Ratio of Finished Goods (Capital Goods)
59	Index of Producer's Inventory Ratio of Finished Goods (Construction Goods)
60	Index of Producer's Inventory Ratio of Finished Goods (Consumer Goods)
61	Index of Producer's Inventory Ratio of Finished Goods (Durable Consumer Goods)
62	Index of Producer's Inventory Ratio of Finished Goods (Nondurable Consumer Goods)
63	Index of Producer's Inventory Ratio of Finished Goods (Producer Goods)
64	Index of Producer's Inventory Ratio of Finished Goods (Producer Goods for Mining and Manufacturing)
65	Index of Producer's Inventory Ratio of Finished Goods (Producer Goods for Others)
66(L2)	Index of Raw Materials Inventory Ratio (Manufacturing)
67	Index of Producer's Inventory of Finished Goods (Mining and Manufacturing)
68	Index of Producer's Inventory of Finished Goods (Final Demand Goods)
69	Index of Producer's Inventory of Finished Goods (Investment Goods)
70	Index of Producer's Inventory of Finished Goods (Capital Goods)
71	Index of Producer's Inventory of Finished Goods (Construction Goods)
72	Index of Producer's Inventory of Finished Goods (Consumer Goods)
73	Index of Producer's Inventory of Finished Goods (Durable Consumer Goods)
74	Index of Producer's Inventory of Finished Goods (Nondurable Consumer Goods)
75	Index of Producer's Inventory of Finished Goods (Producer Goods)
76	Index of Producer's Inventory of Finished Goods (Producer Goods for Mining and Manufacturing)
77	Index of Producer's Inventory of Finished Goods (Producer Goods for Others)
78	Index of Inventory (Final Demand Goods)

Investments

79	Index of Producer's Shipments (Investment Goods Excluding Transport Equipments)
80	Index of Producer's Shipments (Producer Goods)
81	Index of Industrial Production (Investment Goods)
82	Index of Industrial Production (Capital Goods)
83	Index of Industrial Production (Construction Goods)
84	Index of Production Capacity (Manufacturing)
85	Machinery Orders (Total, Excluding Ships)
86(L4)	Machinery Orders (Private Sector, Excluding Volatile Orders)
87	Machinery Orders (Manufacturing)
88	Machinery Orders (Non-Manufacturing, Excluding Volatile Orders)
89	Machinery Orders (Government)
90	Order Received for Construction (Grand Total)
91	Order Received for Construction (Private)
92	Order Received for Construction (Manufacturing)
93	Order Received for Construction (Non-Manufacturing)
94	Order Received for Construction (Public)
95	Total Floor Area of Building Construction Started (Grand Total)
96(L5)	Total Floor Area of Building Construction Started (Mining, Manufacturing and Commercial Use)
97	Total Floor Area of Building Construction Started (Mining)
98	Total Number of New Housing Construction Started (Total)
99	Total Number of New Housing Construction Started (Owned)
100	Total Number of New Housing Construction Started (Rented)
101	Total Number of New Housing Construction Started (Built for Sale)
102	Total Number of New Housing Construction Started (Government Housing Loan Corporation)
103(L6)	Total Floor Area of New Housing Construction Started (Total)
104	Total Floor Area of New Housing Construction Started (Owned)
105	Total Floor Area of New Housing Construction Started (Rented)
106	Total Floor Area of New Housing Construction Started (Built for Sale)

Employment

107	Index of Non-Scheduled Worked Hours (All Industries, 30 or More Persons)
108(HOUR)	Index of Non-Scheduled Worked Hours (Manufacturing)
109	Index of Total Worked Hours (All Industries, 30 or More Persons)
110	Index of Total Worked Hours (Manufacturing)
111	Ratio of Non-Scheduled to Total Worked Hours (All Industries, 30 or More Persons)
112	Ratio of Non-Scheduled to Total Worked Hours (Manufacturing)
113	New Job Offers
114	Effective Job Offers
115(L3)	New Job Offer Rate

116 Effective Job Offer Rate
117 New Job Offers (Part-Time)
118 Effective Job Offers (Part-Time)
119 New Job Offer Rate (Part-Time)
120 Effective Job Offer Rate (Part-Time)
121 Index of Regular Workers Employment (All Industries, 30 or More Persons)
122 Index of Regular Workers Employment (All Industries, Excluding Services)
123 Index of Regular Workers Employment (Mining)
124 Index of Regular Workers Employment (Construction)
125 Index of Regular Workers Employment (Manufacturing)
126 Index of Regular Workers Employment (Electricity, Gas, Heat Supply)
127 Index of Regular Workers Employment (Transport and Communication)
128 Index of Regular Workers Employment (Wholesale and Retail Trade)
129 Index of Regular Workers Employment (Finance and Insurance)
130 Index of Regular Workers Employment (Real Estate)
131 Index of Regular Workers Employment (Services)
132 Number of Unemployment
133 Unemployment Rate
134 Number of Beneficiaries of Unemployment Insurance (Initial Claimants)
135 Number of Beneficiaries of Unemployment Insurance (Total)
136 Number of Persons with Unemployment Insurance
137 Real Wage Index (Contractual Cash Earnings in All Industries, 30 or More Persons)

Consumption

138 Sales at Department Stores (Total)
139 Sales at Department Stores (Per Square Meter Floor Space)
140 Index of Sales (Total)
141 Index of Sales (Wholesale)
142 Index of Sales (General Merchandise Retail)
143(L7) Number of New Passenger Car Registrations and Reports (Total)
144 Number of New Passenger Car Registrations and Reports (Excluding Cars Under 550cc)
145 Household Consumption Expenditure (Workers)
146 Household Consumption Expenditure (Food)
147 Household Disposable Income (Workers)
148 Index of Industrial Production (Consumer Goods)
149 Index of Industrial Production (Durable Consumer Goods)
150 Index of Industrial Production (Non-Durable Consumer Goods)
151 Index of Producer's Shipments (Consumer Goods)
152 Index of Producer's Shipments (Durable Consumer Goods)
153 Index of Producer's Shipments (Non-Durable Consumer Goods)

Firms

154(L10) Index of Investment Climate (Manufacturing)
155 Corporation Tax Revenue
156 Suspension of Business Transaction with Bank

Money, Stock Price and Interest Rate

157(L9) Money Supply (M2+CD, Average Outstanding)
158 Money Supply (M1, Average Outstanding)
159 Monetary Base (Average Outstanding)
160 Bank Notes Issued (Average Outstanding)
161 Bank Clearings (Number)
162 Bank Clearings (Value)
163 Nikkei Stock Average 225 Selected Stocks (Average of Month)
164 Nikkei Stock Average 500 Selected Stocks
165 Stock Price Index (TOPIX)
166 Stock Price Average (Tokyo Stock Market, First Section)
167 Stock Price Index (Fisheries, Agriculture and Forestry)
168 Stock Price Index (Mining)
169 Stock Price Index (Construction)
170 Stock Price Index (Foods)
171 Stock Price Index (Textiles)
172 Stock Price Index (Pulp and Paper)
173 Stock Price Index (Oil and Coal Products)
174 Stock Price Index (Rubber Products)
175 Stock Price Index (Glass and Ceramics Product)

176 Stock Price Index (Iron and Steel)
177 Stock Price Index (Non-Ferrous Metals)
178 Stock Price Index (Metal Products)
179 Stock Price Index (Machinery)
180 Stock Price Index (Electrical Machinery)
181 Stock Price Index (Transportation Equipment)
182 Stock Price Index (Precision Instrument)
183 Stock Price Index (Other Products)
184 Stock Price Index (Electric and Gas)
185 Stock Price Index (Land Transportation)
186 Stock Price Index (Marine Transportation)
187 Stock Price Index (Air Transportation)
188 Stock Price Index (Warehouse and Transport-Related)
189 Stock Price Index (Communication)
190 Stock Price Index (Real Estate)
191 Stock Price Index (Service)
192 Sales Volume (Daily Average, Tokyo Stock Market, First Section)
193 Sales Value (Daily Average, Tokyo Stock Market, First Section)
194 Official Discount Rates
195 Short-Term Prime Lending Rates
196 Long-Term Prime Lending Rates
197 Average Contracted Interest Rate on Loans and Discounts (Domestically Licensed Bank)
198 Yields of Bond Traded with Repurchase Agreement (3 Months, Month Average)
199 Call Rates (Collateralized Overnight, Month Average)
200 Bill Rates (2 Months, Month Average)
201 Yields of Short-Term Government Securities (13 Weeks)
202 Yields of Interest-Bearing Bank Debentures (5 Years)
203 Yields of Interest-Bearing Government Bonds (10 Years)
204 Yields of Government Guaranteed Bonds (10 Years)
205 Yields of Local Government Bonds (10 Years)
206 Yields to Maturity of Listed Government Bonds (Longest Term until Redemption Day)

Price Indexes

207(L8) Nikkei Commodity Price Index (17 items)
208 Nikkei Commodity Price Index (42 items)
209 Wholesale Price Index (All Commodities)
210 Wholesale Price Index (Manufacturing Industry Products)
211 Wholesale Price Index (Raw Materials)
212 Wholesale Price Index (Intermediate Materials)
213 Wholesale Price Index (Final Goods)
214 Wholesale Price Index (Capital Goods)
215 Wholesale Price Index (Consumer Goods)
216 Wholesale Price Index (Durable Consumer Goods)
217 Wholesale Price Index (Nondurable Consumer goods)
218 Consumer Price Index (General)
219 (CPI) Consumer Price Index (General, Excluding Fresh Food)
220 Consumer Price Index (General, Excluding Fresh Food and Imputed Rent)
221 Consumer Price Index (Food)
222 Consumer Price Index (Housing)
223 Consumer Price Index (Fuel Light and Water Charges)
224 Consumer Price Index (Furniture and Household Utensils)
225 Consumer Price Index (Clothes and Footwear)
226 Consumer Price Index (Medical Care)
227 Consumer Price Index (Transportation and Communication)
228 Consumer Price Index (Reading and Recreation)
229 Consumer Price Index (Miscellaneous)

Trade

230 Terms of Trade Index (All Commodities)
231 Quantum Index of Exports (Total)
232 Quantum Index of Imports (Total)
233 Customs Clearance (Value of Exports, Grand Total)
234 Foreign Exchange Rate (Yen per US Dollar, Spot)

References

Andrews, Donald W.K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59, 817-858.

Bai, Jushan. (2003). "Inferential Theory for Factor Models of Large Dimensions." *Econometrica* 71, 135-171.

Bai, Jushan, and Serena Ng. (2002). "Determining the Number of Factors In Approximate Factor Models." *Econometrica* 70, 191-221.

Cabinet Office of Japan. (1997). "30 Keiretsu Ni Yoru Keiki Doukou Shisuu (Business Cycle Index Based on 30 Series)." Economic Social Research Institute (formerly Economic Planning Agency), in Japanese.

Chao, John, Valentina Corradi, and Norman R. Swanson. (2001). "An Out of Sample Test for Granger Causality," *Macroeconomic Dynamics* 5, 598-620.

Chen, Xiaohong, Jeffrey Racine, and Norman R. Swanson. (2001). "Semiparametric ARX Neural-Network Models with an Application to Forecasting Inflation." *IEEE Transactions on Neural Networks* 12, 674-683.

Clark, Todd E., and Michael W. McCracken. (2001). "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics* 105, 85-110.

Diebold, Francis X. (2003)."'Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, edited by M. Dewatripont, L.P. Hansen, and S. Turnovsky, pp. 115-122. Cambridge: Cambridge University Press.

Diebold, Francis X., and Roberto S. Mariano. (1995). "Comparing Predictive Accuracy," *Journal of Business Economics and Statistics* 13, 253-263.

Fan, Yanqin, and Qi Li. (1997). "A Consistent Nonparametric Test for Linearity of AR(p) Models." *Economics Letters* 55, 53-59.

Forni, Mario, Marc Hallin, Marco Lippi and Lucrezia Reichlin. (2000). "The Generalized Dynamic-Factor Model: Identification and Estimation." *Review of Economics and Statistics* 82, 540-554.

Franses, Phillips Hans, and Dick van Dijk. (2000). *Nonlinear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press.

Fukuda, Shin-ichi, and Takashi Onodera. (2001). "A New Composite Index of Coincident Economic Indicators In Japan: How Can We Improve the Forecast Performance?" *International Journal of Forecasting* 17, 483-498.

Geweke, John. (1977). "The Dynamic Factor Analysis of Economic Time-Series Models." In *Latent Variable in Socioeconomic Models*, edited by D. J. Aigner and A. S. Goldberger, pp. 365-387. Amsterdam: North-Holland.

Hamilton, James D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." Econometrica 57, 357-384.

Hess, Gregory D., and Shigeru Iwata. (1997). "Measuring and Comparing Business-Cycle Features," Journal of Business Economics and Statistics 15, 432-444.

Hong, Yongmiao, and Tae-Hwy Lee. (2002). "Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models." Review of Economics and Statistics, forthcoming.

Kanoh, Satoru. (1990). "The Statistical Reconsideration of the EPA Diffusion Index." Journal of the Japanese and International Economies 4, 139-156.

Kariya, Takeaki. (1993). Quantitative Methods of Portfolio Analysis: MTV Model Approach, Dordrecht: Kluwer Academic Publishers.

Kim, Chang-Jin, and Charles R. Nelson. (1998). "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime-Switching." Review of Economics and Statistics 80, 188-201.

Lee, Tae-Hwy. (2001). "Neural Network Test and Nonparametric Kernel Test for Neglected Nonlinearity in Regression Models." Studies in Nonlinear Dynamics and Econometrics 4, 169-182.

Lee, Tae-Hwy, Halbert White, and Clive W. J. Granger. (1993). "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests." Journal of Econometrics 56, 269-290.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson. (2003). "Macroeconomic Forecasting In the Euro Area: Country Specific Versus Area-Wide Information." European Economic Review 47, 1-18.

Ramsey, J.B. (1969). "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis." Journal of the Royal Statistical Society B 31, 350-371.

Sargent, Thomas J., and Christopher A. Sims. (1977). "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory." In New Methods of Business Cycle Research, edited by C. A. Sims, pp. 45-109. Minneapolis: Federal Reserve Bank of Minneapolis.

Shintani, Mototsugu. (2003). "Finite Sample Performance of Principal Components Estimators for Dynamic Factor Models." Manuscript, Vanderbilt University.

Stock, James H., and Mark W. Watson. (1989). "New Indexes of Coincident and Leading Economic Indicators." In NBER Macroeconomics Annual, edited by O. Blanchard and S. Fischer, pp. 351-394. Cambridge: MIT Press.

____. (1998). "Diffusion Indexes." NBER Working Paper no. 6702.

____. (2002). "Macroeconomic Forecasting Using Diffusion Indexes." Journal of Business and Economic Statistics 20, 147-162.

Swanson, Norman R., and Halbert White. (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.

Teräsvirta, Timo, Chien-Fu Lin, and Clive W. J. Granger. (1993). "Power of The Neural Network Linearity Test." *Journal of Time Series Analysis* 14, 209-220.

Timmermann, Allan. (2000). "Moments of Markov Switching Models." *Journal of Econometrics* 96, 75-111.

Watanabe, Toshiaki. (2001). "Measuring Business Cycle Turning Points In Japan With a Dynamic Markov Switching Factor Model." Manuscript, Tokyo Metropolitan University.

West, Kenneth D. (1996). "Asymptotic Inference About Predictive Ability." *Econometrica* 64, 1067-1084.

White, Halbert. (1989). "An Additional Hidden Unit Test for Neglected Nonlinearity In Multilayer Feedforward Networks." *Proceedings of the International Joint Conference on Neural Networks*, pp. 451-455. New York: IEEE Press.

Table 1
Testing for Linearity

	RESET	NN	NN-HAC	NN-LM	KERNEL	RESET	NN	NN-HAC	NN-LM	KERNEL
			$p = 1$					$p = 2$		
$k = 1$	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**0.02**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**
$k = 2$	**0.07**	0.16	0.39	**0.04**	0.78	0.29	0.18	0.99	**0.03**	**0.06**
$k = 3$	**0.03**	**0.09**	0.57	**0.02**	0.89	**<0.01**	**0.03**	0.99	**<0.01**	0.90
$k = 4$	**0.04**	**0.04**	0.29	**0.02**	0.42	0.26	0.33	0.99	0.17	0.63
$k = 5$	**0.02**	**<0.01**	0.12	**0.01**	**0.08**	0.47	**0.08**	0.97	**<0.01**	0.78
$k = 6$	**0.03**	**0.03**	0.24	**0.03**	0.52	0.22	0.22	0.99	0.13	0.87
			$p = 3$					$p = 4$		
$k = 1$	0.54	**0.04**	0.97	**0.02**	0.18	0.16	0.30	0.98	**0.02**	0.32
$k = 2$	**<0.01**	**<0.01**	0.90	**<0.01**	**0.07**	**<0.01**	**<0.01**	0.87	**<0.01**	**<0.01**
$k = 3$	**<0.01**	**<0.01**	0.93	**<0.01**	0.34	**<0.01**	**0.03**	0.96	**<0.01**	0.37
$k = 4$	0.11	**<0.01**	0.93	**0.02**	0.42	**0.06**	**<0.01**	0.83	**<0.01**	0.44
$k = 5$	0.95	**<0.01**	0.76	**<0.01**	0.46	0.83	**<0.01**	0.73	**<0.01**	0.42
$k = 6$	0.33	0.78	0.99	0.56	0.86	0.54	0.97	1.00	**0.10**	0.80

Notes: Numbers are p-values of the tests for the null hypothesis of linearity applied to each k-th diffusion index ($\widetilde{f}_t^{(k)}$). See Ramsey (1969) for the RESET, White (1989) for the neural network tests (NN and NN-HAC), Teräsvirta, Lin and Granger (1993) for the LM type neural network test (NN-LM), and Fan and Li (1997) for the kernel test (KERNEL), respectively. The Bartlett kernel with an automatic lag selection procedure of Andrews (1991) is used in the HAC covariance estimation.

Table 2
Linear Diffusion Index Forecast

Series	Model	MSFE						MAFE					
		vs. AR			vs. LI			vs. AR			vs. LI		
		Ratio	DM	p-value	Ratio	DM	p-value	Ratio	DM	p-value	Ratio	DM	p-value
IP	LI	1.11	-0.70	0.76	—	—	—	1.09	-1.34	0.74	—	—	—
	DI1	0.95	0.48	0.32	0.86	1.12	0.13	0.94	1.34	**0.09**	0.86	2.42	**<0.01**
	DI2	0.80	1.71	**0.04**	0.72	2.52	**<0.01**	0.87	2.41	**<0.01**	0.80	4.26	**<0.01**
SHIP	LI	1.16	-0.98	0.84	—	—	—	1.05	-0.77	0.78	—	—	—
	DI1	0.99	0.08	0.47	0.86	1.10	0.14	0.98	0.37	0.36	0.94	0.91	0.18
	DI2	0.89	1.55	**0.06**	0.74	2.37	**<0.01**	0.92	1.73	**0.04**	0.88	2.31	**0.01**
CAP	LI	1.15	-0.97	0.83	—	—	—	1.03	-0.58	0.74	—	—	—
	DI1	0.86	1.56	**0.06**	0.75	1.64	**0.05**	0.86	2.69	**<0.01**	0.84	2.31	**0.01**
	DI2	0.72	3.42	**<0.01**	0.63	2.71	**<0.01**	0.80	4.60	**<0.01**	0.78	3.99	**<0.01**
SALE	LI	1.08	-0.34	0.64	—	—	—	1.08	-0.62	0.73	—	—	—
	DI1	0.98	0.12	0.45	0.91	0.53	0.30	0.99	0.08	0.47	0.92	0.94	0.17
	DI2	0.78	1.13	0.13	0.73	1.69	**0.05**	0.90	0.93	0.18	0.83	1.95	**0.03**
HOUR	LI	1.20	-1.76	0.96	—	—	—	1.12	-1.71	0.96	—	—	—
	DI1	1.04	-0.55	0.71	0.87	1.54	**0.06**	1.03	-0.56	0.72	0.92	1.35	**0.09**
	DI2	1.05	-0.44	0.67	0.78	1.08	0.14	1.01	-0.10	0.54	0.90	1.61	**0.05**
CPI	LI	1.01	-0.03	0.51	—	—	—	1.03	-0.35	0.64	—	—	—
	DI1	0.76	1.53	**0.06**	0.76	1.32	**0.09**	0.83	2.05	**0.02**	0.81	1.99	**0.02**
	DI2	0.70	1.50	**0.07**	0.70	1.89	**0.03**	0.83	1.87	**0.03**	0.80	1.99	**0.02**

Notes: The MSFE (MAFE) is the mean squared forecast error (mean absolute forecast error) of the 120 out-of-sample 6-month-ahead forecasts. Ratio is the MSFE (MAFE) of the forecasting model relative to that of benchmark models, AR and LI. DM is Diebold and Mariano's (1995) test for equal forecast ability. AR lag order is fixed to two. Forecasting series are: index of industrial production (IP); index of producer's shipments (SHIP); index of capacity utilization ratio (CAP); index of sales in small and medium-sized enterprises (SALE); index of non-scheduled worked hours (HOUR); and consumer price index (CPI).

Table 3
Nonlinear Diffusion Index Forecast

Series	Model	MSFE			MAFE		
		Ratio	DM	p-value	Ratio	DM	p-value
IP	NN1 vs. DI1	1.09	-1.39	0.92	1.07	-1.84	0.97
	NN2 vs. DI2	1.27	-1.54	0.94	1.10	-1.75	0.96
	COMB1 vs. DI1	1.03	-0.78	0.78	1.02	-1.27	0.90
	COMB2 vs. DI2	1.09	-1.45	0.93	1.03	-1.24	0.89
SHIP	NN1 vs. DI1	1.15	-1.51	0.93	1.07	-1.01	0.84
	NN2 vs. DI2	1.19	-1.48	0.93	1.11	-1.49	0.93
	COMB1 vs. DI1	1.03	-0.74	0.77	1.01	-0.44	0.67
	COMB2 vs. DI2	1.05	-0.83	0.80	1.04	-1.04	0.85
CAP	NN1 vs. DI1	1.02	-0.59	0.72	1.00	-0.08	0.53
	NN2 vs. DI2	1.11	-2.31	0.99	1.09	-2.85	1.00
	COMB1 vs. DI1	0.99	0.39	0.35	0.99	0.46	0.32
	COMB2 vs. DI2	1.01	-0.36	0.64	1.03	1.57	0.94
SALE	NN1 vs. DI1	1.54	-1.10	0.86	1.12	-1.30	0.90
	NN2 vs. DI2	1.19	-0.87	0.81	1.05	-0.66	0.75
	COMB1 vs. DI1	1.11	-0.66	0.75	1.00	-0.09	0.53
	COMB2 vs. DI2	0.96	0.46	0.32	0.98	0.59	0.28
HOUR	NN1 vs. DI1	0.91	0.63	0.26	0.90	1.41	**0.08**
	NN2 vs. DI2	0.87	1.24	0.11	0.92	1.27	0.10
	COMB1 vs. DI1	0.90	1.47	**0.07**	0.92	2.22	**0.01**
	COMB2 vs. DI2	0.86	2.32	**0.01**	0.90	2.36	<**0.01**
CPI	NN1 vs. DI1	1.78	-0.93	0.82	1.00	0.02	0.49
	NN2 vs. DI2	3.21	-1.85	0.97	1.28	-1.52	0.94
	COMB1 vs. DI1	1.15	-0.58	0.72	0.95	0.83	0.20
	COMB2 vs. DI2	1.56	-1.55	0.94	1.09	-0.98	0.84

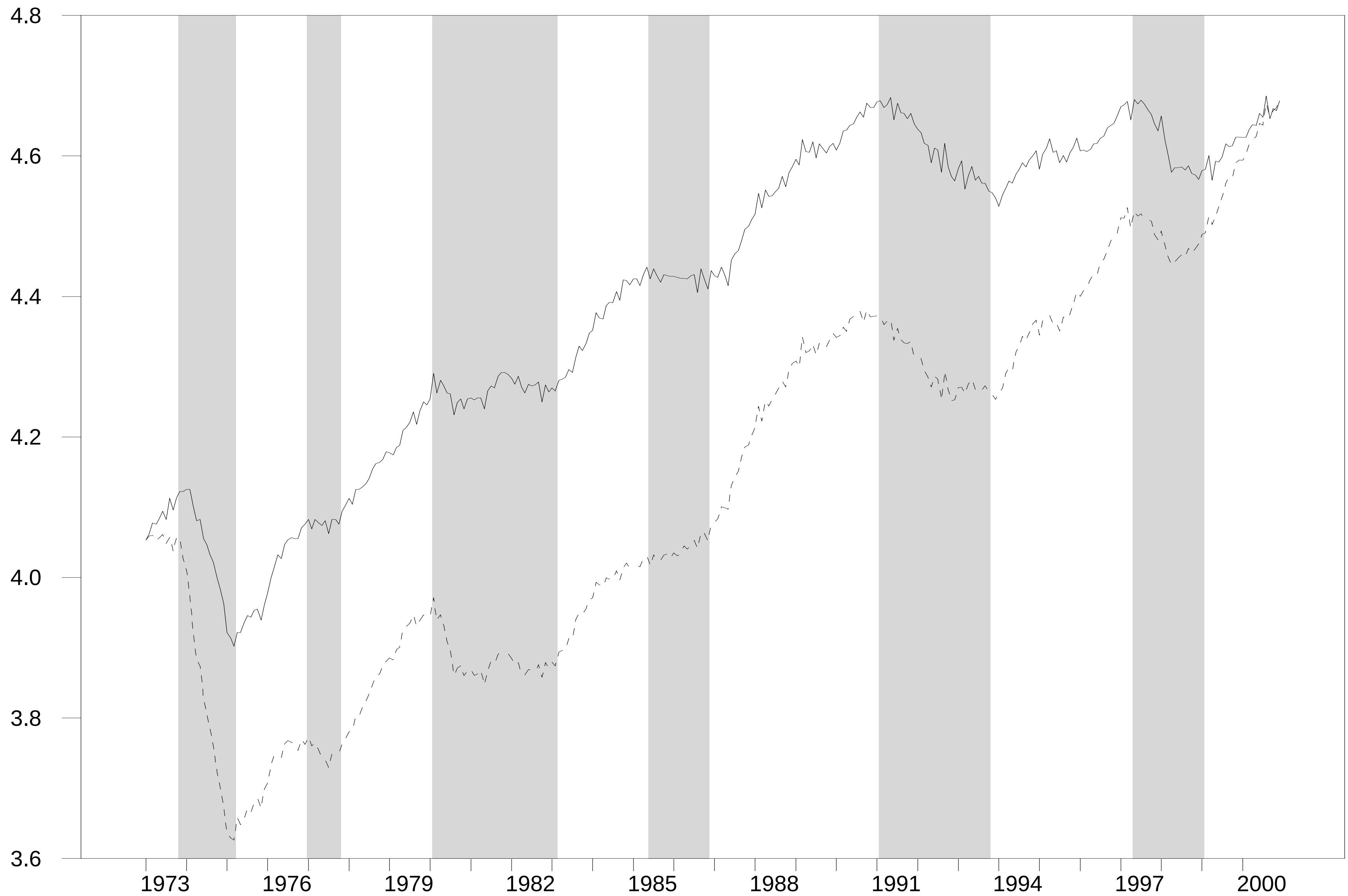
Notes: NN1 and NN2 are nonlinear versions of DI1 and DI2, respectively. Nonlinear models are based on ANN estimation where the number of hidden units is selected by BIC. COMB1 (COMB2) is the combination forecast of DI1 (DI2) and NN1 (NN2). See notes to Table 2.

Table 4
Test for Nested Forecasting Models

Series	Model	Linear test (DI1/DI2 vs. AR)			Nonlinear test (NN1/NN2 vs. DI1/DI2)		
		CCS	df	p-value	CCS	df	p-value
IP	DI1	9.74	6	0.14	7.03	3	0.11
	DI2	18.79	12	**0.09**	7.14	3	**0.06**
SHIP	DI1	8.81	6	0.18	4.18	3	0.57
	DI2	24.05	12	**0.02**	6.31	3	0.12
CAP	DI1	21.34	6	$<$**0.01**	6.15	3	0.22
	DI2	31.80	12	$<$**0.01**	6.11	3	0.12
SALE	DI1	11.65	6	**0.07**	7.19	3	0.13
	DI2	18.03	12	0.12	9.61	3	0.40
HOUR	DI1	25.76	6	$<$**0.01**	7.06	3	**0.06**
	DI2	42.78	12	$<$**0.01**	10.88	3	$<$**0.01**
CPI	DI1	62.71	6	$<$**0.01**	10.35	3	$<$**0.01**
	DI2	83.48	12	$<$**0.01**	8.22	3	$<$**0.01**

Notes: CCS is Chao, Corradi and Swanson's (2001) test for equal forecast ability of nested models based on MSFE. The mean of CCS from five draws is shown for the nonlinear tests.

Figure 1. Industrial Production (IP) and 1st Principal Component



Note: In logarithms. Industrial production (solid line) and 1st principal component (dotted line).

Figure 2. Smoothed Probability of a Recession

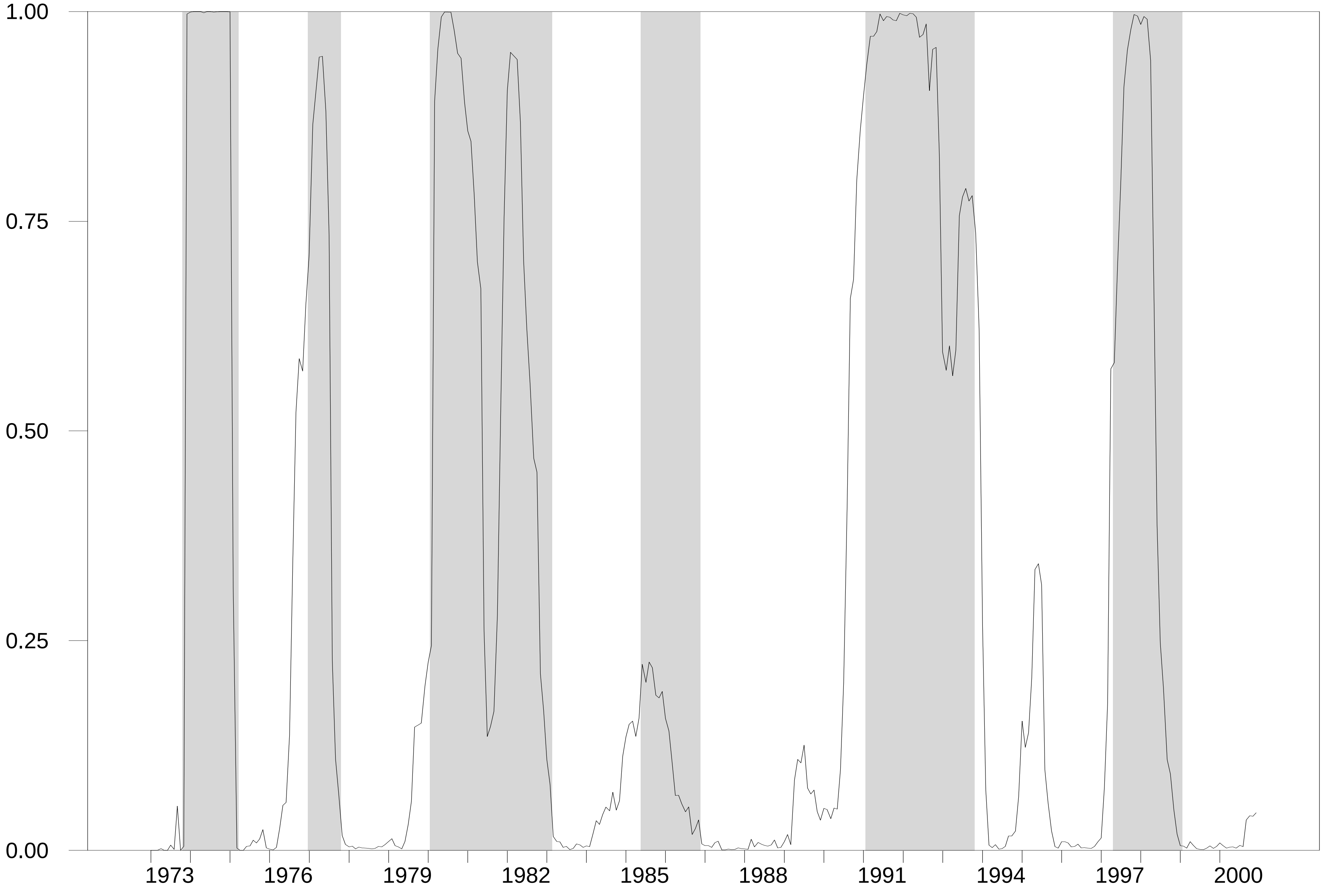


Figure 3
Linear AR(2) Estimate of 1st Principal Component

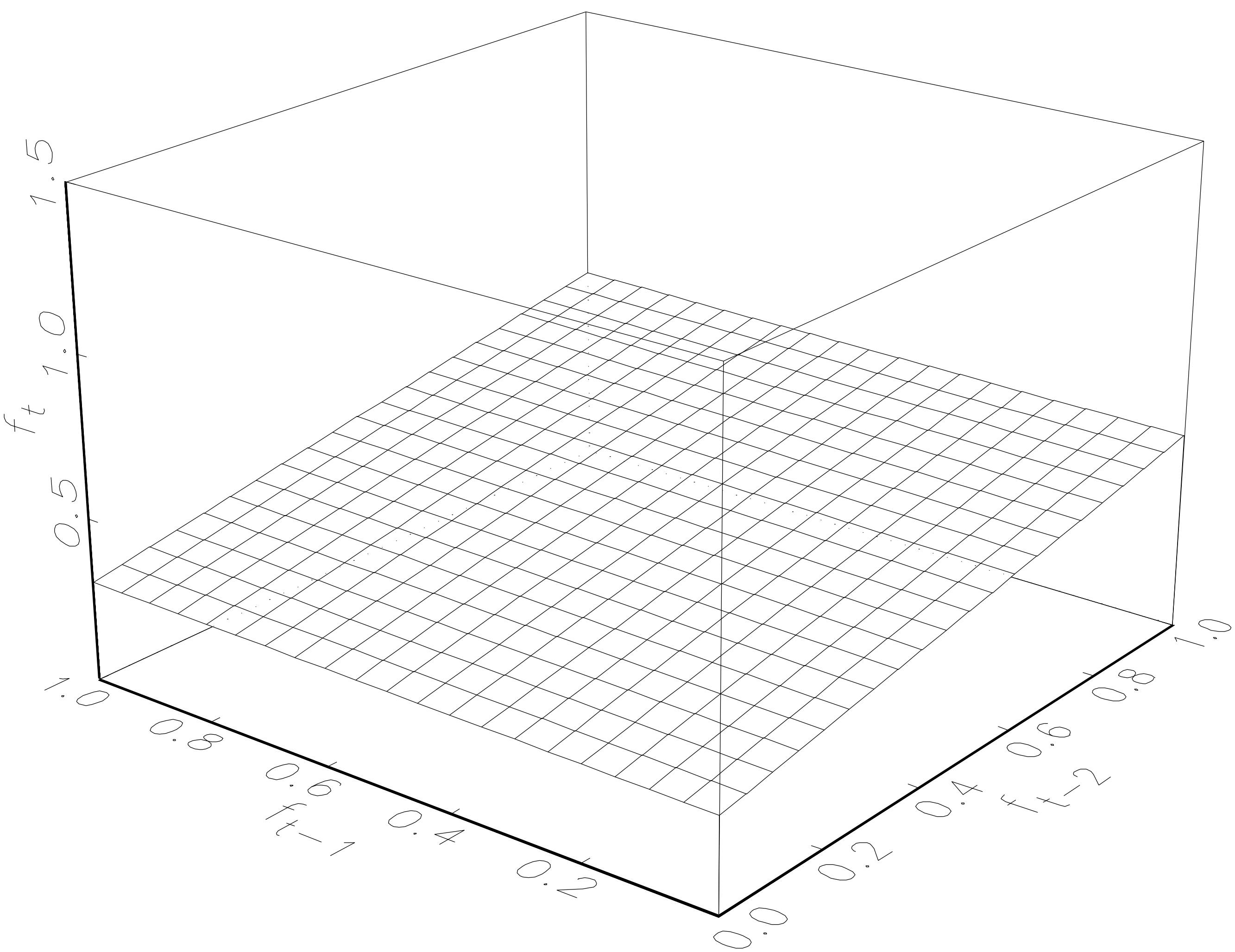


Figure 4
Nonlinear AR(2) Estimate of 1st Principal Component