Trend Time–Series Modeling and Forecasting With Neural Networks

Min Qi and G. Peter Zhang

Abstract—**Despite its great importance, there has been no general consensus on how to model the trends in time–series data. Compared to traditional approaches, neural networks (NNs) have shown some promise in time-series forecasting. This paper investigates how to best model trend time series using NNs. Four different strategies (raw data, raw data with time index, detrending, and differencing) are used to model various trend patterns (linear, nonlinear, deterministic, stochastic, and breaking trend). We find that with NNs differencing often gives meritorious results regardless of the underlying data generating processes (DGPs). This finding is also confirmed by the real gross national product (GNP) series.**

Index Terms—**Difference-stationary (DS) series, forecasting, neural networks (NNs), trend-stationary (TS) series, trend time series.**

I. Introduction

TRADITIONAL analyses of time series were mainly concerned with modeling the autocorrelation structure in a time series, and they typically require that the data under study be stationary. Trends in time series clearly violate the condition of stationarity. Thus, the removal of the trend is often desirable in the time–series analysis and forecasting. For example, the well-known Box–Jenkins [4] approach to time–series modeling relies entirely on the stationarity assumption. The classic decomposition technique decomposes a time series into trend, cycle, and irregular components. The trend is often estimated and removed from the data first before other components are estimated.

A trend-stationary (TS) series can be made stationary by removing its deterministic trend, whereas difference-stationary (DS) series or series with stochastic trend can be made stationary by differencing (TS, DS, and detrending related issues will be discussed in detail in Section II) These two detrending approaches are not equivalent and should not be used interchangeably. Spurious autocorrelations can be induced in a trend time series by either mistakenly removing a deterministic trend from DS data, or differencing TS data (see, e.g., [6], [14], [24], and [35]). In addition, the distinction between TS and DS is also critical from the economic forecasting perspective because these models imply very different dynamics and hence different point forecasts [5], [13], [35]. Using the incorrect detrending form can have markedly different implications for forecasting [8], [13], [15].

It is, however, often difficult to determine whether a given series is TS or DS. Examples can be easily constructed to illustrate the arbitrary closeness of TS and DS models [7]. Although statistical tests such as the Dickey–Fuller test and the Phillips–Perron test, among others, have been developed in the time–series econometrics literature, these tests suffer from very low power in distinguishing between a unit-root and a near-unit-root process. In addition, the testing procedure can be confounded by the presence of the deterministic regressors (i.e., the intercept and deterministic trend), which may further reduce the power of the test.

This practical difficulty is perhaps the reason that more than 20 years after the seminal work of Nelson and Plosser [25] who find that a large number of macroeconomic aggregates follow the DS process, the issue of deterministic versus stochastic trends in the U.S. gross national product (GNP) series remains controversial [23], [26], [33]. For example, Franses and Kleibergen [15] find that the DS processes can adequately describe a large number of macroeconomic aggregates, whereas in [9], [10], [14], [28], and [29], it was found that the case for TS is stronger. Despite these conflicting findings on unit roots, Diebold and Kilian [11] find that pretesting for unit root can improve forecasting accuracy for linear models compared to routinely differencing or never differencing. Thus, it remains unclear what is the best approach to model and forecast trending time series.

Because "data generating processes are unknown and inherently unknowable" [30], it might be beneficial to use robust models that are able to capture the often unidentifiable underlying structure of a time series with few assumptions about the mechanisms of the trend [7]. One such model is the neural networks (NNs). Compared to most traditional forecasting approaches, NN are nonlinear nonparametric adaptive models. They are able to approximate complex relationships without a prespecified model form. During the last decade, NNs have received attention from both practitioners and academics across a wide range of disciplines. They are found to be a viable contender among various time–series models [3], [19], [37] and have been successfully applied to different areas (see [16], [21], and [34], for recent examples). Research efforts in NN forecasting are considerable and the literature has been steadily growing. Several recent studies have attempted to model time series with trend. However, most of them are application specific (e.g., [17], [18], and [27]), and are limited to a particular trend pattern or trend modeling approach (e.g., [18] and [36]). No systematic effort has been devoted to studying the general

Manuscript received February 28, 2007; revised August 16, 2007; accepted September 25, 2007. This work was supported in part by the grant from SAS and International Institute of Forecasters to support research on principles of forecasting. The work of G. P. Zhang was supported by the J. Mack Robinson College of Business, Georgia State University.

M. Qi is with the Risk Analysis Division, Office of the Comptroller of the Currency, Washington, DC 20219 USA (e-mail: min.qi@occ.treas.gov).

G. P. Zhang is with the J. Mack Robinson College of Business, Georgia State University, Atlanta, GA 30303 USA (e-mail: gpzhang@gsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2007.912308

issue of how to use NNs to best model and forecast time series of various trend patterns.

The main purpose of this paper is to explicitly investigate the issue of how to best use NNs to model trend time series. Three research questions are of interest. First, are NNs able to directly model trend time series? Up to now, there has been no theoretical answer to this question. The currently available asymptotic result of NN estimator relies on the assumption of weak stationarity, which does not cover the trend time series. Due to this lack of theoretical rigor plus many practical factors such as data limitation, noise level in the data, complexity of the patterns, and suboptimal training of NNs, it is not clear whether NN models can provide a mistake-proof direct modeling of the trend time series. Second, is inclusion of time index helpful in improving forecasting performance? Traditional regression-based forecasting techniques rely on the inclusion of the time index for trend time–series forecasting. Therefore, it is natural to see if the inclusion of the time index is able to improve forecasting accuracy with NN-based models. Finally and perhaps most importantly, what is the effect of incorrect and correct detrending on NN forecasting performance? More specifically, we would like to know what happens when a detrending method does not match the underlying trend mechanism.

The remainder of this paper is organized as follows. In Section II, we provide a brief discussion on the theoretical issues of trend time series. Section III gives details on research design and methodology. Results are discussed in Section IV. Finally, concluding remarks are offered in Section V.

II. DS and TS Processes in Time–Series Modeling

For illustrative purposes, consider the simple cases of deterministic and stochastic trend models often seen in economic and finance literature

$$y_t = a + bt + \varepsilon_t \tag{1}$$

$$y_t = y_{t-1} + b + \varepsilon_t \tag{2}$$

where $\varepsilon_t \sim iidN(0, \sigma^2)$ and a and b are constant parameters. Model (1) is a TS process as the trend-eliminated series $\hat{\varepsilon}_t$, the least squares residuals, are stationary. Model (2) represents a unit-root process which is DS as first differencing of $y_t, y_t - y_{t-1} = b + \varepsilon_t$, is a stationary process.

We can express model (2) via recursively replacing lagged y_t as

$$y_t = y_0 + bt + \sum_{k=1}^{t} \varepsilon_k \tag{3}$$

where y_0 is the starting point of y. Comparing (1) and (3), one can see that both DS and TS series can be written as a linear function of time (a deterministic trend component) plus a stochastic term. In a TS process (1), however, the intercept a is a constant parameter while in a DS process (3), it is a random starting point. The impact of shocks ε_t is temporary for the TS process, but permanent for the DS process as they accumulate over time. Because of this key difference, forecasts as well as forecast errors can be quite different from the theoretical perspective. For example, for TS processes, forecasts are independent of past shocks while forecasts of the DS process depend on the past shocks. Forecast errors in a TS process are constant over forecasting horizons, but are increasing for a DS process.

As discussed in the introduction, the controversy around trend time–series analysis is mainly about whether economic particularly macroeconomic time series should be modeled as a DS or a TS process. In their seminal work, Nelson and Plosser [25] conclude that most macroeconomic time series can be adequately modeled as DS processes. DeJong and Whiteman [9], [10], however, conclude just the opposite. Mixed findings can be found in many places including [13], [22], and [32].

Although theoretically DS and TS processes have strikingly different properties thus should have potentially very different implications for forecasting [6], [8] [[35], empirical studies give mixed findings regarding the effect of model misspecification. For example, reexamining the Nelson–Plosser data from the out-of-sample forecasting perspective, Franses and Kleibergen [15] find that DS models are regularly preferred to the TS models even if a DS process is not necessarily the true data generating process (DGP). On the other hand, based on a Monte Carlo simulation study, Clements and Hendry [8] find in almost all practical conditions that these two models are "indistinguishable in terms of their implications for predictability" regardless which one is the true DGP.

III. Research Methodology

In practice, NNs are used even when the DGP is nonstationary in which case the asymptotic results for universal approximation are not established in the literature. In order to address research questions expressed previously and provide empirical evidence on the issue of trend time–series modeling with NNs, we perform a Monte Carlo simulation experiment. We investigate a variety of different scenarios in order to represent real-world situations where NNs may be used. We examine relatively simple processes in order to provide manageable and interpretable analysis and insights.

A. Data

We consider six broad types of time–series DGPs: linear deterministic trend (LDT), linear stochastic trend (LST), nonlinear deterministic trend (NDT), nonlinear stochastic trend (NST), deterministic trend with structure break (SBD), and stochastic trend with structure break (SBS). To see the effect of near unit root, we also consider two near-unit-root processes NUR1 and NUR2 with autoregressive coefficient of 0.99 and 0.97, respectively. Specifically, data are generated according to the following DGPs:

- process 1 (LDT)
$$y_t = bt + \varepsilon_t; \tag{4}$$
- process 2 (LST)
$$y_t = b + y_{t-1} + \varepsilon_t; \tag{5}$$
- process 3 (NDT)
$$y_t = bt^2 + \varepsilon_t; \tag{6}$$
- process 4 (NST)
$$y_t = b + y_{t-1} + cy_{t-1}\varepsilon_{t-1} + \varepsilon_t; \tag{7}$$
- process 5 (SBD)
$$y_t = [b_1 \cdot I(t \le \mathrm{TB}) + b_2 \cdot I(t > \mathrm{TB})]t + \varepsilon_t; \tag{8}$$
- process 6 (SBS)
$$y_t = [b_1 \cdot I(t \le \mathrm{TB}) + b_2 \cdot I(t > \mathrm{TB})] + \rho y_{t-1} + \varepsilon_t; \tag{9}$$

- process 7 (NUR1)

$$y_t = b + 0.99y_{t-1} + \varepsilon_t; \tag{10}$$

- process 8 (NUR2)

$$y_t = b + 0.97y_{t-1} + \varepsilon_t; \tag{11}$$

where $\varepsilon_t \sim iidN(0, \sigma^2)$; TB is the time period in which a structural break has happened; I is a logic operator, $I(\text{true}) = 1$, and $I(\text{false}) = 0$. We set $b = 0.0083, \sigma = 0.0435, c = 0.40, b_1 = 0.0051, b_2 = 0.0115$, and TB $= 100$. Some of these parameter values are calibrated from fitting a linear deterministic model to the natural logarithm of the quarterly U.S. GNP (billions of chained 1996 dollars, seasonally adjusted). The data run from the first quarter of 1947 to the third quarter in 2001.

For each DGP, we generate a total of 250 observations for model building and testing. The starting point is randomly generated from a uniform distribution within $[0, 1]$. Using different random seeds for the random shock ε_t, the previous process is replicated 100 times, generating 100 different time series for the same DGP. For consistency and comparison purposes, the same starting point for all series as well as common random errors ε_t is used across models.

The natural logarithm of the real U.S. GNP is also used to test the findings from the simulation study. The entire sample consists of seasonally adjusted quarterly data from the first quarter of 1947 to the third quarter in 2001, a total of 219 observations (billions of chained 1996 dollars; source: St. Louis Federal Reserve Bank). The whole sample is divided into three parts: training (the first 169 observations), validation (the next 25 observations), and testing (the last 25 observations).

B. Neural Networks

The NN models used in this study are the standard three-layer feedforward NNs with only one output node and with nodes in adjacent layers fully connected. The transfer function for hidden nodes is the logistic function and for the output node, the identity function. Bias (or intercept) terms are used in both hidden and output layers. Our NN model can be written as

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j g\left(\sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j}\right) + \varepsilon_t \tag{12}$$

where k is the number of input nodes, n is the number of hidden nodes, and g is a sigmoid transfer function such as the logistic: $g(x) = 1/(1 + \exp(-x))$. $\{\alpha_j, j = 0, 1, \ldots, n\}$ is a vector of weights from the hidden to output nodes and $\{\beta_{ij}, i = 0, 1, \ldots, m; j = 1, 2, \ldots, n\}$ are weights from the input to hidden nodes. α_0 and β_{0j} are weights of arcs leading from the bias terms which have values always equal to 1. Equation (12) can also be written as

$$y_t = f(x_1, \ldots, x_k, \theta) + \varepsilon_t \tag{13}$$

where $f(\cdot)$ represents the NN model and θ is a vector that contains all the parameters in (12).

The Levenberg and Marquardt algorithm (LMA) provided by the MATLAB NN toolbox is employed in training. LMA interpolates between the Gauss–Newton algorithm (GNA) and the steepest descent method. Like GNA, it does not require calculation of the second-order derivatives thus offers fast convergence. It converges to the steepest descent in case the Jacobian has deficient rank thus reduces the conditioning problem of the GNA [20]. The LMA is one of the most popular curve-fitting algorithms, and is especially powerful for the nonlinear least square problem.

To determine the best NN structure for each time series, we use the common practice of cross validation in NN modeling. Each time series is divided into three portions of training, validation, and testing. The training sample is used to estimate the parameters for any specific model architecture. The validation set is then used to select the best model among all models considered. Finally, the selected model is tested with the testing data set.

There is no specific rule governing the splitting of the data in the literature. However, it is generally agreed that most data points should be used for model building. In the simulation study, we use the first 200 observations for training, the next 25 points for validation, and the last 25 points for out-of-sample testing, the same as the validation and testing sample sizes chosen for our real GNP data. Because most macroeconomic time series are quarterly, 25 observations cover more than six years. In addition, the average length of expansion for the U.S. real GDP is 20 quarters and the average length of recession is around three quarters. On average, 25 quarters are long enough to cover an entire U.S. business cycle thus should be enough for out-of-sample testing. Furthermore, with 100 replications, there is ample amount of data points for out-of-sample testing.

C. Research Design

We apply four methods to model each trend time series:

1) modeling with raw original data (denoted as "original"), $y_t = f(y_{t-1}, \ldots, y_{t-k}) + \varepsilon_t$;
2) modeling original data with time index (denoted as "time index"), $y_t = f(t) + \varepsilon_t$;
3) modeling with detrended data (denoted as "detrending"), $y_t^* = f(y_{t-1}^*, \ldots, y_{t-k}^*) + \varepsilon_t, y_t = \hat{a} + \hat{b}t + y_t^*$, where y_t^* is detrended from a linear regression model;
4) modeling with differenced data (denoted as "differencing"), $\Delta y_t = f(\Delta y_{t-1}, \ldots, \Delta y_{t-k}) + \varepsilon_t$, where $\Delta y_t = y_t - y_{t-1}$.

The first two methods are direct modeling approaches in that raw data are used in both input and output layers. The difference between methods (1) and (2) is that the latter is similar to a time–series regression model and we use the time index t associated with each observation as the sole input variable while in the former we try to model directly the relationship between the future time–series value and the past lagged observations. Methods (3) and (4) are indirect approaches as data are detrended either with linear detrending method or with differencing. Therefore, for each data set, we build four models corresponding to the previous four methods.

All NN models are built with the usual cross-validation approach. For time–series methods (1), (3), and (4), we experiment for each time series with 1–5 input nodes and 1–5 hidden nodes, a total of 25 candidate models. On the other hand, for method (2) with time index, only a number of hidden nodes need to be varied and we experiment with 1–5 hidden layer nodes. The performance of in-sample fit and out-of-sample forecast is judged by two commonly used error measures: the root mean squared error (RMSE) and the mean absolute error (MAE). The best model based on the validation set results is retained for forecasting performance evaluation. We convert the forecasts from

TABLE I
SUMMARY RESULTS FOR LINEAR DETERMINISTIC AND STOCHASTIC TREND SERIES

This table reports the mean and standard deviation of the RMSE and MAE from the Monte Carlo simulation of 100 replications for the linear deterministic trend (LDT) and linear stochastic trend (LST) time series, respectively, for the training, validation, and testing sub-samples, and for each modeling strategy: raw original data, original data with time index, detrending, and differencing.

			Method							
			Original Data		Time Index		Detrending		Differencing	
Series	Sample	Measure	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	Training	RMSE	0.0486	0.0031	0.0430	0.0020	0.0412	0.0025	0.0472	0.0038
		MAE	0.0387	0.0025	0.0344	0.0017	0.0327	0.0021	0.0376	0.0032
LDT	Validation	RMSE	0.0487	0.0076	0.0447	0.0071	0.0412	0.0070	0.0455	0.0074
		MAE	0.0393	0.0066	0.0358	0.0060	0.0327	0.0058	0.0362	0.0061
	Testing	RMSE	0.0549	0.0084	0.0516	0.0128	0.0457	0.0083	0.0496	0.0083
		MAE	0.0445	0.0070	0.0423	0.0118	0.0366	0.0064	0.0403	0.0070
	Training	RMSE	0.0420	0.0023	0.0816	0.0253	0.0414	0.0023	0.0411	0.0025
		MAE	0.0335	0.0019	0.0652	0.0210	0.0329	0.0019	0.0326	0.0021
LST	Validation	RMSE	0.0463	0.0126	0.1360	0.0958	0.0454	0.0109	0.0413	0.0070
		MAE	0.0373	0.0108	0.1171	0.0908	0.0369	0.0098	0.0331	0.0059
	Testing	RMSE	0.0630	0.0351	0.2930	0.1831	0.0623	0.0335	0.0464	0.0096
		MAE	0.0541	0.0339	0.2747	0.1812	0.0516	0.0301	0.0367	0.0065

the indirect methods (detrending and differencing) back to the original scale, thus, the model fitting and forecasting performance are directly comparable to those of the direct methods (raw data and raw data with time index).

As each type of time series has 100 replications, we repeat the previous modeling process for each replication yielding 100 RMSE and MAE for each type. Analysis of variance (ANOVA) test procedures are employed to determine if the mean performance measures are statistically different among the four methods. Tukey's honesty significant difference (HSD) tests [31] are then used to identify the significantly different methods in multiple pairwise comparisons.

IV. RESULTS

Tables I–IV report the summary statistics from the Monte Carlo simulation for linear, near unit roots, nonlinear, and breaking trend time series, respectively. Each table gives the mean and standard deviation of the RMSE and MAE of 100 replications for the training, validation, and testing subsamples and for each modeling strategy. Following [1], we report results for training, validation, and testing to reveal the effectiveness of model implementation. It is not surprising that across Tables I–IV, model fitting and forecasting performance generally gets worse from training to validation to testing subsamples for each of the four methods. However, the degradation in error measures is generally within a reasonable range, indicating our model implementation is adequate.

Table I shows that for LDT series, modeling with detrended data is the best approach overall judging by both RMSE and MAE in all three samples of training, validation, and testing. On the contrary, for LST series, modeling with differenced data consistently outperforms other modeling approaches. These results are expected given the different properties of trend stationary and difference stationary time series.

Considering the underlying DGP, it is not surprising to observe that using time index t for modeling and predicting LST is much worse than for LDT as reflected by both error measures in all three samples. In addition, the variation measured by the standard deviation is also much higher for LST than for LDT. What is important to note is that using a time–series regression structure with time index t as the sole input is not the best way to model an NN forecaster even for LDT series. Our results suggest that differencing or detrending helps improve modeling and forecasting accuracy.

Theoretically, with the autoregressive coefficients of 0.99 and 0.97, the two near-unit-root time series are stationary, however, in practice, it is often difficult to distinguish near-unit-root time series from those with unit root. Table II presents summary statistics from the simulation study for these two near-unit-root situations. It is interesting to find that results for both situations are remarkably similar in terms of both average performance measures and standard deviations in all three subsamples across four different modeling strategies. Similar to the unit-root case, using time index is not helpful at all in both modeling and forecasting these stochastic processes. It generates considerably larger errors in all three samples, particularly in the testing sample. The performances of the other three methods (i.e., differencing, detrending, and original) are close, especially for the training and validation subsamples. The best performer across all subsamples, though, is modeling with the differenced data, the same finding as in the unit-root situation. Therefore, for stochastic linear trending time series, even though the process is near unit root, differencing can still be used.

The practical significance of this finding is clear as unit-root tests often have low power in distinguishing between the unit-root and the near-unit-root processes, yet, from the NN modeling and forecasting perspective, it does not really matter. The

TABLE II
SUMMARY RESULTS FOR NEAR-UNIT-ROOT PROCESSES

This table reports the mean and standard deviation of the RMSE and MAE from the Monte Carlo simulation of 100 replications for near unit root 1 (NUR1) and near unit root 2 (NUR2) time series, respectively, for the training, validation, and testing sub-samples, and for each modeling strategy: raw original data, original data with time index, detrending, and differencing.

			Method							
			Original Data		Time Index		Detrending		Differencing	
Series	Sample	Measure	Mean	Std	Mean	Std	Mean	Std	Mean	Std
NUR1	Training	RMSE	0.0414	0.0025	0.0756	0.0194	0.0416	0.0025	0.0413	0.0023
		MAE	0.0331	0.0022	0.0597	0.0156	0.0332	0.0022	0.0326	0.0019
	Validation	RMSE	0.0440	0.0094	0.1210	0.0682	0.0465	0.0139	0.0415	0.0071
		MAE	0.0351	0.0080	0.1040	0.0640	0.0375	0.0121	0.0331	0.0060
	Testing	RMSE	0.0555	0.0302	0.2034	0.1216	0.0705	0.0540	0.0461	0.0082
		MAE	0.0456	0.0278	0.1875	0.1209	0.0600	0.0510	0.0367	0.0063
NUR2	Training	RMSE	0.0414	0.0025	0.0707	0.0180	0.0416	0.0024	0.0410	0.0025
		MAE	0.0330	0.0022	0.0558	0.0147	0.0332	0.0021	0.0326	0.0022
	Validation	RMSE	0.0445	0.0106	0.1208	0.0688	0.0464	0.0142	0.0415	0.0070
		MAE	0.0356	0.0094	0.1039	0.0644	0.0375	0.0126	0.0333	0.0058
	Testing	RMSE	0.0546	0.0254	0.2023	0.1300	0.0713	0.0554	0.0461	0.0079
		MAE	0.0448	0.0229	0.1870	0.1297	0.0609	0.0527	0.0371	0.0064

TABLE III
SUMMARY RESULTS FOR NONLINEAR DETERMINISTIC AND STOCHASTIC TREND SERIES

This table reports the mean and standard deviation of the RMSE and MAE from the Monte Carlo simulation of 100 replications for the nonlinear deterministic trend (NDT) and nonlinear stochastic trend (NST) time series, respectively, for the training, validation, and testing sub-samples, and for each modeling strategy: raw original data, original data with time index, detrending, and differencing.

			Method							
			Original Data		Time Index		Detrending		Differencing	
Series	Sample	Measure	Mean	Std	Mean	Std	Mean	Std	Mean	Std
NDT	Training	RMSE	0.0750	0.0188	0.0445	0.0030	0.0899	0.0183	0.0624	0.0050
		MAE	0.0596	0.0150	0.0355	0.0024	0.0716	0.0144	0.0497	0.0040
	Validation	RMSE	0.5316	0.1810	0.4055	0.2005	0.3298	0.0473	0.0620	0.0100
		MAE	0.4496	0.1631	0.3352	0.1613	0.2735	0.0403	0.0501	0.0086
	Testing	RMSE	2.6098	0.9617	2.2360	1.4709	1.6797	0.2104	0.0691	0.0137
		MAE	2.4087	0.8730	2.0592	1.3221	1.5446	0.1973	0.0561	0.0118
NST	Training	RMSE	0.0429	0.0032	0.0979	0.0294	0.0435	0.0040	0.0430	0.0032
		MAE	0.0341	0.0026	0.0769	0.0232	0.0344	0.0031	0.0341	0.0026
	Validation	RMSE	0.0542	0.0183	0.2261	0.1333	0.0561	0.0179	0.0488	0.0115
		MAE	0.0436	0.0155	0.1963	0.1228	0.0453	0.0159	0.0395	0.0094
	Testing	RMSE	0.0814	0.0498	0.4298	0.2725	0.0938	0.0665	0.0560	0.0153
		MAE	0.0685	0.0461	0.3990	0.2693	0.0776	0.0583	0.0447	0.0122

best modeling strategy for both unit-root and near-unit-root processes is the same: first differencing.

Table III summarizes the results for two nonlinear trend series. It is clear that using time index is not able to model and forecast the nonlinear time series well. For both deterministic trend case (NDT) and stochastic trend process (NST), differencing is the best approach in all three samples judged by both mean and standard deviation of the two error measures. In addition, differencing is particularly effective for predicting the NDT process as the testing sample performance of differencing is considerably better than that of other methods. The reason that detrending does not do well for this deterministic nonlinear trend series may be that the linear detrending method is used while the underlying DGP is a nonlinear trend. Given the practical difficulty in identifying correct nonlinear form and the very good performance of the differencing method, it is reasonable to recommend differencing for nonlinear trend time series.

For breaking trend processes, Table IV shows that time index method again is the worst among all methods in both model fit-

TABLE IV
SUMMARY RESULTS FOR DETERMINISTIC AND STOCHASTIC BREAKING TREND SERIES

This table reports the mean and standard deviation of the RMSE and MAE from the Monte Carlo simulation of 100 replications for the deterministic trend with structural break (SBD) and stochastic trend with structural break (SBS) time series, respectively, for the training, validation, and testing sub-samples, and for each modeling strategy: raw original data, original data with time index, detrending, and differencing.

			Method							
			Original Data		Time Index		Detrending		Differencing	
Series	Sample	Measure	Mean	Std	Mean	Std	Mean	Std	Mean	Std
	Training	RMSE	0.0599	0.0111	0.0444	0.0031	0.0514	0.0082	0.0643	0.0069
		MAE	0.0416	0.0048	0.0353	0.0025	0.0388	0.0041	0.0418	0.0034
SBD	Validation	RMSE	0.0506	0.0080	0.0703	0.0352	0.0449	0.0067	0.0464	0.0075
		MAE	0.0408	0.0070	0.0588	0.0329	0.0361	0.0057	0.0370	0.0061
	Testing	RMSE	0.0601	0.0108	0.1488	0.0687	0.0504	0.0072	0.0521	0.0096
		MAE	0.0491	0.0095	0.1368	0.0683	0.0410	0.0063	0.0421	0.0073
	Training	RMSE	0.0656	0.0095	0.0594	0.0042	0.0560	0.0088	0.0657	0.0073
		MAE	0.0451	0.0045	0.0474	0.0034	0.0419	0.0045	0.0434	0.0037
SBS	Validation	RMSE	0.0542	0.0084	0.0945	0.0394	0.0485	0.0076	0.0486	0.0088
		MAE	0.0435	0.0072	0.0787	0.0351	0.0385	0.0065	0.0388	0.0076
	Testing	RMSE	0.0635	0.0123	0.1799	0.0892	0.0537	0.0100	0.0532	0.0090
		MAE	0.0512	0.0096	0.1643	0.0884	0.0434	0.0085	0.0430	0.0074

ting and forecasting. The two detrended methods are very close in performance along all comparison dimensions, though detrending is slightly better than differencing for the SBD process while differencing is slightly better than detrending for the SBS series. (The Tukey's HSD tests show that these two strategies are not statistically significantly different from each other.) Relative to preprocessing strategies (differencing or detrending), direct modeling with the raw data generates average error measures that are almost 20% larger.

For each combination of DGP and performance measure, we perform an ANOVA procedure to determine if there exists statistically significant difference among the four approaches in out-of-sample forecasting. The results are omitted to save space. All ANOVA results are highly significant (p-value < 0.001), suggesting that there are significant differences among the methods for trend time–series forecasting.

To identify the significant difference between any two methods, we use Tukey's HSD tests to compare all pairwise differences simultaneously. Results of these multiple comparison tests are reported in Table V. To facilitate presentation, for each type of series, we rank order the methods from 1 (the best) to 4 (the worst), with the best method (with the lowest overall error measure) ranked as 1, and the next one as 2, and so on. Several observations can be made from Table V. First, for each DGP, based on either RMSE or MAE, the ranking of each modeling strategy stays the same, thus our findings are robust to the choice of error measures. Second, the time index method ranks the fourth for all DGPs except for LDT and NDT where it ranks the third. Third, methods based on the preprocessed data by either detrending or differencing outperform methods based on raw data (original and time index). Finally, for stochastic processes such as LST (including the near-unit-root cases), NST, and SBS, the best method is differencing, although in most cases, the differences among differencing, detrending, and original are not significant at the 0.05 level. To our surprise, differencing is also the best approach for the NDT process and it is significantly different from the other three methods. On the other hand, detrending is the best method for two deterministic processes of LDT and SBD. In the LDT process, detrending significantly outperforms differencing. However, the difference between them is not significant in the other case.

Real GNP time series is used to examine if the results from our experimental investigation match with those from the real-world application. Table VI gives RMSE and MAE for the training, validation, and testing samples. In addition to the entire testing sample, we also report results for forecasting horizons of two years and four years. It is clear that differencing is the best method in all three subsamples judged by both performance measures of RMSE and MAE. From the out-of-sample forecasting performance, we find that time index and detrending methods are much worse than differencing. In addition, although the difference between differencing and original appears not large, differencing still achieves 22–25% and 10–21% reductions in RMSE and MAE, respectively, over the method that relies on the original data at various forecasting horizons. Given the dominant traditional view of the stochastic rather than deterministic nature of GNP data, the results in Table VI reinforce our earlier findings on the effectiveness of differencing for the stochastic processes.

We also apply Diebold and Mariano [12] test with Andrews [2] optimal fixed bandwidth parameters to check the significance of the difference between two adjacent methods. Results of the pairwise comparison tests are reported in Table VII. Just as in Table V, we rank order the methods from 1 (the best) to 4 (the worst), with the best method (with the lowest overall error measure) ranked as 1, and the next one as 2, and so on.

TABLE V

MULTIPLE COMPARISON RESULTS WITH RANKED METHOD FOR OUT-OF-SAMPLE FORECASTING

This table reports the results of the Tukey's HSD test to identify the significant difference between any two methods, with all pair-wise differences compared simultaneously. For each type of series, we rank order the methods from 1 (the best) to 4 (the worst), with the best method (with the lowest overall error measure) ranked as 1, and the next one as 2, and so on. * indicates the mean difference between the two adjacent methods is significant at the 0.05 level.

		Rank of Methods						
Series	Measure	1		2		3		4
LDT	RMSE	Detrending	<*	Differencing	<*	Time Index	<	Original Data
	MAE	Detrending	<*	Differencing	<*	Time Index	<	Original Data
LST	RMSE	Differencing	<	Detrending	<	Original Data	<*	Time Index
	MAE	Differencing	<	Detrending	<	Original Data	<*	Time Index
NUR1	RMSE	Differencing	<	Original Data	<	Detrending	<*	Time Index
	MAE	Differencing	<	Original Data	<	Detrending	<*	Time Index
NUR2	RMSE	Differencing	<	Original Data	<	Detrending	<*	Time Index
	MAE	Differencing	<	Original Data	<	Detrending	<*	Time Index
NDT	RMSE	Differencing	<*	Detrending	<*	Time Index	<*	Original Data
	MAE	Differencing	<*	Detrending	<*	Time Index	<*	Original Data
NST	RMSE	Differencing	<	Original Data	<	Detrending	<*	Time Index
	MAE	Differencing	<	Original Data	<	Detrending	<*	Time Index
SBD	RMSE	Detrending	<	Differencing	<	Original Data	<*	Time Index
	MAE	Detrending	<	Differencing	<	Original Data	<*	Time Index
SBS	RMSE	Differencing	<	Detrending	<	Original Data	<*	Time Index
	MAE	Differencing	<	Detrending	<	Original Data	<*	Time Index

TABLE VI

RESULTS FOR THE U.S. REAL GNP SERIES

This table gives RMSE and MAE for the training, validation, and testing samples of the real GNP time series for each of the four modeling strategies: raw original data, original data with time index, detrending, and differencing.

Sample	Measure	Original Data	Time Index	Detrending	Differencing
Training	RMSE	0.0104	0.0201	0.0111	0.0097
	MAE	0.0079	0.0152	0.0085	0.0071
Validation	RMSE	0.0047	0.0286	0.0219	0.0046
	MAE	0.0038	0.0271	0.0203	0.0037
Testing	RMSE	0.0068	0.0315	0.0411	0.0051
(2-yr horizon)	MAE	0.0061	0.0310	0.0409	0.0048
Testing	RMSE	0.0073	0.0243	0.0513	0.0057
(4-yr horizon)	MAE	0.0059	0.0220	0.0501	0.0052
Testing	RMSE	0.0080	0.0203	0.0665	0.0062
(25-qtr horizon)	MAE	0.0062	0.0167	0.0630	0.0056

It is apparent from Table VII that regardless of the forecasting horizons and the squared or absolute differences, the ranking of each modeling strategy stays the same—differencing is the best method, followed by original data, then time index, and finally detrending.

V. CONCLUDING REMARKS

Trend time–series modeling and forecasting is an important topic in many business and economic settings. Although numerous studies find that DS and TS models can imply very different predictions, research findings regarding whether a particular time series is DS or TS process are often inconsistent. Because of the controversy around the nature of the trend in economic and business time series, it is often difficult for applied forecasters to build an effective model for forecasting.

In this paper, we conducted a Monte Carlo study to address the question: what is the most effective way to model and forecast trend time series with NNs, a recent popular nonlinear modeling tool? We examined a variety of different underlying DGPs that have different trend mechanisms (linear, nonlinear, deterministic, stochastic, and breaking trends). Four strategies of direct modeling with the raw data, using raw data and the time index, modeling with linearly detrended data, and modeling with differenced data were considered. While the results do not give a clear-cut universal answer to the previous

TABLE VII
SIGNIFICANCE OF DIFFERENCE IN OUT-OF-SAMPLE FORECASTING ACCURACY BETWEEN TWO ALTERNATIVE METHODS

This table reports the results of the Diebold and Mariano (1995) test for the significant difference between two forecasting methods. For each forecasting horizon, we rank order the methods from 1 (the best) to 4 (the worst), with the best method (with the lowest overall error measure) ranked as 1, and the next one as 2, and so on. * indicates the mean difference between the two adjacent methods is significant at the 0.05 level, and ** indicates significance at the 0.01 level.

		Rank of Methods						
Series	Measure	1		2		3		4
2-yr	Squared Error	Differencing	<*	Original Data	<**	Time Index	<*	Detrending
Horizon	Absolute Error	Differencing	<**	Original Data	<**	Time Index	<**	Detrending
4-yr	Squared Error	Differencing	<**	Original Data	<**	Time Index	<**	Detrending
Horizon	Absolute Error	Differencing	<**	Original Data	<**	Time Index	<**	Detrending
25-qtr	Squared Error	Differencing	<**	Original Data	<*	Time Index	<**	Detrending
Horizon	Absolute Error	Differencing	<**	Original Data	<**	Time Index	<**	Detrending

question, they do provide insights on the modeling issues for trend time–series forecasting.

Our results show that only for the LDT case, linear detrending is the most effective way for NNs to significantly outperform other methods in out-of-sample forecasting performance. Except for the deterministic structure break process for which detrending and differencing are not significantly different, in all other situations including both nonlinear trend and stochastic trend cases, differencing is the most effective. As most real-world time series are nonlinear and/or stochastic, we may be able to conclude that differencing data first is the best practical approach to building an effective NN forecasting model. This is an important suggestion for applied forecasters as one of the most significant but controversial practical issues to them is whether a time series under study contains unit root and whether a specific unit-root test has sufficient power to detect it. The findings in our study imply that from the NN forecasting perspective, the power of unit-root tests should not be a concern because whether a time series contains unit root or not, differencing is always the best approach. Our recommendation for NN modeling and forecasting thus complements those of Franses and Kleibergen [15] who recommend always differencing for linear models.

The limitations of this study lie in several aspects. First, although we have examined a variety of simulated trend time series that are likely encountered in real applications, there are many other possible types of trend time series, which may shed a different light on the modeling issue. Second, our simulation study focuses on relatively simple lag structure of the DGP. Further research is needed to verify the robustness of the findings for processes with richer dynamics. Third, we have used only one real data series to verify our findings in the Monte Carlo study. Future research could examine more real data with different frequencies to substantiate our findings. Finally, although it is highly desirable to calibrating the simulation parameters to real data series (quarterly real GNP in our case), it is not clear whether the same conclusion will hold if different simulation parameters are used. Future research could also experiment with different simulation parameters.

REFERENCES

[1] M. Adya and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation," *J. Forecasting*, vol. 17, pp. 481–495, 1998.

[2] D. W. K. Andrews, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, vol. 59, no. 3, pp. 817–858, 1991.

[3] S. D. Balkin and J. K. Ord, "Automatic neural network modeling for univariate time series," *Int. J. Forecasting*, vol. 16, no. 4, pp. 509–515, 2000.

[4] G. E. P. Box and G. Jenkins*, Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.

[5] J. Y. Campbell and P. Perron, "Pitfalls and opportunities: What macroeconomists should know about unit roots," in *NBER Macroeconomics Annual*, O. Blanchard and S. Fischer, Eds. Cambridge, MA: MIT Press, 1991, pp. 141–201.

[6] K. H. Chan, J. C. Hayya, and J. K. Ord, "A note on trend removal methods: The case of polynomial regression versus variate differencing," *Econometrica*, vol. 45, no. 3, pp. 737–744, 1977.

[7] C. Chatfield*, Time-Series Forecasting*. Boca Raton, FL: Chapman & Hall/CRC, 2001.

[8] M. P. Clements and D. F. Hendry, "Forecasting with difference-stationary and trend-stationary models," *Econometrics J.*, vol. 4, pp. S1–S19, 2001.

[9] D. N. DeJong and C. H. Whiteman, "Reconsidering 'Trends and random walks in macroeconomic time series'," *J. Monetary Econom.*, vol. 28, no. 2, pp. 221–254, 1991.

[10] D. N. DeJong and C. H. Whiteman, "The case for trend-stationarity is stronger than we thought," *J. Appl. Econometrics*, vol. 6, no. 4, pp. 413–421, 1991.

[11] F. X. Diebold and L. Kilian, "Unit-root tests are useful for selecting forecasting models," *J. Business Econom. Statist.*, vol. 18, no. 3, pp. 265–273, 2000.

[12] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Business Econom. Statist.*, vol. 13, no. 3, pp. 253–263, 1995.

[13] F. X. Diebold and A. Senhadji, "The uncertain unit root in real GNP: Comment," *Amer. Econom. Rev.*, vol. 86, pp. 1291–1298, 1996.

[14] S. N. Durlauf and P. C. B. Phillips, "Trends versus random walks in time series analysis," *Econometrica*, vol. 56, no. 6, pp. 1333–1354, 1988.

[15] P. H. Franses and F. Kleibergen, "Unit roots in the Nelson-Plosser data: Do they matter for forecasting?," *Int. J. Forecasting*, vol. 12, pp. 283–288, 1996.

[16] R. Gencay and R. Gibson, "Model risk for European-style stock index options," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 193–202, Jan. 2007.

[17] F. Gonzalez-Romera, M. A. Jaramillo-Moran, and D. Carmona-Fernandez, "Monthly electric energy demand forecasting based on trend extraction," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1946–1953, Apr. 2006.

[18] J. V. Hansen and R. D. Nelson, "Forecasting and recombining time-series components by using neural networks," *J. Operat. Res. Soc.*, vol. 54, pp. 307–317, 2003.

[19] T. Hill, M. O'Connor, and W. Remus, "Neural network models for time series forecasts," *Manage. Sci.*, vol. 42, pp. 1082–1092, 1996.

[20] K. L. Judd, *Numerical Methods in Economics*. Cambridge, MA: MIT Press, 1998, pp. 117–119.

[21] Y.-K. Kwon and B.-R. Moon, "A hybrid neurogenetic approach for stock forecasting," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 851–864, May 2007.

[22] A. Lucas, "An outlier robust unit root test with application to the extended Nelson-Plosser data," *J. Econom.*, vol. 66, pp. 153–173, 1995.

[23] C. J. Murray and C. R. Nelson, "The uncertain trend in U.S. GDP," *J. Monetary Econom.*, vol. 46, no. 1, pp. 79–95, 2000.

[24] C. R. Nelson and H. Kang, "Spurious periodicity in inappropriately detrended time series," *Econometrica*, vol. 49, pp. 741–751, 1981.

[25] C. R. Nelson and C. I. Plosser, "Trends and random walks in macroeconomic time series," *J. Monetary Econom.*, vol. 10, pp. 139–162, 1982.

[26] P. Newbold, S. Leybourne, and M. E. Wohar, "Trend-stationarity, difference-stationarity, or neither: Further diagnostic tests with an application to U.S. real GNP," *1875–1993, J. Econom. Business*, vol. 53, no. 1, pp. 85–102, 2001.

[27] K. Papagianaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of internet backbone traffic," *IEEE Trans. Neural Netw.*, vol. 16, no. 5, pp. 1110–1124, Sep. 2005.

[28] P. Perron, "The great crash, the oil price shock and the unit root hypothesis," *Econometrica*, vol. 57, pp. 1361–1401, 1989.

[29] P. Perron, "Further evidence on breaking trend functions in macroeconomic variables," *J. Econom.*, vol. 80, pp. 355–385, 1997.

[30] P. C. B. Phillips, "Laws and limits of econometrics The Ohio State Univ.. Columbus, 2002, speech given at G.S. Maddala Memorial Lecture.

[31] F. L. Ramsey and D. W. Shafer, *The Statistical Sleuth*. Belmont, CA: Duxbury, 1997.

[32] G. R. Rudebusch, "Trends and random walks in macroeconomic time series: A re-examination," *Int. Econom. Rev.*, vol. 33, pp. 661–680, 1992.

[33] P. C. Schotman and J. K. van Dijk, "On Bayesian routes to unit roots," *J. Appl. Econometrics*, vol. 6, pp. 387–401, 1991.

[34] Z. Shi and M. Han, "Support vector echo-state machine for chaotic time-series prediction," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 359–372, Mar. 2007.

[35] J. H. Stock and M. W. Watson, "Variable trends in economic time series," *J. Econom. Perspectives*, vol. 2, no. 3, pp. 147–174, 1988.

[36] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *Eur. J. Operat. Res.*, vol. 160, pp. 501–514, 2005.

[37] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks," *Int. J. Forecasting*, vol. 14, pp. 35–62, 1998.

Min Qi received the M.A. and Ph.D. degrees in economics from the Ohio State University, Columbus, in 1992 and 1996, respectively.

Currently, she is a Senior Financial Economist in the Risk Analysis Division of the Office of the Comptroller of the Currency (OCC), Washington, DC. She has been involved in the development of policy and guidance related to downturn LGD of the Basel II capital framework. She has served as a Quant Expert in bank exams on credit scoring, pricing, loan loss reserves, and Basel II IRB. She has published extensively on economics and finance related topics in academic journals, such as the IEEE TRANSACTIONS ON NEURAL NETWORKS, *Journal of Money, Credit, and Banking*, *Journal of Futures Markets*, *Journal of Economic Psychology*, *Journal of Forecasting*, and *Journal of Business and Economic Statistics*. Prior to joining the OCC in 2004, She was an Associate Professor of Economics at Kent State University.

Dr. Qi won the second place in the Financial Series Competition at the 1999 International Joint Conference on Neural Networks.

G. Peter Zhang received the B.S. degree in mathematics and the M.S. degree in statistics from East China Normal University, Shanghai, China, in 1985 and 1987, respectively, and the Ph.D. degree in operations management from Kent State University, Kent, OH, in 1998.

Currently, he is an Associate Professor of Managerial Sciences at Georgia State University, Atlanta. His research interests include neural networks, forecasting, and supply chain management. His research was published in the IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *Decision Sciences*, *European Journal of Operational Research*, *IIE Transactions*, *International Journal of Forecasting*, *International Journal of Production Economics*, *Journal of the Operational Research Society*, *Neurocomputing*, among others.

Dr. Zhang currently serves as an Associate Editor of *Neurocmputing* and *Forecasting Letters* and is on the editorial review board of *Production and Operations Management* and *International Journal of E-Business Research*.