EUROPEAN JOURNAL OF MATHEMATICAL SCIENCES
Vol. 2, No. 2, 2013, 235-245
ISSN 2147-5512 – www.ejmathsci.com

Parameter Optimization for Combining Lognormal Antithetic Time Series

D. Ridley[1], P. Ngnepieba [2,*], D. Duke[3]

[1] *SBI, Florida A&M University and Department of Scientific Computing, Florida State University, Tallahassee, FL., USA.*
[2] *Department of Mathematics, Florida A&M University, Tallahassee, FL, 32307, USA.*
[3] *Department of Physics, Florida State University, Tallahassee, FL., USA*

Abstract. Time series models of serially correlated variables yield biased fitted values and forecasts. Antithetic time series can be created so as to be negatively correlated with the original time series. The original and antithetic time series can be combined so as to eliminate said bias. A system of equations that define the optimal combining parameters is derived.

2010 Mathematics Subject Classifications: 62E20, 62F10

Key Words and Phrases: Antithetic time series theory, antithetic random variables, bias reduction, inverse correlation, serial correlation, sampling bias.

---

1. Introduction

Time series models have wide spread applications in science, engineering and economics. A mathematical model of a time series variable may be fitted to historical data from the time series for the purpose of gaining a better understanding of the underlying process that generates the data and for extrapolating future values. If the time series variable is random then its probability distribution must be known or assumed. Even if the distribution is known and the perfect model is selected, the data are almost always a limited sample. That is, the population is truncated and distorted. Therefore, a sampling bias occurs. The absence of any relevant information from a model will express itself in the patterns of the error term. If complete avoidance of bias requires normally distributed data, then the absence of normality is like missing information. The errors can become serially correlated [Griliches 3]. The parameters of the model are biased and fitted values from the model are biased. That is, in addition to random errors, the fitted values contain a component of error that is systematically

*Corresponding author.

Email addresses: dridley@fsu.edu (D. Ridley), Pierre.Ngnepieba@famu.edu (P. Ngnepieba), dduke@fsu.edu (D. Duke)

biased. Likewise, future values extrapolated from the model are biased. The bias grows with the forecast horizon. This diminishes the usefulness of the forecast when applied to long range economic projections.

Antithetic time series modeling is a method for creating new fitted values that are inversely correlated with the original fitted values (Ridley [4], Ridley and Ngnepieba [5]). Reversal of correlation is accomplished by raising the fitted values to an infinitesimally small negative exponent. The original and antithetic fitted values are then combined so as to eliminate the biased component of error, leaving only purely random error. In theory, the mean square error (MSE) of the unbiased combined forecast is constant throughout the forecast horizon. The purpose of this paper is to investigate the sensitivity of the MSE to the combining parameters. The remainder of the paper is organized as follows. The antithetic time series model is defined in section 2. The population statistics that describe the original and antithetic time series are defined in terms of combining parameters in Section 3. The corresponding sample statistics are given in section 4. A system of equations that define the optimal combining parameters is derived in section 5. An empirical example that illustrates how the combining parameters affect MSE is given in section 6. Some suggestions for future research are given in section 7.

2. The Antithetic Time Series Model

Consider the lognormal time series $X_t, -\infty \le t \le \infty, \ln X_t \sim N\left(\mu, \sigma^2\right)$, from which a sample $x_t, t = 1, 2, 3, \ldots, n$, is taken. The reversal of correlation between X_t and X_t^p as $p \to 0^-, \sigma \to 0$ was proved by the Ridley [4] antithetic time series theorem for the lognormal distribution. Now, suppose that $x_t = f\left(x_{t-1}\right) + \epsilon_t, t = 2, 3, \ldots, n$, is a time series model which is biased due to serial correlation or due to data sampling bias, such that the covariance $\operatorname{Cov}\left(\epsilon_t, x_{t-1}\right) \neq 0$. Let the fitted values be $\widehat{x}_t$. Denoting the standard deviation in x_t by s_x, and the sample correlation coefficient by r, the combined antithetic fitted model is given (see appendix A) by

$$\widehat{x}_{c,t} = \omega \widehat{x}_t + (1-\omega)\left\{\overline{x} + r_{\widehat{x}\widehat{x}^p}\left(s_{\widehat{x}}/s_{\widehat{x}^p}\right)\left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)\right\},\ p \to 0^-.$$

When applying this model to actual data, typical of what can be expected, the model is restated as

$$\widehat{x}_{c,t} = \omega \widehat{x}_t + (1-\omega)\left\{\overline{x} + \left(1 - k\sqrt{n+1-t}\right) r_{\widehat{z}\widehat{z}^p}\left(s_{\widehat{z}}/s_{\widehat{z}^p}\right)\left(\widehat{z}_t^p - \overline{\widehat{z}^p}\right)\right\}, \tag{1}$$

where $p = -0.001$ and $z_t = x_t + \lambda$, where λ is used to facilitate the power transformation, and k is an empirical factor to correct for heteroscedasticity in the data. The values of ω, λ and k are chosen to minimize the fitted MSE. Whereas $\widehat{x}_t$ is biased, $\widehat{x}_{c,t}$ is unbiased [see 4]. In Ridley [4], ω, λ were found by grid search[†]. Also, only stationary time series were considered so the parameter k was not included. In this paper we derive the system of analytical equations that specify the combining parameters exactly. This allows ω, λ and k to be determined iteratively, and therefore more rapidly than the previous grid search. With the help of graphs obtained

[†]Using FOURCAST application, `http://www.fourcast.net`

from these equations we are better able to show and interpret the effects of ω, λ and k on the MSE.

3. Population Statistics

Given X_t lognormally distributed such that $lnX_t \sim N\left(\mu, \sigma^2\right)$ then, $\mathbb{E}\left[X_t\right] = \exp\left(\mu + \sigma^2/2\right)$ and $\mathbb{E}\left[X_t^p\right] = \exp\left(p\mu + p^2\sigma^2/2\right)$. Next, consider $Z_t = X_t + \lambda$. Then, $\ln Z_t \sim N\left(\mu + \lambda, \sigma^2\right)$, and $\mathbb{E}\left[Z_t^p\right] = \mathbb{E}\left[\left(X_t + \lambda\right)^p\right] = \exp\left(p\left(\mu + \lambda\right) + p^2\sigma^2/2\right)$. The correlation between Z_t and Z_t^p, ρ_{zz^p} is given by

$$\rho_{zz^p} = \frac{\mathbb{E}\left[Z_t^{p+1}\right] - \mathbb{E}\left[Z_t^p\right]\mathbb{E}\left[Z_t\right]}{\sigma_z \sigma_{z^p}}, \tag{2}$$

where σ_z and σ_{z^p} are the standard deviations of Z_t and Z_t^p, respectively. We then have

$$\begin{aligned}
\mathbb{E}\left[Z_t^{p+1}\right] - \mathbb{E}\left[Z_t^p\right]\mathbb{E}\left[Z_t\right] = & \exp\left\{(p+1)\left(\mu+\lambda\right) + (p+1)^2\sigma^2/2\right\} \\
& - \exp\left\{p\left(\mu+\lambda\right) + p^2\sigma^2/2\right\}\exp\left(\mu + \lambda + \sigma^2/2\right) \\
& = \exp\left\{(p+1)\left(\mu+\lambda\right) + (p+1)^2\sigma^2/2\right\} \\
& - \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2\right\} \\
& = \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+2p+1\right)\sigma^2/2\right\} \\
& - \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2\right\} \\
& = \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2 + p\sigma^2\right\} \\
& - \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2\right\} \\
& = \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2\right\}\exp\left(p\sigma^2\right) \\
& - \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2\right\} \\
& = \exp\left\{(p+1)\left(\mu+\lambda\right) + \left(p^2+1\right)\sigma^2/2\right\}\left(\exp\left(p\sigma^2\right) - 1\right).
\end{aligned} \tag{3}$$

Also,

$$\begin{aligned}
\sigma_{z^p}^2 &= \mathbb{E}\left[Z^{2p}\right] - \left(\mathbb{E}\left[Z^p\right]\right)^2 \\
&= \exp\left\{2p\left(\mu+\lambda\right) + (2p)^2\sigma^2/2\right\} - \left[\exp\left\{p\left(\mu+\lambda\right) + p^2\sigma^2/2\right\}\right]^2 \\
&= \exp\left\{2p\left(\mu+\lambda\right) + 2p^2\sigma^2\right\} - \exp\left\{2p\left(\mu+\lambda\right) + p^2\sigma^2\right\} \\
&= \exp\left\{2p\left(\mu+\lambda\right) + p^2\sigma^2\right\}\exp\left\{p^2\sigma^2\right\} - \exp\left\{2p\left(\mu+\lambda\right) + p^2\sigma^2\right\} \\
&= \exp\left\{2p\left(\mu+\lambda\right) + p^2\sigma^2\right\}\left(\exp\left(p^2\sigma^2\right) - 1\right),
\end{aligned}$$

$$\sigma_{z^p} = \exp\left\{p\left(\mu+\lambda\right) + p^2\sigma^2/2\right\}\sqrt{\exp\left(p^2\sigma^2\right) - 1}, \tag{4}$$

and

$$\sigma_z = \exp\left\{(\mu+\lambda)+\sigma^2/2\right\}\sqrt{\exp\left(\sigma^2\right)-1}. \tag{5}$$

From equations (2), (3), (4) and (5)

$$\rho_{zz^p} = \frac{\exp\left\{(p+1)(\mu+\lambda)+(p^2+1)\sigma^2/2\right\}\left\{\exp\left(p\sigma^2\right)-1\right\}}{\exp\left\{(\mu+\lambda)+\sigma^2/2\right\}\sqrt{\exp\left(\sigma^2\right)-1}\cdot\exp\left\{p\left(\mu+\lambda\right)+p^2\sigma^2/2\right\}\sqrt{\exp\left(p^2\sigma^2\right)-1}}$$

$$\rho_{zz^p} = \frac{\left\{\exp\left(p\sigma^2\right)-1\right\}}{\sqrt{\exp\left(\sigma^2\right)-1}\cdot\sqrt{\exp\left(p^2\sigma^2\right)-1}}. \tag{6}$$

It is an interesting observation from (6) that for the lognormal distribution, the correlation is independent of λ.

From equation (3) and (4),

$$\sigma_z/\sigma_{z^p} = \frac{\exp\left\{(\mu+\lambda)+\sigma^2/2\right\}\sqrt{\exp\left(\sigma^2\right)-1}}{\exp\left\{p(\mu+\lambda)+p^2\sigma^2/2\right\}\sqrt{\exp\left(p^2\sigma^2\right)-1}}. \tag{7}$$

4. Sample Statistics

From equation (1), writing $\widehat{x}_{c,t}$ in terms of the combining parameters ω, k and λ,

$$\widehat{x}_{c,t} = \omega\widehat{x}_t + (1-\omega)\left[\overline{x} + \left(1-k\sqrt{n+1-t}\right)r_{\widehat{z}\widehat{z}^p}\left(s_{\widehat{z}}/s_{\widehat{z}^p}\right)\left\{\left(\widehat{x}_t+\lambda\right)^p - \mathbb{E}\left[\left(\widehat{x}_t+\lambda\right)^p\right]\right\}\right].$$

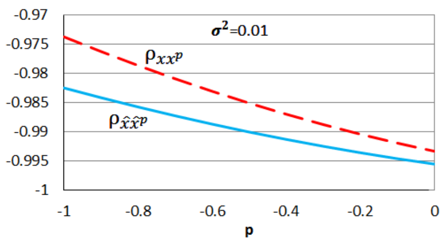
Substituting σ_z/σ_{z^p} from (7) for $s_{\widehat{z}}/s_{\widehat{z}^p}$, substituting for $\mathbb{E}\left[\left(X_t+\lambda\right)^p\right]$, and substituting sample fitted values $\overline{\ln\left(\widehat{x}+\lambda\right)}$ and s for population variables $\mu+\lambda$ and σ respectively,

$$\begin{aligned}\widehat{x}_{c,t} =& \omega\widehat{x}_t + (1-\omega)\left[\overline{x} + \left(1-k\sqrt{n+1-t}\right)r_{\widehat{z}\widehat{z}^p}\frac{\exp\left\{\overline{\ln\left(\widehat{x}+\lambda\right)}+s^2/2\right\}\sqrt{\exp\left(s^2\right)-1}}{\exp\left\{p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)}+p^2s^2/2\right\}\sqrt{\exp\left(p^2s^2\right)-1}}\right.\\ &\left.\times\left\{\left(\widehat{x}_t+\lambda\right)^p - \exp\left(p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)}+p^2s^2/2\right)\right\}\right]\\ =& \omega\widehat{x}_t + (1-\omega)\left[\overline{x} + \left(1-k\sqrt{n+1-t}\right)\varphi\left(p,\lambda,\widehat{x}_t\right)\right],\end{aligned} \tag{8}$$

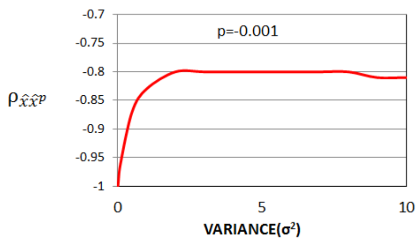
where

$$\begin{aligned}\varphi\left(p,\lambda,\widehat{x}_t\right) \quad = \quad & r_{\widehat{z}\widehat{z}^p}\frac{\exp\left\{\overline{\ln\left(\widehat{x}+\lambda\right)}+s^2/2\right\}\sqrt{\exp\left(s^2\right)-1}}{\exp\left\{p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)}+p^2s^2/2\right\}\sqrt{\exp\left(p^2s^2\right)-1}}\times\\ &\left[\left(\widehat{x}_t+\lambda\right)^p - \exp\left(p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)}+p^2s^2/2\right)\right].\end{aligned}$$

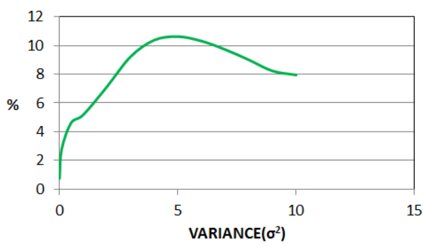
Discussion. *The values of* s^2 *and* σ^2 *depend on the particular data. Typically,* s^2 *and* σ^2 *being the standard deviations of the logarithms of* $\widehat{x}_t$ *and* X_t*, respectively, are naturally small. If* σ^2 *were zero, because there are no variations in the time series to be explained by the time series model, then there would be no MSE and therefore no MSE to be improved by antithetic combining. As the variance is increased, the amount of variance unexplained by the fitted model increases, contributing to the amount of MSE improvement that is possible. On the other hand,* $r_{\widehat{x}\widehat{x}^p}$ *starts out at approximately* -1 *(see Figure 1a), with near maximum combining contribution to MSE improvement, but as* $r_{\widehat{x}\widehat{x}^p}$ *moves away from* -1 *(see Figure 1b), the error in estimating* $\widehat{x}_t$ *from* $\widehat{x}_t^p$ *increases (see Appendix A), and the combining contribution falls. The joint effect is an improvement in MSE that peaks and then declines.*



(a) Correlation coefficient vs. p



(b) Correlation coefficient vs. σ^2



(c) Percentage reduction in MSE due to antithetic combining vs σ^2

Figure 1: Computer Simulations

To illustrate this, the results of a hypothetical simulation for $X_t = \exp(Y_t)$, where $Y_t = 0.8Y_{t-1} + \epsilon_t$, $t = 2, 3, \cdots, 1000$, and $\epsilon_t \sim N(0, \sigma^2)$, is shown in Figure 1c. The maximum improvement in MSE is about 10.6% where the variance is about 5. In practice, antithetic combining will perform more or less well depending on the variance of an actual time series. It is a simple matter to apply pre and post antithetic model fitting transformations $X_t^* = X_t^g$ and $\left(X_t^*\right)^{1/g}$, where g is chosen to accomplish any desired variance.

From Appendix B,

$$\lim_{p \to 0^-, s \to 0} \varphi\left(p, \lambda, \widehat{x}_t\right) = \exp\left\{\overline{\ln\left(\widehat{x} + \lambda\right)}\right\}\left\{-\overline{\ln\left(\widehat{x} + \lambda\right)} + \ln\left(\widehat{x}_t + \lambda\right)\right\}, \tag{9}$$

therefore, as $p \to 0^-$, (8) becomes

$$\lim_{p\to 0^-} \widehat{x}_{c,t} = \omega \widehat{x}_t + (1-\omega)\left[\overline{x} + \left(1 - k\sqrt{n+1-t}\right) \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\} \times \left\{-\overline{\ln(\widehat{x}+\lambda)} + \ln(\widehat{x}_t + \lambda)\right\}\right]. \tag{10}$$

5. Optimizing ω, λ and k

Let the error in $\widehat{x}_{c,t}$be $\xi_t = \widehat{x}_{c,t} - x_t$. We are interested in finding the optimal ω, λ and k that minimize $\sum_{t=2}^{n} \xi_t^2$ by setting the limit as p approaches 0 from the left of the partial derivatives $\lim_{p\to 0^-, s\to 0} \frac{\partial}{\partial \omega,\lambda,k} \sum_{t=2}^{n} \xi_t^2 = 0$ and solving for ω, λ and k. This can be simplified by permutation of the limit and the derivatives (see Buck [1]), and solving $\frac{\partial}{\partial \omega,\lambda,k} \lim_{p\to 0^-, s\to 0} \sum_{t=2}^{n} \xi_t^2 = 0$. The limit as p approaches zero from the left of the sum of squared errors is

$$\lim_{p\to 0^-, s\to 0} \sum_{t=2}^{n} \xi_t^2 = \lim_{p\to 0^-, s\to 0} \sum_{t=2}^{n} \left[\widehat{x}_{c,t} - x_t\right]^2.$$

Applying the power rule for limits $\lim_{p\to 0^-, s\to 0} \sum_{t=2}^{n} \xi_t^2 = \sum_{t=2}^{n} \left\{\lim_{p\to 0^-, s\to 0} \left(\widehat{x}_{c,t} - x_t\right)\right\}^2$, and from (9)

$$\lim_{p\to 0^-, s\to 0} \sum_{t=2}^{n} \xi_t^2 = \sum_{t=2}^{n} \left[\omega \widehat{x}_t + (1-\omega)\left[\overline{x} + \left(1 - k\sqrt{n+1-t}\right) \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\} \times \left\{-\overline{\ln(\widehat{x}+\lambda)} + \ln(\widehat{x}_t + \lambda)\right\}\right] - x_t\right]^2,$$

$$\frac{\partial \sum_{t=2}^{n} \xi_t^2}{\partial \omega} = 2\sum_{t=2}^{n} \left[\omega \widehat{x}_t + (1-\omega)\left[\bar{x} + (1 - k\sqrt{n+1-t}) \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\} \times \left\{-\overline{\ln(\widehat{x}+\lambda)} + \ln(\widehat{x}_t + \lambda)\right\}\right] - x_t\right] \left[\widehat{x}_t - \bar{x} - (1 - k\sqrt{n+1-t}) \times \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\} \left\{-\overline{\ln(\widehat{x}+\lambda)} + \ln(\widehat{x}_t + \lambda)\right\}\right],$$

$$\frac{\partial \sum_{t=2}^{n} \xi_t^2}{\partial k} = 2\sum_{t=2}^{n} \left[\omega \widehat{x}_t + (1-\omega)\left[\bar{x} + (1 - k\sqrt{n+1-t}) \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\} \times \left\{-\overline{\ln(\widehat{x}+\lambda)} + \ln(\widehat{x}_t + \lambda)\right\}\right] - x_t\right] \left[-(1-\omega)\sqrt{n+1-t} \times \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\} \left\{-\overline{\ln(\widehat{x}+\lambda)} + \ln(\widehat{x}_t + \lambda)\right\}\right],$$

$$\frac{\partial \sum_{t=2}^{n} \xi_t^2}{\partial \lambda} = 2\sum_{t=2}^{n} \left[\omega \widehat{x}_t + (1-\omega)\left[\bar{x} + (1 - k\sqrt{n+1-t}) \exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\}\right.\right.$$

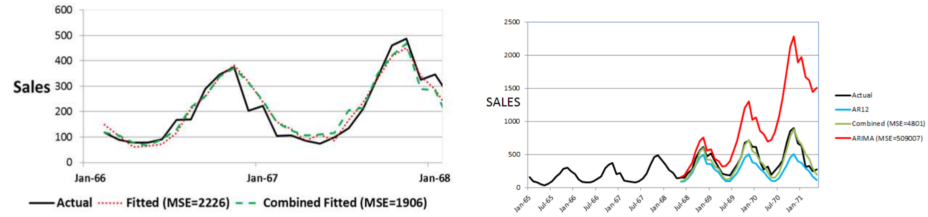
$$\times \left\{-\overline{\ln(\widehat{x}+\lambda)}+\ln(\widehat{x}_t+\lambda)\right\}\Big] - x_t\Big]\Big[(1-\omega)\left(1-k\sqrt{n+1-t}\right)$$
$$\times \overline{1/(\widehat{x}+\lambda)}\exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\}\left\{-\overline{\ln(\widehat{x}+\lambda)}+\ln(\widehat{x}_t+\lambda)\right\}$$
$$+(1-\omega)\left(1-k\sqrt{n+1-t}\right)\exp\left\{\overline{\ln(\widehat{x}+\lambda)}\right\}\left\{-\overline{1/(\widehat{x}+\lambda)}+1/(\widehat{x}_t+\lambda)\right\}\Big].$$

6. Empirical Example

For example, consider the CompanyX case study by Chatfield and Prothero [2]. The 77 months of data for the study (given in that paper) are charted below in Figures 2a & 2b. That paper also reported the results of fitting an ARIMA model to the data: $(1+.37\mathfrak{B})\nabla\nabla_{12}x_t^{.34}=(1-.79\mathfrak{B}^{12})a_t$, where x_t are data, ∇ is the differencing operator, $\mathfrak{B}$ is the backward shift operator, and a_t are random errors. The discussants identified many ways, including model misspecification, in which the results may have been biased. The authors acknowledged that forecasts from the ARIMA model tended to be biased high and divergent. We recognize that outcome as resulting from sampling bias that antithetic time series analysis was created to address. The distribution of this data is skewed to the right, similar to the lognormal distribution.

This gives us an independent benchmark with which to compare combined antithetic fitted values. In this paper we fit a twelfth order autoregressive model AR12: $x_t=\sum_{l=1}^{12}\Phi_l x_{t-l}+e_t$, $t=1,2,\cdots$ to the first forty data values from January 1965 to April 1968. That makes thirty seven data values from May 1968 to May 1971 available for comparison with a forecast for that period. The model parameter estimates are $\widehat{\Phi}_1=.123$, $\widehat{\Phi}_2=.025$, $\widehat{\Phi}_3=-.055$, $\widehat{\Phi}_4=.019$, $\widehat{\Phi}_5=.005$, $\widehat{\Phi}_6=.007$, $\widehat{\Phi}_7=-.012$, $\widehat{\Phi}_8=-.010$, $\widehat{\Phi}_9=.004$, $\widehat{\Phi}_{10}=.039$, $\widehat{\Phi}_{11}=.007$, $\widehat{\Phi}_{12}=.884$. The actual and fitted values are shown in Figure 2a. The combining procedure described in Appendix A is then applied to the fitted values (the calculations are performed using the computer program FOURCAST[‡]). The optimal combining model parameters are $\omega=1.17, \lambda=432, k=-0.153$. The AR12 model original and combined fitted MSE values are 2226 and 1906 respectively, a reduction of 14.4%. The forecasts are shown in Figure 2b. The April 1967 base month May 1968 to May 1971 ex ante 37 month ARIMA, AR12 (not shown) and combined forecast MSE values are 509007, 30759 and 4801 respectively. When the AR12 model is refitted to $x_t^{.34}$ to correct it for heteroscedasticity, the resulting ex ante 37 month forecast MSE is 30683. Either way, the combined forecast MSE is 99% less than that for the ARIMA model and 84% less than that for the AR12 model. The ARIMA and AR12 forecasts are consistently high and low respectively. While the ARIMA forecasts diverge, the combined antithetic forecasts converge.

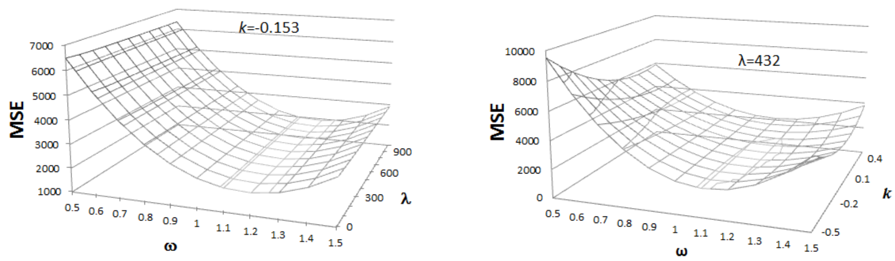
[‡]`http://www.fourcast.net`, CompanyX data available here: `http://www.fourcast.net/fourcast/CompanyX.zip`



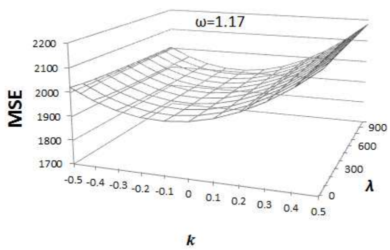
(a) Actual, fitted, & combined fitted values (b) Forecasts for May 1968 to May 1971

Figure 2: CompanyX Data

To obtain an improved understanding of the combining function, we explore the effects of ω, λ, k on MSE. The combined fitted MSE values are plotted in Figures 3a, 3b, 3c. In each case MSE is a smooth continuous surface. This supports our proposal that the derivatives $\frac{\partial}{\partial_{\omega,\lambda,k}}$ exist. In this example, MSE is not very sensitive to changes in λ. It does no harm and will be necessary for cases when the data contain negative numbers.



(a) MSE vs. ω, λ (b) MSE vs. ω, K



(c) MSE vs. K, λ

Figure 3: MSE vs ω, λ, K

7. Conclusions

Given that the parameters of a time series model are to be estimated. It is well known that for a *specified MSE*=Variance +Bias^2, the smaller the bias of the estimate, the larger is its variance and vice versa. Antithetic time series theory is intended to eliminate the bias and reduce the MSE for a *specified variance* of the fitted values that are obtained from the mathematical model. The specified variance is associated with the purely random error component. The bias is associated with a systematic error component. Antithetic time series analysis can be applied to reduce bias in fitted values from autoregressive time series models so as to reduce MSE. Where there is no bias, the combining parameter ω is simply $= 1.0$ and the combined values are the original values. The combined MSE is a smooth, continuous, differentiable function of the combining parameters w, λ and k. A suggestion for future research is to investigate the applicability of antithetic combining to data distributions other than the lognormal distribution.

Appendix A: Bias Reduction

Since $\lim_{p\to 0^-, s\to 0} \mathrm{Corr}(x, x^p) = -1$, we can express x_t^p in the original units of x_t by means of the linear regression of x_t on x_t^p: $x_t = c_0 + c_1 x_t^p + \varepsilon_t$, where ε_t is an error term that approaches zero, so that a near perfect estimate of x_t is obtained from

$$x_t' = c_0 + c_1 x_t^p. \tag{A1}$$

Now, suppose that $x_t = f\left(x_{t-1}\right) + \epsilon_t,\ t = 2, \cdots, n$ is a time series model that is biased due either to serially correlated errors ϵ_t or sampling error or both. Estimates $\widehat{x}_t$ of x_t from this model will be biased. To remove this bias, we power transform $\widehat{x}_t$ to obtain $\widehat{x}_t^p$. Then, we use equation (A1) to convert $\widehat{x}_t^p$ back to the original units of x_t. Hence

$$\widehat{x}_t' = \widehat{c}_0 + \widehat{c}_1 \widehat{x}_t^p, \tag{A2}$$

where $\widehat{c}_0$and $\widehat{c}_1$are least squares estimates obtained from the regression of $\widehat{x}_t$on $\widehat{x}_t^p$, and the error approaches zero.

Both estimates $\widehat{x}_t$ and $\widehat{x}_t'$ contain errors. These errors contain two components. One component is purely random and one component is systematic bias. Combining the estimates cancels the systematic bias component, leaving only the purely random component. The combined estimate $\widehat{x}_{c,t}$ is obtained from $\widehat{x}_{c,t} = \omega \widehat{x}_t + (1-\omega)\widehat{x}_t'$, where $-\infty \le \omega \le \infty$, and ω is chosen to minimize the MSE $= \sum_{t=2}^{n} \left(x_t - \widehat{x}_{c,t}\right)^2/(n-1)$. Likewise, the unbiased combined estimate of a future value at time $\tau$is obtained from $\widehat{x}_{c,n}(\tau) = \omega \widehat{x}_n(\tau) + (1-\omega)\widehat{x}_n'(\tau)$. Substituting for $\widehat{c}_0$ in (A2), $\widehat{x}_t' = \overline{\widehat{x}} - \widehat{c}_1 \overline{\widehat{x}^p} + \widehat{c}_1 \widehat{x}_t^p = \overline{x} + \widehat{c}_1 \left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)$. Denoting sample variance and covariance in x_t by s_x^2, $\widehat{x}_t' = \overline{x} + \left(s^2_{\widehat{x}\widehat{x}^p}/s^2_{\widehat{x}^p}\right)\left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)$, $\widehat{x}_t' = \overline{x} + \left(s^2_{\widehat{x}\widehat{x}^p}/s^2_{\widehat{x}^p}\right)\left(s_{\widehat{x}}/s_{\widehat{x}}\right)\left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)$ and $\widehat{x}_t' = \overline{x} + r_{\widehat{x}\widehat{x}^p}\left(s_{\widehat{x}}/s_{\widehat{x}^p}\right)\left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)$. The combined fitted values are given by
$\widehat{x}_{c,t} = w\widehat{x}_t + (1-\omega)\left\{\overline{x} + r_{\widehat{x}\widehat{x}^p}\left(s_{\widehat{x}}/s_{\widehat{x}^p}\right)\left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)\right\}$.

The steps for obtaining the combined antithetic fitted values are outlined as follows

Step 1: Estimate the model parameters and fitted values $\widehat{x}_t = f\left(x_{t-1}\right), t = 2, 3, \cdots, n$

Step 2: Set $p = -0.001$ to approximate $p \to 0^-$

Step 3: Calculate $\omega = \sum_{t=2}^{n}\left(x_t - \widehat{x}'_t\right)\left(\widehat{x}_t - \widehat{x}'_t\right) \div \sum_{t=2}^{n}\left(\widehat{x}_t - \widehat{x}'_t\right)^2$

Step 4: Calculate $\widehat{x}_{c,t} = \omega\widehat{x}_t + (1-\omega)\left\{\overline{x} + r_{\widehat{x}\widehat{x}^p}\left(s_{\widehat{x}}/s_{\widehat{x}^p}\right)\left(\widehat{x}_t^p - \overline{\widehat{x}^p}\right)\right\},\ t = 2,3,\cdots,n$

Appendix B: Limit as $p \to 0^-$, $s \to 0$

We wish to find

$$\lim_{p\to 0^-, s\to 0} \varphi\left(p,\lambda,\widehat{x}_t\right) = \lim_{p\to 0^-, s\to 0} r_{\widehat{z}\widehat{z}^p} \frac{\exp\left\{\overline{\ln\left(\widehat{x}+\lambda\right)} + s^2/2\right\}\sqrt{\exp\left(s^2\right)-1}}{\exp\left\{p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2\right\}\sqrt{\exp\left(p^2s^2\right)-1}} \times\left[\left(\widehat{x}_t+\lambda\right)^p - \exp\left(p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2\right)\right].$$

Consider the Taylor expansion at $p = 0$ of

$$\exp(p) = 1 + p + \frac{p^2}{2} + \frac{p^3}{6} + \frac{p^4}{24} + o\left(p^4\right). \tag{A3}$$

Using (A3), we derive the Taylor expansion at $p = 0$ of $\exp(p^2s^2) - 1 = p^2s^2 + \frac{p^4s^4}{2} + \frac{p^6s^6}{6} + o\left(p^6\right)$. Hence,

$$\frac{1}{\sqrt{\exp\left(p^2s^2\right)-1}} = \frac{1}{\sqrt{p^2s^2 + \frac{p^4s^4}{2} + \frac{p^6s^6}{6} + o\left(p^6\right)}} = \frac{1}{s|p|\sqrt{1 + \frac{p^2s^2}{2} + \frac{p^4s^4}{6} + o\left(p^4\right)}}. \tag{A4}$$

Using equation (A3), we derive the Taylor expansion at $p = 0$ of

$$\begin{aligned}\exp\left(p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2\right) &= 1 + p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2 + \frac{1}{2}\left\{p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2\right\}^2 + o\left(p^2\right)\\ &= 1 + p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2 + \frac{p^2}{2}\cdot\overline{\ln\left(\widehat{x}+\lambda\right)^2} + o\left(p^2\right).\end{aligned} \tag{A5}$$

The Taylor expansion at $p = 0$ of $\left(\widehat{x}_t+\lambda\right)^p$ is

$$\left(\widehat{x}_t+\lambda\right)^p = 1 + p\cdot\ln\left(\widehat{x}+\lambda\right) + \frac{p^2}{2}\cdot\ln^2\left(\widehat{x}+\lambda\right) + o\left(p^2\right). \tag{A6}$$

By subtracting (A5) from (A6), we obtain the Taylor expansion at $p = 0$ of

$$\begin{aligned}\left[\left(\widehat{x}_t+\lambda\right)^p - \exp\left\{p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)} + p^2s^2/2\right\}\right] =& p\left\{-\overline{\ln\left(\widehat{x}+\lambda\right)} + \ln\left(\widehat{x}_t+\lambda\right)\right\}\\ &+ \frac{p^2}{2}\left\{-\overline{\ln\left(\widehat{x}+\lambda\right)^2} - s^2 + \ln^2\left(\widehat{x}_t+\lambda\right)\right\} + o\left(p^2\right).\end{aligned} \tag{A7}$$

The Taylor expansion of $\varphi\left(p, \lambda, \widehat{x}_t\right)$ at $p = 0$ may now be obtained by multiplication of $r_{\widehat{z}\widehat{z}^p} \frac{\exp\left\{\overline{\ln(\widehat{x}+\lambda)}+s^2/2\right\}\sqrt{\exp(s^2)-1}}{\exp\left\{p\cdot\overline{\ln(\widehat{x}+\lambda)}+p^2s^2/2\right\}}$ and the expansions (A4) and (A7) as follows.

$$\varphi\left(p, \lambda, \widehat{x}_t\right) = r_{\widehat{z}\widehat{z}^p} \frac{\exp\left\{\overline{\ln\left(\widehat{x}+\lambda\right)}+s^2/2\right\}\sqrt{\exp\left(s^2\right)-1}}{\exp\left\{p\cdot\overline{\ln\left(\widehat{x}+\lambda\right)}+p^2s^2/2\right\}} \cdot \frac{1}{s|p|\sqrt{1+\frac{p^2s^2}{2}+\frac{p^4s^4}{6}+o\left(p^4\right)}}$$
$$\times\left[p\left\{-\overline{\ln\left(\widehat{x}+\lambda\right)}+\ln\left(\widehat{x}_t+\lambda\right)\right\}+\frac{p^2}{2}\left\{-\overline{\ln\left(\widehat{x}+\lambda\right)^2}-s^2+\ln^2\left(\widehat{x}_t+\lambda\right)\right\}+o\left(p^2\right)\right].$$

From (6) and Ridley [4], $r_{\widehat{z}\widehat{z}^p} \downarrow -1$ as $p \downarrow 0^-$, $s \downarrow 0$ and $\lim_{s\to 0} \frac{\sqrt{\exp(s^2)-1}}{s} = 1$, therefore

$$\lim_{p\to 0^-, s\to 0} \varphi\left(p, \lambda, \widehat{x}_t\right) = \exp\left\{\overline{\ln\left(\widehat{x}+\lambda\right)}\right\}\left\{-\overline{\ln\left(\widehat{x}+\lambda\right)}+\ln\left(\widehat{x}_t+\lambda\right)\right\}.$$

References

[1] C. R. Buck. *Advanced Calculus, 3rd Edition*. McGraw-Hill, New York, 1978.

[2] C. Chatfield and D. L. Prothero. Box-jenkins seasonal forecasting: Problems in a case study. *Journal of the Royal Statistical Society: Series A*, 136:275–336, 1973.

[3] Z. Griliches. A note on serial correlation bias in estimates of distributed lags. *Econometrica*, 29:65–73, 1961.

[4] A. D. Ridley. Optimal antithetic weights for lognormal time series forecasting. *Computers & Operations Research*, 26(3):189–209, 1999.

[5] A. D. Ridley and P. Ngnepieba. Antithetic time series analysis and the CompanyX data. *Journal of the Royal Statistical Society, A*, 176(4):1–12, 2013.