Working Paper Series: Business and Law

Working Paper No. 05

Combination of Forecasts across Estimation Windows:
An Application to Air Travel Demand

Andre Jungmittag

July 2014

Combination of Forecasts across Estimation Windows: An Application to Air Travel Demand

Andre Jungmittag

FB 3: Business and Law

Frankfurt University of Applied Sciences

Frankfurt am Main

July 2014

This paper applies combining forecasts of air travel demand generated from the same model but over different estimation windows. The combination approach used resorts to Pesaran and Pick (2011), but the empirical application is extended in several ways. The forecasts are based on a seasonal Box-Jenkins model (SARIMA), which is adequate to forecast monthly air travel demand with distinct seasonal patterns at the largest German airport Frankfurt am Main. Furthermore, forecasts with forecast horizons from one to twelve months-ahead, which are based on different average estimation windows, expanding windows and single rolling windows, are compared with baseline forecasts based on an expanding window of the observations after a structural break. The forecast exercise shows that the average window forecasts mostly outperform the alternative single window forecasts.

Keywords: Air travel demand; Combination of forecasts; Estimation windows; Structural breaks.

JEL Classifikations: C22, C53, L93.

1 Introduction

Forecasting of air travel demand has a long tradition in time series analysis and applied econometrics. Long-term as well as short-term forecasts provide important information for a wide range of economic and business decisions (Carson, Cenesizoglu and Parker, 2011). Long-term forecasts are important for decisions with regard to the expansion of

airport facilities, research and development, airplane design and production planning, but also for regional regional planning of policy makers. Short-term forecasts provide important inputs for decisions with regard to the capacity and resource planning of airlines and airport operators, but also for their marketing measures. Furthermore, the air travel market is very sensitive to the prevailing business cycle and demands frequent updating of forecasts (Uddin, McCullough and Crawford, 1985).

A popular approach to forecast air travel demand in the short-run is the seasonal Box-Jenkins model (SARIMA). In this paper, it is shown that forecasts from such a model can be improved considerably by combining forecasts across different estimation windows. This new approach, proposed by Pesaran and Timmermann (2007), augments the idea of combining forecasts into a new direction. Since the seminal paper of Bates and Granger (1969), a extensive literature on the advantages of combining forecasts from different models has evolved. This literature shows empirically and – as far as possible – theoretically that, if forecasts are based on different explanatory variables and/or different assumptions about the relations between the variables, an averaging of forecasts with equal or estimated weights can outperform the individual forecasts (Jungmittag, 1998 and 2010).[1] Insofar, the combination of forecasts is a method to deal with uncertainty about the proper specification and choice of a model.

Pesaran and Timmermann (2007) argue that the forecast combination procedure can be extended to deal with other types of model uncertainty, such as uncertainty over the size of the estimation window. They propose the idea of averaging forecasts from the same model, but computed over different estimation windows. Using Monte Carlo experiments, they show that this type of forecast averaging reduces the mean square forecast error (MSFE) in many cases when the underlying relations are subject to structural breaks. Pesaran and Pick (2011) extend the approach of Pesaran and Timmermann (2007) and apply it to financial market data before and after the credit crunch in 2007 and 2008.

The main contributions of my paper to the existing literature can be summarized as follows. The empirical application in Pesaran and Pick (2011) is limited to a very simple model (random walk with drift) and one-step ahead forecasts. In my paper, I apply the more sophisticated seasonal Box-Jenkins model (SARIMA), which is adequate to forecast monthly air travel demand with distinct seasonal patterns at the largest German airport Frankfurt am Main. Furthermore, I compare multi-step forecasts up to twelve months. Since, in the forecasting exercise, the average window forecasts mostly outperform the alternative single window forecasts, the paper also contributes to the more general literature on forecasting air travel.

2 Methodological Issues

Pesaran und Pick (2011) argue that an attractive feature of combining forecasts across estimation windows is that no exact information about structural breaks is needed. This

---

[1]The literature is reviewed e.g. by Clemen (1989), Jungmittag (1998), Stock and Watson (2004), Timmermann (2006) and Jungmittag (2010).

feature distinguishes the combination approach from the usual approach of assessing or estimating the break points and basing the forecasts only on the post-break data. However, Pesaran and Timmermann (2007) showed that it is not always optimal to base forecasts only on the post-break observations. Actually, the use of pre-break observations biases the forecasts, but also decreases the forecast error variance. With regard to this trade-off between bias and variance of forecast errors, Pesaran and Timmermann (2007) as well as Pesaran and Pick (2011) showed that the overall effect of using pre-break data on the mean square forecast error (MSFE) depends on the size and the point of the break — and it is hard to assess the size of the break since it involves estimating the model over the pre- and post-break periods.[2] However, if the distance to the break (with viewing direction from the beginning of the forecast period) is short, the post-break parameters are likely to be poorly estimated.

To set up the basic approach of forecast averaging across estimation windows, let's consider the sample $\{y_t, x_t\}_{t=T_i+1}^{T}$, with $0 \leq T_i < T$, which provides us with an observation window of the size $W_i = T - T_i$.[3] Then, the fraction of observations in a single window is $w_i = (T - T_i)/T$. Starting the estimation process with a minimum window $\{y_t, x_t\}_{t=T_{\min}+1}^{T}$ of the size $w_{\min} = (T - T_{\min})/T$, larger windows can be considered successively with $T_i = T_{\min} - j, \ldots, T_{\min} - j(m-1)$, providing us with m separate estimation windows with j observations apart. Thus, the fractions of observations in the individual estimation windows evolve as

$$w_i = w_{\min} + \left(\frac{i-1}{m-1}\right)(1 - w_{\min}), \text{ for } i = 1, 2, \ldots, m, \tag{1}$$

so that $w_i \in [w_{\min}, 1]$, and $w_m = 1$ captures the whole sample. The number of estimation windows m can be kept constant as T changes or can be allowed to rise with T. In both cases it must hold $m \leq T(1 - w_{\min}) + 1$. In this case is

$$w_i = w_{\min} + \frac{i-1}{T}. \tag{2}$$

Based on this definition of estimation windows, the rule for combining forecasts can be set up straightforward. The one step-ahead average window forecast is given by

$$\hat{y}_{m,T+1} = \frac{1}{m} \sum_{i=1}^{m} \hat{y}_{T+1}(w_i), \tag{3}$$

where $\hat{y}_{T+1}(w_i)$ is the one step-ahead forecast from an estimation window w_i, and forecasts from all windows receive equal weights.

The aim of the following analysis is to compare single window forecasts and the average window forecasts. In recursive estimation, the single window can be an expanding or a rolling window, and average window forecasts can be computed by averaging over sub-windows within a given expanding or rolling window. Thus, the average window forecast is not an alternative to rolling forecasts (Pesaran und Pick, 2011).

---

[2]Further theoretical results with regard to this trade-off can be found in Clark and McCracken (2009).
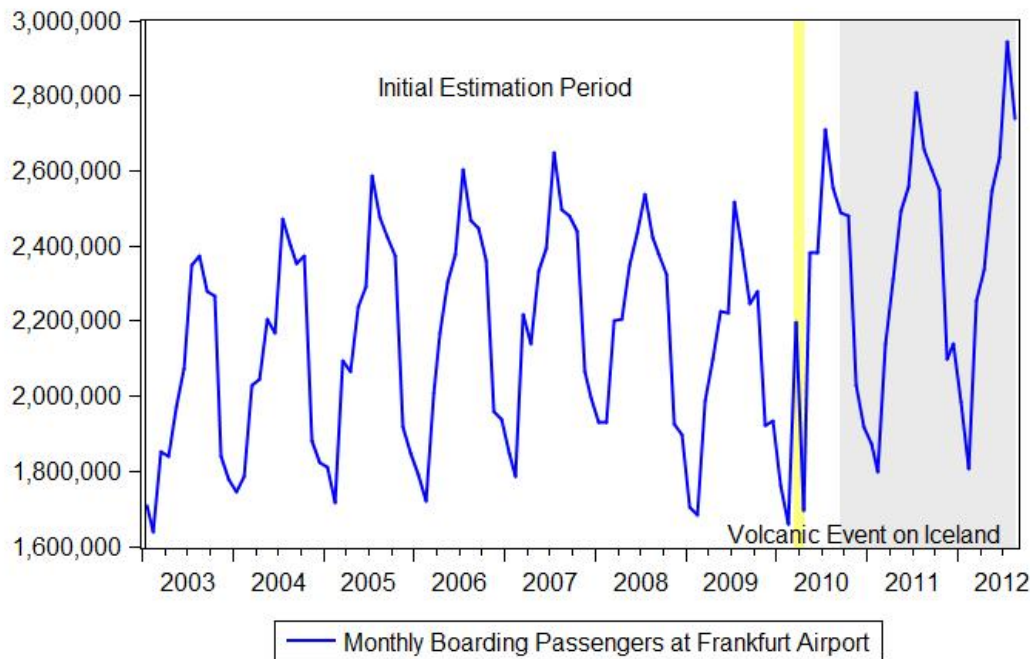[3]With regard to the notations and presentation of the basic model, I follow Pesaran and Pick (2011).

Figure 1: Boarding passengers per month at Frankfurt airport, January 2003 until August 2012

3 Specification of the SARIMA-Model and the Forecasting Exercise

In the following the average windows forecast approach is applied to the monthly air travel demand at the largest German airport Frankfurt am Main. Figure 1 shows the boarding passengers per month at this airport from January 2003 to August 2012. The time span up to September 2010 is used as initial estimation period and the grey shadowed time span is used to compute recursively one to twelve-month-ahead forecasts applying various forecasting methods for the specified seasonal autoregressive integrated moving average (SARIMA) model. The data show distinct seasonal patterns. Furthermore, there is an outlier in April 2010, caused by a volcanic event on Iceland. Finally, the time series also shows a temporary decrease of air travel demand in 2008 and 2009 associated with the worldwide financial crisis and reduction of economic growth.

The distinct seasonal patterns of air travel demand suggests the use of a SARIMA model which adopts the ARIMA framework to seasonal data. Since its introduction at the beginning of the 1970s, the SARIMA approach became a very popular class of models to forecast airline passengers. Thus, it is not surprising that the often used SARIMA(0,1,1)(0,1,1) model is also called airline model.[4] For monthly data, the

---

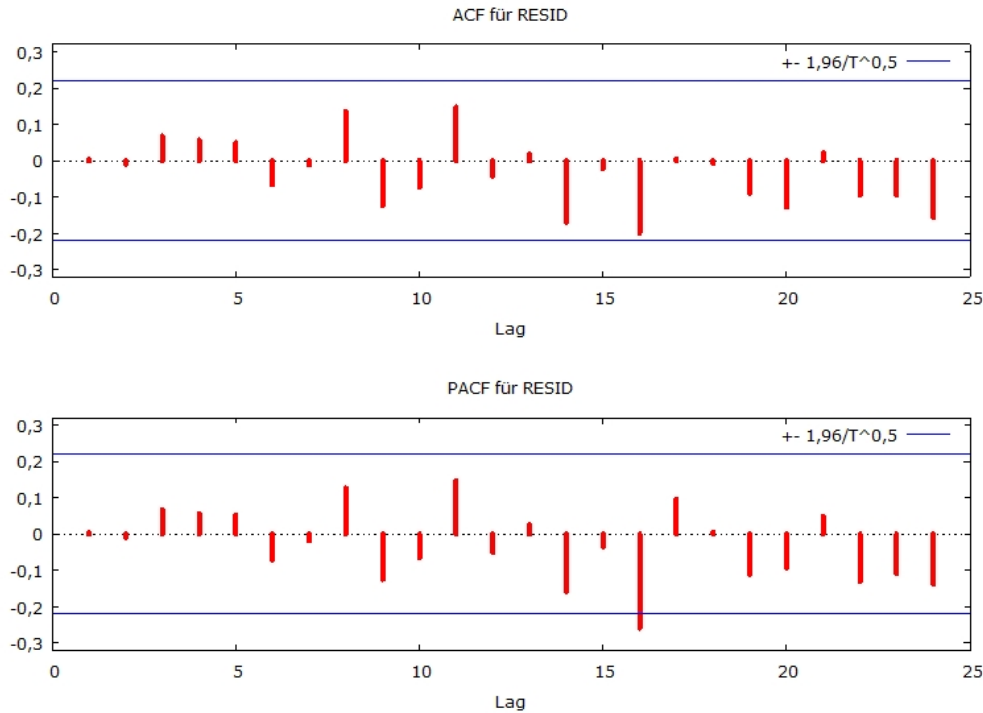[4]This model was originally proposed by Box and Jenkins (1970).

Figure 2: ACF and PACF for the residuals of the SARIMA model for the whole estimation period

SARIMA methodology examines the year-to-year relationships for each month of a time series (Box, Jenkins and Reinsel, 1994), thus capturing the seasonal relationship between observations for the same month (Y_t and Y_{t-12}) in successive years. Initially, the original time series is transformed linearly until the data is stationary and the estimated autocorrelation function (ACF) and partial autocorrelation function (PACF) show only a few significant, easily interpretable autocorrelations (Schulze and Prinz, 2009). The linear transformation is done by differencing the data. Non-seasonal as well as seasonal differencing can be applied. In the case of monthly data, seasonal differencing consists in subtracting the values of two observations for the same month in two successive years. Afterwards, the ACF and PACF can be used to identify the remaining AR- and MA-terms, which have to be estimated. In the end, the residuals of the model should have a zero mean, a constant variance, and should be serially independent.

Initially, a model for the seasonally differenced logarithmized data over the whole estimation period up to September 2010 is specified and estimated. It includes a constant, a dummy variable for the volcanic event on Iceland, a non-seasonal AR(1)- and MA(1)-term, as well as a seasonal MA(12)-term, which are all statistically highly significant. Figure 2 shows the ACF and PACF for the residuals of this model. They allow the conclusion that there is no remaining autocorrelation. However, a view on the actual and fitted data as well as the residuals reveals that there is a structural break after June 2008 (Figure 3).
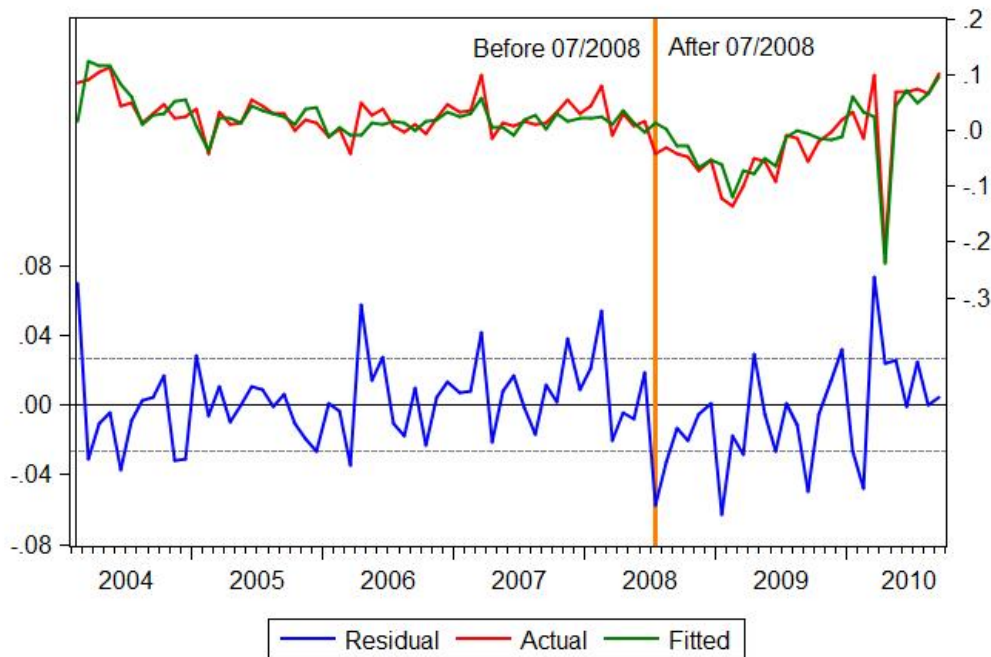
Figure 3: Actual and fitted data as well as residuals of the SARIMA model for the whole estimation period

Thus, the model was re-estimated for the post-break period from July 2008 to September 2010. Again, all coefficient estimates are statistically highly significant and the ACF as well as the PACF indicate the absence of autocorrelation for the residuals (Figure 4). Nevertheless, it can be assumed that the coefficients are only poorly estimated with the remaining 27 observations.

To compare the performance of the different single and average window forecasts, a forecasting exercise is carried out for short-term to medium-term forecasts horizons, 1 to 12 months ahead. The evaluation is based on recursive forecasts that involve an average of the respective horizon forecasts over twelve recursive windows. The one-month-ahead forecasts are starting with the forecast for October 2010 based on an estimation window up to September 2010, the second forecast is for November 2010 based on an estimation window up to October 2010, etc., the twelfth forecast for September 2011 is then based on an estimation window up to August 2011. Similarly, two- to twelve-months-ahead forecasts are carried out for twelve recursive windows. E.g. for the twelve-months-ahead forecasts, the forecast for September 2011 is based on an window up to September 2010, whereas the last twelve-months-ahead forecast is carried out for August 2012 based on an estimation window until August 2011.

In the forecasting exercise, forecasts from seven different methods are compared. The baseline forecasts (i) are based on an expanding window of the observations after the structural break (initially 27 observations). Average window forecasts from a mini-
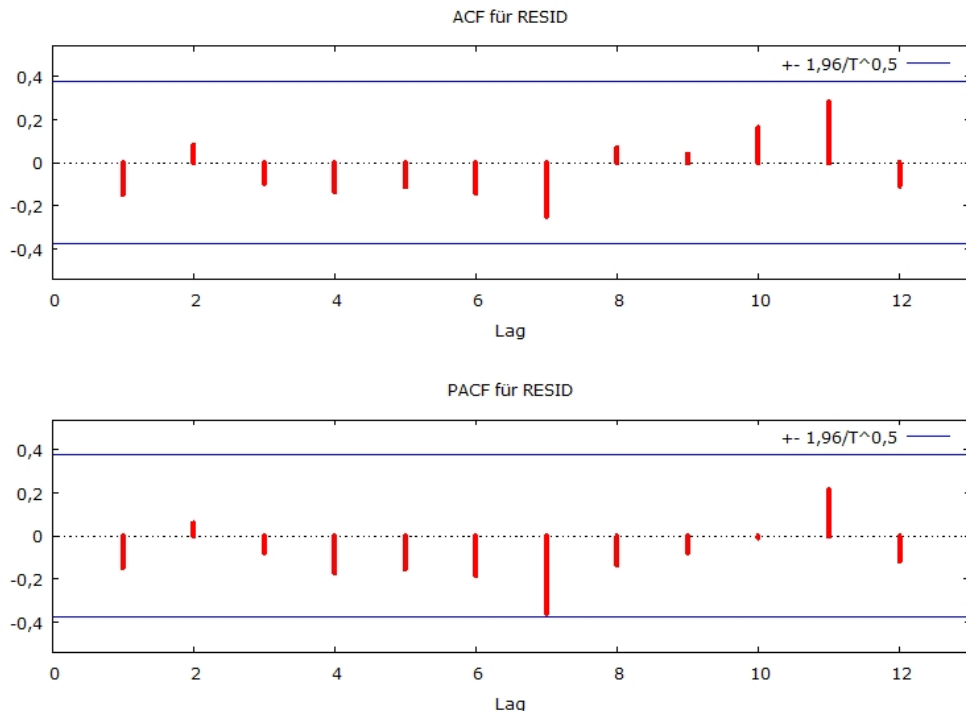
6

Figure 4: ACF and PACF for the residuals of the SARIMA model for the post-break estimation period

mum number of observations $W_{\min} = 27$ months and $m = 5$ sub-windows within (ii) a rolling window of length $W_a = 80$ months and (iii) an expanding window. Furthermore, forecasts from single expanding and rolling windows are included in the comparison. Forecasts from the single expanding window (iv) are always based on the available whole observation period. For the rolling windows, three variants are considered: (v) a minimum rolling window of size $W_{\min} = 27$ months, (vi) a rolling window of size $W_a = 80$ months, and a (vii) rolling window of (average) effective window size $\bar{W} = W_a(1/5 + 2/5 + \ldots + 5/5)/5 = 48$ months. The latter is included, as it could be argued that the average window forecasts are performing better because they are actually based on a smaller average window (compared to W_a) (see Pesaran and Pick, 2011).

For each series of forecasts, the mean absolute forecast error (MAFE), the root mean square forecast error (RMSFE), as well as the relative MAPE and RMSFE are computed. The MAFE is defined as

$$MAFE = \frac{1}{N} \sum_{n=1}^{N} \left| \hat{y}_{T-1+n+i|T-1+n} - y_{T-1+n+i} \right|, \tag{4}$$

where n is the number forecasts, T is the end of the estimation window for the first forecast, and i is the step size of a forecast. Thus, $\hat{y}_{T-1+n+i|T-1+n}$ is the i-months-ahead

7

Table 1: Performance of the alternative forecasts: MAFE

Steps ahead	Post-break	Average windows		Expanding windows	Rolling windows		
		Rolling	Expanding		$W_{\min} = 27$	$\bar{W} = 48$	$W_a = 80$
1 month	64277	49339	47614	43396	60794	51210	44854
2 months	79597	57010	56559	54164	83141	67610	54828
3 months	83537	55388	56437	61410	75559	69548	57995
4 months	73579	73159	73331	84527	77351	91605	83210
5 months	87276	78934	77843	98291	88956	100092	92956
6 months	108776	75016	78903	109425	106511	103383	100409
7 months	101596	65773	70827	117793	92595	117097	107684
8 months	123635	70153	72195	116980	111720	123557	108958
9 months	146187	70084	64992	117594	125096	121806	102977
10 months	153646	57786	50345	117040	135178	105738	99431
11 months	159308	45988	42217	123194	144905	89125	104865
12 months	173808	46803	50210	127939	164074	75914	110093

forecast for month $T - 1 + n + i$ based on the observations up to $T - 1 + n$. Similarly, the RMSFE is given by

$$RMSFE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(\hat{y}_{T-1+n+i|T-1+n} - y_{T-1+n+i} \right)^2}. \qquad (5)$$

4 Forecasting Results

Table 1 summarizes the mean absolute forecast errors (MAFE) for the various forecast procedures and different months-ahead forecasts. Looking at the one month-ahead forecasts, it is obvious that all other forecast procedures outperform the baseline forecasts (procedure i), based on an expanding window of the observations after the structural break, and the single minimum rolling windows forecasts (v), based on observation periods of 27 months. When the forecast horizon increases successively from one month-ahead to twelve months-ahead, the strongest increase of the MAFE can be observed for the baseline forecasts (i) and the single minimum rolling windows forecasts (v). On the other hand, both average window forecasts show the smallest increase of the MAFE. While the MAFE of the baseline forecasts and the minimum rolling windows forecasts both increase by 170 percent from the one month-ahead to the twelve months-ahead forecasts, the MAFE of the rolling average window forecasts decrease by five percent and the MAFE of the expanding average window forecasts only increase by 5 percent. Compared to their overall performance, the average window four and five months-ahead forecasts are a little bit flagging, but that holds also for most other forecast procedures.

Additionally, Table 2 reports the relative MAFE, i.e. the MAFE of a certain forecast procedure in relation to the MAFE of the baseline forecasts (procedure i). Again, it is obvious that the average window forecasts unfold their merits with a rising forecast horizon. For the one month-ahead forecasts the MAFE of the rolling and expanding average window forecasts (ii and iii) are 23.2 and 25.9 percent lower than the MAFE

Table 2: Performance of the alternative forecasts: Relative MAFE

Steps ahead	Post-break	Average windows		Expanding	Rolling windows		
		Rolling	Expanding	windows	$W_{\min} = 27$	$\bar{W} = 48$	$W_a = 80$
1 month	1	0.768	0.741	0.675	0.946	0.797	0.698
2 months	1	0.716	0.711	0.680	1.045	0.849	0.689
3 months	1	0.663	0.676	0.735	0.904	0.833	0.694
4 months	1	0.994	0.997	1.149	1.051	1.245	1.131
5 months	1	0.904	0.892	1.126	1.019	1.147	1.065
6 months	1	0.690	0.725	1.006	0.979	0.950	0.923
7 months	1	0.647	0.697	1.159	0.911	1.153	1.060
8 months	1	0.567	0.584	0.946	0.904	0.999	0.881
9 months	1	0.479	0.445	0.804	0.856	0.833	0.704
10 months	1	0.376	0.328	0.762	0.880	0.688	0.647
11 months	1	0.289	0.265	0.773	0.910	0.559	0.658
12 months	1	0.269	0.289	0.736	0.944	0.437	0.633

Table 3: Performance of the alternative forecasts: RMSFE

Steps ahead	Post-break	Average windows		Expanding	Rolling windows		
		Rolling	Expanding	windows	$W_{\min} = 27$	$\bar{W} = 48$	$W_a = 80$
1 month	90816	64098	61128	61668	84312	67355	60563
2 months	103286	73672	71984	77775	103306	96760	75277
3 months	112188	74856	75890	90780	104359	99615	85043
4 months	99680	84139	84097	107060	101631	123152	102443
5 months	125019	92025	90159	115322	128209	129041	109101
6 months	156889	90692	92873	121611	154600	128613	111330
7 months	136701	80408	82844	126842	132205	139295	117477
8 months	158673	80088	82336	126465	152381	135402	118884
9 months	175420	87329	83083	129755	161870	137623	120016
10 months	201446	73081	68540	126337	194643	122990	114611
11 months	209968	60674	61659	131191	204749	102250	114845
12 months	230733	68609	70230	137322	226936	100881	119217

of the baseline forecasts (i). From the three months-ahead to the twelve months-ahead forecasts the average window procedures outperform all other forecast procedures to an increasing degree. For the twelve months-ahead forecasts, the MAFE of the two average window forecast procedures are 73.1 and 71.1 percent lower than the MAFE of the baseline forecasts.

The root mean square forecast errors (RMSFE) displayed in Table 3 can provide same further insight into the performance of the different forecast procedures, because this measure weights larger forecast errors stronger than smaller ones, while the MAFE weights each error identically. On the whole, the RMSFE largely confirm the conclusions drawn from the MAFE. The only difference is perhaps that the average window four and five months ahead forecasts are not flagging so strongly as suggested by the MAFE. Finally, Table 4 shows the relative RMSFE, again calculated as the RMSFE of a certain in relation to the RMSFE of the baseline forecasts (i). Referred to this forecast

Table 4: Performance of the alternative forecasts: Relative RMSFE

Steps ahead	Post-break	Average windows		Expanding windows	Rolling windows		
		Rolling	Expanding		$W_{\min} = 27$	$\bar{W} = 48$	$W_a = 80$
1 month	1	0.706	0.673	0.679	0.928	0.742	0.667
2 months	1	0.713	0.697	0.753	1.000	0.937	0.729
3 months	1	0.667	0.676	0.809	0.930	0.888	0.758
4 months	1	0.844	0.844	1.074	1.020	1.235	1.028
5 months	1	0.736	0.721	0.922	1.026	1.032	0.873
6 months	1	0.578	0.592	0.775	0.985	0.820	0.710
7 months	1	0.588	0.606	0.928	0.967	1.019	0.859
8 months	1	0.505	0.519	0.797	0.960	0.853	0.749
9 months	1	0.498	0.474	0.740	0.923	0.785	0.684
10 months	1	0.363	0.340	0.627	0.966	0.611	0.569
11 months	1	0.289	0.294	0.625	0.975	0.487	0.547
12 months	1	0.297	0.304	0.595	0.984	0.437	0.517

performance measure, both average window forecast procedures outperform all other forecast procedures already from the two months-ahead forecasts upwards. Furthermore again, both average window forecast procedures achieve the best relative performance at the eleven and twelve months-ahead forecasts. Here, the RMSFE are approximately 70 percent lower that the RMSFE of the baseline forecasts, and still 30 to 40 percent lower than the RMSFE of the rolling windows forecasts based on a window size of 48 months.

5 Conclusion

In this paper, I analyzed whether forecasts of air travel demand—which is very sensitive to business cycles and perhaps structural breaks—can be improved by combining forecasts across different estimation windows. One result is very obvious: In relation to the MAFE and RMSFE, nearly every thing provides better forecasts than just using the observations from the post-break period. However, the proposed average window forecasts mostly outperform the alternative single window forecasts. Furthermore, the results reveal that the average window forecasts unfold their whole merits with a rising forecast horizon: the relative performance of these forecasts is the stronger, the longer the forecast horizon.

More generally, averaging of forecasts over different estimation windows offers a simple approach to generating forecasts that are reasonably robust to structural breaks of unknown dates and sizes. Therefore, the average windows approach surely would be fruitful for other market or sales forecast tasks. The approach also will likely improve the forecast accuracy of other models, e.g. with independent explanatory variables. This is left for future research, but first results in this direction give reason for well-founded hope (cf. Giraitis, Kapetanios and Price, 2013; Pesaran, Pick and Pranovich, 2013; Tian and Anderson, 2014).

References

Bates, J. M., Granger, C. W. J. (1969), The Combination of Forecasts, Operational Research Quarterly 20, 452-468.

Box, G. E. P., Jenkins, G. M. (1970), Time Series Analysis: Forecasting and Control, San Francisco, Holden Day.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994), Time Series Analysis: Forecasting and Control (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall, Inc.

Carson, R. T., Cenesizoglu, T., Parker, R. (2011), Forecasting (Aggregate) Demand for US Commercial Air Travel, International Journal of Forecasting 27, 923-941.

Clark, T. E., McCracken, M. W (2009), Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts, International Economic Review 50, 363-395.

Clemen, R. T. (1989), Combining Forecasts: A Review and Annotated Bibliography, International Journal of Forecasting 5, 559-583.

Giraitis, L., Kapetanios, G., Price, S. (2013), Adaptive Forecasting in the Presence of Recent and Ongoing Structural Change, Journal of Econometrics 177, 153-170.

Jungmittag, A. (1998), Combination of Forecasts, in: Kotz, S. et al. (Eds.), Encyclopedia of Statistical Sciences, Update Volume 2, New York et al.: John Wiley and Sons, 258-263.

Jungmittag, A. (2010), Combination of Forecasts, in: Balakrishnan, N. (Ed.), Methods and Applications of Statistics in Business, Finance and Management Sciences, New York et al.: John Wiley and Sons, 116-121.

Pesaran, M. H., Pick, A. (2011), Forecast Combination across Estimation Windows, Journal of Business Economics and Statistics 29, 307-318.

Pesaran, M. H., Pick, A., Pranovich, M. (2013), Optimal Forecasts in the Presence od Structural Breaks, Journal of Econometrics 177, 134-152.

Pesaran, M. H.,Timmermann, A. (2007), Selection of Estimation Window in the Presence of Breaks, Journal of Econometrics 137, 134-161.

Schulze, P., Prinz, A. (2009), Forecasting Container Transshipment in Germany, Applied Economics 41, 2809-2815.

Stock, J. H., Watson, M. W. (2004), Combination Forecasts of Output Growth in a Seven-Country Data Set, Journal of Forecasting 26, 405-430.

Tian, J., Anderson, H. M. (2014), Forecast Combinations under Structural Break Uncertainty, International Journal of Forecasting 30, 161-175.

Uddin, M., McCullough, B. F., Crawford, M. M. (1985), Methodology for Forecasting Air Travel and Airport Expansion Needs, Transportation Research Record 1025, 7-14.