Online Supplement: Can Financial Markets Inform Operational Improvement Efforts? Evidence from the Airline Industry

Kamalini Ramdas*, Jonathan Williams†, and Marc Lipson‡

February 2013

Estimation of Abnormal Returns

Our discussion in the main text focused on abnormal returns estimated using the Carhart (1997) four-factor model,

$$ExRet_{it} = \gamma_{0i} + \gamma_{1i}Mkt_t + \gamma_{2i}Smb_t + \gamma_{3i}Hml_t + \gamma_{4i}Mom_t + AbnRet_{it},$$

where $ExRet_{it}$ is airline i's excess return in month t, which is defined as airline i's return in month t minus the risk-free return; Mkt_t is the excess return (over the risk-free return) on the market portfolio in month t; Smb_t is the average return on the three small cap portfolios minus the average return on the three large cap portfolios; Hml_t is the average return on the two value portfolios minus the average return on the two growth portfolios;[1] and Mom_t is the average of the returns on two high prior return portfolios minus the average of the returns on two low prior return portfolios. The residuals from the above regression, $\widehat{AbnRet}_{it}$, are the abnormal returns that we use as the dependent variable in our empirical analysis.

In examining the sensitivity of our results, we experimented with three alternative ways to estimate abnormal returns. We first use the 3-factor model of Fama and French (1993),

$$ExRet_{it} = \gamma_{0i} + \gamma_{1i}Mkt_t + \gamma_{2i}Smb_t + \gamma_{3i}Hml_t + AbnRet_{it},$$

which drops the Mom_t factor from the four-factor model. We also estimate a more parsimonious market model,

$$ExRet_{it} = \gamma_{0i} + \gamma_{1i}CRSPIndex_t + AbnRet_{it},$$

*Management Science and Operations, London Business School, kramdas@london.edu.
†Department of Economics, Terry College of Business, University of Georgia, jonwms@uga.edu.
‡Darden School of Business, University of Virginia, mlipson@virginia.edu.

[1] See Fama and French (1993) for a complete description of the Fama-French factors.

where $CRSPIndex_t$ is the CRSP US Total Market Index. Finally, we consider a simple industry model where abnormal returns are calculated as the deviation from average returns in the industry,

$$AbnRet_{it} = ExRet_{it} - \frac{1}{I_t}\sum_{i=1}^{I_t} ExRet_{it},$$

where I_t is the number of firms at time t.

We ran eleven regressions, one per carrier, to estimate abnormal returns using each model. We use the entire time series for each carrier (1990-2009) to calibrate the parameters of the model. This is efficient if the parameters are stable over time, or airlines tend to load similarly on each of the factors throughout our panel. To formally test this assumption, we use Andrews (1993) test for parameter instability (with an unknown break date). Delta is the only carrier for which we are able to reject the null of time-invariant parameters, at the 10% level. Repeating the entire analysis allowing for a break at the estimated break date yields nearly identical second-stage results. We also performed the analysis allowing for parameters to differ before and after mergers (e.g. U.S. Airways and America West), and find no change in our results. For this reason, we focus our discussion of the model with time-invariant parameters. We also tested for serial correlation and found none.

We observed that in most cases the airlines load positively on Mkt (excess returns on market portfolio) and $CRSPIndex$, positively on Hml (returns on a portfolio long in high market-to-book ratio stocks and short in low market-to-book ratio stocks) and negatively on Mom (momentum). None of the airlines load on Smb (returns on a portfolio long in small stocks and short in large stocks), which is not surprising since companies in this industry are fairly large.

For example, using the four-factor model for American Airlines, we observe the results below (standard errors under parameter estimates):

$$E[ExRet_{it} \mid X_t] = \underset{(0.010)}{0.001} + \underset{(0.244)}{1.892} Mkt_t + \underset{(0.299)}{0.477} Smb_t + \underset{(0.327)}{1.195} Hml_t - \underset{(0.196)}{0.691} Mom_t$$

In these carrier-level regressions, the adjusted R-squared values range from 0.158 for Alaska Airlines to 0.425 for Northwest Airlines. American is in the middle of this range with an adjusted R-squared of 0.315. For American Airlines, the estimates for the three-factor model are

$$E[ExRet_{it} \mid X_t] = \underset{(0.010)}{-0.005} + \underset{(0.235)}{2.189} Mkt_t + \underset{(0.305)}{0.431} Smb_t + \underset{(0.326)}{1.461} Hml_t,$$

while the market model yields

$$E[ExRet_{it} \mid X_t] = \underset{(0.010)}{-0.001} + \underset{(0.231)}{1.951} CRSPIndex_t.$$

The three-factor and market models explain less variation than the four-factor model, with adjusted R-squared values ranging from 0.143 for Alaska Airlines to 0.348 for United Airlines and 0.136 for

Alaska Airlines to 0.331 for Northwest Airlines, respectively. The results of the industry model are difficult to describe, as there are no coefficients to report.

For our purposes, all that matters is the degree to which the alternative models give similar predictions of abnormal returns. In particular, if the results from each model are highly correlated (high degree of linear dependence), then which ones are used as the dependent variable in the regression analysis of abnormal returns on surprises in operational metrics is not an issue. The correlations between the alternative measures of abnormal returns are

Correlation among Alternative Abnormal Returns

	$4-$factor	$3-$factor	*Market*	*Industry*
$4-$factor	1			
$3-$factor	0.981	1		
Market	0.951	0.969	1	
Industry	0.830	0.831	0.836	1

Each measure of abnormal returns for our sample is very highly correlated with the others, with the industry model having the lowest correlation with all other measures. Consistent with the results of our sensitivity analysis in Table 7, the particular measure used in our analysis does not change the results in a meaningful way.

Prediction of Operational Variables

Much of our analysis relies on accurately forecasting operational variables. Our estimates of Equation 3 in Table 5 of the paper suggest that our forecasts are consistent with investors' expectations of operational performance and priced into the market (equal and opposite signs on predicted and actual performance). However, as with abnormal returns, it is important to examine the sensitivity of our quantitative and qualitative conclusions to the particular forecasting model used.

Consider the prediction of on-time performance. The focus of the discussion in the paper is an autoregressive model, specifically an AR-4 model,

$$P(>15)_{it} = \sum_{k=2}^{5} \phi_{ik} P(>15)_{i(t-k)} + \alpha_i \mathbf{x}_{it} + \xi_{it}^{P(>15)}.$$

While it varied by the operational metric being forecast, 4 lags most frequently maximized the adjusted R-squared. We also estimated AR-5 and AR-6 models with 5 and 6 lags, respectively. To check for the possibility that investors update their forecasting model over time, we also estimate a model allowing for parameters to vary with time,

$$P(>15)_{it} = \sum_{k=2}^{5} \left(\phi_{1ik} + \phi_{2ik} t\right) P(>15)_{i(t-k)} + \left(\alpha_{1i} + \alpha_{2i} t\right) \mathbf{x}_{it} + \xi_{it}^{P(>15)}.$$

Finally, we estimated an AR-1 model using one-year lags,

$$P(>15)_{it} = \phi_i P(>15)_{i(t-12)} + \alpha_i \mathbf{x}_{it} + \xi_{it}^{P(>15)},$$

which captures any year-on-year dependence that carries over from the same month in the previous year. The models above were estimated for each of the operational measures. We also tested for serial correlation, and found none. The models are calibrated using the entire sample, which assumes that the parameters are stable over the period we study. We formally tested this assumption using the structural-break test of Andrews (1993), and are only able to reject the null of time-invariant parameters for one carrier, Delta. The break date suggested by the test is shortly after Delta's emergence from bankruptcy in 2007. Allowing for this break, or dropping Delta from the analysis completely, has almost no impact on our results.

Relative to the baseline AR-4 model, we find that the adjusted R-squared is almost always lower for the alternative models. For between 2 and 3% of the regressions, across both carriers and operational measures, the AR-5 and AR-6 models had a better adjusted fit. However, the range of adjusted R-squared was very tight across the models, varying much more by the operational metric being forecasted. The operational metrics forecasted most accurately were the delay measures and revenue passenger miles. For example, on-time performance and long delays had (across all carriers) adjusted R-squared measures ranging from 0.63-0.94 and 0.61-0.87, respectively. Adjusted R-squared measures for revenue passenger miles were between 0.81 and 0.96. Missed bags and denied boardings were forecasted with the least accuracy, with adjusted R-squared measures between 0.41-0.78 and 0.47-0.83, respectively. The inability to forecast these variables as accurately may, in part, explain the lack of market responsiveness, along with the small number of passengers impacted by these operational metrics.

Despite a lack of improvement in fit which accounts for the complexity of the respective models, the alternative models may yield forecasts that differ in important ways from those of the AR-4 model. As in the case of abnormal returns, the degree of linear dependence between the predictions from each of the models will determine how much the results of our second-stage analysis will vary with the forecasting model. While we performed the analysis for each of the alternative models, the hundreds of resulting regressions are too many to concisely summarize here. Instead, as with abnormal returns, the table below reports the correlation in the predictions, across all operational metrics, from the five alternative models.

Correlation among Alternative Forecasts

	$AR-4$	$AR-5$	$AR-6$	$AR-t$	$AR-yr$
$AR-4$	1				
$AR-5$	0.98	1			
$AR-6$	0.97	0.98	1		
$AR-t$	0.92	0.91	0.91	1	
$AR-yr$	0.82	0.83	0.83	0.81	1

Each of the models yields very similar predictions. The model with the lowest correlation with all other models is the model using year lags (i.e., AR-yr). This model also provides the least accurate forecasts of any of the models. Besides the auto-regressive models above, we also explored moving-average (MA) forecasting models. The high degree of linear dependence in the forecasts between the MA and AR models makes it difficult to distinguish one model as superior, and yields very similar second-stage results.

References

[1] Andrews, D. 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*. **61** 821-856.

[2] Fama, E. and K. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*. **33** 3-56.