Journal of Forecasting, *J. Forecast.* **32**, 32–40 (2013)
Published online 10 October 2011 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.1248

Nonlinear Forecasting Using Factor-Augmented Models

BRUNO CARA GIOVANNETTI*
[1] *Department of Economics, University of São Paulo, Brazil*

ABSTRACT
Using factors in forecasting exercises reduces the dimensionality of the covariates set and, therefore, allows the forecaster to explore possible nonlinearities in the model. For an American macroeconomic dataset, I present evidence that the employment of nonlinear estimation methods can improve the out-of-sample forecasting accuracy for some macroeconomic variables, such as industrial production, employment, and Fed fund rate. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS factor models; forecasting; nonlinear

INTRODUCTION

A reasonably recent advance in forecasting is the so-called factor-augmented model. In a first step one estimates a few latent common factors from a large number of explanatory variables. Then, in a second step, the estimated common factors are used as covariates in a forecasting regression.

Assuming forecast variables to be linearly related to both their lags and common factors, Stock and Watson (2002) show that out-of-sample forecasting for macroeconomic variables can be improved with such a methodology. Ludvigson and Ng (2009a,b) use factor-augmented linear regressions to argue that bond excess returns are forecastable by macroeconomic variables. They present evidence that the inclusion of a macro factor in the model developed by Cochrane and Piazzesi (2005) improves the forecasting results.

Given the dimensionality reduction provided by the factor approach, a natural extension of this literature would be to consider the possible existence of nonlinearity in the models. Could the employment of a factor-augmented nonlinear model improve forecasting results even more?

Shintani (2005) was the first to investigate this possibility. Using artificial neural networks to predict Japanese output and inflation, he concludes that the linear specification is a good approximation of the true model for his data. However, as the present paper shows, such an issue does deserve further investigation. Other datasets have to be tested and other nonlinear estimation methods have to be employed.

I use additive semi-parametric techniques to estimate the factor-augmented forecasting equation in order to predict American macroeconomic data. The results indicate that for industrial production, employment, and Fed fund rate, nonlinearity—when considered—helps the forecaster to get a better prediction out of the sample. Indeed, the mean square forecasting error (MSFE) can be reduced as much as 30% in some cases.

The paper is organized as follows: Section 2 presents the forecasting model and the estimation method, Section 3 gives the empirical analysis and Section 4 concludes.

THE FORECASTING MODEL

In accordance with the factor-augmented model literature (e.g. Chamberlain and Rothschild, 1983; Stock and Watson, 1999, 2002; Forni *et al.*, 2002; Breitung and Eickmeier, 2006) the model consists of both a dynamic factor model and a forecasting equation.

Let y_{t+n} denote the scalar series to be predicted and let X_t be an N-dimensional vector with all the explanatory variables at time t, both sets weakly stationary. In addition, define f_t as an $(R \times 1)$ vector representing the R common dynamic factors of X_t, $F_t = \left(f_t, f_{t-1}, \dots, f_{t-q}\right)'$, and both e_t and ϵ_{t+n} as idiosyncratic disturbances, where the first is an N-dimensional vector and the second is a scalar.

Defining the parameter vectors Γ and β with dimensionality equal to $(N \times (q+1)R)$ and $((q+1)R \times 1)$ respectively, and the lag operator $\gamma(L)$, if the pair (y_{t+h}, X_t) admits a factor model *linear* representation with R dynamic factors, the model is [1]

* Correspondence to: Bruno Cara Giovannetti, Av. Prof. Luciano Gualberto, 908, Prédio FEA 2, Cidade Universitária, São Paulo - SP, CEP: 05508-010. E-mail: bgiovannetti@fipe.org.br

[1] Γ contains the factor loadings, and the product ΓF_t is called the common component.

$$X_t = \Gamma F_t + e_t \tag{1}$$

$$y_{t+h} = \beta' F_t + \gamma(L) y_t + \epsilon_{t+h} \tag{2}$$

According to equation (2), there is a linear relation between the dependent variable and both the common factors and its lags. The goal of the present paper is to lift such a restriction. Hence, for some unknown function g, the forecasting model becomes

$$X_t = \Gamma F_t + e_t \tag{3}$$

$$y_{t+h} = g\left(F_t, y_t, y_{t-1}, \ldots\right) + \epsilon_{t+h} \tag{4}$$

The most common estimation method for the factor-augmented model is a two-step procedure. The first step consists of using the sample data $\{X_t\}_{t=1}^{T}$ to estimate the time series of factors. The method of principal components produces consistent estimators of the factors under general conditions. Particularly, e_t is allowed to be correlated and heteroskedastic in both the time and cross-sectional dimensions. Bai (2003) derives the rate of convergence and the limiting distributions for the estimated factors, factor loadings, and common components, estimated by the principal component method. A standard way to determine the number of factors is the information criteria developed by Bai and Ng (2002). In order to simulate real-time forecasting and perform an out-of-sample analysis in the second step, the principal components should be extracted recursively.

Regarding the second step, in the most general case, one would estimate equation (4). However, intermediate cases between the full parametric linear model and the full nonparametric model can also be considered. An example is the additive nonparametric model

$$\begin{aligned} y_{t+h} &= g\left(\widehat{F}_t, y_t, y_{t-1}, \ldots, y_{t-S}\right) + \epsilon_{t+h} \\ &= \Phi_1\left(\widehat{f}_t\right) + \ldots + \Phi_{Q+1}\left(\widehat{f}_{t-Q}\left(\widehat{f}_{t-Q}\right)\right) + \Psi_1(y_t) + \ldots + \Psi_{S+1}(y_{t-S}) + \epsilon_{t+h} \end{aligned} \tag{5}$$

where $\Phi_i\left(\widehat{f}_{t-i+1}\right) = \phi_{1,i}\left(\widehat{f}_{1,t-i+1}\right) + \ldots + \phi_{R,i}\left(\widehat{f}_{R,t-i+1}\right)$, for $i = 1, \ldots, Q+1$, and $\phi_{r,i}$ and Ψ_s are unknown functions.

In fact, model (5) is the one that, according to the out-of-sample forecasting results, has the best fit to all series considered here.[2] Given that, I impose g to be a linear combination of unknown functions from now on.

A standard way to estimate (5) is penalized splines. It was first proposed by O'Sullivan (1986), refined by Eilers and Marx (1996), and made popular through the book by Ruppert *et al.* (2003).[3]

To illustrate the method, suppose first that there is only one factor and no lag, i.e. $y_{t+h} = \phi\left(\widehat{f}_{t1}\right) + \epsilon_{t+h}$, where the only assumption about ϕ is smoothness. Then, to estimate ϕ one can use some basis function parametrizing ϕ as

$$\phi\left(\widehat{f}_1, \beta\right) = \sum_{j=1}^{P+K} B_j\left(\widehat{f}_1\right)\beta_j$$

where P and K are positive integers, $\beta = \left(\beta_0, \beta_1, \ldots, \beta_{P+K}\right)'$ is a vector of regression coefficients, and $B_0(\cdot), \ldots, B_{P+K}(\cdot)$ is the basis. A common choice for the basis is the cubic spline, where $K = 2$:

$$B_j\left(\widehat{f}_1\right) = \left|\widehat{f}_1 - \kappa_j\right|^3, \ \text{ for } j = 1, \ldots, P$$

$$b_{P+1}\left(\widehat{f}_1\right) = 1$$

$$b_{P+2}\left(\widehat{f}_1\right) = \widehat{f}_1$$

and $\kappa = \left\{\kappa_j : j = 1, \ldots, P\right\}$ is a set of points in the range of $\widehat{f}_1$, the so-called knots.

---

[2] Two other possibilities were also considered: a full nonparametric model as in equation (4) and a single-index model. The full nonparametric model was estimated by local linear regression and neural networks and the single-index model by Ichimura's (1993) method.

[3] Other celebrated methods to estimate additive models are the classical backfitting of Buja *et al.* (1989) and Hastie and Tibshirani (1989), and its extension, the smooth backfitting of Mammen *et al.* (1999). We also employ the last one in our estimations in the next section, to check the robustness of the results.

Defining $\widehat{z}_t = \left(\left| \widehat{f}_{t1} - \kappa_1 \right|^3, \left| \widehat{f}_{t1} - \kappa_2 \right|^3, \ldots, \left| \widehat{f}_{t1} - \kappa_P \right|^3, 1, \widehat{f}_{t1} \right)'$ and $\widehat{Z} = (\widehat{z}_1 \widehat{z}_2 \ldots \widehat{z}_T)$, we have

$$y_{t+h} = \widehat{Z}\beta + \epsilon \tag{6}$$

where $\epsilon = (\epsilon_{1+h}, \epsilon_2, \ldots, \epsilon_T)'$.

In principle, if P was chosen to be large enough to approximate ϕ well, this model could be fitted by minimizing $\epsilon'\epsilon$. However, if P is too large, the estimation is going to over-fit the data; i.e. it will begin to fit the sample noise (which will cause, in the limit, the perfect interpolation of the data points). An option would be to choose P by cross-validation methods, as in Friedman and Silverman (1989) and Stone *et al.* (1997), where the knots would be selected from a set of candidate knots in a way similar to stepwise regression. However, this can be computationally too expensive.

A solution is to use penalized splines (P-splines). The idea of O'Sullivan (1986) and Eilers and Marx (1996) was to set P intentionally large (it can even be a knot at each unique value of the support, which, in this case, would be called smoothing splines) and control over-fitting by using least squares estimation with a roughness penalty. The penalty is on the integral of the square of a specified derivative, usually the second. In this case, one would minimize

$$\epsilon'\epsilon + \lambda \int_{\widehat{f}_{1\,\min}}^{\widehat{f}_{1\,\max}} \left[\phi''\left(\widehat{f}_1\right) \right]^2 \mathrm{d}\widehat{f}_1 \tag{7}$$

The first term in (7) is the traditional sum of the square of the residuals. The second term is the roughness penalty, which increases as the cubic splines become rougher, i.e. when their slope changes very rapidly—the integrated second derivative of the regression function is a measure for it. Therefore, λ defines the degree of smoothing: the larger λ, the larger is the smoothness of the estimator (the estimator's bias increases and its variance decreases). For $\lambda = 0$, one tends to the perfect interpolation of all data points as P increases; on the other hand, for $\lambda \to \infty$, one has the linear least squares estimator. The advantage of the P-splines is that now one has to choose a single parameter value to determine the smoothness of the estimator, in contrast to having to define the number and location of the knots. Cross-validation is now computationally cheaper.

Note that object (7) is equal to $\varepsilon'\varepsilon + \lambda\beta'\Lambda\beta$, where Λ is a matrix of known values. In the case of the cubic splines:

$$\Lambda = \begin{pmatrix} 12 & 12 - \frac{\kappa_2}{2} - \frac{\kappa_1}{2} + \kappa_1\kappa_2 & \cdots & 12 - \frac{\kappa_P}{2} - \frac{\kappa_1}{2} + \kappa_1\kappa_P & 0 & 0 \\ 12 - \frac{\kappa_2}{2} - \frac{\kappa_1}{2} + \kappa_1\kappa_2 & 12 & \cdots & 12 - \frac{\kappa_P}{2} - \frac{\kappa_2}{2} + \kappa_2\kappa_P & 0 & 0 \\ \vdots & \cdots & \ddots & \vdots & \vdots & \vdots \\ 12 - \frac{\kappa_P}{2} - \frac{\kappa_1}{2} + \kappa_1\kappa_P & 12 - \frac{\kappa_P}{2} - \frac{\kappa_2}{2} + \kappa_2\kappa_P & \cdots & 12 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}$$

Minimizing (7), the following closed-form estimator of $\phi\left(\widehat{f}_1\right)$ is obtained:

$$\widehat{\phi}\left(\widehat{f}_1\right) = \widehat{Z}\left(\widehat{Z}'\widehat{Z} + \lambda\Lambda\right)^{-1}\widehat{Z}'y \tag{8}$$

where $\widehat{Z}$ and Λ were defined above and $y = (y_{1+h}, y_{2+h}, \ldots, y_T)'$.

The estimator in (8) was derived imposing the existence of a single factor with no lag and no autoregressive component in the forecasting model. It is straightforward to generalize it for the case of the complete model, as in equation (5). If the model has R factors, Q lags for the factors and S lags for the dependent variable, the function to be minimized is

$$\begin{aligned} \epsilon'\epsilon &+ \sum_{q=1}^{Q+1} \lambda_{1,q} \int_{\widehat{f}^{\min}_{1,t-q+1}}^{\widehat{f}^{\max}_{1,t-q+1}} \left[\phi''_{1,q}\left(\widehat{f}_{1,t-q+1}\right) \right]^2 \mathrm{d}\widehat{f}_{1,t-q+1} + \ldots \\ &+ \sum_{q=1}^{Q+1} \lambda_{R,q} \int_{\widehat{f}^{\min}_{R,t-q+1}}^{\widehat{f}^{\max}_{R,t-q+1}} \left[\phi''_{R,q}\left(\widehat{f}_{R,t-q+1}\right) \right]^2 \mathrm{d}\widehat{f}_{R,t-q+1} \\ &+ \sum_{s=0}^{S} \lambda_{y,s} \int_{y^{\min}_{t-s}}^{y^{\max}_{t-s}} \left[\Psi''_s(y_{t-s}) \right]^2 \mathrm{d}y_{t-s} \end{aligned} \tag{9}$$

where $\epsilon = (\epsilon_{1+h}, \epsilon_2, \ldots, \epsilon_T)$ from equation (5).

Analogously to the univariate case, each unknown smooth function in (9) is approximated using a cubic spline basis. Given that, equation (6) also holds. The difference is that $\widehat{Z}$ now has many more columns. More precisely, dividing $\widehat{Z}$ into blocks:

$$\widehat{Z} = \left(\widehat{Z}_{1,1} \ldots \widehat{Z}_{1,Q+1} \ldots \widehat{Z}_{R,1} \ldots \widehat{Z}_{R,Q+1} \widehat{Z}_{y,0} \ldots \widehat{Z}_{y,S} \;\; \mathbf{1}\right)$$

where

- $\widehat{Z}_{r,q}$, for $q = 1, \ldots, Q+1$, and $r = 1, \ldots R$, is a $T \times P$ matrix with columns equal to

$$\left(\left|\widehat{f}_{r,q} - \kappa_{r,q,1}\right|^3, \left|\widehat{f}_{r,q} - \kappa_{r,q,2}\right|^3, \ldots, \left|\widehat{f}_{r,q} - \kappa_{r,q,P}\right|^3, \widehat{f}_{r,q}\right)'$$

- $\widehat{Z}_{y,s}$, for $s = 0, \ldots S$, is a $T \times P$ matrix with columns equal to

$$\left(\left|y_{t-s} - \kappa_{y,1}\right|^3, \left|y_{t-s} - \kappa_{y,2}\right|^3, \ldots, \left|y_{t-s} - \kappa_{y,P}\right|^3, y'_{t-s}\right)$$

- and $\mathbf{1}$ is a $T \times 1$ vector with ones.

Hence $\widehat{Z}$ has $(R(Q+1)+S)(P+1)+1$ columns and β is an $((R(Q+1)+S)(P+1)+1) \times 1$ vector of parameters. We have now to set $R(Q+1)+S+1$ smoothing parameters by cross-validation, after having defined an intentionally large number of knots within the support of each of the explanatory variables. Therefore, the larger R, Q and S, the longer is the computational time for the cross-validation procedure.

Minimizing (9) with respect to β, the following estimator in closed form is produced:

$$\widehat{g}\left(\widehat{F}_t, y_t, y_{t-1}, \ldots, y_{t-S}\right) = \widehat{Z}\left(\widehat{Z}'\widehat{Z} + \sum_{q=1}^{Q+1} \lambda_{1,q}\Lambda_{1,q} + \ldots + \sum_{q=1}^{Q+1} \lambda_{R,q}\Lambda_{R,q} + \sum_{s=1}^{S} \lambda_{y,s}\Lambda_{y,s}\right)^{-1} \widehat{Z}'y \qquad (10)$$

In the next section, I apply the estimator defined in (10) recursively, as in the first step, in order to simulate real-time forecasting.

EMPIRICAL ANALYSIS

In this section, results on the prediction of US macroeconomic variables (industrial production, employees on nonfarm payroll, housing starts, and Fed fund rates) and Treasury bonds future returns are presented.

The empirical work is based on the macroeconomic panel used in Ludvigson and Ng (2009b), which consists of monthly observations of 131 macroeconomic time series, ranging from January 1964 to December 2007. Some series are transformed to become $I(0)$ stationary, according to the appendix in Ludvigson and Ng (2009b). The data on Treasury bonds are taken from the Fama–Bliss dataset, available from the Center for Research in Securities Prices (CRSP).

The 131 explanatory variables used to construct the macroeconomic factors in the first step represent broad categories of macroeconomic time series: real output and income, employment and hours, real retail, manufacturing and sales data, international trade, consumer spending, housing starts, inventories and inventory sales ratios, orders and unfilled orders, compensation and labor costs, capacity utilization measures, price indexes, interest rates and interest rate spreads, stock market indicators, and foreign exchange measures. A complete list of the series is given in the appendix in Ludvigson and Ng (2009b).

The number of factors in the first step, R, is equal to eight, in accordance with Bai and Ng's (2002) information criteria. The first factor accounts for 18% of the total variance in the macroeconomic panel as, all together, factors 1–8 represent 50% of the total variance. To try to get an interpretation of each factor, a common practice is to regress each series in the macroeconomic panel on the factor, and to report the R^2. Ludvigson and Ng (2009b) present such plots. Accordingly, the first factor is related to measures of employment, production, capacity utilization and new manufacturing orders; the second factor is correlated with several interest rate spreads; the third factor correlates with measures of inflation; factors 4, 5, 6 and 7 display less clear correlation patterns; and factor 8 loads heavily on measures of the aggregate stock market.

However, instead of using all eight factors in the second step (which would make the number of different possible specifications too large), I follow Ludvigson and Ng (2009a,b) and compute a single factor as a linear combination of them. The parameters of the linear combination are defined at each forecasting period, given by the optimal (in the Bayesian information criterion (BIC) sense) linear regression of the dependent variable on the eight factors. Subsequently, the number of lags for the dependent variable and for the single factor are set in accordance with the BIC, considering the full sample.

Besides estimating the additive models using cubic splines, I also employ thin plate splines. Additionally, I use another popular estimation method for additive models: the backfitting of Mammen *et al.* (1999). This is done to check the robustness of the results.

Tables I–IV present the results for the macroeconomic variables, and are divided into four blocks. In the first block, I compare the forecasting power of a full nonlinear model and of a full linear model. For example, if the forecasting

Table I. Industrial production

Months ahead	Cubic	Thin plate	Backfitting
Block 1: MSFE (nonlinear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	0.98	0.97	0.96
6	0.90	0.88	0.85
12	0.91	0.92	0.90
24	0.99	1.01	1.00
Block 2: MSFE (nonlinear AR + linear factor)/ MSFE(linear AR + linear factor)			
1	0.98	0.98	0.98
6	0.88	0.88	0.86
12	0.88	0.88	0.86
24	0.95	0.95	0.94
Block 3: MSFE (linear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	1.00	1.00	0.99
6	1.06	1.05	1.04
12	1.05	1.06	1.06
24	1.05	1.06	1.05
Block 4: MSFE (nonlinear AR)/ MSFE(linear AR)			
1	0.99	0.99	0.98
6	1.00	1.01	1.02
12	1.00	1.00	1.03
24	1.01	1.01	1.03

Note: The entries of block 1 are the ratios between the MSFE of the full nonlinear model and the MSFE of the full linear model. Hence a value less than one is an indication of some nonlinearity in the model. Blocks 2, 3 and 4 indentify where the possible nonlinearity comes from. In the case that the entry in block 2 is less than one, this is an indication of nonlinearity in the autoregressive structure, when this is controlled for the factor structure. In the case that the entry in block 3 is less than one, this is an indication of nonlinearity in the factor structure, when controlled for the autoregressive structure. Block 4 verifies the existence of nonlinearity in the autoregressive structure when not controlled for the factor.

Table II. Employment

Months ahead	Cubic	Thin plate	Backfitting
Block 1: MSFE (nonlinear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	1.02	1.03	0.97
6	0.82	0.80	0.78
12	0.85	0.85	0.81
24	0.99	0.99	1.00
Block 2: MSFE (nonlinear AR + linear factor)/ MSFE(linear AR + linear factor)			
1	1.05	1.05	0.98
6	0.74	0.75	0.73
12	0.70	0.70	0.71
24	0.95	0.95	0.96
Block 3: MSFE (linear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	1.02	1.03	1.04
6	1.13	1.12	1.10
12	1.16	1.15	1.13
24	1.05	1.06	1.07
Block 4: MSFE (nonlinear AR)/ MSFE(linear AR)			
1	1.10	1.12	1.03
6	0.98	0.98	0.95
12	0.90	0.90	0.88
24	0.97	0.97	0.97

Note: See Table I

Table III. Housing starts

Months ahead	Cubic	Thin plate	Backfitting
Block 1: MSFE (nonlinear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	1.04	1.03	1.02
6	1.09	1.06	1.06
12	1.27	1.26	1.16
24	1.07	1.06	1.04
Block 2: MSFE (nonlinear AR + linear factor)/ MSFE(linear AR + linear factor)			
1	1.02	1.01	1.02
6	1.07	1.05	1.04
12	1.15	1.12	1.08
24	0.99	0.98	0.98
Block 3: MSFE (linear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	1.01	1.01	1.02
6	1.05	1.06	1.06
12	1.16	1.16	1.13
24	1.11	1.10	1.07
Block 4: MSFE (nonlinear AR)/ MSFE(linear AR)			
1	1.03	1.03	1.03
6	1.07	1.06	1.10
12	1.11	1.05	1.05
24	0.96	0.93	0.92

Note: See Table I

Table IV. Fed fund rate

Months ahead	Cubic	Thin plate	Backfitting
Block 1: MSFE (nonlinear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	0.86	0.92	0.79
6	0.81	0.79	0.78
12	0.94	0.88	0.83
24	1.23	1.20	1.06
Block 2: MSFE (nonlinear AR + linear factor)/ MSFE(linear AR + linear factor)			
1	1.13	0.94	1.02
6	0.86	0.86	0.85
12	0.85	0.85	0.84
24	1.00	1.00	1.00
Block 3: MSFE (linear AR + nonlinear factor)/ MSFE(linear AR + linear factor)			
1	0.88	0.94	0.90
6	0.99	0.99	0.96
12	1.11	1.06	1.00
24	1.19	1.16	1.06
Block 4: MSFE (nonlinear AR)/ MSFE(linear AR)			
1	0.98	0.98	0.92
6	0.80	0.80	0.80
12	0.93	0.95	0.92
24	1.25	1.22	1.16

Note: See Table I

model has three lags for the dependent variable and one lag for the factor, I compute the ratios between the MSFEs of models

$$y_{t+h} = \Phi_1\left(\widehat{f}_t\right) + \Psi_1(y_t) + \Psi_2(y_{t-1}) + \Psi_3(y_{t-2}) + \text{ error}$$

and

$$y_{t+h} = \beta_0 + \beta_1 \widehat{f}_t + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 y_{t-2} + \text{ error}$$

As we will see, for three out of the four macroeconomic variables considered, the nonlinear model produces smaller MSFEs. Given that, the second, third and forth blocks investigate where the possible nonlinearity comes from. In the second block, Φ_1 above is assumed to be linear. In the third block, Ψ_1, Ψ_2 and Ψ_3 above are assumed to be linear. Finally, in the forth block, the macro factor is excluded from both models in order to compare a nonlinear autoregressive model and its linear version.

For all blocks, an entry of less than 1 indicates a better performance of the nonlinear model. Of course, providing statistical evidence on whether the MSFEs of the nonlinear models are smaller than those of linear models, by reporting Diebold and Mariano (1995) type t-statistics, would be ideal. However, as is well known, the limiting distribution of t-statistics under the null hypothesis of the same MSFE depends on a non-nested structure of the competing models. When competing models are nested, as here, the limiting distribution of the test statistic degenerates. For example, Clark and McCraken (2001) have to compute by simulation the statistical distribution for a test of equal forecast accuracy for linear and nested models. Unfortunately, for nonlinear (nonparametric and semi-parametric) models, no accepted testing procedure for equal forecast accuracy has been developed so far, to the best of my knowledge. Given that, we are left with a simple comparison of MSFEs, and this is standard in the nonlinear forecasting literature.

The out-of-sample forecasting exercises are performed for the period between January 1985 and December 2007. For the macroeconomic variables, 1-, 6-, 12- and 24-month-ahead forecasts are produced (the same periods as in Stock and Watson, 2002). Regarding the bonds, the returns of holding (during 1 year) bonds with maturities equal to 2–5 years are forecast, following Ludvigson and Ng (2009a).

For industrial production (Table I), block 1 indicates the existence of some sort of nonlinearity for 6 and 12 months ahead. According to blocks 2 and 3, the nonlinearity is in the autoregressive structure. Interestingly, according to block 4, the nonlinear relation between industrial production and its lags exists only when the model is controlled for the macro factor.

With respect to employment (Table II), the conclusion is qualitatively the same. Block 1 provides evidence for the existence of some sort of nonlinearity for 6 and 12 months ahead. Blocks 2 and 3 indicate that the nonlinearity is in the autoregressive structure. Finally, according to block 4, the nonlinear autoregressive structure is much more important once the model is controlled for the macro factor (although there is some source of nonlinearity in the simple autoregressive structure as well).

Regarding housing starts (Table III), there are no signs of nonlinearity.

Finally, for the Fed fund rate (Table IV), there are strong indications of nonlinearity in the autoregressive structure, even when the model does not include the macro factor.

We now turn to the Treasury bonds returns, and some additional explanation is necessary. For $t = 1, \ldots, T$, let $rx_{t+1}^{(n)}$ denote the continuously compounded (log) excess return on an n-year discount bond in year $t+1$. Excess returns are defined $rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}$, where $r_{t+1}^{(n)}$ is the log holding period return from buying an n-year bond in year t and selling it as an $(n-1)$ -year bond in year $t+1$, and $y_t^{(1)}$ is the log yield on the 1-year bond in year t. In other words, $rx_{t+1}^{(n)}$ is the return of the following trading strategy: one borrows at time t the necessary amount to buy an n-year discount bond, buys the n-year discount bond, sells such a bond 1 year later, and pays back the loan.

Fama and Bliss (1987) show that it is possible to forecast n-year excess bond returns by the spread between the n-year forward rate and the 1-year yield. Campbell and Shiller (1991) report that excess bond returns are forecastable by Treasury yield spreads. Cochrane and Piazzesi (2005) show that a linear combination of five forward spreads, the so-called CP factor, explains between 30% and 35% of the variation in next year's excess returns on bonds with maturities ranging from 2 to 5 years. More recently, Ludvigson and Ng (2009a,b) reported that the inclusion

Table V. Bond returns

	Cubic	Thin plate	Backfitting
Block 1: MSFE (nonlinear CP + nonlinear factor) / MSFE (linear CP + linear factor)			
2-year exc. return	1.10	1.09	1.06
3-year exc. return	1.10	1.10	1.06
4-year exc. return	1.10	1.10	1.05
5-year exc. return	1.03	1.04	1.01
Block 2: MSFE (nonlinear CP) / MSFE (linear CP)			
2-year exc. return	1.05	1.05	1.04
3-year exc. return	1.03	1.02	1.03
4-year exc. return	1.03	1.03	1.02
5-year exc. return	1.08	1.08	1.05

Note: The entries in block 1 are the ratios between the MSFE of the full (CP factor and macro factor) nonlinear model and the MSFE of the full linear model. Hence a value less than one is an indication of some nonlinearity in the model. Block 2 considers the model with the CP factor only.

of macroeconomic factors in Cochrane and Piazzesi's forecasting models could improve even more both in- and out-of-sample bonds excess return forecasts (they get R^2 values of the order of 40%). Their model is the same two-step factor-augmented linear model discussed here, with the difference that instead of lags of the dependent variable the right-hand side gives the Cochrane and Piazzesi CP factor (besides the macroeconomic factors). Given that, I also investigate whether a more general second-step procedure could improve even more the out-of-sample bonds excess return forecasts obtained by Ludvigson and Ng (2009a,b).

Table V presents the ratios of the MSFE of the nonlinear model and the MSFE of the linear model, for 2-, 3-, 4- and 5-year bonds, for models with only the CP factor and models with both the CP factor and macroeconomic factors. In contrast to the macroeconomic forecast variables, the results indicate that there is no nonlinearity in the models.

CONCLUSIONS

Combining factor-augmented models and nonlinear estimation methods should be a natural forecasting strategy, given the dimensionality reduction provided by the factor approach.

In spite of that, little attention has been given to the investigation of possible nonlinearities in factor-augmented forecasting models.

This study presented some empirical evidence that the linear assumption in these models may be restrictive. Indeed, for industrial production, employment and Fed fund rate, the MSFE was reduced as much as 30% using a generalized additive model for the factor-augmented forecasting equation.

ACKNOWLEDGEMENTS

This study was initiated while I was affiliated to Columbia University as a PhD student. I am grateful to an anonymous referee and a departmental editor (David Stoffer) for their constructive comments, which led to significant improvement in the paper. I also thank Dennis Kristensen, Alexei Onatski and Marcos Nakaguma for valuable comments and Serena Ng and Sydney Ludvigson for making their dataset available. In addition, this paper has benefited from suggestions received from the Columbia University Econometrics Colloquium. Any errors are my own.

REFERENCES

Bai J. 2003. Inferential theory for factor models of large dimension. *Econometrica* **71**: 135–171.

Bai J, Ng S. 2002. Determining the number of factors in approximate factor models. *Econometrica* **70**: 191–221.

Breitung J, Eickmeier S. 2006. Dynamic factor models. In *Modern Econometric Analysis: Surveys on Recent Developments*, Hübler O, Frohn J (eds). Springer: Berlin; 25–40.

Buja A, Hastie T, Tibshirani R. 1989. Linear smoothers and additive models. *Annals of Statistics* **17**: 453–555.

Campbell J, Shiller R. 1991. Yield spreads and interest rate movements: a bird's eye view. *Review of Economics Studies* **58**: 495–514.

Chamberlain G, Rothschild M. 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**: 1305–1324.

Clark T, McCracken M. 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* **105**: 85–110.

Cochrane J, Piazzesi M. 2005. Bond risk premia. *American Economic Review* **95**: 138–160.

Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Eilers P, Marx B. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* **11**: 89–121.

Fama E, Bliss R. 1987. The information in long-maturity forward rates. *American Economic Review* **77**: 680–692.

Forni M, Hallin M, Lippi F, Reichlin L. 2002. The generalized dynamic factor model: consistency and convergence rates. *Journal of Econometrics* **82**: 540–554.

Friedman J, Silverman B. 1989. Flexible parsimonious smoothing and additive modeling. *Technometrics* **31**: 3–39.

Hastie B, Tibshirani R. 1989. Linear Smoothers and Additive Models. *Annals of Statistics* **17**: 453–555.

Ichimura H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**: 71–120.

Ludvigson S, Ng S. 2009a. Macro factors in bond risk premia. *Review of Financial Studies* **22**: 5027–5067.

Ludvigson S, Ng S. 2009b. A factor analysis of bond risk premia. In *Handbook of Empirical Economics and Finance*, Ulah A, Giles D (eds). Chapman & Hall: Boca Raton, FL; 313–372.

Mammen E, Linton O, Nielsen J. 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* **27**: 1443–1490.

O'Sullivan F. 1986. A statistical perspective on ill-posed inverse problems. *Statistical Science* **1**: 505–527.

Ruppert D, Wand M, Carroll R. 2003. *Semi-parametric regression*. Cambridge University Press: Cambridge, UK.

Shintani M. 2005. Nonlinear forecasting analysis using diffusion indexes: an application to Japan. *Journal of Money, Credit and Banking* **37**: 517–538.

Stock J, Watson M. 1999. Forecasting inflation. *Journal of Monetary Economics* **44**: 293–335.

Stock J, Watson M. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* **20**: 147–162.
Stone C, Hansen M, Kooperberg C, Truong Y. 1997. Polynomial splines and their tensor products in extended linear modeling. *Annals of Statistics* **25**: 1371–1470.

Author's biography:
Bruno Cara Giovannetti is affiliated to the Economics Deparment at University of São Paulo since April 2011 as a visiting professor and post-doctorate researcher. He holds a Ph.D. in Economics from Columbia University (2011). His research fields are Financial Economics and Applied Econometrics.

Author's address:
Bruno Cara Giovannetti, Department of Economics, University of São Paulo, Brazil.

