UNIVERSITY OF ROME
"TOR VERGATA"

SCHOOL OF ECONOMICS

Ph.D. Program in

Econometrics and Empirical Economics

(XXII Cycle)

Nonlinear Forecasting Using a Large Number of Predictors

Alessandro Giovannelli

A.Y. 2009/2010

Advisor: Professor Marco Lippi

Coordinator: Professor Franco Peracchi *to my parents*

to Francesca

to my syster

Abstract

This dissertation aims to introduce a nonlinear model to forecast macroeconomic time series using a large number of predictors, namely the Feedforward Neural Network - Dynamic Factor Model (FNN-DF). The technique used to summarize the predictors in a small number of factors is Generalized Dynamic Factor Model, while the method used to capture nonlinearity is artificial neural networks, specifically Feedforward Neural Network. Commonly in GDFM literature, forecasts are made using linear models. However linear techniques are often misspecified and the resulting forecasts provide only a poor approximation to the best possible forecast. In an effort to address this issue, the technique we propose is FNN-DF. To determine the practical usefulness of the model, we conducted several pseudo forecasting exercises on 8 series of the United States economy. The series we were interested in forecasting were grouped in real and nominal categories. This method was used to construct the forecasts at 1-, 3-, 6-, and 12-month horizons for monthly U.S. economic variables using 131 predictors. The empirical study shows that FNN-DF has good ability to predict the variables under study in the period before the start of the "Great Moderation", namely 1984. After 1984, FNN-DF has the same accuracy in forecasting with respect to the benchmark.

Keywords: Factor model; Principal components analysis; Artificial neural networks; Nonlinear modeling; Bayesian Regularization; Forecasting.

Contents

Introduction

The main objective of this dissertation is to provide a new forecasting technique for macroeconomics and financial variables obtained from the combination of two different methods of data modeling already well-established in economic forecasting literature. The proposed technique is related to factor analysis and artificial neural networks. In the contest of forecasting, we restrict our discussion to point forecast constructed as an approximations of the conditional expectation of a target variable, say y_t, given a set of informative predictors, say x_t, namely $\mu(x_t) = \mathrm{E}\,(y_t|x_t)$. In general $\mu(x_t)$ provides the best possible prediction of y_t given x_t in terms of the mean squared forecast error, where the function μ solves the

$$\arg\min_{f\in\mathcal{F}} \mathrm{E}\left[(y_t - f(x_t))^2\right]$$

where $\mathcal{F}$ is the collection of function f of x_t. As pointed out by Lee et al. (1993), μ is not known and the forecaster typically choses a model for μ; if μ belongs to $\mathcal{F}$ we can affirm that the model is correctly specified instead, if μ does not belong to $\mathcal{F}$ we can affirm that the model is not correctly specified. Since we rarely have enough information to correctly specify μ, our discussion will be concentrated on the construction of forecast as the best possible approximation for μ. Then one of the main objectives in this dissertation is to construct a model that best approximates μ.

The increasing availability of macroeconomic variables provides the opportunity to explore a much richer base of information which can be used to improve accuracy in forecasting key macroeconomic variables such as inflation or industrial production. However, many of the statistical tools available in literature are not able to efficiently use the information available. One of the main problems encountered is the "curse of dimensionality", where the number of parameters to be estimated increases dramatically when the number of economic variables is increased. The result is that the estimates are not efficient and the forecast error grows proportionally to the number of variables used.

The aim of factor analysis is to provide a tool that summarizes the information contained in a large dataset in a small number of factors. This type of information extraction has been successfully applied to many fields of economic research, ranging from asset pricing theory to business cycle. The basic idea that stands behind the factor model is that the movement of a time series can be characterized as the sum of two mutually orthogonal components, the common component is a linear combination of the common factors and should explain the main part of the variance of the time series; the second, the idiosyncratic component, contains the remaining variable-specific information and is only weakly correlated across the dataset. However, neither the common nor the idiosyncratic component can be observed directly and have to be estimated.

In this dissertation we open with a discussion on a recent identification and estimation scheme proposed by Forni and Lippi (2001). The common components in their so-called generalized dynamic factor model (GDFM) are derived by assuming a dataset with an infinite number of series and an infinite number of observations. It thereby combines the so-called approximate static factor model proposed by Chamberlain and Rothschild (1983) and recently by Stock and Watson (2002a,b), and the dynamic factor model proposed by Geweke (1977) and Sargent and Sims (1977) for which cross-sectional and serial correlation was allowed. The model is called dynamic since the common shocks may not impact a series simultaneously, as in the static model, but they can propagate with leads or lags. In the GDFM the common components are inherently unobservable and are estimated by using the technique of dynamic principal components. While the familiar static principal components are based on an eigenvalue decomposition of the contemporaneous covariance matrix, dynamic principal components are based on the spectral density matrix (i.e. dynamic covariations) of the data and consequently are averages of the data, weighted and shifted through time.

In this type of literature one of the possible limitations is the usual assumption that the common shocks have a linear relationship with each series contained in the panel. We could have, for the discussion made at the beginning of this introduction, the situation where the assumption of linearity is not adequate to capture some economic behavior, since as previously stated we do not have sufficient information for it. Indeed, it may happen that in some periods a particular economy is characterized by the presence of two or more regimes, for instance phases of expansion or recession; the same reasoning is valid for financial variables for which it is very common to have periods of high and low volatility. For the reason described above, the common components could have an impact that is transferred on each series in the panel in a nonlinear way. Then, in these periods it is possible to improve accuracy in forecasting, using nonlinear techniques. The number of possible models that can be applied is enormous and is constantly expanding; although many of these have been successfully applied in the literature; the use of neural networks is motivated by their important mathematical properties, in particular the approximation capabilities, since our goal is to build a model that best approximates μ. The use of such a technique involves theoretical complications that often inhibit practical use. The main issues that are often encountered are the computational burden and the danger of overfit. In this framework we introduce a new model, feedforward neural network-dynamic factor model (FNN-DF).

The dissertation continues by presenting artificial neural network which is a parallel distributed statistical model made up of simple units that process information in currently available data and makes generalizations for future observations. In the context of artificial neural network literature, the input and the output can be interpreted respectively as regressors and regressands as in a regression model. Estimation of the parameters in a neural network is often called training, which is equivalent to the parameter estimation in a regression model. In this dissertation a particular network structure called Feedforward Neural Network (FNN) is used, to have more informations on FNN see Fine (1999). In the FNN the processing units or neurons are organized in the form of layers. We have at least three layers: an input layer, an output layer, and hidden layers, which are the layers between the input and the output. The processing units which correspond to the hidden layers are called hidden units. The source nodes in the input layer constitute the input signals to the neurons in the second layer.

Among neural networks, FNN is one of the most recent techniques used and its use is motivated by the results from the Universal Approximation Theorem. This theorem is very important since it has been shown by Hornik et al. (1989) and Cybenko (1989) that every continuous functions defined on a compact set can be arbitrarily well approximated with a FNN even if it is formed by a single layer.

A FNN derives approximation capabilities from its parallel structure. The hidden units are used to process information and include an activation function which describes the nature between input and hidden units. The universal approximation capabilities emerge when a number of activation functions are connected through a set of parameters. This parallel structure has the advantage of decomposing a large problem into a number of simpler problems. The universal approximation property is an important result theoretically and has immediate implications for financial and economic modeling. Finally the new theory is applied to demonstrate the practical usefulness of the FNN-DF model. In particular we conducted several forecasting exercises on different series of the U.S. economy.

Chapter 1

The Generalized Dynamic Factor Models

In the macroeconomic literature, traditional factor analysis is based on theoretical support where the number of variables (n) and the number of observations (T) are small. When we have a dataset where n and T tend to be very large, this exhibits a computational problems due to the increasing number of parameters to estimate. In this chapter a new concept of factor analysis particularly suited to situations where the available dataset is very large is presented. From a theoretical point of view large factor analysis is particularly interesting, since it is possible to assume, under certain conditions, that the panel of available variables can be decomposed into the sum of two components said the common and idiosyncratic components respectively. From a practical perspective it can be observed that an economy composed of many sectors, regions or individuals is characterized by a small number of variations common to all the variables under study.

1.1 Notations and Basic Assumptions

Let $\mathcal{P} = (\Omega, \mathcal{I}, P)$ be a probability space and let $L_2(\mathcal{P}, \mathbb{C})$ be the linear space of all complex-valued, zero mean, square-integrable random variables defined on Ω.

In this chapter we deal with a double sequence $\mathbf{x} = \{x_{it},\ i \in \mathbb{N}, t \in \mathbb{Z}\}$, where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ and with $\mathbf{x}_{nt} = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})'$ we denote the n-dimensional column vector for the observation made at time t. Given a complex matrix $\mathbf{D}$, we denote $\mathbf{D}^{'}$ as the transpose of $\mathbf{D}$ and $\mathbf{D}^{*}$ as the complex conjugate of $\mathbf{D}^{'}$. With θ we denote the real interval $[-\pi, \pi]$. Given the subset $G \subseteq L_2(\mathcal{P}, \mathbb{C})$ we denote the closed span of G as $\overline{span}(G)$ which is the minimum closed subspace of $L_2(\mathcal{P}, \mathbb{C})$ containing G. If S is a closed linear subspace of $L_2(\mathcal{P}, \mathbb{C})$ and $\mathbf{x} \subseteq L_2(\mathcal{P}, \mathbb{C})$, we denote $proj\,(\mathbf{x}|S)$ as the orthogonal projection of $\mathbf{x}$ on S.

Moreover we assume that for any $n \in \mathbb{N}$ the process $\mathbf{x}_{nt}$ is covariance stationary, that is $E\left[\mathbf{x}_{nt}\mathbf{x}'_{nt-k}\right] = \mathbf{\Gamma}^x_{nk}$ and $\mathbf{x}_{nt}$ has spectral density $\mathbf{\Sigma}^x_n$ with its entries $\sigma_{ij}(\theta)$ bounded in modulus and it is absolutely continuous with respect to the Lebesgue measure on θ

$$\mathbf{\Gamma}^x_{nk} = \int_{-\pi}^{\pi} e^{ik\theta}\mathbf{\Sigma}^x_n(\theta)\,d\theta$$

Given a double sequence $\mathbf{x} = \{x_{it},\, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ the model proposed by Forni and Lippi (2001) defined by the following.

Definition 1 (Generalized Dynamic Factor Model) *Let q be a nonnegative integer. The double sequence $\mathbf{x}$ is a q-dynamic factor sequence if $L_2(\mathcal{P}, \mathbb{C})$ contains an orthonormal q-dimensional white-noise vector process $\mathbf{u} = \left\{(u_{1t}\; u_{2t}\; \cdots\; u_{qt})',\, t \in \mathbb{Z}\right\} = \{\mathbf{u}_t,\, t \in \mathbb{Z}\}$, and a double sequence $\boldsymbol{\xi} = \{\xi_{it},\, i \in \mathbb{N},\, t \in \mathbb{Z}\}$ such that*

1. *For any $i \in \mathbb{N}$,*

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t \tag{1.1}$$

$$\chi_{it} = \underline{\mathbf{b}}_1(L)\,u_{1t} + \underline{\mathbf{b}}_2(L)\,u_{2t} + \cdots + \underline{\mathbf{b}}_q(L)\,u_{qt} = \underline{\mathbf{B}}(L)\,\mathbf{u}_t,$$

   where $\underline{\mathbf{b}}_i \in L_2^q(\theta, \mathbb{C})$.

2. *For any $i \in \mathbb{N}$, $j = 1, 2, \ldots, q$ and $k \in \mathbb{Z}$, we have $\xi_{it} \perp \xi_{it-k}$, then $\xi_{it} \perp \chi_{st-k}$ for any $i \in \mathbb{N}$, $s \in \mathbb{N}$ and $k \in \mathbb{Z}$.*

3. *$\boldsymbol{\xi}$ is idiosyncratic.*

4. *Putting $\boldsymbol{\chi} = \{\chi_{it},\, i \in \mathbb{N},\, t \in \mathbb{Z}\}$, $\lambda^{\chi}_q(\theta) = \infty$ a.e. in θ.*

The double sequences $\boldsymbol{\chi}$ and $\boldsymbol{\xi}$ are referred to as the common and the idiosyncratic component of representation (1.1).

The corresponding statistical model for the vector of observable $\mathbf{x}_{nt}$ is

$$\begin{aligned}\mathbf{x}_{nt} &= \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} \\ &= \underline{\mathbf{B}}_n(L)\,\mathbf{u}_t + \boldsymbol{\xi}_{nt}\end{aligned} \tag{1.2}$$

where $\underline{\mathbf{B}}_n(L) = \left(\underline{\mathbf{b}}_{n1}(L) \cdots \underline{\mathbf{b}}_{nq}(L)\right)$ is an $n \times q$ matrix.

1.2 Identification of the Model

The problem of identification of model (1.1) refers to conditions on the variance-covariance of the data for which the common and idiosyncratic components are identified. In Forni et al. (2000) conditions on the spectral density matrix of $\mathbf{x}$ have been defined under which the components are identified as n goes to infinity. The asymptotic identification is the precondition to develop an estimator for the components which are consistent for n and T going to infinity. The essential assumptions for identification are the following.

Assumption 1 (Identification) *Given a double sequence* $\mathbf{x} = \{x_{it},\, i \in \mathbb{N}, t \in \mathbb{Z}\}$*, where* $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ *and*

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \underline{\mathbf{B}}(L)\, \mathbf{u}_t + \boldsymbol{\xi}_t$$

we suppose that

(i) *the q-dimensional vector process* $\left\{(u_{1t}\, u_{2t}\ \cdots u_{qt})',\ t \in \mathbb{Z}\right\}$ *is an orthonormal white noise. That is,* $\mathrm{E}(u_{jt}) = 0$*;* $\mathrm{var}(u_{jt}) = 1$ *for any* j *and* t*;* $u_{jt} \perp u_{jt-k}$ *for any* j*,* t*, and* $k \neq 0$*;* $u_{jt} \perp u_{s,t-k}$ *for any* $s \neq j$*,* t *and* k*.*

(ii) $\boldsymbol{\xi} = \{\xi_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$ *is a double sequence such that,* $\boldsymbol{\xi}_n = \left\{(\xi_{1t}\, \xi_{2t}\ \cdots \xi_{nt})',\ t \in \mathbb{Z}\right\}$ *is a zero-mean stationary vector process for any* n*, and* $\xi_{it} \perp u_{j,t-k}$ *for any* i*,* j*,* t*,* k*;*

(iii) *the filters* $\underline{\mathbf{B}}(L)$ *are one-sided in* L *and their coefficients are square summable.*

The hypothesis assumed implies that the vector $\mathbf{x} = \{x_{it},\, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where $x_{it} \in L_2(\mathcal{P}, \mathbb{C})$ is stationary with mean equally zero for all n. Another consequence of Assumption (1) is the possibility to write the spectral density of $\boldsymbol{\Sigma}_n^x(\theta)$ as the sum of the spectral density matrix of common component $\boldsymbol{\Sigma}_n^{\chi}(\theta)$, and the spectral density matrix of idiosyncratic component, $\boldsymbol{\Sigma}_n^{\xi}(\theta)$. Moreover, in order to identify the latent variables above defined, the model needs an additional assumption.

Assumption 2 *For any* $i \in \mathbb{N}$*, there exist a real* $c_i > 0$ *such that* $\sigma_{ii}(\theta) \leq c_i$ *for any* $\theta \in [-\pi, \pi]$*. The first idiosyncratic dynamic eigenvalues* λ_{n1}^{ξ} *is uniformly bounded. That is, there exist a real*

Λ such that $\lambda^{\xi}_{n1}(\theta)$ for any $\theta \in [-\pi, \pi]$ and $n \in \mathbb{N}$. The first q common dynamic eigenvalues diverge almost everywhere in $[-\pi, \pi]$. That is $\lim_{n\to\infty} \lambda^{\chi}_{nj}(\theta) = \infty \quad$ for $j \leq q$, a.e. in $[-\pi, \pi]$.

As exposed by the authors, there is some intuition behind Assumption (2). Specifically, the two statements implies the following: First, the bound to the dynamic eigenvalues of the spectral density of idiosyncratic components indicates that the dynamic eigenvalues have effects concentrated on a limited number of variables. These tend to zero when the number of variables tends to infinite. Second, the divergence in the spectral density matrix of common components implies that the dynamic eigenvalues are present in a large number of observational units with non-decreasing importance among them.

If the Assumptions (1) and (2) are satisfied, Forni and Lippi (2001) show that the double sequence $\mathbf{x} = \{x_{it},\, i \in \mathbb{N}, t \in \mathbb{Z}\}$ is a **q-generalized dynamic factor model**.

Forni et al. (2000) propose the following method for consistently recovering the common components χ_{it}. Given the spectral density $\boldsymbol{\Sigma}^{x}_{n}$ of $\mathbf{x}_{nt}$, there exist n vectors of complex-valued functions

$$\mathbf{p}_{nj}(\theta) = \left(\mathrm{p}_{nj,1}(\theta)\ \ \mathrm{p}_{nj,2}(\theta)\ \ \cdots\ \ \mathrm{p}_{nj,n}(\theta)\right)$$

for $j = 1, \ldots, n$ such that

(a) $\mathbf{p}_{nj}(\theta)$ is a row eigenvector of $\boldsymbol{\Sigma}^{x}_{n}$, i.e.

$$\mathbf{p}_{nj}(\theta)\,\boldsymbol{\Sigma}^{x}_{n} = \lambda^{x}_{nj}(\theta)\,\mathbf{p}_{nj}(\theta) \quad \text{for any} \quad \theta \in [-\pi, \pi]\,;$$

(b) $|\mathbf{p}_{nj}(\theta)|^2 = 1$ for any j and $\theta \in [-\pi, \pi]$;

(c) $\mathbf{p}_{nj}(\theta)\,\mathbf{p}^{*}_{ns}(\theta) = 0$ for any $j \neq s$ and any $\theta \in [-\pi, \pi]$;

(d) $\mathbf{p}_{nj}(\theta)$ is θ-measurable on $[-\pi, \pi]$.

Through the properties (a)-(d), each eigenvector $\mathbf{p}_{nj}(\theta)$ can be expanded in Fourier series

$$\mathbf{p}_{nj}(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \mathbf{p}_{nj}(\theta)\, e^{ik\theta} d\theta\right] e^{-ik\theta},$$

thus allowing for the construction of a square-summable, n-dimensional, bilateral filter

$$\underline{\mathbf{p}}_{nj}(L) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left[\int_{-\pi}^{\pi}\mathbf{p}_{nj}(\theta)\, e^{ik\theta}d\theta\right] L^k.$$

Moreover, for $j = 1, \ldots, n$ the scalar process $\pi_{jt} = \left\{\underline{\mathbf{p}}_{nj}(L)\,\mathbf{x}_{nt},\ t \in \mathbb{Z}\right\}$ is called the j-th dynamic principal component of $\mathbf{x}_{nt}$. Now consider the minimal closed subspace of $L_2(\Omega, \mathcal{I}, P)$ containing the first q principal components

$$\mathcal{U}_n = \overline{span}\left(\underline{\mathbf{p}}_{nj}(L)\,\mathbf{x}_{nt},\ j = 1, \ldots, q,\ t \in \mathbb{Z}\right)$$

and the orthogonal projections

$$\begin{aligned}\chi_{it,n} &= \text{proj}\left(x_{it}|\mathcal{U}_n\right)\\ &= \underline{\mathbf{K}}_{ni}(L)\,\mathbf{x}_{nt} \end{aligned} \tag{1.3}$$

with $\underline{\mathbf{K}}_{ni}(L) = \text{p}^*_{n1,i}(L)\,\underline{\mathbf{p}}_{n1}(L) + \text{p}^*_{n2,i}(L)\,\underline{\mathbf{p}}_{n2}(L) + \cdots + \text{p}^*_{nq,i}(L)\,\underline{\mathbf{p}}_{nq}(L)$. Then, under the Assumptions (1) and (2) for all i and t the $\lim_{n\to\infty}\chi_{it,n} = \chi_{it}$ in mean square. This result indicates that the common component χ_{it} can be recovered asymptotically from the sequence $\underline{\mathbf{K}}_{ni}(L)\,\mathbf{x}_{nt}$.

1.3 The Static Representation

An alternative model for the large n case was developed by Stock and Watson (2002b). Their model, in time invariant formulation, can be written in the following form

$$\begin{aligned}\mathbf{x}_{nt} &= \mathbf{B}_n(L)\mathbf{u}_t + \boldsymbol{\xi}_{nt}\\ &= \boldsymbol{\Lambda}_n\mathbf{F}_t + \boldsymbol{\xi}_{nt}\end{aligned} \tag{1.4}$$

where $\mathbf{F}_t$ is an $r \times 1$ vector of common factors. Contrary to the specification by Forni and Lippi (2001), the common factors are not required to be uncorrelated in time, and they can also be correlated with the idiosyncratic components. In this case only var $[\mathbf{F}_t] = \mathbf{I}$ is required for identification. Suppose that the filter has finite order $m \geq 0$, i.e. $\mathbf{B}_n(L) = \mathbf{B}^n_0 + \mathbf{B}^n_1 L +$

$\ldots + \mathbf{B}_m^n L^m$, then the model in (1.1) can be written as in (1.4), where $\mathbf{F}_t = (\mathbf{u}_t^{'}, \mathbf{u}_{t-1}^{'}, \ldots, \mathbf{u}_{t-m}^{'})$ and the i-th row of $\mathbf{\Lambda}_n$ has elements $(\mathbf{B}_0^n, \mathbf{B}_1^n, \ldots, \mathbf{B}_m^n)$. The dimension of $\mathbf{F}_t$ is always equal to $r = q(m+1)$, where q is the dimension of $\mathbf{u}_t$. Although the relation between $\mathbf{x}_{nt}$ and $\mathbf{F}_t$ is static, $\mathbf{F}_t$ itself can be a dynamic process, depending on the dynamics of $\mathbf{u}_t$. Then, the static method makes use of representation (1.4)without taking into account the dynamic structure of $\mathbf{F}_t$. This implies that the common factors are dynamically singular and the spectral density matrix of $\mathbf{F}_t$ has rank q, which is smaller than r if $m > 0$.

Following Stock and Watson (2002b), let $\mathbf{\Gamma}_{n0}^{\chi}$ and $\mathbf{\Gamma}_{n0}^{\xi}$ be the covariance matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ respectively. Let μ_{nj}^{χ} and μ_{nj}^{ξ} be the largest eigenvalues, in descending orders, of $\mathbf{\Gamma}_{n0}^{\chi}$ and $\mathbf{\Gamma}_{n0}^{\xi}$ respectively.

Assumption 3 *We assume that the following hold:*

A) $\lim_{n\to\infty} \mu_{nj}^{\chi} = \infty$ *for* $1 \leq j \leq r$*;*

B) *there exists a real* M*, such that* $\mu_{nj}^{\xi} \leq M$ *for any* n*.*

Assumption (3)A establishes that, as n increases, the variance of $\mathbf{x}_{nt}$ explained by the first r eigenvalues of the common component increases to infinity. This means that as n goes to infinity the weight of the idiosyncratic component in explaining $\mathbf{\Gamma}_{n0}^{x}$ becomes smaller and smaller. Assumption (3)B sets out that the idiosyncratic components can be correlated, but the assumption puts a limit to the amount of correlation. As n increases, the variance of the vector $\mathbf{x}_{nt}$ captured by the largest eigenvalue of the idiosyncratic component, μ_{nr}^{χ}, remains bounded. Then, under the Assumptions (3)A and (3)B, the static projection on the first r static principal components of $\mathbf{x}_{nt}$ converge in mean square to the common component in equation (1.4) for $n \to \infty$.

To derive the form of the static principal components (SPC), we consider the finite realization of the form $\mathbf{x}_n^T = \{x_{it}\ i = 1 \ldots n,\ t = 1 \ldots T\}$ with the estimated contemporaneous variance-covariance $\check{\mathbf{\Gamma}}_0^x = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{nt}^T \mathbf{x}_{nt}^{T'}$. Now consider first the quantity $\check{\boldsymbol{\alpha}}_1 \mathbf{x}_{nt}$ where the $1 \times n$ vector $\check{\boldsymbol{\alpha}}_1$ maximizes the variance $\mathrm{var}\left[\check{\boldsymbol{\alpha}}_1 \mathbf{x}_{nt}\right] = \check{\boldsymbol{\alpha}}_1 \check{\mathbf{\Gamma}}_0^x \check{\boldsymbol{\alpha}}_1'$. Since the maximum will not be achieved for finite $\check{\boldsymbol{\alpha}}_1$ a normalization constraint must be imposed. The constraint used in derivation is $\check{\boldsymbol{\alpha}}_1 \check{\boldsymbol{\alpha}}_1' = 1$, namely the sum of squared elements of $\check{\boldsymbol{\alpha}}_1$ equals one. To maximize $\check{\boldsymbol{\alpha}}_1 \mathbf{\Gamma}_0^x \check{\boldsymbol{\alpha}}_1'$ subject to $\check{\boldsymbol{\alpha}}_1 \check{\boldsymbol{\alpha}}_1' = 1$, the standard approach is to use the technique of Lagrange multipliers. Maximize

$$\check{\boldsymbol{\alpha}}_1 \check{\mathbf{\Gamma}}_0^x \check{\boldsymbol{\alpha}}_1' - \mu_1 \left(\check{\boldsymbol{\alpha}}_1 \check{\boldsymbol{\alpha}}_1' - 1 \right)$$

where μ_1 is a Lagrange multiplier. Differentiation with respect to $\check{\boldsymbol{\alpha}}_1$ results in $\check{\boldsymbol{\alpha}}_1 \left(\check{\boldsymbol{\Gamma}}_0^x - \mu_1 \mathbf{I}\right) = 0$. Thus, $\check{\mu}_1$ is an eigenvalue of $\check{\boldsymbol{\Gamma}}_0^x$ and $\check{\boldsymbol{\alpha}}_1$ is the corresponding eigenvector. To decide which eigenvector results in $\check{\boldsymbol{\alpha}}_1 \mathbf{x}_{nt}$ with maximum variance, the quantity to be maximized is

$$\check{\boldsymbol{\alpha}}_1 \check{\boldsymbol{\Gamma}}_0^x \check{\boldsymbol{\alpha}}_1' = \check{\boldsymbol{\alpha}}_1 \check{\mu}_1 \check{\boldsymbol{\alpha}}_1' = \check{\mu}_1 \check{\boldsymbol{\alpha}}_1 \check{\boldsymbol{\alpha}}_1' = \check{\mu}_1.$$

So $\check{\mu}_1$ must be as large as possible. Thus, $\check{\boldsymbol{\alpha}}_1$ is the eigenvector corresponding to the largest eigenvalue of $\check{\boldsymbol{\Gamma}}_0^x$, and $\text{var}\left[\check{\boldsymbol{\alpha}}_1 \mathbf{x}_{nt}\right] = \check{\mu}_1$ the largest eigenvalue.
In general, the r-th PC is $\check{\boldsymbol{\alpha}}_r \mathbf{x}_{nt}$ and $\text{var}\left[\check{\boldsymbol{\alpha}}_r \mathbf{x}_{nt}\right] = \check{\mu}_r$, where $\check{\mu}_r$ is the largest eigenvalue of $\check{\boldsymbol{\Gamma}}_0^x$ and $\check{\boldsymbol{\alpha}}_r$ is the corresponding eigenvector. Then for $j = 1, 2, \ldots, r$ ordering the eigenvalues $\check{\mu}_j$ in descending order and taking the eigenvectors corresponding to the largest eigenvalue we define $\check{\mathbf{S}}_t = (\check{\boldsymbol{\alpha}}_1 \mathbf{x}_{nt} \ \check{\boldsymbol{\alpha}}_2 \mathbf{x}_{nt} \ \cdots \ \check{\boldsymbol{\alpha}}_r \mathbf{x}_{nt})'$ the j-th static principal component of $\mathbf{x}_{nt}$.

1.4 One Sided Estimation and Forecasting

Let us consider a finite realizations of the form $\mathbf{x}_n^T = \{x_{it} \; i = 1 \ldots n, \, t = 1 \ldots T\}$; the filters $\underline{\mathbf{K}}_{nj}(L)$ obtained as functions of the spectral density matrices $\boldsymbol{\Sigma}_n^x(\theta)$ are unknown and have to be estimated. Let us assume $\mathbf{x}_{nt}$ admits a linear representation of the form

$$\mathbf{x}_{nt} = \sum_{k=-\infty}^{\infty} \mathbf{c}_k \mathbf{Z}_{t-k} \tag{1.5}$$

where $\{\mathbf{Z}_t; \, t \in \mathbb{Z}\}$ is second-order white noise with nonsingular covariance matrix and finite fourth-order moments, and $\sum_{k=-\infty}^{\infty} |c_{ij,k}||k|^{1/2} < \infty$. Under equation (1.5) any periodogram smoothing or lag-window estimator $\check{\boldsymbol{\Sigma}}^x(\theta)$ is a consistent estimator of $\boldsymbol{\Sigma}_n^x(\theta)$ for T going to ∞. Now, in the following we provide a description of the spectral estimate $\check{\boldsymbol{\Sigma}}^x(\theta)$ considered throughout the section. The estimation of the spectral density is constructed using a Bartlett lag-window estimator of size $M = M(T)$. The sample covariance matrix is

$$\check{\boldsymbol{\Gamma}}_k^x = (T-k)^{-1} \sum_{t=k+1}^{T} \mathbf{x}_{nt}^T \mathbf{x}_{nt}^{T\prime}$$

Then we compute the $(2M+1)$ points discrete Fourier transformation of the truncated two-sided sequence $\check{\boldsymbol{\Gamma}}^x_{-M}, \ldots, \check{\boldsymbol{\Gamma}}^x_0, \ldots, \check{\boldsymbol{\Gamma}}^x_M$, where $\check{\boldsymbol{\Gamma}}^x_{-k} = \check{\boldsymbol{\Gamma}}^x_k$, that is

$$\check{\boldsymbol{\Sigma}}^x(\theta_h) = \sum_{k=-M}^{M} \check{\boldsymbol{\Gamma}}^x_k \omega_k e^{-ik\theta_h}$$

where $\theta_h = 2\pi h/(2M+1)$, for $h = 0, 1, \ldots, 2M$ and $\omega_k = 1 - \frac{|k|}{(M+1)}$ are the weights corresponding to the Bartlett lag-window of size $M = M(T)$. The choice of M represents the trade off between small bias (large M) and small variance (small M). Now, we can observe that the filters $\underline{\check{\mathbf{K}}}_{ni}(L)$ are infinite two-sided, that is

$$\underline{\check{\mathbf{K}}}_{ni}(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left[\int_{-\pi}^{\pi} \check{\mathbf{K}}_{ni}(\theta) e^{-ik\theta} d\theta \right] L^k,$$

while $\mathbf{x}_{nt}$ is not available, neither for $t \leq 0$ nor for $t > T$, then the projection $\underline{\check{\mathbf{K}}}_{ni}(L)\mathbf{x}_{nt}$ onto the space spanned by the q dynamic principal components cannot be computed. Therefore a truncated version of the estimated filter

$$\underline{\check{\mathbf{K}}}_{ni}(L) = \sum_{k=-M}^{M} \underline{\check{\mathbf{K}}}_{ni,k} L^k,$$

where the conditions $M(T) \to \infty$ and $M(T)/T \to \infty$ must be fulfilled, is considered.

The method discussed above produces an estimator to the common component which is a two-sided filter of the observations. As seen before, this method has the advantage of exploring the dynamic structure of the data and needs few dynamic aggregates to approximate the space spanned by the common factors, but the performance of the estimator $\check{\chi}_t$ deteriorates as t approaches T or 1. Indeed to compute the estimator for the last observation, one needs M future observations which are not available. For this reason, this makes the estimation procedure inappropriate for prediction.

Forni et al. (2005) propose a refinement of the original procedure which retains the advantages of the dynamic approach while obtaining a consistent estimate of the optimal forecast as a one-sided filter of the observations. The method consists of two steps. In the **first step**, they follow Forni et al. (2000) and obtain the cross-covariances for common and idiosyncratic components at all leads and lags from the inverse Fourier transformation of the estimated spectral density matrices. In the **second step**, they use these estimates to obtain the r contemporaneous linear combinations of the observations with the smallest idiosyncratic common variance ratio. The resulting aggregates can be obtained as the solution of a generalized principal component problem.

First Step. To estimate the common and idiosyncratic cross-covariance $\mathbf{\Gamma}^{\chi}_{nk}$ and $\mathbf{\Gamma}^{\chi}_{nk}$ respectively, we start with the matrix $\check{\mathbf{\Sigma}}^{x}(\theta)$, defined as a periodogram smoothing or lag window estimator of the spectral density $\mathbf{\Sigma}^{x}_{n}(\theta)$. Now, using Assumption (1), the spectral density $\check{\mathbf{\Sigma}}^{x}(\theta)$ can be decomposed in a spectral density matrix of the common component and idiosyncratic component

$$\check{\mathbf{\Sigma}}^{\chi}(\theta) = \sum_{j=1}^{q} \check{\mathbf{p}}^{x*}_{j}(\theta)\, \check{\lambda}^{x}_{j}(\theta)\, \check{\mathbf{p}}^{x}_{j}$$

and

$$\check{\mathbf{\Sigma}}^{\xi}(\theta) = \sum_{l=q+1}^{n} \check{\mathbf{p}}^{x*}_{l}(\theta)\, \check{\lambda}^{x}_{l}(\theta)\, \check{\mathbf{p}}^{x}_{l}.$$

Therefore, applying an inverse discrete Fourier transformation to these density matrices, the covariance matrices of $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ can be estimated as

$$\check{\mathbf{\Gamma}}^{\chi}_{k}(\theta) = \int_{-\pi}^{\pi} \check{\mathbf{\Sigma}}^{\chi}(\theta)\, e^{ik\theta} d\theta$$

and

$$\check{\mathbf{\Gamma}}^{\xi}_{k}(\theta) = \int_{-\pi}^{\pi} \check{\mathbf{\Sigma}}^{\xi}(\theta)\, e^{ik\theta} d\theta$$

Second Step. The estimated covariance matrix of the common components is used to solve the generalized principal component (GPC) problem. More precisely the objective is to find r independent linear combinations $\check{W}_{jt} = \check{\mathbf{Z}}_j \mathbf{x}_{nt}$, where the weights $\check{\mathbf{Z}}_j$ are defined as

$$\begin{aligned} \check{\mathbf{Z}}_j &= \underset{\mathbf{g}\in\mathbb{R}^n}{\arg\max}\ \mathbf{g}\check{\mathbf{\Gamma}}^{\chi}_{0}\mathbf{g}' \\ \text{s.t.} &\quad \mathbf{g}\check{\mathbf{\Gamma}}^{\xi}_{0}\mathbf{g}' = 1 \\ \text{and} &\quad \mathbf{g}\check{\mathbf{\Gamma}}^{\xi}_{0}\check{\mathbf{Z}}_l' = 0 \qquad \text{for} \quad 1 \le l < j. \end{aligned}$$

The solutions to the problem are

$$\check{\mathbf{Z}}_j \check{\mathbf{\Gamma}}_0^{\chi} = \check{v}_j \check{\mathbf{Z}}_j \check{\mathbf{\Gamma}}_0^{\xi} \qquad j = 1, \dots, r$$

where $\check{\mathbf{Z}}_j$ are the generalized eigenvectors associated with the generalized eigenvalues $\check{v}_j$, with the normalization conditions

$$\begin{aligned} \check{\mathbf{Z}}_l \check{\mathbf{\Gamma}}_0^{\chi} \check{\mathbf{Z}}_j' &= 1 \qquad \text{for} \quad l = j \\ &= 0 \qquad \text{for} \quad l \neq j. \end{aligned}$$

Ordering the eigenvalues $\check{v}_j$ in descending order and taking the eigenvectors to the largest r eigenvalues, we define $\check{\mathbf{G}}_t = \left(\check{\mathbf{Z}}_1 \mathbf{x}_{nt}\ \check{\mathbf{Z}}_2 \mathbf{x}_{nt}\ \cdots\ \check{\mathbf{Z}}_r \mathbf{x}_{nt}\right)'$ as the first r generalized principal component of $\mathbf{x}_{nt}$.

Forecasting with GDFM. Since we are interested in forecasting a single variable, we call y_t the variable of interest contained in $\mathbf{x}_n^T$. The two types of estimated factors, static and dynamic, will be used for prediction. For forecasting purposes a single equation is estimated with the one-step approach. The forecasting equation is estimated as a linear projection of h-step aheads of y_t, i.e. y_{t+h}, on t-dated predictors. In general a factor based forecast is specified as follows

$$y_{t+h} = \alpha + \beta \check{F}_t + \delta(L) y_t + \epsilon_{t+h} \tag{1.6}$$

where $\check{F}_t$ are the factors estimated using GPC as in Sec. (1.4) or SPC as in Sec. (1.3); α is the constant term and β the coefficient vector for the factors. They are estimated by ordinary least squares for each forecast horizon h. The autoregressive term is introduced by the coefficients $\delta(L)$, which is a polynomial with non-negative power of lag operator L. The variable y_{t+h} is defined as the growth rate of the chosen time series between period t and period $t + h$.

1.5 Determining the Number of Factors

As mentioned in the introduction, the most important feature of factor models is to summarize the information contained in a large panel of variables using a small number of factors. However, the exact number of factors to use is not known "a priori". Indeed, a controversial issue in the analysis of approximate factor models is the preliminary identification of the optimal numbers r and q of static and dynamic factors. In this dissertation the optimal number r of static factors is determined by the criterion proposed by Bai and Ng (2002), whereas the optimal number q of dynamic factors is determined by the criterion proposed by Hallin and Liska (2007).

Since, in empirical applications, we observe only a finite sequence of length T of a finite number n of variables, these two criteria are described using a finite realization of the form $\mathbf{x}_n^T = \{x_{it}\ i = 1 \dots n,\ t = 1 \dots T\}$.

Determining the Number of Static Factors. Bai and Ng (2002) propose using an information criteria to determine the optimal number of static factors r as a trade-off between goodness-of-fit and overfitting.
Formally, let $V(k) = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(x_{it} - \check{\boldsymbol{\Lambda}}_i^{(k)} \check{\mathbf{F}}_t^{(k)} \right)^2$ be the variance of the idiosyncratic term when the factor loadings $\check{\boldsymbol{\Lambda}}_i^{(k)}$ and the common factors $\check{\mathbf{F}}_t^{(k)}$ are estimated using k static factors by the method of static principal components described in Section (1.3). They define the information criterion

$$
\begin{aligned}
\mathrm{IC}(k) &= \log\left(V(k)\right) + k\, g(n,T) \\
\check{r}_{\mathrm{IC}} &= \underset{0 \le k \le r_{\max}}{\arg\min}\ \mathrm{IC}(k)
\end{aligned}
$$

The term $V(k)$ represents the goodness-of-fit which depends on the estimate of common factors and the number of factors. When the number k of factors is increased, the variance explained by the factors increases too, then $V(k)$ decreases. However in order to avoid overfitting they introduce the penalty term $g(n,T)$ which is an increasing function of n and T. The information criterion $\mathrm{IC}(k)$ has to be minimised in order to determine the optimal number of static factors. In empirical application we have to fix a maximum number of static factors, say $r_{\max}$, and estimate the model for all numbers of factors $k = 1, \dots, r_{\max}$. As a penalty function Bai and Ng (2002) propose to use $g(n,T) = \left(\frac{n+T}{nT}\right) \log\left(\min\{n,T\}\right)$.

Determining the Number of Dynamic Factors. Hallin and Liska (2007) proposed a method for determining the number of factors in GDFM. In the generalized dynamic factor models the criterion proposed by Hallin and Liska exploits the relation between the number of dynamic factors and the number of diverging eigenvalues of the spectral density matrix of $\mathbf{x}_n^T$.

The information criterion proposed, associated with the estimated spectral density $\check{\boldsymbol{\Sigma}}^x$ and its eigenvalues λ_{ni}^T, is

$$
\begin{aligned}
\mathrm{IC}(k) &= \log\left[\frac{1}{n}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{h=-M_T}^{M_T}\lambda_{ni}^T(\theta_h)\right] + c\,k\,\bar{g}(n,T)\\
\check{q}_{\mathrm{IC}} &= \underset{0\le k\le q_{\max}}{\arg\min}\ \mathrm{IC}(k)
\end{aligned}
$$

The authors suggest using $M_T = [0.5\sqrt{T}]$ or $[0.7\sqrt{T}]$ and as penalty function $\bar{g}(n,T) = (M_T^{-2} + M_T^{1/2}T^{-1/2} + n^{-1})\log\,A_T$ with $A_T = (\min\{n, M_T^2, M_T^{-1/2}T^{1/2}\})$. Therefore, the penalty function should be large enough to avoid overestimation of $\check{q}_{\mathrm{IC}}$, but at the same time it should not over penalize. Multiplying the penalty function by a constant c is a way to tune the penalizing power of $\bar{g}(n,T)$. Hallin and Liska propose an automatic procedure for selecting $\check{q}_{\mathrm{IC}}$ which basically explores the behavior of the variance of the selected $\check{q}_{\mathrm{IC}}$ for the whole region of values of the constant c.

Chapter 2

The Feedforward Neural Network Dynamic Factor

The aim of this chapter is to introduce a new technique, called Feedforward Neural Network Dynamic Factor model, FNN-DF, which forecasts macroeconomic time series using a large number of predictors. The model proposed to summarize information contained in the whole set of predictors is GDFM. Commonly, in the GDFM literature, the forecasts are made using linear model, namely the relationship between the variables to be predicted and the common factors is supposed to be linear. However, linear forecasting models are often misspecified and the resulting forecast provides only a poor approximation to the best possible forecast. It is possible to obtain superior approximations to the optimal forecast using nonlinear methods. The nonlinearity needs to be described through an adequate model. Unfortunately, for many applications, theory does not guide the model building process by suggesting the relevant functional form. This particular difficulty makes it attractive to consider an "atheoretical" but flexible class of statistical models. Artificial Neural Networks of the FNN type are essentially semi-parametric regression estimators and are well suited for this purpose, as they can approximate any function up to an arbitrary degree of accuracy.

2.1 Motivations

In general, as pointed by McNelis (2004) in the Preface, we assume that the economy under study comes from a linear data generating process and shocks are from an underlying normal distribution and represent small deviations around a steady state. In this case standard tools such as classical linear regression are perfectly appropriate. However, making use of the linear model with normally generated disturbances may lead to poor results if the real world deviates significantly from these assumptions of linearity and normality. For this reason it is important to verify if, also in GDFM literature, the use of linear techniques are justified not only for their simplicity but also by empirical evidence. We propose the FNN-DF model as an alternative to linear models.

Here we report some important dates in the history of artificial neural networks. Historically, neural network theory was motivated by the idea that certain key properties of the human brain can be extracted and applied in order to create, in simplified form, an artificial brain. The neural network era began with the pioneering work of McCulloch and Pitts proposed in 1943, in particular see Arbib (1998). It is a multiple input summing device that consists of different weighting for each input and a threshold before the output. The significance at that time was its ability to compute any arithmetic or logical function. During the 1950s numerous neural networks results were reported. Nevertheless during the mid 1960s research into neural networks was abandoned for about two decades.
In 1982, J. Hopfield published two important articles regarded as the beginning of the current neural network era. J. Hopfield presented a novel idea in which he stated that the approach to artificial intelligence should not be purely to imitate the human brain but, instead, to use its concepts to build machines that could solve dynamic problems.
Cybenko (1989) published a very important piece proving the universal functional approximation ability of neural networks. In the same period, Hornik et al. (1989) also reported their findings on proving multilayered perceptron networks as universal approximators. Rumelhart et al. (1988) reported on the developments of the backpropagation algorithm. The paper discussed how backpropagation learning had emerged as the most popular learning set for the training of multilayered perceptrons. Subsequently, neural networks have been widely applied in many different scientific areas. Currently, neural networks have extended from speech recognition to time series forecasting.

Regarding the estimation of the FNN-DF model, we have to distinguish two stages. The first is the estimation of the factors and the second refers to the architectural selection and the estimation of the FNN, which reveals the relationship between the estimated factors and the variables to predict. Regarding the estimation of the factors, we assume that each series can be divided into a common part, which depends on some dynamic factors, and an idiosyncratic part, which is variable specific. Using the technique proposed by Forni et al. (2005), described in Section (1.4), and that proposed by Stock and Watson (2002a), described in Section (1.3), we obtain an estimate of dynamic and static factors, respectively. Once the estimation of the factors is obtained, we use those as regressors in FNN.

We proceed to the second stage of the proposed model, namely the architectural selection and estimation of FNN. However, at this stage it is important to clarify some concepts. Indeed, as has been pointed out, the FNN has the important property of being very flexible in approximating an arbitrary function. This ability of approximation requires a specification of the architecture (or model selection) of the FNN, which often is very complicated to obtain. Architectural selection requires choosing both the appropriate number of hidden units and the connections therein. Indeed, as reported in Anders and Korn (1999), one of the main problems in the literature on neural networks is which architecture should be used for a given problem. In general, a desirable architecture considers the trade-off between estimation bias and variability due to estimation errors and contains as few hidden units and connections as necessary for a good approximation of the true function. Consequently it is necessary to have a methodology to select the appropriate architecture. The usual approaches pursued in the network literature are regularization, pruning, and stopped training. The strategy we adopt in this dissertation, which turned out to be quite successful in a number of applications in time series forecasting, is Bayesian regularization proposed by Mackay (1992a,b). The fundamental idea is finding a balance between the number of parameters and goodness-of-fit by penalizing large models. The objective function is modified in such a way that the estimation algorithm effectively prunes the network by driving irrelevant parameter estimates to zero during the estimation process.

There are many types of estimating algorithms in the literature on neural networks and in general it is very difficult to know which estimating algorithm will be most efficient for a given problem. In literature, one of the most widely accepted techniques is the backpropagation algorithm. This algorithm employs only the first-order partial derivatives of the object function and has proved it usefulness in dealing with a large number of classification and function approximation problems. However, in practical applications, the large number of iterations needed to converge to the optimal parameters of the FNN becomes prohibitive for several applications. An alternative way to speed up the estimation phase is by using higher-order optimization methods that utilize second-order partial derivatives. The algorithm used to estimate FNN in this dissertation is

Marquardt-Levenberg. This algorithm is widely accepted as the most efficient in realization accuracy. It was designed to approach second order training speed without having to compute the Hessian matrix. Under the assumption that the object function is the squared sum of residuals, the Hessian matrix can be approximated using the Jacobian matrix that contains only first derivatives of the FNN errors with respect to parameters.

2.2 Nonlinear Dynamic Model

Before introducing the model it is important to define the problem that we discuss in this chapter. As presented by (Terasvirta, 2006, pag. 417), a general nonlinear dynamic model with an additive noise component can be defined as follows:

$$y_t = f\left(\mathbf{s}_t; \boldsymbol{\vartheta}\right) + \varepsilon_t \tag{2.1}$$

We define as $\mathbf{s}_t = (\mathbf{w}_t', \mathbf{z}_t')'$ the l-dimensional vector of explanatory variables, where $\mathbf{w}_t' = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})'$ is a $(p+1)$-dimensional vector of the lags of the variable of interest, $\mathbf{z}_t' = (z_{1t}, z_{2t}, \ldots, z_{kt})'$ is a k-dimensional vector of exogenous variables, and $l = (p+1) + k$. Furthermore, the random term $\varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$ and $\boldsymbol{\vartheta}$ is a Q-dimensional vector of real parameters. It is assumed that y_t is a stationary process. In general, few "a priori" assumptions can be made about the functional form of $f\left(\cdot\right)$, for this reason it is necessary to construct an estimator for $f\left(\cdot\right)$ from a large class of functions $\mathcal{F}$ known to have good approximation properties.

In literature, there are different methods to obtain approximators for $f\left(\cdot\right)$. However, we have chosen FNN because, as we shall see following this chapter it is very flexible and has a good ability to approximate any arbitrary function.

2.2.1 The FNN Model

In the FNN model, as stated in the introduction, the hidden units (or activation functions) are organized in layers. The layer that contains the regressors (or input) is called **Input Layer**. The layer where the regressand (or output) of the network is located is called the **Output Layer**. The layers between the input and the output are called **Hidden Layers**. In general there can be more than one layer in the FNN model, due to the complexity of the network or the nature of the problem. In this dissertation we deal only with one hidden layer.

The mathematical structure of a representation of an arbitrary function as proposed in equations (2.1) with a single hidden layer network having one or more hidden units is characterized as follows: The class $\mathcal{N}$ of real valued functions using a single hidden layer feedforward neural network has the following form

$$\mathcal{N}\left(\mathbf{s}_t;\boldsymbol{\vartheta}\right) \equiv \left\{ f:\mathbb{R}^l \to \mathbb{R} \mid f(\mathbf{s}_t) = \omega_0 + \sum_{j=1}^{s} \omega_j \Psi\left(\boldsymbol{\gamma}_j' \mathbf{s}_t\right) \right\} \tag{2.2}$$

where ω_0 is the constant term in the output layer, the parameter ω_j corresponds to the parameters from the hidden to the output layer, the $l \times 1$ vector $\boldsymbol{\gamma}_j$ corresponds to the parameters from the input to the hidden layer; the $Q \times 1$ vector $\boldsymbol{\vartheta} = (\omega_0, \omega_1, \ldots, \omega_s, \boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_s')'$ collects all network parameters, where $Q = s(l+2)+1$ is the total number of parameters to estimate.

The function Ψ is called the activation function. The activation function plays an important role in a neural network framework, because it introduces nonlinearity. Since a composition of linear functions is again a linear function, FNN in equation (2.2), without introducing nonlinearity would not be able to perform nonlinear separation. The choice of nonlinear activation function has a key influence on the complexity and performance of FNN. In this dissertation we deal with a nonlinear activation function that belongs to the class of sigmoid functions. In general they are defined as follows.

Definition 2 (Sigmoid Function) *Let $x \in \mathbb{R}$, a function $\Psi : \mathbb{R} \to \mathbb{R}$ belongs to the class of Sigmoid function if*

$$\lim_{x \to +\infty} \Psi(x) = a \quad \textit{and} \quad \lim_{x \to -\infty} \Psi(x) = b \quad \textit{with} \quad (a \neq b)$$

and having the following properties

(a) $\Psi(x)$ *is a continuously differentiable function;*

(b) $\Psi'(x) = \dfrac{d\,\Psi(x)}{d\,x} > 0$*;*

(c) $\Psi'(x) \to 0$ *as* $x \to \pm\infty$*;*

(d) $\Psi'(x)$ *takes a global maximal value at unique point* $x = 0$*;*

(e) *A sigmoidal function has only one inflection point, preferably at* $x = 0$*;*

(f) *From* (c)*, function* Ψ *is monotonically nondecreasing, i.e. if* $x_1 < x_2$ *for each* x_1*,* $x_2 \in \mathbb{R} \Rightarrow \Psi(x_1) \leq \Psi(x_2)$*;*

Examples of sigmoid functions are the following:

Logistic Function

$$\Psi(x) = \frac{1}{1 + e^{-\beta x}} \tag{2.3}$$

Hyperbolic Tangent Function

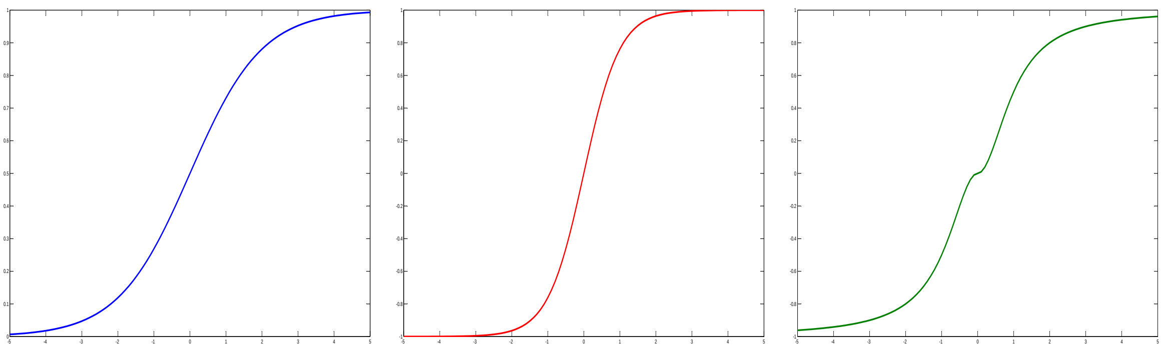
$$\Psi(x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}} \tag{2.4}$$

Sign Function

$$\Psi(x) = \frac{x^2}{1 + x^2} \operatorname{sign}(x) \tag{2.5}$$

A graphical representation of these functions are reported in Figure (2.1). In this dissertation we use the activation function reported in equation (2.4). Although sigmoid activation functions are the most common choice in FNN literature, there is no strong "a priori" justification as models based on this class of functions should be preferred over others. However thanks to universal approximation properties, based on the Stone-Weierstrass theorem, as we shall see in the next section, any sigmoid function is a suitable candidate for an activation function.

We will briefly discuss some necessary requirements for the sigmoid activation function, as reported in (Mandic and Chambers, 2001, pag. 51). First, as listed in the Definition (2), the property (a) is important for the estimation algorithm, which requires the existence of the Hessian matrix. The properties (b) and (c) require that the sigmoid function have a positive first derivative, which in turn means that the optimization algorithm have gradient vectors pointing towards the bottom of the bowl shaped error surface of the object function. Property (d) means that the point around which the first derivative is centered is the origin. This is connected with property (e) which means that the second derivative of the activation function should change its sign at the origin. Monotonicity, required by (f), is useful for uniform convergence of estimation algorithms.



(a) Logistic Function. (b) Hyperbolic Tangent Function. (c) Sign Function.

Figure 2.1: Different Representations of Activation Functions.

2.2.2 The Universal Approximation

The functional approximation capability of an FNN architecture is one of the most important properties and has potentials for application in different disciplines. This issue has been investigated by many authors. Here, the universal approximation capabilities of FNN are studied mainly using the well-known Stone-Weierstrass theorem, using the results obtained by Hornik et al. (1989).

To present the ability of FNN to approximate functions, it is important to choose a function space in which we operate and a metric space which is associated with the function space and used to measure the distance between two functions. In this dissertation, we consider the density property in the spaces of continuous functions $\mathcal{C}(K)$ endowed with the supremum norm and $\mathcal{L}_p(K)$, where K is a compact set of $\mathbb{R}^l$. When the normed linear space is the space of continuous functions with the supremum norm, in neural network terminology the density property is also called the *Universal Approximation Property.* More formally, a set of functions $\mathcal{F}$ can arbitrarily closely approximate a set of functions $\mathcal{G}$, in the sense of the metric $\mathcal{L}_p(K)$, if for any $g \in \mathcal{G}$ and to any positive ϵ there is an $f \in \mathcal{F}$ that is close to g.

Now we can introduce the Stone-Weierstrass theorem, which is the basis of our discussion.

Theorem 1 (Stone-Weierstrass) *Let domain K be a compact set with l dimensions, and let $\mathcal{F}$ be a set of continuous real-valued functions on K satisfying the following conditions:*

1. **Identity function**: *the constant function $f(\mathbf{x}) = \mathbf{I}$ belongs to $\mathcal{F}$;*

2. **Separability**: *for any two points* $\mathbf{x}_1 \neq \mathbf{x}_2 \in K$ *there exist a function* $f \in \mathcal{F}$ *such that* $f(\mathbf{x}_1) \neq f(\mathbf{x}_2)$*;*

3. **Algebraic Closure**: *for any* f*,* $g \in \mathcal{F}$ *the function* fg *and* $(\alpha f + \beta g)$ *are in* $\mathcal{F}$ *for any two real numbers* α *and* β*.*

Then $\mathcal{F}$ *is dense in* $\mathcal{C}(K)$*, the set of continuous real-valued functions on* K*. In others words, for any* $\epsilon > 0$ *and any function* $g \in \mathcal{C}(K)$*, there is a function* $f \in \mathcal{F}$ *such that*

$$\sup_{\mathbf{x} \in K} \left| f(\mathbf{x}) - g(\mathbf{x}) \right| < \epsilon.$$

Applying Theorem (1) to neural networks is not immediate but requires some explanation which will be discussed below. In order to establish the function approximation capabilities of FNN that are described by nonlinear mapping from input to output space directly using the Stone-Weierstrass theorem, one has to verify that it satisfies the following three conditions:

1. The ability of the approximating network to generate $f(\mathbf{x}) = \mathbf{I}$. This is always satisfied in feedforward neural networks that use the constant parameters.

2. The second condition which requires the separability of the function is satisfied since the activation functions of the neural networks are strictly monotonic. In fact, the neural networks generate different outputs for different inputs.

3. The algebraic closure condition requires that the nonlinear mappings of the neural networks are able to generate sums and products of functions. This condition is more difficult to satisfy and is further discussed below.

If an FNN spans a function space that satisfies the conditions of the Stone-Weierstrass theorem, it can be simply concluded that this network structure has the capability, on a compact set, to approximate arbitrary continuous real-valued functions to any desired degree of accuracy. However not all networks verify the universal approximation property. This depends on the activation function used. A typical example of this group are the sigmoid activation functions introduced in Definition (2), where the multiplication condition is not satisfied. However the universal approximation capabilities of this network structure can be ensured[1].

[1] The purpose of this section is not to give a rigorous proof of the ability of FNN as universal approximators, but simply give an intuition of how it is obtained.

Hornik et al. (1989), proposed an intuitive approach to indirectly prove the denseness of the space spanned by FNN with sigmoidal activation functions in continuous function space. The first step shows that a single-variable cosine activation function can be uniformly approximated by a single input, FNN with a sigmoidal function. In the second step, they prove that the arbitrary cosine network can be uniformly approximated by an FNN with a sigmoidal activation function. Finally, the denseness of the space spanned by the cosine network implies the denseness of the space of the FNN with sigmoidal functions. More formally the result that they obtained can be expressed by the following theorem

Theorem 2 (Universal Approximation) *Let $\mathcal{N}$ be the class of feedforward neural network functions $\mathcal{N} \equiv \left\{ f : \mathbb{R}^l \to \mathbb{R} \mid f(\mathbf{x}) = \omega_0 + \sum_{j=1}^{s} \omega_j \Psi\left(\boldsymbol{\gamma}_j' \mathbf{x}\right) \right\}$, where Ψ is any sigmoid activation function. Then $\mathcal{N}$ is uniformly dense on $\mathcal{C}(K)$, that is for every $g \in \mathcal{C}(K)$ and every $\epsilon > 0$, there exist $f \in \mathcal{N}$ such that*

$$\sup_{\mathbf{x} \in K} \left| f(\mathbf{x}) - g(\mathbf{x}) \right| < \epsilon.$$

These results establish FNN with sigmoid activation function as a class of universal approximators. Moreover, any lack of success in the application of an FNN that is a universal approximator must arise from inadequate estimation phase, an insufficient number of hidden units, or the lack of a deterministic relationship between the regressors and the regressand.

However, the proof given by Hornik et al. (1989) is merely "existential": they neither provide an algorithm to construct a model nor estimate the number of hidden units necessary to guarantee the desired approximation accuracy.

2.3 The FNN-DF Model

The use of a nonlinear forecasting factor model is new in the macroeconometric literature. However because of the complexity of implementation of both factor models and nonlinear models, this appears to be, from a practical point of view, difficult. One objective of this dissertation is to make construction the forecasting model FNN-DF as coherent as possible. In this section we introduce the FNN-DF model. We denote by $\mathbf{x}_t = (x_{1t}\ x_{2t}\ \cdots\ x_{nt})'$ the n-dimensional vector processes of observed data. Following the GDFM representation, $\mathbf{x}_t$ is represented as the sum of two components: the common component and the idiosyncratic component, namely we have

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \mathbf{B}(L)\mathbf{u}_t + \boldsymbol{\xi}_t \tag{2.6}$$

Where $\mathbf{u}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is a q-dimensional orthonormal white noise process, where u_{jt} has unit variance and is orthogonal to u_{st} for any $j \neq s$; $\mathbf{B}(L)$ is a $n \times q$ polynomial of order m in the lag operator L. Regarding to the idiosyncratic components we assume that $\boldsymbol{\xi}_t$ is orthogonal to all components of $\mathbf{u}_t$. Note that for a given finite lag order m, the model in equation (2.6) can be written in the following form

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{2.7}$$

where $\mathbf{F}_t = (u_{1t}\ u_{2t}\ \cdots\ u_{qt})'$ is a $r = q(m+1)$ dimensional vector of stacked dynamic factors and $\boldsymbol{\Lambda}$ is an $n \times r$ dimensional parameter matrix which contains the coefficients of $\mathbf{B}(L)$.
Let $y_t \in \mathbf{x}_t$ be the variable of interest we want to predict, then making use of the factors $\mathbf{F}_t$ as regressors in equation (2.2), we have the FNN-DF

$$y_t = \omega_0 + \sum_{j=1}^{s} \omega_j \Psi\left(\boldsymbol{\gamma}_j'\mathbf{s}_t\right) + \varepsilon_t \tag{2.8}$$

where now $\mathbf{s}_t = (\mathbf{w}_t', \mathbf{F}_t')'$ is composed by $\mathbf{w}_t'$ a $(p+1)$-dimensional vector of the lags of the variable of interest as in equation (2.2), and $\mathbf{F}_t'$ a r-dimensional vector of factor; the number of variables used in the FNN-DF model becomes $l = (p+1) + r$. Furthermore, the random term $\varepsilon_t \sim \text{iid}\left(0, \sigma^2\right)$. The $Q \times 1$ vector $\boldsymbol{\vartheta} = (\omega_0, \omega_1, \ldots, \omega_s, \boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_s')'$ collects all network parameters, where $Q = s(l+1) + 1$ is the total number of parameters to estimate.

2.4 Building Procedure for FNN-DF

The construction of the FNN-DF requires two stages. The first concerns the estimation phase of dynamic factors (DF), while the second the construction of the FNN. Lets first consider the estimation phase of the DF: since the factors are unobservable components, it is necessary for these to be estimated in order to be used during the forecasting phase. As described in Chapter 1, in this dissertation we focus on the techniques proposed by Stock and Watson and

Forni et al. Moreover, from a practical point of view, it is important to know the optimum number of factors necessary to summarize the information contained in the dataset. To do this, we use the techniques proposed by Bai and Ng (2002) to choose the optimum number of r static factors in the model (2.7) and the one proposed by Hallin and Liska (2007) to choose the optimum number of dynamic factors q in the model (2.6).

The construction of the FNN requires two basic steps. First, we need a technique to estimate the parameters. In our case, we use the method of nonlinear least squares proposed by Hagan and Menhaj (1994). Second, we need a procedure to select the number of hidden units (the architecture) of FNN. The critical issue in developing an FNN is to select the optimum number of hidden units. Since the number of hidden units measures the complexity of FNN we may have an FNN not sufficiently complex, which may fail to detect the underlying process in a dataset, leading to underfitting. An FNN that is too complex may look not just at the underlying process, but also at the noise, leading to overfitting. The method we adopt to solve this issue is the Bayesian Regularization proposed by Mackay (1992a) and Foresee and Hagan (1997).

Finally, since in the FNN-DF model the lags of the variable to predict are also included, the optimum number, p, of lags will be selected using techniques such as AIC or BIC.

2.4.1 Estimation of the Alternative Factor Models

The technique used to estimate the factors affects the precision of the estimates. As pointed in Chapter 1, basically two different methods are employed in the literature to estimate factors with large dataset, namely those proposed by Stock and Watson (2002a) and Forni et al. (2005).

Estimating the Factors according to Stock and Watson (2002a). Stock and Watson propose estimating $\mathbf{F}_t$ with static principal component analysis applied to $\mathbf{x}_t$. The factor estimates are simply the first r principal components of $\mathbf{x}_t$, as showed in Section (1.3). They are defined as $\hat{\mathbf{F}}_t^{\mathrm{SW}} = \boldsymbol{\alpha}'\mathbf{x}_t$ where $\boldsymbol{\alpha}$ is the $n \times r$ matrix of the eigenvectors corresponding to the r largest eigenvalues of the estimate sample covariance matrix $\check{\Gamma}_0^x$ of $\mathbf{x}_t$.

Estimating the Factors according to Forni et al. (2005). The estimator proposed by Stock and Watson (2002a), being based on contemporaneous covariances only, fails to exploit the dynamic relations between the variables of the panel. Forni et al. propose a weighted version of the principal component estimator proposed by Stock and Watson, where time series are weighted according to their signal-to-noise ratio, which is estimated in the frequency domain. As shown in Section (1.4) the authors proceed in two steps. In the first step, the covariance matrices of common and idiosyncratic components of $\mathbf{x}_t$ are estimated. This involves estimating the spectral density matrix $\mathbf{\Sigma}^x(\theta)$ of $\mathbf{x}_t$. Since $\check{\mathbf{\Sigma}}^x(\theta) = \check{\mathbf{\Sigma}}^\chi(\theta) + \check{\mathbf{\Sigma}}^\xi(\theta)$, we obtain the estimates of the spectral density matrix of the common and the idiosyncratic component respectively. Inverse Fourier transformation provides the time domain autocovariances of the common and the idiosyncratic components $\check{\mathbf{\Gamma}}_k^\chi$ and $\check{\mathbf{\Gamma}}_k^\xi$ respectively. In the second step, the authors search for the r linear combinations of $\mathbf{x}_t$ that maximize the contemporaneous covariance explained by the common factors which can be formulated as the generalized eigenvalue problem, $\check{\mathbf{Z}}_j\check{\mathbf{\Gamma}}_0^\chi = \check{v}_j\check{\mathbf{Z}}_j\check{\mathbf{\Gamma}}_0^\xi$, where $\check{\mathbf{Z}}_j$ are the generalized eigenvectors associated with the generalized eigenvalues $\check{v}_j$ and the factor estimates are obtained as $\hat{\mathbf{F}}_t^{\mathrm{FHLR}} = \check{\mathbf{Z}}'\mathbf{x}_t$.

2.4.2 Parameter Estimation

Several optimization methods have been developed for the estimation of parameters in FNN. In particular, two classes of algorithms are widely used in literature. On the one side we have the backpropagation technique proposed by Rumelhart et al. (1988) and its variants such as backpropagation with momentum term and learning rate. On the other side numerical optimization techniques such as conjugate gradient methods or quasi-Newton have been used. Backpropagation is a recursive estimation technique in which the parameters are updated in the opposite direction of the gradient of the objective function to minimize. In each step the gradient contributes to the reduction of the error until the minimum is reached. The computational complexity of backpropagation is mainly due to the calculation of first-order partial derivatives and is of the order of $O(Q)$. However, using the first-order partial derivatives, the backpropagation algorithm appears to only be linearly convergent and therefore slow in its convergence. Some improvements are obtained by introducing the momentum term and the learning rate. In general the use of these two techniques reduces the possibility of the FNN getting stuck in a local minimum of the objective function, effectively reducing the convergence time. However, the momentum term and the learning rate are free parameters and should be carefully selected, which is not always an easy task. An alternative way to speed up the estimation phase is by using higher order optimization methods that utilize the second-order partial derivatives such as Gauss-Newton methods or conjugate gradient.

In this dissertation we make use of the Marquardt-Levenberg algorithm. This algorithm is an iterative technique that locates the minimum of an objective function, expressed as the sum of squares of non-linear real-valued functions. It has become a standard technique for nonlinear least-square problems. The Marquardt-Levenberg can be thought of as a combination of gradient descent and the Gauss-Newton method. When the current solution is far from the correct one, the algorithm behaves as a gradient descent method, which converges everywhere, albeit slowly. When the current solution is close to the correct one, it becomes a Gauss-Newton method, which converges very quickly. A comparison study is reported in Hagan and Menhaj (1994) where the Marquardt-Levenberg method significantly outperforms the conjugate gradient and the backpropagation methods with momentum term and learning rate, in terms of convergence time and accuracy. Whereas the other algorithms are designed to work with a wide range of objective functions, the Marquardt-Levenberg is designed specifically to minimize a particular objective function said sum of squares error.
In order to describe the Marquardt-Levenberg algorithm we define the objective function as

$$\widehat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}}\ E_T\left(\boldsymbol{\vartheta}\right) = \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}} \sum_{t=1}^{T} \Big(y_t - \mathcal{N}\left(\mathbf{s}_t;\boldsymbol{\vartheta}\right)\Big)^2 \tag{2.9}$$

where equation (2.9) represents the sum of squares error. Given a vector $\boldsymbol{\vartheta}_0$, called nominal point, where the error function has a local minimum, a second-order Taylor series approximation of the error function described in equation (2.9) around this vector is expressed as

$$E_T\left(\boldsymbol{\vartheta}\right) = E_T\left(\boldsymbol{\vartheta}_0\right) + \mathbf{g}'\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) + \frac{1}{2}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)'\mathbf{H}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) \tag{2.10}$$

where $\mathbf{g}$ and $\mathbf{H}$ are the gradient vector and the Hessian matrix, respectively. The minimums of the function E_T are located where the gradient of E_T expressed by equation (2.10) is zero:

$$\frac{\partial E}{\partial \boldsymbol{\vartheta}} = \mathbf{g} + \mathbf{H}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) = 0. \tag{2.11}$$

Therefore, the optimal value of $\boldsymbol{\vartheta}$ is given by

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 - \mathbf{H}^{-1}\,\mathbf{g} \tag{2.12}$$

equation (2.12) is a basic formulation for second-order optimization methods. The key issue related to the second-order methods is computing the inversion of the Hessian matrix $\mathbf{H}$. In fact the rank of $\mathbf{H}$ is equal to $O(Q^2)$. As the number of parameters of the network increases, the demand for memory to work with such large matrix increases exponentially. The method proposed by Marquardt-Levenberg is to approximate the Hessian that can be written in this form $\mathbf{H} = \mathbf{J}'\mathbf{J} + \eta\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{J}$ is also called Jacobian, whose elements can be calculated directly using the first order partial derivatives. Then the updating rule for the parameters vector $\boldsymbol{\vartheta}$ is

$$\boldsymbol{\vartheta}^{(k+1)} = \boldsymbol{\vartheta}^{(k)} + \left(\mathbf{J}'\mathbf{J} + \eta^{(k)}\mathbf{I}\right)^{-1} \mathbf{g}^{(k)}. \tag{2.13}$$

The matrix $\mathbf{J}'\mathbf{J}$ is symmetric and defined nonnegative; hence, any positive η will ensure that $(\mathbf{J}'\mathbf{J} + \eta\mathbf{I})^{-1}$ is defined positive, as required by gradient algorithm. In practice a value must be chosen for η and this value should vary appropriately during the minimization process. One common approach for setting η is to use the method proposed by Hagan and Menhaj (1994). They propose to beginning with $\eta = 0.1$, and at each step if $E_T^{(k+1)} > E_T^{(k)}$, η is increased by a factor of 10, the old parameters vector is restored, and a new parameters update computed. This is repeated until a decrease in E_T is obtained. If, however, $E_T^{(k+1)} < E_T^{(k)}$ after taking the step described by equation (2.13) the new parameter vector is retained, the value of η is decreased by a factor of 10, and the process repeated.

2.4.3 Determining the Number of Hidden Units

One of the most serious problems that arises in estimation of FNN is overfitting. This means that the estimated function fits the presented data very closely however it does not generalize well, that is, it does not yield the most accurate forecast possible. This problem is also known as Bias-Variance dilemma. In the context of neural network bias measures how well a model estimates the process underlaying the data. This accounts only for the accuracy of the estimation process, but not for the level of generalization. Variance measures the deviation of the accuracy of an estimation process from one sample to another sample generated by the same process, without regard to the specifics of the provided data. To avoid overfitting, we have to use some criteria that, during the estimation process, enable us to balance the statistical bias and variance in order to achieve the smallest possible generalization error. In literature, there are several criteria to avoid the problem of overfitting. The usual approaches pursued in the network literature are pruning, early stopping, and regularization.

The aim of the pruning methods is to identify those parameters which do not contribute to the overall network performance. However, identifying these parameters is not usually judged on the basis of statistical test. Instead, pruning methods use so-called saliency as a measure of a parameter's importance. The saliency of a parameter is defined as the increase in network model error incurred by setting this parameter to zero. The idea is to remove the parameters with low saliency; however, the method does not provide any guidance on how to judge saliency as low.

In the application of early stopping the dataset is split into an estimation set and a validation set. If the errors in the validation set grow too much during the estimation process, the procedure is stopped. In statistical terms, the method tries to make up for the model being over parameterized by stopping the estimation algorithm before the minimum of the network error function is reached. In general this does not lead to reasonable estimates of the network parameters. Instead, the growing errors in the validation set should be seen as an indication to reduce the network's complexity.

The method we use in this dissertation is the Bayesian Regularization, see Mackay (1992a) and Foresee and Hagan (1997). The Bayesian Regularization tries to find a balance between the number of parameters and goodness-of-fit by penalizing large models. The objective function is modified in such a way that the estimation algorithm effectively prunes the network by driving irrelevant parameters to zero during the estimation process.
The parameter vector $\boldsymbol{\vartheta}$ is estimated as

$$\widehat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}}\, F\left(\boldsymbol{\vartheta}\right) = \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}}\, \zeta E_T\left(\boldsymbol{\vartheta}\right) + \kappa E_{\vartheta}\left(\boldsymbol{\vartheta}\right) \tag{2.14}$$

where

$$E_T\left(\boldsymbol{\vartheta}\right) = \sum_{t=1}^{T} \big(y_t - \mathcal{N}\left(\mathbf{s}_t; \boldsymbol{\vartheta}\right)\big)^2 \quad \text{and} \quad E_{\vartheta}\left(\boldsymbol{\vartheta}\right) = \sum_{j=0}^{s} \omega_j^2 + \sum_{j=1}^{s} \sum_{h=0}^{l} \gamma_{jh}^2,$$

whereas ζ and κ are scalar objective function parameters. These parameters are very important because their relative size dictates the result of forecast. Indeed, if $\kappa \ll \zeta$, then the optimization algorithm will make the errors small, this means we may have a large variance. If $\kappa \gg \zeta$, the optimization algorithm will emphasize parameter size reduction producing a large errors network, this means we may have large bias. In both cases the resulting forecast will be very inaccurate. The approach we use to optimally determine the regularization parameters ζ and κ is the Bayesian framework reported in Mackay (1992a). The steps required for Bayesian optimization of the regularization parameters in conjunction with the Marquardt-Levenberg algorithm as in Foresee and Hagan (1997) are as follows.

In the Bayesian framework, the parameters of the FNN are considered random variables. Let $\mathbf{D}_t = (y_t, \mathbf{s}_t)$ represent the dataset and $\mathcal{N}$ a particular FNN model. After the data are collected, the distribution function for the parameters is updated according to Bayes' rule

$$\mathrm{P}\left(\boldsymbol{\vartheta}|\mathbf{D}_t,\zeta,\kappa,\mathcal{N}\right) = \frac{\mathrm{P}\left(\mathbf{D}_t|\boldsymbol{\vartheta},\zeta,\kappa,\mathcal{N}\right)\,\mathrm{P}\left(\boldsymbol{\vartheta}|\kappa,\mathcal{N}\right)}{\mathrm{P}\left(\mathbf{D}_t|\zeta,\kappa,\mathcal{N}\right)}, \tag{2.15}$$

where $\mathrm{P}\left(\boldsymbol{\vartheta}|\kappa,\mathcal{N}\right)$ is the prior distribution, which represent our knowledge of the parameters before any data is collected, and $\mathrm{P}\left(\mathbf{D}_t|\boldsymbol{\vartheta},\zeta,\kappa,\mathcal{N}\right)$ is the likelihood function, which is the probability of the data occurring given the parameters. $\mathrm{P}\left(\mathbf{D}_t|\zeta,\kappa,\mathcal{N}\right)$ is a normalization factor, which guarantees that the total probability is equal to one. If we assume that the noise and prior distribution for the parameters are both Gaussian then we have

$$\mathrm{P}\left(\boldsymbol{\vartheta}|\kappa,\mathcal{N}\right) = \left(\frac{\pi}{\zeta}\right)^{-T/2} \exp\left(-\zeta E_T\right) \tag{2.16}$$

and

$$\mathrm{P}\left(\mathbf{D}_t|\boldsymbol{\vartheta},\zeta,\kappa,\mathcal{N}\right) = \left(\frac{\pi}{\kappa}\right)^{-Q/2} \exp\left(-\kappa E_{\vartheta}\right) \tag{2.17}$$

Once a prior distribution is chosen for the parameters and for the likelihood, using equation (2.15) we obtain the posterior distribution for the parameters in the form

$$\mathrm{P}\left(\boldsymbol{\vartheta}|\mathbf{D}_t,\zeta,\kappa,\mathcal{N}\right) = \frac{1}{Z_F} \exp\left(-\zeta E_T - \kappa E_{\vartheta}\right) \tag{2.18}$$

where $Z_F = \int \exp\left(-\zeta E_T - \kappa E_{\vartheta}\right) \mathrm{d}\boldsymbol{\vartheta}$. In the Bayesian framework, the optimal parameters maximize the posterior probability. Since the normalization factor is independent of the parameters, maximizing the posterior probability is equivalent to minimizing the objective function in equation (2.14).

The regularization parameters are optimized by applying the Bayes' rule

$$\mathrm{P}\left(\zeta,\kappa|\mathbf{D}_t,\mathcal{N}\right)=\frac{\mathrm{P}\left(\mathbf{D}_t|\boldsymbol{\vartheta},\zeta,\mathcal{N}\right)\,\mathrm{P}\left(\zeta,\kappa|\mathcal{N}\right)}{\left(\mathbf{D}_t|\mathcal{N}\right)}. \tag{2.19}$$

Assuming an uniform prior distribution $\mathrm{P}\left(\zeta,\kappa|\mathcal{N}\right)$ for the regularization parameters, then maximizing the posterior is achieved by maximizing the likelihood function $\mathrm{P}\left(\mathbf{D}_t|\boldsymbol{\vartheta},\zeta,\mathcal{N}\right)$. Since all probabilities have a Gaussian form, the normalization factor can be expressed as

$$\mathrm{P}\left(\mathbf{D}_t|\boldsymbol{\vartheta},\zeta,\kappa,\mathcal{N}\right)=\left(\frac{\pi}{\zeta}\right)^{-T/2}\left(\frac{\pi}{\kappa}\right)^{-Q/2}Z_F^{-1}. \tag{2.20}$$

Assuming that the objective function has a quadratic shape in a small area surrounding a minimum point, we can expand $F\left(\boldsymbol{\vartheta}\right)$ in a Taylor series around the minimum point of the posterior density, where the gradient is zero. Solving equation (2.20) for the normalization factor yields

$$Z_F=\left(2\pi\right)^{Q/2}\left[\det\left(\mathbf{H}^{-1}\right)\right]^{1/2}\exp\left(-E_\vartheta\left(\boldsymbol{\vartheta}\right)\right), \tag{2.21}$$

where $\mathbf{H}$ is the Hessian matrix of the objective function $F\left(\boldsymbol{\vartheta}\right)$. Substituting equation (2.21) into equation (2.20), we can solve for the optimal value of ζ and κ at the minimum point. This is done by taking the derivative with respect to the log of equation (2.20) and setting it equal to zero. This yields

$$\hat{\zeta}=\frac{\varrho}{2E_\vartheta\left(\boldsymbol{\vartheta}\right)}\quad\text{and}\quad\hat{\kappa}=\frac{T-\varrho}{2E_T\left(\boldsymbol{\vartheta}\right)}$$

where $\varrho=Q-2\zeta\mathrm{tr}\left(\mathbf{H}\right)^{-1}$ is called the effective number of parameters.

The Bayesian optimization of the regularization parameters requires the computation of the Hessian matrix of $F\left(\boldsymbol{\vartheta}\right)$. Foresee and Hagan (1997) proposed using the Gauss-Newton approximation to the Hessian matrix, which is available if the Marquardt-Levenberg optimization algorithm is used to locate the minimum point, as showed in Section (2.4.2). Here are reported the steps required for Bayesian optimization of the regularization parameters in conjunction with the Marquardt-Levenberg optimization algorithm:

Bayesian Regularization Procedure

1. Initialize ζ and κ and the parameters. We set $\zeta=1$ and $\kappa=0$ and use the Nguyen and Widrow (1990) method to initialize the parameters;

2. Use the Marquardt-Levenberg optimization algorithm to minimize the objective function $F(\boldsymbol{\vartheta})$ in (2.14) and the parameters vector $\boldsymbol{\vartheta}$;

3. Compute the effective number of parameters $\varrho = Q - 2\zeta \mathrm{tr}\,(\mathbf{H})^{-1}$, where $\mathbf{H}$ is the Hessian computed by the Marquardt-Levenberg optimization algorithm;

4. Compute new estimates for the regularization parameters $\hat{\zeta}$ and $\hat{\kappa}$;

5. Iterate steps 2 through 4 until convergence.

2.4.4 Forecasting with FNN-DF

Forecasting with nonlinear models such as FNN-DF for more than one period ahead can be achieved by using the iterated or the direct approach. Regarding the iterated approach, analytically obtaining the forecasts is not always possible and often are required numerical techniques.

In this regard, suppose the FNN-DF is correctly specified and $\mathbf{s}_t = \mathbf{A}\mathbf{s}_{t-1} + \boldsymbol{\eta}_t$ follows a first order vector autoregressive representation, where $\mathbf{A}$ is a $l \times l$ vector of parameters and $\boldsymbol{\eta}_t \sim \mathrm{iid}\,(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ the $l \times 1$ vector of error terms. The one step ahead forecast for y_{t+1} is equal to $\hat{y}_{t+1|t} = E\,(y_{t+1}|\mathbf{s}_t) = n(\mathbf{s}_t; \hat{\boldsymbol{\vartheta}})$, where the vector parameters $\hat{\boldsymbol{\vartheta}}$ are estimated up to time t. Forecasting two or more periods ahead is much more complicated. Indeed since we do not know the value for $\mathbf{s}_{t+1}$ we need to compute a forecast of this. Suppose we can forecast $\mathbf{s}_{t+1} = \mathbf{A}\mathbf{s}_t + \boldsymbol{\eta}_{t+1}$, then the two steps ahead is given by $\hat{y}_{t+2|t} = E\,(y_{t+2}|\mathbf{s}_t) = n(\mathbf{A}\mathbf{s}_t + \boldsymbol{\eta}_{t+1}; \hat{\boldsymbol{\vartheta}}) = \int_{\eta_1} \cdots \int_{\eta_l} n(\mathbf{A}\mathbf{s}_t + \boldsymbol{\eta}_{t+1}; \hat{\boldsymbol{\vartheta}}) dF(\eta_1, \ldots, \eta_l)$, where $F(\eta_1, \ldots, \eta_l)$ is the joint cumulative distribution function of $\boldsymbol{\eta}_t$. This means that for periods grater than one, we have to solve a multidimensional integral. In Terasvirta (2006) paper two numerical techniques have been proposed in order to avoid analytical integration: the simulation and the bootstrap technique; these techniques could be computationally demanding, especially when using nonlinear models. Moreover, in empirical applications, very often the model assumed for the observations is not the true data generating process. This misspecification can lead to less than optimal predictions.

This situation is also reported in (Bhansali, 2002, pages 206-221), which has shown the direct approach is preferable to the iterated approach either computationally or as forecast performance, when the model for the observations is misspecified.

Therefore, in this dissertation we adopt the direct approach. The forecasting equation is given by the projection of h steps ahead of y_t on the t-dated predictors. Then assuming as the data generating process the FNN-DF in equation (2.8)

$$y_{t+h} = \omega_0 + \sum_{j=1}^{s} \omega_j \Psi \left(\boldsymbol{\gamma}_j' \mathbf{s}_t \right) + \varepsilon_{t+h}. \tag{2.22}$$

Here, as already mentioned, we prefer the direct method as it has the advantage that no numerical generation of forecasts is necessary. A disadvantage is that a separate model has to be specified and estimated for each forecast horizon.

2.4.5 Forecasting Evaluation

There are many ways to compare the forecasting performance of a model, ranging from magnitude measures to directional measures. Regarding the magnitude measures, we use two statistics commonly adopted in literature. They are the mean squared error and mean absolute error. For an h-steps ahead forecast, they are defined as follows

(C1) Mean Squared Error (MSE)

$$MSE = \frac{1}{T_1 - h - T_0 - 1} \sum_{\tau=T_0}^{T_1-h} \left(\hat{y}_{\tau+h} - y_{\tau+h} \right)^2$$

(C2) Mean Absolute Error (MAE)

$$MAE = \frac{1}{T_1 - h - T_0 - 1} \sum_{\tau=T_0}^{T_1-h} \left| \hat{y}_{\tau+h} - y_{\tau+h} \right|$$

where $h = 1, 3, 6, 9, 12$, and T_0 is the first point in time for out of sample evaluation and T_1 is the last point in time. In addition to MSE and MAE, we use also the Directional Accuracy test (DA) proposed by Pesaran and Timmermann (1992). The DA test considers the future direction (up or down) implied by the model.

(C3) We compute the indicator for the correct sign of the model used in forecasting, i.e. $I_T = 1$ if $\hat{y}_{T+h} \cdot y_{T+h} > 0$ otherwise $I_T = 0$. Once computed the indicator for the correct sign, we calculate the success ratio (SR)

$$\text{DA} = \frac{1}{T_1 - h - T_0 - 1} \sum_{\tau=T_0}^{T_1-h} I_T$$

SR indicates the percentage in predicting the correct sign for the proposed model. Finally, we use the test proposed by Pesaran and Timmermann (1992) to verify if the model outperforms the chance of random choice.

2.5 Montecarlo Experiments

In this section we apply the FNN-DF model presented in Section (2.3) to data generated artificially. The motivation is to show the potential of the FNN-DF approach, relative to linear dynamic factor models, in predicting relatively complex stochastic processes.

Unlike the literature regarding GDFM, we are not interested in verifying the properties of the FNN-DF model when the panel of simulated data differs in the cross and time dimensions, in the number of dynamic and static factors, or in the amount of variance explained by the common part with respect to the total. The goal here is to determine whether the proposed model is able to identify the function that exists between the static and dynamic factors as well as the variable of interest. In this regard two processes are simulated: the Markov regime switching factor model and the stochastic chaos factor model. The panel of data from which the factors are extracted is common to both.

Data Generating Process. DGP. We consider the following data-generating process, common to both models

$$x_{it} = \sum_{k=0}^{2} b_{ik}\, u_{1,t-k} + \xi_{it} \tag{2.23}$$

where u_{1t} is an univariate first-order AR process with AR(1) parameter $\alpha = 0.6$, namely $u_{1t} = \alpha\, u_{1,t-1} + v_t$ and $\xi_{it} = \iota\, \epsilon_{it}$. The parameters b_{ik}, with $i = 1, \ldots, n$ and $k = 0, 1, 2$, the shocks u_{1t} and ϵ_{it} with $t = 1, \ldots, T$ are standard normal variables. The constant ι is chosen

$$\iota = \frac{1-\varsigma}{\varsigma} \left(\frac{\sum_{i=1}^{N} \mathrm{var}(b_i u_{1t})}{\sum_{i=1}^{N} \mathrm{var}(\epsilon_{it})} \right)$$

where $\varsigma = 0.5$, so that on average, 50% of the variation in x_{it} is explained by the common component. Following Forni et al. (2005), since $\epsilon \sim N(0, \sigma_i^2)$, we use $\sigma_i^2 \sim U(0.1, 1.1)$. This means that even though $\varsigma = 0.5$ on average, there is a good deal of variation in the size of the common component. Moreover as seen in Section (1.3), the simulated GDFM in (2.23) can be written in its static form as a r-factor model where $r = q(m+1)$, namely

$$x_{it} = \sum_{j=1}^{3} \Lambda_{ij} F_{jt} + \xi_{it},$$

where $F_{1t} = u_{1t}$, $F_{2t} = u_{1,t-1}$ and $F_{3t} = u_{1,t-2}$, whereas $\Lambda_{i1} = b_{i0}$, $\Lambda_{i2} = b_{i1}$ and $\Lambda_{i3} = b_{i2}$.

M1. Markov regime switching model. The variable of interest to be forecast is generated according to the Markov Switching model (MRS)

$$y_t = \begin{cases} c_1 + \sum_{j=1}^{r} \phi_{1j} F_{jt} + e_{1t}, & \text{if } S_t = 1 \\ c_2 + \sum_{j=1}^{r} \phi_{2j} F_{jt} + e_{2t}, & \text{if } S_t = 2 \end{cases} \tag{M1}$$

where S_t assumes values in $\{1, 2\}$ and is a first-order Markov chain with transition probabilities

$$\mathbf{P} = \begin{bmatrix} (S_T = 1|S_{t-1} = 1) & (S_T = 1|S_{t-1} = 2) \\ (S_T = 2|S_{t-1} = 1) & (S_T = 2|S_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} (1 - w_2) & w_2 \\ w_1 & (1 - w_1) \end{bmatrix}$$

The error terms e_{1t} and e_{2t} are sequences of iid random variables with mean zero, finite variance, and are independent of each other. A small w_i means that the model tends to stay longer in state i. In fact, $1/w_i$ is the expected duration of the process in State i. In simulation of the model, we use the following parameters:

$$\begin{cases} c_1 = 1,\ \phi_{11} = 1.5,\ \phi_{12} = 0.7,\ \phi_{13} = 0.05,\ w_1 = 0.3 & \text{if } S_t = 1 \\ c_2 = -0.5,\ \phi_{21} = 2,\ \phi_{22} = 0.9,\ \phi_{23} = -0.15,\ w_2 = 0.5 & \text{if } S_t = 1. \end{cases}$$

and the transition probability matrix $\mathbf{P}$ is generated from a uniform distribution.

M2. Nonlinear Factor Model. The variable of interest to be forecast is generated according to Nonlinear Factor model (NF)

$$y_t = \alpha\, \upsilon_t\, F_{1t}\, (1 + F_{1t}) + \sum_{j=2}^{r} \beta_j F_{jt} \tag{M2}$$

where the error term υ_t is generated according to a uniform distribution, namely $\upsilon_t \sim U(0, 1)$. When the time series y_t is characterized by periods of high volatility followed by flat stable intervals, the presence of nonlinear events could be considered.

2.5.1 Montecarlo Experiment Results

We simulate 350 observations of the variables and discard the first 50, leaving $T = 300$ observations for evaluation, while for the cross section we use $n = 150$. The first estimation uses data from $t = 1, \ldots, 120$ to perform a one-step ahead forecast, namely, $T + h = 121$. Then T is incremented by 1, the estimation is repeated using data from $t = 2, \ldots, 121$, and a forecast for $T+h = 122$ is performed. The last forecast of $T+h = 300$ is based on estimation using data up to $t = 179, \ldots, 299$. The same procedure is applied for $h = 3, 6, 9, 12$ and the whole experiment repeated 500 times. Each forecast $\hat{y}_{T+h}$ is then compared to y_{T+h}. Regarding the criterion (C1), in Table (2.1) we refer to the ratio of the MSE for a given method to the MSE of GDFM; a relative MSE bigger than one means that the GDFM outperforms the method considered. For evaluation criterion (C2), Table (2.1) report the ratio of the MAE for each given method. A relative MAE greater than one means that the GDFM outperforms the method considered. The final criterion used (C3) is the Directional Accuracy (DA) test where the ratio of correct prediction is reported. GDFM estimated using the generalized principal component technique is noted $\mathrm{GDFM}^{\mathrm{G}}$ in Table (2.1), whereas GDFM estimated using the standard principal component technique is reported $\mathrm{GDFM}^{\mathrm{S}}$

The results for the model (**M1**) are shown in Table (2.1)-Panel A. The rows marked GDFM are the benchmarks and refer to linear prediction as in equation (1.6) using the static and dynamic factors as regressors described in Section (1.3). The rows marked FNN-DF refer to the nonlinear forecasting technique described in Section (2.3) using as regressors the static factors described in Section (2.3). For the FNN-DF models, the number of hidden units is fixed to $s = 10$, which is equivalent to a number of initial parameters $Q = 51$.

In Table (2.1)-Panel A there is no method that outperforms the benchmarks, whatever the criterion for evaluating forecasts used. This result is not surprising, since the nonlinearity is not driven by the factors themselves but by a latent process. Regarding the relationship that exists in its two states, namely between the variable of interest and the factors, is completely linear. What is interesting to observe is how the model FNN-DF is able to adjust its complexity depending on the problem under study. Indeed we observe that, for both nonlinear models as compared with $Q = 51$ initial parameters to estimate, the same number of parameters are used as in the linear model. This peculiarity is constant for the five forecasting horizons proposed.

Regarding the model (**M2**), it can be seen in the Table (2.1)-Panel B, nonlinear models appear to have the best performance compared to the linear benchmarks, whatever the criterion adopted. This indicates that the FNN-DF models respond according to changes in the functional form between the predictors and the variable of interest, however this is evident only if the variation is determined by variables that can be directly observed. Another observation concerns the length of forecasting horizon. In particular, as the horizon is more distant in time, more the FNN-DF loses strength to capture the nonlinearity, predicting only the unconditional mean of the process.

Table 2.1: Montecarlo Experiments results for the models proposed in (**M1**) and in (**M2**)

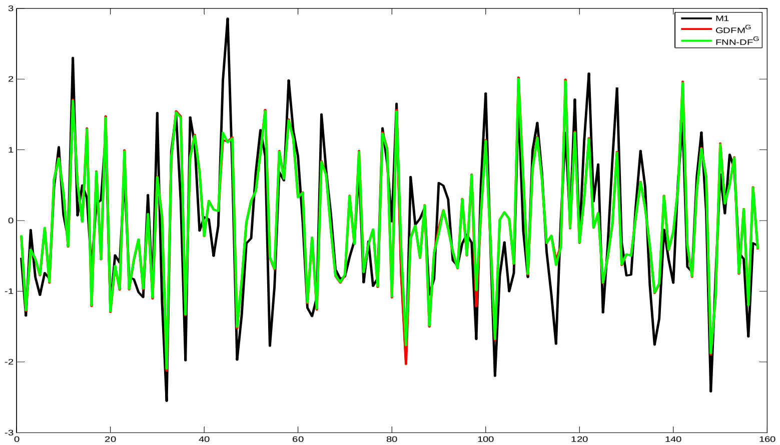
Panel A: Model M1

	H = 1			**H=3**			**H=6**			**H=9**			**H=12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
$GDFM^{G}$	1.000	1.000	0.828	1.000	1.000	0.735	1.000	1.000	0.768	1.000	1.000	0.801	1.000	1.000	0.775
$GDFM^{S}$	1.096	1.123	0.662	1.111	1.061	0.728	1.081	1.020	0.742	1.076	1.009	0.781	1.056	1.024	0.742
Nonlinear															
FNN-DF^{G}	1.004	1.008	0.828	1.028	1.012	0.728	0.956	0.970	0.762	1.020	1.017	0.801	1.002	1.001	0.775
FNN-DF^{S}	1.114	1.127	0.676	1.118	1.058	0.735	1.123	1.038	0.742	1.098	1.026	0.781	1.051	1.019	0.742

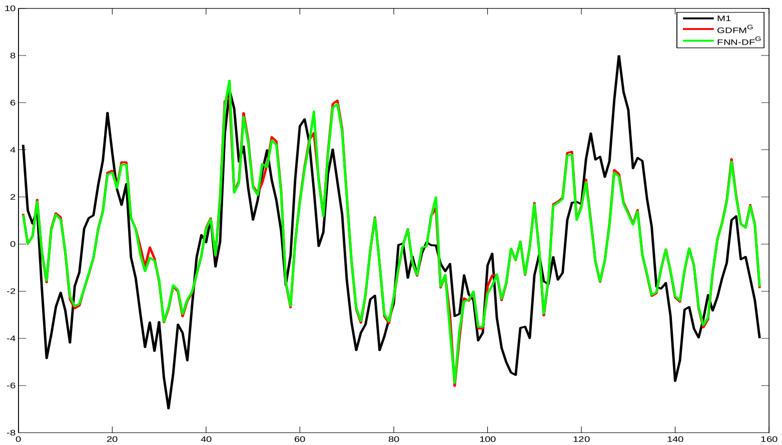
Panel B: Model M2

	H = 1			**H=3**			**H=6**			**H=9**			**H=12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
$GDFM^{G}$	1.000	1.000	0.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$GDFM^{S}$	1.085	1.060	0.981	1.010	1.033	1.000	0.959	1.016	1.000	0.969	1.003	1.000	0.963	0.993	1.000
Nonlinear															
FNN-DF^{G}	0.234	0.453	0.968	0.457	0.689	1.000	0.853	0.952	1.000	0.951	0.984	1.000	0.969	0.998	1.000
FNN-DF^{S}	0.723	0.862	0.975	0.752	0.869	1.000	0.855	0.985	1.000	0.949	0.997	1.000	0.936	0.983	1.000

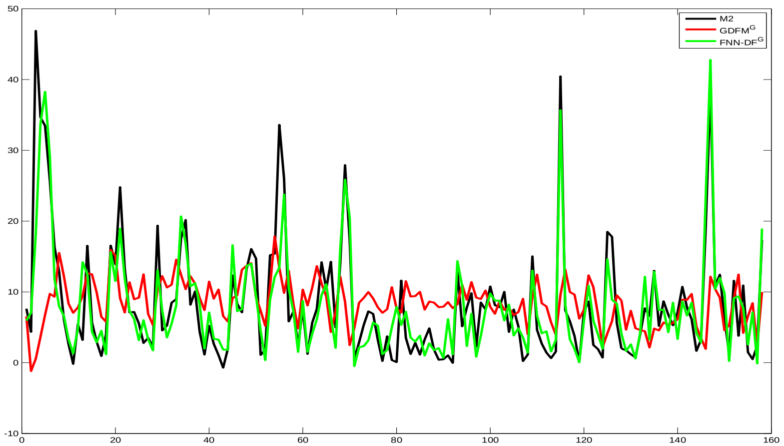
Notes: Entries are the relative MSE described in (C1) relative to the benchmark $GDFM^{G}$, relative MAE described in (C2), and DA test described in (C3). The Panel A corresponds to the model (**M1**) and the Panel B corresponds to the model (**M2**). We simulated 350 observations of the variables and discard the first 50 observations, leaving $T = 300$ observations for evaluation, while for the predictors we set $n = 150$. The simulated experiments were repeated 500 times.



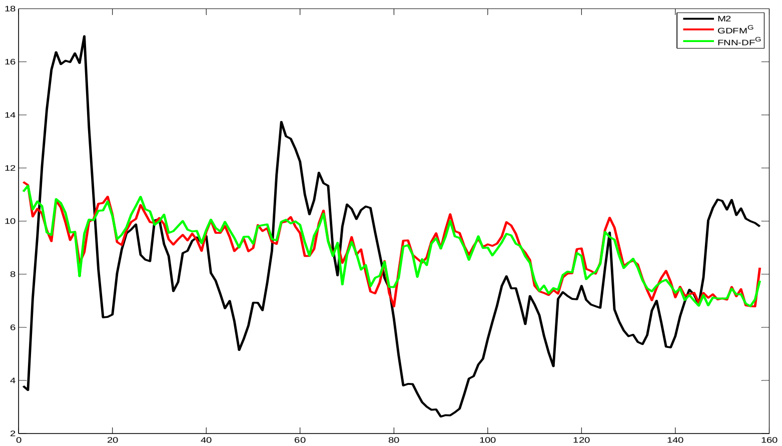
(a) (**M1**) Plot of forecasting results for H = 1.



(b) (**M1**) Plot of forecasting results for H =12.



(c) (**M2**) Plot of forecasting results for H = 1.



(d) (**M2**) Plot of forecasting results for H = 12.

Figure 2.2: Plot of forecasting results of one of the 500 realizations for **M1** (top) and **M2** (bottom). **Black line**: True Model, **Red line**: Forecasting for GDFM^G and **Green line**: Forecasting for FNN-DF^G.

Chapter 3

Forecasting U.S. Economic Time Series

In this chapter we present the use of the FNN-DF model proposed in Chapter 2 on real data using the benchmark datasets proposed by Stock and Watson (2002a) and reported in Appendix (A). The study is done on real and nominal macroeconomic variables. Forecasts obtained with the FNN-DF model are extended to another two methods, which have often exhibited more accurate prediction.

As described in the introduction, in empirical applications of both linear and nonlinear models, one of the main issues is that the model used is not correctly specified. For this as noted by Granger (1989) each model used may have specific information that other models do not possess. The use of techniques that allow the combination of information may provide better results than each individual model. In this regard the forecast combination of the linear and nonlinear models used in prediction are presented.

3.1 Motivations and Related Literature

Forecasting is very important in obtaining useful information to produce correct economic choices. However, the specification of the model is not always correct for the problem we are studying. This is quite evident when we make a choice between linear and nonlinear models. In fact, linear models tend to be used because they are easier to implement, may have an economic interpretation and are less time consuming than nonlinear models. Moreover, from an empirical point of view in recent years, the results of forecasts obtained with nonlinear models, especially using neural networks, are quite controversial. As we discuss below, in some cases the use of nonlinear techniques did not provide significant benefits, while in other cases provided encouraging results. In general, a question that often arises in forecasting field is as follows: if nonlinear models do not provide more accurate forecasts than those provided by linear models, why are they used? Since there is no reason to suppose that the economy is linear, a partial answer to this problem may come from the use of highly flexible forecasting techniques, such as techniques that are able to adapt quickly to functional changes. When the functional form between the variable to predict and the predictors is linear, the prediction obtained by nonlinear models are certainly not better than linear techniques, nor are they worse. Instead, if the functional form differs significantly from the linear one, then the results should be more accurate.

Remaining in the research field proposed by Stock and Watson (1998) and later re-examined by Terasvirta et al. (2005), we try to answer the same questions as in the case of GDFM. First, do nonlinear models produce forecasts that improve upon linear models in simulated real time? Second, if there are benefits to using nonlinear models, are the benefits great enough to justify their use? Finally, do combination forecasts outperform forecasts based on a single method?

In literature, the study of flexible forecasting techniques in conjunction with GDFM is not very large. Shintani (2005) addresses the problem of nonlinearity for the Japanese economy, using feedforward neural networks. The result he obtains is that in most series considered the feedforward neural networks have forecasting performance comparable to linear methods. Other studies however address the problem in the univariate case. Since the literature is readily available for this type of analysis, here we mention the most important results. Swanson and White (1997a,b) compared various methodologies for forecasting nine U.S. macroeconomic variables. The methods included linear autoregressive models, feedforward neural networks, professional consensus forecast and others techniques. The results that they obtain is the feedforward neural networks gives performance similar to linear models. In another forecasting comparison proposed by Stock and Watson (1998) a total of 49 univariate forecasting methods, including 15 feedforward neural networks, and various forecast pooling procedures, were used to forecast 215 U.S. monthly macroeconomic time series at three forecasting horizons. The various pooling procedures provided the most accurate forecasts, suggesting that neural networks may help improve forecasting accuracy when combined with other forecasts. However, when comparing the forecasting accuracy of individual models, the neural networks performed poorly relative to a "naive" AR(BIC) forecast and relative to most other methods in the comparison. Terasvirta et al. (2005) examine the forecast accuracy of linear autoregressive, smooth transition autoregressive (STAR) and neural network time series models for 47 monthly macroeconomic variables of the G7 economies. The forecast results indicate the STAR model generally outperforms linear autoregressive models, but does not dominate the feedforward neural networks forecast obtained using the Bayesian regularization approach. Heravi et al. (2004) considers 24 series measuring the annual change in monthly seasonally unadjusted industrial production for important sectors of the German, French and the U.K. economies. They found that linear models generally produce more accurate forecasts than neural network at horizons up to one year. This applies overall and also to the sub-group of series with substantial sample period with evidence of nonlinearity. However, they found that the neural networks give better results when the forecast results are measured in terms of direction accuracy. Dhal and Hylleberg (2004) used four alternative flexible nonlinear regression model approaches. The class of flexible regression model considered includes feedforward neural networks, two methods of projection pursuit and the random field approach. They found that linear models for the U.S. unemployment rate and the growth rate in U.S. industrial production cannot outperform the best flexible nonlinear regression models. Finally, Marcellino (2004) fits a variety of nonlinear and time varying models to aggregate European Monetary Union (E.M.U.) macroeconomic variables, and compares them with linear models. He found that several variables, linear models are beaten by nonlinear specifications.

3.2 The Data and Forecasting Procedure

In this section, to determine the practical usefulness of the model presented in Section (2.3), we conducted several forecasting exercises on different series of the U.S. economy. In particular, the series we are interested in forecasting are grouped in real and nominal categories: Personal Income, Real Consumption, Industrial Production and Unemployment Rate for real variables; Producer Price Index, Consumer Price Index, Money Supply and Interest Rate for nominal variables. These variables are modeled as follows, for real variables we assume that the $\log(RV) \sim I(1)$, for nominal variables we assume that the $\log(NV) \sim I(2)$, except for Interest Rate which is modeled as $\log(IR) \sim I(1)$ and Money Supply which is modeled as $\log(MS) \sim I(0)$. For $\mathbf{x}_n^T = \{x_{it}\ i = 1 \ldots n,\ t = 1 \ldots T\}$, the panel that includes the predictors, following the standard procedure in Stock and Watson (2002a), we consider 131 monthly U.S. economic time series observed from January, 1959 to December, 2003. The predictors are divided into 14 categories: real output and income; employment and hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories; orders; stock prices; exchange rates; interest rates and spreads; money and credit quantity aggregates; price indexes;

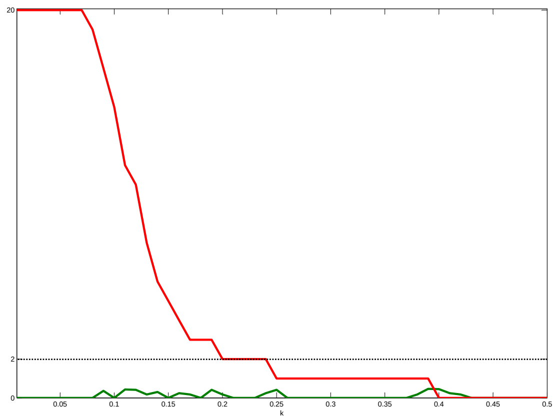
average hourly earnings and miscellaneous. The series are all transformed to be stationary by taking first or second differences, logarithms, first or second differences of logarithms, then the factors are estimated.

The exercise is based on simulated real time forecasting. This exercise begins with data from t =1960:1 ... 1971:1$-h$ for the estimation, then the values of the estimated parameters at T =1971:1 are used to forecast $\widehat{y}_{1971:1+h}$. When a new observation is available, the sample is updated by one month and moves from 1960:2 to 1971:2, the factors and parameters are both re-estimated and we obtain the second forecast $\widehat{y}_{1971:2+h}$. The process continues until the end of the sample is reached. The idea of this procedure is that as data become more distant in the past, we assume that they have little or no predictive relevance, so they can be removed from the sample. The last observation used in estimation is 2002:12, when $h = 12$.

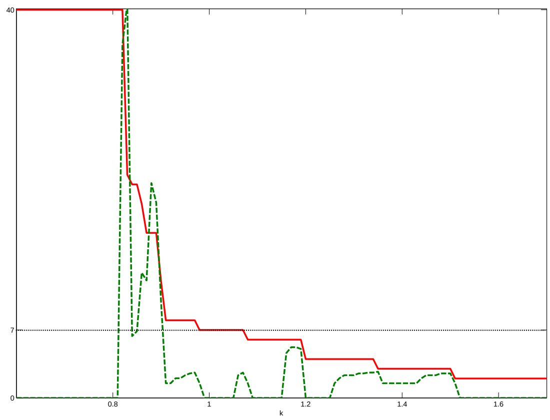
Finally it is important that the variables to be predicted and the estimated factors are scaled within the range of the activation function used. Without scaling, a great deal of information from the data is likely to be lost, since the hidden units are not able to recognize the data outside the limits of its activation function. Since in the FNN-DF model we use the activation function describe in equation (2.4), which has limits between $[-1, 1]$ we scale the variable of interest and the estimated factor into this range. Moreover in order to have comparable results in out of sample with other methods, once the forecast is obtained we rescale this to the original range, applying the inverse procedure.

3.3 Determining the Number of Factors

To determine the number of dynamic and static factors, we use the criteria described in Section (1.5). Regarding the criterion proposed by Bai and Ng (2002), as noted in Alessi et al. (2007), it is highly dependent on the variance of the residuals associated with principal component estimates, having as a practical consequence that it is very sensible to the choice of $r_{\max}$. To avoid this problem, Alessi et al. (2007) proposed, as similarrly did Hallin and Liska (2007) for the dynamic factor, to multiply the penalty function by a positive constant c. Then the procedure for selecting the number of static factors is similar to the procedure proposed by Hallin and Liska (2007).



(a) Number of Dynamic Factors.



(b) Number of Static Factors.

Figure 3.1: Plot of the Criteria proposed to determine the number of factors.

Following the empirical procedure proposed by Hallin and Liska (2007), we use the Figure (3.1) to determine the number of factors to use in the forecast exercise. In Figure (3.1) we look for the first zero variance interval for c (green line), corresponding to a stable value of $\hat{q}_{IC} < q_{\max}$ (red line); this represents the number of dynamic factors. In the Figure (3.1(a)) the criterion sets the number of dynamic factors equal to 2. The same reasoning is valid for determining the number of static factors, as recommended in the modified version by Alessi et al. (2007). In this case looking at the Figure (3.1(b)), the first zero variance interval for c corresponding to a stable value of $\hat{r}_{IC} < r_{\max}$ is located at 7 static factors.

3.4 Pooling Forecast

In forecast accuracy the main objective is to find which model has better performance with respect to a particular loss of function, such as MSE or MAE. In general the "best" model, as described in Chong and Hendry (1986), should be able to take into account the findings of alternative models. Moreover, as noted by Timmermann (2006), testing whether one forecast dominates another is neither sufficient nor adequate to establish if it is useful to combine them. Since the encompassing test to asses which models better represent the data are not appropriate in deciding whether combining the forecasts is useful to improve forecast accuracy, Chong and Hendry (1986) formalized a test which allows for the determination of whether a certain forecast incorporates (or encompasses) all the relevant futures in alternative forecast models. In order to explain pooling forecast, consider, as usual, the future value of the target variable is indicated by y_{t+h}. Suppose now that for a time t we have a set of information, denoted by $\mathcal{I}_t$, which is an F-vector of the forecasts for y_{t+h}, denoted by $\hat{\mathbf{y}}_{t+h} = (\hat{y}_{t+h,1}, \hat{y}_{t+h,2}, \ldots, \hat{y}_{t+h,F})$. The forecast combination problem tries to find an aggregator that reduces the information in a potentially high dimension vector of forecasts, $\hat{\mathbf{y}}_{t+h} \in \mathbb{R}^F$, to a lower dimensional summary measure, $\mathcal{C}(\hat{\mathbf{y}}_{t+h}; \boldsymbol{\pi}_c) \in \mathbb{R}^c \subset \mathbb{R}^F$, where $\boldsymbol{\pi}_c$ are the parameters associated with the combination. Since we are interested in a point forecast, we let $\hat{y}^c_{t+h} = \mathcal{C}(\hat{\mathbf{y}}_{t+h}; \boldsymbol{\pi})$ be the combined point forecast as a function of the underlying forecasts $\hat{\mathbf{y}}_{t+h}$ and the parameters of the combination, $\boldsymbol{\pi} \in \Pi$, where Π is often assumed a compact subset of $\mathbb{R}^F$. In this dissertation we make use of two methods

1. Equal weight combination, reported as (EW), meaning that the proposed methods have equal informative power. The weights are determined using the simple average;

2. Predictive Least Squares, reported as (PLS), proposed by Granger (1989), the weight of each forecast used in the pooling model is obtained using ordinary least squares, namely regressing the true value of target variable y_τ, on the F-vector of forecasts, $\hat{\mathbf{y}}_{\tau+h}$, that is

$$\hat{\boldsymbol{\pi}} = \left(\sum_{\tau=T_0}^{T_1-h} \hat{\mathbf{y}}_{\tau+h}\, \hat{\mathbf{y}}'_{\tau+h} \right)^{-1} \sum_{\tau=T_0}^{T_1-h} \hat{\mathbf{y}}_{\tau+h}\, y_{\tau+h}$$

The pooling forecast is obtained as

$$\hat{y}^c_{t+h} = \hat{\pi}_1\, \hat{y}^{\text{GDFM}^{\text{G}}}_{t+h} + \hat{\pi}_2\, \hat{y}^{\text{GDFM}^{\text{S}}}_{t+h} + \hat{\pi}_3\, \hat{y}^{\text{FNN-DF}^{\text{G}}}_{t+h} + \hat{\pi}_4\, \hat{y}^{\text{FNN-DF}^{\text{G}}}_{t+h} \tag{3.1}$$

The weights are re-estimated in each period, in order to simulate real time forecasting, using a rolling window of 36 months of the out of sample forecast computed between 1971:1 and 1973:12, as described in Section 3.2. We report the results for the pooling methods in Table (3.1), (3.3) and (3.5).

3.5 Forecast Comparison

There are many ways to evaluate the forecasting performance of a model. We use the same criteria as described in Section (2.4.5), results are reported in Table (3.1) to Table (3.6). We use as the benchmark an autoregressive model $y_{t+h} = \phi_0 + \phi 1 y_t + \cdots + \phi_p y_{t-p+1} + \epsilon_{t+h}$, where p is selected using the Bayesian Information Criterion (BIC) with $0 \leq p \leq 6$. The GDFM proposed in (1.6), where the factors used in prediction are estimated using static and dynamic principal components and are labeled as GDFM$^{\mathrm{S}}$ and GDFM$^{\mathrm{G}}$ respectively; for the comparison model we report results for the FNN-DF model described in (2.8) with factors estimated using both static and generalized principal components, and are labeled as FNN-DF$^{\mathrm{S}}$ and FNN-DF$^{\mathrm{G}}$ respectively. We report results for the choice of $r = 7$ static and $q = 2$ generalized principal components, whereas for the number of hidden units we fix $s = 5$. In the Tabless we will refer to the ratio of the MSE for a given method to the MSE of AR(BIC). An entry less than one indicates that the specified method is superior to the AR(BIC) forecast. For each series we compare all models at $h = 1$, 3, 6, 9, 12 month forecasting horizons.

It is useful in addition to the full sample analysis also consider two subsamples 1971:1-1984:12 and 1985:1-2002:12. The results of those subsamples help us to understand whether the methods proposed are helpful in improving forecast performance of the variable where its predictability is low; this phenomenon is known as "great moderation".

3.5.1 Test of Predictive Ability

To investigate the forecasting accuracy of the proposed models, we adopt the recent conditional predictive ability test of Giacomini and White (2006). Their test is based on an out-of-sample evaluation using a rolling window scheme. As described before, the in-sample size T used for estimation remains constant, while the sample itself and the points at which the forecast is evaluated, move with time. It is assumed that the number of out-of-sample forecasts tends to infinity while the in-sample size T remains constant.

One of the main motivations for using a conditional test versus an unconditional methodology of West (1996) is that it can be applied in a more general setting. The test evaluates not only the model itself, but the whole forecasting method, which includes the choice of the in-sample size T. While traditional tests of forecast equivalence answer the question of which forecast was more accurate on average, the Giacomini and White (2006) test answers the question of whether one can predict which forecast will be more accurate at a future date. The conditional methodology can be applied to the comparison of a wide range of models, such as parametric, semi-parametric, nonparametric and Bayesian models.

Suppose for simplicity that two alternative models are used to forecast the variable of interest h steps ahead, y_{t+h}. As described before, the forecasts formulated at time t are based on information set $\mathcal{I}_t$ and are denoted as $\mathrm{M}_i : \hat{y}_{t+h,i} = f(\mathbf{s}_t; \boldsymbol{\vartheta})$ with $i = 1, 2$. We evaluate the sequence of out of sample using a loss function $L_{t+h}(y_{t+h}, \hat{y}_{t+h,i})$ with $i = 1, 2$. The loss function for out-of-sample evaluation used in this dissertation are the squared error loss and the absolute error loss, as described in Section (2.4.5). For a given loss function and the σ-field $\mathcal{I}_t$, we write the null hypothesis of equal conditioned predictive ability of forecasts $\hat{y}_{t+h,1}$ and $\hat{y}_{t+h,2}$ for the horizon $t + h$ as

$$\mathrm{H}_0 : \ \mathrm{E}[\Delta L_{t+h} | \mathcal{I}_t] = 0$$

where $\Delta L_{t+h} = L_{t+h}(y_{t+h}, \hat{y}_{t+h,1}) - L_{t+h}(y_{t+h}, \hat{y}_{t+h,2})$. Nevertheless, before describing the test, it is important to clarify some concepts. In particular, if we are interested in testing which forecast is better on average, then we let $\mathcal{I}_t = (\emptyset, \Omega)$. If, on the other hand, we are interested in producing a forecast for specific data h periods in the future, then conditioning on $\mathcal{I}_t$ could be more appropriate, as it allows us to ask whether there is additional information that can helps identify which forecast is more appropriate for that date. This situation is particularly suited for real time forecasting exercise. When $\mathcal{I}_t$ is the σ-field $\mathcal{I}_t = (\emptyset, \Omega)$ and $h \geq 1$ the null hypothesis can be viewed as in Diebold and Mariano (1995) and West (1996); the test is based on the statistics

$$t_{\tau,h} = \frac{\Delta \bar{L}_\tau}{\hat{\sigma}_\tau / \sqrt{\tau}}$$

where $\Delta \bar{L}_\tau = \tau^{-1} \sum_{\tau=T_0}^{T_1-h} \Delta L_{\tau+h}$ and $\hat{\sigma}_\tau^2$ is a consistent estimator of the asymptotic variance σ_τ^2. For the level α test rejects the null hypothesis of equal unconditional predictive ability whenever $|t_{\tau,h}| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of a standard normal distribution.

3.6 Empirical Results

The forecasting results for the variables are reported in Table 3.1. Three sets of statistics are reported. The first is the MSE of the proposed forecasting model, computed relative to the MSE of the AR(BIC) forecast, so the autoregressive forecast has a relative mean square error (hereafter rMSE) of 1.000. The second is the relative mean absolute error (hereafter rMAE).

An entry of less than one indicates that the specified model is superior to the simple AR(BIC) forecast. The third is the directional accuracy (DA), which indicates the percentage of forecasts that correctly predict the direction of the change. We assess the significance of the observed differences in MSE between models by applying the pairwise Giacomini and White test (hereafter GW) of equal forecast accuracy, the test results are reported in Table 3.2. The entries in the table are the p-values of pairwise tests of equal forecast accuracy; the plus or minus sign indicates that the method in the row outperforms or underperforms, the method in the column at the 5% significance level. Therefore, in addition to the full sample analysis we report the results also for the subsamples 1974:1-1984:12 and 1985:1-2002:12. The results in these subsamples help us to understand whether the method proposed is useful for improving forecast performance of the variable where its predictability is decreased.

3.6.1 Forecasting Results

The results reported in Table 3.1 for Personal Income show that the forecasts obtained using both GDFM and FNN-DF, in terms of MSE, have equal performance with respect to the benchmark. Comparing the forecasts obtained from GDFM to those of FNN-DF, the nonlinear models outperform the linear. However, an improvement in terms of MSE or MAE does not reflect an improvement in correctly predicting the direction of the change of the series. The pooling forecasts obtained by using equal weights or by using the technique PLS do not provide significant improvement with respect to each single model, except for H=12. These results are confirmed in Table 3.2, where the test demonstrates that the three models: AR, GDFM, and FNN-DF, have equal prediction accuracy. However, when FNN-DF is compared to GDFM, the hypothesis of equal forecast accuracy is rejected in favor of nonlinear models. For Personal Income, complex models do not seem to provide an improvement in prediction accuracy.

For Real Consumption, in Table 3.1 when H=1, the GDFM and FNN-DF models have equal ability to forecast with respect to the benchmark. However, for H = 3, 6, 9, and 12, forecasts obtained using these models provide a small improvement. Comparing the forecasts obtained by linear and nonlinear techniques, FNN-DF models are more accurate in terms of MSE and MAE. Looking at the DA, the percentage of correctly predicting the sign of the series is in favor of the FNN-DF$^{\mathrm{G}}$ model, especially for H = 6, 9 and 12. The pooling forecasts do not provide a significant advantage with respect to each single model for H = 1, 3, 6, whereas for H=9 and

12, the use of PLS provides an improvement of about 10%. From Table 3.2, the GW test does not reject the hypothesis of equal forecast accuracy with respect to the benchmark. Comparing linear and nonlinear techniques, the results show that there is an improvement in MSE, which appears to be relevant only for $H = 3$ and 6. When predicting Real Consumption, there appears to be an advantage in using the nonlinear models, in comparison to linear models. However, using more complex models does not necessarily offer a significant improvement with respect to simple models such as the benchmark.

For Industrial Production Index, the results indicate that forecasts obtained with factor models using linear or nonlinear techniques are better compared to the benchmark. The improvement over the benchmark is approximately 25% for $H = 1$, while for $H = 3$, 6, 9, 12 the improvement increased as much as 40%. Comparing the forecast obtained by linear and nonlinear techniques, the forecasting performance of FNNDF models are worse than those of GDFM, especially for the horizon $H = 3$ and 9. However, observing the DA, there is no model which dominates the others. The equal weights pooling forecast seems to provide an advantage only when $H = 12$, however the PLS method does not seem to provide any advantage in forecast accuracy. Looking at the tests results reported in Table 3.2, the difference in terms of MSE for both linear and nonlinear models is statistically significant for all forecast horizons compared to the benchmark. The test does not indicate which model provides more accurate forecasts, instead indicating that the GDFMG model at $H = 1$ and 3 there is improved accuracy. For this variable, both linear and nonlinear models seem to have equally superior ability in prediction accuracy compared to the benchmark.

For Unemployment Rate, the results are very similar to Industrial Production Index. Here the improvements are in the order of 17% for $H = 1$, while for $H = 3$, 6, 9, and 12, the accuracy increased up to 46%. Comparing the forecasts obtained by linear and nonlinear techniques, the performance is very similar. The pooling forecasts do not provide significant improvement when compared to each single model. The directional accuracy, there is no particular model which dominates the others. Looking that the results reported in Table 3.2, the GW test results are statistically significant for all forecasting horizons for both linear and nonlinear models with respect to the benchmark. As for Industrial Production Index, the test does not indicate which model provides more accurate forecasts. Nonlinear models do not seem to provide more accurate forecasts for Unemployment Rate.

Also in the case of Money Supply (M2 series), the forecast obtained with the models using linear and nonlinear techniques have better performance with respect to the benchmark (Table 3.1). The models have very similar ability to correctly predict the direction of the series. Some evidence in favor of GDFM^{S} is found for $H = 12$. Comparing the forecasts obtained by linear and nonlinear techniques, the performance of FNN-DF models are worse than the GDFM models, especially for $H = 6$ and 9. The pooling forecast using PLS provides marginal improvement over the single models for $H = 1$ and 3, whereas for $H = 6$, 9, and 12 the improvement is more evident. Table 3.2, the poor performance of nonlinear models is confirmed by the GW test, particularly, for the horizon $H = 6$ and 9. For Money Supply (M2 Series), the pooling forecasts seem to mitigate the errors found in the nonlinear models when H is greater than 6.

For Interest Rate, as shown in Table 3.1, the forecasts obtained with linear and nonlinear models have better performance only for $H = 1$ and 3. For the horizon $H = 6$, 9, and 12, the performance is similar to the benchmark. Comparing rMSE to rMAE, the rMSE is larger for $H = 6$, 9, and 12; this result indicates that the forecast obtained by GDFM and FNN-DF are more sensitive to the outliers in the series than AR(BIC). Looking at DA, both GDFM and FNN-DF have better performance than the benchmark. Comparing linear and nonlinear models, FNN-DF^{G} outperforms linear models. The pooling forecast using PLS provide marginal improvement with respect to each single model with $H = 1$ and 3, whereas for $H = 6$, 9, and 12, it provides an improvement of about 10%. In Table 3.2, the GW test yields results which are statistically significant for the benchmark for $H = 6$, 9, and 12. Comparing the GDFM and the FNN-DF, the results FNN-DF^{G} are statistically significant. As for Money Supply, the pooling forecast seems to provide more accurate results.

For Producer Price Index, the linear and nonlinear models have the same ability as the benchmark to forecast for $H = 1$ and 3. For $H = 6$, 9, and 12, the forecasts obtained using these models provide some improvement over the benchmark. The directional accuracy for both linear and nonlinear models is superior to the benchmark, as is evident for $H = 9$ and 12. Comparing the forecasts obtained using linear and nonlinear techniques, the performance of both GDFM and FNN-DF is very similar, except for $H = 6$, 9, and 12, where the linear models are superior. The pooling forecasts do not provide significant improvement with respect to each individual model. The GW test results (3.2) confirm the results described above and contained in Table 3.1. The more complex models have better performance in short horizons for Producer Price Index forecasting.

For Consumer Price Index, the improvement of both linear and nonlinear techniques are in the order of 10% for $H = 1$, while for $H = 3$, 6, 9, and 12 it increased up to 60% compared to the benchmark. The directional accuracy both GDFM and FNN-DF dominate the benchmark, especially for $H = 9$ and 12 where the prediction of correct sign reaches 90%. Comparing the forecasts obtained by linear and nonlinear techniques, the GDFM models lead in accuracy compared to the FNN-DF models. The pooling forecasts do not provide significant improvement over the individual models. Looking at the GW test results both linear and nonlinear models perform better than the benchmark. Also comparing the GDFM and FNN-DF models, the results are statistically significant for GDFM. The linear models seem to perform better than the alternatives for Consumer Price Index.

Forecasting Results for 1974 - 1984. The results obtained for the period between 1974 and 1984 are very similar to those described above. However, for Personal Income the nonlinear model appears to be even better than AR(BIC); this result is evident for $H = 1, 3, 6$ as shown in Tables 3.3 and 3.4. For all other series, the results do not change.

Forecasting Results for 1985 - 2002. In the period between 1985 and 2002 we observe a substantial decrease in the forecasting ability of GDFM and FNN-DF compared to simpler models such as AR (BIC). Looking at Tables 3.5 and 3.6, this result is evident using either linear and nonlinear models. For Unemployment Rate and Money Supply is it possible obtain improvements.

Table 3.1: Forecasting Results for the period between 1974 - 2002

PERSONAL INCOME

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.509	1.000	1.000	0.560	1.000	1.000	0.497	1.000	1.000	0.431	1.000	1.000	0.440
GDFM^{G}	1.045	1.026	0.537	0.975	0.989	0.606	1.005	0.998	0.546	1.008	1.019	0.552	0.995	0.993	0.598
GDFM^{S}	1.072	1.045	0.555	0.993	1.006	0.581	1.037	1.043	0.549	1.081	1.054	0.552	1.075	1.042	0.624
Nonlinear															
FNN-DF^{G}	1.013	1.007	0.517	0.927	0.967	0.601	0.945	0.977	0.540	0.944	0.990	0.535	0.911	0.974	0.572
FNN-DF^{S}	1.015	1.000	0.575	0.929	0.959	0.572	0.947	0.990	0.546	0.942	0.988	0.560	0.971	1.013	0.581
Pooling															
PLS	1.026	1.002	0.537	0.939	0.971	0.592	0.947	0.976	0.537	0.938	0.982	0.526	0.888	0.968	0.586
EW	1.018	1.003	0.543	0.934	0.963	0.598	0.953	0.981	0.546	0.932	0.992	0.552	0.906	0.971	0.601

REAL CONSUMPTION

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.448	1.000	1.000	0.437	1.000	1.000	0.592	1.000	1.000	0.480	1.000	1.000	0.388
GDFM^{G}	1.035	1.032	0.535	0.897	0.964	0.618	0.895	0.974	0.569	0.917	1.027	0.546	0.929	1.037	0.500
GDFM^{S}	1.050	1.049	0.535	0.946	1.013	0.575	0.854	0.958	0.569	0.891	0.971	0.589	0.862	0.973	0.552
Nonlinear															
FNN-DF^{G}	0.989	1.007	0.549	0.848	0.934	0.629	0.866	0.957	0.603	0.885	0.998	0.603	0.904	1.015	0.580
FNN-DF^{S}	1.012	1.022	0.512	0.867	0.959	0.578	0.817	0.933	0.589	0.928	0.965	0.572	0.964	1.021	0.526
Pooling															
PLS	1.001	1.014	0.535	0.848	0.944	0.612	0.825	0.935	0.583	0.807	0.918	0.572	0.795	0.927	0.532
EW	1.004	1.017	0.549	0.868	0.956	0.615	0.820	0.938	0.581	0.850	0.965	0.572	0.850	0.973	0.523

INDUSTRIAL PRODUCTION INDEX

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.575	1.000	1.000	0.606	1.000	1.000	0.552	1.000	1.000	0.500	1.000	1.000	0.477
GDFM^{G}	0.741	0.887	0.658	0.623	0.835	0.713	0.561	0.811	0.681	0.572	0.811	0.704	0.603	0.773	0.675
GDFM^{S}	0.784	0.923	0.652	0.674	0.880	0.681	0.577	0.843	0.626	0.619	0.835	0.658	0.629	0.776	0.687
Nonlinear															
FNN-DF^{G}	0.736	0.873	0.658	0.619	0.827	0.713	0.551	0.810	0.672	0.637	0.853	0.670	0.630	0.783	0.664
FNN-DF^{S}	0.772	0.910	0.644	0.652	0.859	0.684	0.547	0.826	0.621	0.668	0.866	0.649	0.638	0.789	0.652
Pooling															
PLS	0.740	0.883	0.670	0.657	0.849	0.693	0.571	0.810	0.667	0.658	0.823	0.690	0.611	0.767	0.678
EW	0.738	0.884	0.672	0.625	0.843	0.687	0.541	0.811	0.641	0.594	0.825	0.684	0.588	0.756	0.675

UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.644	1.000	1.000	0.661	1.000	1.000	0.598	1.000	1.000	0.549	1.000	1.000	0.466
GDFM^{G}	0.830	0.944	0.672	0.748	0.891	0.718	0.573	0.799	0.727	0.555	0.759	0.753	0.544	0.725	0.739
GDFM^{S}	0.850	0.955	0.606	0.735	0.897	0.713	0.521	0.793	0.733	0.510	0.739	0.741	0.551	0.704	0.733
Nonlinear															
FNN-DF^{G}	0.816	0.937	0.670	0.742	0.888	0.718	0.617	0.815	0.724	0.591	0.786	0.730	0.592	0.749	0.730
FNN-DF^{S}	0.821	0.941	0.647	0.726	0.891	0.716	0.526	0.789	0.730	0.540	0.763	0.767	0.536	0.703	0.744
Pooling															
PLS	0.811	0.933	0.664	0.721	0.876	0.721	0.574	0.813	0.667	0.529	0.735	0.704	0.548	0.723	0.652
EW	0.815	0.936	0.655	0.722	0.876	0.730	0.537	0.786	0.747	0.516	0.738	0.744	0.514	0.694	0.739

– Continued on next page –

MONEY SUPPLY - M2

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.678	1.000	1.000	0.690	1.000	1.000	0.695	1.000	1.000	0.701	1.000	1.000	0.595
$GDFM^G$	0.785	0.888	0.721	0.765	0.898	0.707	0.762	0.852	0.672	0.675	0.797	0.733	0.709	0.851	0.736
$GDFM^S$	0.749	0.854	0.733	0.749	0.873	0.741	0.788	0.861	0.733	0.716	0.802	0.753	0.758	0.865	0.747
Nonlinear															
$FNN\text{-}DF^G$	0.776	0.884	0.730	0.767	0.898	0.713	0.790	0.863	0.678	0.728	0.821	0.721	0.756	0.866	0.701
$FNN\text{-}DF^S$	0.742	0.851	0.736	0.749	0.874	0.739	0.913	0.918	0.716	0.771	0.824	0.741	0.740	0.861	0.721
Pooling															
PLS	0.726	0.847	0.741	0.703	0.879	0.741	0.634	0.794	0.739	0.602	0.785	0.773	0.660	0.854	0.741
EW	0.738	0.851	0.733	0.718	0.865	0.721	0.753	0.851	0.724	0.679	0.794	0.750	0.687	0.840	0.736

INTEREST RATE: FEDERAL FUNDS (EFFECTIVE)

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.626	1.000	1.000	0.612	1.000	1.000	0.560	1.000	1.000	0.494	1.000	1.000	0.376
$GDFM^G$	0.828	0.963	0.695	0.948	0.933	0.690	0.992	0.926	0.667	1.022	0.921	0.658	0.961	0.881	0.598
$GDFM^S$	0.854	0.988	0.684	0.976	0.957	0.687	1.068	0.978	0.690	1.137	0.992	0.661	1.049	0.944	0.626
Nonlinear															
$FNN\text{-}DF^G$	0.785	0.933	0.698	0.875	0.902	0.684	0.963	0.922	0.661	0.973	0.918	0.638	0.956	0.877	0.612
$FNN\text{-}DF^S$	0.805	0.950	0.684	0.924	0.937	0.693	0.999	0.947	0.684	1.055	0.965	0.661	1.088	0.946	0.632
Pooling															
PLS	0.797	0.933	0.672	0.875	0.910	0.698	0.911	0.916	0.667	0.909	0.915	0.672	0.851	0.906	0.618
EW	0.796	0.935	0.693	0.908	0.923	0.687	0.975	0.931	0.684	1.012	0.933	0.655	0.972	0.897	0.606

PRODUCER PRICE INDEX

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.535	1.000	1.000	0.506	1.000	1.000	0.526	1.000	1.000	0.497	1.000	1.000	0.486
$GDFM^G$	0.996	1.026	0.529	0.991	1.015	0.549	0.877	0.953	0.615	0.859	0.947	0.552	0.740	0.893	0.586
$GDFM^S$	0.982	0.999	0.563	0.968	1.017	0.558	0.864	0.948	0.609	0.844	0.945	0.598	0.785	0.926	0.566
Nonlinear															
$FNN\text{-}DF^G$	0.975	1.011	0.540	0.975	1.002	0.543	0.889	0.962	0.595	0.859	0.937	0.558	0.828	0.912	0.558
$FNN\text{-}DF^S$	0.953	0.988	0.552	0.933	0.987	0.558	0.843	0.934	0.578	0.911	0.960	0.575	0.844	0.932	0.543
Pooling															
PLS	0.958	0.999	0.535	0.933	0.984	0.543	0.848	0.941	0.601	0.787	0.902	0.606	0.740	0.869	0.578
EW	0.959	1.002	0.540	0.949	0.993	0.558	0.831	0.930	0.618	0.844	0.941	0.569	0.773	0.906	0.572

CPI-U: ALL ITEMS

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.558	1.000	1.000	0.583	1.000	1.000	0.552	1.000	1.000	0.506	1.000	1.000	0.500
$GDFM^G$	0.876	0.960	0.618	0.691	0.866	0.641	0.636	0.824	0.664	0.506	0.739	0.618	0.474	0.718	0.621
$GDFM^S$	0.841	0.917	0.638	0.711	0.856	0.652	0.678	0.836	0.664	0.531	0.754	0.624	0.517	0.760	0.644
Nonlinear															
$FNN\text{-}DF^G$	0.874	0.956	0.592	0.683	0.859	0.641	0.716	0.861	0.655	0.595	0.785	0.586	0.597	0.799	0.586
$FNN\text{-}DF^S$	0.845	0.932	0.644	0.694	0.851	0.675	0.744	0.861	0.647	0.725	0.825	0.592	0.733	0.860	0.575
Pooling															
PLS	0.840	0.930	0.624	0.687	0.852	0.649	0.653	0.832	0.658	0.491	0.723	0.632	0.477	0.727	0.632
EW	0.839	0.931	0.638	0.685	0.849	0.664	0.649	0.817	0.667	0.534	0.743	0.618	0.532	0.768	0.612

Notes: Entries are relative MSE and relative MAE, relative to the benchmark AR. DA is the directional accuracy criterion. All the criteria were computed over the period 1974:1 - 2002:12-h. The pooling forecast is obtained applying the simulated forecasting exercise described in Section (3.2). The various columns correspond to forecasts of 1, 3, 6, 9 and 12-month growth, where all the multiperiod forecasts were computed using direct methods.

Table 3.2: Tests of Equal Forecast Accuracy for the Period 1974 - 2002.

PANEL A. FORECASTING HORIZON H = 1

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**
GDFM^G	0.872				0.809				0.013$^+$				0.016$^+$			
GDFM^S	0.959$^-$	0.870			0.880	0.685			0.046$^+$	0.967$^-$			0.032$^+$	0.817		
FNN-DF^G	0.668	0.008$^+$	0.015$^+$		0.375	0.000$^+$	0.032$^+$		0.003$^+$	0.393	0.109		0.005$^+$	0.158	0.113	
FNN-DF^S	0.656	0.095	0.002$^+$	0.527	0.652	0.186	0.019$^+$	0.828	0.015$^+$	0.869	0.327	0.956$^-$	0.006$^+$	0.335	0.050$^+$	0.617
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^G	0.006$^+$				0.046$^+$				0.464				0.053			
GDFM^S	0.000$^+$	0.182			0.042$^+$	0.799			0.389	0.348			0.038$^+$	0.151		
FNN-DF^G	0.003$^+$	0.230	0.760		0.013$^+$	0.047$^+$	0.040$^+$		0.266	0.116	0.431		0.028$^+$	0.444	0.807	
FNN-DF^S	0.000$^+$	0.143	0.247	0.176	0.018$^+$	0.303	0.051	0.773	0.156	0.067	0.092	0.211	0.015$^+$	0.144	0.575	0.123

PANEL B. FORECASTING HORIZON H = 3

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**
GDFM^G	0.384				0.138				0.004$^+$				0.009$^+$			
GDFM^S	0.471	0.658			0.291	0.875			0.012$^+$	0.953$^-$			0.006$^+$	0.313		
FNN-DF^G	0.159	0.006$^+$	0.069		0.033$^+$	0.008$^+$	0.011$^+$		0.002$^+$	0.403	0.069		0.006$^+$	0.258	0.600	
FNN-DF^S	0.166	0.099	0.005$^+$	0.517	0.052	0.219	0.003$^+$	0.764	0.005$^+$	0.802	0.102	0.877	0.004$^+$	0.210	0.181	0.262
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^G	0.037$^+$				0.307				0.468				0.015$^+$			
GDFM^S	0.013$^+$	0.391			0.420	0.629			0.045$^+$	0.313			0.023$^+$	0.745		
FNN-DF^G	0.029$^+$	0.542	0.633		0.045$^+$	0.016$^+$	0.119		0.386	0.305	0.554		0.010$^+$	0.221	0.162	
FNN-DF^S	0.006$^+$	0.406	0.497	0.381	0.231	0.377	0.025$^+$	0.767	0.229	0.146	0.114	0.149	0.013$^+$	0.531	0.079	0.675

PANEL C. FORECASTING HORIZON H = 6

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**
GDFM^G	0.521				0.174				0.011$^+$				0.009$^+$			
GDFM^S	0.632	0.665			0.116	0.286			0.022$^+$	0.657			0.006$^+$	0.176		
FNN-DF^G	0.239	0.001$^+$	0.103		0.103	0.071	0.559		0.009$^+$	0.238	0.247		0.009$^+$	0.953$^-$	0.898	
FNN-DF^S	0.275	0.144	0.003$^+$	0.513	0.048$^+$	0.083	0.088	0.191	0.017$^+$	0.380	0.017$^+$	0.456	0.005$^+$	0.195	0.619	0.103
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^G	0.049$^+$				0.484				0.224				0.014$^+$			
GDFM^S	0.044$^+$	0.665			0.629	0.888			0.045$^+$	0.953$^-$			0.036$^+$	0.759		
FNN-DF^G	0.048$^+$	0.899	0.513		0.415	0.259	0.048$^+$		0.232	0.721	0.955$^-$		0.047$^+$	0.961$^-$	0.658	
FNN-DF^S	0.288	0.957$^-$	0.960$^+$	0.903	0.498	0.542	0.026$^+$	0.758	0.141	0.284	0.252	0.230	0.070	0.956$^-$	0.875	0.661

– Continued on next page –

PANEL D. FORECASTING HORIZON H = 9

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**
GDFM^{G}	0.535				0.246				0.012^{+}				0.001^{+}			
GDFM^{S}	0.718	0.810			0.233	0.340			0.038^{+}	0.961^{-}			0.002^{+}	0.189		
FNN-DF^{G}	0.225	0.013^{+}	0.077		0.138	0.071	0.466		0.021^{+}	0.997^{-}	0.634		0.003^{+}	0.851	0.958^{-}	
FNN-DF^{S}	0.124	0.036^{+}	0.009^{+}	0.171	0.322	0.551	0.689	0.673	0.046^{+}	0.978^{-}	0.959^{-}	0.751	0.003^{+}	0.393	0.954^{-}	0.143
	MONEY SUPPLY - M2				**INTEREST RATE**				**PRODUCER PRICE INDEX**				**CPI-U: ALL ITEMS**			
GDFM^{G}	0.009^{+}				0.539				0.043^{+}				0.001^{+}			
GDFM^{S}	0.033^{+}	0.734			0.706	0.885			0.048^{+}	0.095			0.002^{+}	0.813		
FNN-DF^{G}	0.022^{+}	0.990^{-}	0.567		0.451	0.195	0.054		0.136	0.508	0.587		0.003^{+}	0.955^{-}	0.824	
FNN-DF^{S}	0.063	0.953^{-}	0.924	0.749	0.591	0.636	0.032^{+}	0.846	0.233	0.726	0.967^{+}	0.905	0.055	0.964^{-}	0.959^{-}	0.957^{-}

PANEL E. FORECASTING HORIZON H = 12

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**	**AR**	**GDFM^{G}**	**GDFM^{S}**	**FNN-DF^{G}**
GDFM^{G}	0.482				0.302				0.006^{+}				0.000^{+}			
GDFM^{S}	0.666	0.814			0.196	0.152			0.021^{+}	0.710			0.001^{+}	0.560		
FNN-DF^{G}	0.185	0.032^{+}	0.044^{+}		0.213	0.137	0.703		0.007^{+}	0.866	0.505		0.000^{+}	0.974^{-}	0.801	
FNN-DF^{S}	0.426	0.404	0.084	0.755	0.409	0.657	0.929	0.741	0.020^{+}	0.758	0.612	0.563	0.001^{+}	0.434	0.271	0.117
	MONEY SUPPLY - M2				**INTEREST RATE**				**PRODUCER PRICE INDEX**				**CPI-U: ALL ITEMS**			
GDFM^{G}	0.018^{+}				0.415				0.029^{+}				0.001^{+}			
GDFM^{S}	0.064	0.793			0.589	0.834			0.048^{+}	0.972^{-}			0.001^{+}	0.952^{-}		
FNN-DF^{G}	0.028^{+}	0.809	0.490		0.413	0.451	0.122		0.084	0.971^{-}	0.963^{+}		0.005^{+}	0.988^{-}	0.894	
FNN-DF^{S}	0.035^{+}	0.661	0.393	0.405	0.640	0.889	0.790	0.950^{-}	0.086	0.958^{-}	0.977^{+}	0.642	0.057	0.984^{-}	0.964^{-}	0.958^{-}

Notes: Results of tests of equal forecast accuracy as described in Section 3.5.1. Entries are the *p*-values of the test for the forecast in the corresponding row and column. A plus or minus sign indicates that the method in the row outperforms or underperforms the method in the column at the 5% significance level. For example, for Personal Income at the H = 1 horizon, equal forecast accuracy of the GDFM^{G} and the FNN-DF^{G} is rejected with a *p*-value of 0.008, then FNN-DF^{G} outperforms GDFM^{G}.

Table 3.3: Forecasting Results for the Period 1974 - 1984

PERSONAL INCOME

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.511	1.000	1.000	0.474	1.000	1.000	0.511	1.000	1.000	0.436	1.000	1.000	0.361
GDFM^{G}	1.015	1.004	0.564	0.816	0.876	0.684	0.834	0.893	0.639	0.797	0.922	0.752	0.776	0.864	0.805
GDFM^{S}	0.998	0.971	0.624	0.778	0.854	0.692	0.840	0.943	0.677	0.879	0.961	0.767	0.912	0.945	0.827
Nonlinear															
FNN-DF^{G}	0.967	0.989	0.564	0.777	0.852	0.677	0.801	0.894	0.624	0.791	0.921	0.707	0.785	0.885	0.782
FNN-DF^{S}	0.912	0.944	0.594	0.746	0.809	0.669	0.806	0.910	0.662	0.711	0.880	0.752	0.939	0.977	0.767
Pooling															
PLS	0.976	0.966	0.579	0.766	0.829	0.707	0.782	0.879	0.662	0.758	0.905	0.699	0.574	0.755	0.790
EW	0.954	0.968	0.587	0.760	0.828	0.692	0.786	0.881	0.647	0.737	0.902	0.767	0.770	0.872	0.812

REAL CONSUMPTION

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.466	1.000	1.000	0.421	1.000	1.000	0.571	1.000	1.000	0.414	1.000	1.000	0.278
GDFM^{G}	1.060	1.040	0.549	0.781	0.891	0.692	0.670	0.780	0.654	0.675	0.803	0.707	0.678	0.809	0.609
GDFM^{S}	1.071	1.055	0.541	0.836	0.938	0.647	0.678	0.799	0.692	0.712	0.791	0.714	0.694	0.799	0.677
Nonlinear															
FNN-DF^{G}	1.000	1.013	0.541	0.742	0.872	0.699	0.664	0.781	0.662	0.676	0.799	0.722	0.682	0.810	0.587
FNN-DF^{S}	1.047	1.044	0.504	0.755	0.883	0.662	0.661	0.788	0.707	0.801	0.816	0.692	0.860	0.910	0.602
Pooling															
PLS	1.021	1.028	0.541	0.747	0.887	0.654	0.638	0.766	0.677	0.608	0.739	0.707	0.608	0.753	0.684
EW	1.025	1.032	0.534	0.764	0.889	0.654	0.635	0.773	0.692	0.665	0.781	0.714	0.665	0.785	0.639

INDUSTRIAL PRODUCTION INDEX

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.556	1.000	1.000	0.647	1.000	1.000	0.481	1.000	1.000	0.308	1.000	1.000	0.293
GDFM^{G}	0.630	0.814	0.707	0.550	0.743	0.790	0.398	0.626	0.790	0.360	0.593	0.805	0.346	0.534	0.850
GDFM^{S}	0.647	0.840	0.714	0.567	0.760	0.782	0.375	0.624	0.797	0.396	0.612	0.797	0.381	0.568	0.842
Nonlinear															
FNN-DF^{G}	0.646	0.813	0.707	0.548	0.737	0.805	0.392	0.630	0.790	0.444	0.656	0.744	0.383	0.547	0.835
FNN-DF^{S}	0.663	0.840	0.692	0.553	0.738	0.782	0.348	0.606	0.805	0.457	0.655	0.790	0.400	0.576	0.790
Pooling															
PLS	0.635	0.812	0.714	0.576	0.737	0.812	0.420	0.635	0.805	0.506	0.663	0.820	0.422	0.601	0.827
EW	0.634	0.816	0.737	0.544	0.740	0.797	0.363	0.608	0.797	0.386	0.610	0.797	0.338	0.528	0.842

UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.632	1.000	1.000	0.639	1.000	1.000	0.466	1.000	1.000	0.391	1.000	1.000	0.384
GDFM^{G}	0.752	0.920	0.684	0.701	0.840	0.677	0.497	0.712	0.699	0.463	0.651	0.767	0.433	0.596	0.835
GDFM^{S}	0.747	0.923	0.669	0.661	0.812	0.714	0.429	0.682	0.729	0.421	0.624	0.759	0.470	0.604	0.850
Nonlinear															
FNN-DF^{G}	0.743	0.913	0.692	0.695	0.839	0.684	0.551	0.728	0.699	0.496	0.679	0.737	0.491	0.625	0.842
FNN-DF^{S}	0.740	0.910	0.662	0.663	0.815	0.699	0.433	0.684	0.722	0.441	0.650	0.782	0.433	0.579	0.880
Pooling															
PLS	0.732	0.908	0.669	0.678	0.827	0.684	0.513	0.734	0.699	0.468	0.651	0.722	0.488	0.650	0.752
EW	0.735	0.910	0.669	0.669	0.816	0.699	0.460	0.696	0.744	0.431	0.637	0.744	0.428	0.581	0.857

– Continued on next page –

MONEY SUPPLY - M2

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.677	1.000	1.000	0.669	1.000	1.000	0.729	1.000	1.000	0.609	1.000	1.000	0.278
GDFM^G	0.683	0.835	0.744	0.650	0.840	0.774	0.615	0.782	0.744	0.575	0.714	0.790	0.560	0.744	0.812
GDFM^S	0.668	0.802	0.774	0.705	0.841	0.774	0.773	0.862	0.782	0.713	0.772	0.797	0.677	0.814	0.774
Nonlinear															
FNN-DF^G	0.674	0.825	0.767	0.656	0.837	0.767	0.650	0.789	0.744	0.631	0.743	0.774	0.638	0.800	0.737
FNN-DF^S	0.655	0.792	0.774	0.701	0.830	0.759	0.993	0.967	0.744	0.812	0.820	0.752	0.669	0.813	0.744
Pooling															
PLS	0.649	0.796	0.767	0.550	0.762	0.812	0.562	0.679	0.752	0.501	0.712	0.790	0.510	0.678	0.774
EW	0.650	0.794	0.752	0.635	0.808	0.752	0.677	0.822	0.767	0.630	0.743	0.797	0.579	0.770	0.782

INTEREST RATE: FEDERAL FUNDS (EFFECTIVE)

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.624	1.000	1.000	0.541	1.000	1.000	0.451	1.000	1.000	0.459	1.000	1.000	0.263
GDFM^G	0.791	0.897	0.752	0.951	0.943	0.647	0.992	0.910	0.699	1.086	0.964	0.632	0.951	0.853	0.594
GDFM^S	0.836	0.965	0.707	0.976	0.977	0.624	1.030	0.939	0.722	1.117	0.990	0.669	0.951	0.869	0.662
Nonlinear															
FNN-DF^G	0.763	0.891	0.752	0.882	0.920	0.662	0.962	0.905	0.692	1.028	0.955	0.602	0.953	0.849	0.609
FNN-DF^S	0.793	0.931	0.714	0.927	0.961	0.632	0.973	0.911	0.722	1.049	0.962	0.662	1.021	0.881	0.677
Pooling															
PLS	0.784	0.906	0.722	0.881	0.927	0.647	0.886	0.875	0.684	0.892	0.888	0.692	0.701	0.771	0.654
EW	0.782	0.910	0.722	0.915	0.946	0.639	0.970	0.908	0.692	1.048	0.951	0.654	0.941	0.847	0.624

PRODUCER PRICE INDEX

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.617	1.000	1.000	0.602	1.000	1.000	0.677	1.000	1.000	0.564	1.000	1.000	0.444
GDFM^G	1.094	1.079	0.504	0.925	1.021	0.662	0.859	1.008	0.714	0.763	0.898	0.752	0.564	0.744	0.812
GDFM^S	1.108	1.066	0.526	0.843	0.983	0.677	0.804	0.949	0.714	0.683	0.863	0.782	0.587	0.771	0.797
Nonlinear															
FNN-DF^G	1.064	1.061	0.496	0.917	1.002	0.639	0.889	1.026	0.692	0.809	0.909	0.737	0.736	0.803	0.774
FNN-DF^S	1.064	1.038	0.534	0.834	0.963	0.684	0.808	0.948	0.692	0.888	0.932	0.782	0.735	0.824	0.759
Pooling															
PLS	1.069	1.047	0.504	0.867	0.977	0.662	0.796	0.967	0.707	0.666	0.824	0.827	0.603	0.737	0.790
EW	1.068	1.055	0.519	0.859	0.981	0.677	0.781	0.951	0.722	0.756	0.895	0.759	0.626	0.780	0.797

CPI-U: ALL ITEMS

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.677	1.000	1.000	0.669	1.000	1.000	0.639	1.000	1.000	0.564	1.000	1.000	0.466
GDFM^G	0.871	0.976	0.654	0.547	0.756	0.759	0.516	0.755	0.774	0.361	0.587	0.790	0.322	0.539	0.865
GDFM^S	0.898	0.954	0.684	0.569	0.752	0.767	0.576	0.764	0.759	0.395	0.608	0.759	0.388	0.599	0.842
Nonlinear															
FNN-DF^G	0.883	0.987	0.624	0.554	0.769	0.759	0.643	0.826	0.722	0.461	0.656	0.729	0.491	0.679	0.805
FNN-DF^S	0.887	0.966	0.684	0.567	0.759	0.782	0.687	0.829	0.729	0.624	0.722	0.714	0.656	0.765	0.737
Pooling															
PLS	0.869	0.963	0.654	0.562	0.765	0.774	0.554	0.781	0.752	0.377	0.611	0.790	0.372	0.604	0.827
EW	0.868	0.961	0.662	0.553	0.756	0.782	0.553	0.758	0.759	0.397	0.610	0.752	0.412	0.631	0.812

Notes: Entries are relative MSE and relative MAE, relative to the benchmark AR. DA is the directional accuracy criterion. All the criteria were computed over the period 1974:1 - 1985:12-h. The pooling forecast is obtained applying the simulated forecasting exercise described in Section (3.2). The various columns correspond to forecasts of 1, 3, 6, 9 and 12-month growth, where all the multiperiod forecasts were computed using direct methods.

Table 3.4: Tests of Equal Forecast Accuracy for the Period 1974 - 1984.

Models	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$
PANEL A. FORECASTING HORIZON H = 1																
	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
$GDFM^G$	0.561				0.789				0.011^{+}				0.021^{+}			
$GDFM^S$	0.497	0.366			0.770	0.557			0.024^{+}	0.708			0.020^{+}	0.425		
$FNN\text{-}DF^G$	0.320	0.034^{+}	0.302		0.504	0.015^{+}	0.163		0.004^{+}	0.707	0.500		0.010^{+}	0.292	0.444	
$FNN\text{-}DF^S$	0.049^{+}	0.011^{+}	0.044^{+}	0.039^{+}	0.773	0.382	0.271	0.854	0.010^{+}	0.853	0.657	0.759	0.010^{+}	0.326	0.337	0.443
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
$GDFM^G$	0.013^{+}				0.024^{+}				0.848				0.143			
$GDFM^S$	0.003^{+}	0.381			0.052	0.956^{-}			0.844	0.617			0.240	0.725		
$FNN\text{-}DF^G$	0.007^{+}	0.343	0.541		0.008^{+}	0.207	0.040^{+}		0.841	0.216	0.269		0.128	0.703	0.383	
$FNN\text{-}DF^S$	0.001^{+}	0.296	0.238	0.333	0.016^{+}	0.514	0.093	0.953^{-}	0.822	0.306	0.153	0.511	0.154	0.665	0.380	0.564
PANEL B. FORECASTING HORIZON H = 3																
	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$
$GDFM^G$	0.098				0.069				0.005^{+}				0.012^{+}			
$GDFM^S$	0.058^{+}	0.286			0.157	0.975^{-}			0.006^{+}	0.666			0.004^{+}	0.090		
$FNN\text{-}DF^G$	0.034^{+}	0.041^{+}	0.048^{+}		0.022^{+}	0.049^{+}	0.041^{+}		0.002^{+}	0.471	0.328		0.009^{+}	0.320	0.960^{-}	
$FNN\text{-}DF^S$	0.014^{+}	0.047^{+}	0.205	0.221	0.028^{+}	0.192	0.021^{+}	0.653	0.003^{+}	0.531	0.239	0.555	0.004^{+}	0.107	0.580	0.110
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
$GDFM^G$	0.053				0.326				0.345				0.013^{+}			
$GDFM^S$	0.068	0.830			0.436	0.626			0.216	0.164			0.023^{+}	0.767		
$FNN\text{-}DF^G$	0.045^{+}	0.579	0.155		0.112	0.029^{+}	0.144		0.283	0.445	0.797		0.012^{+}	0.770	0.321	
$FNN\text{-}DF^S$	0.041^{+}	0.717	0.470	0.722	0.272	0.415	0.047^{+}	0.758	0.149	0.188	0.440	0.129	0.018^{+}	0.779	0.460	0.708
PANEL C. FORECASTING HORIZON H = 6																
	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$
$GDFM^G$	0.128				0.008^{+}				0.004^{+}				0.008^{+}			
$GDFM^S$	0.191	0.517			0.018^{+}	0.533			0.006^{+}	0.307			0.004^{+}	0.153		
$FNN\text{-}DF^G$	0.047^{+}	0.109	0.371		0.005^{+}	0.397	0.443		0.003^{+}	0.343	0.651		0.007^{+}	0.959^{-}	0.952^{-}	
$FNN\text{-}DF^S$	0.104	0.378	0.226	0.524	0.008^{+}	0.442	0.302	0.480	0.005^{+}	0.176	0.043^{+}	0.177	0.003^{+}	0.162	0.599	0.085
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
$GDFM^G$	0.048^{+}				0.500				0.335				0.018^{+}			
$GDFM^S$	0.206	0.957^{-}			0.556	0.719			0.279	0.301			0.047^{+}	0.768		
$FNN\text{-}DF^G$	0.097	0.805	0.115		0.438	0.272	0.165		0.363	0.818	0.755		0.070	0.953^{-}	0.684	
$FNN\text{-}DF^S$	0.485	0.992^{-}	0.950^{-}	0.983^{-}	0.457	0.377	0.081	0.571	0.262	0.324	0.531	0.251	0.105	0.958^{-}	0.904	0.676

– Continued on next page –

PANEL D. FORECASTING HORIZON H = 9

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}
GDFM^{G}	0.086				0.012^{+}				0.002^{+}				0.001^{+}			
GDFM^{S}	0.320	0.730			0.066	0.715			0.011^{+}	0.747			0.001^{+}	0.237		
FNN-DF^{G}	0.108	0.420	0.286		0.004^{+}	0.509	0.341		0.005^{+}	0.999^{-}	0.772		0.002^{+}	0.782	0.909	
FNN-DF^{S}	0.066	0.206	0.126	0.236	0.170	0.874	0.809	0.843	0.012^{+}	0.954^{-}	0.953^{-}	0.590	0.002^{+}	0.359	0.895	0.147
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^{G}	0.031^{+}				0.603				0.154				0.000^{+}			
GDFM^{S}	0.148	0.958^{-}			0.649	0.683			0.106	0.220			0.001^{+}	0.971^{+}		
FNN-DF^{G}	0.050^{+}	0.962^{-}	0.204		0.525	0.217	0.173		0.180	0.811	0.847		0.002^{+}	0.967^{-}	0.778	
FNN-DF^{S}	0.239	0.995^{-}	0.969^{-}	0.979^{-}	0.568	0.421	0.095	0.653	0.295	0.814	0.865	0.802	0.050^{+}	0.957^{-}	0.958^{-}	0.957^{-}

PANEL E. FORECASTING HORIZON H = 12

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}
GDFM^{G}	0.131				0.026^{+}				0.000^{+}				0.000^{+}			
GDFM^{S}	0.394	0.831			0.069	0.567			0.004^{+}	0.704			0.001^{+}	0.756		
FNN-DF^{G}	0.111	0.602	0.214		0.015^{+}	0.556	0.466		0.000^{+}	0.906	0.515		0.000^{+}	0.967^{-}	0.644	
FNN-DF^{S}	0.422	0.871	0.595	0.835	0.241	0.957^{-}	0.961^{-}	0.957^{-}	0.004^{+}	0.790	0.699	0.594	0.000^{+}	0.492	0.058	0.122
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^{G}	0.015^{+}				0.397				0.008^{+}				0.000^{+}			
GDFM^{S}	0.093	0.969^{-}			0.412	0.508			0.010^{+}	0.666			0.000^{+}	0.991^{-}		
FNN-DF^{G}	0.024^{+}	0.824	0.385		0.414	0.523	0.511		0.061	0.967^{-}	0.832		0.002^{+}	0.997^{-}	0.954^{-}	
FNN-DF^{S}	0.065	0.834	0.472	0.637	0.519	0.740	0.893	0.818	0.042^{+}	0.972^{-}	0.867	0.497	0.036^{+}	0.991^{-}	0.967^{-}	0.960^{-}

Notes: Results of test of equal forecast accuracy as described in Section 3.5.1. Entries are the p-values of the test for the forecast in the corresponding row and column. A plus or minus sign indicates that the method in the row outperforms or underperforms the method in the column at the 5% significance level. For example, for Personal Income at the $H = 1$ horizon, equal forecast accuarcy of the GDFM^{G} and the FNN-DF^{G} is rejected with a p-value of 0.034, then FNN-DF^{G} outperforms GDFM^{G}.

Table 3.5: Forecasting Results for the Period 1984 - 2002

PERSONAL INCOME

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.507	1.000	1.000	0.614	1.000	1.000	0.488	1.000	1.000	0.428	1.000	1.000	0.488
$GDFM^G$	1.060	1.042	0.521	1.142	1.089	0.558	1.185	1.085	0.488	1.212	1.094	0.428	1.247	1.104	0.470
$GDFM^S$	1.109	1.098	0.512	1.220	1.139	0.512	1.245	1.124	0.470	1.276	1.124	0.419	1.264	1.127	0.498
Nonlinear															
$FNN\text{-}DF^G$	1.036	1.019	0.488	1.085	1.067	0.554	1.097	1.045	0.488	1.092	1.042	0.428	1.057	1.051	0.442
$FNN\text{-}DF^S$	1.065	1.041	0.563	1.120	1.090	0.512	1.095	1.055	0.474	1.047	1.035	0.442	1.009	1.045	0.465
Pooling															
PLS	1.051	1.027	0.512	1.121	1.095	0.521	1.121	1.056	0.461	1.112	1.041	0.419	1.034	1.022	0.461
EW	1.049	1.028	0.516	1.116	1.082	0.540	1.129	1.063	0.484	1.120	1.061	0.419	1.107	1.058	0.470

REAL CONSUMPTION

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.437	1.000	1.000	0.447	1.000	1.000	0.605	1.000	1.000	0.521	1.000	1.000	0.456
$GDFM^G$	1.018	1.026	0.526	1.078	1.038	0.572	1.400	1.217	0.516	1.520	1.319	0.447	1.518	1.328	0.433
$GDFM^S$	1.036	1.045	0.530	1.117	1.088	0.530	1.250	1.158	0.493	1.339	1.206	0.512	1.257	1.195	0.474
Nonlinear															
$FNN\text{-}DF^G$	0.982	1.002	0.554	1.012	0.997	0.586	1.318	1.178	0.502	1.406	1.257	0.465	1.422	1.276	0.414
$FNN\text{-}DF^S$	0.986	1.007	0.516	1.043	1.035	0.526	1.168	1.115	0.516	1.246	1.159	0.498	1.206	1.163	0.479
Pooling															
PLS	0.986	1.003	0.530	1.004	1.002	0.586	1.243	1.147	0.526	1.236	1.152	0.488	1.166	1.149	0.437
EW	0.989	1.006	0.558	1.029	1.024	0.591	1.235	1.145	0.512	1.311	1.206	0.484	1.285	1.214	0.451

INDUSTRIAL PRODUCTION INDEX

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.586	1.000	1.000	0.581	1.000	1.000	0.595	1.000	1.000	0.619	1.000	1.000	0.591
$GDFM^G$	0.974	0.960	0.628	0.955	0.957	0.665	1.110	1.081	0.614	1.208	1.135	0.642	1.221	1.094	0.567
$GDFM^S$	1.068	1.006	0.614	1.012	1.039	0.619	1.257	1.164	0.521	1.290	1.167	0.572	1.226	1.054	0.591
Nonlinear															
$FNN\text{-}DF^G$	0.973	0.933	0.628	0.944	0.947	0.656	1.091	1.074	0.600	1.218	1.145	0.623	1.223	1.098	0.558
$FNN\text{-}DF^S$	0.998	0.979	0.614	0.966	1.020	0.623	1.220	1.148	0.507	1.304	1.180	0.563	1.210	1.074	0.567
Pooling															
PLS	0.960	0.953	0.642	0.911	0.997	0.619	1.081	1.066	0.581	1.114	1.061	0.609	1.066	0.989	0.586
EW	0.954	0.952	0.633	0.885	0.980	0.619	1.142	1.107	0.544	1.219	1.144	0.614	1.188	1.061	0.572

UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.651	1.000	1.000	0.674	1.000	1.000	0.679	1.000	1.000	0.647	1.000	1.000	0.516
$GDFM^G$	0.945	0.967	0.665	0.880	0.956	0.744	0.887	0.943	0.744	0.892	0.936	0.744	0.839	0.909	0.679
$GDFM^S$	1.001	0.986	0.567	0.942	1.006	0.712	0.901	0.975	0.735	0.834	0.926	0.730	0.764	0.848	0.661
Nonlinear															
$FNN\text{-}DF^G$	0.924	0.960	0.656	0.872	0.950	0.740	0.890	0.958	0.740	0.936	0.961	0.726	0.863	0.926	0.661
$FNN\text{-}DF^S$	0.941	0.969	0.637	0.905	0.989	0.726	0.910	0.962	0.735	0.902	0.947	0.758	0.811	0.881	0.661
Pooling															
PLS	0.927	0.957	0.661	0.844	0.940	0.744	0.827	0.944	0.647	0.752	0.871	0.693	0.704	0.828	0.591
EW	0.932	0.960	0.647	0.870	0.954	0.749	0.854	0.933	0.749	0.824	0.903	0.744	0.745	0.855	0.665

– Continued on next page –

MONEY SUPPLY - M2

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.679	1.000	1.000	0.702	1.000	1.000	0.674	1.000	1.000	0.758	1.000	1.000	0.791
$GDFM^G$	0.875	0.926	0.707	0.905	0.945	0.665	0.916	0.902	0.628	0.789	0.865	0.698	0.938	0.952	0.688
$GDFM^S$	0.821	0.892	0.707	0.803	0.899	0.721	0.804	0.861	0.702	0.719	0.827	0.726	0.881	0.913	0.730
Nonlinear															
$FNN\text{-}DF^G$	0.866	0.926	0.707	0.903	0.947	0.679	0.939	0.917	0.637	0.839	0.885	0.688	0.935	0.928	0.679
$FNN\text{-}DF^S$	0.820	0.894	0.712	0.808	0.908	0.726	0.828	0.882	0.698	0.723	0.827	0.735	0.849	0.906	0.707
Pooling															
PLS	0.795	0.884	0.726	0.889	0.971	0.698	0.815	0.879	0.730	0.718	0.845	0.763	0.935	1.018	0.721
EW	0.817	0.891	0.721	0.820	0.910	0.702	0.834	0.872	0.698	0.734	0.835	0.721	0.853	0.906	0.707

INTEREST RATE: FEDERAL FUNDS (EFFECTIVE)

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.628	1.000	1.000	0.656	1.000	1.000	0.628	1.000	1.000	0.516	1.000	1.000	0.447
$GDFM^G$	1.227	1.117	0.661	0.920	0.912	0.716	0.991	0.955	0.647	0.926	0.955	0.674	1.002	0.927	0.600
$GDFM^S$	1.045	1.042	0.670	0.977	0.913	0.726	1.278	1.045	0.670	1.212	0.995	0.656	1.425	1.062	0.605
Nonlinear															
$FNN\text{-}DF^G$	1.023	1.031	0.665	0.914	0.965	0.698	0.971	0.952	0.642	0.904	0.961	0.661	0.966	0.921	0.614
$FNN\text{-}DF^S$	0.933	0.996	0.665	0.892	0.883	0.730	1.140	1.010	0.661	1.078	0.970	0.661	1.345	1.047	0.605
Pooling															
PLS	0.940	0.996	0.642	0.924	0.972	0.730	1.048	0.988	0.656	0.971	0.956	0.661	0.943	0.936	0.595
EW	0.951	0.994	0.674	0.936	0.974	0.716	1.002	0.972	0.679	0.955	0.965	0.656	1.091	0.976	0.595

PRODUCER PRICE INDEX

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.484	1.000	1.000	0.447	1.000	1.000	0.433	1.000	1.000	0.456	1.000	1.000	0.512
$GDFM^G$	0.913	0.990	0.544	1.053	1.010	0.479	0.894	0.914	0.554	0.971	0.990	0.428	1.034	1.045	0.447
$GDFM^S$	0.877	0.953	0.586	1.087	1.040	0.484	0.918	0.948	0.544	1.034	1.017	0.484	1.116	1.084	0.423
Nonlinear															
$FNN\text{-}DF^G$	0.900	0.978	0.567	1.030	1.002	0.484	0.889	0.916	0.535	0.918	0.962	0.447	0.982	1.023	0.423
$FNN\text{-}DF^S$	0.861	0.955	0.563	1.027	1.004	0.479	0.875	0.923	0.507	0.938	0.984	0.447	1.025	1.043	0.409
Pooling															
PLS	0.865	0.966	0.554	0.996	0.989	0.470	0.896	0.923	0.535	0.929	0.970	0.470	0.968	1.004	0.447
EW	0.868	0.966	0.554	1.034	1.001	0.484	0.875	0.914	0.554	0.948	0.981	0.451	1.018	1.035	0.433

CPI-U: ALL ITEMS

	H = 1			H = 3			H = 6			H = 9			H = 12		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.484	1.000	1.000	0.530	1.000	1.000	0.498	1.000	1.000	0.470	1.000	1.000	0.521
$GDFM^G$	0.882	0.945	0.595	0.938	0.993	0.567	0.865	0.904	0.595	0.863	0.953	0.512	1.008	1.017	0.470
$GDFM^S$	0.768	0.884	0.609	0.953	0.976	0.581	0.872	0.919	0.605	0.866	0.961	0.540	0.971	1.030	0.521
Nonlinear															
$FNN\text{-}DF^G$	0.862	0.929	0.572	0.902	0.964	0.567	0.857	0.901	0.614	0.926	0.967	0.498	0.973	0.999	0.451
$FNN\text{-}DF^S$	0.792	0.900	0.619	0.909	0.957	0.609	0.852	0.898	0.595	0.973	0.972	0.516	1.005	1.020	0.474
Pooling															
PLS	0.804	0.901	0.605	0.900	0.952	0.572	0.844	0.892	0.600	0.772	0.882	0.535	0.845	0.935	0.512
EW	0.801	0.903	0.623	0.909	0.957	0.591	0.834	0.886	0.609	0.870	0.931	0.535	0.956	0.997	0.488

Notes: Entries are relative MSE and relative MAE, relative to the benchmark AR. DA is the directional accuracy criterion. All the criteria were computed over the period 1985:1 - 2003:12-h. The pooling forecast is obtained applying the simulated forecasting exercise described in Section (3.2). The various columns correspond to forecasts of 1, 3, 6, 9 and 12-month growth, where all the multiperiod forecasts were computed using direct methods.

Table 3.6: Tests of Equal Forecast Accuracy for the Period 1984 - 2002.

PANEL A. FORECASTING HORIZON H = 1

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$
$GDFM^G$	0.966^-				0.654				0.323				0.182			
$GDFM^S$	0.991^-	0.962^-			0.892	0.734			0.820	0.947			0.485	0.921		
$FNN\text{-}DF^G$	0.881	0.017^+	0.004^+		0.317	0.001^+	0.017^+		0.068	0.001^+	0.004^+		0.060	0.195	0.073	
$FNN\text{-}DF^S$	0.952^-	0.588	0.005^+	0.929	0.285	0.155	0.002^+	0.582	0.482	0.694	0.000^+	0.957^-	0.081	0.447	0.042^+	0.716
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
$GDFM^G$	0.101				0.952^-				0.046^+				0.077			
$GDFM^S$	0.006^+	0.189			0.664	0.025^+			0.042^+	0.159			0.009^+	0.003^+		
$FNN\text{-}DF^G$	0.071	0.217	0.780		0.575	0.000^+	0.293		0.023^+	0.105	0.711		0.034^+	0.079	0.985^-	
$FNN\text{-}DF^S$	0.005^+	0.176	0.456	0.204	0.299	0.001^+	0.001^+	0.100	0.009^+	0.026^+	0.171	0.110	0.007^+	0.008^+	0.953^-	0.020^+

PANEL B. FORECASTING HORIZON H = 3

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$
$GDFM^G$	0.972^-				0.713				0.094				0.117			
$GDFM^S$	0.989^-	0.959^-			0.889	0.701			0.565	0.980^-			0.319	0.868		
$FNN\text{-}DF^G$	0.978^-	0.016^+	0.010^+		0.471	0.026^+	0.064		0.051	0.296	0.014^+		0.078	0.309	0.114	
$FNN\text{-}DF^S$	0.951^-	0.324	0.001^+	0.835	0.699	0.375	0.012^+	0.725	0.386	0.966^-	0.091	0.989^-	0.170	0.697	0.039^+	0.765
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
$GDFM^G$	0.210				0.350				0.731				0.316			
$GDFM^S$	0.013^+	0.131			0.380	0.534			0.832	0.765			0.301	0.563		
$FNN\text{-}DF^G$	0.185	0.390	0.870		0.100	0.003^+	0.164		0.685	0.205	0.114		0.175	0.035^+	0.143	
$FNN\text{-}DF^S$	0.009^+	0.128	0.642	0.123	0.159	0.266	0.002^+	0.607	0.651	0.282	0.018^+	0.438	0.121	0.256	0.023^+	0.518

PANEL C. FORECASTING HORIZON H = 6

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$	AR	$GDFM^G$	$GDFM^S$	$FNN\text{-}DF^G$
$GDFM^G$	0.986^-				0.999^-				0.732				0.269			
$GDFM^S$	0.989^-	0.790			0.965^-	0.110			0.907	0.980^-			0.290	0.565		
$FNN\text{-}DF^G$	0.960^-	0.001^+	0.040^+		0.999^-	0.015^+	0.700		0.703	0.166	0.010^+		0.252	0.512	0.439	
$FNN\text{-}DF^S$	0.887	0.064	0.001^+	0.488	0.957^-	0.013^+	0.050^+	0.059	0.892	0.958^-	0.084	0.985^-	0.300	0.592	0.566	0.597
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
$GDFM^G$	0.240				0.415				0.140				0.191			
$GDFM^S$	0.018^+	0.048^+			0.804	0.894			0.209	0.659			0.211	0.541		
$FNN\text{-}DF^G$	0.306	0.886	0.976^-		0.379	0.403	0.077		0.091	0.424	0.324		0.159	0.296	0.410	
$FNN\text{-}DF^S$	0.031^+	0.098	0.970^-	0.047^+	0.680	0.779	0.006^+	0.827	0.059	0.364	0.122	0.378	0.148	0.427	0.197	0.473

– Continued on next page –

PANEL D. FORECASTING HORIZON H = 9

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**
GDFM^G	0.987^-				1.000^-				0.925				0.215			
GDFM^S	0.990^-	0.739			0.989^-	0.084			0.966^-	0.911			0.111	0.279		
FNN-DF^G	0.893	0.006^+	0.053		1.000^-	0.000^+	0.702		0.924	0.591	0.182		0.303	0.933	0.847	
FNN-DF^S	0.751	0.041^+	0.003^+	0.266	0.987^-	0.009^+	0.022^+	0.068	0.970^-	0.892	0.576	0.920	0.213	0.557	0.844	0.383
	MONEY SUPPLY - M2				**INTEREST RATE**				**PRODUCER PRICE INDEX**				**CPI-U: ALL ITEMS**			
GDFM^G	0.031^+				0.171				0.455				0.266			
GDFM^S	0.002^+	0.046^+			0.744	0.918			0.617	0.823			0.254	0.406		
FNN-DF^G	0.086	0.971^-	0.965^-		0.110	0.232	0.067		0.238	0.029^+	0.041^+		0.310	0.720	0.708	
FNN-DF^S	0.001^+	0.109	0.579	0.034^+	0.612	0.869	0.039^+	0.901	0.260	0.191	0.029^+	0.574	0.424	0.754	0.800	0.670

PANEL E. FORECASTING HORIZON H = 12

	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
Models	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**	**AR**	**GDFM^G**	**GDFM^S**	**FNN-DF^G**
GDFM^G	0.967^-				1.000^-				0.948				0.095			
GDFM^S	0.956^-	0.576			0.982^-	0.024^+			0.953^-	0.551			0.045^+	0.209		
FNN-DF^G	0.716	0.018^+	0.019^+		1.000^-	0.001^+	0.889		0.942	0.514	0.475		0.111	0.782	0.840	
FNN-DF^S	0.504	0.032^+	0.005^+	0.205	0.980^-	0.005^+	0.111	0.033^+	0.947	0.415	0.373	0.395	0.083	0.406	0.793	0.323
	MONEY SUPPLY - M2				**INTEREST RATE**				**PRODUCER PRICE INDEX**				**CPI-U: ALL ITEMS**			
GDFM^G	0.339				0.569				0.657				0.952^-			
GDFM^S	0.176	0.212			0.895	0.913			0.821	0.926			0.960^-	0.776		
FNN-DF^G	0.351	0.477	0.736		0.490	0.150	0.069		0.513	0.028^+	0.015		0.989^-	0.653	0.461	
FNN-DF^S	0.097	0.082	0.284	0.123	0.885	0.902	0.139	0.926	0.644	0.424	0.043^+	0.814	0.976^-	0.733	0.634	0.626

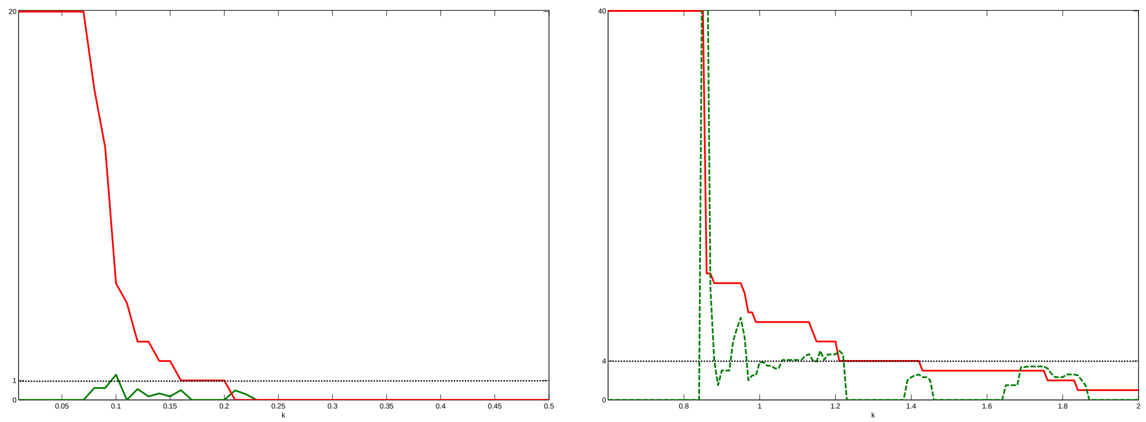
Notes: Results of test of equal forecast accuracy as described in Section 3.5.1. Entries are the p-values of the test of equal unconditional predictive ability for the test proposed by Giacomini and White for the forecast in the corresponding row and column. A plus or minus sign indicates that the method in the row outperforms or underperforms the method in the column at the 5% significance level. For example, for Personal Income at the $H = 1$ horizon, equal forecast accuracy of the GDFM^G and the AR is rejected with a p-value of 0.966, then AR outperforms GDFM^G.

3.7 Changes in U.S. Economic Time Series

In this chapter we present the same analysis as in Section (3.5); however, the difference between the two sections concerns the period in which it was conducted. As seen in Section (3.5) the period under study covers the time span between 1974 and 2002. Here, the analysis will be performed for the period between 2003 and 2008. The reasons for extending the analysis to this period are varied. In general, the vast majority of papers concerning the use of factor models in predicting US macroeconomic time series, are based on and compared to the dataset presented in the previous section found here. It is now established that the forecasts made with these types of models have better performance in the period between 1970 and 1985, compared to very simple models, such as autoregressive models. This evidence, however, changes in the period following the 1985 to 2002, namely during the so-called Great Moderation, where the use of many predictors does not improve the forecasts obtained with very simple models, especially for key series such as Industrial Production Index or Consumer Price Index. For this reason, it will be interesting to see whether the forecasting ability of factor models for the period between 2003 and 2008 are still marked by poor forecasting or if it is reasonable to think that something has changed.

The dataset has been updated taking into account the changes that have elapsed since 2003. In particular the new dataset is no longer composed of 131 series as the original but 130. In this extended analysis the monetary aggregate M3 is excluded; the M3 series includes all physical currency and deposits in checking accounts, deposits in savings accounts, certificates of deposit, institutional money market accounts, repurchase agreements, and other large liquid assets that do not circulate very often. However, On March 23 2006, the Board of Governors of the Federal Reserve System ceased publication of the M3 monetary aggregate, because it does not appear to convey any additional information about economic activity that is not already embodied in M2 and has not played a role in the monetary policy process for many years.
In order to determine the number of factors, we proceed as in Section 1.5. In Figure (3.2) the criterion proposed by Hallin and Liska (2007) sets the number of dynamic factors equal to 1. Whereas the criterion proposed by Alessi et al. (2007) sets the number of static factors equal to 4.

The results obtained confirm that there is no change in the forecasting ability of models GDFM



(a) Number of Dynamic Factors.

(b) Number of Static Factors.

Figure 3.2: Plot of the Criteria proposed to determine the number of factors.

and FNN-DF. Comparing the Tables 3.5 and 3.6 and Tables 3.7 and 3.8, this lack of change in forecasting ability is clear, as the results for the period between 1985 and 2003 are in line with the results for the new period under study.

Table 3.7: Forecasting Results for the period between 2003 - 2008

PERSONAL INCOME

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.561	1.000	1.000	0.526	1.000	1.000	0.649	1.000	1.000	0.632	1.000	1.000	0.561
GDFM^G	0.995	0.960	0.649	0.814	0.902	0.667	0.906	0.952	0.649	0.973	0.992	0.667	1.150	1.115	0.561
GDFM^S	1.016	1.029	0.561	0.963	1.023	0.526	1.254	1.108	0.491	1.377	1.140	0.544	1.309	1.171	0.509
Nonlinear															
FNN-DF^G	0.954	0.950	0.579	0.803	0.873	0.632	0.931	0.980	0.649	0.971	0.974	0.702	1.099	1.072	0.561
FNN-DF^S	0.967	0.976	0.561	0.871	0.931	0.561	1.190	1.068	0.509	1.289	1.108	0.526	1.260	1.132	0.491
Pooling															
PLS	0.973	0.984	0.544	0.861	0.930	0.544	0.994	0.978	0.561	1.086	0.996	0.614	1.154	1.075	0.544
EW	0.967	0.959	0.632	0.841	0.915	0.579	1.017	0.993	0.544	1.102	1.022	0.632	1.184	1.108	0.509

REAL CONSUMPTION

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.421	1.000	1.000	0.561	1.000	1.000	0.614	1.000	1.000	0.491	1.000	1.000	0.579
GDFM^G	0.944	0.989	0.597	0.948	0.984	0.649	1.197	1.096	0.526	1.249	1.098	0.439	1.369	1.304	0.561
GDFM^S	0.988	0.985	0.526	0.973	0.985	0.561	1.188	1.091	0.509	1.262	1.115	0.421	1.425	1.321	0.526
Nonlinear															
FNN-DF^G	0.926	0.966	0.561	0.912	0.977	0.614	1.261	1.093	0.509	1.250	1.093	0.456	1.330	1.270	0.561
FNN-DF^S	0.970	0.975	0.509	0.925	0.931	0.597	1.250	1.093	0.491	1.250	1.091	0.421	1.362	1.279	0.526
Pooling															
PLS	0.957	0.980	0.526	0.846	0.893	0.614	1.257	1.097	0.509	1.245	1.085	0.456	1.324	1.258	0.544
EW	0.950	0.977	0.544	0.781	0.867	0.632	1.214	1.087	0.526	1.240	1.084	0.439	1.352	1.283	0.561

INDUSTRIAL PRODUCTION INDEX

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.614	1.000	1.000	0.614	1.000	1.000	0.561	1.000	1.000	0.614	1.000	1.000	0.754
GDFM^G	0.996	0.994	0.632	1.008	0.998	0.667	0.948	0.999	0.632	1.072	1.187	0.737	1.248	1.315	0.754
GDFM^S	0.939	0.968	0.614	1.061	1.019	0.632	0.915	0.994	0.649	1.196	1.156	0.667	1.333	1.170	0.772
Nonlinear															
FNN-DF^G	0.940	0.943	0.684	0.970	0.963	0.667	0.652	0.920	0.667	0.984	1.077	0.719	0.999	1.135	0.737
FNN-DF^S	0.907	0.917	0.649	1.029	0.989	0.632	0.735	0.935	0.667	1.032	1.055	0.684	1.073	1.096	0.702
Pooling															
PLS	0.940	0.953	0.632	1.005	0.977	0.667	0.927	0.963	0.667	1.039	1.091	0.702	1.071	1.108	0.702
EW	0.939	0.949	0.632	1.007	0.986	0.667	0.919	0.951	0.667	1.023	1.098	0.684	1.100	1.148	0.754

UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.649	1.000	1.000	0.754	1.000	1.000	0.737	1.000	1.000	0.702	1.000	1.000	0.807
GDFM^G	0.922	0.993	0.719	0.693	0.888	0.825	0.610	0.786	0.807	0.722	0.805	0.790	0.648	0.710	0.807
GDFM^S	0.925	0.989	0.719	0.672	0.877	0.842	0.553	0.808	0.825	0.680	0.798	0.772	0.650	0.756	0.807
Nonlinear															
FNN-DF^G	0.918	0.975	0.702	0.712	0.881	0.807	0.610	0.790	0.807	0.708	0.794	0.772	0.665	0.722	0.790
FNN-DF^S	0.918	0.977	0.702	0.717	0.873	0.825	0.566	0.812	0.807	0.665	0.781	0.790	0.655	0.758	0.772
Pooling															
PLS	0.914	0.974	0.754	0.706	0.869	0.825	0.608	0.842	0.772	0.683	0.797	0.790	0.701	0.800	0.772
EW	0.910	0.983	0.719	0.692	0.873	0.842	0.578	0.796	0.807	0.686	0.788	0.790	0.645	0.725	0.807

– Continued on next page –

MONEY SUPPLY - M2

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.421	1.000	1.000	0.333	1.000	1.000	0.351	1.000	1.000	0.351	1.000	1.000	0.298
GDFM^G	1.055	1.010	0.509	1.264	1.033	0.544	1.088	1.063	0.509	0.870	0.897	0.561	0.688	0.820	0.439
GDFM^S	1.183	1.083	0.509	1.137	1.002	0.597	1.361	1.150	0.526	1.077	0.962	0.544	0.757	0.796	0.439
Nonlinear															
FNN-DF^G	1.002	0.987	0.474	1.150	0.989	0.491	1.062	1.061	0.439	0.902	0.963	0.474	0.760	0.900	0.386
FNN-DF^S	1.039	1.020	0.474	1.042	0.971	0.526	1.233	1.124	0.491	1.066	1.015	0.491	0.830	0.893	0.333
Pooling															
PLS	1.047	1.017	0.474	1.129	0.991	0.561	1.171	1.101	0.509	0.869	0.928	0.509	0.674	0.816	0.404
EW	1.055	1.021	0.474	1.134	0.993	0.561	1.164	1.088	0.509	0.957	0.952	0.526	0.731	0.845	0.439

INTEREST RATE: FEDERAL FUNDS (EFFECTIVE)

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.912	1.000	1.000	0.895	1.000	1.000	0.684	1.000	1.000	0.509	1.000	1.000	0.316
GDFM^G	0.702	0.848	0.877	0.565	0.784	0.860	0.550	0.742	0.860	0.418	0.634	0.807	0.405	0.624	0.825
GDFM^S	0.653	0.853	0.877	0.515	0.758	0.860	0.553	0.710	0.895	0.403	0.627	0.842	0.423	0.634	0.877
Nonlinear															
FNN-DF^G	0.721	0.864	0.842	0.587	0.801	0.860	0.569	0.758	0.860	0.439	0.654	0.807	0.440	0.660	0.825
FNN-DF^S	0.669	0.854	0.895	0.538	0.766	0.860	0.533	0.701	0.895	0.419	0.641	0.860	0.442	0.652	0.877
Pooling															
PLS	0.687	0.843	0.895	0.532	0.762	0.877	0.577	0.762	0.895	0.428	0.650	0.860	0.448	0.665	0.860
EW	0.674	0.843	0.860	0.543	0.772	0.877	0.542	0.721	0.895	0.414	0.637	0.842	0.420	0.639	0.842

PRODUCER PRICE INDEX

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.597	1.000	1.000	0.561	1.000	1.000	0.632	1.000	1.000	0.632	1.000	1.000	0.526
GDFM^G	1.241	1.114	0.561	1.016	1.013	0.509	1.200	1.079	0.614	1.278	1.206	0.526	1.505	1.171	0.421
GDFM^S	1.279	1.136	0.509	1.032	1.040	0.491	1.270	1.139	0.649	1.459	1.337	0.509	1.731	1.273	0.404
Nonlinear															
FNN-DF^G	1.140	1.074	0.526	0.980	0.989	0.526	1.103	1.040	0.632	1.192	1.151	0.509	1.420	1.135	0.421
FNN-DF^S	1.162	1.089	0.491	0.993	1.004	0.509	1.158	1.090	0.632	1.357	1.278	0.509	1.634	1.225	0.404
Pooling															
PLS	1.195	1.099	0.491	0.995	1.005	0.509	1.151	1.070	0.632	1.321	1.250	0.509	1.540	1.179	0.404
EW	1.197	1.100	0.526	0.998	1.007	0.544	1.167	1.071	0.632	1.298	1.235	0.509	1.544	1.182	0.421

CPI-U: ALL ITEMS

	H = 1			**H = 3**			**H = 6**			**H = 9**			**H = 12**		
	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA	rMSE	rMAE	DA
Benchmark															
AR	1.000	1.000	0.702	1.000	1.000	0.597	1.000	1.000	0.719	1.000	1.000	0.737	1.000	1.000	0.579
GDFM^G	0.916	0.928	0.702	1.041	0.994	0.667	0.842	0.864	0.667	1.141	1.077	0.719	1.515	1.187	0.561
GDFM^S	0.970	0.975	0.632	1.033	1.031	0.597	0.837	0.901	0.737	1.248	1.139	0.667	1.634	1.232	0.526
Nonlinear															
FNN-DF^G	0.960	1.006	0.702	1.010	1.017	0.561	0.845	0.871	0.684	1.112	1.056	0.754	1.476	1.177	0.579
FNN-DF^S	0.986	1.028	0.614	1.011	1.042	0.491	0.836	0.897	0.719	1.215	1.111	0.649	1.591	1.211	0.561
Pooling															
PLS	0.958	0.990	0.667	1.007	1.014	0.526	0.824	0.883	0.702	1.197	1.100	0.684	1.538	1.193	0.579
EW	0.947	0.974	0.684	1.015	1.015	0.597	0.828	0.881	0.684	1.164	1.087	0.719	1.541	1.193	0.579

Notes: Entries are relative MSE and relative MAE, relative to the benchmark AR. DA is the directional accuracy criterion. All the criteria were computed over the period 2003:1 - 2009:12-h. The pooling forecast is obtained applying the simulated forecasting exercise described in Section (3.2). The various columns correspond to forecasts of 1, 3, 6, 9 and 12-month growth, where all the multiperiod forecasts were computed using direct methods.

Table 3.8: Tests of Equal Forecast Accuracy for the Period 2003 - 2008.

PANEL A. FORECASTING HORIZON H = 1

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}
GDFM^{G}	0.500				0.801				0.286				0.033^{+}			
GDFM^{S}	0.086	0.782			0.929	0.865			0.139	0.235			0.035^{+}	0.467		
FNN-DF^{G}	0.091	0.129	0.093		0.541	0.089	0.020^{+}		0.070	0.019^{+}	0.498		0.015^{+}	0.755	0.747	
FNN-DF^{S}	0.096	0.190	0.139	0.907	0.749	0.339	0.086	0.883	0.029^{+}	0.040	0.255	0.193	0.016^{+}	0.812	0.814	0.868
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^{G}	0.083				0.002^{+}				0.957^{-}				0.043^{+}			
GDFM^{S}	0.291	0.711			0.002^{+}	0.151			0.961^{-}	0.607			0.188	0.865		
FNN-DF^{G}	0.057	0.480	0.292		0.002^{+}	0.960^{-}	0.964^{-}		0.922	0.086	0.076		0.094	0.950^{-}	0.548	
FNN-DF^{S}	0.113	0.574	0.115	0.581	0.002^{+}	0.249	0.881	0.101	0.925	0.161	0.069	0.645	0.193	0.899	0.697	0.626

PANEL B. FORECASTING HORIZON H = 3

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}
GDFM^{G}	0.206				0.327				0.247				0.012^{+}			
GDFM^{S}	0.374	0.738			0.408	0.802			0.356	0.788			0.016^{+}	0.348		
FNN-DF^{G}	0.083	0.119	0.114		0.226	0.466	0.336		0.258	0.609	0.378		0.015^{+}	0.958^{-}	0.952^{-}	
FNN-DF^{S}	0.102	0.448	0.013^{+}	0.698	0.323	0.630	0.422	0.786	0.393	0.812	0.609	0.883	0.017^{+}	0.966^{-}	0.962^{-}	0.825
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^{G}	0.355				0.003^{+}				0.714				0.725			
GDFM^{S}	0.179	0.241			0.003^{+}	0.178			0.750	0.760			0.751	0.458		
FNN-DF^{G}	0.228	0.084	0.626		0.003^{+}	0.953^{-}	0.957^{-}		0.703	0.324	0.254		0.614	0.179	0.133	
FNN-DF^{S}	0.078	0.136	0.050^{+}	0.203	0.002^{+}	0.309	0.865	0.124	0.760	0.502	0.335	0.880	0.623	0.239	0.147	0.490

PANEL C. FORECASTING HORIZON H = 6

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}	AR	GDFM^{G}	GDFM^{S}	FNN-DF^{G}
GDFM^{G}	0.445				0.976^{-}				0.174				0.010^{+}			
GDFM^{S}	0.755	0.740			0.955^{-}	0.689			0.304	0.932			0.014^{+}	0.645		
FNN-DF^{G}	0.489	0.647	0.267		0.967^{-}	0.534	0.396		0.173	0.497	0.073		0.010^{+}	0.887	0.843	
FNN-DF^{S}	0.840	0.788	0.644	0.799	0.970^{-}	0.861	0.687	0.793	0.258	0.827	0.274	0.927	0.017^{+}	0.846	0.919	0.670
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^{G}	0.752				0.009^{+}				0.729				0.066			
GDFM^{S}	0.633	0.342			0.013^{+}	0.672			0.782	0.714			0.045^{+}	0.234		
FNN-DF^{G}	0.722	0.100	0.574		0.008^{+}	0.852	0.464		0.633	0.077	0.056		0.051	0.533	0.740	
FNN-DF^{S}	0.566	0.184	0.093	0.256	0.009^{+}	0.617	0.362	0.464	0.703	0.254	0.071	0.806	0.031^{+}	0.230	0.576	0.213

– Continued on next page –

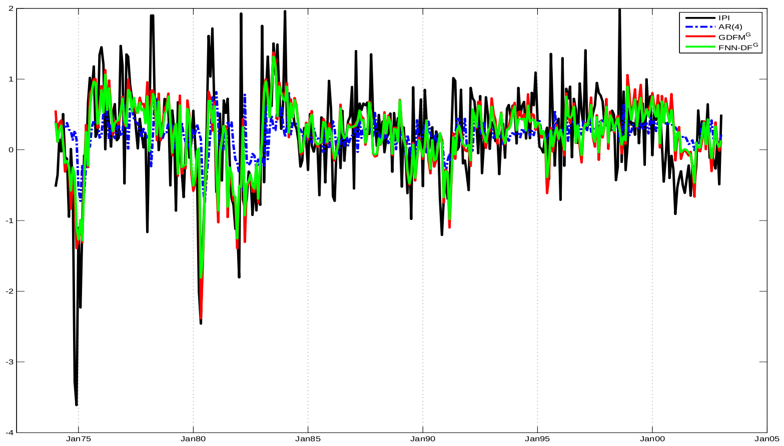
PANEL D. FORECASTING HORIZON H = 9

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^G	GDFM^S	FNN-DF^G	AR	GDFM^G	GDFM^S	FNN-DF^G	AR	GDFM^G	GDFM^S	FNN-DF^G	AR	GDFM^G	GDFM^S	FNN-DF^G
GDFM^G	0.160				0.989$^-$				0.171				0.006$^+$			
GDFM^S	0.587	0.772			0.971$^-$	0.278			0.554	0.963$^-$			0.013$^+$	0.647		
FNN-DF^G	0.059	0.148	0.115		0.986$^-$	0.088	0.420		0.066	0.115	0.005$^+$		0.004$^+$	0.872	0.570	
FNN-DF^S	0.384	0.652	0.127	0.852	0.972$^-$	0.152	0.252	0.475	0.250	0.708	0.004$^+$	0.961$^-$	0.007$^+$	0.745	0.635	0.486
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^G	0.584				0.004$^+$				0.784				0.591			
GDFM^S	0.452	0.333			0.007$^+$	0.760			0.993$^-$	0.969$^-$			0.801	0.767		
FNN-DF^G	0.574	0.345	0.648		0.003$^+$	0.845	0.410		0.588	0.028$^+$	0.001$^+$		0.421	0.039$^+$	0.087	
FNN-DF^S	0.395	0.256	0.196	0.269	0.005$^+$	0.857	0.685	0.727	0.982$^-$	0.779	0.006$^+$	0.991$^-$	0.726	0.665	0.143	0.866

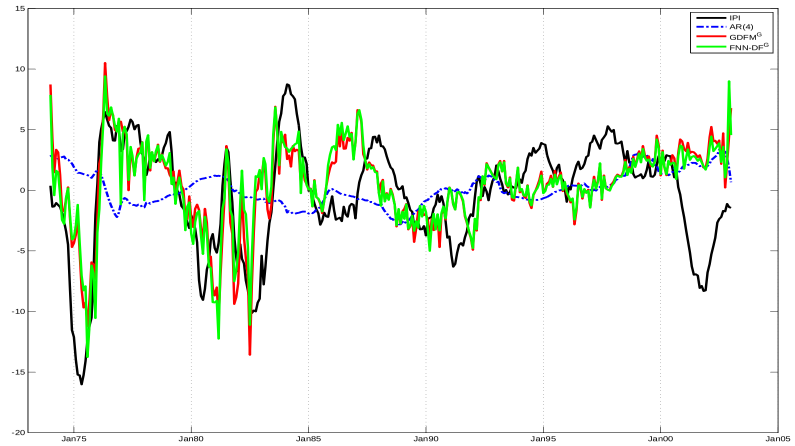
PANEL E. FORECASTING HORIZON H = 12

Models	PERSONAL INCOME				REAL CONSUMPTION				INDUSTRIAL PRODUCTION				UNEMPLOYMENT RATE			
	AR	GDFM^G	GDFM^S	FNN-DF^G	AR	GDFM^G	GDFM^S	FNN-DF^G	AR	GDFM^G	GDFM^S	FNN-DF^G	AR	GDFM^G	GDFM^S	FNN-DF^G
GDFM^G	0.311				0.999$^-$				0.973$^-$				0.017$^+$			
GDFM^S	0.204	0.283			0.995$^-$	0.235			0.980$^-$	0.956$^-$			0.189	0.861		
FNN-DF^G	0.080	0.022$^+$	0.056		0.999$^-$	0.013$^+$	0.304		0.356	0.004$^+$	0.007$^+$		0.008$^+$	0.369	0.162	
FNN-DF^S	0.091	0.074	0.033$^+$	0.466	0.995$^-$	0.081	0.046$^+$	0.446	0.862	0.529	0.005$^+$	0.959$^-$	0.053	0.739	0.085	0.736
	MONEY SUPPLY - M2				INTEREST RATE				PRODUCER PRICE INDEX				CPI-U: ALL ITEMS			
GDFM^G	0.584				0.000$^+$				0.892				0.980$^-$			
GDFM^S	0.531	0.457			0.001$^+$	0.955$^-$			0.992$^-$	0.995$^-$			0.977$^-$	0.837		
FNN-DF^G	0.565	0.330	0.516		0.000$^+$	0.849	0.207		0.865	0.154	0.001$^+$		0.955$^-$	0.057	0.018$^+$	
FNN-DF^S	0.439	0.353	0.180	0.377	0.000$^+$	0.970$^-$	0.460	0.972$^-$	0.985$^-$	0.961$^-$	0.007$^+$	0.996$^-$	0.970$^-$	0.711	0.115	0.977$^-$

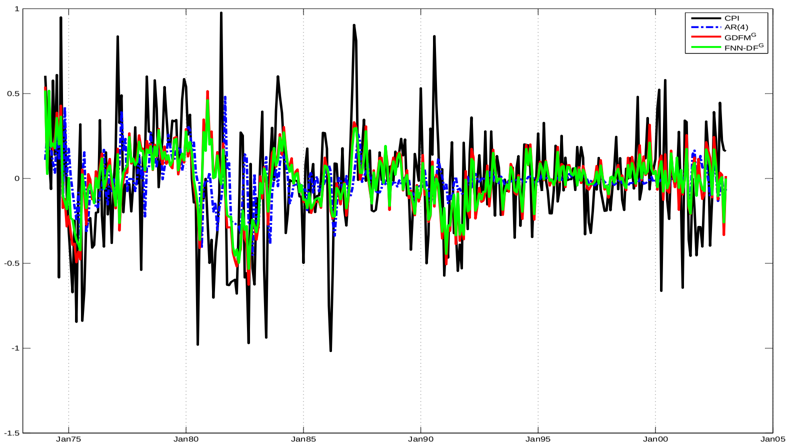
Notes: Results of tests of equal forecast accuracy as described in Section 3.5.1. Entries are the p-values of the test of equal forecast accuracy for the forecast in the corresponding row and column. A plus or minus sign indicates that the method in the row outperforms or underperforms the method in the column at the 5% significance level. For example, for Unemployment Rate at the H = 1 horizon, equal unconditional predictive ability of the GDFM^G and the AR is rejected with a p-value of 0.034, then GDFM^G outperforms AR.



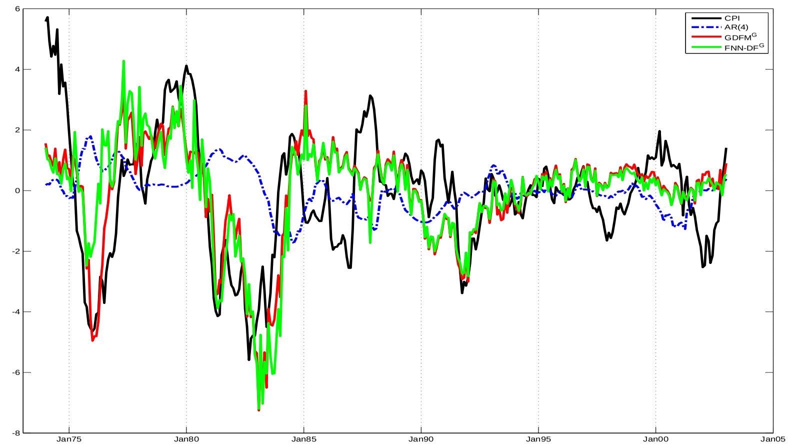
(a) Plot of forecasting results of Industrial Production Index for H=1.



(b) Plot of forecasting results of Industrial Production Index for H=12.



(c) Plot of forecasting results of Consumer Price Index for H=1.



(d) Plot of forecasting results of Consumer Price Index for H=12.

Figure 3.3: Plot of forecasting results of Industrial Production (top) and Consumer Price Index (bottom). **Black line**: true value, **Dotted Blue line**: AR, **Red line**: Forecasting for GDFM^G and **Green line**: Forecasting for FNN-DF^G.

Conclusions

In this dissertation we introduced a new forecasting technique for macroeconomics and financial variables which was obtained from the combination of two different methods of data modeling, both well-established in economic forecasting literature. The methods under study relate to factor analysis and artificial neural networks of the FNN-DF type. The forecasting ability of the model FNN-DF was studied for eight monthly series of the U.S. economy, grouped in real and nominal variables, for the period between 1974 and 2008. In the introduction to Chapter 3 we posed three questions which were designed to verify whether the use of nonlinear techniques is superior and if such an improvement is significant enough to justify their use, lastly we asked if the combined techniques provide a further improvement.

The immediate conclusion which comes from the empirical study is that the FNN-DF model has the same forecasting ability as linear models even when the linearity seems prevalent in the data. However, a careful and comparative analysis of FNN-DF and GDFM can not be comprehensive, because, as it has often been noted, linear models do not provide improved accuracy in forecasting with respect to the benchmark in the dataset examined. It is interesting to note that in constructing FNN-DF further complexities , compared to the GDFM, are not required. In other terms we tried to find a compromise between the numerous techniques already established to construct FNN and their forecasting accuracy. In this sense, the Bayesian Regularization uses less subjective evaluation than other methods considered and was therefore the most advantageous choice.

In general the results obtained do not indicate FNN-DF models have superior performance compared to GDFM, except in some cases such as Real Consumption and Interest Rate. In fact there is no model that dominates the others as can be seen in Tables 3.2 to 3.8; moreover, the models having some significant differences, are not the same for all variables and for all forecasting horizons. Empirical evidence in favor of non-linearity exists only for the period between 1974 and 1985; afterwards, as analyzed by the growing literature on the Great Moderation, the forecasting ability of complex models is reduced drastically.

Lastly the issue of whether the use of combined techniques provides an improvement over each single model, namely when the forecaster is uncertain about which model to use, should he pool the results or verify which of these is most appropriate for the problem at hand? The results suggest that combining the forecasts provides some improvement only for long-term horizons. Long-term forecasting is effected by high uncertainty and the error made in the long run can be mitigated by the forecast combination. A further distinction can be made with respect to the two techniques used. On the one hand the Equal Weight technique does not provide benefits. Whereas in the case of PLS, this technique shows good ability to weight properly different forecast. However, we have combined only four models, the gain may be more evident if the number of forecast is large.

The results obtained in this dissertation are based on the estimation of the factors using linear techniques. This, in general, can have limiting effects, when some of the variables considered are strongly nonlinear. For future research it could be interesting to implement a more general dimensionality reduction technique, which takes into account both the linear and the nonlinear correlations between the series used, without imposing any restrictions on the character of nonlinearity in the data.

Another possibility is to implement a new forecasting technique using the test proposed by Giacomini and White (2006) to automatically determine which model is better to predict the given time period. To our knowledge this particular model selection has not been applied in selecting linear or nonlinear models, for example in the neural networks framework.

Appendix A

Data Sources

	Mnemonic	Description	Transf.
1	a0m052	Personal Income (Ar, Bil. Chain 2000 $)	Δln
2	a0m051	Personal Income Less Transfer Payments (Ar, Bil. Chain 2000 $)	Δln
3	a0m224_r	Real Consumption (Ac) A0M224Gmdc	Δln
4	a0m057	Manufacturing And Trade Sales (Mil. Chain 1996 $)	Δln
5	a0m059	Sales Of Retail Stores (Mil. Chain 2000 $)	Δln
6	ips10	Industrial Production Index - Total Index	Δln
7	ips11	Industrial Production Index - Products, Total	Δln
8	ips299	Industrial Production Index - Final Products	Δln
9	ips12	Industrial Production Index - Consumer Goods	Δln
10	ips13	Industrial Production Index - Durable Consumer Goods	Δln
11	ips18	Industrial Production Index - Nondurable Consumer Goods	Δln
12	ips25	Industrial Production Index - Business Equipment	Δln
13	ips32	Industrial Production Index - Materials	Δln
14	ips34	Industrial Production Index - Durable Goods Materials	Δln
15	ips38	Industrial Production Index - Nondurable Goods Materials	Δln
16	ips43	Industrial Production Index - Manufacturing (Sic)	Δln
17	ips307	Industrial Production Index - Residential Utilities	Δln
18	ips306	Industrial Production Index - Fuels	Δln
19	pmp	Napm Production Index (Percent)	lv
20	a0m082	Capacity Utilization (Mfg)	Δlv
21	lhel	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)	Δlv
22	lhelx	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf	Δlv
23	lhem	Civilian Labor Force: Employed, Total (Thous.,Sa)	Δln
24	lhnag	Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)	Δln
25	lhur	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)	Δlv
26	lhu680	Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)	Δlv
27	lhu5	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)	Δlv
28	lhu14	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)	Δlv
29	lhu15	Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)	Δln
30	lhu26	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)	Δln
31	lhu27	Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)	Δln
32	a0m005	Average Weekly Initial Claims, Unemploy. Insurance (Thous.)	Δln
33	ces002	Employees On Nonfarm Payrolls - Total Private	Δln

34	ces003	Employees On Nonfarm Payrolls - Goods-Producing	Δln
35	ces006	Employees On Nonfarm Payrolls - Mining	Δln
36	ces011	Employees On Nonfarm Payrolls - Construction	Δln
37	ces015	Employees On Nonfarm Payrolls - Manufacturing	Δln
38	ces017	Employees On Nonfarm Payrolls - Durable Goods	Δln
39	ces033	Employees On Nonfarm Payrolls - Nondurable Goods	Δln
40	ces046	Employees On Nonfarm Payrolls - Service-Providing	Δln
41	ces048	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities	Δln
42	ces049	Employees On Nonfarm Payrolls - Wholesale Trade	Δln
43	ces053	Employees On Nonfarm Payrolls - Retail Trade	Δln
44	ces088	Employees On Nonfarm Payrolls - Financial Activities	Δln
45	ces140	Employees On Nonfarm Payrolls - Government	Δln
46	a0m048	Employee Hours In Nonag. Establishments (Ar, Bil. Hours)	Δln
47	ces151	Avg Weekly Hours Of Prod. Or Nonsuperv. Workers On Private Nonfarm	lv
48	ces155	Avg Weekly Hours Of Prod. Or Nonsuperv. Workers On Private Nonfarm	Δlv
49	a0m001	Average Weekly Hours, Mfg. (Hours)	lv
50	pmemp	Napm Employment Index (Percent)	lv
51	hsfr	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-)(Thous.,Sa	ln
52	hsne	Housing Starts:Northeast (Thous.U.)S.A.	ln
53	hsmw	Housing Starts:Midwest(Thous.U.)S.A.	ln
54	hssou	Housing Starts:South (Thous.U.)S.A.	ln
55	hswst	Housing Starts:West (Thous.U.)S.A.	ln
56	hsbr	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)	ln
57	hsbne	Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A	ln
58	hsbmw	Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.	ln
59	hsbsou	Houses Authorized By Build. Permits:South(Thou.U.)S.A.	ln
60	hsbwst	Houses Authorized By Build. Permits:West(Thou.U.)S.A.	ln
61	pmi	Purchasing Managers' Index (Sa)	lv
62	pmno	Napm New Orders Index (Percent)	lv
63	pmdel	Napm Vendor Deliveries Index (Percent)	lv
64	pmnv	Napm Inventories Index (Percent)	lv
65	a0m008	Mfrs' New Orders, Consumer Goods And Materials (Bil. Chain 1982 $)	Δln
66	a0m007	Mfrs' New Orders, Durable Goods Industries (Bil. Chain 2000 $)	Δln
67	a0m027	Mfrs' New Orders, Nondefense Capital Goods (Mil. Chain 1982 $)	Δln
68	a1m092	Mfrs' Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $)	Δln
69	a0m070	Manufacturing And Trade Inventories (Bil. Chain 2000 $)	Δln
70	a0m077	Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000 $)	Δlv
71	fm1	Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ck'Able Dep)(Bil$,Sa)	Δ^2ln
72	fm2	Money Stock:M2(M1+O'Nite Rps,Euro$,Gp&Bd Mmmfs&Sav&Sm Time ,	Δ^2ln
73	fm3	Money Stock: M3(M2+Lg Time Dep,Term Rp'S&Inst Only Mmmfs) Sa	Δ^2ln
74	fm2dq	Money Supply - M2 In 1996 Dollars (Bci)	Δln
75	fmfba	Monetary Base, Adj For Reserve Requirement Changes(Mil$,Sa)	Δ^2ln
76	fmrra	Depository Inst Reserves:Total,Adj For Reserve Req Chgs(Mil$,Sa)	Δ^2ln
77	fmrnba	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil$,Sa)	Δ^2ln

78	fclnq	Commercial & Industrial Loans Oustanding In 1996 Dollars (Bci)	Δ^2ln
79	fclbmc	Wkly Rp Lg Com'L Banks:Net Change Com'L & Indus Loans(Bil$,Saar)	Δln
80	ccinrv	Consumer Credit Outstanding - Nonrevolving(G19)	Δ^2ln
81	a0m095	Ratio, Consumer Installment Credit To Personal Income (Pct.)	Δlv
82	fspcom	S&P'S Common Stock Price Index: Composite (1941-43=10)	Δln
83	fspin	S&P'S Common Stock Price Index: Industrials (1941-43=10)	Δln
84	fsdxp	S&P'S Composite Common Stock: Dividend Yield (% Per Annum)	Δlv
85	fspxe	S&P'S Composite Common Stock: Price-Earnings Ratio (%,Nsa)	Δln
86	fyff	Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)	Δlv
87	cp90	Cmmercial Paper Rate (Ac)	Δlv
88	fygm3	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.(% Per Ann,Nsa)	Δlv
89	fygm6	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.(% Per Ann,Nsa)	Δlv
90	fygt1	Interest Rate: U.S.Treasury Const Maturities,1-Yr.(% Per Ann,Nsa)	Δlv
91	fygt5	Interest Rate: U.S.Treasury Const Maturities,5-Yr.(% Per Ann,Nsa)	Δlv
92	fygt10	Interest Rate: U.S.Treasury Const Maturities,10-Yr.(% Per Ann,Nsa)	Δlv
93	fyaaac	Bond Yield: Moody'S Aaa Corporate (% Per Annum)	Δlv
94	fybaac	Bond Yield: Moody'S Baa Corporate (% Per Annum)	Δlv
95	scp90	Cp90-Fyff	lv
96	sfygm3	Fygm3-Fyff	lv
97	sfygm6	Fygm6-Fyff	lv
98	sfygt1	Fygt1-Fyff	lv
99	sfygt5	Fygt5-Fyff	lv
100	sfygt10	Fygt10-Fyff	lv
101	sfyaaac	Fyaaac-Fyff	lv
102	sfybaac	Fybaac-Fyff	lv
103	exrus	United States;Effective Exchange Rate(Merm)(Index No.)	Δln
104	exrsw	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)	Δln
105	exrjan	Foreign Exchange Rate: Japan (Yen Per U.S.$)	Δln
106	exruk	Foreign Exchange Rate: United Kingdom (Cents Per Pound)	Δln
107	exrcan	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)	Δln
108	pwfsa	Producer Price Index: Finished Goods (82=100,Sa)	Δ^2ln
109	pwfcsa	Producer Price Index:Finished Consumer Goods (82=100,Sa)	Δ^2ln
110	pwimsa	Producer Price Index:Intermed Mat.Supplies & Components(82=100,Sa)	Δ^2ln
111	pwcmsa	Producer Price Index:Crude Materials (82=100,Sa)	Δ^2ln
112	psm99q	Index Of Sensitive Materials Prices (1990=100)(Bci-99A)	Δ^2ln
113	pmcp	Napm Commodity Prices Index (Percent)	lv
114	punew	Cpi-U: All Items (82-84=100,Sa)	Δ^2ln
115	pu83	Cpi-U: Apparel & Upkeep (82-84=100,Sa)	Δ^2ln
116	pu84	Cpi-U: Transportation (82-84=100,Sa)	Δ^2ln
117	pu85	Cpi-U: Medical Care (82-84=100,Sa)	Δ^2ln
118	puc	Cpi-U: Commodities (82-84=100,Sa)	Δ^2ln
119	pucd	Cpi-U: Durables (82-84=100,Sa)	Δ^2ln
120	pus	Cpi-U: Services (82-84=100,Sa)	Δ^2ln
121	puxf	Cpi-U: All Items Less Food (82-84=100,Sa)	Δ^2ln

122	puxhs	Cpi-U: All Items Less Shelter (82-84=100,Sa)	Δ^2ln
123	puxm	Cpi-U: All Items Less Midical Care (82-84=100,Sa)	Δ^2ln
124	gmdc	Pce,Impl Pr Defl:Pce (1987=100)	Δ^2ln
125	gmdcd	Pce,Impl Pr Defl:Pce; Durables (1987=100)	Δ^2ln
126	gmdcn	Pce,Impl Pr Defl:Pce; Nondurables (1996=100)	Δ^2ln
127	gmdcs	Pce,Impl Pr Defl:Pce; Services (1987=100)	Δ^2ln
128	ces275	Avg Hourly Earnings Of Prod. Or Nonsuperv. Workers On Private Nonfarm	Δ^2ln
129	ces277	Avg Hourly Earnings Of Prod. Or Nonsuperv Workers On Private Nonfarm	Δ^2ln
130	ces278	Avg Hourly Earnings Of Prod. Or Nonsuperv. Workers On Private Nonfarm	Δ^2ln
131	hhsntn	U. Of Mich. Index Of Consumer Expectations(Bcd-83)	Δlv

Appendix B

Parameters Initialization Technique

In the following appendix we summarize the techniques used in the dissertation for initializing the parameters in FNN-DF and well described in (Reed and Marks, 1998, pages 97-105). In literature there are basically two methods. One consists of methods of choosing parameters by controlling the distribution of random initial parameters. The idea is to avoid sigmoid saturation problems causing slow estimation and poor response in the output of hidden units. The other method consists of initializing the FNN-DF from an approximate solution found by another modeling system, examples are decision tree and discriminant analysis. Here we adopt the method proposed by Nguyen and Widrow (1990), which in principle belongs to the cluster of random initialization but also possesses several characteristics which decrease the chance of the network becoming trapped in a local minimum, as discussed in Wessels and Barnard (1992).

Nguyen-Widrow Initialization. Parameter vectors are chosen with random directions, magnitudes are adjusted so each hidden unit is linear over a fraction of the input space with some overlap of linear regions between the hidden units with similar directions, and constants are set so the hyperplanes have random distances from the origin within the region occupied by the input data.

Let $\bar{\boldsymbol{\gamma}}$ represent the parameter vector excluding the constant term and let $\bar{\gamma}_0$ denote the constant. The sum in an hidden unit is $u_t = \bar{\boldsymbol{\gamma}}'\mathbf{s}_t + \bar{\gamma}_0$.

I. First, set the parameters so each vector has a random direction. A Gaussian distribution is used because this makes all directions equally likely (a uniform distribution tends to favor directions pointing to corners of the hypercube); then $\bar{\boldsymbol{\gamma}} \sim N(0, 1)$;

II. Adjust the magnitude of $\bar{\boldsymbol{\gamma}}$ so the linear region covers a fraction of the input space. The best width for the linear region depends on the number of hidden units. The linear region of the hidden unit with sigmoid activation function covers the region from - u_t^{sat}

to $+\ u_t^{sat}$. The hyperbolic tangent activation function u_t^{sat} is determined as follows. For $y_t = \tanh(u_t)$ nonlinearities, the input u_t needed to produce an output y_t is $u_t = \ln \frac{1+y_t}{1-y_t}$. The hidden unit is saturated for $|y_t| > 0.9$ so $u_t^{sat} = \ln 19 = 2.94$. If the inputs lie in the interior of the unit hypersphere, the maximum hidden unit occurs when $\mathbf{s}_t = \bar{\gamma}$, namely $u_t^{max} = \|\bar{\gamma}\|^2$. To make the linear region approximately $1/5$ of the diameter of the input space this should be about 5 times u_t^{sat}, namely $u_t^{max} = 5u_t^{sat}$. Normalization of $\bar{\gamma}$ to a magnitude $\|\bar{\gamma}\| = \sqrt{5u_t^{sat}} = 3.84$ gives

$$\bar{\gamma} \leftarrow 3.84 \frac{\bar{\gamma}}{\|\bar{\gamma}\|}$$

III. Set the constant term $\bar{\gamma_0}$ so the distance of the hyperplane from the origin has a random distribution between 0 and 1. The distance of the hyperplane from the origin is $d = \frac{\bar{\gamma_0}}{\|\bar{\gamma}\|}$ so choose

$$\bar{\gamma_0} = \|\bar{\gamma}\|\tau$$

where τ is a random number between 0 and 1.

Bibliography

L. Alessi, M. Barigozzi, and M. Capasso. A robust criterion for determining the number of static factors in approximate factor models. LEM Papers Series 2007/19, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, September 2007. URL http://ideas.repec.org/p/ssa/lemwps/2007-19.html. 46, 47, 67

U. Anders and O. Korn. Model selection in neural networks. *Neural Networks*, 12(2):309 – 323, 1999. ISSN 0893-6080. URL http://www.sciencedirect.com/science/article/B6T08-3VW54W9-C/2/e0c306f6dc0a1914ac0f9d5f8607937a. 19

M. A. Arbib. *The Handbook of Brain Theory and Neural Networks (Bradford Book)*. The MIT Press, July 1998. ISBN 0262511029. 18

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002. 15, 27, 46

R. J. Bhansali. *Multi-step Forecasting*, volume 1 of *A Companion to economic Forecasting*, chapter 9, pages 206–221. Blackwell, 2002. 34

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–304, September 1983. 2

Y. Y. Chong and D. F. Hendry. Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, 53(4):671–690, 1986. URL http://www.jstor.org/stable/2297611. 47

G. Cybenko. Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signal & Systems*, 2(4):303–314, 1989. 3, 18

M. C. Dhal and S. Hylleberg. Flexible regression models and relative forecast performance. *International Journal of Forecasting*, 20(2):201–217, 2004. URL http://ideas.repec.org/a/eee/intfor/v20y2004i2p201-217.html. 45

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63, July 1995. URL http://ideas.repec.org/a/bes/jnlbes/v13y1995i3p253-63.html. 50

T. L. Fine. *Feedforward Neural Network Methodology.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1999. ISBN 0387987452. 3

D.F. Foresee and M.T. Hagan. Gauss-newton approximation to bayesian learning. *International Conference on Neural Networks*, 3:1930–1935, 1997. 27, 31, 32, 33

M. Forni and M. Lippi. The generalized dynamic factor model: Representation theory. *Econometric Theory*, 17(06):1113–1141, 2001. URL `http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=93204&fulltextType=RA&fileId=S0266466601176048`. 2, 6, 8, 9

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, November 2000. URL `http://ideas.repec.org/a/tpr/restat/v82y2000i4p540-554.html`. 7, 8, 12

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–840, September 2005. URL `http://ideas.repec.org/a/bes/jnlasa/v100y2005p830-840.html`. 12, 18, 27, 36

J. Geweke. *The dynamic factor analysis of economic time series.* Latent Variables in Socio-Economic Models. North-Holland, Amsterdam, 1977. 2

R. Giacomini and H. White. Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578, November 2006. URL `http://dx.doi.org/10.1002/for.3980030207`. 49, 76

C. W. J. Granger. Invited review combining forecasts - twenty years later. *Journal of Forecasting*, 8(3):167–173, 1989. URL `http://dx.doi.org/10.1002/for.3980080303`. 43, 48

M.T. Hagan and M.B. Menhaj. Training feedforward networks with the marquardt algorithm. *Neural Networks, IEEE Transactions on*, 5(6):989–993, 1994. 27, 29, 30

M. Hallin and R. Liska. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617, 2007. 15, 27, 46, 47, 67

S. Heravi, D. R. Osborn, and C. R. Birchenhall. Linear versus neural network forecasts for european industrial production series. *International Journal of Forecasting*, 20(3):435–

446, 2004. ISSN 0169–2070. URL http://www.sciencedirect.com/science/article/B6V92-49HDT54-4/2/bde6d2513ebb403e8159936e839cfd00. 45

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Netw.*, 2(5):359–366, 1989. 3, 18, 23, 25

T. H. Lee, H. White, and C. W. J. Granger. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56(3):269–290, April 1993. URL http://www.sciencedirect.com/science/article/B6VC0-459J6XM-1G/2/a6b6b6ad923a1834cf398ef00e75b1b8. 1

D.J.C. Mackay. A practical bayesian framework for backpropagation networks. *Neural Computation*, 4:448–472, 1992a. 19, 27, 31, 32

D.J.C. Mackay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992b. URL http://www.mitpressjournals.org/doi/abs/10.1162/neco.1992.4.3.415. 19

D. P. Mandic and J. Chambers. *Recurrent Neural Networks for Prediction: Learning Algorithms,Architectures and Stability*. John Wiley & Sons, Inc., New York, NY, USA, 2001. ISBN 0471495174. 22

M. Marcellino. Forecasting emu macroeconomic variables. *International Journal of Forecasting*, 20(2):359–372, 2004. ISSN 0169–2070. URL http://www.sciencedirect.com/science/article/B6V92-4BP3NWD-1/2/9dd97b8ea9a4498b832d39e3ec81c5d5. 45

P. D. McNelis. *Neural Networks in Finance: Gaining Predictive Edge in the Market (Academic Press Advanced Finance Series)*. Academic Press, Inc., Orlando, FL, USA, 2004. ISBN 0124859674. 17

D. Nguyen and B. Widrow. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proceedings of the International Joint Conference on Neural Networks*, 3:21–26, 1990. 33, 81

M. H. Pesaran and A. Timmermann. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4):561–65, October 1992. URL http://ideas.repec.org/a/bes/jnlbes/v10y1992i4p561-65.html. 35, 36

R. D. Reed and R. J. Marks. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. MIT Press, Cambridge, MA, USA, 1998. ISBN 0262181908. 81

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. pages 673–695, 1988. 18, 28

T.J. Sargent and C.A. Sims. *Business cycle modelling without pretending to have too much a priori economic theory*. Federal reserve Bank of Minneapolis, Minneapolis, new methods in business research edition, 1977. 2

M. Shintani. Nonlinear forecasting analysis using diffusion indexes: An application to japan. *Journal of Money, Credit and Banking*, 37(3):517–38, June 2005. URL http://ideas.repec.org/a/mcb/jmoncb/v37y2005i3p517-38.html. 44

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, December 2002a. 2, 19, 27, 43, 45

J.H. Stock and M.W. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Nber working papers, National Bureau of Economic Research, Inc, June 1998. 44

J.H. Stock and M.W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–62, April 2002b. URL http://ideas.repec.org/a/bes/jnlbes/v20y2002i2p147-62.html. 2, 9, 10

N. R. Swanson and H. White. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13(4):439–461, 1997a. ISSN 0169–2070. URL http://www.sciencedirect.com/science/article/B6V92-3SX4NTF-2/2/ccb11743df0f3448aa8915458857b6e0. 44

N.R. Swanson and H. White. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, 79(4):540–550, November 1997b. URL http://www.jstor.org/stable/1392186. 44

T. Terasvirta. *Forecasting economic variables with nonlinear models*, volume 1 of *Handbook of*

Economic Forecasting, chapter 8, pages 413–457. Elsevier, 2006. URL http://ideas.repec.org/h/eee/ecofch/1-08.html. 20, 34

T. Terasvirta, D. van Dijk, and M.C. Medeiros. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21(4):755–774, 2005. URL http://www.sciencedirect.com/science/article/B6V92-4GMJ8X2-1/2/c960be792af9af88f57deb2e3c9f4885. 44

A. Timmermann. *Forecast Combinations*, volume 1 of *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier, 2006. URL http://www.sciencedirect.com/science/article/B7P5J-4JSMTWJ-7/2/894a363c6779672c8d0a83df231e6c24. 47

L.F.A. Wessels and E. Barnard. Avoiding false local minima by proper initialization of connections. *Neural Networks, IEEE Transactions on*, 3(6):889–905, 1992. 81

K. D. West. Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084, September 1996. URL http://www.jstor.org/stable/2171956. 49, 50