Scientific
Research
Publishing

General Theory of Antithetic Time Series

Pierre Ngnepieba[1], Dennis Ridley[2,3*]

[1]Department of Mathematics, Florida A&M University, Tallahassee, USA
[2]SBI, Florida A&M University, Tallahassee, USA
[3]Department of Scientific Computing, Florida State University, Tallahassee, USA
Email: pierre.ngnepieba@famu.edu, *dridley@fsu.edu

Abstract

A generalized antithetic time series theory for exponentially derived antithetic random variables is developed. The correlation function between a generalized gamma distributed random variable and its pth exponent is derived. We prove that the correlation approaches minus one as the exponent approaches zero from the left and the shape parameter approaches infinity.

Keywords

Antithetic Time Series Theory, Antithetic Random Variables, Bias Reduction, Gamma Distribution, Inverse Correlation, Serial Correlation

1. Introduction

Serially correlated random variables arise in ways that both benefit and bias mathematical models for science, engineering and economics. In one widespread example, a mathematical formula is used to create uniformly distributed pseudo random numbers for use in Monte Carlo simulation. The numbers are serially correlated because they are generated by a formula. The benefit is that the same pseudo random numbers can be recreated at will, and two or more simulation experiments can be compared without regard to the pseudo random numbers. The correlation is designed to be very small so as not to bias the results of a simulation experiment. Still, some bias is unavoidable when using serially correlated numbers (Ferrenberg, Lanau, and Wong [1]).

Another wide spread example is a regression model in which the dependent variable is serially correlated. The result is biased model parameter estimates because the independence assumption of the Gauss-Markov theorem is violated (see Griliches [2], Nerlove [3], Koyck [4], Klein [5]). Similarly, the independence assumption of Fuller and Hasza [6] and Dufour [7] would not apply. 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

---

*Corresponding author.

data, then the absence of normality is like missing information. Bias may also be due to missing data points (Chandan and Jones [8], Li, Nychka and Amman [9]). Assume that a perfect model is postulated for a given application in which the population to which the data belong is known exactly. The model must be fitted to a sample of data, not the population. However, once the sample is taken, the distribution is automatically truncated and distorted, and the fitted model is biased. Regardless of the method of fitting, however small, sampling bias is unavoidable. One approach aimed at improving model performance is to combine the results from different models. For an extensive discussion and review of traditional combining see Bunn [10], Diebold [11], Clemen [12], Makridakis *et al.* [13], and Winkler [14].

Economics researchers have commented on serial correlation bias. Hendry and Mizon [15] and Hendry [16] considered common factor analysis (Mizon [17]) and suggested that serial correlation is a feature for representing dynamic relationships in economic models. This in turn implies that economics allows for serial correlation (see Pindyck and Rubinfield [18]). Time domain methods for detecting the nature and presence of serial correlation were considered by Durbin and Watson [19] and Durbin [20]. Spectral methods were considered by Hendry [16], Osborn [21] and Espasa [22]. Even if serial correlation can be a tool for studying the nature of economics, it is detrimental to long range forecasting models. Whatever the source of bias may be, the only possibility for long range forecasting is to completely eliminate the bias.

1.1. Background

Inversely correlated random numbers were suggested by Hammersley and Morton [23] for use in Monte Carlo computer simulation experiments. In that application, a single computer simulation is replaced by two simulations. One simulation uses uniformly distributed $(0,1)$ random numbers in r. The other simulation uses $1-r$. The expectation is that the average of the results of these two simulations has a smaller variance than for either one. In practice, the variance sometimes decreases, but sometimes it increases. See also Kleijnen [24].

The theory of combining antithetic lognormally distributed random variables that contain negatively correlated components was introduced by Ridley [25]. The Ridley [25] antithetic time series theorem states that "if $X_t > 0, t = 1, 2, 3, \cdots$ is a discrete realization of a lognormal stochastic process, such that $\ln X_t \sim N(\mu, \sigma^2)$, then if the correlation between X_t and X_t^p is ρ, then $\lim_{p \to 0^-, \sigma \to 0^+} \rho = -1$." Antithetic variables can be combined so as to eliminate bias in fitted values associated with any autoregressive time series model (see the Ridley [25] antithetic fitted function theorem, and antithetic fitted error variance function theorem). Similarly, antithetic forecasts obtained from a time series model can be combined so as to eliminate bias in the forecast error. Ridley [26] applied combined antithetic forecasting to a wide range of data distributions. Ridley [27] demonstrated the methodology for optimizing weights for combining antithetic forecasts. See also Ridley and Ngnepieba [28] and Ridley, Ngnepieba, and Duke [29]. The antithetic variables proof in Ridley [25] was for the special case of X_t lognormally distributed.

The implication for using a biased mathematical model to investigate economic, engineering and scientific phenomena is that estimates obtained from the model are biased. Estimates of future values extrapolated from the model are also biased. As the forecast horizon increases, the bias accumulates and the extrapolations diverge from the actual values. This is most pronounced in the case of investigations into global warming phenomena. There, the horizon is by definition very far into the future. The smallest bias will accumulate, so much so that conclusions may be as much an artifact of the mathematical model as they are about climate dynamics. Combining antithetic extrapolations can dynamically remove the bias in the extrapolated values.

1.2. Proposed Research

The antithetic gamma variables discussed in this research are defined as follows.

Definition 1. *Two random variables are antithetic if their correlation is negative. A bivariate collection of random variables is asymptotically antithetic if its limiting correlation approaches minus one asymptotically (see antithetic gamma variables theorem below).*

Definition 2. $\{X(\xi, t)\}$ *is an ensemble of random variables, where ξ belongs to a sample space and t belongs to an index set representing time, such that X_t is a discrete realization of a gamma stationary stochastic process from the ensemble, $X_t \sim \Gamma(\alpha, \beta)$, and $X_t, t = 1, 2, 3, \cdots$ are serially correlated.*

In this paper, we extend the discovery by Ridley [25] beyond the lognormal distribution. The gamma distribution is very important for technical reasons, since it is the parent of the exponential distribution and can

explain many other distributions. That is, a wide range of distributions can be represented by the gamma distribution. We will explore these possibilities by examining the correlation between X and X^p when X is gamma distributed. Of particular interest is the correlation between X and X^p as $p \to 0^-$. A graph of the correlation between X and X^p as $p \to 0^+$ and $p \to 0^-$ is shown in **Figure 1**. We begin by reviewing the obvious results for the case when p is positive. The correlation is positive when p is positive and exactly one when p is one. This is expected. As p moves away from one, the correlation decreases. As p approaches zero from the right, the correlation falls, albeit very slowly. This is also expected. In the case when p is negative, the correlation behaves quite differently. The result is entirely counterintuitive. As expected, the correlation is negative. But, unlike when p is positive, as p approaches zero from the left, the absolute value of the correlation increases. Furthermore, the actual correlation approaches minus one, not zero.

One purpose of this paper is to derive an analytical function for the correlation between X and X^p when X is gamma distributed. A second purpose is to explore by extensive computation, the behavior of the correlation as p approaches zero from the left. The trivial case of p equal zero where the correlation is zero, is of no interest. We are interested in p inside a delta neighborhood of zero, not zero. Finally, we prove that the limiting value of the correlation is minus one.

The paper is organized as follows. In Section 2 we review the gamma distribution. In Section 3 we derive the analytic function for the correlation. In Section 4 we prove its limiting value. In Section 5 we use MATLAB [30] to compute correlations for a wide range of values generated from the gamma distribution. In Section 6 we outline the method for using antithetic variables to dynamically remove bias from the fitted and forecast values obtained from a time series model. Examples include computer simulated data. Section 7 contains conclusions and suggestions for further research.

2. The Gamma Distribution

The gamma distribution is very important for technical reasons, since it is the parent of the exponential distribution and can explain many other distributions. Its probability distribution function (pdf) (see Hogg and Ledolter [31]) is:

$$f(X;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} X^{\alpha-1} \exp\left(-\dfrac{X}{\beta}\right), & \text{if } X \geq 0 \\ 0, & \text{if } X < 0, \end{cases} \tag{1}$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter. The gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty X^{\alpha-1} e^{-X}\, dX. \tag{2}$$

A graph of the gamma probability density function for $\beta = 0.6$ and various values of $\alpha = 1.1, 3, 5, 9, 11$ is shown in **Figure 2**.
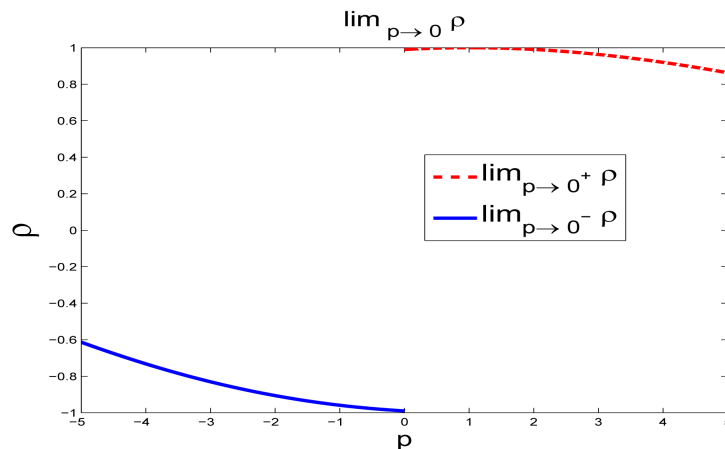

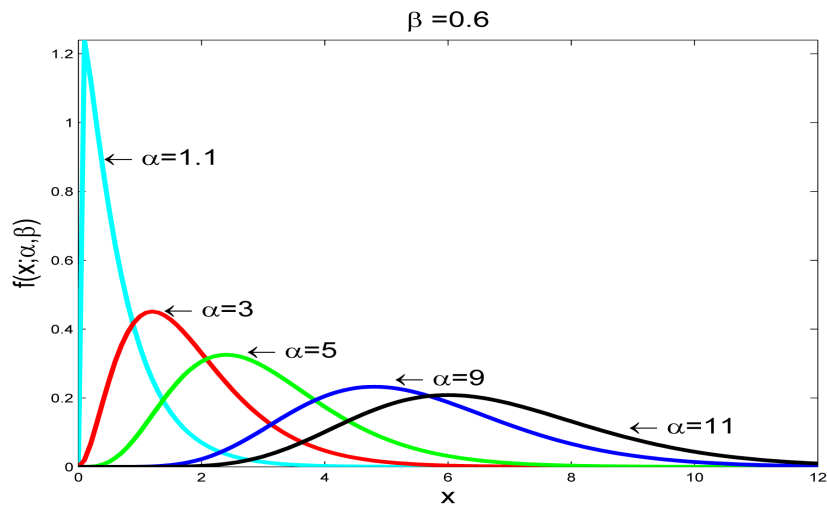**Figure 1.** Behavior of ρ as p approaches 0.

Figure 2. Exploring the effect of varying parameter values in the pdf of the gamma distribution.

3. Correlation between X and X^p

Let $X_t > 0$, $t = 1, 2, 3, \cdots$, be a discrete realization of a generalized gamma stochastic process. For $p \in \mathbb{R}$, from Appendix A, the *pth* moment of the gamma distribution is given by

$$\mathbb{E}\left[X_t^p\right] = \beta^p \frac{\Gamma(\alpha + p)}{\Gamma(\alpha)}. \tag{3}$$

Therefore, since $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, the mean is

$$\mathbb{E}[X_t] = \beta \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)} = \alpha\beta,$$

the second moment is

$$\mathbb{E}\left[X_t^2\right] = \beta^2 \frac{\Gamma(\alpha + 2)}{\Gamma(\alpha)} = \beta^2 \alpha(\alpha + 1),$$

and the variance is

$$\sigma^2 = \mathbb{E}\left[X_t^2\right] - \left(\mathbb{E}[X_t]\right)^2 = \alpha\beta^2. \tag{4}$$

Let ρ be the correlation between X_t and X_t^p. Then

$$\rho = \frac{\mathrm{Cov}\left(X_t^p, X_t\right)}{\left\{\mathrm{Var}\left(X_t^p\right)\mathrm{Var}(X_t)\right\}^{\frac{1}{2}}} = \frac{\mathbb{E}\left[X_t^{p+1}\right] - \mathbb{E}\left[X_t^p\right]\mathbb{E}[X_t]}{\left\{\mathrm{Var}\left(X_t^p\right)\mathrm{Var}(X_t)\right\}^{\frac{1}{2}}}, \tag{5}$$

and

$$\mathrm{Var}\left(X_t^p\right) = \mathbb{E}\left[X_t^{2p}\right] - \left(\mathbb{E}\left[X_t^p\right]\right)^2.$$

Using Equation (3)

$$\mathrm{Var}\left(X_t^p\right) = \beta^{2p} \frac{\Gamma(\alpha + 2p)}{\Gamma(\alpha)} - \beta^{2p} \frac{\left(\Gamma(\alpha + p)\right)^2}{\left(\Gamma(\alpha)\right)^2}. \tag{6}$$

Therefore, using Equations (3) and (6), Equation (5) becomes

$$\rho = \frac{\beta^{p+1}\frac{\Gamma(\alpha+p+1)}{\Gamma(\alpha)} - \alpha\beta^{p+1}\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}}{\left\{\beta^{2p}\frac{\Gamma(\alpha+2p)}{\Gamma(\alpha)} - \left(\beta^p\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}\right)^2\right\}^{\frac{1}{2}}\left(\alpha\beta^2\right)^{\frac{1}{2}}} = \frac{\frac{\Gamma(\alpha+p+1)}{\Gamma(\alpha)} - \alpha\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}}{\left\{\frac{\Gamma(\alpha+2p)}{\Gamma(\alpha)} - \left(\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}\right)^2\right\}^{\frac{1}{2}}(\alpha)^{\frac{1}{2}}}$$

$$= \frac{\frac{\Gamma(\alpha+p+1)}{\Gamma(\alpha)} - \alpha\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}}{\frac{\alpha^{\frac{1}{2}}}{\Gamma(\alpha)}\left\{\Gamma(\alpha+2p)\Gamma(\alpha) - \left(\Gamma(\alpha+p)\right)^2\right\}^{\frac{1}{2}}} = \frac{\Gamma(\alpha+p+1) - \alpha\Gamma(\alpha+p)}{\alpha^{\frac{1}{2}}\left\{\Gamma(\alpha)\Gamma(\alpha+2p) - \left(\Gamma(\alpha+p)\right)^2\right\}^{\frac{1}{2}}}. \tag{7}$$

Since $\Gamma(\alpha+p+1) = (\alpha+p)\Gamma(\alpha+p)$, Equation (7) becomes,

$$\rho = \frac{p\Gamma(\alpha+p)}{\alpha^{\frac{1}{2}}\left\{\Gamma(\alpha)\Gamma(\alpha+2p) - \left(\Gamma(\alpha+p)\right)^2\right\}^{\frac{1}{2}}}. \tag{8}$$

From Equation (4) $\alpha^{\frac{1}{2}} = \frac{\sigma}{\beta}$, and the correlation can be expressed in terms of σ as

$$\rho = \frac{p\beta\Gamma(\alpha+p)}{\sigma\left\{\Gamma(\alpha)\Gamma(\alpha+2p) - \left(\Gamma(\alpha+p)\right)^2\right\}^{\frac{1}{2}}},$$

or

$$\rho = \frac{p\beta\Gamma\left(\left(\frac{\sigma}{\beta}\right)^2 + p\right)}{\sigma\left[\Gamma\left(\left(\frac{\sigma}{\beta}\right)^2\right)\Gamma\left(\left(\frac{\sigma}{\beta}\right)^2 + 2p\right) - \left\{\Gamma\left(\left(\frac{\sigma}{\beta}\right)^2 + p\right)\right\}^2\right]^{\frac{1}{2}}}. \tag{9}$$

The gamma function $\Gamma(\alpha)$ results in a complex number when the argument is negative. This is avoided if $p < \frac{\alpha}{2}$. In any case, since we are only interested in p approaching zero from the left, this condition will always be satisfied when $\alpha > 0$.

4. Antithetic Gamma Variables Theorem

Theorem 1. *If* $X_t > 0$, $t = 1, 2, 3, \cdots$, *is a discrete realization of a generalized gamma stochastic process with shape parameter* α, *then if* ρ *is the correlation between* X_t *and* X_t^p, *then*

$$\lim_{p\to 0^-,\,\alpha\to\infty} \rho = \lim_{p\to 0^-,\,\alpha\to\infty} \frac{p\Gamma(\alpha+p)}{\alpha^{\frac{1}{2}}\left\{\Gamma(\alpha)\Gamma(\alpha+2p) - \left(\Gamma(\alpha+p)\right)^2\right\}^{\frac{1}{2}}} = -1.$$

See proof in Appendix B.

5. Correlation versus *p*

The effect of *p* on the correlation is demonstrated by calculating the correlation coefficient from Equation (8) for various values of α and β. The correlation coefficients are listed in **Table 1** and plotted in **Figure 3** and **Figure 4**. From **Figure 3** and **Figure 4**, for all values of α, the correlation coefficient gets closer to -1 as $p \to 0^-$. For all values of *p*, the correlation coefficient gets closer to -1 as α increases. From **Figure 3**, smaller values of α produce distributions that are more asymmetrical, and larger values produce distributions
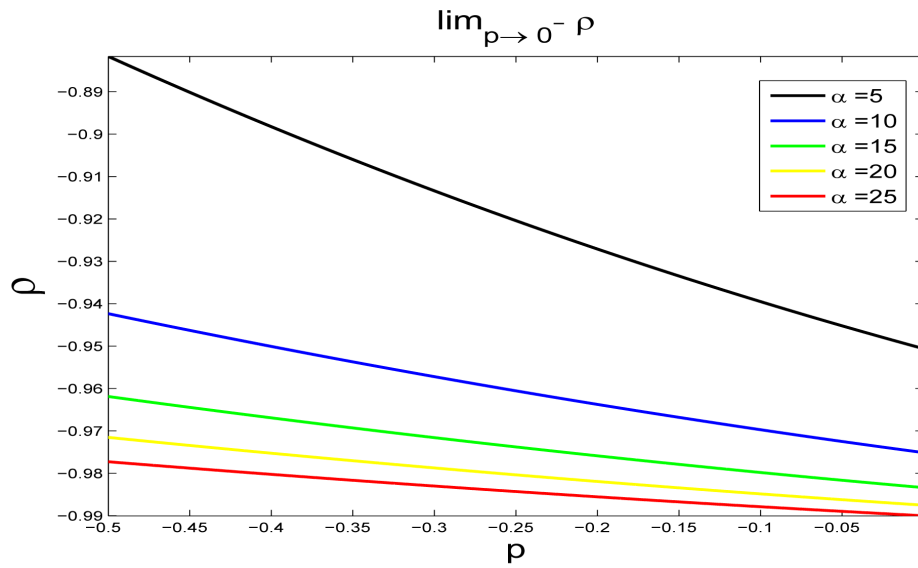
Figure 3. $\lim\limits_{p\to 0^-} \rho$ using Equation (8) for $\alpha = 5, 10, 15, 20$ and 25; $p = -0.5$ to -0.005.
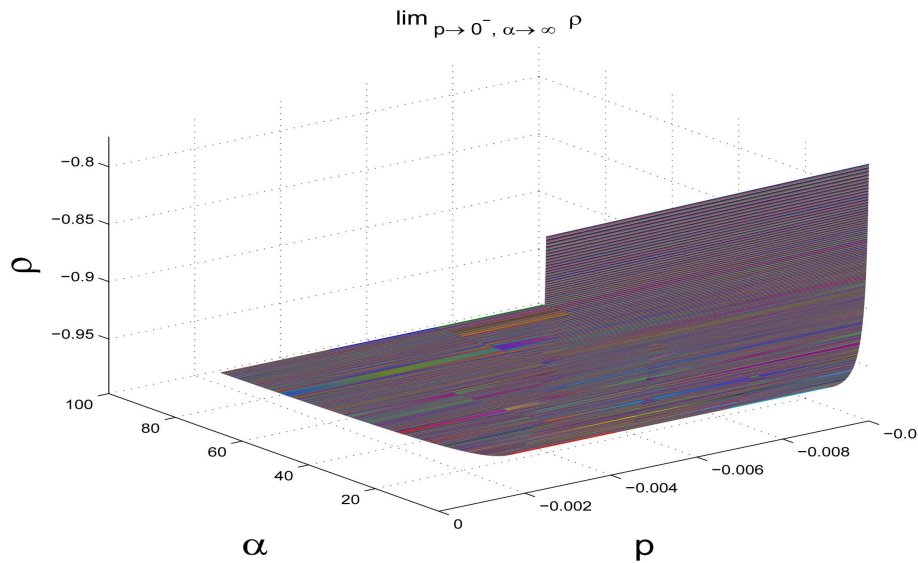

Figure 4. $\lim\limits_{p\to 0^-, \alpha\to\infty} \rho$ using Equation (8).

that are more symmetrical. Also, as α increases the standard deviation increases. This is indicative of greater spread about both sides of the mode of *X*. Equation (9) expresses the correlation in terms of σ. In practice the value of α will be that for the actual data under study. It cannot be modified. Still, one might say that it appears that the effect of p on reversing the correlation is greatest for symmetrical distributions.

To validate Equation (8), the MATLAB [30] random number generator GAMRND (α, β, *n*) is used to generate $n = 10000$ random numbers x_t, from the gamma distribution in Equation (1) with parameters $\alpha = 25$, $\beta = 0.5$. The correlation is estimated from the sample correlation coefficient ($\hat{\rho}$). One application of the correlation reversal is to remove bias in values extrapolated from a time series model (see Appendix C). The gamma distribution is immediately applicable when $\alpha \geq 25$, such that ρ is approximately minus one. For $\alpha < 25$, the difference between ρ and minus one may introduce an error in estimating values extrapolated from the time series model. The sample correlation coefficient is obtained from

$$\hat{\rho} = \sum_{t=1}^{n}\left(x_t - \overline{x}\right)\left(x_t^p - \overline{x^p}\right) \Bigg/ \left\{\sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2 \sum_{t=1}^{n}\left(x_t^p - \overline{x^p}\right)^2\right\}^{\frac{1}{2}}, \text{ where, } \overline{x} = \sum_{t=1}^{n} x_t \Big/ n, \ \overline{x^p} = \sum_{t=1}^{n} x_t^p \Big/ n, \ t = 1, 2, \cdots, n.$$

The results are shown in **Table 2**. The coefficients are almost identical to the theoretical values obtained from Equation (8) and listed in **Table 1**. In practice, the data may include relatively few observations. To investigate the small sample correlation coefficient, the correlation coefficient is calculated for $n = 100$.

6. Bias Reduction

Consider an autoregressive time series x_t of n discrete observations obtained from a gamma distribution with a large shape parameter to which a least squares model $x_t = \Phi x_{t-1} + \epsilon_t$, $t = 2, \cdots, n$ is fitted. Let the fitted values be $\hat{x}_t$. Next, consider the combined weighted average fitted values $\hat{x}_{c,t} = \omega\hat{x}_t + (1-\omega)\hat{x}_t'$, $-\infty \le \omega \le \infty$. The parameter ω is a combining weight. The fitted values $\hat{x}_t$ and $\hat{x}_t'$ are antithetic in the sense that they contain components of error $\hat{\epsilon}_t$ and $\hat{\epsilon}_t'$, respectively, that are biased and when weighted, $\omega\hat{\epsilon}_t$ and $(1-\omega)\hat{\epsilon}_t'$ are perfectly negatively correlated. The antithetic component $\hat{x}_t'$ is estimated from $\hat{x}_t' = \overline{x} + r_{\hat{x}x^p}\left(s_{\hat{x}}/s_{\hat{x}^p}\right)\left(\hat{x}_t^p - \hat{x}^p\right), t = 1, 2, 3, \cdots, n,$

Table 1. Behavior of ρ as $p \to 0^-$ using Equation (8) with various values of α.

$p \to 0^-$	$\alpha = 5$	$\alpha = 10$	$\alpha = 15$	$\alpha = 10$	$\alpha = 25$
$p = -0.5$	−0.8817	−0.9423	−0.9618	−0.9715	−0.9772
$p = -0.25$	−0.9204	−0.9605	−0.9737	−0.9803	−0.9843
$p = -0.1$	−0.9395	−0.9697	−0.9789	−0.9848	−0.9878
$p = -0.05$	−0.9452	−0.9724	−0.9816	−0.9862	−0.9888
$p = -0.025$	−0.9479	−0.9738	−0.9825	−0.9868	−0.9895
$p = -0.015$	−0.9490	−0.9743	−0.9828	−0.9871	−0.9897
$p = -0.005$	−0.9500	−0.9748	−0.9832	−0.9874	−0.9899
$p = -0.0025$	−0.9503	−0.9759	−0.9833	−0.9833	−0.9875
$p = -0.0013$	−0.9504	−0.9751	−0.9833	−0.9874	−0.9900
$p = -0.0005$	−0.9505	−0.9751	−0.9833	−0.9875	−0.9900
$p = -0.00025$	−0.9505	−0.9751	−0.9833	−0.9872	−0.9900
$p = -0.0001$	−0.9505	−0.9751	−0.9833	−0.9875	−0.9900

Table 2. Values of $\hat{\rho}$ for $p \to 0^-$, $\alpha = 25$, $\beta = 0.5$.

$p \to 0^-$	$n = 100$	$n = 10000$
$p = -0.5000$	−0.9785	−0.9772
$p = -0.4500$	−0.9799	−0.9802
$p = -0.3500$	−0.9826	−0.9816
$p = -0.3000$	−0.9839	−0.9830
$p = -0.2500$	−0.9851	−0.9842
$p = -0.2000$	−0.9863	−0.9855
$p = -0.1500$	−0.9874	−0.9867
$p = -0.1000$	−0.9888	−0.9879
$p = -0.0500$	−0.9905	−0.9890
$p = -0.0001$	−0.9905	−0.9900

where the exponent of the power transformation is set to the small negative value $p = -0.001$, r denotes sample correlation coefficient and s denotes sample standard deviation (see Appendix C for an outline of how inverse correlation can be used to eliminate bias in $\hat{x}_{c,t}$). The expectation is that if $\hat{x}_t$ are biased, then $\hat{x}_{c,t}$ will exhibit diminishing bias as $p \to 0^-$. If $\hat{x}_t$ are unbiased then $\omega = 1$ and the combined fitted values are just the original fitted values. The corresponding combined forecast values are $\hat{x}_{c,n}(\tau) = \omega \hat{x}_n(\tau) + (1 - \omega) \hat{x}'_n(\tau)$.

A shift parameter λ similar to that discussed by Box and Cox [32] is used to facilitate the power transformation and further improve the combined fitted mse. λ (determined by grid search) can be added to each value of x_t to obtain $z_t = x_t + \lambda$ prior to applying the power transformation and subtracted after conversion back to their original units, leaving the mean unchanged. While the data may be from stationary time series, they are of necessity a truncated sample. Any truncated data sample will fall short of the complete distributional properties of the population from which they are drawn, and therefore the property of stationary data. The antithetic time series is rewritten and computed from $\hat{x}'_t = \bar{x} + \hat{\rho}_{\hat{z}\hat{z}^p}\left(s_{\hat{z}}/s_{\hat{z}^p}\right)\left(\hat{z}^p_t - \overline{\hat{z}^p}\right)$, $t = 1, 2, 3, \cdots, n$, where

$p = -0.001$, $\hat{\rho}_{\hat{z}\hat{z}^p}$ is the sample correlation between $\hat{z}$ and $\hat{z}^p$, where $s_{\hat{z}}$ and $s_{\hat{z}^p}$ are sample standard deviations in $\hat{z}_t$ and $\hat{z}^p_t$ respectively, and where λ and ω are chosen so as to minimize the combined fitted mse for $\hat{x}_{c,t} = \omega \hat{x}_t + (1 - \omega) \hat{x}'_t$. The antithetic forecast values are computed from

$$\hat{x}'_n(\tau) = \bar{x} + \hat{\rho}_{\hat{z}\hat{z}^p}\left(s_{\hat{z}}/s_{\hat{z}^p}\right)\left(\hat{z}^p_n(\tau) - \overline{\hat{z}^p}\right), \quad \tau = 1, 2, 3, \cdots, N \ .$$

Of the terms p, λ, and ω, only p is unique to antithetic time series analysis, and it is not a fitted parameter. When implemented, p is actually a constant set to -0.001, an approximation of 0^-. Also, since $\hat{\rho}_{zz^p} = -1$, the transformation involving p is linear, and does not imply that the original model should have been non-linear. Like the use of λ here, it is common practice to apply various transformations such as logarithm, square root and Box and Cox [32] that add no new information, but make the data better conform to the assumptions of a postulated model. If there were no bias, or if antithetic combining did not reduce bias, then ω would simply be equal to one and the original postulated model only would apply.

Computer Simulation

To illustrate, consider a model fitted to computer simulated data based on stationary autoregressive processes, containing 1060 observations generated from $x_t = y_{t+250} + e_t$, $t = 1, 2, 3, \cdots, 1060$, where to avoid initialization problems, the first 250 values are dropped from $y_t = 0.8 y_{t-1} + e_t$, $t = 1, 2, 3, \cdots, 1310$, $y_0 = 0$ and $e_t \sim \Gamma(\alpha, \beta)$, $\beta = 0.6$, $\alpha = 5, 10, 15, 20, 25$ are obtained from MATLAB [30]. From the 1060 values, different models are fitted from the first 50, 51, $\cdots$, 60 values. Each model is used to forecast 1000 one-step-ahead forecast values corresponding to periods 51 - 1050, 52 - 1051, $\cdots$, 61 - 1060. This simple first order autoregressive model is chosen for its ease of understanding and transparency. It is perfect for the population from which the data are sampled. The sample sizes are typical of what can be expected in practice, and the outcomes from model fitting are subject to sampling bias.

The results are shown in **Table 3**. As α increases, the fitted mse's increase, indicative as expected, of the increase in the variance in the data. The combined fitted mse's are all lower than the original fitted mse's. The average gain is a reduction in fitted mse of 5.5%. This demonstrates that for a wide range of gamma distributions, combining antithetic fitted values can reduce the component of error that is due to systematic bias, leaving only random error. The fitted mse and 1000 period forecast horizon mse sensitivities to forecast origin (n) are shown in **Table 4**. As n increases from 51 to 60, the combined fitted mse's are lower than the original fitted mse's. The average gain is a reduction of 11.1%. The average gain in the combined forecast mse over the original forecast mse is a reduction of 6.9%. The forecast mse sensitivities to forecast horizon (N) are shown in **Table 5**. As N increases from 100 to 700, the combined forecast mse's are lower than the original forecast mse's. The average gain is a reduction of 6.1%.

7. Conclusion

The correlation between a gamma distributed random variable and its pth power was derived. It was proved that the correlation approaches minus one as p approaches zero from the left and the shape parameter approaches infinity. This counterintuitive result extends a previous finding of the similar result for lognormally distributed random variables. The gamma distribution was modified so as to emulate a range of distributions, showing that

Table 3. Fitted mean square error (mse) for gamma distributed autoregressive processes of length $n = 50$, $p = -0.001$, $\beta = 0.6$ and various values of α.

α	λ	ω	Fitted mse		
			Original	Combined	Reduction %
5	370	−46	1.377	1.209	12.2
10	520	−39	4.753	4.394	7.6
15	481	−17	6.892	6.732	2.3
20	575	−11	6.922	6.824	1.4
25	529	−43	9.686	9.304	3.9
		Average			5.5

Table 4. Fitted mean square error (mse) and one thousand period forecast mean square error (mse) for gamma distributed autoregressive processes of length n, $p = -0.001$, $\beta = 0.6$, $\alpha = 5$.

Forecast origin n	λ	ω	Fitted mse		Forecast mse	
			Original	Combined	Original	Combined
50	370	−46	1.377	1.209	6.158	5.760
51	411	−46	1.351	1.185	6.230	5.407
52	252	−46	1.354	1.202	6.493	5.641
53	446	−39	1.397	1.266	6.071	5.943
54	310	−55	1.371	1.160	6.070	5.611
55	261	−37	1.356	1.254	5.941	5.549
56	465	−41	1.345	1.189	6.108	5.720
57	400	−39	1.331	1.218	6.234	5.592
58	507	−36	1.312	1.191	6.142	6.224
59	405	−32	1.290	1.175	6.044	5.687
60	354	−62	1.434	1.207	5.600	5.300
	Average		1.356	1.205	6.099	5.676
	Combined reduction %		11.1		6.9	

Table 5. Forecast mean square error (mse) for gamma distributed autoregressive processes of length $n = 50$, $p = -0.001$, $\beta = 0.6$, $\alpha = 5$.

Forecast horizon N	Forecast mse		
	Original	Combined	Reduction%
100	5.973	5.581	6.6
150	6.529	6.059	7.2
200	5.426	5.220	3.8
250	6.559	6.152	6.2
300	6.455	6.181	4.2
350	6.390	6.051	5.3
400	5.982	5.630	5.9
450	5.906	5.554	6.0
500	5.972	5.607	6.1
550	6.114	5.699	6.8
600	6.063	5.616	7.4
650	5.818	5.401	7.2
700	5.632	5.243	6.9
Average	6.063	5.692	6.1

antithetic time series analysis can be generalized to all data distributions that are likely to occur in practice. The gamma distribution is unimodal. A suggestion for future research is to investigate the correlation between a random variable and its pth power when its distribution is multimodal. Another suggestion is to compare the effectiveness of the Hammersley and Morton [23] antithetic random numbers with antithetic random numbers constructed from the method described in this paper. Combining antithetic extrapolations can dynamically reduce bias due to model misspecifications such as serial correlation, non-normality or truncation of the distribution due to data sampling. Removing bias will eliminate the divergence between the extrapolated and actual values. In the particular case of climate models, removing bias can reveal the true long range climate dynamics. This will be most useful in models designed to investigate the phenomenon of global warming. Beyond the examples discussed here, antithetic combining has broad implications for mathematical statistics, statistical process control, engineering and scientific modeling.

Acknowledgements

References

[1] Ferrenberg, A.M., Lanau, D.P. and Wong, Y.J. (1992) Monte Carlo Simulations: Hidden Errors from Good Random Number Generators. *Physical Review Letters*, **69**, 3382-3384. http://dx.doi.org/10.1103/PhysRevLett.69.3382

[2] Griliches, Z. (1961) A Note on Serial Correlation Bias in Estimates of Distributed Lags. *Econometrica*, **29**, 65-73. http://dx.doi.org/10.2307/1907688

[3] Nerlove, M. (1958) Distributed Lags and Demand Analysis for Agricultural and Other Commodities. U.S.D.A Agricultural Handbook No. 141, Washington.

[4] Koyck, L.M. (1954) Distributed Lags and Investment Analysis. North-Holland Publishing Co., Amsterdam.

[5] Klein, L.R. (1958) The Estimation of Distributed Lags. *Econometrica*, **26**, 553-565. http://dx.doi.org/10.2307/1907516

[6] Fuller, W.A. and Hasza, D.P. (1981) Properties of Predictors from Autoregressive Time Series. *Journal of the American Statistical Association*, **76**, 155-161. http://dx.doi.org/10.1080/01621459.1981.10477622

[7] Dufour, J. (1985) Unbiasedness of Predictions from Estimated Vector Autoregressions. *Econometric Theory*, **1**, 381-402. http://dx.doi.org/10.1017/S0266466600011270

[8] Chandan, S. and Jones, P. (2005) Asymptotic Bias in the Linear Mixed Effects Model under Non-Ignorable Missing Data Mechanisms. *Journal of the Royal Statistical Society*: *Series B*, **67**, 167-182. http://dx.doi.org/10.1111/j.1467-9868.2005.00494.x

[9] Li, B., Nychka, D.W. and Ammann, C.M. (2010) The Value of Multiproxy Reconstruction of Past Climate. *Journal of the American Statistical Association*, **105**, 883-911. http://dx.doi.org/10.1198/jasa.2010.ap09379

[10] Bunn, D.W. (1979) The Synthesis of Predictive Models in Marketing Research. *Journal of Marketing Research*, **16**, 280-283. http://dx.doi.org/10.2307/3150692

[11] Diebold, F.X. (1989) Forecast Combination and Encompassing: Reconciling Two Divergent Literatures. *International Journal of Forecasting*, **5**, 589-592. http://dx.doi.org/10.1016/0169-2070(89)90014-9

[12] Clemen, R.T. (1989) Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, **5**, 559-583. http://dx.doi.org/10.1016/0169-2070(89)90012-5

[13] Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982) The Accuracy of Extrapolation (Times Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, **1**, 111-153. http://dx.doi.org/10.1002/for.3980010202

[14] Winkler, R.L. (1989) Combining Forecasts: A Philosophical Basis and Some Current Issues. *International Journal of Forecasting*, **5**, 605-609. http://dx.doi.org/10.1016/0169-2070(89)90018-6

[15] Hendry, D.F. and Mizon, G.E. (1978) Serial Correlation as a Convenient Simplification, Not a Nuisance: A Commentary on a Study of the Demand for Money by the Bank of England. *The Economic Journal*, **88**, 549-563. http://dx.doi.org/10.2307/2232053

[16] Hendry, D.F. (1976) The Structure of Simultaneous Equations Estimators. *Journal of Econometrics*, **4**, 551-588. http://dx.doi.org/10.1016/0304-4076(76)90017-8

[17] Mizon, G.E. (1977) Model Selection Procedures. In: Artis, M.J. and Nobay, A.D., Eds., *Studies in Modern Economic Analysis*, Basil Blackwell, Oxford.

[18] Pindyck, R.S. and Rubinfeld, D.L. (1976) Econometric Models and Economic Forecasts. McGraw-Hill, New York.

[19] Durbin, J. and Watson, G.S. (1950) Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, **37**, 409-428.

[20] Durbin, J. (1970) Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables. *Econometrica*, **38**, 410-421. http://dx.doi.org/10.2307/1909547

[21] Osborn, D.R. (1976) Maximum Likelihood Estimation of Moving Average Processes. *Journal of Economic and Social Measurement*, **5**, 75-87.

[22] Espasa, D. (1977) The Spectral Maximum Likelihood Estimation of Econometric Models with Stationary Errors. 3, Applied Statistics and Economics Series. Vanderhoeck and Ruprecht, Gottingen.

[23] Hammersley, J.M. and Morton, K.W. (1956) A New Monte Carlo Technique: Antithetic Variates. *Mathematical Proceedings of the Cambridge Philosophical Society*, **52**, 449-475. http://dx.doi.org/10.1017/S0305004100031455

[24] Kleijnen, J.P.C. (1975) Antithetic Variates, Common Random Numbers and Optimal Computer Time Allocation in Simulations. *Management Science*, **21**, 1176-1185. http://dx.doi.org/10.1287/mnsc.21.10.1176

[25] Ridley, A.D. (1999) Optimal Antithetic Weights for Lognormal Time Series Forecasting. *Computers & Operations Research*, **26**, 189-209. http://dx.doi.org/10.1016/s0305-0548(98)00058-6

[26] Ridley, A.D. (1995) Combining Global Antithetic Forecasts. *International Transactions in Operational Research*, **4**,387-398. http://dx.doi.org/10.1111/j.1475-3995.1995.tb00030.x

[27] Ridley, A.D. (1997) Optimal Weights for Combining Antithetic Forecasts. *Computers & Industrial Engineering*, **2**, 371-381. http://dx.doi.org/10.1016/s0360-8352(96)00296-3

[28] Ridley, A.D. and Ngnepieba, P. (2014) Antithetic Time Series Analysis and the CompanyX Data. *Journal of the Royal Statistical Society*: *Series A*, **177**, 83-94. http://dx.doi.org/10.1111/j.1467-985x.2012.12001.x

[29] Ridley, A.D., Ngnepieba, P. and Duke, D. (2013) Parameter Optimization for Combining Lognormal Antithetic Time Series. *European Journal of Mathematical Sciences*, **2**, 235-245.

[30] MATLAB (2008) Application Program Interface Reference, Version 8. The Math Works, Inc.

[31] Hogg, R.V. and Ledolter, J. (2010) Applied Statistics for Engineers and Physical Scientists. 3rd Edition, Prentice Hall, Upper Saddle River, 174.

[32] Box, G.E.P. and Cox, D.R. (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society*: *Series B*, **26**, 211-252.

[33] Abramowitz, M. and Stegun, I.A. (1964) Handbook of Mathematical Functions. Dover Publications, New York, 260 p.

[34] Bernado, J.M. (1976) Algorithm AS 103: Psi (Digamma) Function. *Journal of the Royal Statistical Society*: *Series C* (*Applied Statistics*), **25**, 315-317.

[35] Fuller, W.A. (1996) Introduction to Statistical Times Series. Wiley, New York.

Appendix

Appendix A: *p*th Order Moment for the Gamma Distribution

The *p*th moment of the gamma distribution is derived as follows:

$$\mathbb{E}\left[X_t^p\right] = \int_0^\infty x^p \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+p-1} e^{-\frac{x}{\beta}} dx. \tag{A.1}$$

Multiplying and dividing by $\beta^{\alpha+p} \Gamma(\alpha+p)$, Equation (A.1) becomes

$$\begin{aligned}
\mathbb{E}\left[X_t^p\right] &= \frac{\beta^{\alpha+p}\Gamma(\alpha+p)}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{1}{\beta^{\alpha+p}\Gamma(\alpha+p)} x^{\alpha+p-1} e^{-\frac{x}{\beta}} dx \\
&= \beta^p \frac{\Gamma(\alpha+p)}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\beta^{\alpha+p}\Gamma(\alpha+p)} x^{\alpha+p-1} e^{-\frac{x}{\beta}} dx.
\end{aligned} \tag{A.2}$$

Since $\dfrac{1}{\beta^{\alpha+p}\Gamma(\alpha+p)} x^{\alpha+p-1} e^{-\frac{x}{\beta}} = f(x;\alpha+p,\beta)$, is the pdf for a gamma function with the parameter $\alpha+p$,

$$\int_0^\infty \frac{1}{\beta^{\alpha+p}\Gamma(\alpha+p)} x^{\alpha+p-1} e^{-\frac{x}{\beta}} dx = 1,$$

and equation (A.2) becomes

$$\mathbb{E}\left[X_t^p\right] = \beta^p \frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}.$$

Appendix B: Proof of the Antithetic Gamma Variables Theorem

By applying the Taylor expansion around $p = 0$ to Equation (8) we have

$$\Gamma(\alpha+p) = \Gamma(\alpha) + p\Gamma'(\alpha) + \frac{p^2}{2}\Gamma''(\alpha) + O(p^3), \tag{B.1}$$

where Γ', Γ'' are the first and second derivatives of Γ and $O(p^3)$ represents the remainder.

$$\begin{aligned}
\left(\Gamma(\alpha+p)\right)^2 &= \left(\Gamma(\alpha)\right)^2 + p^2\left(\Gamma'(\alpha)\right)^2 + 2p\Gamma(\alpha)\Gamma'(\alpha) \\
&\quad + p^2\Gamma(\alpha)\Gamma''(\alpha) + p^3\Gamma'(\alpha)\Gamma''(\alpha) + O(p^3)
\end{aligned} \tag{B.2}$$

$$\Gamma(\alpha)\Gamma(\alpha+2p) = \left(\Gamma(\alpha)\right)^2 + 2p\Gamma(\alpha)\Gamma'(\alpha) + 2p^2\Gamma(\alpha)\Gamma''(\alpha) + O(p^3) \tag{B.3}$$

The combination of equations (B.1)-(B.3) reduces Equation (8) to

$$\begin{aligned}
\rho &= \frac{p\left(\Gamma(\alpha) + p\Gamma'(\alpha) + \frac{p^2}{2}\Gamma''(\alpha) + O(p^3)\right)}{\left\{p^2\alpha\left(\Gamma(\alpha)\Gamma''(\alpha) - \left(\Gamma'(\alpha)\right)^2 - p\Gamma'(\alpha)\Gamma''(\alpha) + O(p)\right)\right\}^{\frac{1}{2}}} \\
&= \frac{p\left(\Gamma(\alpha) + p\Gamma'(\alpha) + \frac{p^2}{2}\Gamma''(\alpha) + O(p^3)\right)}{|p|\left\{\alpha\left(\Gamma(\alpha)\Gamma''(\alpha) - \left(\Gamma'(\alpha)\right)^2 - p\Gamma'(\alpha)\Gamma''(\alpha) + O(p)\right)\right\}^{\frac{1}{2}}}.
\end{aligned}$$

Therefore,

$$\lim_{p \to 0^-} \rho = -\frac{\Gamma(\alpha)}{\left\{\alpha\left(\Gamma(\alpha)\Gamma''(\alpha)-(\Gamma'(\alpha))^2\right)\right\}^{\frac{1}{2}}}. \tag{B.4}$$

By using the polygamma function (see Abramowitz and Stegun [33])

$$\Psi^{(1)}(\alpha) = \frac{d}{d\alpha}\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{\Gamma(\alpha)\Gamma''(\alpha)-(\Gamma'(\alpha))^2}{(\Gamma(\alpha))^2},$$

Equation (B.4) is transformed into

$$\lim_{p \to 0^-} \rho = -\frac{1}{\left\{\alpha\Psi^{(1)}(\alpha)\right\}^{\frac{1}{2}}}. \tag{B.5}$$

The digamma function for real $\alpha > 0$, as $\alpha \to \infty$ is

$$\Psi(\alpha) = \log(\alpha) - \frac{1}{2\alpha} - \frac{1}{12\alpha^2} + \frac{1}{120\alpha^4} - \frac{1}{252\alpha^6} + O\left(\left(\frac{1}{\alpha}\right)^8\right)$$

(see also Bernado [34]).

Its derivative is the polygamma function

$$\Psi^{(1)}(\alpha) = \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \frac{1}{6\alpha^3} - \frac{1}{30\alpha^5} + \frac{1}{42\alpha^7} + O\left(\left(\frac{1}{\alpha}\right)^9\right).$$

And,

$$\alpha\Psi^{(1)}(\alpha) = 1 + \frac{1}{2\alpha} + \frac{1}{6\alpha^2} - \frac{1}{30\alpha^4} + \frac{1}{42\alpha^6} + O\left(\left(\frac{1}{\alpha}\right)^8\right).$$

From which, $\lim_{\alpha \to \infty} \alpha\Psi^{(1)}(\alpha) = 1$. Finally, the limit in Equation (B.5) is

$$\lim_{p \to 0^-, \alpha \to \infty} \rho = \lim_{\alpha \to \infty}\left(\lim_{p \to 0^-} \rho\right) = \lim_{\alpha \to \infty}\left[-\frac{1}{\left\{\alpha\Psi^{(1)}(\alpha)\right\}^{\frac{1}{2}}}\right] = -1.$$

Appendix C: Inverse Correlation and Bias Elimination

Consider a gamma distributed time series X_t with a large shape parameter from which x_t are observations. We have shown that for very small negative p, x_t and x_t^p are nearly perfectly correlated, albeit negatively, so 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 as follows:

$$x_t = c_0 + c_1 x_t^p + \varepsilon_t,$$

where ε_t is an error term.

As p approached zero from the left, near perfect correlation between x_t and x_t^p ensures that the error term becomes negligible, and a near perfect estimate is obtained from

$$x_t' = c_0 + c_1 x_t^p. \tag{C.1}$$

Now, suppose that

$$x_t = f(x_{t-1}) + \epsilon_t, t = 2, \cdots, n,$$

is a time series model. If there is any bias due either to serial correlation in ϵ_t or sampling error in estimating the model, the estimated model will be biased such that $\text{Cov}(\epsilon_t, x_{t-1}) \neq 0$. The estimated parameters of this model will be biased. That is unavoidable. Therefore, any estimate $\hat{x}_t$ of x_t from this model will also be

biased.

To remove this bias, we power transform $\hat{x}_t$ to obtain $\hat{x}_t^p$. Then, we use Equation (C.1) to convert $\hat{x}_t^p$ back to the original units of $\hat{x}_t$. Hence

$$\hat{x}_t' = \hat{c}_0 + \hat{c}_1 \hat{x}_t^p,$$

where $\hat{c}_0$ and $\hat{c}_1$ are least squares estimates obtained from the regression of $\hat{x}_t$ on $\hat{x}_t^p$, and the error approaches 0.

$$\hat{x}_t' = \hat{x} - \hat{c}_1 \overline{x}^p + \hat{c}_1 \hat{x}_t^p = \overline{x} + \hat{c}_1 \left(\hat{x}_t^p - \overline{\hat{x}^p} \right).$$

Denoting sample standard deviation by s and correlation coefficient by $\hat{\rho}_{\hat{x}\hat{x}^p}$,

$$\hat{x}_t' = \overline{x} + \hat{\rho}_{\hat{x}\hat{x}^p} \left(s_{\hat{x}} / s_{\hat{x}^p} \right) \left(\hat{x}_t^p - \overline{\hat{x}^p} \right)$$

(see also the Ridley [25] antithetic fitted function theorem).

Both estimates $\hat{x}_t$ and $\hat{x}_t'$ contain errors. These errors contain two components. One component is purely random and one component is bias. Combining the estimates dynamically cancels the bias components, leaving only the purely random components. See Appendix D for the proof of how this can occur. The combining weights discussed in Appendix D are theoretical, expressed in terms of errors that are unknown and un-observable, so we must rely on the approximation as follows. The combined estimate $\hat{x}_{c,t}$ is obtained from

$$\hat{x}_{c,t} = \omega \hat{x}_t + (1 - \omega) \hat{x}_t',$$

where $-\infty < \omega < \infty$, and the value of ω is chosen so as to minimize the mse

$$\frac{\sum_{t=2}^{n} \left(x_t - \hat{x}_{c,t} \right)^2}{n-1}.$$

Consider the error in $\hat{x}_{c,t}$, $\hat{e}_t = \hat{x}_{c,t} - x_t$. Then $\sum_{t=2}^{n} \hat{e}_t^2 = \sum_{t=2}^{n} \left\{ \omega \hat{x}_t + (1-\omega) \hat{x}_t' - x_t \right\}^2$. Differentiating with respect to ω, $\mathrm{d} \sum_{t=2}^{n} \hat{e}_t^2 / \mathrm{d}\omega = \sum_{t=2}^{n} 2 \left\{ \omega \left(\hat{x}_t - \hat{x}_t' \right) + \hat{x}_t' - x_t \right\} \left(\hat{x}_t - \hat{x}_t' \right)$. Setting the derivative to zero and solving for ω, $\omega^\star = \sum_{t=2}^{n} \left(x_t - \hat{x}_t' \right) \left(\hat{x}_t - \hat{x}_t' \right) / \sum_{t=2}^{n} \left(\hat{x}_t - \hat{x}_t' \right)^2$. This optimal $\omega^\star$ yields the minimum mse, because $\mathrm{d}^2 \sum_{t=2}^{n} \hat{e}_t^2 / \mathrm{d}\omega^2 = 2 \sum_{t=2}^{n} \left(\hat{x}_t - \hat{x}_t' \right)^2 = 0$ iff $\hat{x}_t' = \hat{x}_t \forall t$, in which case $\omega = 1$, and $\mathrm{d}^2 \sum_{t=2}^{n} \hat{e}_t^2 / \mathrm{d}\omega^2 > 0$ otherwise.

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

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

Step 2: Set $p = -0.001$.

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

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

Likewise, the unbiased combined estimate of a future value at time τ is obtained from $\hat{x}_{c,n}(\tau) = \omega \hat{x}_n(\tau) + (1-\omega) \hat{x}_n'(\tau)$.

Appendix D: Antithetic Fitted Error Variance Reduction

Consider a gamma distributed time series X_t with a large shape parameter. Next, consider a minimum mean square error fitted value obtained from a stationary first-autoregressive process $X_t = \mu_X + \phi(X_{t-1} - \mu_X) + \epsilon_t$, given by $\hat{X}_t = \hat{\mu}_X + \hat{\phi}(X_{t-1} - \hat{\mu}_X)$ where $\mu_X = \mathbb{E}[X_t]$ and $\hat{\mu}_X$ and $\hat{\phi}$ are least-squares estimates of μ and ϕ, respectively, such that

$$\hat{\phi} = \sum_{t=2}^{n} \left(X_t - \mu_X \right) \left(X_{t-1} - \mu_X \right) \bigg/ \sum_{t=2}^{n} \left(X_{t-1} - \mu_X \right)^2,$$

$$\hat{\phi} = \sum_{t=2}^{n} \left\{ \mu_X + \phi(X_{t-1} - \mu_X) + \epsilon_t - \mu_X \right\} \left(X_{t-1} - \mu_X \right) \bigg/ \sum_{t=2}^{n} \left(X_{t-1} - \mu_X \right)^2,$$

$$\hat{\phi} = \left\{ \phi \sum_{t=2}^{n} \left(X_{t-1} - \mu_X \right)^2 + \sum_{t=2}^{n} \epsilon_t \left(X_{t-1} - \mu_X \right) \right\} \bigg/ \sum_{t=2}^{n} \left(X_{t-1} - \mu_X \right)^2,$$

$$\hat{\phi} = \phi + \sum_{t=2}^{n} \epsilon_t \left(X_{t-1} - \mu_X \right) \bigg/ \sum_{t=2}^{n} \left(X_{t-1} - \mu_X \right)^2.$$

Therefore, as $n \to \infty$, and since X_t is stationary so that $\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1})$, and since the errors are serially correlated so that $\mathrm{Cov}(\epsilon_t, X_{t-1}) \neq 0$,

$$\hat{\phi} = \phi + \mathrm{Cov}(\epsilon_t, X_{t-1}) / \mathrm{Var}(X_t) \tag{D.1}$$

(see also Fuller [35], p. 404). Consider $\hat{X}_t$ as an estimate of X_t. From (D.1), and given that the time series is stationary, then as $n \to \infty$ and $\hat{\mu}_X \to \mu_X$,

$$\hat{X}_t = \mu_X + \phi \left(X_{t-1} - \mu_X \right) + \left\{ \mathrm{Cov}(\epsilon_t, X_{t-1}) / \mathrm{Var}(X_t) \right\} \left(X_{t-1} - \mu_X \right) = \mathbb{E}[X_t] + u_t,$$

where

$$u_t = \left\{ \mathrm{Cov}(\epsilon_t, X_{t-1}) / \mathrm{Var}(X_t) \right\} \left(X_{t-1} - \mu_X \right) \tag{D.2}$$

due only to errors resulting from serial correlation. Therefore,

$$\mathrm{Var}(u_t) = \left\{ \mathrm{Cov}(\epsilon_t, X_{t-1}) / \mathrm{Var}(X_t) \right\}^2 \mathrm{Var}(X_t). \tag{D.3}$$

Next, consider another fitted value $\hat{X}'_t$, obtained from the linear projection of the asymptotically antithetic series X_t^p on X_t, without the introduction of any new error,

$$\hat{X}'_t = \lim_{\sigma \to 0, p \to 0^-} \mathbb{E}[X_t] + \rho \left\{ \mathrm{Var}(X_t) / \mathrm{Var}(X_t^p) \right\}^{1/2} \left\{ \hat{X}_t^p - \mathbb{E}[X_t^p] \right\}.$$

Substituting for $\hat{X}_t$

$$\begin{aligned}
\hat{X}'_t &= \lim_{\sigma \to 0, p \to 0^-} \mathbb{E}[X_t] + \rho \left\{ \mathrm{Var}(X_t) / \mathrm{Var}(X_t^p) \right\}^{1/2} \left\{ \left(\hat{\mu}_X + \hat{\phi} \left(X_{t-1} - \hat{\mu}_X \right) \right)^p - \mathbb{E}[X_t^p] \right\} \\
&= \mathbb{E}[X_t] + u'_t,
\end{aligned} \tag{D.4}$$

where ρ is the correlation between X_t and X_t^p, and

$$u'_t = \lim_{\sigma \to 0, p \to 0^-} \rho \left\{ \mathrm{Var}(X_t) / \mathrm{Var}(X_t^p) \right\}^{1/2} \left\{ \left(\hat{\mu}_X + \hat{\phi} \left(X_{t-1} - \hat{\mu}_X \right) \right)^p - \mathbb{E}[X_t^p] \right\} \tag{D.5}$$

is the antithetic error due to the serial correlation, but corresponding to $\hat{X}'_t$.

The expansion of $\left(\hat{\mu}_X + \hat{\phi} \left(X_{t-1} - \hat{\mu}_X \right) \right)^p = \left(\hat{\mu}_X + \hat{\phi} X_{t-1} - \hat{\phi} \hat{\mu}_X \right)^p$ will

contain the constant $\left(\hat{\mu}_X - \hat{\phi} \hat{\mu}_X \right)^p$, the product of p and some function of X_{t-1} and $\left(\hat{\phi} X_{t-1} \right)^p$ as follows:

$\left(\hat{\mu}_X + \hat{\phi} X_{t-1} - \hat{\phi} \hat{\mu}_X \right)^p = \left(\hat{\mu}_X - \hat{\phi} \hat{\mu}_X \right)^p + p f \left(X_{t-1} \right) + \hat{\phi}^p X_{t-1}^p \to \left(\hat{\mu}_X - \hat{\phi} \hat{\mu}_X \right)^p + \left(\hat{\phi} X_{t-1} \right)^p$ as $p \to 0$. Substituting into Equation (D.5),

$$u'_t = \lim_{\sigma \to 0, p \to 0^-} \rho \left\{ \mathrm{Var}(X_t) / \mathrm{Var}(X_t^p) \right\}^{1/2} \left\{ \left(\hat{\mu}_X - \hat{\phi} \hat{\mu}_X \right)^p + \hat{\phi}^p X_{t-1}^p - \mathbb{E}[X_t^p] \right\}. \tag{D.6}$$

Now

$$\mathrm{Var}\left\{ \omega u_t + (1-\omega) u'_t \right\} = \omega^2 \mathrm{Var}(u_t) + (1-\omega)^2 \mathrm{Var}(u'_t) + 2\omega(1-\omega) \mathrm{Cov}(u_t, u'_t).$$

Substituting from Equation (D.2) and (D.6) and since $\hat{\mu}_X$ and $\hat{\phi}$ are fixed for the data and model,

$$\begin{aligned}
\mathrm{Var}\left\{ \omega u_t + (1-\omega) u'_t \right\} = \lim_{\sigma \to 0, p \to 0^-} &\; \omega^2 \mathrm{Var}(u_t) + (1-\omega)^2 \rho^2 \left\{ \mathrm{Var}(X_t) / \mathrm{Var}(X_t^p) \right\} \\
&\times \hat{\phi}^{2p} \mathrm{Var}(X_{t-1}^p) + 2\omega(1-\omega) \left\{ \mathrm{Cov}(\epsilon_t, X_{t-1}) / \mathrm{Var}(X_t) \rho \right\} \\
&\times \rho \left\{ \mathrm{Var}(X_t) / \mathrm{Var}(X_t^p) \right\}^{1/2} \times \hat{\phi}^p \mathrm{Cov}(X_{t-1}, X_{t-1}^p).
\end{aligned}$$

Substituting for $\mathrm{Cov}\left(X_{t-1}, X_{t-1}^{p}\right)$ and $\mathrm{Var}\left(X_{t-1}\right)$

$$\mathrm{Var}\left\{\omega u_{t}+(1-\omega)u_{t}'\right\} = \lim_{\sigma \to 0, p \to 0^{-}} \omega^{2}\mathrm{Var}\left(u_{t}\right)+(1-\omega)^{2}\rho^{2}\hat{\phi}^{2p}\mathrm{Var}\left(X_{t}\right)$$

$$+2\omega(1-\omega)\left\{\mathrm{Cov}\left(\epsilon_{t}, X_{t-1}\right)/\mathrm{Var}\left(X_{t}\right)\right\}\rho\left\{\mathrm{Var}\left(X_{t}\right)/\mathrm{Var}\left(X_{t}^{p}\right)\right\}^{1/2}\hat{\phi}^{p}\rho\left\{\mathrm{Var}\left(X_{t}\right)\mathrm{Var}\left(X_{t}^{p}\right)\right\}^{1/2}.$$

Substituting from (D.3) and factoring out $\mathrm{Var}\left(X_{t}\right)$

$$\mathrm{Var}\left\{\omega u_{t}+(1-\omega)u_{t}'\right\} = \lim_{\sigma \to 0, p \to 0^{-}}\left[\omega^{2}\left\{\mathrm{Cov}\left(\epsilon_{t}, X_{t-1}\right)/\mathrm{Var}\left(X_{t}\right)\right\}^{2}+(1-\omega)^{2}\rho^{2}\hat{\phi}^{2p}\right.$$

$$\left.+2\omega(1-\omega)\left\{\mathrm{Cov}\left(\epsilon_{t}, X_{t-1}\right)/\mathrm{Var}\left(X_{t}\right)\right\}\rho^{2}\hat{\phi}^{p}\right]\mathrm{Var}\left(X_{t}\right)$$

and since $\rho^{2} \to 1$ (see Appendix B), and $\hat{\phi}^{p}$ and $\hat{\phi}^{2p} \to 1$ as $p \to 0^{-}$, then

$\mathrm{Var}\left\{\omega u_{t}+(1-\omega)u_{t}'\right\} = \left\{\omega\mathrm{Cov}\left(\epsilon_{t}, X_{t-1}\right)/\mathrm{Var}\left(X_{t}\right)+1-\omega\right\}^{2}\mathrm{Var}\left(X_{t}\right)$ from which we see that there are many ways in which the combined error variance can be less than the original error variance in Equation (D.3). In particular when $\omega = \left\{1-\mathrm{Cov}\left(\epsilon_{t}, X_{t-1}\right)/\mathrm{Var}\left(X_{t}\right)\right\}^{-1}$, the error variance due to systematic serial correlation vanishes. The only error variance remaining will be due purely to random error unexplained by the original model.