Forecasting Australian port throughput: Lessons and Pitfalls in the era of Big Data

Yuriy Tyshetskiy, Soumya Banerjee, George Mathews, Thomas Vitsounis*

Data61 | CSIRO, Australia

***Corresponding author. Thomas.Vitsounis@data61.csiro.au**

Abstract

Modelling and forecasting port throughput enables stakeholders to make efficient decisions ranging from management of port development, to infrastructure investments, operational restructuring and tariffs policy. Accurate forecasting of port throughput is also critical for long-term resource allocation and short-term strategic planning. In turn, efficient decision-making enhances the competitiveness of a port. However, in the era of big data we are faced with the enviable dilemma of having too much information. We pose the question: is more information always better for forecasting? We suggest that more information comes at the cost of more parameters of the forecasting model that need to be estimated. We compare multiple forecasting models of varying degrees of complexity and quantify the effect of the amount of data on model forecasting accuracy. Our methodology serves as a guideline for practitioners in this field. We also enjoin caution that even in the era of big data more information may not always be better. It would be advisable for analysts to weigh the costs of adding more data: the ultimate decision would depend on the problem, amount of data and the kind of models being used.

Keywords: Australia port throughput, forecasting, big data, machine learning.

1. Introduction

Seaports are the main gateways for the flow of freight traded between countries. Reliable forecasts of port activity are highly desirable both at a strategic and operational level. Governments, planning agencies, regulators and port authorities use forecasts to develop efficient future strategies (national port policies, pursue free trade agreements, etc.), increase attractiveness of ports, outperform competition and decide on future investment plans. Terminal operators, transport companies (shipping, trucking, rail) and numerous port stakeholders (freight forwarders etc.) are also in need of reliable forecasts to guide future investments, and design and deploy strategies to increase their market share and profitability.

Forecasting is a complex process aimed at predicting future behaviour through modelling past data. Naturally, the success of forecasting is highly dependent on: a) the characteristics of the specific system, b) the available data, and c) the models used. A variety of techniques have been applied to forecast port throughput and we briefly review these techniques in Section 2. In general terms, the majority of existing forecasting applications in port throughput employ univariate or multi-variate forecasting methods.

In terms of data, univariate forecasting uses mainly port throughput data while multivariate models are usually enhanced with economic indicators (with gross domestic product being the dominant type of indicators used). In general, data availability is a major constraint of port throughput forecasting models since port activity data are considered commercial and kept confidential or simply not kept at all. However, this will change in the near future due to the overall trend towards "open data", the expansion of the Internet of Things and the overall contemporary trend of capturing and storing every possible data available.

We suggest that selection of appropriate data and models in forecasting ports throughput deserves great attention especially since relevant studies are quite limited. We challenge the logic "the more data the better" and aim to develop a framework enabling forecasters to make better decisions about what models to use. We shed light on the following questions: a) How many data points and how many time series or indicators are necessary to make good predictions? b) Is more information always better? More information comes at a cost since potentially more parameters encapsulating the interactions between multiple time series have to be estimated. We note however that merely the number of time points is not an adequate measure and the amount of data required to adequately learn a model is very model-, problem- and data-specific.

We compare a suite of techniques to forecast Australian imports using openly available and commercial data. We investigate if more complex models incorporating commonly used socio-economic indicators such as gross domestic product (GDP) produce more accurate forecasts of Australian port imports.

First, we suggest a computational framework and methodology to guide port throughput forecasting exercises. Second, we make recommendations on how to determine the best performing forecasting methodology based on the amount of available data. Our results provide a guide to forecasters and challenge the logic "the more data the better the outcome".

The remainder of the paper proceeds as follows: Section 2 reviews the literature on forecasting port throughput. Section 3 describes the data and the applied methodology. Section 4 presents the results of the different forecasting models for Australian port imports and discusses the amount of data required by these methods. Finally Section 5 provides concluding remarks.

2. Related Work

Port throughput projection is an emerging discipline in maritime studies. A variety of methodologies and techniques are employed in an effort to increase the predictive power of the forecasting analysis (Pallis et al., 2011). The majority of port throughput forecasting studies found in the relevant literature employ: 1) univariate methods, and 2) multivariate methods.

2.1. Univariate Methods

In univariate time series forecasting, the past values of the quantity of interest, or "target variable" are used to predict future values with dependence on trends, seasonal cycles and irregularities. The techniques that have been applied are statistical autoregressive methods such as Auto-Regressive Integrated Moving Average (ARIMA) (Kim et al., 2011), moving averages, seasonal decomposition and exponential smoothing models (Abraham and Ledolter, 2009; McCarthy et al., 2006). On a regional level, Maloni and Jackson (2005) forecast container flows in North America using the exponential smoothing methodology. On a port level, scholars have also applied Seasonal Auto-Regressive Integrated Moving Average (SARIMA) to examine trans-shipment flows (Schulze and Prinze, 2009).

2.2. Multivariate Methods

More complex multivariate models encapsulate information about the target variable and its interaction with other predictors or "indicator variables". Recent studies have used Multivariable Adaptive Regression Splines (MARS), Dynamic Factor Models (DFM) and Auto-Regressive Integrated Moving Average with eXogenous variables (ARIMAX) with DFM to forecast macroeconomic and seaport data (Geng et al., 2015; Angelopoulos and Chlomoudis, 2015, Intihar et al., 2015).

Vector Auto-Regressive Moving Average (VARMA) have been used to forecast commodity (steel) volumes (Gooijer and Klein, 1989) and chaotic models have been used to predict future container flows (Goulielmos and Kaselimi, 2011). To forecast port throughput scholars also use Grey models (Guo et al., 2005) and neural network models (Mostafa 2004; Li et al., 2008; Chen and Chen, 2010).

Multivariate cause-and-effect methods model the relationship between the forecasted variable and other variables (Anderson, et al., 2009). Techniques falling within this category are multivariate regression analysis, econometric modeling and input-output models (Hanke et al, 2001; Abraham and Ledolter, 2005). Cause-and-effect models aim to interpret the relationship between port throughput and certain economic determinants before proceeding to future projections. Port throughput is considered to be influenced by: a) fundamental macro-economic determinants of the country the port is located in such as economic activity, international and maritime trade, and b) micro-economic variables like generalized costs of the logistic chain and cost of fuel (Meersman, 2009). Generally on a macro level, GDP and trade by value are systematically used as the main driving factors of port throughput (on both country and port levels).

Recent studies also suggest that additional macroeconomic variables such as industrial production and financial determinants have the potential to lead to better forecasting (Chou et al., 2008; Gosasang et al., 2011; Paflioti et al., 2015). On a micro level, qualitative and quantitative characteristics of ports such as port tariffs (Fung, 2002), generalized costs (De Langen et al., 2012), number of berths (Hui et al., 2004) or even expenditure on building and construction (Seabrook et al., 2003) are taken into account for throughput forecasting. The majority of the causal and effect studies employ Ordinary Least Squares (OLS) (Seabrook et al., 2003; Lehto et al., 2006), or Error Correction Model (Fung, 2002; Hui et al., 2004, etc.), to reveal the causal relations between the target and the indicator variables. Finally, Van Dorsser et al., (2012) and De Langen et al., (2012) apply a combination of System Dynamic Modeling, judgment, causal relations, commodity specific research and freight transport modeling to estimate ports throughput in the La Havre-Hamburg range.

3. Methodology

In this section we outline the different datasets we use and our computational methodology. We combine both commercial (Australian imports) and openly available data (GDPs of different countries). We separate the data into a training set to fit our models and a testing set to determine forecasting accuracy. After subjecting the data to various transformations we fit models of varying degrees of complexity and select the best performing models.

3.1. Data

The data used for this study is based on detailed imports information for Australian seaports. The data are from the Australian Bureau of Statistics derived from Trade Data International Ltd. and are aggregated over each quarter from 1995:Q1 - 2014:Q3. The total value of the goods has been aggregated over all ports and commodities (imports by sea), resulting in the single time-series shown in Figure 1.

To help model and forecast this data, supporting economic indicators are also obtained. This is in the form of the GDP of Australia and its five largest trading partners in terms of imports. We aggregated the value of all imports to Australia from each country over the period 1995 to 2014. It was found that the top 5 countries consisted of China, United States, Japan, Singapore and Germany; these countries' GDPs were chosen for the subsequent analysis. GDP data was acquired from the World Bank (World Bank GDP Data, 2016) for the period 1995-2014. The data was sampled at a yearly frequency, and adjusted for purchasing power parity in constant 2011 international dollars (which has the same purchasing power over GDP as the U.S. dollar has in the United States).

We combine the data on port imports and GDP of countries into a multivariate time series.

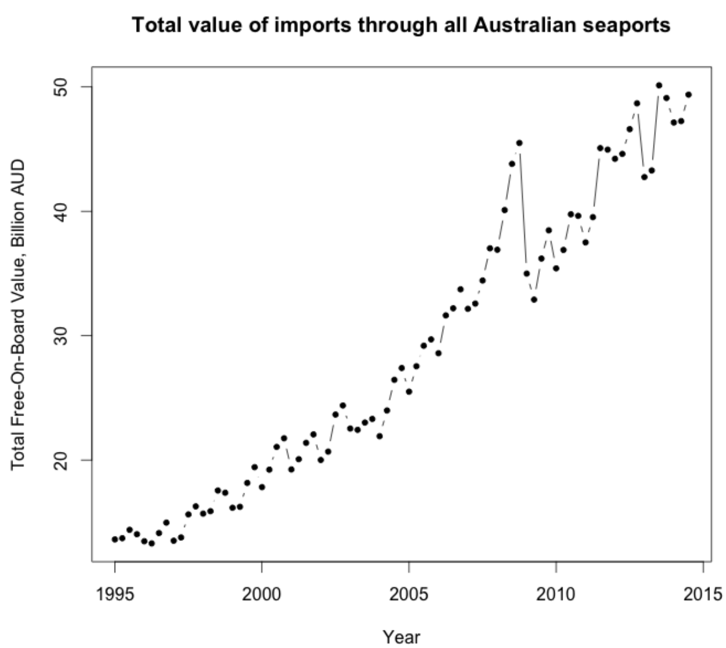

Figure 1. Total Free-On-Board value of Australian imports by sea through all ports.

3.1.1. Separation into training and testing set

We separate the multivariate time series data into a training set (from Q1 1995 till Q4 2007) and testing set (from Q1 2008 till Q3 2014). The training data is used to fit forecasting models. Each of the fitted models is then used to generate a forecast over the testing period (Q1 2008 till Q3 2014). The accuracy of the model forecast is then evaluated by comparing the model predicted values to the actual values by accuracy metrics such as out-of-sample root mean squared error (RMSE) on the test set.

3.2. Modeling Approach

3.2.1. Summary of Computational Framework

Our modelling framework involves the following steps:

1) Exploring a range of transformations to pre-process the data (Section 3.2.2).
2) Fitting models of varying degrees of complexity and number of predictors, and performing model selection (Section 3.2.4).
3) Using synthetic datasets to assess whether there is sufficient data to learn these models (Section 3.2.6).

3.2.2. Data Preprocessing

The statistical forecasting models employed in this work (detailed in the next subsection), assume that the system generating the data is stationary, and the noise component of the time series is white noise. To ensure these assumptions hold and to ensure the positivity of quantities, pre-transforming the data is usually necessary (Chatfield, 2013). The main transforms considered in this work are:

- Log transform: This transform is typically used in econometrics to ensure positive values, and allow larger variations when the value is higher (an aspect that is present in the throughput data displayed in Figure 1).
- Interpolation: Increases the temporal resolution of the GDP data by linearly interpolating from yearly data to a quarterly period.
- Temporal differencing: Converts the data into increments per time unit, used to remove non-stationarity.
- Standardizing (z-scoring): This shifts and scales the data to produce data that has zero mean and unit standard deviation.

In this paper three different options for pre-transforming the data are considered: (i) no transformations are applied at all, (ii) interpolation is applied first, followed by temporal differencing, followed by standardization (iii) log transform is applied first, followed by the sequence (ii) of transforms.

3.2.3. Model Definitions

We consider two general classes of models, described below.

Seasonal autoregressive integrated moving average (SARIMA) model:
This is a univariate model, which models and forecasts the Australian port throughput using only its own past (lagged) values. This model is a generalisation of a standard auto-regressive moving average model. For a given maximum autoregressive lag p and moving average lag q, an ARMA model is defined by the equation:

$$y_t = C + \sum_{i=1}^{p} \phi_i \cdot y_{t-i} - \sum_{i=1}^{q} \theta_i \cdot \epsilon_{t-i}$$

Here, y_t denotes the port throughput at time t and C, ϕ_i and θ_i are the coefficient parameters of the model. The stochastic residual or error terms ϵ_t are assumed to be a zero mean white noise sequence, with a constant variance denoted by the parameter Σ. The coefficients C, ϕ_i, θ_i and variance Σ parameters must be estimated from the historic training data, further details are given in the next section. The addition of an integrated (I) component introduces an additional temporal differencing stage to the data. The order of this differencing is denoted by d. Furthermore, the Seasonal ARIMA model has an additional seasonal model component of length m which includes separate AR, MA and I components, each with lags denoted by P, Q and D respectively. The full model structure is denoted by SARIMA$(p,d,q)(P,D,Q)_m$. Full details of the model can be found in Asteriou and Hall (2011).

Vector Auto-Regressive (VAR) model:
This is a multivariate model, which is capable of modelling the joint dependencies between the throughput and the supporting GDP data, and uses these dependencies to forecast the imports along with the GDP in the future.
A p-th order VAR(p) is represented by the following equation:

$$\mathbf{y_t} = \mathbf{c} + \sum_{i=1}^{p} A_i \cdot \mathbf{y_{t-i}} + \mathbf{\epsilon_t}$$

The variables (port throughput and GDP of all countries) are subsumed in the vector y_t, and y_{t-i} is the ith lag of y_t. The coefficient matrices A_i are time-invariant and represent a set of model parameters, ϵ is a vector of error terms with mean 0 and covariance Σ, and c is a vector of constant intercept terms. Fitting the VAR model involves estimating the matrix of interactions A_i, vector c and the covariance matrix Σusing the training data.
In this work, two different types of VAR model classes are considered: a two dimensional class which incorporates the imports time series, and the GDP of Australia only; and a seven dimensional model class that in addition to the port throughput data and Australian GDP

also includes the GDPs of Australia's five largest trading partners in terms of value of imports. These models will be referred to as VAR2 and VAR7 respectively.

3.2.4. Parameter Estimation and Model Selection

Here we outline the sequence of steps in estimating the parameters of the models and selecting the best performing models from a set of model classes and model structures. We consider three model classes: SARIMA, VAR2 and VAR7. For each model class, a candidate model structure can be considered that specifies p for VAR and p, d, q, P, D, Q, m for SARIMA. The performance of a candidate model structure is quantified on the training data by:

1) Determining the best-fit model parameters (coefficients and noise variance) by solving a maximum likelihood optimisation problem based on the training data.
2) Checking that the residual errors of the model fit to data are normal using a Portmanteau Ljung-Box test (Ljung and Box, 1978). Model structures that fail normality tests are rejected.
3) Quantifying the performance of the model on the training data using the Akaike Information Criterion (AIC). This balances model complexity with the ability of the model to fit the data (Brockwell and Davis, 1996).

This procedure has been used to determine the best model structure for each of the three model classes. It is noted that the use of the AIC also allows a comparison between the best performing model structure of each of the three model classes and an overall best performing model determined.

It is noted that the rejection step in 2) is required as we observed that some VAR model structures with low AIC scores had residuals which differed significantly from a normal distribution (these models also had very high lags).

3.2.5. Forecasting

After the models were fit to the historical data, forecasts were generated by rolling the models forward in time. To enable the effects of the unknown future disturbances (error terms) to be explicitly included, the mean and covariance of the forecast's probability distribution was generated. This was performed using the estimated covariance matrix of the model error terms. Lastly, the effects of any data transforms applied in the pre-processing stage were reversed. As these transforms are in general nonlinear, this was performed by first extracting the median forecast and the upper and lower 5 and 20 percentiles of the distribution. Each of these forecast time series was then untransformed separately.

3.2.6. Synthetic Data Generation

We generated synthetic time series data from a data-generating model. The synthetic data-generating model was the VAR7 model with 6 additional predictors (GDPs of Australia and of top-5 importers to Australia) fit to the actual training data of port throughput. This synthetic data was split into a training set and a held-out test set. Then we trained three forecasting models - a univariate ARIMA model, a multivariate 2-dimensional VAR2 model, and a multivariate 7-dimensional VAR7 model on training subsets of different lengths. These trained models were then used to forecast the target variable over the held-out test set.

3.3. Software

All our computational procedures were implemented in the R programming language (version 3.2.2) (R Core Team, 2015). For SARIMA model selection, parameter estimation and forecasting, we used the functions "auto.arima", "Arima" and "forecast" from the "forecast" R package. For VAR model selection and fitting, we used the functions "VARSelect" and "VAR" from the "vars" R package, and for forecasting with VAR models we used the "forecast" function from the "forecast" package. Finally, we implemented our own synthetic data generator by VAR models, and a function that generates the learning curves shown in Figure 5.

4. Results and Discussion

4.1. Forecasts

We compared the forecasts of three models: 1) a univariate SARIMA model that only relies on lagged values of the quantity of interest (Imports) for forecasting, and two multivariate models: 2) the 2-dimensional VAR2 model that incorporates the GDP of Australia as an additional predictor, and 3) the 7-dimensional VAR7 model that incorporates GDPs of Australia and the top-5 importers to Australia as additional predictors.

The forecasts of Australian imports for the test set, made by each of these three models over the period Q1 2008 to Q3 2014, using the information available from the training data, are shown in the plots below in Figures 2-4. (The training data has been pre-transformed using interpolation of the yearly GDPs to quarterly GDPs, followed by temporal differencing and standardization of all the variables.) We observe that the trajectory of the actual imports values (test set) lies well within the estimated uncertainty ranges (the 80% and 95% confidence intervals) of the forecasts, for all three models, i.e., the estimated uncertainties of the forecasts appear adequate.

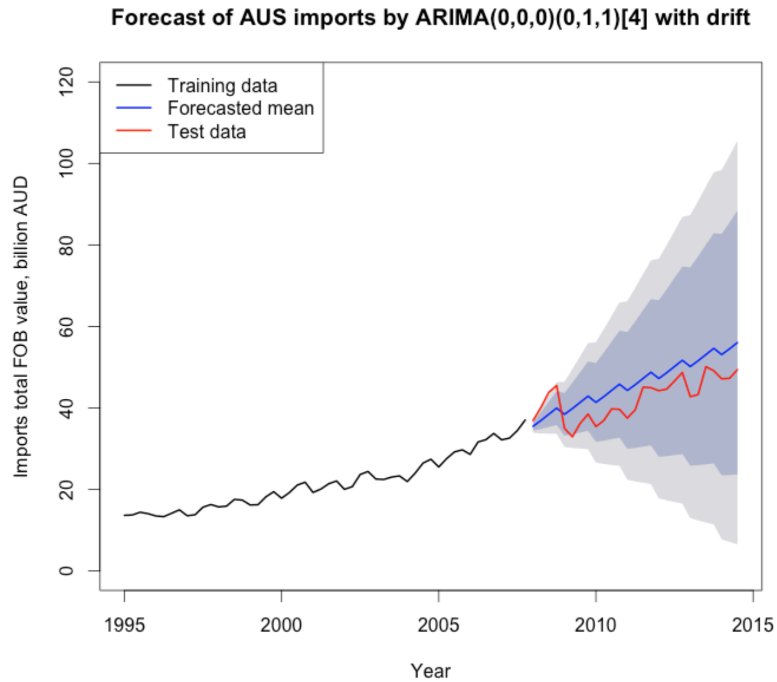

Figure 2. Forecast of port throughput (imports) by best fitting Seasonal ARIMA (SARIMA) model. In-sample (training set) AIC=37.66, out-of-sample (test set) Root Mean Squared Error RMSE=10190.

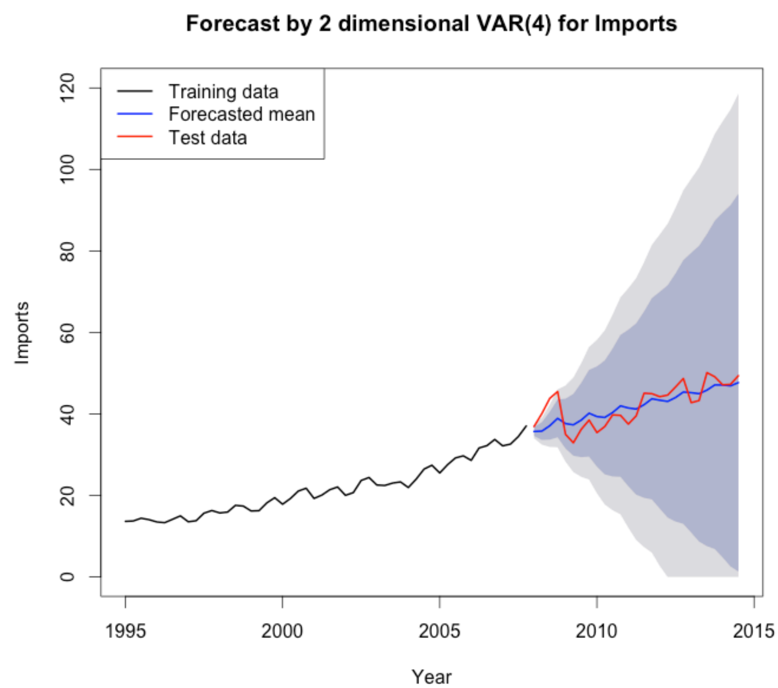

Figure 3. Forecast of port throughput by best fitting 2D VAR model (using GDP of Australia as an additional predictor of imports). In-sample (training set) AIC=141.95, out-of-sample (test set) Root Mean Squared Error RMSE=3057.

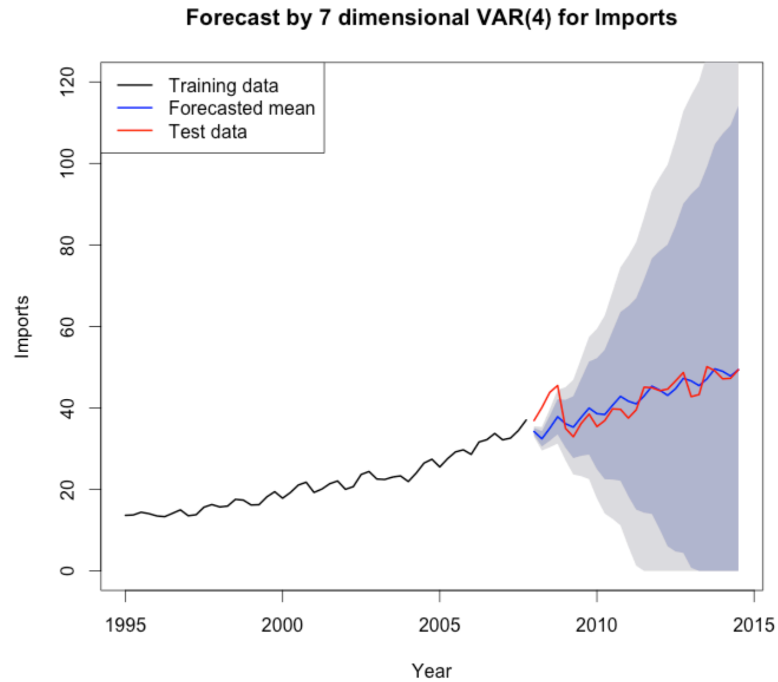

Figure 4. Forecast of port throughput by best fitting 7D VAR model (using GDPs of Australia and of top-5 importers as additional predictors of imports). In-sample (training set) AIC=37.73, out-of-sample (test set) Root Mean Squared Error RMSE=3343.

4.2 Effect of including additional predictors

Comparing the out-of-sample (test set) RMSE of the three forecasts, we see that adding Australian GDP as a predictor of imports greatly improves the forecast accuracy (compare Figure 2 and Figure 3); this improvement is in favor of the "more data the better" intuition. However, counter-intuitively, adding even more additional predictors (GDPs of top-5 importers to Australia in this particular case) does not necessarily lead to further improvement of the forecasting model, and in fact may even have an adverse effect, as seen from Figures 3-4 and from Table 1. We discuss possible reasons for this in Section 4.4.

Table 1. Metrics of model quality (in-sample Akaike Information Criterion) and of forecast accuracy (out-of-sample RMSE) for different forecasting models with and without data pre-processing transformations. Highlighted are cells corresponding to best models out of three, according to AIC or RMSE.

in-sample AIC, out-of-sample RMSE	No transformations applied to data		Transformations Interpolation → temporal differencing (lag 1) → z-score		Transformations Log10 → interpolation → temporal differencing (lag 1) → z-score	
	AIC	**RMSE**	**AIC**	**RMSE**	**AIC**	**RMSE**
SARIMA	768.3	5271	37.66	10190	70.09	19223
VAR2	2921.0	8358	141.95	3057	172.95	4486
VAR7	14094.5	12561	37.73	3343	133.3	7090

4.3 Effect of data transformations

We observe that performing pre-processing transformations on the data also affects the accuracy of the forecasts, as seen by comparing RMSE values in the three columns of Table 1. From the nine combinations of data transforms and model classes shown in Table 1, the best forecast (in terms of out-of-sample RMSE) is produced by a 2-dimensional VAR2(4) model that uses GDP of Australia as an additional predictor of imports. This model was trained on pre-processed time series data with interpolation from yearly to quarterly GDP applied, followed by time differencing with lag 1 of both imports value and GDP, and finally by standardizing the data. Applying the log transform to ensure positivity of the forecasted quantities has an adverse effect on the accuracy (RMSE of 4486 compared to RMSE of 3057 if no log transform is applied).

Thus we see that the choice of transforms applied to data during pre-processing, as well as the choice of the class and structure of the forecasting model, all have a significant effect on the quality (out-of-sample accuracy) of the resulting forecasts.

4.4. Is more data always better?

We have seen that including additional indicator variables (GDPs of top-5 importers in our case) can have an adverse effect on forecast accuracy, contrary to the “big data” intuition of “the more data (predictors) we use for forecasting, the better the forecast will be”. To explore why simpler models (with fewer additional predictors) can perform better than more complex models (with more additional predictors), we performed a series of additional experiments on synthetic data.

We generated a long synthetic time series data from a fixed 7-dimensional multivariate model (VAR7) with 6 additional predictors (GDPs of Australia and of top-5 importers to Australia) of port throughput. The synthetic data-generating model was the VAR7 model fit to the actual training data; hence the resulting synthetic data has similar statistical properties to the actual data (but the synthetic data is much longer in duration). This synthetic multivariate time series was split into a training set and a held-out test set. The training set was then divided into smaller training subsets of different lengths, all subsets having the same endpoint in time. Then we trained three forecasting models - a univariate ARIMA model, a multivariate 2-dimensional VAR2 model, and a multivariate 7-dimensional VAR7 model (the last model is of the same class and structure as the model used to generate the synthetic data) - on these training subsets of different lengths. We then used these trained models to generate forecasts of the target (imports) over the time range of the held-out test set. Using the held-out test set, we evaluated out-of-sample RMSE accuracies of those forecasts, and plotted them against the training subset length in Figure 5. Thus Figure 5 demonstrates the effect of the amount of time-series data available for model training on forecasting accuracy, for models of different complexity (simple univariate model with no additional predictors compared to more complex multivariate models with one and with six additional predictors).

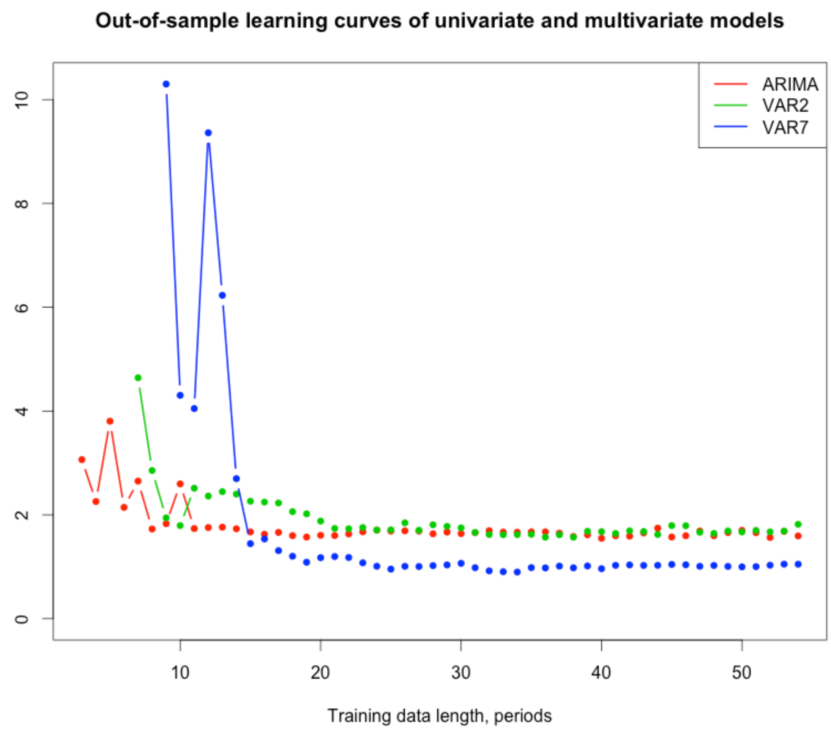

Figure 5. Forecast accuracy (out-of-sample forecast RMSE accuracies measured on the held-out synthetic test set) for a univariate model (ARIMA), a 2-dimensional multivariate VAR2 model, and a 7-dimensional multivariate VAR7 model, all models trained on synthetic data of varying lengths.

It is seen from Figure 5 that, despite the fact that the VAR7 model has the structure of the "true" data-generating model, in data-poor cases it performs worse at forecasting than simpler models (univariate ARIMA and 2-dimensional VAR2) whose structure does not properly capture the "true" data-generating process. The reason is that more complex models have more parameters to estimate from the training data, and thus require more data to estimate them accurately. In data-poor situations (when the length of time series available for model training is limited), there may not be enough data for more complex models to learn properly, unlike for simpler models. As a result the forecasts generated by more complex models can turn out to be less accurate than the forecasts generated by simpler models.

If sufficient training data is available, more complex models should eventually be able to learn their parameters well enough to be able to realize their potential to produce more accurate forecasts than simpler models. In data-poor situations (for small training set lengths) the simple univariate ARIMA model and the slightly more complex bivariate VAR2 model both yield more accurate forecasts than the complex multivariate VAR7 model, despite the latter model having the exact structure of the data generating process (Figure 5). However in data-rich situations, the more complex multivariate model outperforms the simple models, as its structure better reflects the true structure of the underlying data-generating process. Figure 5 illustrates the fact that there is a crossover between data-poor to data-rich situations. We note that the location of this crossover depends on the statistical properties of the data generating process and quantity of data, i.e., it is problem- and data-dependent.

4.5. Discussion

Our analysis highlights the difficulties in forecasting even in the age of big data. Complex models with additional information like socio-economic indicators require more training data. In the absence of enough data, complex models can be outperformed by much simpler models. Time series prediction poses significant challenges even in the era of big data since the amount of data of a target variable (like imports) or a predictor (like GDP) cannot be increased other than by the passage of time.

Additionally, our models implicitly encapsulate a very complex and dynamical socio-economic system with many components like the economies of Australia and other countries. Our results point to the difficulty of fitting models of a dynamical system composed of multiple interconnected components.

The pitfalls of making predictions from big data have also been highlighted in another context: predicting cases of influenza using Google Flu Trends (Lazer et al., 2014). Google Flu Trends predicted twice as many influenza cases as happened in reality. Paradoxically simple models using just 2 week lagged data from the Center for Disease Control have better prediction accuracy than Google Flu Trends which used 50 million search terms to fit 1152 data points. This is not to say that big data coupled with complex models is always bad practice. However we advocate a more nuanced approach that takes a critical look at the data, problem domain and algorithms. Simple models may sometimes be better suited to some problems than complex multivariate models.

Another difficulty in forecasting a complex socio-economic system is that some of the underlying components of the system may also have been perturbed during the global financial crisis (GFC) of 2008. Such a perturbation may cause a shift in the covariate structure encapsulating relationships between different economies (which are assumed to be time-invariant in models we examined). We expect significant practical difficulties in estimating parameters from a model if the training set has a major event like a global financial crisis which could cause a covariate shift or completely change the underlying system and network connections (Tsay, 1988). Hence methods to automatically detect covariate shifts and change points can be useful in guiding proper choice of forecasting algorithms (Shimodaira 2000, Davis 2006, Cho and Fryzlewicz, 2015).

Performance of the predictive models could be sensitive to whether the training set includes the period during and after the GFC. To explore this, we built and compared two models of the same class (VAR7): Model 1 is trained on 13.5 years of pre-GFC data from Q1 1995 to Q2 2008, and is tested on a 3-year post-GFC data from Q3 2008 to Q2 2011, while Model 2 is trained on 13.5 years of both pre- and post-GFC data from Q1 1998 to Q2 2011, and is tested on a 3-year post-GFC data from Q3 2011 to Q2 2014. As shown in Figs. 6-7, Model 1 fails to predict the oncoming drop in the imports that occurs at the start of 2009. Predictions generated by Model 2 that has captured 3 post-GFC years, agree with the future test imports data much better. Note that the high volatility in the imports during and immediately after the GFC is captured and reflected in larger estimated uncertainty interval around the prediction generated by Model 2.

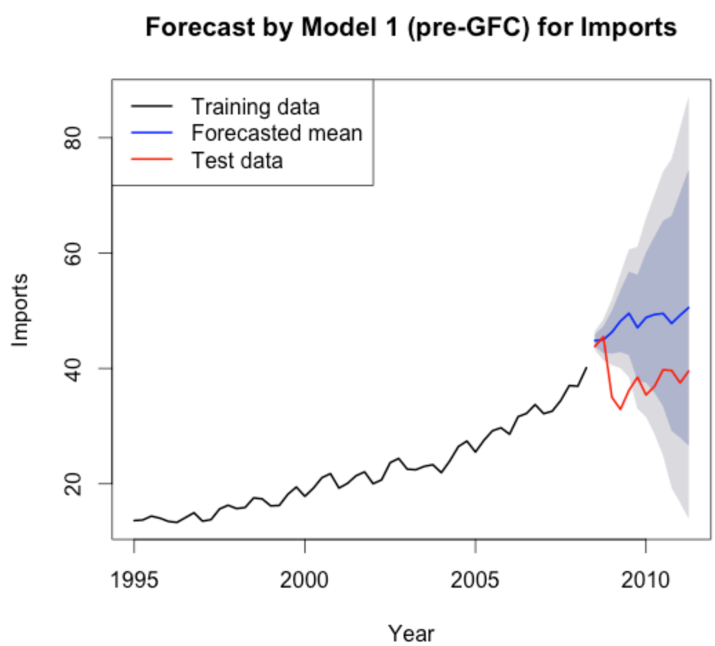

Figure 6. Forecast of port throughput by Model 1 (7D VAR model using GDPs of Australia and of top-5 importers as additional predictors of imports) trained on pre-GFC data from Q1 1995 to Q2 2008. Out-of-sample (test set from Q3 2008 to Q2 2011) Root Mean Squared Error RMSE=10685.

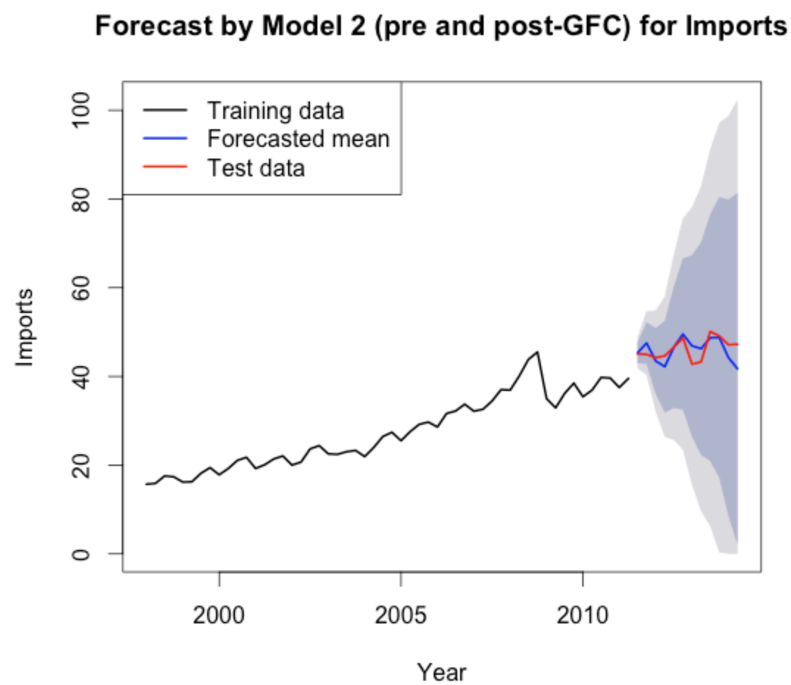

Figure 7. Forecast of port throughput by Model 2 (7D VAR model using GDPs of Australia and of top-5 importers as additional predictors of imports) trained on pre- and post-GFC data from Q1 1998 to Q2 2011. Out-of-sample (test set from Q3 2011 to Q2 2014) Root Mean Squared Error RMSE=2594.

Model 2 that has seen only 3 years' worth of post-GFC data on top of what Model 1 has seen, yet its performance on predicting post-GFC imports is surprisingly good. This might indicate that the perturbation that the GFC introduced into the Australian trade relation with its major trading partners was localized in time, and did not introduce a permanent covariate shift into the underlying structure of the system. In other words, the system's underlying structure (at least the part describing the trading relationships of Australia) reverted back to its pre-GFC state shortly after the GFC passed. This conjecture, however, would require further analysis to confirm, which is beyond the scope of this paper.

5. Concluding Remarks

This paper challenges the general rule of thumb: "the more data we use for forecasting, the better the forecast will be". Our analysis suggests that the forecasting process is very problem and data specific. Using additional information like socio-economic indicators can have either a beneficial or adverse effect on the model forecasting accuracy. We also observed that the choice of data transformations affects the accuracy of the forecasts (Table 1 and Section 4.2). This makes forecasting a complex subjective process, in which the forecaster is faced with a multitude of choices that depend on the peculiarities of the problem and data, and each of these choices affects the result. Therefore, a methodology that would make the forecast-building process more objective is required. We suggest the following methodology as a first step towards a more objective forecasting process:

1) Exploring a range of transformations to pre-process the data,
2) Fitting models of varying degrees of complexity and number of predictors, and performing model selection based on in-sample and out-of-sample model performance (e.g., in-sample AIC, and out-of-sample RMSE).
3) Assessing whether there is enough data to learn these models by assessing performance on synthetic datasets.

More complex models require more training data to estimate their parameters, and in data-poor cases may yield worse forecasts than simpler models (univariate models or multivariate models with fewer predictors). As such, amount of data available for training the forecasting model is a critical factor in the choice of models used for forecasting.

Our approach should be of interest to practitioners and policy makers in deciding which techniques to apply to their problems.

In summary we caution that more information may not always be better, and it is advisable for analysts to weigh the cost of adding more data; the ultimate decision depends on the problem, amount and quality of data and the kind of models being used. We hope that our work will start a discussion around picking the right models for forecasting and change the prevalent perception that complex models with more information are always better.

References

Abraham, B and Ledolter, J., 2009, Statistical methods for forecasting. John Wiley & Sons.

Anderson, D.R., Sweeney, D.J., Williams, T.A., Jeffrey D. Camm, J.D. and Martin, R.K., 2009, Quantitative Methods for Business. 12th edition. Canada: South-Western Cendage Learning

Asteriou, D. and Hall, Stephen G., 2011, ARIMA Models and the Box–Jenkins Methodology. Applied Econometrics (Second ed.). Palgrave MacMillan. 265–286. ISBN 978-0-230-27182-1.

Angelopoulos, J. and Chlomoudis, K., 2015. Towards a sport services indicator- The case of US Ports, paper presented at the IAME Conference, Marseilles, France, 1-21.

Brockwell, P. J. and Davis, R. A., 1996, Introduction to Time Series and Forecasting. Springer, New York.

Cho, H. and Fryzlewicz, P., 2015, Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77: 475–507.

Chatfield, C., 2013, The analysis of time series: an introduction, CRC press.

Chen, S.H. and Chen, J.N., 2010, Forecasting container throughputs at ports using genetic programming. Expert Systems with Applications, 37(3), 1054-2058.

Chou, C.C, Wu, C.C. and, Liang. S.G., 2008, A modified regression model for forecasting the volumes of Taiwan's import containers. Mathematical and Computer Modelling 47(9-10), 797–807.

De Langen, P.W., Van Meijeren, J. and Tavasszy, A.L., 2012, Combining Models and Commodity Chain Research for Making Long-Term Projections of Port-Throughput: An Application to the Hamburg-Le Havre Range. European Journal of Transport and Infrastructure Research, 12(3), 310-331.

Fung, M., 2002. Forecasting Hong Kong's Container Throughput: An Error correction Model. International Journal of Forecasting, 21(1), 69–80.

Geng, J., Li, M.W., Dong, Z.H. and Liao, Y.S., 2015, Port Throughput Forecasting by MARS-RSVR with chaotic simulated annealing particle swarm optimization algorithm. Neurocomputing, 147, 239-250.

Ljung, G.M. and Box, G.E.P., 1978, On a Measure of a Lack of Fit in Time Series Models. Biometrika 65(2): 297–303.

Goulielmos, A.M. and Kaselimi, E., 2011, A nonlinear forecasting of container traffic: the case study of the port of Piraeus, 1973-2008. International Journal of Shipping and Transport Logistics, 3(1), 72-99.

Gooijer J.G. and Klein, A., 1989. Forecasting the Antwerp maritime steel traffic flow: A case study, Journal of Forecasting, 8(4), 381-398.

Gosasang, V., Chandraprakaikul, W. and Kiattisin, S., 2011. A Comparison of Traditional and Neural Networks Forecasting Techniques for Container Throughput at Bangkok Port, The Asian Journal of Shipping and Logistics, 27(3), 463-382.

Guo, Z., Song, X. and Ye, J., 2005. A Verhulst model on time series error corrected for port throughput forecasting. Journal of the Eastern Asian Society for Transportation Studies, 6, 881-891.

Hanke, J.E., Reitsch, A.G. and Wichern, D.W., 2001, Business Forecasting. 7th edition. Upper Saddle River, New Jersey: Prentice Hall.

Hui, E. C. M., Seabrooke, W. and Wong, G.K.C., 2004, Forecasting Cargo Throughput for the Port of Hong Kong: Error Correction Model Approach, Journal of Urban Planning and Development, 130(4), 195-203.

Intihar, M., Krnberger, T. and Dragan, D., 2015, The relationship between the economic indicators and the accuracy of container port throughput forecasting, paper presented at the IAME Conference, Marseilles, France, 1-25.

Kim, J.K., Pak, J.Y., Wang, Y., Park, S.I. and Yeo, G.T., 2011, Study on forecasting container volume of port using SD and ARIMA, Journal of Korean navigation and port research, 35(4), 343-349.

Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014, The Parable of Google Flu: Traps in Big Data Analysis, Science.

Lehto, H.,Venäläinen, P. and Hietala, K., 2006. Growth outlook of seaborne transport between Finland and foreign countries up to 2030. Finnish Maritime Administration Publications 10/2006, Helsinki.

Li, J., Chen, Y. and Cui, X., 2008, Port Throughput Forecast Based on Nonlinear Combination Method. Paper presented at the 7th WSEAS International Conference on applied computer & applied computational science (ACACOS ‘08), Hangzhou, China.

Maloni, M. and Jackson, E.C., 2005, North American Container Port Capacity: An Exploratory Analysis. Transportation Journal, 44(3), 1–22.

McCarthy, T., Davis, T. and SL Mentzer J.T., 2006, The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. Journal of Forecasting 25: 303-324.

Meersman, H., 2009, Maritime traffic and the world economy., In: Meersman, H., Van de Voorde, E. and Vanelslander, T. (eds.), Future challenges for the port and shipping sector, Informa, London.

Mostafa M.M., 2004, Forecasting the Suez Canal traffic: a neural network analysis. Maritime Policy and Management, 31(2), 139-156.

Paflioti, P., Roumpis, E., Vitsounis, T., Tsamourgelis, I. and Bell, M., 2015, Characterizing the cycle in seaports: An Australian case study, paper presented at the IAME Conference, Marseilles, France, 1-25.

Pallis, A.A., Vitsounis, T.K., De Langen, P.W. and Notteboom, T.E., 2011, Port economics, policy and management: Content classification and survey. Transport Reviews, 31(4), 445-471.

R Core Team, 2015, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Seabrooke, W., Hui, E.C.M., Lam, W.H.K., and Wong, G.K.C., 2003, Forecasting cargo growth and regional role of the port of Hong Kong. Cities, 20(1), 51-64.

Schulze, P.M. and Prinze, A., 2009, Forecasting container transshipment in Germany, Applied Economics, 41(22), 2809-2815.

Shimodaira, H., 2000, Improving predictive inference under covariate shift by weighting the log-likelihood function, Journal of Statistical Planning and Inference, 90(2), 227-244.

Tsay, R. S., 1988, Outliers, level shifts, and variance changes in time series. Journal of Forecasting, 7(1), 1–20.

Van Dorsser, C., Wolters, M. and van Wee, G.P., 2012, A Very Long Term Forecast of the Port Throughput in the Le Havre – Hamburg Range up to 2100. European Journal of Transport and Infrastructure Research, 12(1), 88-110.

World Bank GDP Data, http://data.worldbank.org/indicator/NY.GDP.MKTP.PP.KD. Date of access: 1/2/2016.