Forecasting monthly air passenger flows from Sweden

Evaluating forecast performance using the Airline model as benchmark

by

Fredrik Robertson & Max Wallin

Bachelor thesis
Department of Statistics
Uppsala University

Supervisor: Anders Ågren

Fall 2014
Abstract

In this paper two different models for forecasting the number of monthly departing passengers from Sweden to any international destination are developed and compared. The Swedish transport agency produces forecasts on a yearly basis, where net export is the only explanatory variable controlled for in the latest report. More profound studies have shown a relevance of controlling for variables such as unemployment rate, oil price and exchange rates. Due to the high seasonality within passenger flows, these forecasts are based on monthly or quarterly data. This paper shows that a seasonal autoregressive integrated moving average model with exogenous input outperforms the benchmark model forecast in seven out of nine months. Thus, controlling for oil price, the SEK/EUR exchange rate and the occurrence of Easter reduces the mean absolute percentage error of the forecasts from 3,27 to 2,83 % on Swedish data.

Keywords: airline model, SARIMAX, forecast, passengers
# Table of contents

1. Introduction .................................................................................................................. 3
2. Previous research ........................................................................................................... 4
   2.1 Forecasting (aggregate) demand for US commercial air travel................................. 4
   2.2 Forecasting of Hong Kong airport’s passenger throughput ........................................ 5
   2.3 Air Passenger Flows: Evidence from Sicily and Sardinia ........................................ 6
   2.4 Forecast 2013-2018 by the Swedish Transport Agency .............................................. 6
3. Method .......................................................................................................................... 7
   3.1 Benchmark - The airline model .................................................................................. 7
   3.2 Causal model ........................................................................................................... 8
   3.3 Method for model and forecast evaluation ................................................................. 9
4. Data ............................................................................................................................... 10
5. The naive model ........................................................................................................... 11
   5.1 Data transformation ................................................................................................. 11
   5.2 Model identification ............................................................................................... 13
   5.3 Parameter estimation .............................................................................................. 14
   5.4 Residual diagnostics ............................................................................................. 15
   5.5 Naive model forecast ............................................................................................. 18
6. The causal model .......................................................................................................... 22
   6.1 Explanatory variables ............................................................................................. 22
   6.2 Model estimation .................................................................................................... 24
   6.3 Causal model forecast ........................................................................................... 25
7. Evaluation ..................................................................................................................... 27
9. Analysis ......................................................................................................................... 31
   9.1 Results .................................................................................................................... 31
   9.2 Differences from previous studies ......................................................................... 32
   9.3 Possible improvements of the model ....................................................................... 33
10. Conclusion .................................................................................................................... 34
References ....................................................................................................................... 35
Appendix A – Hypothesis tests ....................................................................................... 37
Appendix B – Time series plots of the explanatory variables .......................................... 39
Appendix C – Estimated causal models ......................................................................... 41
1. Introduction

Air travel is a vital part of the Swedish infrastructure. It is a function of, and a contributing factor to, the ongoing process of globalization where people in different countries become more integrated in economic, cultural, social and political aspects. International institutions, such as the European Union, continue to harmonize trade regulations in order to facilitate intra border movement for people as well as for goods. Trade with goods and services are facilitated by abolished customs, and the introduction of a common currency along with the freedom of movement has increased the number of people travelling between the countries within the union. The deregulation of the European aviation market has changed the market situation for airline companies. Since the beginning of the 21:st century, national airlines face competition from private actors. This led to the introduction of low-cost companies which made air travel a good available for a larger share of the population. (Transportstyrelsen, 2009).

The number of passengers from Sweden has increased steadily during the last decade, why it is of interest for the airports, airline companies and authorities to obtain forecasts in order to plan future production and logistics. The Swedish transport agency produces forecasts on a yearly basis regarding the number of passengers to and from Sweden. The aim of the forecasts is to provide external stakeholders the assessment of aviation development in Sweden and for planning logistics in future years. A number of airport related fees, such as terminal- and luggage fees, are based on these forecasts why the accuracy is of great importance (Transportstyrelsen, 2013).

The purpose of this paper is to find a model that accounts for exogenous variables, and by doing so improve a regular time series forecast of monthly airline passengers departing from Sweden to any international destination.

A short review of previous research will be presented in section 2. This is followed by a methodology part where the estimated models are described. Section 4 contains a description of the data. The naive and causal models are presented separately in section 5 and 6, and then evaluated and compared considering forecast performance in section 7. A forecast until June 2015 is provided in section 8. Analysis and conclusions are presented in section 9 and 10.
2. Previous research

In previous research, many techniques have been used to forecast passenger flows with various results. Two common approaches are the *top-down* and *bottom-up strategies*. The top-down strategy is based on a single aggregated forecast, i.e. a forecast of the total number of passengers in a country. This forecast can then be distributed to the individual airports within the country based on their historical passenger development. The bottom-up strategy is the opposite, where forecasting is done for individual airports and is then summed to obtain an aggregate forecast.

2.1 Forecasting (aggregate) demand for US commercial air travel

In a study by Carson et al. (2011) the disaggregated approach was found to produce more accurate forecasts of air travel demand than the bottom-up approach. In the study, data for 179 individual airports were used to compute a model where the aggregated demand was forecasted by the sum of all individual forecasts, taking into account the oil price and the unemployment rates in the regions served by the different airports.

This approach was suitable for U.S. data where the economic structure of the country is very heterogeneous, i.e. there are large differences in the economic structure within the country. The unemployment rate in New Jersey differs considerably from the unemployment rate in Washington, why it is reasonable to treat each region separately when computing forecasts.

The final model was expressed as:

\[ y_{it} = \alpha_i + \beta_{i1}t + \beta_{i2}unemp_{it} + \beta_{i3}Ci_{it} + \beta_{i4}CI_{it}^2 + \beta_{i5}jetfuel_{t} \]

\[ + \beta_{i6}\Delta oilfutures_t + \beta_{i7}sept11_{t} + \sum_{k=1}^{11} \theta_{ik}sd_k + \phi_{i1}y_{i,t-1} + \varepsilon_{it} \]

where \( \beta_{i2}unemp_{it} \) is the unemployment rate in region \( i \) in period \( t \), and \( \sum_{k=1}^{11} \theta_{ik}sd_k + \phi_{i1}y_{i,t-1} + \varepsilon_{it} \) is the weighted sum of the seasonal components and an AR(1) term.

The main finding was that the disaggregated approach outperforms the aggregate approach in terms of forecast error measurements.
2.2 Forecasting of Hong Kong airport’s passenger throughput

Since the opening of the Hong Kong International Airport (HKIA) in 1998 the volume of air passengers and cargo have grown steadily, except for the post 9/11 period and the SARS outbreak. A forecast of the airport passenger flows allows for short- and long term planning regarding airport facilities and flight network.

Kan Tsui et al. (2014) use the Box-Jenkins methodology to estimate a seasonal ARIMA and a seasonal ARIMAX to forecast the number of passengers using data from 1993 – 2011.

In contrast to the study by Carson et.al, passengers were categorized into different groups depending on destinations. Eleven groups, such as Africa, Europe, Japan etc., were identified and separate forecasts were computed for each region. The sum of these formed the aggregated airport forecast of the passenger throughput. The final model was a SARIMAX \((p, d, q)(P, D, Q)_{12}\) where the explanatory variables \(GDP_{t-4},\) \(Connecting Traffic_t,\) and \(Fuel Price_t\) were significant. Instead of using the actual fuel price, a dummy variable set to 1 was used if the fuel price in month \(t\) was more than 80 dollars per barrel, and 0 otherwise.

The authors used three different scenarios when producing the forecasts to account for changes in the explanatory variables;

1) The explanatory variables are assumed to take on values made by forecasts from external sources,
2) GDP is assumed to decrease at a 5 % annual rate while the oil price remains as in 1),
3) The oil price remains below 80 dollars per barrel.

The key finding of the paper was that a SARIMAX model, taking into account GDP, oil price and connecting flights, made the best forecast with an average monthly deviation of 3.6 %.
2.3 Air Passenger Flows: Evidence from Sicily and Sardinia

In a study by Castellani et al. (2010), passenger flows to Italy were used as a proxy to measure the tourism demand. A set of VAR model specifications were constructed in order to investigate the monthly time series 2003-2008 of arrivals to the Italian islands Sardinia and Sicily. Arrivals were treated as one group, i.e. the origin of the tourists was not accounted for. The model found to be the best in forecasting the tourism flows was;

\[ y_t = A_1 y_{t-1} + A_{12} y_{t-12} + B x_t + u_t \]

where \( y_t \) is a vector containing the dependent variables, i.e. arrivals to the airports, and the right hand side of the equation consists of vectors for the lagged dependent and explanatory variables.

The findings reveal significant impact of both meteorological variables, (atmospheric temperatures and raining days) and lagged values of the exchange rates USD/EUR and YEN/EUR. These factors were controlled for, which improved the explanatory and forecasting power of estimated VAR models (Castellani et al. 2010).

2.4 Forecast 2013-2018 by the Swedish Transport Agency

Compared with the previous models, the one constructed by the Swedish transport agency seems rather basic;

\[ lnY_t = 7.7 + 1 * lnNet\ trade_t \]

where \( Y_t \) is the number of passengers travelling to an international destination at time \( t \), and \( Net\ trade_t \) is the change in the Swedish net export in time \( t \).

This model has several flaws. It is of great importance that the mean and variance of the time series do not change over time, which is not the case in this study. This means that neither the dependent nor explanatory variable is stationary, which is desirable. The model does not account for any lag in the explanatory variable, thus a change in the net trade is assumed to affect the number of passengers instantly. Moreover, the coefficient in front of the net export is 1, meaning that a percentage change of net trade in year \( t \) gives rise to an equal percentage change of passengers in year \( t \).
The only explanation given is that the correlation between net export and departing passengers is 0.82. (Swedish transport agency, 2013). The correlation between the variables does not necessarily imply that the net export has a causal effect on the number of passengers, but the authors seem to be satisfied by the fact that the correlation is high. Since the data consist of yearly observations on departing passengers, there is no need to account for seasonal variation in the passenger flows, even though this is of interest when making such forecasts.

3. Method

In order to forecast the monthly passenger flow, two models for forecasting the number of air passengers departing from Sweden will be estimated and forecasted using EViews 8. The forecasts will be made for the total number of passengers departing from Sweden to any international destination. Arriving passengers will not be included since the aim is to estimate the number of international air travel passengers. Airports will not be treated individually, and no distinction will be made regarding destination. To evaluate the forecasts, monthly observations from 2014 will be treated as unknown, and used to compare the point predictions with the actual values. The first model is a time series forecast, which only takes previous values of the passenger data into account, in the paper referred to as a naive model. This model will then be expanded by exogenous explanatory variables, in the paper referred to as a causal model.

3.1 Benchmark - The airline model

The naive model is estimated for the monthly passenger data process, and is done by applying the Box-Jenkins four-step methodology for stationary time series. These are model identification, parameter estimation, residual diagnostics and forecasting. When modelling monthly time series with recurring patterns every year, Box & Jenkins (1976) developed a two-coefficient time series model that is known as the airline model since the data were monthly passenger flow (Findley et al. 2014). The model is often used to forecast passenger flow since it is both accurate and straightforward to estimate. The model has the form:

\[(1 - B)(1 - B^7)Y_t = (1 - \theta B)(1 - \Theta B^7)e_t\]
It is a SARIMA(0,1,1)x(0,1,1), where B is the backshift operator, \(Y_t\) is the number of passengers in time \(t\), \(\theta\) is a moving average (1) term and \(\Theta\) is a seasonal moving average term. \(e_t \sim WN(0,\sigma^2)\). Thus, for monthly data, \(s = 12\). To forecast a SARIMA(0,1,1)x(0,1,1) \(k\) steps ahead, the following representation is used:

\[
(1 - B)(1 - B^s)Y_{t+k} = (1 - \theta B)(1 - \Theta B^s)e_{t+k}
\]

In this paper, it will be found that the airline model will serve as a suitable benchmark model that is to be improved.

3.2 Causal model

The causal model, which aims to improve the forecasting performance of air passengers, is an expansion of the airline model with explanatory variables. Thus, the model is defined as a SARIMAX(0,1,1)x(0,1,1)\(_{12}\) and it is outlined as;

\[
(1 - B)(1 - B^s)Y_t = (1 - \theta B)(1 - \Theta B^s)e_t + \sum_{i=1}^{b} \beta_i X_{i,t-v}
\]

where \(b\) is the number of exogenous variables included in the model and \(v\) is the lag. The explanatory variables can be either lagged, not lagged or a mix of the two depending on the delay of the effect in the dependent variable given a change in the specific explanatory variable. The forecast of the SARIMAX model is obtained from;

\[
(1 - B)(1 - B^s)Y_{t+k} = (1 - \theta B)(1 - \Theta B^s)e_{t+k} + \sum_{i=1}^{b} \beta_i \hat{X}_{i,t-v+k}
\]

The causal model uses available or unknown future values of the explanatory variables depending on the lag structure in order to forecast. Actual values on the explanatory variables are used for the observations out of sample in this paper, since finding a forecast model with significant exogenous variables is the purpose. This is not possible in a real life situation, but if proven that the causal model significantly exceeds the naive model in forecast performance, it is advantageous to use exogenous variables as input.
3.3 Method for model and forecast evaluation

The naive- and causal model estimations will be evaluated by the coefficient of determination, $R^2$. This statistic tells how much variation in the dependent variable the model is able to capture. The $F$-statistic of the estimated models will also be evaluated since this statistic determines whether the specific model has some explanatory power.

Hypothesis tests for the procedure can be found in Appendix A, where all tests are carried out on the 5%-level. Thus, a probability of 5% of committing a type 1-error, i.e. wrongly rejecting the null hypothesis, is accepted. This is reasonable since the results should serve as an indication of which exogenous factors that are relevant to control for when forecasting airline passengers.

Two error measures will be calculated and used to compare the forecast performance of the models. The root mean square error (RMSE) is defined as the square root of the average of the squared forecast errors. The mean absolute percentage error (MAPE) determines the magnitude of the deviations in absolute percentage terms. The measures are calculated by:

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2}{n}}
\]

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|
\]

These will be used to determine whether the causal model has a better model fit than the benchmark model.
4. Data

The data come from the Swedish transport agency in association with the aviation authority. It consists of monthly passengers departing from Sweden to any international destination during the time period January 1996 – September 2014. In Figure 1, only observations until December 2013 are included since the remaining ones are saved to evaluate the forecasts.

Figure 1. Monthly departing passengers (in 1000:s)

In September 2001 there is a distinct decline in the number of passengers as a result of the terrorist attacks on World Trade Center. During 2002/2003, the SARS pandemic forced a shutdown of large parts of the Asian air space, and resulted in a decrease in air traffic worldwide (Ruwantissa, 2003). The decline in 2008 is due to the financial crisis. At this point, uncertainty regarding the economic future was large why people postponed consumption which included reduced air travel.

As expected, there is a clear seasonal pattern where the summer months are the busiest. Since the pattern is identical regardless of year, this indicates that the behavior of passengers does not change over time. There is a positive trend, and the variance increases with time. The series follows the expected pattern and, not surprisingly, it resembles an airline passenger process.
5. The naive model

The methodology of Box-Jenkins will be applied in order to determine the fit of the airline model to Swedish data.

5.1 Data transformation

To make inference regarding the structure of a stochastic process, an important condition is that the observed series is stationary, meaning that the properties of the process do not change over time. This is a crucial assumption in the Box-Jenkins methodology, why it is often necessary to transform the data until this condition is fulfilled (Cryer & Chan, 2008:14).

The original data displayed in Figure 1 show a positive trend and increasing seasonal variation. Since stationarity is required, the data is log-transformed to smooth out the seasonal variation with no loss of information.

*Figure 2. Log-transformed data*
The plot in Figure 2 indicates a non-stationary process with a linear trend, but the log-transformation made the seasonal variance constant over time. To make the series stationary, first- and seasonal differences are applied according to the following formula;

$$\nabla_1 \nabla_{12} \ln Y_t = \ln Y_t - \ln Y_{t-1} - \ln Y_{t-12} + \ln Y_{t-13}$$

If all terms in the equation above are multiplied by 100, it is to be interpreted as the percentage change in passengers between two months year $t$, minus the percentage change in passengers between the same months the previous year. The transformed data series is illustrated in Figure 3.

*Figure 3. First and seasonal difference of logged passengers*

After the transformation, the mean is centered on 0, but the variance look far from constant due to the heavy deviations in 2010 and 2011. These outliers are the result of the volcano eruption on Iceland in April 2010, which resulted in an ash cloud covering large parts of Europe and the Swedish airspace was forced to shut down during half of April (Transportstyrelsen, 2013). In terms of passenger flows in April, there was a decline by approximately 200 000 passengers. Since the series is seasonally differenced, this deviation affects April 2011 as well. Such an unusual external chock is pointless to model for. It resulted in a limited disruption in the series which did not affect neither the passenger flows
nor the behavior of passengers in future time periods. This is the definition of an *additive outlier*, why it is reasonable to adjust the values for April 2010 from the time series (Cryer & Chan, 2008:259). This is done by taking the mean between March and May 2010.

*Figure 4. First and seasonal difference were the ash cloud effect is adjusted*

After adjusting for the ash cloud effect, the time series in Figure 4 look stationary which is confirmed by the Augmented-Dickey-Fuller test of a unit root, see Appendix A (1). Since the series is difference stationary, the Box-Jenkins methodology for time series forecasting is applicable.

5.2 Model identification

The main tools in the identification of the time series are the sample autocorrelation function $\hat{\rho}_k$, and the sample partial autocorrelation function (SPACF). $\hat{\rho}_k$ is simply the correlation between the transformed $Y_t$ and $Y_{t+k}$. The partial autocorrelation measures the correlation between observations in the time series that are $k$ time periods apart, after controlling for the correlations at intermediate lags (less than $k$). These functions are used to identify which kind of pattern the series follow, and thus are tools to determine a model for the series. (Gujarati, 2004:842).

After the necessary data transformations, the sample consists of 203 observations. The first 19 lag autocorrelations are illustrated in the correlogram below.
The significant spikes at the first and twelfth lags in Figure 5 suggest a non-seasonal MA(1) component and a seasonal MA(12) component. The autocorrelation structure is in line with previous studies on passenger time series data, indicating a seasonal ARIMA\((0,1,1) \times (0,1,1)_{12}\). Thus the airline model seems to be the best model to be fitted.

5.3 Parameter estimation

Having identified an appropriate model to fit the data, the parameters can be estimated. The estimation in this paper will be done by using the non-linear least squares method. The SARIMA\((0,1,1) \times (0,1,1)_{12}\) model that is estimated is outlined as:

\[
\nabla_{12} \nabla_1 \ln Y_t = \alpha + e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta \Theta e_{t-13}
\]
Table 1. Estimated naive model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.00</td>
<td>0.00</td>
<td>0.83</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.23</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>SMA(12)</td>
<td>0.58</td>
<td>0.06</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample size</th>
<th>203</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.22</td>
</tr>
<tr>
<td>F-statistic</td>
<td>28.20</td>
</tr>
</tbody>
</table>

The moving average and the seasonal moving average part are statistically significant when interpreting the corresponding p-values while the constant lacks a meaningful interpretation. The $R^2$ is quite high even though there are possible omitted factors not controlled for. The $F$-statistic significantly states that the estimated parameters have some explanatory power, indicating a good model fit of SARIMA(0,1,1)$\times$(0,1,1)$_{12}$.

5.4 Residual diagnostics

In order to determine the quality of the fitted model, several assumptions regarding the error terms need to be evaluated. If the assumptions are fulfilled, the estimated model is accepted. The following assumptions will be assessed;

1) White noise of the error terms

The remaining error terms should follow a white noise process, i.e. an uncorrelated process with mean zero and finite variance. This is tested for using the $Q_{LB}$ statistic, see Appendix 1 (2) for details. The first 19 autocorrelations of the residuals are displayed in Figure 6.
It is found that the model is acceptable since there is no significant autocorrelation between any lags, thus the structure corresponds to a white noise process and the white noise hypothesis can not be rejected.

2) Normality of the error terms

The residuals of the fitted model should ideally indicate a Gaussian distribution of the error terms. A quantile-quantile plot of the residuals is displayed in Figure 7.
The quantiles of the probability distribution that generated the residuals are compared to the quantiles of a standard normal distribution. A clear linear pattern is visible, indicating normality. However, two outliers are evident at the beginning of the range. When formally tested for, the Jarque-Bera statistic rejects the null hypothesis of normality, but the distribution corresponds well to a normal distribution, see Appendix A (3) for further details.

3) Homoscedasticity of the error terms

If the standardized residuals do not have constant variance, the estimates will be unbiased but no longer have minimum variance among the unbiased estimators. A time series plot of the standardized residuals is illustrated in Figure 8.

Figure 7. Quantile-Quantile plot of residuals

Figure 8. Time series of standardized residuals of the estimated model
The variance in the time series appears constant in Figure 8, except from September 2001, October 2008 and the respective months the following years due to the differencing. The magnitude of these residuals is larger than 3, which is unusual in a standard normal distribution. (Cryer & Chan, 2008:177). As addressed earlier, these deviations occurred due to the 9/11 attacks, SARS and the financial crisis.

Heteroscedasticity in a time series often occur due to omitted variable bias, i.e. shocks to the series depend on factors not being controlled for. Since this model only takes the actual time series of the investigated variable into account, this bias is allowed. Hence, by definition, this is a naive model.

To summarize, the diagnostics of the estimated model fulfill the necessary conditions, with exception for the slight deviation from normality of the error terms. Hence, the forecasts should be unbiased but the efficiency might be damaged due to the outliers given the model used.

5.5 Naive model forecast

Forecasting can be done in two ways, static and dynamic forecasting, where the latter uses the forecasted values in period \( t+1 \) to forecast \( t+2 \) instead of the actual values. In order to produce accurate forecasts of future values for a time series, as much information as possible until time \( t \) when the forecast starts should be used. Data is assumed to be available until December 2013, and the forecasts will be made until September 2014. The forecast generates point predictions for each month in terms of logged yearly and monthly differences in thousands of passengers.
The naive forecast seems to capture the monthly trend of the process. The actual deviations are tedious to evaluate since the data is transformed, but the largest deviation seems to occur in April. To obtain the forecast in monthly passengers in 1000:s, the difference procedure is done in reversed order;

\[ \hat{Y}_{t+i} = \exp(\nabla_{12} \nabla_1 \ln Y_{t+i} + \ln Y_{t+i-1} + \ln Y_{t+i-12} - \ln Y_{t+i-13}) \]

By doing so, the forecasts are still estimated on a logged difference stationary process, but the actual predictions can be interpreted as numbers of passengers.
The prediction of April is grossly underestimated, but the forecast series follow the actual values quite well. The underestimation in April is due to the presence of Easter holidays. Easter occurred in March 2013, and since the forecast is carried out on yearly differenced data, the model assumes Easter to occur in March 2014. This results in an overestimation in the number of passengers in March while underestimating the number of passengers in April.
Table 2. Table of naive forecasts and actual values

<table>
<thead>
<tr>
<th>Month 2014</th>
<th>Actual values</th>
<th>Naive forecast</th>
<th>Difference</th>
<th>Difference in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>787 621</td>
<td>809 013</td>
<td>+ 21 392</td>
<td>+ 2,7</td>
</tr>
<tr>
<td>February</td>
<td>817 609</td>
<td>819 159</td>
<td>+ 1 550</td>
<td>+ 0,2</td>
</tr>
<tr>
<td>March</td>
<td>915 594</td>
<td>948 790</td>
<td>+ 33 196</td>
<td>+ 3,6</td>
</tr>
<tr>
<td>April</td>
<td>1 054 043</td>
<td>961 108</td>
<td>- 92 935</td>
<td>- 8,8</td>
</tr>
<tr>
<td>May</td>
<td>1 159 433</td>
<td>1 205 708</td>
<td>+ 46 275</td>
<td>+ 4,0</td>
</tr>
<tr>
<td>June</td>
<td>1 372 414</td>
<td>1 345 022</td>
<td>- 27 392</td>
<td>- 2,0</td>
</tr>
<tr>
<td>July</td>
<td>1 339 337</td>
<td>1 310 178</td>
<td>- 29 159</td>
<td>- 2,2</td>
</tr>
<tr>
<td>August</td>
<td>1 212 215</td>
<td>1 240 058</td>
<td>+ 27 843</td>
<td>+ 2,3</td>
</tr>
<tr>
<td>September</td>
<td>1 208 302</td>
<td>1 251 982</td>
<td>+ 43 680</td>
<td>+ 3,6</td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td></td>
<td></td>
<td>42 978</td>
</tr>
<tr>
<td>MAPE</td>
<td></td>
<td></td>
<td></td>
<td>3,27</td>
</tr>
</tbody>
</table>

February is the month where the forecast deviation is the smallest, 1 550 passengers. Thus, February seems to be a month were the number of passengers has little variation among years. The square root of the mean of the squared errors of the forecasted number of passengers is 42 978, which corresponds to an absolute error percentage of 3,27.
6. The causal model

To account for omitted factors with possible effect on passenger flow, the causal model is defined as an extension of the airline model, thus a SARIMA(0,1,1)x(0,1,1)_{12} including explanatory variables in different time lags. This model is commonly denoted as a SARIMAX(0,1,1)x(0,1,1)_{12}.

6.1 Explanatory variables

There are several variables that might affect the number of monthly passengers. Latent factors important to incorporate in the model are determinants of ticket price, indicators of economic development and variations in occurrences of holidays. Since there are no close substitutes to international air travel, cross price elasticities are not considered. Possible explanatory variables, for which data are available during the investigated time period, are listed below. These are illustrated in Appendix B.

*Price of crude oil*

Previous research has found the price of crude oil to be almost perfectly correlated with the price of jet fuel. Since the ticket prices depend on factors such as destination, number of stop overs etc., it is not possible to compare one ticket to another. However, the fuel price is an operating cost for all companies and thus comparable. In 2007, jet fuel expenditure for the airline industry accounted for 29% of the operating costs, compared to approximately 15% in 2002 (IATA, 2014). The oil price might serve as a proxy for the ticket price; an increase in oil price increases the airline companies operating cost, which might be reflected in the ticket prices. Data on the variable comes from the World Bank, and is the monthly average price on Brent crude oil which originates from the North Sea and is the most commonly used benchmark for crude oil prices (U.S. EIA, 2014). However, the lag structure of the variable is not obvious since the refining of crude oil to jet fuel normally takes time, and thus the effects on ticket price could occur with delay. Furthermore, airlines adjust ticket prices today if the expected price of crude oil rises. In the latter case the effect is thus immediate. Airlines set the prices of the tickets far in advance, sometimes up to a year ahead, where forecasts of the oil price are taken into account.
Euro exchange rate

A great share of the international flights departing from Sweden has the euro area as destination. Thus, from the passengers’ point of view, if the SEK/EURO exchange rate depreciates, the cost of holidays decreases. The data is obtained from the Swedish central bank and is the monthly average. Variations in the exchange rate should reasonably affect air travel demand with delay since passengers normally buy tickets in advance. In the study by Castellani et al. (2010), it was found that the exchange rate GBP/EUR was to be lagged three months in order to best explain variation in arrivals of tourists from UK to Sicily.

Net trade

Previous studies have shown a strong correlation between net trade in a specific country and number of passengers. The Swedish aviation authority controls for the variable when predicting airline passengers. However, the report was done on yearly data, which might give a larger significance of the variable. The variable could be interpreted as a measure of economic growth in Sweden. The data is obtained from SCB as the monthly aggregate.

Unemployment rate

When a large share of the population lacks income, consumption is likely to be postponed according to basic economic theory. Problems may arise since the unemployment rate does not normally fluctuate much on a monthly basis, whereas the passenger flow does. Data on the variable is obtained from Eurostat and is given in monthly averages.

Holiday dummy variables

As seen in the naive model, occurrence of Easter had a large impact on the passenger flow in April, which led to an underestimation of the passengers in April and overestimation of the passengers in March. Since other holidays occur in specific months, Easter is the only holiday that needs to be accounted for. In order to control for this, an Easter dummy variable is constructed which takes on the value 1 if Easter occurs in April one year when the past Easter occurred in March and 0 otherwise. This is done since an SARIMA(0,1,1)x(0,1,1)_{12} only remembers the values in the previous month and value at the same months in the previous year.
6.2 Model estimation

Several models with different sets of explanatory variables and different lags have been estimated and evaluated and are found in Appendix C. The best fitted causal model is when the price of crude oil is not lagged, the exchange rate is lagged by four months and when Easter is controlled for. Surprisingly, neither net trade nor unemployment rate was significant on any reasonable level when estimated with different lags. This might be due to the lack of variation between months. The equation of the best estimated model is outlined below, from now on referred to as the causal model.

\[
\nabla_{12}\nabla_1 \ln \text{Passengers}_t = \alpha + e_t - \theta e_{t-1} - \Theta e_{t-12} + \Theta e_{t-13} + \beta_1 \nabla_{12}\nabla_1 \ln \text{Oil price}_t \\
+ \beta_2 \nabla_{12}\nabla_1 \ln \text{Exchange rate}_{t-4} + \beta_3 \text{Easter}_t + w_t
\]

As for passengers, all explanatory variables have been transformed in the same way, thus by first taking the logarithm and then yearly and monthly differences respectively. This is because the estimated relationships should be comparable to a situation where none of the variables were transformed. The estimated causal model is described in Table 3.

Table 3. Estimated causal model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.26</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>SMA(12)</td>
<td>0.53</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Oil price</td>
<td>0.06</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Exchange rate (-4)</td>
<td>-0.25</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>Easter dummy</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Sample size 203

\( R^2 \) 0.30

F-statistic 17.58
Important to notice from Table 3 is that the estimated coefficients are related to logged and differenced variable values since the forecast should be done on stationary data. Thus, interpretation of the coefficients is in terms of partial elasticities, before the anti-transformation of data once the forecasts are made.

When the euro is cheap, travelling to countries using the currency becomes cheaper why air passengers to such destinations should increase, thus the coefficient is negative and implying that a 1% increase in the exchange rate yields 0.25% decrease in passengers.

The effect of oil price on passenger flow is positive, which is contrary to expectations. A possible explanation of the positive coefficient might be that the oil price serves as an indicator of the overall economic development in the world, which yields a positive coefficient in the model with a 1% increase indicating a 0.06% increase in passengers.

6.3 Causal model forecast

The same data transformation as for the naive model has been applied in order to obtain point predictions in terms of passengers. A time series graph of the causal forecasted values and the actual number of passengers is displayed in Figure 11.

*Figure 11. Time series of causal forecast vs. actual values*
Point predictions of the causal model follow the actual values to quite a good extent. The largest deviations seem to occur during the summer months and in April. The model underestimates the number of passengers in the summer of 2014; hence people appear to have travelled more during the summer of 2014 compared to previous summer months. The model continues to underestimate the passengers in April, even though Easter was controlled for.

Table 4. Causal forecasts vs. actual values

<table>
<thead>
<tr>
<th>Month 2014</th>
<th>Actual values</th>
<th>Causal forecast</th>
<th>Difference</th>
<th>Difference in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>787 621</td>
<td>804 094</td>
<td>+ 16 473</td>
<td>+ 2,1</td>
</tr>
<tr>
<td>February</td>
<td>817 609</td>
<td>818 077</td>
<td>+ 468</td>
<td>+ 0,1</td>
</tr>
<tr>
<td>March</td>
<td>915 594</td>
<td>943 451</td>
<td>+ 27 857</td>
<td>+ 3,0</td>
</tr>
<tr>
<td>April</td>
<td>1 054 043</td>
<td>983 074</td>
<td>- 70 969</td>
<td>- 6,7</td>
</tr>
<tr>
<td>May</td>
<td>1 159 433</td>
<td>1 210 272</td>
<td>+ 50 839</td>
<td>+ 4,4</td>
</tr>
<tr>
<td>June</td>
<td>1 372 414</td>
<td>1 345 300</td>
<td>- 27 114</td>
<td>- 2,0</td>
</tr>
<tr>
<td>July</td>
<td>1 339 337</td>
<td>1 305 858</td>
<td>- 33 479</td>
<td>- 2,5</td>
</tr>
<tr>
<td>August</td>
<td>1 212 215</td>
<td>1 226 971</td>
<td>+ 14 756</td>
<td>+ 1,2</td>
</tr>
<tr>
<td>September</td>
<td>1 208 302</td>
<td>1 250 713</td>
<td>+ 42 411</td>
<td>+ 3,5</td>
</tr>
</tbody>
</table>

RMSE       | 37 329         |
MAPE       | 2,83           |

As for the naive model forecast, the point prediction of February is fairly good with a difference of only 468 passengers. The largest deviation occurs in April, as was shown in Figure 11, where the difference is -70 969 passengers. The root mean square error is 37 329 passengers, which corresponds to a MAPE of 2,83 %.
7. Evaluation

*Figure 12. Time series graph of comparison between actual and forecasted values*

Both models’ point predictions are plotted with the actual series in Figure 12, where it can be seen that both forecasts underestimate the number of passengers in April, June and July while overestimating them in other months. In general, the causal forecast lies between the naive forecast and the actual number of passengers, indicating a forecast improvement.
When forecasting data with distinct seasonal patterns and an overall increasing trend, the naive model predicts fairly well. The RMSE gives a root of the squared average deviation of 42 978 passenger per month, or 3.2 % in mean absolute percentage terms, compared to the causal model of 37 329 or 2.8 %. The causal model forecast outperforms the naive in seven out of nine months with an improvement of 5649 passengers per month in root mean square errors or with 0.42 error percentage units.

The causal model is used to forecast the passenger flows 12 months ahead;

\[ \nabla_{12} \nabla_1 \ln \text{Passengers}_t \\
= \alpha + e_t - \theta e_{t-1} - \Theta e_{t-12} + \Theta \theta e_{t-13} + \beta_1 \nabla_{12} \nabla_1 \ln \text{Oilprice}_t \\
+ \beta_2 \nabla_{12} \nabla_1 \ln \text{Exchange rate}_{t-4} + \beta_3 \text{Easter dummy} + w_t \\
\]

Since the oil price is non-lagged, the forecasts require predictions of the explanatory variables for each month. These are conducted by the U.S. Energy Information Administration. The exchange rate is lagged by four months, why predictions of the last five forecasts will be required for the variable. These are conducted by, and obtained from, Handelsbanken.

The forecast is carried out in the same way as previously, the only difference being the longer sample length and, of course, the lack of actual values to evaluate the predictions by.

Figure 13. Monthly departing passenger forecast in Sweden. October 2014 – September 2015
Table 6. Forecasted values for October 2014- September 2015 and actual values for corresponding month the previous year

<table>
<thead>
<tr>
<th>Month</th>
<th>Forecast</th>
<th>Previous year</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 2014</td>
<td>1 154 259</td>
<td>1 140 912</td>
<td>+ 13 347</td>
</tr>
<tr>
<td>November 2014</td>
<td>887 616</td>
<td>868 770</td>
<td>+ 18 846</td>
</tr>
<tr>
<td>December 2014</td>
<td>918 301</td>
<td>926 444</td>
<td>- 8 143</td>
</tr>
<tr>
<td>January 2015</td>
<td>793 918</td>
<td>787 621</td>
<td>+ 6 297</td>
</tr>
<tr>
<td>February 2015</td>
<td>824 350</td>
<td>817 609</td>
<td>+ 6 741</td>
</tr>
<tr>
<td>March 2015</td>
<td>939 204</td>
<td>915 594</td>
<td>+ 23 610</td>
</tr>
<tr>
<td>April 2015</td>
<td>1 046 683</td>
<td>1 054 043</td>
<td>- 7 360</td>
</tr>
<tr>
<td>May 2015</td>
<td>1 170 292</td>
<td>1 159 433</td>
<td>+ 10 859</td>
</tr>
<tr>
<td>June 2015</td>
<td>1 369 714</td>
<td>1 372 414</td>
<td>- 2 700</td>
</tr>
<tr>
<td>July 2015</td>
<td>1 333 193</td>
<td>1 339 337</td>
<td>- 6 144</td>
</tr>
<tr>
<td>August 2015</td>
<td>1 220 346</td>
<td>1 212 215</td>
<td>+ 8 131</td>
</tr>
<tr>
<td>September 2015</td>
<td>1 215 738</td>
<td>1 208 302</td>
<td>+ 7 436</td>
</tr>
</tbody>
</table>

Forecast sample: Jun 1997–Sep 2014

The model captures the effect of Easter holidays, which will occur in April 2015. The point prediction for this month is adjusted by the forecast of April 2015, in comparison to the 2014 forecast where the estimated number of passengers was 981 000. The other point predictions lie within the trend of the series. An overall increase in passengers is predicted for August 2015 to November 2015, which might be due to increasing weekend travelling which is popular during the fall. The point prediction for March is less than the 2014 forecast, but there is no way of determining the accuracy.
9. Analysis

9.1 Results

The oil price was assumed to serve as a proxy for the ticket price since the airlines’ operational costs increase when the oil price does. This relationship was expected to be reflected in the ticket prices, and thus have a negative effect on the number of passengers, but the positive coefficient proved us wrong.

Previous studies where the oil price has been included as an explanatory variable have faced the same problem, with the possible explanation that the oil price serves as an indicator of the overall state of the global market. When economies grow, demand for oil increases which puts an upward pressure on the price. In the data used to estimate the model, increases in the oil price occur during global periods of economic prosperity. In such a period, the number of passengers is likely to increase simultaneously with the oil price. Data for historical oil prices shows a peak in August 2008, followed by declining prices. Up to this point in time, both the number of monthly passengers and the oil price increased steadily. After the financial crisis, the oil price as well as the number of passengers decreased dramatically, resulting in an even stronger positive correlation, although the decline in passengers was a result of other factors.

If a pre- and post-2008 model were to be constructed, the effect of oil price could possibly differ.

The exchange rate had the expected negative coefficient since it is measured in SEK/EUR. When the euro becomes more expensive relative to SEK, this reduces the number of monthly passengers with a lag of 4 months. The effect of an increase in the exchange rate goes in two ways; when it becomes cheaper to travel to countries using euro as the currency, Swedes are more likely to travel to such countries. The overall vacation cost decreases as hotel stays become cheaper etc. Meanwhile, it becomes more expensive for tourists to visit Sweden. Since tourists are included in the data when they leave Sweden, the negative coefficient indicates that the effect is dominant for Swedes travelling abroad.
The variable controlling for Easter also has the expected positive sign. If a year when Easter occurred in March is followed by a year when Easter occurred in April, this yields a positive effect on the number of passengers in April. When scrutinizing the original data for monthly passengers, March is barely affected by the occurrence of Easter, while April is. A possible explanation might be that April is a warmer month than March, which is why holiday planners are more likely to adjust vacation plans depending on which month Easter occurs in. A week’s leisure in April makes people more likely to travel abroad than a week’s leisure in March.

During the time period 1996 to 2013, only four Easters have occurred in March. The model captures the Easter effect, but underestimates it. When making the 2015 forecast, all observations are included in the model, hence another rare Easter observation. This results in an Easter effect significant on the 1% level, compared to the forecast for 2014 where the effect was significant on the 10% level. Since future Easter dates are known, this is a simple way to improve forecasts.

9.2 Differences from previous studies

Kan Tsui et al. (2014) and Carson et al. (2011) made individual forecasts for airports’ passenger flow, which allowed for more specific explanatory variables. In the Hong-Kong forecast, passenger flow was categorized into 11 regions, why it was possible to include explanatory variables for these specific routes. Furthermore, HKIA is a major international airport hub for Asia, why connecting flights play an important role when estimating the passenger flow. Sweden lacks such airports, due to the isolated location in comparison to Hong-Kong.

Carson et al. (2011) made a similar forecast as for Hong-Kong in the sense that disaggregated data were used. When each airport is treated individually, the region served by a specific airport has a number of geo-economic factors that differ from other regions.

Due to the more homogeneous economic structure in Sweden, it is complicated to differentiate among any region specific criteria. Even though such factors are not equal for all regions in Sweden, there are larger economic differences among the states in the US. In addition to these differences, passengers in Sweden are more dependent on the largest airports. Passengers from northern Sweden are likely to use Arlanda as a transit airport, and
thus would be counted as if they belonged to the region served by Arlanda if the disaggregate approach was to be used.

The results from previous studies are presented in various forms, and most of them are not easily comparable with the results in this paper since forecast performances are evaluated individually for different airports. The forecast made on passenger flows from Hong-Kong airport is most similar in the instruments of forecast evaluation, where the average monthly deviation was 3.6%, compared to MAPE in this paper of 2.83.

9.3 Possible improvements of the model

A variable that most certainly affects the number of passengers, but intentionally has been excluded from the model, is the weather conditions during the summer months. A rainy July is likely to increase the number of passenger travelling abroad and vice versa. Since it is impossible to forecast the amount of rainfall or the average temperature next summer, this variable has intentionally been excluded. The best weather forecast for next summer is this summer’s weather, meaning that such model would assume constant weather conditions, which probably would do more harm than good. Therefore, the chosen model makes the forecasts regardless of weather conditions.

A more accurate forecast would focus on specific regions to which passengers travel. Consider the route Arlanda – Zurich. The ticket prices would be comparable over time, thus the price elasticity would be possible to estimate. Since ticket prices are set in advance, such a forecast can probably be rather accurate. The effect of a change in the exchange rate SEK/CHF would probably be larger than the SEK/EUR used in this paper, and the CPI of the countries would give an accurate reflection of the differences in price levels. The GDP development could be used as a measure of the aggregated income, but since data is available only for quarterly observations, it would not be possible to use for monthly data.

The disaggregated approach would probably increase the accuracy of the forecast, and could be of interest for airline companies to forecast specific routes. It could also be used to compute aggregated forecasts of interest for the Swedish transport agency and other stakeholders. However, such an approach is often time- and data consuming, and beyond the ambitions of this paper.
10. Conclusion

The forecast of departing passengers is improved by taking into account changes in oil price, the SEK/EUR exchange rate while controlling for the occurrence of Easter. The naive model forecast is outperformed in seven out of nine months. The actual values of the explanatory variables were used to determine whether it is advantageous to use them as input. This should be considered when evaluating the forecast, since it is impossible in out of sample forecasts.
References


University of Bamberg, Department of Liturgy Sciences, ”Easter Dates from 1901-2078”. http://www.maa.mhn.de/StarDate/publ_holidays.html (Obtained 2014-12-03).

Appendix A – Hypothesis tests

The following tests have been made for all data used in this paper, but only presented for the passenger series. For the exchange rate, oil price and the causal model error terms, all tests were fulfilled.

1. ADF unit root test of time series

The ADF is not approximately t-distributed under the null hypothesis; it has a non-standard large-sample distribution under the null hypothesis of a unit root.

\[ H_0 : \text{The series has a unit root} \]
\[ H_a : \text{The series is stationary} \]

\( H_0 \) is rejected if the p-value of the ADF test statistic < \( \alpha \)

Significance level: \( \alpha = 0,05 \)

<table>
<thead>
<tr>
<th>Null Hypothesis: DIFF1_12LOGP has a unit root</th>
<th>t-Statistic</th>
<th>Prob.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exogenous: Constant</td>
<td>Augmented Dickey-Fuller test statistic</td>
<td>-17.16426</td>
</tr>
<tr>
<td>Lag Length: 0 (Automatic - based on SIC, maxlag=14)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test critical values:

- 1% level: -3.462737
- 5% level: -2.875680
- 10% level: -2.574385

The null hypothesis is rejected in favour of the alternative hypothesis. Thus the series is claimed to be stationary.

2. Ljung-Box test for autocorrelation of error terms

The LB-Q test statistic follows a \( \chi^2(m) \) distribution under the null hypothesis.

\[ H_0 : \text{The data is independently distributed} \]
\[ H_a : \text{The data is not independently distributed} \]

Test statistic: \( Q_{LB} = n(n + 2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k} \)
where \( n \) is the sample size, \( \hat{\rho}_k^2 \) is the sample autocorrelation at lag \( k \), and \( m \) is the number of lags being tested.

\( H_0 \) is rejected if \( Q_{LB} > \chi^2_{1-\alpha, m} \)

Since the \( Q_{LB} < \chi^2_{1-\alpha, m} \), the null hypothesis is not rejected and thus the data are considered to be independently distributed.

3. Jarque-Bera test of error terms

The test statistic approximately follows a chi-square distribution with two degrees of freedom.

\[
H_0 : \text{The error terms follow a normal distribution} \\
H_a : \text{The error terms do not follow a normal distribution}
\]

Test statistic: \( JB = \frac{n\hat{g}_1^2}{6} + \frac{n\hat{g}_2^2}{24} \)

\( H_0 \) is rejected if \( p < \alpha = 0.05 \).

The test rejects the null hypothesis of normally distributed residuals. However, the distribution shows signs of a normal distribution. The negative skewness is probably due to the large deviations in 2001, 2003 and 2008 (9/11, SARS and the financial crisis). Since the Q-Q plot presented in the figure above indicates normality among the residuals, it is not reasonable that assuming normality of the error terms is to cause large bias for the estimation.
Appendix B – Time series plots of the explanatory variables

OILPRICE

UNEMPLOYMENT
Appendix C – Estimated causal models

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td>(0.73)</td>
<td>(0.73)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.26</td>
<td>0.27</td>
<td>0.27</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>SMA(12)</td>
<td>0.53</td>
<td>0.55</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Oil Price</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Oil Price (-1)</td>
<td></td>
<td></td>
<td></td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.64)</td>
</tr>
<tr>
<td>Exchange rate(-1)</td>
<td></td>
<td></td>
<td></td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.56)</td>
</tr>
<tr>
<td>Exchange rate (-2)</td>
<td></td>
<td></td>
<td>-0.22</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Exchange rate (-4)</td>
<td>-0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easter dummy</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td></td>
<td></td>
<td></td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.22)</td>
</tr>
<tr>
<td>Unemployment(-1)</td>
<td></td>
<td></td>
<td></td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.42)</td>
</tr>
<tr>
<td>Unemployment (-2)</td>
<td></td>
<td></td>
<td></td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Net trade</td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.37)</td>
</tr>
<tr>
<td>Net trade(-1)</td>
<td></td>
<td></td>
<td></td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.42)</td>
</tr>
<tr>
<td>Net trade (-2)</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.69)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.30</td>
<td>0.24</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>F-statistic</td>
<td>17.58</td>
<td>17.29</td>
<td>11.59</td>
<td>7.71</td>
</tr>
</tbody>
</table>

41