OPERATIONS RESEARCH AND DECISIONS

No. 3 2016

DOI: 10.5277/ord160304

Dennis RIDLEY[1]

ADVANCES IN ANTITHETIC TIME SERIES ANALYSIS. SEPARATING FACT FROM ARTIFACT

The problem of biased time series mathematical model parameter estimates is well known to be insurmountable. When used to predict future values by extrapolation, even a de minimis bias will eventually grow into a large bias, with misleading results. This paper elucidates how combining antithetic time series' solves this baffling problem of bias in the fitted and forecast values by dynamic bias cancellation. Instead of growing to infinity, the average error can converge to a constant.

Keywords: *combining, antithetic, time series, bias correction, serial correlation*

1. Introduction

Time series mathematical models are subject to bias in parameter estimates. The bias is an artifact of the model. It is an obstacle to accurate extrapolation. This is particularly true in the case of economic science where long range extrapolation is required and even more so in climate science time series analysis of global warming. Both effects are global and occur over a very long time period. Therefore, the elimination of bias is paramount.

A time series is the name given to a sequence of numbers that occur over time. If the sequence is correlated, that is, if current values are related to past values, then future values can be predicted by extrapolation from a record of historical values. A time series model is a mathematical model for fitting historical time series data. The model is comprised of a variable that represents the value of the time series at any point in time and one or more lagged variables and corresponding coefficients. The coefficients are called parameters. The fitting process is the estimation of the values of the parameters. The

[1]SBI, Florida A&M University and Department of Scientific Computing, Florida State University, Tallahassee, Fl., USA, e-mail address: dridley@fsu.edu

model can then be used to predict future values by extrapolation. This feature can be exploited for the purpose of prediction in engineering, economics, medicine and social, chemical and physical sciences. However, there is a well-known insurmountable obstacle to realizing the true value of time series models. Time series mathematical models are subject to bias in the estimates of the parameters. The bias is an artifact of the model. It is an obstacle to accurate extrapolation. This is particularly true in the case of econometrics where long range extrapolation is required and even more so in climate science analysis of global warming. Combining antithetic time series (not to be confused with antithetic variates [9, 12]) is a recent discovery that permits unbiased model fitting and extrapolation in the presence of bias in parameter estimates. Combining antithetic time series' yields unbiased fitted and forecast values even when the estimates of the model parameters are biased.

2. Methodology

2.1. Deterministic versus probabilistic models

Mathematical models have a long history of service to mathematics, engineering, econometrics, medicine and social, chemical and physical science. There are two types of models, deterministic models and probabilistic ones. Probabilistic models are also referred to as stochastic. Deterministic models assume that model variables can be determined exactly with no uncertainty. Therefore, they are simply specified. They are relatively easy to interpret and appear to work well when their use is limited to their intended purpose and the model formulation does in fact represent the true relationships between the variables represented.

Probabilistic models include random variables, the values of which cannot be determined exactly. The outcomes of probabilistic models are understood to be uncertain. Random variables are not entirely arbitrary. They can follow a pattern, resulting in a well-defined probability distribution, with constant parameters such as mean and variance. The objective of probabilistic models is to estimate means and variances from which statements of probability can be made. An example of such a model is the well-known cross sectional regression model. In that model, a dependent variable (regressand) such as volume is regressed on independent variables (regressors) such as temperature and pressure. The objective is to estimate two parameters that relate temperature and pressure to volume. Any values of the regressand that are not reproducible by the model are the model errors. For the estimates of the parameters to be unbiased, inter alia, the observed values of the dependent variable must be independent of each other and the errors must be independent of the regressors. Physical phenomena of this type are described well by mathematical models.

Described as an enigma [6], despite themselves not being physical, mathematical models may be used to make close and accurate numerical connections between the elements of complex physical systems. However, the smallest deviation from the prescribed assumptions can lead to insurmountable problems. Some such problems are not curable by simply employing more data. For, example, it is well known [8] that when the above mentioned independence requirement (of the Gauss–Markov theorem) is violated, the parameter estimates of the model will be biased. This is unavoidable in the case of longitudinal time series models. The reason is because in time series models, the dependent variable is regressed on past values of the same variable. That is, they are autoregressive. In order to have any predictive value, the values of the dependent variable must depend on past values. That is, the observations are not independent. The model errors can become correlated with past values of the time series variable.

The biases in the model parameter estimates are aptly described as artifacts of the model. They are not real and they do not exist anywhere in the real world or in the physical system under study. Of course, fitted values and predicted values obtained by extrapolation from such a biased model will also be biased. In multiperiod forecasting by extrapolation, the bias will accumulate and the forecast and actual values will diverge. In the case of climate science, where a time series model is used to predict future temperatures by extrapolation, positive and negative biases will over predict warming and cooling, respectively. These scenarios are unrelated to the real world physical system.

Of course, bias in the time series model can be avoided by selecting a time series whose values are independent. Unfortunately, such a model could not possibly have any predictive value. In that case, mathematics has greatly diminished value or no value at all. Worse still, biased estimation can be misleading, so much so that analysis may be better off without it. So, we see that while mathematics can model complex connections of cross sectional relationships, when called on to model longitudinal connections, the rules of mathematics continue to work, but with respect to bias, mathematics exhibits some complexities of its own.

2.2. Physical science versus time series models

In a repeatable physical science experiment where each outcome yields a sample, each sample may be designed such that there is an equal opportunity of including any data point. In time series, the researcher gets one shot at a sample. Unless the sample size is infinite, there is little chance that it is perfectly representative of the population. There is no repetition that would conceptually average out bias. There is no meaning in talking about an unbiased sample from a time series. The real problem, of course, is bias in parameter estimates.

2.3. The sources, mechanism and impact of bias

Let us consider the simplest time series model. A practical model may be more complicated, but the issue of bias will apply similarly. Suppose a discrete stationary stochastic autoregressive time series model $X_t = \phi X_{t-1} + \varepsilon_t$, $t = 2, 3, ..., n$, where ϕ is a parameter of unknown value and ε_t is an unobservable random error, is fitted to a stationary time series of n observations: X_t, $t = 1, 2, 3, \ldots, n$, where t represents time. When X_t are obtained by measurement, randomness is defined as an ensemble of possibilities, each of which is assigned a probability. When X_t are obtained by simulation, the series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself (the complexity of the series of digits is approximately equal to its size in bits [1, 13].

If future values $X_n(\tau)$, $\tau = 1, 2, 3, \ldots, N$, are to be predicted by extrapolation at time $t = n$, then the past values X_t, $t \leq n$, must be serially correlated. The estimation of the parameter ϕ in the model requires that the data be stationary. If a normal distribution is assumed, stationarity requires that the data be defined completely by a constant mean and variance. There are various ways in which this assumption may not apply. One way is if the population is not normally distributed to begin with. Many data assumed to be normally distributed are better described by a lognormal distribution. For example, in the case of economic data such as the production or sales of a product, while the minimum value is zero, the upper limit can be very large. Another way is if the population is normally distributed but sampling truncates the data such that the sample does not appear to the model fitting process as being normally distributed. Yet another way is if other relevant variables exist but are missing because they are not measurable or are otherwise unavailable. Any one of these departures from the model requirement will create a sequential pattern in the errors ε_t. That is, the errors are serially correlated. Another way of stating this is that the correlation between the current and past errors is not zero (Corr(ε_t, ε_{t-j}) $\neq 0$ for some $j \neq 0$). From the above model, X_t and ε_t are correlated. Likewise, so are X_{t-1} and ε_{t-1}. That is, Corr(X_{t-1}, ε_{t-1}) $\neq 0$. Therefore, if ε_t is correlated with ε_{t-1}, then ε_t is correlated with X_{t-1}. That is, Corr(X_{t-1}, ε_t) = 0. This is a clear violation of the requirements of any regression model.

Unfortunately, there is no way to fit an unbiased AR(p) model to serially correlated data wherein Corr(X_{t-1}, ε_t) $\neq 0$ and $p \geq 1$. The Gauss–Markov theorem that requires independence will be violated. This was demonstrated by Griliches [8] and discussed by other researchers [4, 11, 16]. The conventional wisdom is that this and many other problems are easily overcome by simply obtaining more data. Of course, as a practical matter, the size of the data sample may be limited. Notwithstanding the possibilities of additional sampling, it turns out that more data will not solve the problem. As $n \to \infty$, the variance of the estimator of ϕ decreases but the bias does not decrease to zero. It decreases to a finite value greater than zero so that the estimate is more precise. That is, the estimate becomes more and more precisely wrong. This idiopathic bias is an artifact that is introduced by the mathematical model. Such a biased model will generate biased predictions that are obtained by extrapolation. Obviously, if the past values are independent, no future values can be predicted by extrapolation. The future cannot be connected to the past. This is a hopeless paradox.

2.4. Consequences for forecasting

The persistent occurrence of bias has led to the conventional wisdom that the farther in time that a model is used to predict by extrapolation, the greater the error will be [14]. The forecasts and actual values will diverge over time. An etiology of the forecast error gives the appearance of an emergent property that cannot be reduced to the elements of the model. While this may be intuitive, it is only true if there is bias. If the bias can be eliminated, then at least in theory, the forecast mean square error (MSE) will be constant into the future.

A similar problem would exist in a cross sectional model if the observations were not serially independent. However, in a cross sectional model, there is no attempt at finding longitudinal connections between the future and the past. The connections sought are between a dependent variable and one or more independent variables. Excellent connections between independent observations are possible.

There are negative implications for biased extrapolations in engineering, science, medicine and economics. Manufacturing quality control data are correlated, economic data are more correlated and biomedical data are highly correlated. The computer program MyPulse [17] was developed expressly for the purpose of separating systematic internal biological effects from random external environmental effects in biomedical data. Bias poses special problems for long range extrapolation. For example, the investigation of global warming requires long range forecasts. Even a small bias will eventually grow into a large bias, with misleading results. Missing data are also a source of bias [15]. So, better climate models, more data and better data are important. However, such efforts are futile if there is inherent bias in the mathematical model [14].

It is reasonable to consider man made time series (outcomes from engineering, scientific, and computer simulation experiments, etc.) as comprising deterministic and random components. Koutsoyiannis [14] suggested that naturally occurring time series are purely non-deterministic. For all practical purposes, he threw the towel in and gave up on the practice of extrapolation. Despite that, the antithetic time model postulates that natural time series also comprise deterministic and random components. Antithetic time series analysis assumes that the fitted values from a mathematical model are deterministic and that the errors comprise random and systematic bias components. Perceived on both ontological and epistemological grounds, all three components can exist together, and can be separated into corresponding individual mathematical components. As long as the effects of the error randomness and error bias are confounded, it is impossible to arrive at the correct mathematical model.

As a time series evolves over time, future values are created by the mechanism involved. So long as the time series is not observed, the evolution continues unaffected and is as it should be. However, once the time series is observed, that is, once a measurement is taken via a mathematical model, the time series as seen through the instrument of the model is automatically biased, and a change in the time series appears to be effected. This is an event horizon. This has a physical parallel in the concept of quantum mechanics where observation also affects the observed. Even if the parameter bias is de minimis, the forecast bias will accumulate and become very large. The bias must be corrected in the forecast value and it must be corrected dynamically. Forecast bias correction must be performed dynamically.

The purpose of this article is to inform on a major directional change, a paradigm that solves this problem. It is recognized that mathematics can model complex real world systems. Furthermore, it is recognized that mathematics can have complexities of its own. So, it appears that we need a mathematical concept to model the complexity of the mathematics of the autoregressive time series model. It turns out that the solution, involving the bifurcation of error into random and systematic bias components, while entirely counterintuitive, illusive, novel, and accompanied by non-trivial proof, is relatively simple to implement.

2.5. Reversing the correlation

The key to the portal between serial data and unbiased extrapolation is the reversal of correlation. This is made possible by the Ridley [20] antithetic time series theorem

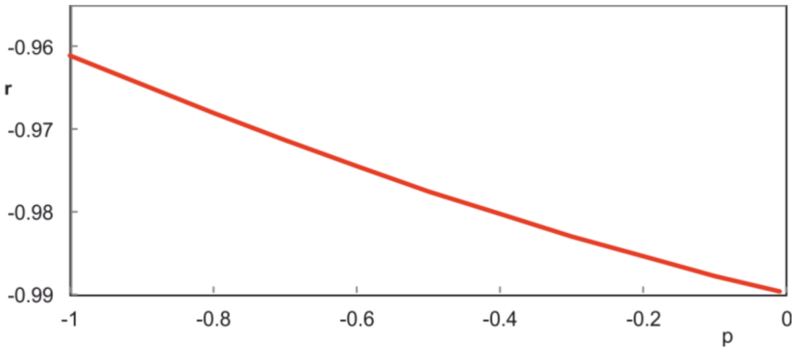

Fig. 1. The sample correlation r between X_t and its pth power X_t^p as p approaches 0^- for the sample data. When p is positive, the correlation must be positive, and when $p = 1$, the correlation must be exactly plus one (not shown). When p is negative, the correlation is negative. When $p = 0$, the correlation must be zero. However, as p approached zero from the left (p inside an infinitesimal neighborhood of zero but not zero), the correlation does not approach zero, instead, it approaches -1. The nature of this correlation is counterintuitive

which states that if $X_t > 0$, $t = 1, 2, 3, \ldots$ is a discrete realization of a lognormal stochastic process, such that $\ln X_t \sim N(\mu, \sigma)$, then if the correlation between X_t and X_t^p is ρ_{XX^p}, then $\lim_{p \to 0^-,\ \sigma \to 0} \rho_{XX^p} = -1$.

The requirement that the standard deviation $\sigma \to 0$ is easily met in practice because it applies to the logarithm of X_t and the logarithm is always relatively de minimis. Intuition might suggest to us that the limit as $p \to 0^-$ of the correlation between X_t and X_t^p is zero since X_t^0 is the constant 1. However, we are interested in an infinitesimal neighborhood of zero for p, not zero. The limiting value of the correlation turns out to be –1. The reversal of correlation is illustrated in Fig. 1.

2.6. The distribution matters

It is important to recognize that the probability distribution of the time series does matter. The foregoing analysis of correlation was based on the positive valued lognormal distribution. The normal distribution is chosen quite often by researchers. One reason is because it occurs quite often in nature. The other reason is because of the availability of normal probability theory for analysing data. If a random variable has a lognormal distribution, then its natural logarithm has a normal distribution. If a random variable has a normal distribution, then after exponentiation it has a lognormal distribution. This fact may come in handy since technicaly speaking, the normal distribution supports negative numbers, and negative numbers cannot be raised to the negative and fractional values of p. One way of handling normally distributed data is to add a constant to each number in the data sample prior to constructing the antithetic time series model, then subtracting it later. It can be proven that this has no adverse impact on the correlation [20]. In passing, purely out of curiousity, it is noted that when the distribution is uniform, the limiting value of the correlation is not –1, it is $-(3/2)^{1/2}$. Since uniformly distributed time series are either very rare or non-existent in nature, one need not be concerned with them in practice.

2.7. Antithetic combining

Finally, we can now combine the two antithetic time series. The antithetic combining model [20] is $\hat{X}_{ct} = \omega \hat{X}_t + (1-\omega)\,\hat{X}'_t$, $t = 1, 2, 3, \ldots, n$, where $\hat{X}_t$ are fitted values obtained from the above time series model $X_t = \Phi X_{t-1} + \varepsilon_t$, $t = 2, 3, \ldots, n$. The parameter $-\infty < \omega < \infty$ 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 $\hat{\varepsilon}_t$ and $\hat{\varepsilon}'_t$, respectively, that are biased and when weighted, $\omega\hat{\varepsilon}_t$ and $(1-\omega)\hat{\varepsilon}'_t$ are perfectly negatively correlated. The antithetic component $\hat{X}'_t$ is estimated from $\hat{X}'_t = \bar{X} + r_{\hat{X}\hat{X}^p}(s_{\hat{X}}/s_{\hat{X}^p})(\hat{X}_t^p - \overline{\hat{X}^p})$, $t = 1, 2, 3, \ldots, n$, 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. The weight ω is chosen so as to minimize the combined model fitted MSE in $\hat{X}_{c,t}$. How the equation for $\hat{X}'_t$ works is as follows. The units of $\hat{X}_t^p$ are unknown and must be converted back to the original units of $\hat{X}_t$ by rescaling. This is accomplished by multiplying the variation of $\hat{X}_t^p$ about its own mean $\overline{\hat{X}_t^p}$ by the ratio of the standard deviations of $\hat{X}_t$ and $\hat{X}_t^p$ respectively, and by the coefficient of correlation between them. Since $\hat{X}_t$ and $\hat{X}_t^p$ are perfectly correlated, albeit negatively, this rescaling by inverse transformation will not introduce any new errors. As discussed earlier, when reversed, the correlation is approximately –1.

If the original model is biased, then the combined MSE will be smaller than the MSE of the original model. Also, the combined model error will be completely random and contain no systemic component. If the original model is unbiased, then ω will simply be equal to 1, and the combined fitted values will just be the original fitted values.

Notice that antithetic combining involves only one time series. Other researchers [3] considered combining several independent time series. If each component model contributes something that the others do not, then such an approach may produce an improved combined model. However, it does not give any consideration to the presence of bias that is inherent in all models. Combining antithetic time series focuses on identifying the best single method and model, then eliminating the bias.

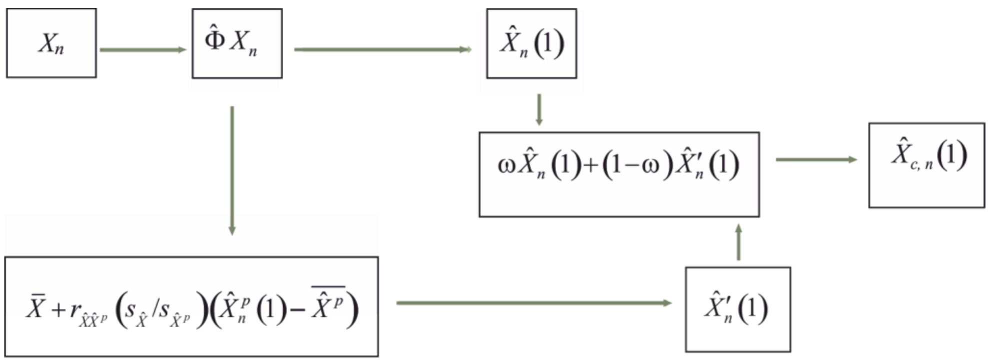

Fig. 2. Dynamic reversal of correlation and combining of antithetic forecasts

By extension, the τ period forecasts at time n are given by $\hat{X}_{c,n}(\tau) = \omega\hat{X}_n(\tau) + (1-\omega)\hat{X}'_n(\tau)$, where $\hat{X}'_n(\tau) = \bar{X} + r_{\hat{X}\hat{X}^p}(s_{\hat{X}}/s_{\hat{X}^p})(\hat{X}_n^p(\tau) - \overline{\hat{X}^p})$, $\tau = 1, 2, 3, \ldots$ (Fig. 2).

3. Empirical demonstration

To demonstrate the effect of bias and how to correct it, consider the well-known CompanyX data that was published by Chatfield and Prothero [2]. The series of recorded sales of electric heaters is a surrogate for what is charted in Fig. 3 as demand.

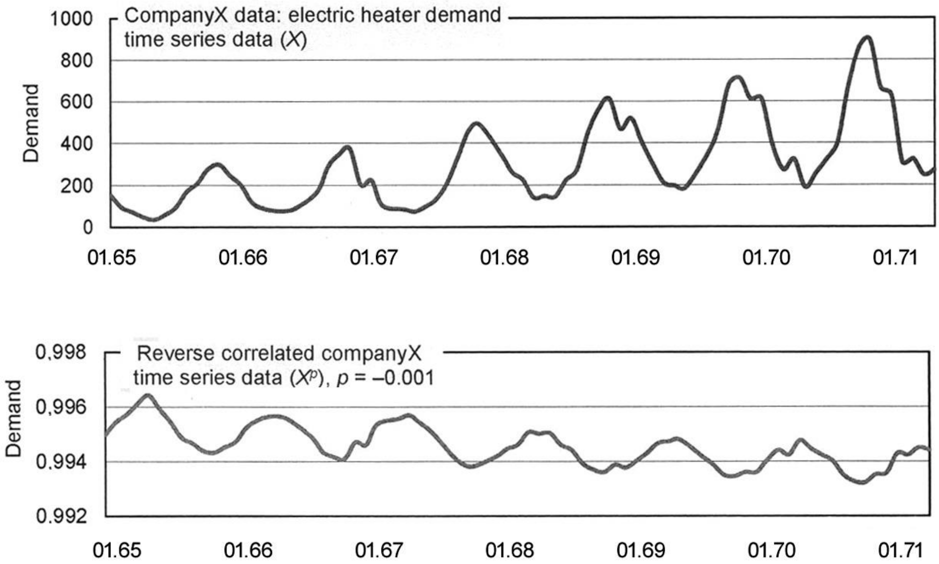

Fig. 3. Original and reverse correlated antithetic time series. The original time series is upward trending while the reverse correlated time series is downward trending. The cyclical and other seasonal patterns in the two time series are also reversed. The value $p = -0.001$ is used to approximate $p \to 0^-$

There are other data that are more current, but the reason that the CompanyX time series is selected for this demonstration is because of the several expert attempts that were made to forecast it and that resulted in failure. The failures occurred in the face of what is a seeming regular and predictable time series. Furthermore, this series contains all the elements of growth in mean and variance, and seasonal and other cyclical elements that are typical of naturally occurring time series. The apparent correspondence between the growth in mean and the variance is a source of what is known as multiplicative errors. The reversal of correlation applied to this time series is demonstrated in Fig. 3.

The forecasts shown in Fig. 4 are the best of the expert forecasts reported and discussed extensively [2, 21]. The data were described only as the sales of electric heaters in England. The forecasts were performed by the best autoregressive integrated moving average (ARIMA) model. Despite the seeming regularity and predictability of the time series, the forecasts diverge from the actual values. The combined antithetic forecasts converge. Their average forecast errors remain constant over the forecast horizon. The combined antithetic forecast calculations were performed by the computer program Fourcast [7].

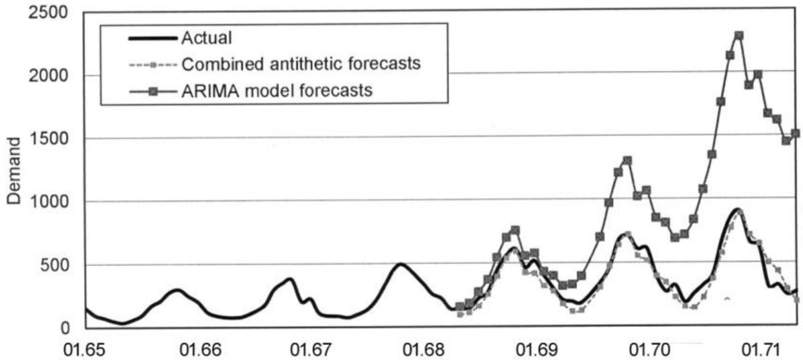

Fig. 4. Based on actual data X_t from January 1965 to April 1968, an autoregressive moving average (ARIMA) model is used to make forecasts for May 1968 to May 1971. The traditional ARIMA model forecasts $\hat{X}_n(\tau)$ diverge over time. The combined antithetic forecasts $\hat{X}_{c,n}(\tau)$ eliminate the bias, follow the pattern of the data, and converge to a constant average error

4. Discussion

The well-known principle of statistics that the smaller the bias of an estimate, the larger its variance is and vice versa applies to estimates of the parameters of the time series model for a specified MSE = Variance + Bias^2. 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, while idiopathic, is associated with a systematic error component. There is a resemblance of this statistical principle to the physical concept that the more that is known about the position of a particle, the less is known about its momentum and vice versa [19, 10, 5]. The resemblance is even greater when it is restated as: the smaller the variance of the estimate of the position of a particle, the larger its bias is and vice versa. Antithetic time series theory as applied here is used to separate the complexity of mathematical statistics from the complexity of a system to be analyzed by mathematical statistics. In this article we see the efficacy of antithetic combining that was demonstrated on a seemingly regular economic time series. The series is typical of what is observed in nature. While the theoretical and measurement construction for demand is manmade, the fundamental provenance of demand is natural. Although it appears to be regular, the series is known to pose insurmountable forecast bias [2]. Other methods combine predicted values that are extrapolated from different independent models. Antithetic combining is applied to the best single model. The original series of values from the model are obtained. A new antithetic series is created so as to be perfectly negatively correlated with the original series. Both the original and antithetic series contain systematic error components. The original and antithetic series are combined, such that the systematic error components dynamically cancel. All the features of the original series are retained while the bias, which is an artifact of the mathematical model, is removed. There is no need to know the amount or source of the bias. Since the bias is unknown and is not measurable prior to its cancellation, the conceptualization of bias cancellation can be regarded as a mind experiment. A range of possible distribtions and variances are investigated and discussed further [18, 21].

References

[1] CHAITIN G.J., *Randomness and mathematical proof*, Sci. Amer., 1975, 232, 47.

[2] CHATFIELD C., PROTHERO D.L., *Box–Jenkins seasonal forecasting: problems in a case study*, J. Royal Stat. Soc. (A), 1973, 136, 295.

[3] CLEMEN R.T., *Combining forecasts. A review and annotated bibliography*, Int. J. Forecastng, 1989, 5, 559.

[4] COPAS J.B., *Monte Carlo results for estimation in a stable Markov time series*, J. Royal Stat. Soc. (A), 1966, 129, 110.

[5] EINSTEIN A., PODOLSKY B., ROSEN N., *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev., 1935, 47, 777.

[6] FOOTE R., *Mathematics and complex systems*, Science, 2007, 318, 410.

[7] FOURCAST, *Application program*, EMC, Inc., Version 2010.12, http://www.fourcast.net/fourcast. File: CompanyX.zip., 2010,

[8] GRILICHES Z., *A note on serial correlation bias in estimates of distributed lags*, Econometrica, 1961, 29, 65.

[9] HAMMERSLEY J.M., MORTON K.W., *A new Monte Carlo technique. Antithetic variates*, Math. Proc. Cambridge Philosophical Society, 1956, 52, 449.

[10] HEISENBERG W., *Über den anschaulichen Inhalt der Quanten theoretischen Kinematik und Mechanik*, Z. Phys., 1927, 43, 172.

[11] KENDALL M.G., *Note on bias in the estimation of autocorrelation*, Biometrika, 1954, 41, 403.

[12] KLIEJNEN J.P.C., *Antithetic variates, common random numbers and optimal computer time allocation in simulation*, Manage. Sci., 1975, 21, 1176.

[13] KOLMOGOROV A.N., *Three approaches to the definition of the quantity of information*, Problems Inf. Trans., 1965, 1, 1.

[14] KOUTSOYIANNIS D., *A random walk on water*, Hydr. Earth Syst. Sci., 2010, 14, 585.

[15] LI B., NYCHKA D.W., AMMANN C.M., *The value of multi-proxy reconstruction of past climate*, J. Am. Stat. Assoc., 2010, 105, 883.

[16] MARRIOTT F.H.C., POPE J.A., *Bias in the estimation of autocorrelations*, Biometrika, 1954, 41, 390.

[17] MyPulse Smart Monitors, Inc., Version 2010.12, http://mypulsemonitor.com/downloadsoftware.html., 2010.

[18] NGNEPIEBA P., RIDLEY A.D., *General theory of antithetic time series*, J. Appl. Math. Phys., 2015, 3 (12), 1726.
[19] POPPER K., *Quantum Theory and the Schism in Physics*. Routledge, London 1992, http://www.scirp.org/journal/jamp http://dx.doi.org/10.4236/jamp.2015.312197
[20] RIDLEY A.D., *Optimal antithetic weights for lognormal time series forecasting*, Comput. Oper. Res., 1999, 26, 189.
[21] RIDLEY A.D., NGNEPIEBA P., *Antithetic time series analysis and the CompanyX data*, J. Royal Stat. Soc. (A), 2014 (1), 177, 83.

Received 27 May 2016
Accepted 13 September 2016