Bangor Business School
Working Paper

BBSWP/13/004

Forecasting multivariate time series with the Theta Method

By

Dimitrios D. Thomakos*,**
Konstantinos Nikolopoulos*+**

***University of Peloponnese**
****Rimini Centre for Economic Analysis**
***** Bangor Business School**
+ corresponding author

July, 2013

Bangor Business School
Bangor University
Hen Goleg 2.06
College Road
Bangor
Gwynedd LL57 2DG
United Kingdom
Tel: +44 (0) 1248 382277
E-mail: k.nikolopoulos@bangor.ac.uk

Forecasting multivariate time series with the Theta Method[1]

Abstract

In this study building on earlier work on the properties and performance of the univariate Theta method for a unit root data generating process we: (a) derive new theoretical formulations for the application of the method on multivariate time series, (b) investigate the conditions for which the multivariate Theta method is expected to forecast better than the univariate one, (c) evaluate through simulations the bivariate form of the method, (d) evaluate this latter model in real macroeconomic and financial time series. The study provides sufficient empirical evidence to illustrate the suitability of the method for vector forecasting; furthermore it provides the motivation for further investigation of the multivariate Theta method for higher dimensions.

Keywords: Theta method; univariate; multivariate time series; unit roots; vector forecasting;

[1] This paper has been submitted for publication in the *European Journal of Operational Research*.

1. Introduction

In this study we propose new theoretical formulations for the application of the Theta method on multivariate time series and discuss the conditions for which the multivariate Theta method forecasts better than the univariate one. The motivation for this investigation comes: (a) from the performance of the Theta method in terms of forecasting accuracy in the M3 forecasting (Makridakis and Hibon, 2000), (b) from the simplicity, and robustness of the method since it can been integrated in the theoretical framework of the quite popular family of exponential smoothing models (Hyndman and Billah, 2003), and (c) last but not least, from the recent theoretical developments (Thomakos and Nikolopoulos, 2013) that illustrate the potential for superior performance of the method for unit root data, and thus making the method applicable to a whole new and wider range of data series coming from finance and economics.

Is the proposed investigation a good idea in the first place? The intuition behind this extension of a univariate model in higher dimension, comes from the notion that if there is a common element that drives the evolution of the series (in the multivariate dataset), call it *causality*, *cointergration*, *co-movement* or…, there should be an advantage if that can be estimated directly from the multivariate data rather than the univariate series and then used for the benefit of forecasting accuracy. In essence it is another episode of the classic anecdote of the 'the tree and the forest'; and through this metaphor we want to emphasize the inability of local (univariate) models to capture properties that are only evident in the multivariate structure of the data.

So now that it is clear what we do, why we do it and what we expect…did we actually make it? By and large the results that will follow in the next few sections, we believe they provide sufficient evidence to illustrate the suitability of the method for vector forecasting; but even more importantly it provides the motivation for further investigation of the multivariate theta model for higher dimensions, as well as against more prominent benchmarks.

The rest of the article is structured as follows: the following two sections provide a brief background on forecasting multivariate data with exponential smoothing and on the Theta method. The new theoretical propositions are presented in section four followed by simulations and empirical results in sections five and six respectively. We conclude the article and highlight roadmaps for future research in the final section.

2. Forecasting multivariate time series with Exponential Smoothing models

Forecasting with exponential smoothing methods has been very popular both among practitioners as well as academics for the most of the last six decades, with the origins of the method being attributed to the seminar work of Bob Brown (1959). Despite the success of the univariate exponential smoothing models, the multivariate forms were developed independently and at later stages, with most of the work appearing in the mid-80s mostly through the work of Harvey in 1986 (and a follow-up book that came out in 1989 capturing most of the developments to that date), although the origins of the multivariate framework are dated back in mid-sixties.

More recently, a lot of attention in the use of exponential smoothing for vector forecasting was created through the work of Hyndman et al. (2002) that proposed a state space framework for forecasting with exponential smoothing methods and a technical book later in 2008 jointly with Koehler, Ord and Snyder. Following that stream of research Silva et al. (2010) have shown that for non-seasonal economic data by using a multivariate exponential smoothing framework, greater forecasting accuracy may be achieved; the proposed concept in that study – the Vector Innovations Structural Time-Series (VISTS) framework is start becoming popular in the field, and has been recently adopted and extended by Athanasopoulos and Silva (2012) via explicitly integrating seasonality. In essence, this latter model is another extension in the class of Hyndman et al. models (2002) with the introduction of stochastic seasonal components.

In a similar fashion, Athanasopoulos and Vahid (2008) evaluated ARIMA specifications on macroeconomic data sets where there is evidence Granger causality and therefore one can claim that some of the series in the dataset could be useful for forecasting some others. In that study the potential of the use of VARMA models is favorably argued – given the improvements in computing power in the last two decades, and the argument is further supported from empirically proven improved forecasting accuracy of VARMAs over VARs.

3. The Theta method

Assimakopoulos' and Nikolopoulos' (2000) Theta method is a univariate forecasting method based on the concept of modifying the second differences of a time series. This modification is managed through the θ coefficient which is applied directly to the second differences of the time series. When $\theta<1$ the second differences are reduced while when $\theta>1$ the second differences are increased. The new time series that are produced via this procedure are called Theta-lines (or θ-lines) which are then extrapolated and combined to produce the forecast of the time series.

Hyndman and Billah (2003) made the first significant advance in theory for the method by first showing that for the specific Theta model applied in the M3 competition data, the forecasting function is equivalent to simple exponential smoothing forecasts with drift; a drift being equal to half the slope of a linear regression line on time. Secondly they proved that under the assumption of a state-space model data generating process, the forecast function is that of an ARIMA(0,1,1) model with drift.

More recently Nikolopoulos et al. (2012) provided several empirical results, but also a very intuitively appealing formula for the calculation of θ-lines as a linear combination of the data and a fitted time trend as: $\theta X_t + (1-\theta)\left(\hat{a} + \hat{b}t\right)$, were $\hat{a}, \hat{b}$ are the usual least squares estimators and X_t the actual data.

Starting from that later result, Thomakos and Nikolopoulos (T&N hereafter, 2013) provided the second major theoretical advancement in the method by providing the full theoretical underpinnings and properties of the method for a unit root data generating processes. They also provided for the first time theoretical formulations for when a single 'Theta line' is used, rather than a combination of two 'Theta lines' as in the aforementioned studies. Furthermore they calculated an optimal value for the theta parameter which coincides with the first order autocorrelation of the innovations and they proved that the optimal forecast function for the model run in the M3 competition was identical with that of ARIMA(1,1,0). Finally they provided formulae for optimal weights when combining two 'Theta lines' as in the model of Assimakopoulos and Nikolopoulos (2000) rather than an optimal value for the drift as in Hyndman and Billah (2003). The paper concluded with a series of simulations and empirical results on the M3 yearly data that confirmed the theoretical propositions.

4. Extending the Theta method in a multivariate unit root time series Model Specification

Consider a multivariate time series $\widetilde{\boldsymbol{X}}_t$, of dimension *k*, that follows a process with unit roots and with possible non-zero drift vector $\boldsymbol{\mu} \neq 0$ as:

$$\widetilde{\boldsymbol{X}}_t \stackrel{\text{def}}{=} \boldsymbol{\mu} + \widetilde{\boldsymbol{X}}_{t-1} + \boldsymbol{u}_t = \widetilde{\boldsymbol{X}}_0 + \boldsymbol{\mu} t + \boldsymbol{S}_t \qquad \text{(1a)}$$

where the innovations $\boldsymbol{u}_t$ are assumed to follow a zero mean, stationary time series with finite second moments and we write $\boldsymbol{S}_t = \sum_{j=1}^{t} \boldsymbol{u}_t$ for the stochastic trend of the cumulated innovations. Adjusting the series by removing the initial value $\boldsymbol{X}_t \stackrel{\text{def}}{=} \widetilde{\boldsymbol{X}}_t - \widetilde{\boldsymbol{X}}_0$ we obtain:

$$\boldsymbol{X}_t = \boldsymbol{\mu} t + \boldsymbol{S}_t \qquad \text{(1b)}$$

where the deterministic trend line $E(\boldsymbol{X}_t) = \boldsymbol{\mu} t$ is the unconditional mean of $\boldsymbol{X}_t$. This specification is a direct extension of the univariate Theta method in a series with a unit root as presented in T&N.

Let us now look as to what sort of changes occur in the multivariate setting vis-à-vis the univariate one, while maintaining the simplicity of the method. This later point needs to be emphasized as part of the success of the Theta method is attributed to its simplicity and robustness, characteristics that should be present in any generalization of the method in higher dimensions. Having that in mind, we define the *multivariate Theta line* to depend on a parameter matrix $\mathbf{\Theta}$ rather than a single parameter θ. We now have:

$$\boldsymbol{Q}_t(\mathbf{\Theta}) = \mathbf{\Theta}\boldsymbol{X}_t + (\mathbf{I} - \mathbf{\Theta})\boldsymbol{\mu}t \tag{2α}$$

where $\mathbf{I}$ is the k-dimensional identity matrix.

The Bivariate Theta method

Without any loss of generality, for the rest of our analysis we will consider the bivariate case k=2 for ease of presentation. Expanding equation (2α) above we get that in the bivariate setting the Theta line for each time series is the sum of the original univariate Theta line in T&N plus a factor that depends on the other time series:

$$Q_{t1}(\theta_{11}, \theta_{12}) = \theta_{11}X_{t1} + \theta_{12}X_{t2} + (1 - \theta_{11})\mu_1 t - \theta_{12}\mu_2 t \tag{2b}$$

$$Q_{t2}(\theta_{21}, \theta_{22}) = \theta_{21}X_{t1} + \theta_{22}X_{t2} - \theta_{21}\mu_1 t + (1 - \theta_{22})\mu_2 t \tag{2c}$$

which can equivalently be written as:

$$Q_{t1}(\theta_{11}, \theta_{12}) = Q_{t1}(\theta_{11}) + \theta_{12}(X_{t2} - \mu_2 t) \tag{2d}$$

$$Q_{t2}(\theta_{21}, \theta_{22}) = Q_{t2}(\theta_{22}) + \theta_{21}(X_{t1} - \mu_1 t) \tag{2e}$$

We see, therefore, that in extending from the univariate to the bivariate Theta line we are merging the univariate component with an additional one which has an interpretation similar to that of *causality*. That is, the bivariate Theta line would be different from the univariate for at least one of the two time series if and only if $\theta_{ij} \neq 0, i \neq j$, for at least one of the off-diagonal parameters. When this (these) condition(s) hold we have that the bivariate Theta line is different from the univariate one and that one can expect different forecasting performance between the univariate and bivariate approaches.

It is straightforward to show that the form of the forecasting function that one obtains by using the bivariate Theta line is a direct extension of the univariate results of T&N and collapses to that of a first order vector autoregression (VAR), in either levels or in differences. One should not, however, dismiss this as simplistic without further qualifications. This is because we can have a couple of interesting twists in the current case. Let's examine them in turn.

In the univariate case T&N provide the justification as to why the data generating process combined with the simplicity of the Theta method requires that we consider the following forecast function:

$$\begin{aligned}\widehat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{F}_{t+1}(\boldsymbol{\Theta}) &\stackrel{\text{def}}{=} \boldsymbol{\mu} + \boldsymbol{Q}_t(\boldsymbol{\Theta}) \\ &= \boldsymbol{\mu} + \boldsymbol{\Theta}\boldsymbol{X}_t + (\mathbf{I} - \boldsymbol{\Theta})\boldsymbol{\mu}t \\ &= \boldsymbol{\mu}(t+1) + \boldsymbol{\Theta}(\boldsymbol{X}_t - \boldsymbol{\mu}t)\end{aligned} \tag{3a}$$

with forecast error given by:

$$\begin{aligned}\boldsymbol{X}_{t+1} - \boldsymbol{F}_{t+1}(\boldsymbol{\Theta}) &= \boldsymbol{\mu}(t+1) + \boldsymbol{S}_{t+1} - \boldsymbol{\mu}(t+1) - \boldsymbol{\Theta}S_t \\ &= \boldsymbol{S}_{t+1} - \boldsymbol{\Theta}\boldsymbol{S}_t\end{aligned} \tag{3b}$$

Similarly, if we take differences we obtain the corresponding forecast function and forecast error given by:

$$\begin{aligned}\widehat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{Y}_{t+1}(\boldsymbol{\Theta}) &\stackrel{\text{def}}{=} \boldsymbol{X}_t + \Delta\boldsymbol{F}_{t+1}(\boldsymbol{\Theta}) \\ &= \boldsymbol{X}_t + \Delta\boldsymbol{Q}_t(\boldsymbol{\Theta}) \\ &= \boldsymbol{\mu} + \boldsymbol{X}_t + \boldsymbol{\Theta}(\Delta\boldsymbol{X}_t - \boldsymbol{\mu})\end{aligned} \tag{4a}$$

$$\begin{aligned}\boldsymbol{X}_{t+1} - \boldsymbol{Y}_{t+1}(\boldsymbol{\Theta}) &= \boldsymbol{\mu} + \boldsymbol{X}_t + \boldsymbol{u}_{t+1} - \boldsymbol{\mu} - \boldsymbol{X}_t - \boldsymbol{\Theta}\boldsymbol{u}_t \\ &\quad \boldsymbol{u}_{t+1} - \boldsymbol{\Theta}\boldsymbol{u}_t\end{aligned} \tag{4b}$$

While in the univariate case the forecast function in levels in equation (3a) was not the preferred one, we can easily see here that we cannot immediately dismiss it because of the possible presence of cointegration in the detrended series, which would make the forecast error series a stationary one.

We can thus have a special case of a model where the linear deterministic trend is not annihilated by the cointegrating relationship which is then present only in the forecast error component of equation (3b). This naturally dictates that one can consider a reduced rank regression in order to estimate the parameter matrix $\mathbf{\Theta}$. If cointegration is indeed present one should (at least in theory) prefer the forecast function of equation (3a) over that of (4a) which is specified in differences. However, this should pre-suppose that the first order VAR that is implicitly formed in equation (3b) is correctly specified for cointegration estimation for if it's not then one can have misspecification problems that will result in decreased forecasting performance.

But even more can be glimpsed on the potential improvement of the bivariate Theta method over the univariate one if we consider the forecast error that comes from the forecast function in differences, i.e. from equation (4b). Take the first of the two equations and write the mean square forecast error:

$$MSE_1(\theta_{11}, \theta_{12}) = E\left(u_{t+1,1} - \theta_{11}u_{t1} - \theta_{12}u_{t2}\right)^2 =$$
$$E\left(u_{t+1,1} - \theta_{11}u_{t1}\right)^2 + E(\theta_{12}u_{t2})^2 - 2\theta_{12}E\left(u_{t+1,1}u_{t2}\right) + 2\theta_{11}\theta_{12}E(u_{t1}u_{t2}) =$$
$$MSE_1(\theta_{11}) + E(\theta_{12}u_{t2})^2 - 2\theta_{12}E\left(u_{t+1,1}u_{t2}\right) + 2\theta_{11}\theta_{12}E(u_{t1}u_{t2}) \quad (5)$$

which is the sum of the univariate Theta method mean-squared error and extra components that depend on the serial and cross correlations of the two series. If we find the value of θ_{12} that minimizes the extra components in equation (5):

$$\theta_{12}(\theta_{11}) = \left[E\left(u_{t+1,1}u_{t2}\right) - \theta_{11}E(u_{t1}u_{t2})\right]/Eu_{t2}^2 \quad (6)$$

we can see that θ_{12} will be different from zero if the innovations are either contemporaneously correlated and $\theta_{11} \neq 0$ or are first order cross-correlated with series 2 leading series 1 (and θ_{11} can take any value including zero) or both. Thus, the bivariate Theta method applied in differences should provide smaller forecast errors that the univariate one for series which have some degree of linear co-dependence. Only if the innovations are contemporaneously and serially uncorrelated we would only need to use the univariate method.

Estimation of parameters

The estimation of the all parameters in the bivariate case is similar to the univariate case and straightforward. The drift terms are estimated by the sample means of the first differenced-series, i.e. $\widehat{\boldsymbol{\mu}} = n^{-1} \sum_{t=2}^{n} \Delta \boldsymbol{X}_t$. Then, the detrended/demeaned series are formed as $\widehat{\boldsymbol{S}}_t = \boldsymbol{X}_t - \widehat{\boldsymbol{\mu}} t$ and $\widehat{\boldsymbol{u}}_t = \Delta \boldsymbol{X}_t - \widehat{\boldsymbol{\mu}}$ and the $\boldsymbol{\Theta}$ matrix is estimated either via reduced rank regression or via multivariate least squares respectively.

Note that as far as possible cointegration is concerned we do not pre-test on the type of trend. This is part of the simplicity of the method and the type of forecast function we consider and it should not be seen as a methodological drawback - if one accepts as a starting point the very general process of equation (1b).

5. Simulations

In this section we conduct a simulation experiment to assess the relative performance of the bivariate setting against the univariate setting of the Theta method. We consider three data generating processes which are all described by the general form of the model in equation (1b), that is a multivariate unit root process with an added deterministic trend.

- In the first model we have that $\boldsymbol{u}_t$ is a vector (Gaussian) white noise process with contemporaneously correlated innovations; in this case the optimal value of both θ_{ii} is zero and, therefore, so is the optimal value for θ_{ij}.
- In the second model we have that $\boldsymbol{u}_t$ follows a (Gaussian) VAR(1) process and therefore the (differenced-based) Theta method forecasts should coincide with the minimum mean-squared error forecasts.
- Finally, in the third model we have that $\boldsymbol{S}_t$ follows a (Gaussian) cointegrated VAR(1) model with a single unit root. Here we are interested to see whether imposing cointegration during estimation and using the levels-based forecasts leads to any improvement over the differenced-based forecasts.

In all three models we consider we see the contemporaneous correlation of the innovations at 0.5 and the drift terms at $\boldsymbol{\mu}$= $(0.05, 0.01)^T$. For the second and third modes we consider the following $\boldsymbol{\Theta}$ matrices respectively:

$$\mathbf{\Theta} = \begin{bmatrix} 0.5 & -0.3 \\ 0.3 & 0.5 \end{bmatrix} \text{ and } \mathbf{\Theta} = \begin{bmatrix} 1.0 & 0.0 \\ 0.8 & 0.1 \end{bmatrix} \quad (7)$$

so that in the third model we have a triangular structure, with roots 1.0 and 0.1 and only the second series is influenced by the first (i.e. we expect performance improvements only for the second series).

We use a sample of 200 observations for estimation and a sample of 60 observations for evaluation. Based on the 60 evaluation points we compute the Root Mean-Squared Error (RMSE) as a proxy of the *uncertainty* of the forecasts and the Mean Absolute Error (MAE) as a proxy of the statistical symmetric *accuracy* of the forecasts. We perform a recursive simulation[2] and compute 60 one-step-ahead forecast errors per series. These values are then averaged across 500 replications from which we obtain our final results, so we compute the AverageRMSE and AverageMAE respectively (where averaging is done across the 500 series). Thus, our comparisons are based on 500 average forecast errors and 30.000 one-step-ahead forecast errors overall.

For easier interpretation of the results we do perform scaling and present in the tables the RelativeAverageRMSE (Rel RMSE) - where we just divide the Average RMSE for every forecasting model with the Average RMSE of the Naïve method: so if the metric is less than 1 that indicates that the method performs better than the naïve method. In a similar fashion the RelativeAverageMAE (Rel MAE) is computed. The methods competing is the Univariate Theta method versus three versions of the forecasting function of the Bivariate Theta method: in Levels, in Levels (estimated with a Reduced Rank regression) and in Differences.

[Insert Table 1 about here]

[2] Similar results have been obtained for rolling simulations, that for the sake of the economy of the presentation we do not report here

The results for this simulation are summarized in Table 1 where it is worth noting the following:

- In the first simulation setup, the Naïve method as expected is performing best but the other methods are reaching quite similar performances.
- In the second simulation setup the univariate Theta method performs better than Naïve as expected (since Theta forecasts should coincide with the minimum mean-squared error forecasts) but also the Bivariate Theta in Diffs performs even better for both X1 and X2 which is an encouraging result for the potential use of the multivariate Theta method.
- In the third simulation setup the univariate Theta method performs better than Naïve. The Bivariate Theta performs even better, however for X1 is the Levels and RR-Levels versions that perform best while for X2 (where we were expecting performance improvements) it is the Diffs that gives the best results. So yet again another promising result in favor of the multivariate form of the method.

6. Empirical evaluations

To avoid being accused of presenting evaluations on artificially created data series we will also employ the competing methods for forecasting real series. We have selected three well known 'duets' of economic and financial series, that is commonly accepted in the literature that they are related to each other in one way or another (all data have been obtained from the FRED database http://research.stlouisfed.org/fred2/ accessed on the 29-06-2013) :

- Quarterly data from 01-01-1947 to 01-01-2013 (in total 265 observations) of Real Gross Domestic Product (RGDP) vs. Real Gross Private Domestic Investment (RPINV) in Billions of Chained 2005 Dollars [Source: U.S. Department of Commerce: Bureau of Economic Analysis]. The quarterly data series is illustrated in figure 1.
- Monthly data from 01-01-1995 to 01-04-2013 (in total 220 observations) of Real Personal Consumption Expenditures vs. Real Disposable Personal Income (in Billions of Chained 2005 Dollars) [Source: U.S. Department of Commerce: Bureau of Economic Analysis]. The monthly data series is illustrated in figure 2.

- Weekly data from 08-01-1999 to 21-06-2013 (in total 755 weeks) of EURUSD vs. GBPUSD foreign exchange rates [Source: Board of Governors of the Federal Reserve System]. The weekly data series is illustrated in figure 3.

So we will be using in our empirical investigation a factorial setup evaluating series with three different frequencies of data (quarterly, monthly, weekly) and two different lengths of series (short, long) per frequency; this latter is achieved by using two different starting points when modeling and forecasting the series. Note that the real world data we use have a common characteristic: they tend to move together. In particular, the economic time series not only follow the same economic cycle but they exhibit theoretical "causality": personal disposable income should drive personal consumption (income is causal in the sense that is part of the consumption equation, it is understood that both series evolve simultaneously; however, note that in a monthly setting one can easily argue that lagged income does "cause" current consumption); similarly, private investment is affected by gross domestic product, certainly in a lagged sense as in standard macro equations, although the simultaneity is much more pronounced in the quarterly data. Then, the financial time series are also driven by common factors and should exhibit positive feedback most of the time as they are benchmarks against the power of the US dollar. We will later see that these meaningful a priori considerations are fully supported by the empirical results.

[Insert Figure 1 about here]

By looking in figure 1 one can notice that the series seem to be 'trending' together although RPINV is more volatile and it tends to fluctuate around RGDP especially after the early 90s (while earlier it was consistently trending at lower levels).

[Insert Figure 2 about here]

By looking in figure 2 one can notice that the series seem to be 'trending' together, almost in parallel, however RPCE moves always at lower levels with a more-or-less constant difference of about 500 units (that is 500 billions USD) for the period we examine. This of course is expected as people tend to spend less money than what they earn… though not always! (as evidenced from the recent economic crisis…) Both seem to have small variance with RPDI illustrating a few more mini 'shocks', about 4-5 of them in the 18 years' worth of data we analyse. The mini 'shocks' in RPDI seem to be followed by smaller shocks in RPCE a few months later.

[Insert Figure 3 about here]

In figure 3 one can notice that the series seem to be 'moving' together but not necessarily trending as the series are closer to stationary processes both in terms of mean and variance (although we are not contesting that there is a stochastic trend present in these data too). Also it is clear that there is some level of cross-correlation although after mid-2007 the GBPUSD has a level shift and moves at lower levels consistently than EURUSD.

In Table 2 we see the relative errors for forecasting RGDP and RPINV independently (via Naïve and Theta Univariate) and simultaneously via the three versions of the Bivariate Theta model. Both for RGDP and RPINV the Theta method (either univariate or bivariate) performs better than Naïve. For the RPINV series it is the bivariate Theta in differences that performs best for both metrics RMSE (for uncertainty) and MAE (for accuracy), and for both lengths of the series (sort and long); so the bivariate version gives more accurate and less volatile forecasts in this case. The RR-Level version of the Bivariate Theta seems to be performing worst.

For RGDP it is the univariate Theta that performs best for both metrics and for both length of the series (sort and long) however for the latter case the bivariate Theta in difference performs the same as the univariate one. Yet again the RR-Level version of the Bivariate Theta seems to be performing worst. It is worth noting that for this second series the gains over Naïve are much higher than what observed for RPINV. It seems that the less volatile the series (i.e. for RGDP) the more the gain for the bivariate Theta in differences over the univariate one

[Insert Table 2 about here]

In Table 3 we see the relative errors for forecasting RPDI and RPCE independently (via Naïve and Theta Univariate) and simultaneously via the three versions of the Bivariate Theta model. Naïve is a much harder benchmark to beat in this couple of series. Both for RPDI and RPINV the Theta method (either univariate or bivariate) performs better than Naïve but this time it is the univariate that wins for RPDI while it is the bivariate in Levels that wins for RPCE; it is the only instance of all the results we present that the bivariate Theta in levels gives the best forecasts. Yet again the RR-Level version of the Bivariate Theta seems to be performing worst for both series while it seems that we experience yet again here the phenomenon of less variance in the series (as in RPCE) to lead to more gains in terms of forecasting accuracy for the bivariate Theta in differences over the univariate one.

[Insert Table 3 about here]

In Table 4 we see the relative errors for forecasting EURUSD and GPBUSD independently (via Naïve and Theta Univariate) and simultaneously via the three versions of the Bivariate Theta model. Both for EURUSD and GPBUSD the Theta method (either univariate or bivariate) performs better than Naïve and this time it is always the bivariate Theta in diffs that performs the best; this is the only couple of series that the data are not trending and seem to be stationary for level, however we have not tested if it is that characteristic of the data that drives the success of the forecasting function in differences. The RR-Level version of the Bivariate Theta gives the worst performance also for these series.

[Insert Table 4 about here]

Summing up our empirical findings from this section:

- There is always a version of Theta performing better than Naïve
- Both the univariate Theta and the bivariate Theta in diffs consistently perform better than Naïve
- There is almost always a version of bivariate Theta performing better than the univariate one:
    - Usually this one is the bivariate forecasting function in differences.
    - The gain seems to be bigger when the series is less volatile.
- We do not experience any differences in forecasting accuracy that can be attributed to either the length of the series or the frequency of observations
- The results of this section (on real data) are consistent with the results on the simulated data (in the previous section)

7. Concluding remarks and future research

In this study, extending the work of the last fifteen years on the univariate Theta forecasting method, we provided new theoretical formulations for the application - for the first time - of the method on multivariate time series. The theoretical developments presented come with their own merit, and stand as is as true and accurate mathematically-derived formulations.

On top of this new theory, and for the sake of illustrating the use of the method in practice, we employ extensive simulations (with three alternative scenarios) as well as empirical evaluations on real data – from well know macroeconomic and financial series (with three frequencies (quarterly – monthly - weekly) and two different series-lengths (long-short)). All these results were consistent with our theoretical predispositions and intuition. All theory and evidence corroborate to one single fact: the multivariate Theta method works well; (a) better than the random-walk, and (b) seems to perform even better than the univariate one.

A notable result – although out of the initial scope of this investigation, is that the Univariate Theta method also outperforms the Naïve method; this is evidenced in well-known data series that in levels or differences are considered by many academics to be random-walks. This is an important results and corroborates to Thomakos and Nikolopoulos (2013) and Hyndman and Billah (2003) that had advocated for the suitability of the method to forecast financial and economic time series.

An equally interesting result is that the bivariate form of the method performs very well, mostly when the forecasting function in differenced is used, and in most evaluations outperforms the univariate alternative of the method. This is quite important and the gain is amplified when favorable conditions do exist between the two univariate series that form the bivariate data set.

Finally we would like to close this investigation with a technical remark: we need to reiterate and stress the fact that we are looking at a method and not a model. The method is simple to implement and is the same irrespective of the true underlying model or the (statistically correctly specified) estimated model. The use of the unit root data generating process for our analysis is for providing some explicit results for trending economic and finance series and to explain when the Theta forecasts can be optimal. The potential of the method is much greater and we have no reason to believe that the presented results do not stand for other data generating processes; this latter point could be an ideally starting point for further research.

As far as the more distant future of the method is concerned, we believe we have inspired enough motivation for further investigation of the multivariate theta model for higher dimensions, as well as created the necessary attention to motivate further empirical comparisons against more prominent benchmarks in the likes of the VAR and VARMA family of models.

References

Assimakopoulos, V. and Nikolopoulos, K. (2000). "The Theta Model: A Decomposition Approach to Forecasting", *International Journal of Forecasting* 16 (4): 521-530.

Athanasopoulos, G. and de Silva, A. (2012). "Multivariate exponential smoothing for forecasting tourist arrivals", *Journal of Travel Research* 51: 640-652

Athanasopoulos, G., and Vahid, F. (2008). "VARMA versus VAR for Macroeconomic Forecasting", *Journal of Business and Economic Statistics* 26: 237-52.

Brown, R. G. (1959). *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill.

de Silva, A., Hyndman, R. and Snyder, R. D. (2010). "The Vector Innovations Structural Time Series Framework: A Simple Approach to Multivariate Forecasting." *Statistical Modelling* 10: 353-74.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Harvey, A. (1986). "Analysis and Generalisation of a Multivariate Exponential Smoothing Model", *Management Science* 32: 374-80.

Hyndman, R. J., A. B. Koehler, J. K. Ord, and R. D. Snyder. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin-Heidelberg, Germany: Springer-Verlag.

Hyndman, R. J., Koehler, R. D. Snyder, and S. Grose. (2002). "A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods." *International Journal of Forecasting* 18: 439-54.

Hyndman, R. J. and Billah, B. (2003). "Unmasking the Theta Method", *International Journal of Forecasting* 19 (2): 287-290.

Makridakis, S. and Hibon, M. (2000). "The M3-Competition: results, conclusions and implications", *International Journal of Forecasting* 16 (4): 451-476.

Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa , A. and Petropoulos, F. (2012), "The Theta model: An essential forecasting tool For Supply Chain planning", *Lecture Notes in Electrical Engineering* 123: 431-437.

Thomakos, D.D. and Nikolopoulos, K. (2013) "Fathoming the Theta Method for a Unit Root Process", *IMA Journal of Management Mathematics*, first published online December 2, 2012, doi:10.1093/imaman/dps030

Authors Profile

Dimitrios D. Thomakos is Professor of Applied Econometrics in the University of Peloponnese, Greece and Senior Fellow of the Rimini Center for Economic Analysis in Italy. Dimitrios holds an MA, M.Phil and Ph.D from the Department of Economics of Columbia University. His work has appeared in the MIT Review of Economics and Statistics, Canadian Journal of Economics and the International Journal of Forecasting among other prestigious outlets.

Konstantinos (Kostas) Nikolopoulos, Dr. Eng., ITP, P2P, is the Director of Research for the College of Business, Law, Education and Social Sciences at Bangor University, and holds the Chair in Decision Sciences at Bangor Business School, Wales, UK.

Tables & Figures

Table 1. Relative (to Naïve) Performance of the Univariate and Bivariate Theta for Accuracy (MAE) and Uncertainty (RMSE) Forecasting on Simulated series: three scenarios with 500 replications in each.

1st Series		X1				
			Theta	Theta Bivariate		
Simulation setup	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
No serial correlation	Rel RMSE	**1.000**	1.002	1.001	1.001	1.005
	Rel MAE	**1.000**	1.002	**1.000**	**1.000**	1.005
1st order serial correlation	Rel RMSE	1.000	0.928	1.002	1.002	**0.897**
	Rel MAE	1.000	0.930	1.002	1.002	**0.897**
Triangular cointegration	Rel RMSE	1.000	0.897	**0.895**	**0.895**	0.899
	Rel MAE	1.000	0.892	**0.890**	**0.890**	0.894
2nd Series		**X2**				
			Theta	Theta Bivariate		
Simulation setup	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
No serial correlation	Rel RMSE	**1.000**	1.005	1.003	1.003	1.007
	Rel MAE	**1.000**	1.005	1.003	1.003	1.007
1st order serial correlation	Rel RMSE	1.000	0.782	1.002	1.002	**0.727**
	Rel MAE	1.000	0.778	1.002	1.002	**0.725**
Triangular cointegration	Rel RMSE	1.000	0.958	0.965	0.957	**0.817**
	Rel MAE	1.000	0.958	0.965	0.956	**0.816**

Table 2. . Relative (to Naïve) Performance of the Univariate and Bivariate Theta for forecasting *accuracy* (MAE) and forecasting *uncertainty* (RMSE) on two Quarterly series: Real Gross Private Domestic Investment (RPINV) and Real Gross Domestic Product (RGDP).

1st Series		RPINV				
			Theta	Theta Bivariate		
Series Length	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
Long	Rel RMSE	1.000	0.962	0.978	1.006	**0.938**
	Rel MAE	1.000	0.945	0.964	0.991	**0.922**
Short	Rel RMSE	1.000	0.919	1.021	1.060	**0.888**
	Rel MAE	1.000	0.913	0.980	1.013	**0.893**
2nd Series		RGDP				
			Theta	Theta Bivariate		
Series Length	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
Long	Rel RMSE	1.000	**0.745**	0.794	0.921	0.750
	Rel MAE	1.000	**0.649**	0.686	0.847	0.652
Short	Rel RMSE	1.000	**0.713**	0.843	1.145	**0.713**
	Rel MAE	1.000	**0.596**	0.683	1.036	**0.596**

Table 3. Relative (to Naïve) Performance of the Univariate and Bivariate Theta for forecasting *accuracy* (MAE) and forecasting *uncertainty* (RMSE) on two Monthly series: Real Disposable Personal Income (RPDI) and Real Personal Consumption Expenditures (RPCE).

1st Series		RPDI				
			Theta	Theta Bivariate		
Series Length	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
Long	Rel RMSE	1.000	**0.973**	1.004	1.191	0.993
	Rel MAE	1.000	**0.932**	0.954	1.189	0.956
Short	Rel RMSE	1.000	**0.697**	0.730	0.905	0.808
	Rel MAE	1.000	**0.691**	0.751	0.899	0.805
2nd Series		**RPCE**				
			Theta	Theta Bivariate		
Series Length	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
Long	Rel RMSE	1.000	1.018	**0.980**	1.049	1.019
	Rel MAE	1.000	1.032	**0.926**	1.132	1.030
Short	Rel RMSE	1.000	0.980	0.979	1.004	**0.977**
	Rel MAE	1.000	0.946	**0.902**	0.993	0.940

Table 4. Relative (to Naïve) Performance of the Univariate and Bivariate Theta for forecasting *accuracy* (MAE) and forecasting *uncertainty* (RMSE) on two Weekly series: EURUSD vs. GBPUSD foreign exchange rates.

1st Series		EURUSD				
			Theta	Theta Bivariate		
Series Length	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
Long	Rel RMSE	1.000	0.973	1.020	1.024	**0.970**
	Rel MAE	1.000	0.968	1.027	1.029	**0.963**
Short	Rel RMSE	1.000	0.973	0.999	1.000	**0.968**
	Rel MAE	1.000	0.967	1.000	1.000	**0.957**
2nd Series		GBPUSD				
			Theta	Theta Bivariate		
Series Length	Metric	Naïve	Univariate	Levels	RR-Levels	Diffs
Long	Rel RMSE	1.000	0.981	1.001	1.004	**0.978**
	Rel MAE	1.000	0.972	1.002	1.007	**0.960**
Short	Rel RMSE	1.000	0.997	1.002	1.002	**0.991**
	Rel MAE	1.000	0.983	1.002	1.003	**0.973**

Figure 1. Real Gross Domestic Product vs. Real Gross Private Domestic Investment (in Billions of Chained 2005 Dollars), Quarterly data from 01-01-1947 to 01-01-2013 [Source: U.S. Department of Commerce: Bureau of Economic Analysis]

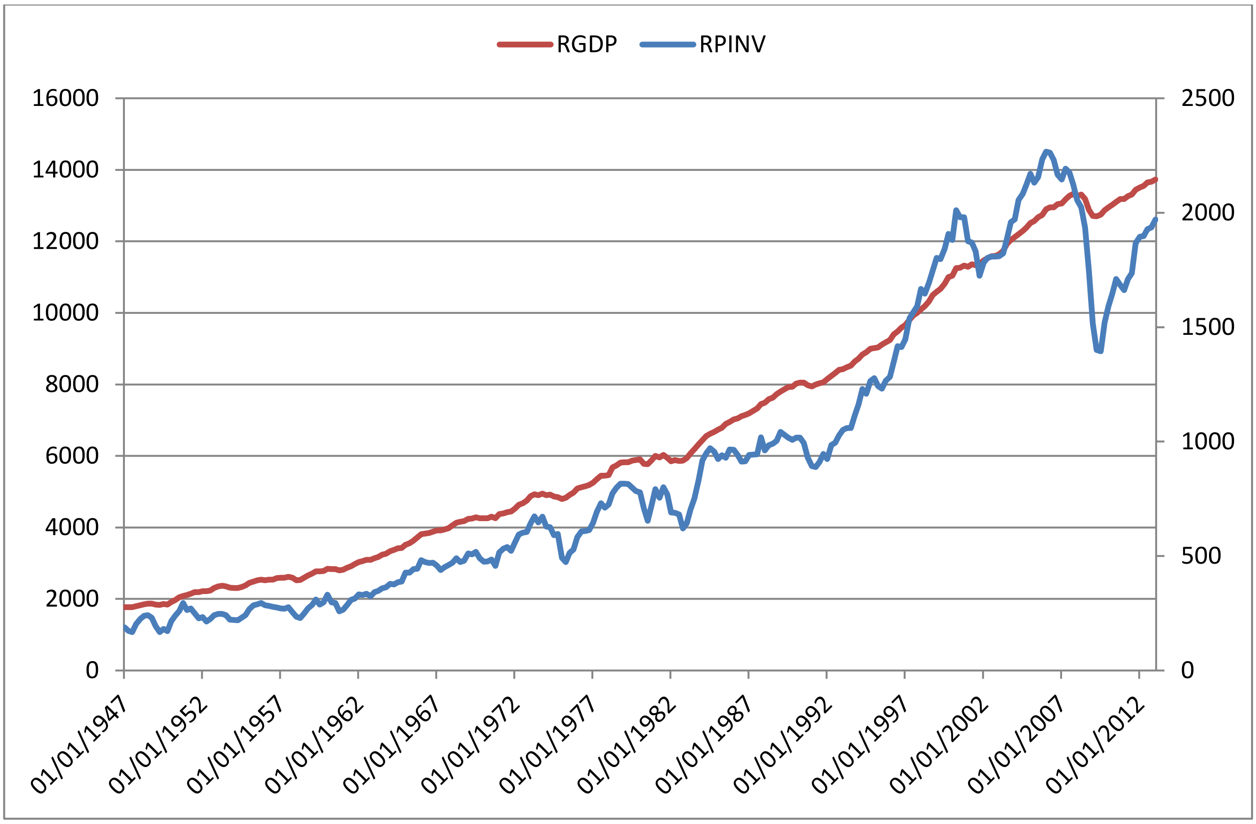

Figure 2. Real Personal Consumption Expenditures vs. Real Disposable Personal Income (in Billions of Chained 2005 Dollars), Monthly data from 01-01-1995 to 01-04-2013 [Source: U.S. Department of Commerce: Bureau of Economic Analysis]

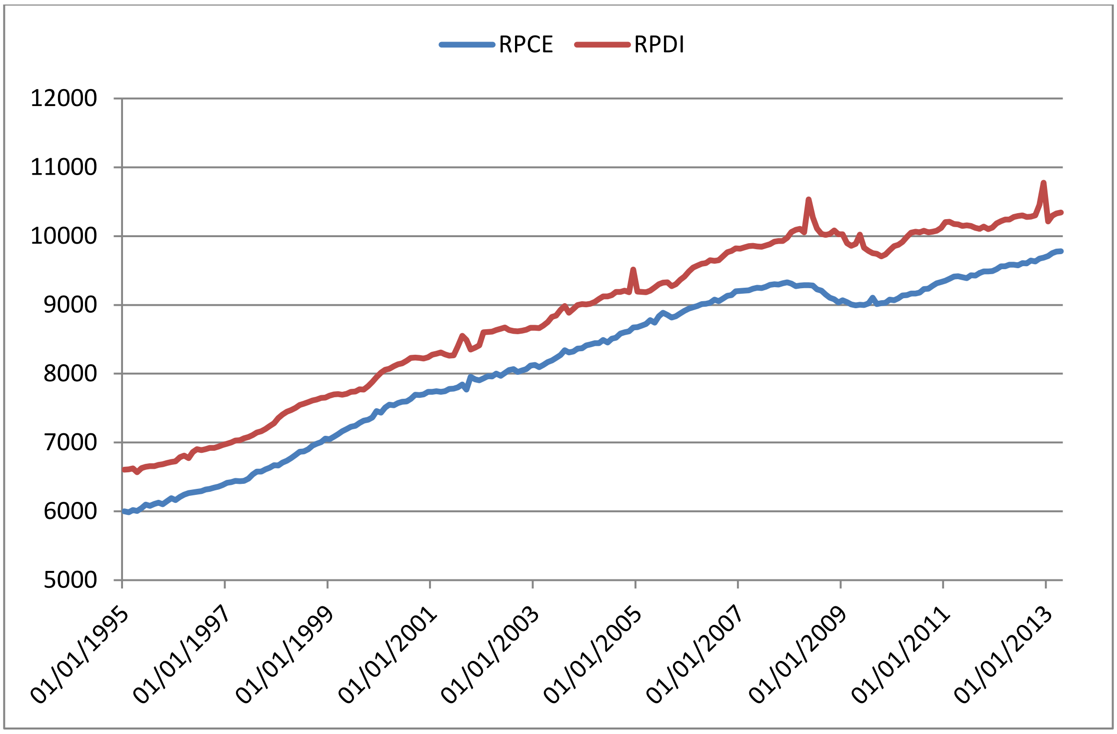

Figure 3. EURUSD vs. GBPUSD foreign exchange rates, Weekly data from 08-01-1999 to 21-06-2013 [Source: Board of Governors of the Federal Reserve System]