Traditional versus Novel Forecasting Techniques: How Much do We Gain?

VIVIANA FERNANDEZ[1,2]*
[1] *Center for Applied Economics (CEA), Department of Industrial Engineering, University of Chile*
[2] *Institute for International Integration Studies, Trinity College, Dublin, Ireland*

ABSTRACT

This article applies two novel techniques to forecast the value of US manufacturing shipments over the period 1956–2000: wavelets and support vector machines (SVM). Wavelets have become increasingly popular in the fields of economics and finance in recent years, whereas SVM has emerged as a more user-friendly alternative to artificial neural networks. These two methodologies are compared with two well-known time series techniques: multiplicative seasonal autoregressive integrated moving average (ARIMA) and unobserved components (UC). Based on forecasting accuracy and encompassing tests, and forecasting combination, we conclude that UC and ARIMA generally outperform wavelets and SVM. However, in some cases the latter provide valuable forecasting information that it is not contained in the former.

KEY WORDS wavelets; support vector machines; forecasting

INTRODUCTION

Over the past 50 years, an increasing number of statistical methods has been developed to predict the evolution of various macroeconomic and financial time series (e.g., Roche, 1995; Diebold, 1998; Clements and Hendry, 1998; Leung *et al.*, 2000; Morana, 2001; Dooley and Lenihan, 2005; Lanza *et al.*, 2005; Mills, 1999; Tsay, 2005; Rapach *et al.*, 2005). Two recently developed mathematical techniques, which have started to gain ground in the economic and financial fields, are wavelets and support vector machines (SVM).

Wavelets, a refinement of Fourier analysis which dates back to the late 1980s, were originally utilized in signal and image processing. A particular feature of wavelet analysis is that it makes it possible to decompose a time series into its high- and low-frequency components, which are localized in time. Recent applications of wavelet methods in economics and finance have dealt with the permanent income hypothesis, the estimation of systematic risk, and the interaction between

*Correspondence to: Viviana Fernandez, Department of Industrial Engineering, University of Chile, Avenida Republica 701, Santiago, Chile. E-mail: vfernand@dii.uchile.cl

emerging and developed stock markets, among other themes (e.g., Ramsey and Lampart, 1998; Ramsey, 1999, 2002; Lin and Stevenson, 2001; Gençay *et al.*, 2002, 2003, 2005; Hong and Kao, 2004; Whitcher, 2004; Connor and Rossiter, 2005; Fernandez, 2005). However, the use of wavelets for forecasting purposes is almost non-existent in the literature. Indeed, one of the first attempts in this area was Arino's (1995) article on forecasting of Spanish car sales. More recently, Conejo *et al.* (2005) applied wavelets and other techniques to forecast electricity prices.

SVM represent a kernel-based learning algorithm, which has arisen as a more user-friendly tool than artificial neural networks (e.g., Burges, 1998; Cristianini and Shawe-Taylor, 2000). Traditionally, SVM have been applied to pattern recognition problems (e.g., text categorization, face recognition), but over the years their use has extended to nonlinear regression models. A key property of SVM is that training SVM boils down to a linearly constrained quadratic programming problem, whose solution is unique and globally optimal. Applications of SVM to forecasting are fairly recent and have dealt primarily with financial and energy issues (e.g., Tay and Kao, 2001, Kim, 2003; Dong *et al.*, 2005; Huang *et al.*, 2005; Lu and Wang, 2005).

Our study focuses on forecasting the value of shipments of the US manufacturing industry for the period January 1958 to December 2000. Some of the categories analyzed include Durable goods, Nondurable goods, Automotive equipment, Consumer staples, Business supplies, and Construction supplies. Providing industry participants with reliable forecasting techniques is, in our view, central to the decision-making process. In particular, an accurate prediction of the future value of shipments can help managers to assess the profitability and growth potential of an ongoing business.

The contribution of our work is twofold. Firstly, we utilize two novel nonlinear forecasting techniques, which are based on wavelet methods and support vector machines, along with two well-known model specifications: seasonal multiplicative autoregressive integrated moving average (ARIMA) and unobserved components. Secondly, we perform encompassing tests for various time horizons. As Fang (2003) illustrates for the case of UK consumption expenditure, forecast encompassing tests are a useful tool to determine whether a composite forecast can be superior to individual forecasts. In addition, forecast encompassing tests are potentially useful in model specification, as forecast combination implicitly assumes the possibility of model misspecification. Our computations, carried out for different time series, show that the time horizon is a key element to decide which model or combination of models can be more suitable in terms of forecast accuracy.

This article is organized as follows. The next section presents the four forecasting techniques utilized in this article. The third section presents forecast accuracy and encompassing tests. The fourth section describes the data, while the fifth section presents our estimation results. The sixth section presents our main findings.

SOME NONLINEAR FORECASTING MODELS

Seasonal multiplicative ARIMA

We consider the parsimonious multiplicative seasonal ARIMA model utilized by Box *et al.* (1994, Ch. 9) in their classical example of airline data. Specifically, let us consider a non-stationary y_t time series and a model which links its behavior one year apart:

$$\Delta_{12} y_t = (1 - \Theta L^{12})\varepsilon_t$$

A similar model is used to link ε_ts one month apart:

$$\Delta\varepsilon_t = (1 - \theta L)\upsilon_t$$

The combination of the two models gives rise to a seasonal multiplicative model:

$$\Delta\Delta_{12}y_t = (1 - \theta L)(1 - \Theta L^{12})\upsilon_t \tag{1}$$

of order $(0, 1, 1) \times (0, 1, 1)_{12}$.

The parameters θ and Θ are estimated by maximum likelihood.

Wavelets

Wavelets allow for decomposing a signal into fine and coarse resolution components (see, for instance, Bruce and Gao, 1996; Percival and Walden, 2000). Wavelets can be classified into father and mother wavelets. Father wavelets (ϕ) represent the smooth and low-frequency parts of a signal, whereas mother wavelets (ψ) characterize its detailed and high-frequency parts. The most widely used wavelets are the orthogonal ones (i.e., haar, daublets, symmelets, and coiflets). In particular, the orthogonal wavelet series approximation to a continuous signal $f(t)$ is given by

$$f(t) \approx \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t) \tag{2}$$

where J is the number of multi-resolution components or scales, and k ranges from 1 to the number of coefficients in the corresponding component. The coefficients $s_{J,k}$, $d_{J,k}, \ldots, d_{1,k}$ are the wavelet transform coefficients, whereas the functions $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are the approximating wavelet functions.

Applications of wavelet analysis usually utilize a discrete wavelet transform (DWT). The DWT calculates the coefficients of the approximation in (2) for a discrete signal of final extent, $f_1, f_2, \ldots, f_n$. That is, it maps the vector $\mathbf{f} = (f_1, f_2, \ldots, f_n)'$ to a vector $\boldsymbol{\omega}$ of n wavelet coefficients, which contains $s_{J,k}$ and $d_{j,k}$, $j = 1, 2, \ldots, J$. The $s_{J,k}$s and $d_{j,k}$s are called the smooth and detail coefficients, respectively. Intuitively, the smooth coefficients represent the underlying smooth behavior of the data at the coarse scale 2^J, whereas the detail coefficients provide the coarse-scale deviations from it. When the length of the data n is divisible by 2^J, there are $n/2$ coefficients $d_{1,k}$ at the finest scale $2^1 = 2$. At the next finest scale, there are $n/2^2$ coefficients $d_{2,k}$. Similarly, at the coarsest scale, there are $n/2^J$ coefficients each for $d_{J,k}$ and $s_{J,k}$.

Expression (2) can be rewritten as

$$f(t) \approx S_J(t) + D_J(t) + D_{J-1}(t) + \ldots + D_1(t) \tag{3}$$

where $S_J(t) = \sum_k s_{J,k}\phi_{J,k}(t)$ and $D_J(t) = \sum_k d_{j,k}\psi_{J,k}(t)$ are denominated the smooth and detail signals, respectively.

The terms in expression (3) represent a multi-resolution decomposition (MRD) of the signal into the orthogonal signal components $S_J(t)$, $D_J(t)$, $D_{J-1}(t), \ldots, D_1(t)$ at different scales. For instance, when analyzing monthly data, wavelet scales are such that scale 1 is associated with 2- to 4-month dynamics, scale 2 with 4- to 8-month dynamics, scale 3 with 8- to 16-month dynamics, scale 4 with 16- to 32-month dynamics, and so on.

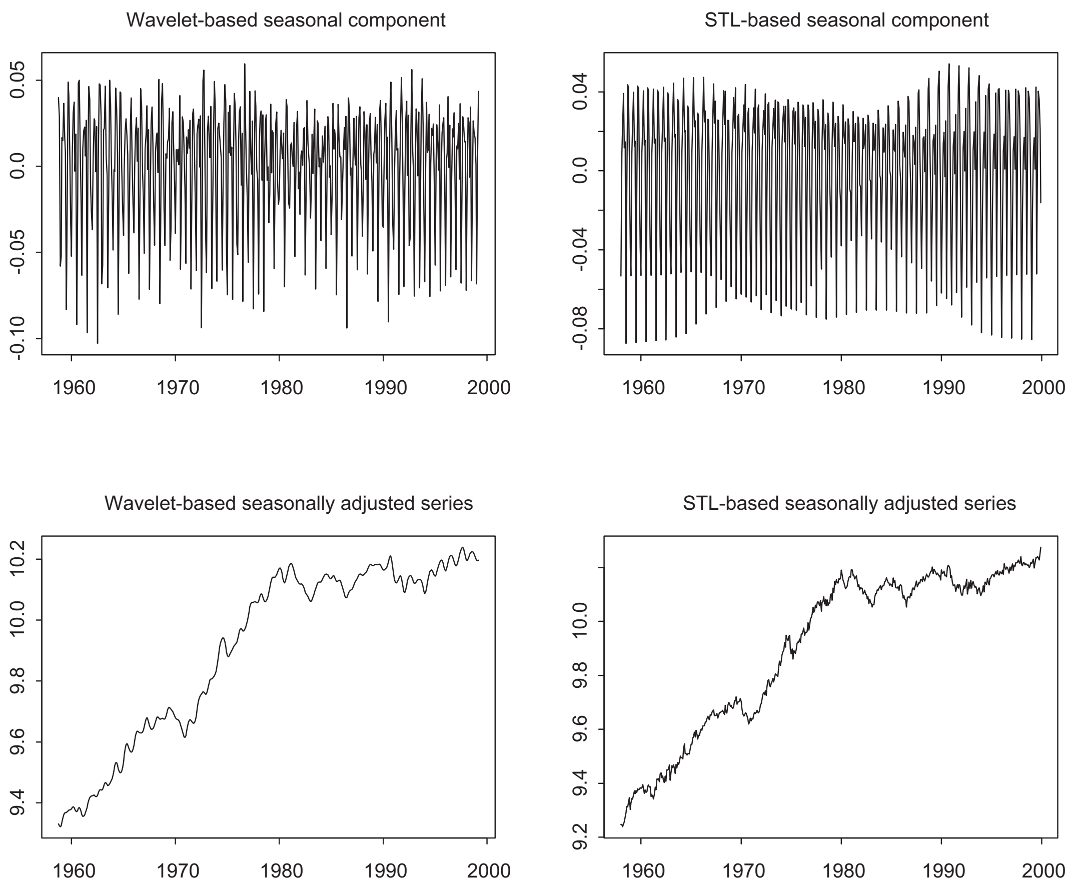

Figure 1. Seasonal decomposition of total manufacturing value of shipments. *Notes*: (1) STL: Cleveland *et al.*'s (1990) trend decomposition procedure based on loess. (2) The wavelet filter utilized is 's8'

In order to decompose the time series of interest into its seasonal and trend components, we utilize an MRD with seven scales and choose the least asymmetric filter (symmlet) 's8'. The seasonal component is reconstructed from scales 1–2, whereas the trend component is obtained from scales 3–7. For illustrative purposes, Figure 1 depicts the seasonal component and the seasonally adjusted series for total manufacturing obtained from wavelets and the STL (seasonal-trend decomposition based on loess) procedure. The latter consists of Cleveland *et al.*'s (1990) trend decomposition procedure, which is based on a locally weighted regression smoother (loess). As we see, the wavelet-based decomposition yields a smoother estimate of the seasonally adjusted series than STL. As a result, a greater proportion of the high-frequency component of the time series is passed on to the seasonal component, relative to STL.

Unobserved components (UC)

The basic univariate UC model is given by

$$y_t = \mu_t + \gamma_t + \xi_t \tag{4}$$

where μ_t is the unobserved trend component, γ_t is the unobserved seasonal component, and ξ_t is the unobserved irregular component (see Harvey, 1993, Ch. 5, or Zivot and Wang, 2003, Ch. 14, for details).

The non-stationary trend component μ_t takes the form of a local linear trend:

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_t \\ \beta_{t+1} &= \beta_t + \varepsilon_t \end{aligned} \tag{5}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ are both white-noise processes, $\mu_1 \sim N(0, \omega)$ and $\beta_1 \sim N(0, \omega)$, with ω large. The stochastic seasonal component is given by

$$\left(1 + L + \ldots + L^{s-1}\right)\gamma_t = \varpi_t \quad \text{or} \quad \gamma_t = -\sum_{j=1}^{s-1}\gamma_{t-j} + \varpi_t \tag{6}$$

where L is the lag operator, s is the number of seasons, and $\varpi_t \sim N(0, \sigma_\varpi^2)$ is a white-noise process. If $\sigma_\varepsilon^2 = 0$, μ_t follows a random-walk process with drift β_1. If $\sigma_\eta^2 = \sigma_\varepsilon^2 = 0$, μ_t is a deterministic trend. Equations (4) and (5) are handled by expressing them in a state-space representation and estimating them by the Kalman filter algorithm.

Support vector machines

SVM is a relatively recent technique within classification methods (e.g., Venables and Ripley, 2002, Ch. 12; Chang and Lin, 2005). It consists of mapping a vector of attributes (i.e., regressors), **x**, into a higher-dimensional space by a function ϕ, and finding a linear maximum-margin hyperplane.[1] That is, we seek a classifying hyperplane of the form $f(\mathbf{x}) = \boldsymbol{\omega}'\phi(\mathbf{x}) + b = 0$. The data points that are located exactly the margin distance away from the hyperplane are denominated the support vectors.[2]

Specifically, the ε-support vector regression (ε-SVR) solves the following quadratic programming problem:

$$\min_{\omega, b, \xi_i^-, \xi_i^+} \frac{1}{2}\omega'\omega + C\sum_{i=1}^{n}(\xi_i^- + \xi_i^+) \tag{7}$$

subject to

$$\begin{gathered} y_i - (\boldsymbol{\omega}'\phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i^- \quad \forall i \\ (\boldsymbol{\omega}'\phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i^+ \\ \xi_i^- \geq 0,\ \xi_i^+ \geq 0 \end{gathered}$$

where $C > 0$ is a penalty parameter and b is a constant term.

The solution to this minimization problem is of the form

$$f(\mathbf{x}) = \sum_{i=1}^{m}(\lambda_i - \lambda_i^*)K(\mathbf{x}_i, \mathbf{x}) + b$$

[1] A maximum-margin hyperplane separates two clouds of points, and it is at equal distance from the two. The smallest distance from the hyperplane is called the margin of separation.
[2] The distance of a point x_i to the hyperplane is given by $|\boldsymbol{\omega}'\phi(\mathbf{x}) + b|/\|\boldsymbol{\omega}\|^2$. The margin distance is given by $2/\|\boldsymbol{\omega}\|$.

where λ_i and λ_i^* are the Lagrange multipliers associated with the constraints $y_i - (\boldsymbol{\omega}'\phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i^-$ and $(\boldsymbol{\omega}'\phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i^+$, respectively. The function $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)'\phi(\mathbf{x}_j)$ represents a kernel, which is the inner product of the two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ in the space $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$. Well-known kernel functions are $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i'\mathbf{x}_j$ (linear), $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\mathbf{x}_i'\mathbf{x}_j + r)^d$, $\gamma > 0$ (polynomial), $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2)$, $\gamma > 0$ (radial basis function), and $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma\mathbf{x}_i'\mathbf{x}_j + r)$ (sigmoid). The radial kernel is a popular choice in the SVM literature. Therefore our computations are based on such a kernel.

FORECASTING ACCURACY EVALUATION

We utilize as forecast evaluation statistics the root mean square error (RMSE) and the mean absolute error (MAE):

$$\text{RMSE}(h) = \sqrt{\frac{1}{T-n-h+1}\sum_{t=n}^{T-h}\hat{u}_{t+h|t}^2} \quad \text{MAE}(h) = \sqrt{\frac{1}{T-n-h+1}\sum_{t=n}^{T-h}|\hat{u}_{t+h|t}|} \tag{8}$$

where T is the sample size, h is the forecast step length and n is the window length.

In order to decide which model is best, we resort to the Diebold–Mariano (1995) test:

$$\text{DM} = \frac{\bar{d}}{\hat{\text{lrv}}(\bar{d})^{1/2}} \xrightarrow{d} N(0,1) \tag{9}$$

where $\bar{d} = \frac{1}{N}\sum_{t=1}^{N} d_t$, $d_t = L(u_{t+h|t}^1) - L(u_{t+h|t}^2)$, $L(.)$ is a loss function, and $\hat{\text{lrv}}(\bar{d})$ is a consistent estimate of the long-run asymptotic variance of $\bar{d}$. In order to compute the latter, Diebold and Mariano recommend using the Newey–West estimator with a rectangular kernel function and a lag truncation parameter equal to $h - 1$. For the RMSE, the loss function is given by $L = (\hat{u}_{t+h|t}^i)^2$, whereas for the MAE it is given by $L = |\hat{u}_{t+h|t}^i|$.

Under the null hypothesis of equal predictive accuracy, $E(d_t) = 0$. For a left-sided test, under the alternative hypothesis model 1 is more accurate than model 2 (i.e., $E(d_t) < 0$). Conversely, for a right-sided test, $E(d_t) > 0$ and model 2 is preferable to model 1.

In order to accommodate heavy-tailed forecasting errors, Harvey *et al.* (1997) suggest utilizing a modified version of the DM statistic:

$$\text{HLN} = \text{DM}\left(\frac{T^*+1-2h+h(h-1)/T}{T^*}\right)^{1/2} \tag{10}$$

which is distributed as $t(T^* - 1)$, $T^* \equiv T - n - h + 1$.

In addition to the DM statistic, we resort to two forecasting evaluation techniques utilized by Fang (2003): forecasting combination and encompassing. The former consists of combining each pair of forecasts (both equally weighted) and computing the RMSE (and/or the MAE) for the combined forecasts. If this is at least 5% less than those of the two corresponding individual forecasts, the combination is preferable.

For forecasting encompassing, one of the specifications utilized by Fang is the following:

$$\Delta_h y_{t+h} = \beta_0 + \beta_1(\hat{y}^{(1)}_{t,t+h} - y_t) + \beta_2(\hat{y}^{(2)}_{t,t+h} - y_t) + u_{t+h} \quad (11)$$

where $\hat{y}_{t,t+h}$ is the forecast of y_{t+h} based on information available at time t, and $\Delta_h y_{t+h} = y_{t+h} - y_t$. (The difference operator is used due to non-stationarity of the time series).[3] When $\beta_1 = 0$ and $\beta_2 \neq 0$, the second model forecast encompasses the first one. Conversely, if $\beta_1 \neq 0$ and $\beta_2 = 0$, the first model forecast encompasses the second one. In the case that both forecasts contain independent information for h-period-ahead forecasting of y_t, both β_1 and β_2 should be different from zero. It is worth noting that no constraint is imposed on the sum $(\beta_1 + \beta_2)$.

Equation (11) can be estimated in principle by ordinary least squares, utilizing standard errors robust to the presence of both heteroskedasticity and serial correlation. Nevertheless, if the two forecast series are highly collinear, Fang advises to resort to ridge regression.

DATA

Our data are US manufacturers' shipments obtained from the US Census Bureau website, www.census.gov. The sectors under consideration are Total manufacturing (MTM), Durable goods total (MDM), Nondurable goods total (MNM), Automotive equipment (AUE), Consumer staples (COS), Business supplies (BUS), Construction supplies (CMS), Total capital goods (TCG), Household durable goods (HDG), and Health care products (HCP). The sample period goes from January 1958 to December 2000. Some descriptive statistics of the seasonally unadjusted series are reported in Table I. Augmented Dickey–Fuller (ADF) and Phillips–Perron unit-root tests show that, at a significance level between 5% and 10%, we cannot reject the presence of a stochastic trend in any of the (seasonally adjusted) time series under consideration.

Table I. Descriptive statistics of US manufacturers' shipments: 1958–2000

	MTM	MDM	MNM	AUE	COS	BUS	CMS	TCG	HDG	HCP
Mean	286,753	151,067	135,687	11,617	62,821	21,059	18,714	44,596	9,129	6,613
Median	301,950	154,873	147,242	11,632	67,740	23,710	19,299	44,359	9,185	5,859
Q_1	243,391	131,248	111,637	9,146	50,370	15,662	15,933	37,669	8,201	3,763
Q_3	331,710	173,009	157,783	14,284	72,569	25,797	21,401	51,487	10,125	9,531
max	416,264	240,960	175,303	19,138	85,395	30,104	26,317	80,528	12,648	15,583
min	149,654	71,435	77,660	3,498	37,687	9,838	9,856	20,244	5,115	2,035
obs	516	516	516	516	516	516	516	516	516	516

Notes:
1. Figures are seasonally unadjusted and expressed in million US dollars of December 2000.
2. The data are measured at a monthly frequency.
3. The data source is the US Census Bureau, www.census.gov.
4. The shipments categories are Total manufacturing (MTM), Durable goods total (MDM), Nondurable goods total (MNM), Automotive equipment (AUE), Consumer staples (COS), Business supplies (BUS), Construction supplies (CMS), Total capital goods (TCG), Household durable goods (HDG), and Health care products (HCP).
5. Q_1 and Q_3 represent the first and third quartiles, respectively.

[3] Given that we utilize the natural logarithm of the shipments series, $\Delta_h y_{t+h} = y_{t+h} - y_t$ represents the return on y between times t and $t + h$.

ESTIMATION RESULTS

The estimation results reported in this section were obtained from routines written by the author in S-Plus 7.0. In addition, the S+Wavelets 2.0 and S+Finmetrics 2.0 modules were utilized for implementing the wavelet-based and unobserved component model spefications, whereas the libsvm S-Plus library, written by David Meyer (based on C/C++ code by Chih-Chung Chang and Chih-Jen Lin from National Taiwan University), was utilized for implementing the SVM technique.[4]

For estimation purposes, the data are expressed in natural logarithms. The ARIMA and UC models estimated are those presented earlier. In order to compute the wavelet-based forecast, we fit an ARIMA$(0, 1, 1)_{12}$ to the seasonal component and an ARIMA(1, 1, 0) to the trend component, which are yielded by the MRD. These two functional forms seem reasonable approximations to the seasonal- and trend-generating processes, respectively. The forecast series is computed as the sum of the seasonal and trend forecasts. For the SVM method, we use 12 lags of the time series as the vector of attributes. Once the separating hyperplane is determined, the fitted model is utilized for out-of-sample forecasting. Given that 12 lags of the dependent variable are needed for forecasting, unknown lagged values are replaced by one-step forecasts. For instance, for a two-step forecast (i.e., $t + 2$), we use as attributes the $t - 10, \ldots, t - 1$ observations, which are known at time t, in addition to the one-step forecast for $t + 1$, which is obtained from the model.

The first exercise we carry out consists of fitting each model to the data series by using the three last years of observations for one-step-ahead forecasting (i.e., 1998–2000). Specifically, at the first stage, we use the observations belonging to 1958–1997 for estimation, and we next add an extra observation at a time to re-estimate the models. We stop once we have utilized the whole sample period for estimation. Our results are summarized in Table II. We observe two patterns. First, the wavelet-based model outperforms the three others in terms of both RMSE and MAE. Second, SVM exhibits the poorest performance, with the exceptions of MDM (Durable goods total) and HDG (Household durable goods), for which the RMSE and MAE of SVM is similar to that of ARIMA. However, if we carry out the HLN test, we find no strong statistical evidence in favor of any model at the conventional significance levels.[5] Indeed, only by considering a 19-percent significance level,

Table II. Assessment statistics of one-step-ahead forecast errors

	MTM	MDM	MNM	AUE	COS	BUS	CMS	TCG	HDG	HCP
(a) RMSE										
ARIMA	0.014	0.020	0.011	0.063	0.011	0.016	0.020	0.026	0.031	0.028
UC	0.012	0.018	0.011	0.061	0.013	0.015	0.020	0.028	0.033	0.028
Wavelet	0.010	0.013	0.008	0.043	0.010	0.011	0.018	0.018	0.020	0.022
SVM	0.021	0.021	0.024	0.085	0.027	0.022	0.033	0.036	0.030	0.035
(b) MAE										
ARIMA	0.012	0.017	0.009	0.052	0.009	0.012	0.016	0.022	0.024	0.022
UC	0.011	0.016	0.009	0.050	0.010	0.012	0.016	0.023	0.026	0.021
Wavelet	0.008	0.010	0.007	0.036	0.008	0.009	0.014	0.014	0.017	0.017
SVM	0.016	0.016	0.020	0.064	0.022	0.018	0.022	0.031	0.024	0.027

[4] Examples on the use of the libsvm library are given in the textbook by Venables and Ripley (2002). Documentation on the SVM technique can be found at Chih-Jen Lin's website, www.csie.ntu.edu.tw/~cjlin/papers/.

[5] When the time series of forecast errors is relatively short, as in this case (=36), the HLN test usually fails to reveal a significance difference between the two models (see Enders, 2004, p. 92).

we would be able to conclude that wavelets outperform SVM for the case of MNM (Nondurable goods total).

In order to have a more complete picture of the forecasting performance of the four models, we next consider from one-step- to 12-step-ahead forecasts. This time, forecasts are computed by using a rolling window of 32 years. Specifically, we hold the last 10 years of data for forecast evaluation (i.e., 1990–2000), obtaining a time series of 132 forecast errors for $h = 1, 2, \ldots, 12$.[6] In order to evaluate the models, in addition to the HLN test, we consider Fang's (2003) forecasting combination and encompassing.

Figure 2 depicts the evolution of the RMSE of forecast errors as the time horizon increases for Durable and Nondurable goods, Consumer staples and Construction supplies. For Durable goods, wavelets outperform the three other specifications for $h < 4$, whereas SVM has the poorest forecast performance for all time horizons. For Nondurable goods, SVM does a better job as it outperforms UC for $h > 6$, while wavelets do worse than the three other methods for $h \geq 4$. For Consumer staples, ARIMA gives the best forecast performance for all $h > 2$, and SVM outperforms wavelets for $h > 4$. Similar results are obtained for Construction supplies, for which ARIMA and UC have almost identical performance, and both do better than SVM and wavelets. Overall, we conclude that the ARIMA and UC specifications tend to outperform the two others, while wavelets may be preferable to SVM in some cases.

If we conduct the HLN test for each category depicted in Figure 2, our above conclusions are, in general, supported. Indeed, for Consumer staples, SVM is rejected in favor of ARIMA and UC at

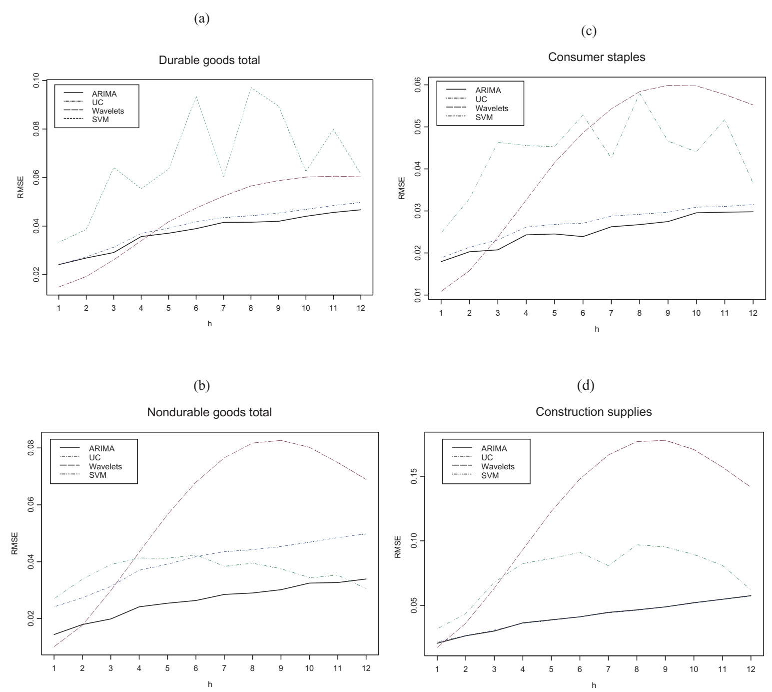

Figure 2. RMSE at different forecasting horizons

[6]The 12 forecasts are computed simultaneously at each iteration.

the 16% and 14% significance levels, respectively, at $h = 6$. For Construction supplies, the wavelet-based specification is rejected in favor of ARIMA and UC at the 15% significance level at $h = 5$ and 10, at the 12% significance level at $h = 6$, and at the 11% significance level at $h = 7$, 8, and 9. For Nondurable and Durable goods, the evidence provided by the HLN test in favor of a particular model is weaker. Truly, for the former, the wavelet-based model is rejected in favor of ARIMA and UC at the 16% significance level at $h = 6$, whereas for the latter ARIMA is preferred to SVM at the 17% significance level at $h = 6$.

As we see, the HLN test helps us to discriminate between models only in some specific cases. Therefore, forecast combination and encompassing arise as complementary tools to discriminate between models. Forecast combination was computed for Total manufacturing, Business supplies, Consumer staples, and Total Capital goods. Our results show that, in terms of RMSE, an equally weighted average of wavelets and ARIMA forecasts can be superior to each single forecast at $h = 3$ (Total manufacturing, Business Supplies, and Consumer staples) and $h = 4$ (Total manufacturing, Consumer staples, and Total capital goods). Likewise for the combination of wavelets and UC forecasts at $h = 3$ (Business supplies), $h = 4$ (Total manufacturing and Consumer staples) and $h = 5$, 10 (Total capital goods); and for the combination of SVM and wavelets forecasts at $h = 4$ (Consumer staples), 5 and 6 (Total manufacturing and Consumer staples).

Forecast-encompassing testing, based on equation (11), was carried out for six shipments categories: Total manufacturing, Durable goods, Nondurable goods, Business supplies, Total capital goods, and Household durable goods.[7] At $h = 6$, ARIMA encompasses UC for all shipment categories, while it encompasses SVM for all categories, except for Business supplies and Total capital goods at the 5% significance level. UC in turn encompasses wavelets for Nondurable goods and Business supplies and SVM for all categories except for Total capital goods. For other pair-wise linear combinations, the two forecasts series are statistically significant, indicating that both contain independent information that is useful to explain the variation in $\Delta_h y_{t+h}$. This is the case, for instance, for UC and wavelets (Total manufacturing, Durable goods, Total capital goods, and Household durable goods), and wavelets and SVM (Total manufacturing and Business supplies).

At $h = 9$, ARIMA's encompassing UC occurs only for Total capital goods. In most cases, the two forecast series contain independent and relevant information. The wavelet-based method is again encompassed by the three others for Nondurable goods and Business supplies. Meanwhile, the performance of SVM relative to the other estimation techniques stays about the same, except for Total capital goods, where it is encompassed by all the other methods at the 1% significance level.

CONCLUSIONS

This article analyzes the forecast performance of four nonlinear methods: seasonal multiplicative ARIMA, unobserved components, wavelets, and SVM. The first two models are well known in the time series literature, whereas the second two models are based on mathematical techniques which

[7] In each case, we computed the condition number of the regressors matrix. This is obtained as the square root ratio of its largest to its smallest characteristic root (after scaling each column so that it has unit length). A condition number exceeding 20 is indicative of the existence of multicollinearity. The greatest condition number we found in our data was around 15. Therefore, we concluded that resorting to ridge regression was unnecessary.

have been developed in recent years. Specifically, wavelets have become popular in economics and finance from the mid 1990s onwards, whereas SVM is a relatively new data classification technique, which is considered as a more user-friendly tool than artificial neural networks.

Our sample is obtained from monthly statistics of US manufacturing shipments for the sample period January 1958 to December 2000. Specifically, we focus on 10 shipment categories, which include, among others, Total manufacturing, Durable goods, Nondurable goods and Consumer staples. Our main findings can be summarized as follows. First, the time horizon is a key element to decide which model or linear combination of models is best in terms of forecast performance. In that regard, forecasting combination and encompassing testing shed more light than the Harvey–Leybourne–Newbold test. Second, in general, the ARIMA and UC techniques are more likely to encompass the wavelet and SVM ones. However, in some cases the latter provide valuable forecasting information that it is not contained in the former. Therefore, pair-wise linear combinations of forecasts can be more informative than those of either ARIMA or UC. Third, it appears that the information provided by SVM forecasts seems most valuable when linear combinations of SVM and wavelets forecasts are considered.

ACKNOWLEDGEMENTS

Financial support from FONDECYT grant no. 1070762 and from an institutional grant of the Hewlett Foundation to CEA is greatly acknowledged. All remaining errors are the author's.

REFERENCES

Arino M. 1995. Time series forecasts via wavelets: an application to car sales in the Spanish market. Discussion Paper 95–30, ISDS, Duke University. http://www.isds.duke.edu/research/papers/ [8 April 2008].

Box G, Jenkins G, Reinsel G. 1994. *Time Series Analysis: Forecasting and Control* (3rd edn). Prentice-Hall, Englewood Cliffs, NJ.

Bruce A, Gao H. 1996. *Applied Wavelet Analysis with S-Plus*. Springer: Berlin.

Burges C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**(2): 955–974.

Chang CC, Lin CJ. 2005. LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin [8 April 2008].

Christianini N, Shawe-Taylor J. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press: Cambridge, UK.

Clements M, Hendry D. 1998. *Forecasting Economic Time Series*. Cambridge University Press: Cambridge, UK.

Cleveland R, Cleveland W, McRae J, Terpening I. 1990. STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* **6**: 3–73.

Conejo A, Contreras J, Espinola R, Plazas M. 2005. Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting* **21**(3): 435–462.

Connor J, Rossiter R. 2005. Wavelet transforms and commodity prices. *Studies in Nonlinear Dynamics and Econometrics* **9**(1): article 6.

Diebold F. 1998. The past, present, and future of macroeconomic forecasting. *Journal of Economic Perspectives* **12**: 175–192.

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

Dong B, Cao C, Lee SE. 2005. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings* **37**(5): 545–553.

Dooley G, Lenihan H. 2005. An assessment of time series methods in metal price forecasting. *Resources Policy* **30**: 208–217.
Enders W. 2004. *Applied Econometric Time Series* (2nd edn). Wiley: Hoboken, NJ.
Fang Y. 2003. Forecasting combination and encompassing tests. *International Journal of Forecasting* **19**(1): 87–94.
Fernandez V. 2005. The international CAPM and a wavelet-based decomposition of value at risk. *Studies of Nonlinear Dynamics and Econometrics* **9**(4): article 4.
Gençay R, Whitcher B, Selçuk F. 2002. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press: San Diego, CA.
Gençay R, Whitcher B, Selçuk F. 2003. Systematic risk and time scales. *Quantitative Finance* **3**(2): 108–116.
Gençay R, Whitcher B, Selçuk F. 2005. Multiscale systematic risk. *Journal of International Money and Finance* **24**(1): 55–70.
Harvey A. 1993. *Time Series Models* (2nd edn). MIT Press: Cambridge, MA.
Harvey D, Leybourne S, Newbold P. 1997. Testing the equality of prediction mean square errors. *International Journal of Forecasting* **13**: 281–291.
Hong Y, Kao C. 2004. Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica* **72**(5): 1519–1563.
Huang W, Nakamori Y, Wang SY. 2005. Forecasting stock market movement direction with support vector machine. *Computers and Operations Research* **32**(10): 2513–2522.
Kim K. 2003. Financial time series forecasting using support vector machines. *Neurocomputing* **55**(1–2): 307–319.
Lanza A, Manera M, Giovannini M. 2005. Modeling and forecasting cointegrated relationships among heavy oil and product prices. *Energy Economics* **27**(6): 831–848.
Leung M, Daouk H, Che A. 2000. Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* **16**(2): 173–190.
Lin S, Stevenson M. 2001. Wavelet analysis of the cost-of-carry model. *Studies in Nonlinear Dynamics and Econometrics* **5**(1): article 7.
Lu WZ, Wang WJ. 2005. Potential assessment of the 'support vector machine' method in forecasting ambient air pollutant trends. *Chemosphere* **59**(5): 693–701.
Mills T. 1999. *The Econometric Modeling of Financial Time Series* (2nd edn). Cambridge University Press: Cambridge, UK.
Morana C. 2001. A semiparametric approach to short-term oil price forecasting. *Energy Economics* **23**(3): 325–338.
Percival D, Walden A. 2000. *Wavelets Analysis for Time Series Analysis.* Cambridge University Press: Cambridge, UK.
Ramsey J, Lampart C. 1998. The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income. *Studies in Nonlinear Dynamics & Econometrics* **3**(1): 23–42.
Ramsey J. 1999. The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions of the Royal Society A* **357**(1760): 2593–2606.
Ramsey J. 2002. Wavelets in Economics and Finance: Past and Future. *Studies in Nonlinear Dynamics & Econometrics* **6**(3): 1–29.
Rapach D, Wohar M, Rangvid J. 2005. Macro variables and international stock return predictability. *International Journal of Forecasting* **21**(1): 137–166.
Roche J. 1995. *Forecasting Commodity Markets.* Probus: London.
Tay F, Cao L. 2001. Application of support vector machines in financial time series forecasting. *Omega* **29**(4): 309–317.
Tsay R. 2005. *Analysis of Financial Time Series* (2nd edn). Wiley: Hoboken, NJ.
Venables W, Ripley B. 2002. *Modern Applied Statistics with S* (4th edn). Springer: New York.
Whitcher B. 2004. Wavelet-based estimation for seasonal long-memory processes. *Technometrics* **46**(2): 225–238.
Zivot E, Wang J. 2003. *Modeling financial times series with S-Plus.* First edition. Insightful Corporation, Seattle, WA.