A New Method for Interval Forecasting of Autoregressive Time Series with a Root Near 1

Staffan A. Fredricsson, PhD.
staffan@pacbell.net, 5278 Salisbury Drive, Newark, CA 94560

Abstract

The Scale Factor (SF) method is presented, to improve the accuracy of multiple step ahead interval forecasts for autoregressive time series with a trend and a root near unity. For this case, the inadequacy of established regression-style methods for model fitting were broadly exposed in the seminal 1982 paper by Nelson and Plosser. When the characteristic polynomial has a root near 1, bias with respect to the parameter estimates as well as the prediction interval width present great problems. The parameter estimate bias has since been addressed by several authors, and the SF method adopts a median-unbiased approach.

The focus in this paper is on the prediction interval width problem. A base width is obtained using GLS and then de-biased using a multiplicative scale factor, determined using simulation and numerical optimization techniques. The substantial benefits of the SF method compared to alternatives, are first demonstrated using simulated $AR(1)$ processes and actual coverage probability accuracy. In addition, the SF method and alternatives are applied to the original Nelson-Plosser AR(p) data set, with forecasts compared to actual data through 2010.

Keywords: forecast, autoregressive, unit root, coverage

1. Background

The nature of an autoregressive process changes quite dramatically as a root of its characteristic equation approaches unity. For the purpose of prediction, it is common to first attempt to classify a process as either difference-stationary (DS) or trend-stationary (TS), depending on if it is judged to possess a unit root or not. In this classification, facilitated by the unit-root test developed by [Dickey and Fuller, 1979] and made popular[1] by [Nelson and Plosser, 1982], it is common to accept the unit-root hypothesis as (at least tentatively) true, unless it is rejected. As a consequence, there is a tendency for processes with a root close to but less than one, to either have that root overestimated to unity if failing to reject, or to have the root underestimated due to bias if rejecting the unit-root hypothesis. For near unit-root processes, the Goldilocks solution of "just right" is clearly missing with this approach.

[1] "One of the most influential papers in macroeconomics during the last decade..." [Rudebusch, 1992]

In virtually all practical forecasting situations, the process parameters are unknown and have to be estimated. The problem of estimating process parameters has been studied extensively, and a comprehensive literature review is provided in [Falk and Roy, 2006]. An issue with many estimators, of particular interest in this paper, is that they produce biased estimates. Glenn Rudebush was an early proponent of using Monte Carlo methods to achieve median-unbiased parameter estimates [Rudebusch, 1992]. Donald Andrews introduced an exactly median-unbiased estimator of the autocorrelation parameter for $AR(1)$ series, based on Monte Carlo methods [Andrews, 1993], and in cooperation with Hong-Yan Chen suggested an approximate method for the more general $AR(p)$ case [Andrews and Chen, 1994]. Furthermore, Anindya Roy and Wayne Fuller introduced an approximately unbiased estimator in [Roy and Fuller, 2001].

For the purpose of interval forecasting, prediction intervals are conventionally determined as if the estimated process parameters equal the true value, typically resulting in overly optimistic confidence levels. The use of median-unbiased parameter estimates offer improvements compared to biased alternatives, but the resulting coverage rates remain severely biased due to the asymmetric effects of parameter estimate variations. The proposed method, which I call the Scale Factor (SF) method, attempts to minimize this problem by selecting a prediction interval based on a range of potential generating processes. As an interesting consequence, the problem of detecting a unit root is eliminated. The Scale Factor method simply considers the unit root and its alternatives concurrently, and crafts a forecast which aims to be acceptable under both hypotheses.

2. Model and Prediction Basics

2.1 Model

Consider *N* observations of a time series $\{y_t\}_0^{N-1}$. For much of this paper, I will use a relatively simple model, assuming that the time series is generated by the equations

$$y_t = \mu + \beta t + x_t \qquad -1 \le t \le N-1 \qquad (2.1a)$$

$$x_t = \rho x_{t-1} + \varepsilon_t \qquad 0 \le t \le N-1 \qquad (2.1b)$$

Equation (2.1a) is readily recognized as the *linear regression model* for $\{y_t\}$ on *t*, with the unobserved errors $\{x_t\}$ formed in equation (2.1b). The model (2.1b) is generally referred to as an $AR(1)$ model with the *autocorrelation parameter* ρ, and the combined model (2.1) is often referred to as an *autoregressive model of order 1 with trend*.

Given $\{y_t\}_0^{N-1}$, the objective of any modeling exercise is to seek parameters $\mu, \beta \text{ and } \rho$ which are such, that the corresponding sequence $\{\varepsilon_t\}_0^{N-1}$ is "likely" (in a statistically well defined sense) to be produced by the pre-selected generator. In this paper I restrict my attention to models (2.1) where the generator behind $\{\varepsilon_t\}_0^{N-1}$ produces stationary white Gaussian noise. Furthermore, I limit my attention to models satisfying $0 \le \rho \le 1$. A slightly more complex model, accommodating $AR(p)$ processes, is discussed in Section 5.

2.2. Interval Prediction

Consider an arbitrary Forecast Method *A*, which provides a nominal confidence level $(1-\alpha)$ prediction interval for an h step ahead observation, based on an observed time series $\{y_t\}_0^{N-1}$ from (2.1). Further assume that Method *A* makes use of an estimate $\hat{\rho}_A$ of the true autocorrelation parameter ρ. The resulting prediction interval, defined by its point estimate $\hat{y}_A$ and its half-width D_A, is written

$$PI_A(\{y_t\}_0^{N-1}, \hat{\rho}_A) = \hat{y}_A \pm D_A \tag{2.2}$$

For the purpose of characterizing the performance of this method, consider the actual probability, conditioned on the true autocorrelation parameter ρ, that a future observation y_{N-1+h} is contained in the prediction interval $PI_A(\{y_t\}_0^{N-1}, \hat{\rho})$, i.e.

$$c_{N,h,\alpha}(\{y_t\}_0^{N-1}, \rho) = P(y_{N-1+h} \in PI_A(\{y_t\}_0^{N-1}, \hat{\rho}_A)|\rho) \tag{2.3}$$

The random variable $c_{N,h,\alpha}(\{y_t\}_0^{N-1}, \rho)$ is here referred to as the *instant coverage probability*.

The probability distribution of the instant coverage probability generally depends on the prediction method, as well as the actual autocorrelation parameter ρ. Its mean, here called the *average coverage probability*[2], is given by

$$C_{N,h,\alpha}(\rho) = E_y\left[c_{N,h,\alpha}(\{y_t\}_0^{N-1}, \rho)\right] \tag{2.4}$$

where the expectation is taken over the ensemble of realizations $\{y_t\}_0^{N-1}$.

In the following, for the purpose of characterizing the performance of interval forecasting methods, I consider the distribution of the error of the instant coverage probability. The focus is on its first two moments, specifically the mean error of the instant coverage probability

$$ME_{N,h,\alpha}^{cvrg}(\rho) = C_{N,h,\alpha}(\rho) - (1-\alpha) \tag{2.5}$$

and the root mean square error of the instant coverage probability

$$RMSE_{N,h,\alpha}^{cvrg}(\rho) = sqrt\left\{E_y\left[(c_{N,h,\alpha}(\{y_t\}_0^{N-1}, \rho) - (1-\alpha))^2\right]\right\} \tag{2.6}$$

respectively.

3. The Scale Factor Method

In this section, I define the *Scale Factor (SF) method* conceptually, and compare it qualitatively to several alternatives. A quantitative performance comparison is presented in subsequent sections.

[2] In other literature, "average coverage probability" is often referred to simply as "coverage probability"

3.1 Scale Factor Concept

On the path to define the Scale Factor (SF) method, let us first consider a related Median-unbiased (MU) prediction method. Figure 1 provides a block diagram for the MU method, implemented in **blocks #1-3**. Although it differs in several details, the MU method utilized here owes much of its general approach to [Andrews, 1993].

In block #1 the MU method makes use of a traditional unconditional maximum likelihood estimate $\hat{\rho}_{ML}$ for the autocorrelation parameter ρ, derived iteratively while deploying GLS to estimate trend parameters [SAS, 2011]. A lookup table in block #2 (sample in Appendix A) provides a median unbiased estimate $\hat{\rho}_{MU}$ of the autocorrelation parameter. The subsequent GLS predictor in block #3 provides an analytical solution to an *h*-step ahead prediction interval with a nominal confidence level of $1-\alpha$:

$$PI_{MU} = \hat{y}_{MU} \pm D_{MU} \tag{3.1}$$

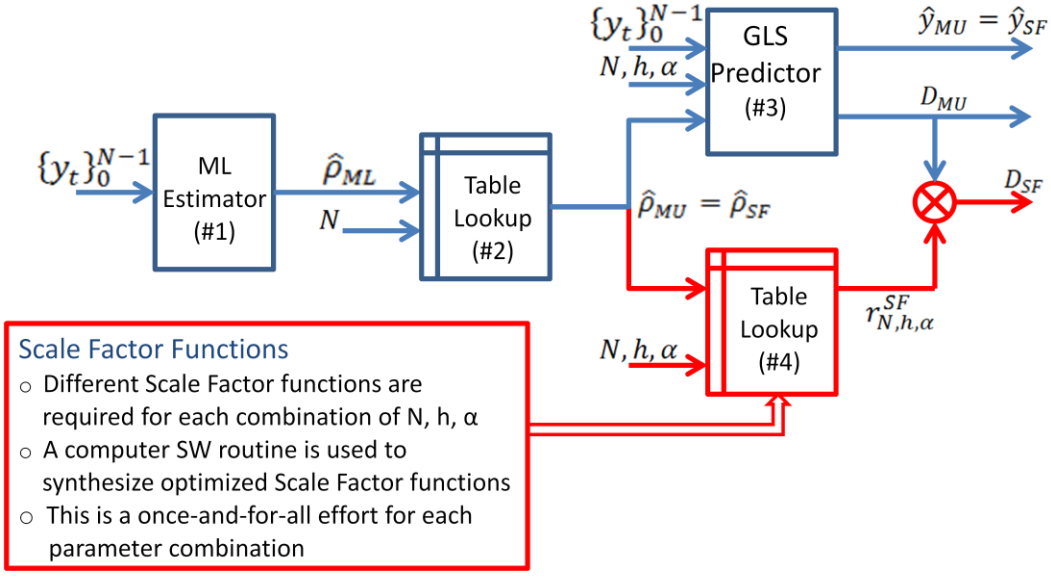

Figure 1: Block diagram for Median-Unbiased (MU) and Scale Factor (SF) methods

The SF method takes the median-unbiased autocorrelation parameter estimate $\hat{\rho}_{MU}$ as its own, i.e. $\hat{\rho}_{SF} = \hat{\rho}_{MU}$. For brevity this estimate is hereafter referenced simply as $\hat{\rho}$ unless there is a possibility of confusion. The SF method builds on the Median Unbiased (MU) forecast, and simply introduces a multiplicative *Scale Factor function* $r_{N,h,\alpha}^{SF}(\hat{\rho})$ to accommodate the uncertainty related to the use of an estimated rather than known autocorrelation parameter. The resulting prediction interval, again with a nominal confidence level of $1-\alpha$, is given by

$$PI_{SF} = \hat{y}_{SF} \pm D_{SF} \tag{3.2a}$$

where

$$\hat{y}_{SF} = \hat{y}_{MU} \tag{3.2b}$$
$$D_{SF} = r_{N,h,\alpha}^{SF}(\hat{\rho}) \cdot D_{MU} \tag{3.2c}$$

It can be shown that the GLS half-width D_{MU} in (3.1) and (3.2c) is given by

$$D_{MU} = t_{\frac{\alpha}{2},N-2}\hat{\sigma}_{\varepsilon,N}(\{y_t\}_0^{N-1},\hat{\rho})\delta_{N,h}(\hat{\rho}) \qquad ;0 \le \hat{\rho} \le 1 \qquad (3.3)$$

where $t_{\frac{\alpha}{2},N-2}$ is the upper $100\alpha/2$ percentile of the *t*-distribution with $N-2$ degrees of freedom, $\hat{\sigma}_{\varepsilon,N}(\{y_t\}_0^{N-1},\hat{\rho})$ is the GLS estimated standard error, and for $0 \le \hat{\rho} \le 1$

$$\delta_{N,h}(\hat{\rho}) = \sqrt{\sum_{i=0}^{h-1}\hat{\rho}^{2i} + \frac{(1-\hat{\rho}^h)\sum_{i=0}^{h-1}\hat{\rho}^i}{2+(N-2)(1-\hat{\rho})} + \frac{1}{N-1}\cdot\frac{\left[\frac{N-1}{2}(1-\hat{\rho}^h)+h\right]^2}{1+\frac{N-3}{2}(1-\hat{\rho})+\frac{(N-2)(N-3)}{12}(1-\hat{\rho})^2}} \qquad (3.4)$$

The SF functions $r_{N,h,\alpha}^{SF}(\hat{\rho})$ in (3.2c) and in the lookup table of **block #4** of Figure 1 are central to the proposed prediction method. They offer considerable flexibility to affect the performance of the prediction function. For a performance criterion, consider the probability distribution of the instant coverage probability $c_{N,h,\alpha}(\{y_t\}_0^{N-1},\rho)$ in (2.3). Ideally, its probability density function should be heavily concentrated around the nominal confidence level $1-\alpha$ for all values of ρ. The first two moments of this distribution, represented by $ME_{N,h,\alpha}^{cvrg}(\rho)$ and $RMSE_{N,h,\alpha}^{cvrg}(\rho)$ in (2.5) and (2.6) respectively, offer simple measures of deviation from this ideal, and their joint minimization over the interval $0 \le \rho \le 1$ represent a reasonable objective for a forecasting method.

The Scale Factor method builds on this observation, and incorporates as a key element a computer software routine to synthesize an optimized Scale Factor function with respect to an $ME_{N,h,\alpha}^{cvrg}(\rho)$ and $RMSE_{N,h,\alpha}^{cvrg}(\rho)$ based objective function over the interval $0 \le \rho \le 1$. Although the details of this procedure are beyond the scope of this paper, a sample of scale factor functions relevant for the results in this paper are documented in Appendix A. It should be clear that over time, improvements to criteria details and optimization methodology may be conceived and considered. However, the usefulness of any specific criteria and methodology is best measured by the prediction performance when compared to alternatives. Such comparisons are the primary objective of subsequent sections.

For the purposes of this paper, each Scale Factor function is defined at 16 predefined values of the autocorrelation parameter $\{\rho_j\}_{j=1}^{16} = \boldsymbol{R}$. For values of $\hat{\rho}$ outside $\boldsymbol{R}$, the Scale Factor functions are defined by piecewise linear interpolation. Scale Factor functions are independent of the observed realization, but depend on the length of the observation span *N*, the forecast horizon *h*, and the nominal confidence level $1-\alpha$. As such, the Scale Factor method requires a significant initial computational investment[3], but once Scale Factor functions have been determined for a range of these parameters, forecasts for individual realizations can be determined with a very small computational effort.

3.2 A Perspective on Scale Factor Functions

This section provides a qualitative comparison of three methods with respect to the relative width of their prediction intervals. The purpose is to offer an intuitive perspective on how the SF method relates functionally to relevant alternatives.

[3]As currently implemented with simulation of 40,000 realizations, computation of one SF function requires about 6 hrs of CPU time in a Windows environment with a 2.5GHz CPU and 8GB RAM

Prediction methods of interest here are:

- *SF - Scale Factor:* Uses the SF method described in Figure 1.
- *MU - Median Unbiased:* Uses the MU method described in Figure 1.
- *UR - Unit Root:* Takes $\hat{\rho}_{UR} = 1$, and uses GLS for estimation of trend parameters and prediction interval.

Consider the half-widths of the prediction intervals of each method, D_{SF}, D_{MU} and D_{UR}, respectively. In combination with the point estimate, it is obvious that the half-widths determine the coverage probability, and therefore strongly influence prediction performance. Figure 2 shows the ratios D_{SF}/D_{MU} ($\equiv r_{N,h,\alpha}^{SF}(\hat{\rho})$), D_{MU}/D_{MU} ($\equiv 1$) and the approximation $\sim D_{UR}/D_{MU}$ ($\equiv \delta_{N,h}(1)/\delta_{N,h}(\hat{\rho})$) as a function of $\hat{\rho}$ for the case $N = 100, h = 10, \alpha = 0.05$. The ratio "$\sim D_{UR}/D_{MU}$" is derived from (3.3):

$$D_{UR}/D_{MU} = \frac{t_{\frac{\alpha}{2},N-2}\,\hat{\sigma}_{\varepsilon,N}(\{y_t\}_0^{N-1},1)\,\delta_{N,h}(1)}{t_{\frac{\alpha}{2},N-2}\,\hat{\sigma}_{\varepsilon,N}(\{y_t\}_0^{N-1},\hat{\rho})\,\delta_{N,h}(\hat{\rho})} \approx \delta_{N,h}(1)/\delta_{N,h}(\hat{\rho})$$

where the approximation assumes that the ratio between the standard errors is close to 1.

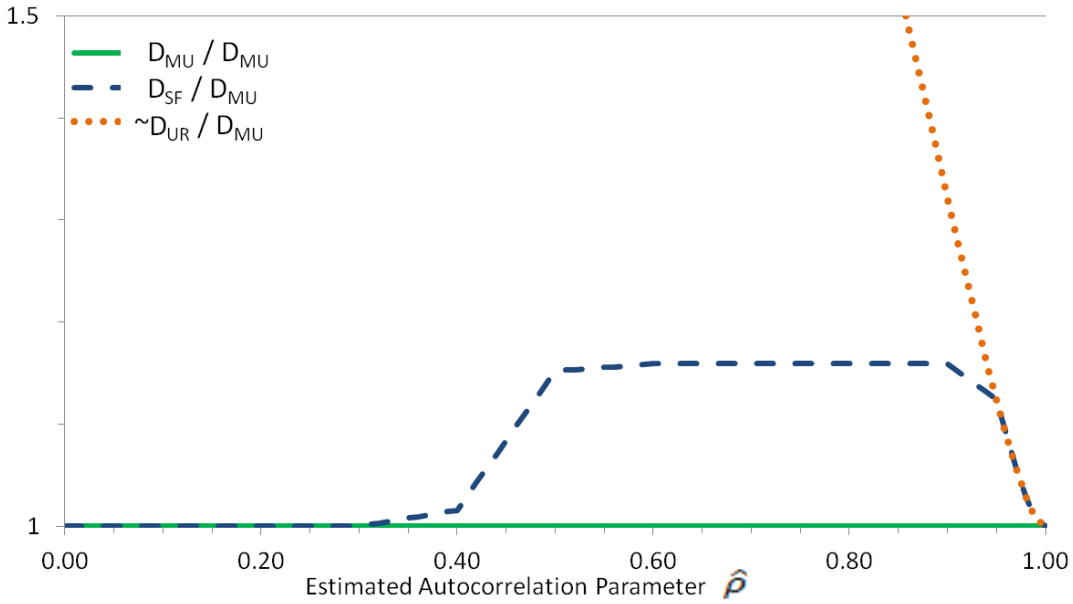

Figure 2: Relative widths of PIs for selected methods for $N = 100, h = 10, \alpha = 0.05$

Here are some observations from Fig. 2:

- For $\hat{\rho} \approx 1$, the SF method is similar to the UR method
- For $\hat{\rho} \approx 0$, the SF method is similar to the MU method
- For autocorrelation parameter estimates between these extremes, the SF method makes use of a compromise Scale Factor value

An intuitive interpretation of Fig. 2, recognizing that $\hat{\rho} \approx \rho$, is that the SF method emulates the UR and MU methods in their respective areas of strength, and seamlessly selects a compromise solution in-between.

Another observation from Fig. 2 relates to an alternative approach of first testing for a unit root, and then choose either the MU or UR method depending on the test outcome. In this case, one is effectively forced to choose either the $\sim D_{UR}/D_{MU}$ or the D_{MU}/D_{MU} alternative. For example, with $\rho = 0.85$ the probability of rejecting the unit root hypothesis is about 50%[4], and the median ratio in half-widths for the two equally likely alternatives is a considerable 1.53. Clearly, at least one of these alternatives provide a less than optimal prediction interval.

It is essential to note that in the Scale Factor method, there is no need to separately consider the case $\rho = 1$ versus $0 \leq \rho < 1$. In simplified terms, instead of forcing a binary choice between "too loose" and "too tight", the Scale factor method offers a gradual choice with the potential to be "just right". So the question of whether the observed subsequence is from a stationary or non-stationary process, which typically cannot be answered with certainty, does not need to be asked.

4. Simulated Forecast Performance

In this section, the performance of the Scale Factor method (SF) is studied through simulation for an observation span $N = 100$, a forecast horizon $h = 10$ and with $\alpha = 0.05$, representing a nominal 95% confidence level. In addition to the previously discussed MU and UR schemes, the performance of the Scale Factor method is compared to two additional forecast techniques:

- *ML - Maximum Likelihood:* Uses a traditional unconditional maximum likelihood estimator for the autocorrelation parameter $\hat{\rho}_{ML}$ (as in Figure 1, block #1), iteratively using the GLS approach to estimate the trend parameters [SAS, 2011].
- *NO - kNOwn Parameter:* Uses GLS directly on the kNOwn autocorrelation parameter to estimate the trend parameters, offering an idealized benchmark for other methods.

For the alternative methods, the prediction intervals are determined by the conventional approach, i.e. by assuming that the estimated $\hat{\rho}_{MTHD}$ equals the true autocorrelation parameter, and using the GLS method to determine the corresponding prediction interval.

Figures 3 and 4 show, for each of the methods of interest, the estimated first two moments of the instant coverage probability error as defined in (2.5) and (2.6), as a function of the actual autocorrelation parameter ρ. The performance is estimated based on 40,000 sample realizations for each value of the autocorrelation parameter in the set $\boldsymbol{R}$ discussed in section 3.1. The same set of sample realizations is used for all methods.

Clearly, the UR method is in a category by itself, for obvious reasons performing magnificently for values of ρ close to 1, and very poorly otherwise. Furthermore, the NO method shows the anticipated zero mean error in Fig. 3, but a perhaps surprisingly large

[4] Using the Dickey-Fuller test at the 5% level

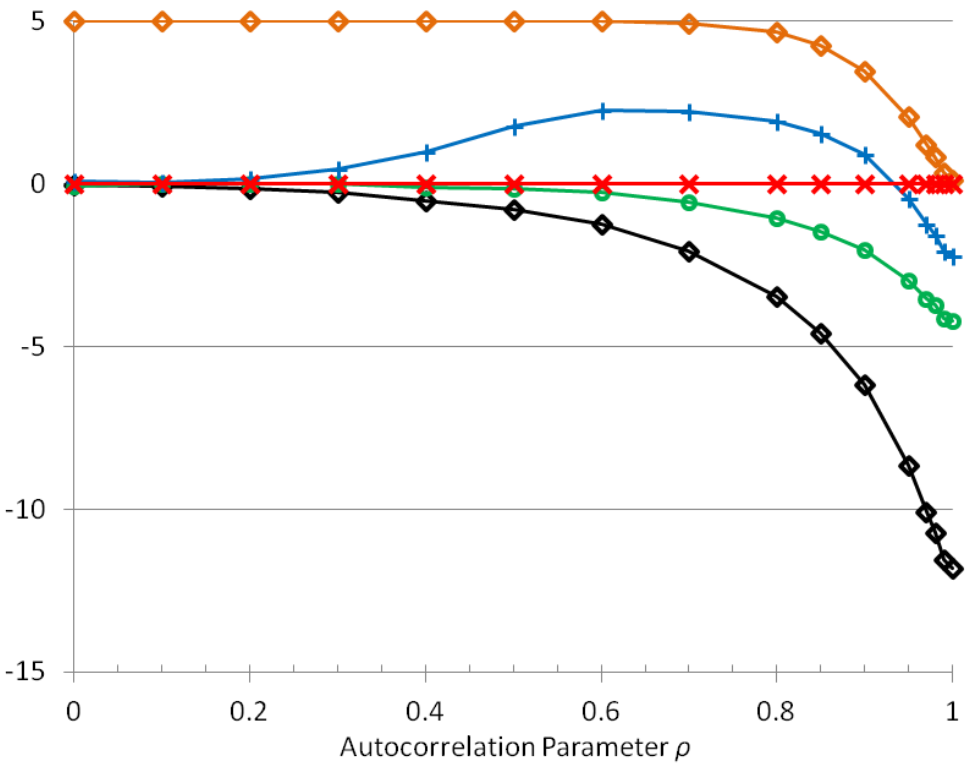

Figure 3: Estimated mean error of coverage probability $ME^{cvrg}_{100,10,0.05}(\rho)$ [%]

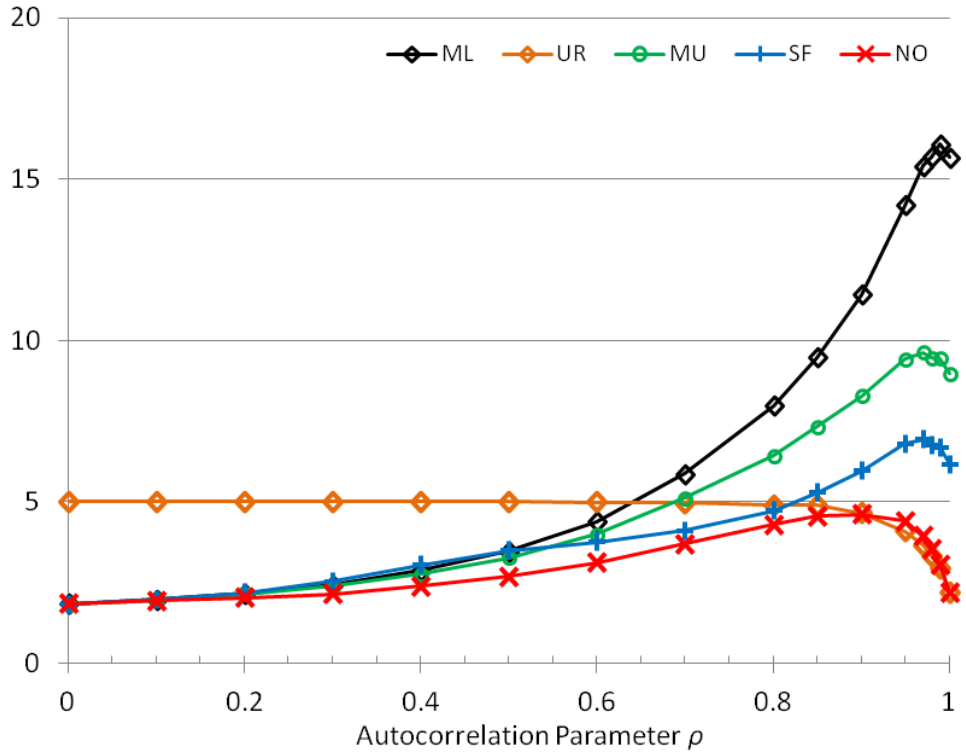

Figure 4: Estimated root mean square error of coverage probability $RMSE^{cvrg}_{100,10,0.05}(\rho)$ [%] for selected forecast methods

4.6% RMSE for values of ρ around 0.9, and even for the benign random walk at $\rho = 0$ it shows a 1.8% RMSE. This serves as a reminder that even when the autocorrelation parameter is known, estimation of the unknown trend parameters still provides significant uncertainty in individual forecasts.

In the forecast environment, the true value of the autocorrelation parameter is often unknown, so in typical applications it makes sense to rank methods based on worst case performance. It is interesting to note that for each of the ML, MU and SF methods, and for both performance measures, the worst case performance is offered for values of ρ quite close to 1. The worst case performance for each method and for both measures is summarized in Table 1. Clearly, the performance improvement offered by the SF method in this case is substantial. It is also worth noting, that with respect to worst case RMSE, the SF method comes remarkably close to the idealized performance of the NO method.

Table 1: Worst Case Performance Measures for Selected Forecast Methods [%}

Measure	UR	ML	MU	SF	NO
Max |ME|	5.00	11.81	4.22	2.24	0.00
Max RMSE	5.00	16.08	9.62	6.93	4.61

5. Application Example: An Extension of the Nelson-Plosser Data Set

In this section, I apply the SF method to the Nelson-Plosser data set. The purpose is simply to seek answers to the following questions:

- Is it possible to make useful forecasts for these series using the SF method?
- Does the SF method offer meaningful performance improvement compared to the ML, MU and UR methods?

5.1 Data

The original Nelson Plosser Data Set [Nelson and Plosser, 1982] contains total of 14 macroeconomic time series with annual observations ending in 1970 and beginnings between 1860 and 1909. All 14 series were extended by [Schotman and van Dijk, 1991]

Table 2: Time Series List, with Selected Data from Original N&P Study

Series	Acronym	Fst Yr	Lst Yr	p	$\hat{t}(\hat{\rho}_{NP})$	$\hat{\rho}_{NP}$	$\hat{\rho}_{ML}$	$\hat{\rho}_{SF}$	Ref
Bond Yield/Interest Rate	BNDYLD	1900	1988	3	0.686	1.030	0.948	1.000	
Consumer Prices	LNCPIE	1860	2010	4	-1.97	0.986	0.980	1.000	[BLS1]
Employment	LNEMP	1890	1988	3	-2.66	0.861	0.858	0.921	
GNP Deflator	LNGNPDFE	1889	2010	2	-2.52	0.915	0.952	0.993	[BEA]
Industrial Production	LNINDPE	1860	2010	6	-2.53	0.835	0.871	0.979	[FED1]
Money Stock	LNMSTK	1889	1988	2	-3.08	0.916	0.913	0.943	
Money Stock Velocity	LNMVEL	1869	1988	1	-1.66	0.941	0.950	1.000	
Wages	LNNPAY	1900	1988	3	-2.09	0.910	0.909	0.977	
Real per Capita GNP	LNRGNPC	1909	1988	2	-3.04	0.818	0.824	0.894	
Real GNP	LNRGNPE	1909	2010	2	-2.99	0.825	0.829	0.888	[FED3]
Real Wages	LNRPAY	1900	1988	2	-3.04	0.831	0.853	0.934	
Common Stock Prices	LNSP500E	1871	2010	3	-2.05	0.913	0.929	0.991	[Shiller]
Unemployment Rate	LNUEMPRE	1890	2010	4	-3.55	0.706	0.682	0.740	[BLS2]

to end in 1988. A subset of 7 time series was extended for this study to end in 2010[5]. Data are spliced to the earlier extended series. The conversion factor is the average ratio of the old series over the new series for the period 1983-1988. The resulting observation period for each series is presented in Table 2, and the data source for each series extended here is given in Table 3.

The order p of the model that [Nelson and Plosser, 1982] assigned to each original time series is shown in Table 2, together with their estimated values $\hat{\rho}_{NP}$ for the parameter ρ, and $\hat{\tau}(\hat{\rho}_{NP})$ which is the "t-statistic" of the ratio of $\hat{\rho}_{NP} - 1$ to its standard error in the Dickey-Fuller Test[6]. Also included are the average estimated autocorrelation parameter values for the ML and SF methods, based on 18 or 40 forecasts for each series, depending on its last year of observation.

Table 3: Data Sources For Series Extended Here

Ref	Title	URL
[BEA]	Table 1.1.9. Implicit Price Deflators for GDP	http://www.bea.gov/national/nipaweb2011
[BLS1]	Consumer Price Index; (CPI-U)	ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt
[BLS2]	Employment status - civilian noninst. pop.	http://bls.gov/opub/ee/2011/cps/annavg1_2010.pdf
[FED1]	Industrial Production Index (INDPRO)	http://research.stlouisfed.org/fred2/series/INDPRO
[FED2]	Gross National Product (GNPA)	http://research.stlouisfed.org/fred2/series/GNPA
[FED3]	Real Gross National Product (GNPCA)	http://research.stlouisfed.org/fred2/series/GNPCA
[Shiller]	Shiller, R., U.S.Stock Price Data, Annual	www.econ.yale.edu/~shiller/data/chapt26.xls

5.2 Methodology

[Nelson and Plosser, 1982] make use of $AR(p)$ models with $p \geq 1$. The time series are considered generated by the equations

$$y_t = \mu + \beta t + x_t \qquad -p \leq t \leq N-1 \qquad (5.1a)$$

$$x_t = \rho x_{t-1} + \sum_{i=1}^{p-1} \psi_i (x_{t-i} - x_{t-i-1}) + \varepsilon_t \qquad 0 \leq t \leq N-1 \qquad (5.1b)$$

which are easily identified as generalizations of (2.1). However, it should be noted that in this case the parameter ρ is not generally equal to the one-step autocorrelation of the time series.

Parameter estimation for the $AR(p)$ model is different from that for the $AR(1)$ model discussed previously. For the MU method, approximately median unbiased parameters are estimated using the simple iterative procedure suggested by [Andrews and Chen, 1994], but with ML estimates of the AR parameters, and GLS estimates of the trend parameters and prediction interval[7]. In each iteration, estimates are based on 40,000 realizations.

The SF method uses the parameter estimates of the MU method, as in the $AR(1)$ case, and the corresponding prediction interval is derived from (3.2), using the Scale Factor function derived for the $AR(1)$ model. The usefulness of this simple approach should be

[5] Extension of remaining series was abandoned due to lack of easily accessible and relevant data
[6] The DF test rejects a unit root hypothesis at the 5% level *only* for the Unemployment Rate series
[7] In place of the OLS estimates used by Andrews and Chen

verified, e.g. through simulation, and such verification has not yet been done. However, several observations suggest that this may indeed be a reasonable approach. First, the outcome of the Augmented Dickey-Fuller Test, which is well established, is independent of estimates for the parameters ψ_i in (5.1b). Second, when the parameters ψ_i are "small", a continuity argument would suggest that the $AR(1)$ data is relevant. Third, the MU method appears to provide interval forecasts which are too optimistic, so using a Scale Factor which is ≥ 1 at least makes the forecast less optimistic.

For the UR method, the parameter estimates are based on the differenced series, with unconditional ML estimates of the AR parameters, and GLS estimates of the trend parameters. The ML method uses the same approach, but on the original series. Both methods use the conventional GLS approach to determine the prediction interval.

Forecasts for a horizon $h = 10$ are produced for each year after 1970 with an available observation, resulting in either 18 or 40 forecasts for each time series, and a total of 406 forecasts for each method. In each case, all available data is used to produce the forecast, i.e. a 10 year ahead forecast for Industrial Production in 1985 is based on all 116 observations from the period 1860-1975. For each method, the forecast is based on an $AR(p)$ process of the order p used in [Nelson and Plosser, 1982].

5.3 Results

Figure 5 presents a scatter plot of all 406 actual observations in relation to the 10 year SF forecast for all 14 time series[8]. In this plot, the vertical axis is scaled such that the endpoints of each 95% prediction interval falls at ± 1. The horizontal axis shows the

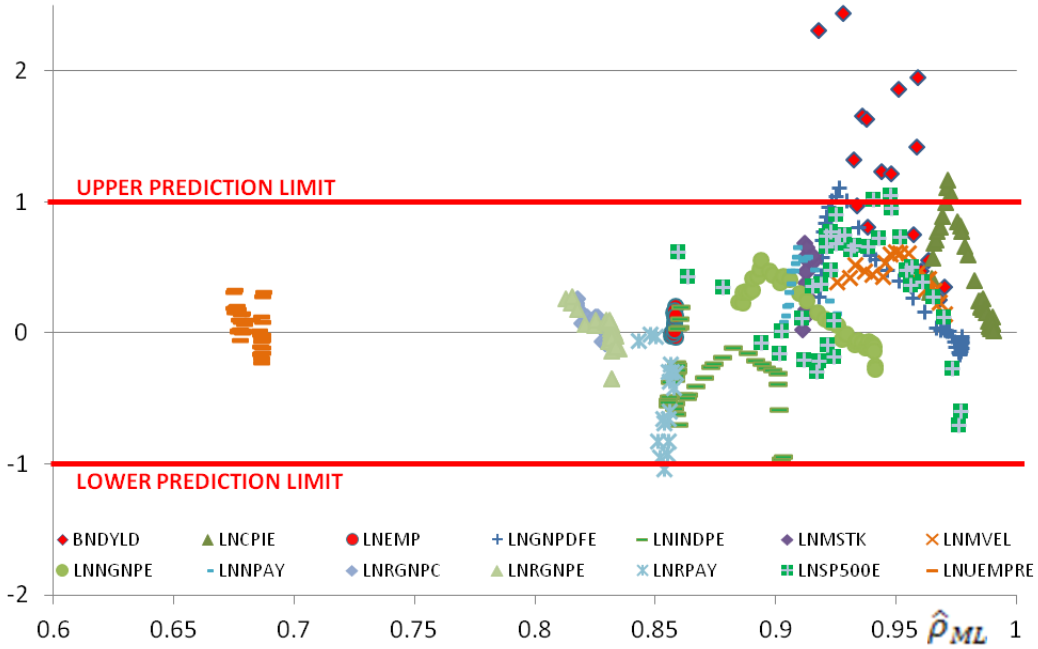

Figure 5: Scatter Plot of Actual Observations in SF Prediction Interval vs. $\hat{\rho}_{ML}$

[8] The acronym used for each time series is listed in Table 1

value of $\hat{\rho}_{ML}$ of the corresponding ML forecast[9]. The clustering of observations clearly indicates the difficulty of drawing strong conclusions from this limited set of time series, despite the collection of a total of 1,565 observations in total for the 14 different series. However, the graph appears to be consistent with the simulated results in Figures 3 and 4, indicating a lower and more variable coverage probability for processes with roots near one, and a higher and more predictable coverage probability with roots further from unity. It is interesting to note that the overall observed coverage rate for the SF method in this case is an "almost too good to be true" 95.1%, but obviously with large variations between the different time series.

Table 4 shows the observed coverage rate error in total and for each series individually, for each of the four methods. The overall results for the ML method are in agreement with the common experience that the confidence intervals provided by that method are too optimistic, while the results for the UR, MU and SF methods appear quite consistent with expectations. For each series, the results for these three methods are also very similar, while the ML method exhibits considerably larger coverage rate errors for several series. The Bond Yield series stands out from the others, with respect to both the extreme position of its data cluster in Fig. 5, and its severe coverage rate error in excess of 50% for all four methods.

Table 4: Observed Coverage Rate Error and Width Ratio RW for Various Methods

		Coverage Rate Error %				Width Ratio RW		
Series	$\hat{\rho}_{SF}$	UR	ML	MU	SF	UR/SF	ML/SF	MU/SF
BNDYLD	1.00	-50.6	-67.2	-50.6	-50.6	1.03	0.94	1.00
LNCPIE	1.00	-5.0	-27.5	-2.5	-2.5	0.99	0.93	1.00
LNMVEL	1.00	5.0	5.0	5.0	5.0	1.00	0.89	1.00
LNGNPDFE	0.99	2.5	-30.0	-5.0	-5.0	0.99	0.83	0.99
LNSP500E	0.99	0.0	-35.0	0.0	0.0	0.99	0.84	0.99
LNNGNPE	0.99	5.0	5.0	5.0	5.0	1.00	0.72	0.98
LNINDPE	0.98	5.0	0.0	5.0	5.0	0.96	0.81	0.96
LNNPAY	0.98	5.0	-6.1	5.0	5.0	0.99	0.72	0.97
LNMSTK	0.94	5.0	5.0	5.0	5.0	1.23	0.74	0.89
LNRPAY	0.93	5.0	-39.4	-11.7	-0.6	1.04	0.66	0.88
LNEMP	0.92	5.0	5.0	5.0	5.0	1.07	0.69	0.86
LNRGNPC	0.89	5.0	5.0	5.0	5.0	1.16	0.62	0.80
LNRGNPE	0.89	5.0	5.0	5.0	5.0	1.25	0.66	0.82
LNUEMPRE	0.74	5.0	5.0	5.0	5.0	1.51	0.78	0.85
MEAN(406)	0.94	0.8	-11.7	-0.4	0.1	1.09	0.78	0.93

Table 4 also considers the ratio RW of the widths of the prediction intervals. For each forecast, the ratio of the prediction interval width of a method and that of the SF method is calculated, and the average of the ratios for all forecasts on each series is presented in the table. A small ratio indicates a tighter prediction interval, which is clearly desirable, provided that the coverage is on target. In Table 4, the series are sorted in order of declining $\hat{\rho}_{SF}$. It is interesting to note, that for the 8 series with $\hat{\rho}_{SF} \geq 0.95$, the ratio RW in the UR/SF and MU/SF case is very close to 1, while RW for the ML/SF case are much smaller, as low as 0.72 for the LNNPAY and LNNGNPE series. The RW similarities and differences in this subsample are reflected in the coverage rate performance of the different methods. For the remaining 6 series with $\hat{\rho}_{SF} \leq 0.95$, there is a trend towards

[9] Due to compression at $\hat{\rho}_{SF} = 1$, I choose the biased $\hat{\rho}_{ML}$ here, for clearer presentation of the data

increased divergence in the ratio RW when $\hat{\rho}_{SF}$ becomes smaller, as would be expected. However, for all but one of the series in this subsample, this has no effect on the coverage rate, perhaps a reflection of the statistical variations that may trouble a small sample.

Returning to the questions asked at the beginning of this section, I believe that the following answers should be drawn from the Nelson-Plosser data set:

- *Is it possible to make useful forecasts for these series using the SF method?* Yes, with an overall coverage rate performance within 0.1% of target for this sample, it is difficult to say no. The results for the Nelson-Plosser data conforms quite well with expectations from simulations. For 13 of the 14 series there is no reason to be concerned about the usefulness of the forecasts, indicating that both the $AR(p)$ models and the Scale Factor method may operate as anticipated. A single series, Bond Yield, failed to produce useful forecasts, but the SF method performed no worse than the alternatives, including the UR method which is generally considered to be quite conservative. The problem may simply be that the historic bond yield record from 1900-1970 contains nothing similar to the interest rate bubble we experienced in the 1970's and 1980's. If so, there may not be any model, AR(p) or otherwise, which could reasonably predict that bubble based on that record.

- *Does the SF method offer meaningful performance improvement compared to the ML, MU and UR methods?* The SF method performs at least as well as any of the alternatives for this sample. However, with the high level of dependency both within and between series, it is clearly difficult to establish statistically significant tests from this data set. Noticeably, there is nothing in the analysis here to suggest that the performance improvement hypothesis should be rejected. More importantly, the results from the Nelson Plosser data set are consistent with expectations from the simulation study, and the simulation study is clearly supporting the performance improvement hypothesis.

6. Conclusion

Is the autoregressive time series stationary or not? In traditional forecasting the answer to that question is important, since the forecast generally will be different in the two cases. The Scale Factor method circumvents the question of stationarity by considering both alternatives concurrently, and crafting a forecast which aims to be acceptable under both hypotheses. The SF method offers a new approach which promises notably improved performance under favorable conditions, and little or no performance degradation under unfavorable ones.

The potential disadvantage of the SF method is rather in its complexity. A decision on the usefulness of the SF method in any particular application may generally come down to a performance vs. cost/complexity trade-off. In case of macroeconomic time series, increasing the sample size and forecast quality by going back further in time may not be feasible. Improving forecast quality by using smaller observation intervals in more recent years has limited benefits, if the time series has time constants of the order of a decade or more. When the application is such that a better forecast is valuable, the Scale Factor method may well be worth considering.

Appendix A: Sample Lookup Tables

Table A.1: Lookup Table for $\hat{\rho}_{MU} = \hat{\rho}_{SF}$. Find $\hat{\rho}_{ML}$ in body of table. Use linear interpolation to determine $0 \leq \hat{\rho}_{SF} \leq 1$

	$\widehat{\boldsymbol{\rho}}_{\boldsymbol{MU}} = \widehat{\boldsymbol{\rho}}_{\boldsymbol{SF}}$															
N	**0.00**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.85**	**0.90**	**0.95**	**0.97**	**0.98**	**0.99**	**1.00**
75	-0.028	0.067	0.164	0.261	0.355	0.453	0.547	0.643	0.736	0.782	0.826	0.867	0.880	0.885	0.889	0.891
80	-0.027	0.069	0.167	0.262	0.360	0.455	0.551	0.646	0.740	0.787	0.832	0.873	0.886	0.892	0.895	0.897
85	-0.025	0.071	0.168	0.265	0.363	0.457	0.554	0.649	0.744	0.791	0.837	0.877	0.892	0.898	0.902	0.903
90	-0.024	0.073	0.171	0.266	0.364	0.460	0.557	0.652	0.748	0.794	0.840	0.882	0.896	0.902	0.907	0.909
95	-0.023	0.074	0.172	0.269	0.367	0.463	0.558	0.655	0.750	0.798	0.843	0.886	0.901	0.907	0.911	0.913
100	-0.020	0.077	0.174	0.271	0.368	0.464	0.561	0.657	0.752	0.800	0.847	0.890	0.904	0.911	0.916	0.917
105	-0.021	0.077	0.175	0.273	0.369	0.466	0.563	0.660	0.756	0.803	0.849	0.892	0.908	0.915	0.920	0.921
110	-0.020	0.078	0.176	0.274	0.372	0.468	0.565	0.662	0.758	0.805	0.852	0.896	0.911	0.918	0.923	0.925
115	-0.019	0.079	0.177	0.275	0.372	0.469	0.566	0.663	0.759	0.807	0.854	0.899	0.914	0.921	0.926	0.928
120	-0.017	0.080	0.178	0.276	0.374	0.470	0.568	0.665	0.761	0.809	0.856	0.901	0.917	0.924	0.929	0.931
125	-0.017	0.081	0.179	0.277	0.374	0.472	0.569	0.666	0.763	0.811	0.857	0.903	0.920	0.926	0.931	0.934

Table A.2: Lookup Table for $r_{N,h,\alpha}^{SF}(\hat{\rho}_{SF})$, in body of table. Use linear interpolation to determine $r_{N,h,\alpha}^{SF}(\hat{\rho}_{SF})$ for $\hat{\rho}_{SF} \notin \boldsymbol{R}$

			$\widehat{\boldsymbol{\rho}}_{\boldsymbol{SF}}$														
N	**h**	$\boldsymbol{\alpha}$	**0.00**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.85**	**0.90**	**0.95**	**0.97**	**0.98**	**0.99**	**1.00**
75	**10**	**0.05**	1.000	1.000	1.000	1.000	1.007	1.220	1.220	1.220	1.220	1.220	1.220	1.106	1.046	1.022	1.005	1.001
80	**10**	**0.05**	1.000	1.000	1.000	1.000	1.002	1.205	1.205	1.205	1.205	1.205	1.205	1.110	1.048	1.023	1.006	1.001
85	**10**	**0.05**	1.000	1.000	1.000	1.000	1.005	1.192	1.192	1.192	1.192	1.192	1.192	1.114	1.050	1.025	1.006	1.000
90	**10**	**0.05**	1.000	1.000	1.000	1.000	1.008	1.179	1.179	1.179	1.179	1.179	1.179	1.118	1.053	1.026	1.006	1.000
95	**10**	**0.05**	1.000	1.000	1.000	1.000	1.000	1.071	1.175	1.175	1.175	1.175	1.175	1.121	1.055	1.027	1.007	1.000
100	**10**	**0.05**	1.000	1.000	1.000	1.000	1.016	1.152	1.159	1.159	1.159	1.159	1.159	1.125	1.057	1.028	1.007	1.000
105	**10**	**0.05**	1.000	1.000	1.000	1.001	1.019	1.127	1.150	1.150	1.150	1.150	1.150	1.128	1.058	1.029	1.008	1.000
110	**10**	**0.05**	1.000	1.000	1.003	1.013	1.042	1.111	1.144	1.144	1.144	1.144	1.144	1.131	1.060	1.031	1.008	1.000
115	**10**	**0.05**	1.000	1.000	1.003	1.013	1.042	1.103	1.135	1.135	1.135	1.135	1.135	1.133	1.062	1.032	1.009	1.000
120	**10**	**0.05**	1.000	1.000	1.003	1.014	1.045	1.107	1.130	1.130	1.130	1.130	1.130	1.130	1.064	1.033	1.009	1.000
125	**10**	**0.05**	1.000	1.000	1.000	1.002	1.015	1.088	1.124	1.124	1.124	1.124	1.124	1.124	1.065	1.034	1.009	1.000

References

Andrews, D. W. K [1993]: “Exactly Median-Unbiased Estimation of First Order Autoregressive / Unit Root Models”; Econometrica Vol. 61, No. 1, pp. 139-165

Andrews, D.W.K and H.-Y. Chen [1994]: "Approximately Median-Unbiased Estimation of Autoregressive Models"; Journal of Business & Economic Statistics, Vol. 12, No. 2, pp. 187-204

Dickey, D.A. and W.A. Fuller [1979]: “Distribution of the Estimators for Autoregressive Time Series with a Unit Root,” Journal of the American Statistical Association, Vol 74, p. 427–431.

Falk, B. and A. Roy [2005]: "Forecasting Using the Trend Model with Autoregressive Errors"; International Journal on Forecasting, Vol. 21, pp. 291-302

Falk, B. and A. Roy [2006]: “Efficiency Tradeoffs in Estimating the Linear Trend Plus Noise Model”; Economics Bulletin, Vol. 3, No. 6, pp. 1-9

Franses, P.H. and F. Kleibergen [1996]: "Unit Roots in the Nelson-Plosser Data: Do They Matter for Forecasting?"; International Journal of Forecasting, Vol. 12, pp. 283-288

Granger, C. W. J. and P. Newbold [1986]: “Forecasting Economic Time Series”, second edition; Academic Press

Harvey, A.C. [1981]: “The Econometric Analysis of Time Series”; John Wiley & Sons, Inc., Chapter 6

Nelson, C. R. and Plosser, C. I. [1982]: “Trends and Random Walks in Macroeconomic Time Series”; Journal of Monetary Economics, vol. 10, pp. 139-162

Roy, A. and W. A. Fuller [2001]: "Estimation of Autoregressive Time Series with a Root Near 1"; Journal of Business & Economic Statistics, Vol. 19, No. 4, pp. 482-493

Roy, A., B. Falk and W. A. Fuller [2004]: “Testing for Trend in the Presence of Autoregressive Errors”; Journal of the American Statistical Association, Vol. 99, No. 468, pp. 1082-1090

Rudebusch, G. D. [1992]: "Trends and Random Walks in Macroeconomic Time Series: A Re-examination"; International Economic Review, Vol. 33, No. 3, pp. 661-680

Schotman, P.C. and H.K. van Dijk [1991]:"On Bayesian Routes to Unit Roots"; Journal of Applied Econometrics, Vol. 6, pp. 387-401

SAS Institute [2011]: "SAS/ETS 9.3 User's Guide/Procedure Reference/The AUTOREG precedure", SAS Institute, ISBN: 978-1-60764-911-3