



CENTRAL BANK OF CYPRUS
EUROSYSTEM

WORKING PAPER SERIES

Estimation and Prediction of Shipping Trends with the Data-Driven Haar Fisz Transform

Antonis A. Michis

Guy P. Nason

August 2015

Working Paper 2015-01

Central Bank of Cyprus Working Papers present work in progress by central bank staff and outside contributors. They are intended to stimulate discussion and critical comment. The opinions expressed in the papers do not necessarily reflect the views of the Central Bank of Cyprus or the Eurosystem

Address

80 Kennedy Avenue
CY-1076 Nicosia, Cyprus

Postal Address

P. O. Box 25529
CY-1395 Nicosia, Cyprus

E-mail

publications@centralbank.gov.cy

Website

<http://www.centralbank.gov.cy>

Fax

+357 22 378153

Papers in the Working Paper Series may be downloaded from:

http://www.centralbank.gov.cy/nqcontent.cfm?a_id=5755

© Central Bank of Cyprus, 2015. Reproduction is permitted provided that the source is acknowledged.

Estimation and Prediction of Shipping Trends with the Data-Driven Haar-Fisz Transform

Antonis A. Michis*

Guy P. Nason**

August 2015

Abstract

We describe the implementation of a computer-based automatic procedure to estimate the trends associated with debit and credit transaction flows in Cyprus's shipping industry. The procedure was also extended to forecasting. Transactions in the shipping industry do not always coincide with the time the service is provided. The transactions are usually completed gradually throughout the financial year and occasionally involve large amounts for balance settlements. In addition, the transactions are subject to several market risks such as the freight rate and exchange rate changes. Consequently, the transactions frequently exhibit large values and changes in variance, which makes trend estimation and forecasting difficult. A key component of the procedure we implemented is a variance stabilization method based on the Data-Driven Haar-Fisz Transform that enables accurate estimation of trends in volatile time series data. This method is sufficiently flexible to accommodate data characteristics such as cyclical changes, shifts in trend and spikes that are frequently encountered in transaction flow data.

Keywords: Shipping; trend; wavelets; Data-Driven Haar-Fisz Transform

JEL classification: C44; C53; C87

* Statistics Department, Central Bank of Cyprus, 80 Kennedy Avenue, CY-1076, Nicosia, Cyprus. Email: AntonisMichis@centralbank.gov.cy, <https://sites.google.com/site/antonismichis/>.

** School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK. Email: g.p.nason@bristol.ac.uk.

The authors would like to thank Thomas Kazakos, Captain Eugene Adami and the Cyprus Shipping Chamber for valuable comments and suggestions. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Central Bank of Cyprus or the Eurosystem.

Introduction

The shipping industry encompasses multiple operations and intermediary services that are necessary, primarily, for the safe and efficient transportation of goods and passengers. The most important operations concern ship owning and ship management activities. These operations utilize inputs of the services of several other shipping related companies as intermediaries, such as ship chandlers and marine insurance brokers. Currently, many ship owning companies employ the services of third party ship management companies to reduce costs and to reduce the risk of disruptions for longer route voyages. As a result of this tendency for outsourcing, the ship management and ship owning operations have become distinct, and specialization in the industry has increased.

Ship owning companies are taxable legal persons that claim the benefits associated with the use of their vessels in economic activities such as chartering (time, voyage or bareboat) and leasing (operational or financial). Ship management companies provide ship owners with the following services: (a) recruitment and training of seafarers necessary for the navigation and operation of vessels, (b) mechanical maintenance and technical supervision of vessels and (c) commercial operations such as logistics and financial management. These services ensure the efficient operation of ships in accordance with maritime regulations (Branch and Robarts, 2014, p. 447).

The shipping industry is one of the most important contributors to Cyprus's Balance of Payments (BOP). Cyprus is considered among the top performing third party ship management centers in the European Union, employing approximately 55000 seafarers of various nationalities and managing more than 2200 ocean-going ships, corresponding to approximately 48 million gross tons (Cyprus Shipping Chamber, 2014). In addition, Cyprus is associated with the third largest ship registry in the European Union, in terms of tonnage, with more than 1000 ships.

In April 2015, we were requested by experts in Cyprus's shipping industry to analyze time-series data that concerned debit and credit transaction flows of resident shipping companies with non-residents. The time series covered the period from May 2004 to December 2014. Specifically, we were requested to recommend a methodology for estimating the long-run trend associated with these transaction flows and to examine possible extensions to forecasting. Working with shipping transaction flows is challenging because these time series are characterized by non-stationarity. Generally, a time series is considered as stationary when it has a stable mean, variance and

autocovariance in time. This is not the case with shipping flows that exhibit large changes in mean and variance.

In the shipping industry, trend and forecasting analyses are useful for monitoring performance and forming expectations regarding upcoming revenues and expenses during the budget planning process (Dickie, 2014). Because shipping markets are volatile, these analyses are also combined with expert opinion and market information derived from networks of professional contacts and associates. After budgets are finalized, trends and forecasts continue to be updated during the financial year as new data becomes available. This is necessary to evaluate progress with respect to specific targets (e.g., key performance indicators) and identify weaknesses in performance sufficiently early for corrective action.

To estimate the underlying trends in debit and credit flows, we used a relatively new but powerful tool for the analysis of time series, the Data-Driven Haar-Fisz Transform (DDHFT). In the remainder of this paper, we explain the procedures we used for estimating and predicting the trends and certain key properties of the DDHFT that are useful for managerial decisions.

Shipping flows

Shipping contributes to a country's BOP (Branch and Robarts, 2014, p. 473). Therefore, the credit (revenues) and debit (expenses) transactions of shipping companies based in Cyprus with non-resident companies are recorded in the Current Account of Cyprus's BOP. Each transaction recorded in the Current Account is assigned an economic activity code based on the classification system outlined in the most recent edition of the *Balance of Payments Vademecum Manual* published by Eurostat. Accordingly, the services component of the Current Account is published with a breakdown by economic activity. For the purpose of this study, we used aggregated, monthly, time series data for revenues and expenses associated with all shipping activities recorded in the Current Account. These data are used in official publications such as the *Ship Management Survey* published, semi-annually, by the Central Bank of Cyprus.

Credit flows consist primarily of revenues derived from the commercial operation of vessels and the provision of ship management services such as: chartering fees, freight rates, passenger revenues, technical management fees, crew management fees and brokerage commissions. Debit flows consist primarily of trading costs (e.g., bunker and port agency costs, brokerage

commissions, and survey fees), operating costs (e.g., crew expenses, repairs, victualing, and dry-docking) and financial costs (e.g., interest payments and loan installments for ship acquisitions).

Payments and receipts concerning shipping contracts do not always coincide with the time the service is provided. Contracts typically involve an initial, partial payment with the signing of the agreement, and the repayment of outstanding balances is usually completed gradually, throughout the financial year, with large payments occurring towards the end of the year. In addition, both revenues and expenses are subject to market risks related to the exchange rate, the freight rate, the interest rate and the bunker price volatility (Branch and Robarts, 2014, p. 459). Therefore, it is of interest for the shipping industry to estimate the underlying long-run trend of the series, which illustrates long-run movements in the volume of revenues and not simply short-term payments. The long-run trend better reflects the prospects of the industry and is useful for forecasting and planning.

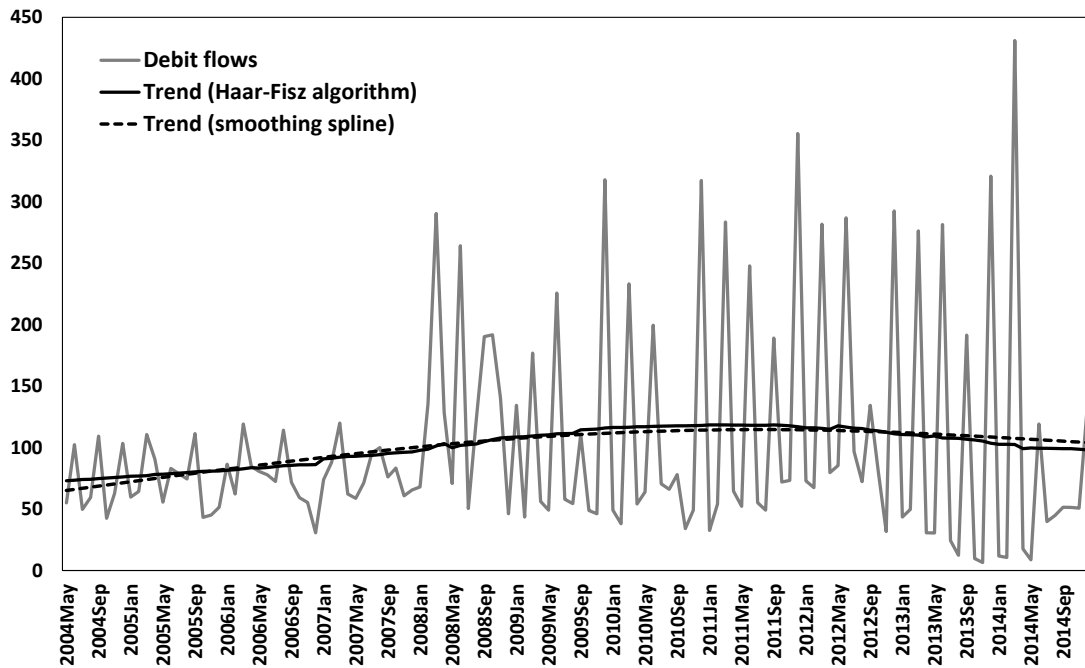


Figure 1: Actual debit flows (aggregated monthly expenses in euros for the period May 2004 - December 2014) of Cyprus's shipping industry and estimates of the trend generated with the DDHFT algorithm and a smoothing spline method. There is a change in the mean value of the series, beginning in 2008, when Cyprus joined the Eurozone, and a corresponding change in the variance of the series. The DDHFT algorithm is designed to model a time series in which the variance is a non-decreasing function of the mean. Both methods provided suitable estimates of the trend. However, the DDHFT algorithm better reflected a small increase in

the mean during the September 2009- March 2011 period, and a small decrease in the mean towards the end of the series.

As a result of these characteristics, shipping transaction flows are characterized by changes in mean and variance (volatility), and frequently, the variance is related to the mean of the series. To observe this, consider the debit flows of Cyprus’s shipping industry depicted in Figure 1. There is a change in the volatility of the series at the beginning of 2008, when Cyprus joined the Eurozone and adopted the euro as its official currency. There is also a change in the mean value of the series after 2008, in accordance with the change in volatility. The DDHFT algorithm presented in the next sections can effectively handle time-series with these non-stationary characteristics, particularly when the variance is a non-decreasing function of the mean of the series.

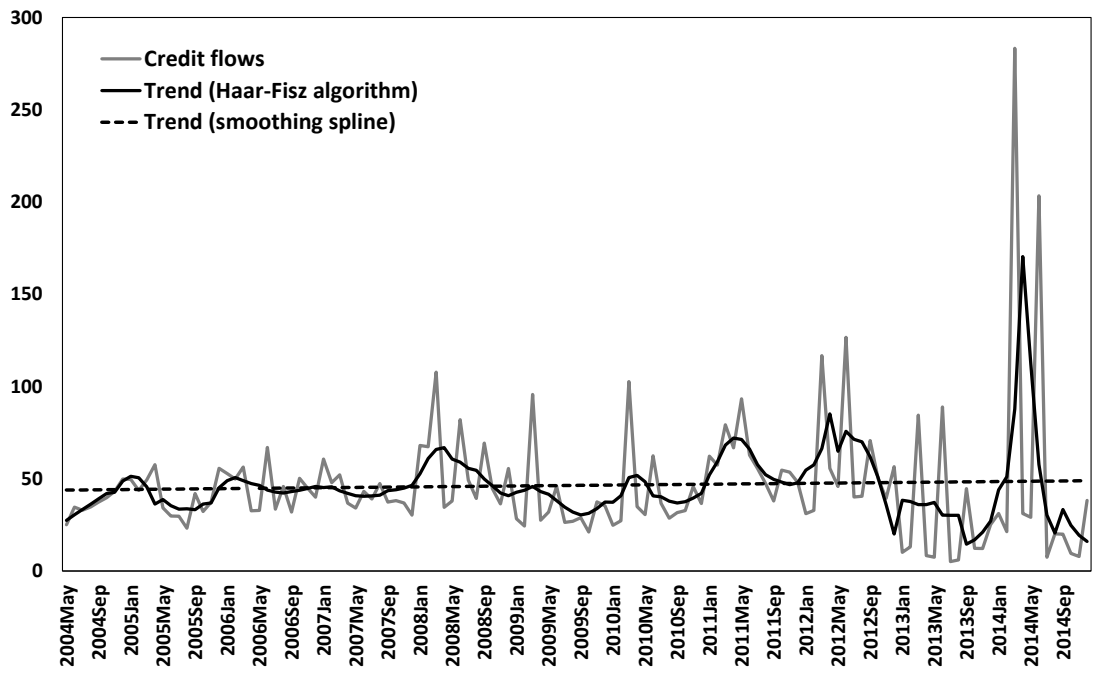


Figure 2: Actual credit flows (aggregated monthly revenues in euros for the May 2004 - Dec 2014 period) of Cyprus’s shipping industry and estimates of the trend generated with the DDHFT algorithm and a smoothing spline method. The DDHFT better reflected the local features of the data when compared with the smoothing spline method. Data characteristics such as cyclical changes, shifts in trend and spikes are not reflected in the smoothing spline estimate. This is most evident during the January 2013 – December 2014 period.

Although the main focus of the analysis is on trend estimation, it is also important to identify exceptionally large transactions in the series that are associated with large deviations from the trend. Such identification is useful either for further investigation of the specific transactions by experts in the industry, or for corrections in cases when large values correspond to classification errors. In contrast to other smoothing methods that tend to mask important characteristics of the time series, the DDHFT algorithm is sufficiently flexible to estimate the long-run trend of the series while, at the same time, capturing exceptionally large deviations from the trend and other local features of the data. This is illustrated in Figure 2. The estimated trend outlines long-run movements in the volume of credit transactions and captures large deviations from the trend and the repetitive, cyclical features of the data.

The DDHFT arises from the combination of the Fisz variance stabilization transform and the Haar wavelet transform to produce a powerful variance stabilizer that works at all scales and locations within a time series. We next review the Haar wavelet transform and then the details of the Data-Driven Haar-Fisz Transform.

The Haar wavelet transform

Because the actual data consists of highly volatile transaction records, a DDHFT was applied to stabilize the variance of the series and to estimate the underlying trend. The DDHFT combines two powerful mathematical methods, which we explain below: the Haar wavelet transform and the Fisz transform.

Wavelets constitute families of mathematical functions that enable the simultaneous representation of sequences of observations over different time-scales and locations. The basic forward wavelet transform extracts the multiscale information content from dyadic length time series. For Haar wavelets this is achieved by first calculating in a series of steps the differences as well as the sums between successive pairs of observations. All outcomes are multiplied by the normalization multiplier $a = 1/2$, which, in the second case, is equal to calculating average values of the observations.

First, the forward Haar wavelet transform starts by calculating the differences between non-overlapping consecutive pairs of observations in the actual data to produce the wavelet detail coefficients (Nason, 2008, p. 16). These coefficients contain information regarding the variability inherent in the series at the finest possible time-scale (J) and at different locations

(corresponding to each pair of observations in the series). Second, the transform moves to the next coarser time-scale ($J - 1$). This is achieved by calculating averages (not differences) of successive, non-overlapping pairs of observations in the actual data to produce the wavelet smoothed coefficients. By calculating the differences between successive pairs of smoothed coefficients (similar to step 1 above), a new set of detail coefficients is generated that contains information regarding the variability inherent in the series at the next coarser time-scale ($J - 1$).

The forward Haar wavelet transform proceeds in the same manner to calculate the detail and smoothed coefficients at successively coarser scales, and each time uses as inputs the smoothed coefficients of the previous finest scale. As a result of the differencing and averaging operations, the number of generated coefficients at each time-scale decreases by a factor of two. Nevertheless, this time-scale decomposition of the data provides a wealth of new information: each wavelet coefficient provides information concerning the variability of the original time series at a particular characteristic scale and location..

Another important property of wavelet transforms refers to their ability to reconstruct the original time series using the generated detail and smoothed coefficients in an inverse wavelet transform (Nason 2008, p. 54-55). The formulas for the inverse Haar wavelet transform have a similar form as the formulas for the forward transform (refer to the Appendix). In empirical studies, it is common to refer to the detail coefficients as mother wavelet coefficients and to the smoothed coefficients as father wavelet coefficients.

The Data-Driven Haar-Fisz transform

The Fisz (1956) transform is an old variance stabilization technique in which, generally, stabilization is achieved by dividing the difference between two suitable random variables with some function of their sum. The Fisz technique can also be used with the father and mother wavelet coefficients, simply because these coefficients are generated using difference and summation operations identical to those required by the Fisz transform.

For the purposes of this study, consider a sequence of observations of dyadic length with a positive mean and variance in which the variance is a non-decreasing function (h) of the mean. We can represent the mean variance relation as follows:

$$variance = h(mean) .$$

Fryzlewicz and Dellouille (2005) and Fryzlewicz et al. (2007) developed the DDHFT, which stabilizes the variance of a time series based on the following four-stage procedure:

Stage 1: A forward Haar wavelet transform is performed on the time series to generate the father and mother wavelet coefficients as described in the previous section.

Stage 2: The mean-variance relation (h) described above, is estimated with a monotonic regression method using local variance and mean estimates that are based on the finest-scale wavelet coefficients (refer to the Appendix).

Stage 3: All father and mother wavelet coefficients generated in Stage 1 are used to form, for each time-scale and location, the respective Fisz coefficients based on the following transform using the mean-variance relation estimates (\hat{h}) from stage 2:

$$Fisz\ coefficient\ ts_{scale,location} = \frac{mother\ coefficient\ ts_{scale,location}}{\hat{h}(mean)^{1/2} \cdot father\ coefficient\ ts_{scale,location}}.$$

Stage 4: An inverse Haar wavelet transform is performed on the Fisz coefficients generated in Stage 3.

Figure 3 illustrates the workflow associated with this four-stage procedure. The procedure generates a new sequence of observations with a stable variance that can be used for forecasting purposes. More importantly, the DDHFT can be used for variables whose distribution is not known a priori, as is usually the case in practice. Further technical details concerning this four-stage procedure are provided in the Appendix. For the analysis of the shipping flow data, we used the DDHFT algorithm described in Motakis et al. (2006). The algorithm can be performed automatically in the R software environment using the “DDHFm” package (available at: <http://cran.r-project.org/web/packages/DDHFm/>) and the commands `ddhft.np.2` and `ddhft.np.inv` therein. To provide mean estimates we apply spline smoothing to the results of the DDHFT before inversion.

To facilitate use by industry experts, these commands have been incorporated into one common R-code that also includes the forecasting command *ets* presented in the next section. This code starts by uploading the aggregated, monthly, time series data associated with debit and credit transaction flows, before proceeding with automatic trend estimation and prediction. The time series data are extracted from a transactions data warehouse that is used to compile the Current Account of Cyprus’s BOP. The graphical output generated by the code is identical to that presented in Figures 1, 2 and 4.



Figure 3: Four stages of the DDHFT algorithm. These stages can be performed automatically in the R software environment using the commands *ddhft.np.2* and *ddhft.np.inv*, which are available in the “DDHFm” package. The DDHFT algorithm generates a new sequence of observations with a stable variance.

Estimation and prediction of shipping trends

Figure 1 includes trend estimates for debit flows using two methods for comparison purposes: the DDHFT algorithm and a smoothing spline method. Both methods provided suitable estimates; however, the DDHFT algorithm is more flexible and better reflects small changes in the mean value of the series. This is evident during the September 2009 – March 2011 (increase in mean) and March 2014 - December 2014 (decrease in mean) periods.

The two methods are also compared in Figure 2, which includes trend estimates for credit flows. The DDHFT algorithm provided superior estimates of the trend in this case. It effectively stabilized the variance of the series but was more accurate in capturing data characteristics such as cyclical changes, shifts in the trend and spikes that are not reflected in the spline smoothing estimate. This is because the DDHFT algorithm is based on wavelet decompositions of the data

that enable the identification of such local characteristics in a time series (refer to Nason, 2008, p. 219).

The DDHFT algorithm provided better behaved estimates of the trend during the months, January – September (particularly for years 2008, 2011 and 2012), when the variance of the series was changing. In addition, the DDHFT algorithm estimate in Figure 2 suggests that the level of credit flows declined towards the end of 2012 (shift in trend), with the exception of certain very large transactions (spikes) that were recorded during the December 2013 - March 2014 period. These data characteristics are not reflected in the spline smoothing estimate.

Having estimated the trends, a logical next step in the analysis is to perform forecasts, which is an important aspect of budgetary control and planning in the shipping industry (Branch and Robarts, 2014, p. 455). The information included in Figures 1 and 2 is updated with semi-annual frequency, which coincides with strategic reviews and other corporate reporting in the industry. Therefore, we were requested by industry experts to augment the DDHFT algorithm with an automatic forecasting procedure and to generate forecasts 10 months ahead. This short-term forecasting procedure will be used every time new data becomes available, and the trend estimates are updated.

Shipping markets are very complex. The main purpose of forecasting in this context is to provide decision makers with useful information to reduce uncertainty and effectively monitor change in fast-moving and volatile environments (Stopford, 2009, p. 701). The results of forecasting models are commonly used as a basis for discussion at board meetings and are always augmented with market intelligence and expert opinions to form reasoned views regarding the industry. The forecasts described in this section are used by professionals in Cyprus's shipping industry for two purposes: (i) budgeting and (ii) strategic planning. In budgeting, the credit and debit forecasts are used as estimates of the expected revenues and expenses for the following year when deciding the allocation of financial resources. With regards to strategic planning, forecasts are used as a basis for discussing the prospects of the industry and for developing business plans with specific policy actions. Frequently, forecasts are also discussed during negotiations with trade unions and the government, whose policies influence the prospects of the shipping industry.

To extend the previous analysis to forecasting, we inserted the trend estimates generated by the DDHFT algorithm into the *ets* command that is included in the 'forecast' package developed by Rob J. Hyndman at Monash University (<http://cran.r->

project.org/web/packages/forecast/). This package operates within the R software environment, and the *ets* command offers a fully automatic procedure for fitting exponential smoothing state space models (including model selection). In addition, this package has performed well in forecasting competitions (see, Hyndman et. al. 2002) and generates the graphical output presented in Figure 4.

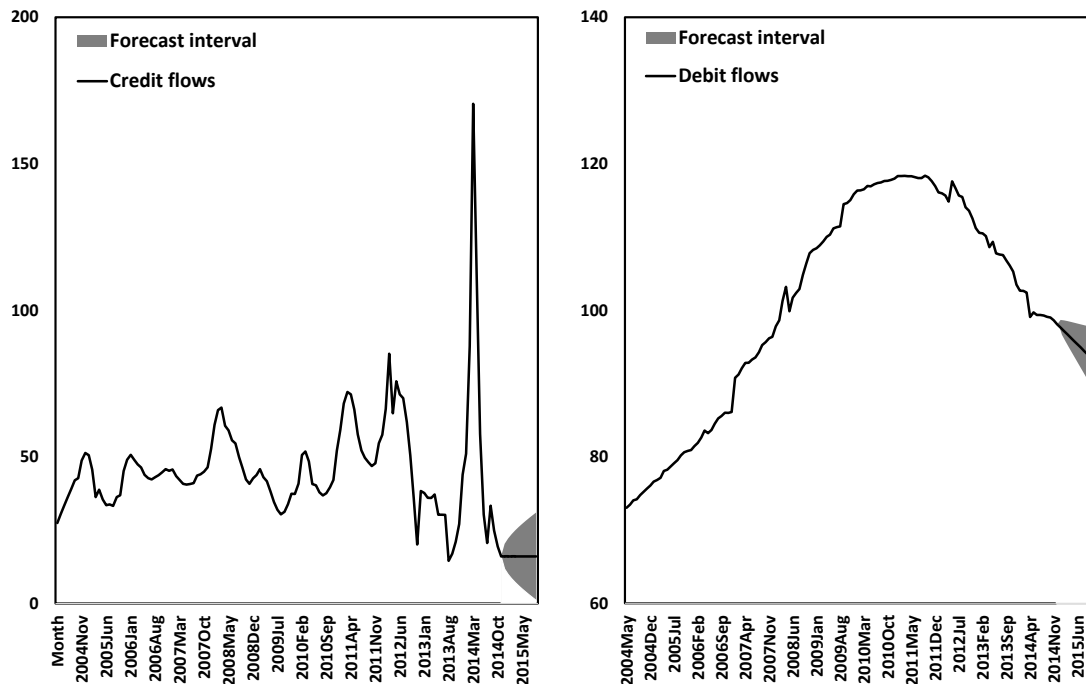


Figure 4: Forecasts of debit and credit trends, 10-months ahead, generated with the *ets* command that is included in the ‘forecast’ package. In both cases, the forecasting horizon covered the period from January 2015 to October 2015. The shaded areas around the forecasted values represent the estimated forecast intervals that are expected to include the future observations with 80% probability.

Figure 4 includes 10-month ahead forecasts and 80% forecast intervals for the debit and credit trends, as generated by the *ets* command. Wider intervals were generated in the case of credit flows because the respective trend is associated with higher volatility. These forecasts can also be used in forecasting combinations (see Michis 2014) or other more computationally intensive smoothing methods (see Michis 2015) that can potentially improve forecasting performance further.

Conclusions

Transactions in the shipping industry are highly irregular in nature and do not always coincide with the time the service is provided. The repayment of outstanding contracts is usually completed gradually, throughout the financial year, with occasional large payments for the settlement of outstanding balances. Consequently, shipping transaction flows are characterized by changes in the mean and the variance, and frequently, the variance is a non-decreasing function of the mean. This complicates the estimation of the long-run trend of the series, which is useful for identifying long-run movements in the volume of transaction revenues (or expenses) and not solely short-term payments.

In the previous sections, we described the implementation of a computer-based automatic procedure to estimate the trends associated with shipping debit and credit transaction flows in Cyprus' shipping industry. The procedure is based on the DDHFT algorithm and includes a variance stabilization method that provides accurate estimates of the trend in cases when the variance is a non-decreasing function of the mean of the series. The procedure is also more accurate in capturing data characteristics such as cyclical changes, shifts in trend and spikes that are usually not reflected in other methods for estimating trends.

In addition, the proposed procedure does not require any prior knowledge of the distribution of the variables of interest, and the trend estimates can be readily used in forecasting applications. The procedure can be easily implemented in the R software environment using the “DDHFm” and ‘forecast’ packages, with the sole input requirement being the actual time series data.

Appendix

In accordance with Fryzlewicz and Nason (2004) or Nason (2008, p. 209-210), the four-stage procedure associated with the DDHFT involves the following computations:

Stage 1: Forward Haar wavelet transform

Let $c^j = [X_1, \dots, X_n]$ be a time series with n observations. The forward Haar wavelet transform proceeds to generate the father (c) and mother (d) wavelet coefficients for each time scale ($j = 1, \dots, J$) and location ($k = 1, \dots, 2^{j-1}$) as follows:

$$c_k^{j-1} = (c_{2k-1}^j + c_{2k}^j)/2 \quad \text{and} \quad d_k^{j-1} = (c_{2k-1}^j - c_{2k}^j)/2.$$

Stage 2: Regression estimation of the mean-variance relation

To estimate the Fisz coefficients, first, it is necessary to obtain an estimate of the non-decreasing mean-variance relation (h), as suggested by the following nonparametric regression model

$$\sigma_i^2 = h(\mu_i) + \varepsilon_i.$$

Because the local means (μ_i) and variances (σ_i^2) are not known a priori, Fryzlewicz et al. (2007) suggested the following estimators based on the finest scale wavelet coefficients and subject to certain suitable smoothness assumptions:

$$\mu_i = (X_i + X_{i+1})/2 \quad \text{and} \quad \sigma_i^2 = (X_i - X_{i+1})^2 / 2.$$

The generated values can be used to fit a monotonic regression model to obtain an estimate of the mean-variance relation h .

Stage 3: Fisz transform

Using the father and mother wavelet coefficients generated in Stage 1 and the estimate of the mean-variance relation generated in stage 2 ($\hat{h}(\mu)$), the Fisz coefficients (f) can be formed using the following function

$$f_k^j = \frac{d_k^j}{\hat{h}(\mu)^{1/2} \cdot c_k^j}.$$

Stage 4: Inverse Haar wavelet transform

For all time-scales $j = 1, \dots, J$ and locations $k = 1, \dots, 2^j$, the inverse Haar wavelet transform generates a final variance-stabilized vector c_k^j , with elements calculated as follows:

$$c_{2k-1}^j = c_k^{j-1} + f_k^{j-1} \quad \text{and} \quad c_{2k}^j = c_k^{j-1} - f_k^{j-1}.$$

References

Branch, A.E. and M. Robarts (2014) *Branch's elements of shipping*, 9th ed., Routledge: Abingdon.

Cyprus Shipping Chamber (2014) *Annual Report: 2013*, Limassol, Cyprus.

Fisz, M. (1955) The limiting distribution of a function of two independent random variables and its statistical application, *Colloquium Mathematicum*, 3, 138-146.

Fryzlewicz, P. and V. Delouille (2005) "A data-driven Haar-Fisz transform for multiscale variance stabilization", Proceedings of the 13th IEEE/SP Workshop on Statistical Signal Processing, Bordeaux, pp. 539-544.

Fryzlewicz, P., Delouille, V. and G.P. Nason (2007) "GOES-8 X-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform", *Journal of the Royal Statistical Society Series C*, 56, 99-116.

Fryzlewicz, P. and G.P. Nason(2004) A Haar-Fisz algorithm for Poisson intensity estimation, *Journal of Computational and Graphical Statistics*, 13, 621-638.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., and S. Grose (2002) "A state space framework for automatic forecasting using exponential smoothing methods", *International Journal of Forecasting*, 18(3), 439–454.

Michis, A.A. (2015) "A wavelet smoothing method to improve conditional sales forecasting", *Journal of the Operational Research Society*, 66, 832–844.

Michis, A.A. (2013) "Do more forecasts always improve averages of model forecasts? Evidence from reports published by HM Treasury", *OR Insight*, 26, 234-241.

Motakis, E.S., Nason, G.P., Fryzlewicz, P. and G.A. Rutter (2006) “Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach”, *Bioinformatics*, 22, 2547-2553.

Nason, G.P. (2008) *Wavelet methods in statistics with R*, Springer: New York.

Stopford, M. (2009) *Maritime economics*, 3rd edition, Routledge: Abingdon.

Dickie, J. W. (2014) *Reeds 21st century ship management*, Bloomsbury Publishing: London.