

ANALYST OPTIMISM IN THE AUTOMOTIVE INDUSTRY: A POST-BAILOUT BOOST AND METHODOLOGICAL INSIGHTS

Barry Hettler.^{1*} Nonna Sorokina.^{2*} Yertai Tanai.^{3*} David Booth.^{4*}

¹ *School of Business Administration and Economics, The College at Brockport – SUNY*

² *School of Business, Wake Forest University*

³ *Department of Management and Information Systems, College of Business Administration, Kent State University*

⁴ *Department of Management and Information Systems, College of Business Administration, Kent State University*

Abstract: This paper empirically investigates the impact of the government bailout on analysts' forecast optimism regarding firms in the automotive industry. We compare the results from M- and MM-robust methodologies to the results from OLS regression in an event study context and find that inferences change. When M- and MM-robust estimation methods are used to estimate the same model, the results for key control variables fall directly in line with those of similar previous studies. Furthermore, an analysis of residuals indicates that the application of M- and MM-estimation methods pulls the main prediction equation towards the main sample data, suggesting a more rigorous fit. Based on robust methods, we observe changes in analyst optimism during the announcement period of the bailout, as evidenced by the significantly positive variable of interest. We support our empirical results with simulations and confirm significant improvements in estimation accuracy when robust regression methods are applied to the samples contaminated by outliers.

Key Words: Financial analysts, M-estimation, optimism, robust regression.

* Corresponding author.

1. Introduction

Government intervention in the marketplace often provides a useful opportunity for researchers to examine the impact of an event on the market and market participants. In the finance and accounting literature, this intervention typically comes in the form of new laws (e.g., Sorokina, Booth, & Thornton [2013]) and regulations (e.g., Barniv, Hope, Myring, & Thomas [2009]). However, the government may also affect markets in other ways. One such event was the bailout of General Motors (GM) and Chrysler in 2009. Due to the size and importance of the automotive industry to the United States' economy, examining this bailout may shed light on the impact of government action on financial analysts. Specifically, we examine the improvement in estimation of the impact of the government bailout on financial analysts' forecast optimism when robust regression methods are applied to the outlier-contaminated data in an event study context (Sorokina, Booth, & Thornton [2013]).

Using the dataset in Hettler (2012), we apply robust methodologies that efficiently treat outliers and high-leverage data points. We employ the M-estimation method, which is robust to outliers, and the MM-estimation method, which is robust to both outliers and high-leverage data points. We compare the results obtained by these robust methodologies with the results obtained by OLS regression and find that inferences change.

When OLS regression is used to estimate the model, the results are opposite to what is expected. Analysts' general experience is found to be significant in explaining forecast optimism, while forecast horizon is not found to be significant. The OLS results are contrary, for example, to those of Groyberg, Healy, Chapman, & Shanthikumar (2007), who find that forecast horizon is strongly and positively correlated with relative optimism and that analyst experience is generally insignificant. In an attempt to reconcile these findings, we first transform our dependent variable using square root and natural log transformations in order to alleviate heteroskedasticity issues. We find that these transformations generally improve inferences, and make our model, as a whole, marginally significant. When M- and MM-robust estimation methods are used to estimate the same model, the results fall directly in line with those of previous studies. Even more importantly, the variable of interest (a post-bailout dummy) is shown to be positive and significant using both robust estimation methods. Model fit, as measured by R^2 , increases compared to OLS when M-estimation is applied, and further improves in MM-estimation.

Additionally, we support our findings with a quasi-simulation experiment. We generate the dependent variable (relative forecast optimism) using high, medium, and low levels of the empirically estimated post-bailout coefficient. We use normally distributed errors in some of the samples, while in others we introduce outliers. We demonstrate that in the datasets with outliers, robust estimators produce much more accurate coefficients than OLS does, and the model's power of correct recognition of the bailout effect improves drastically.

This paper contributes to the analysis of government market intervention by demonstrating that outliers and high-leverage data points can severely disrupt small-sample optimism models. We provide evidence that robust M- and MM-estimation techniques generate theoretically sensible results when OLS does not.

The remainder of the paper is organized as follows: section II contains background information on the auto industry bailout and a review of relevant literature; section III presents the data and methods, and develops the hypotheses; section IV presents our results; section V presents the simulation procedure and results; and section VI offers concluding remarks.

2. Automotive Industry Bailout Background and Literature Review

2.1 Background

Although the economic crisis that began in earnest in 2008 originated in the financial sector, it quickly spread across the economy at large. As consumer confidence waned, the automotive industry was among the first to feel the effect. Burdened with unsustainable debt loads, high union wage rates, and precipitously falling sales, the three major automobile manufacturers in the United States were brought quickly to their knees. Ford, through a risky and well-executed recovery plan, avoided a bankruptcy filing, but GM and Chrysler were not as lucky. On April 30, 2009, Chrysler declared bankruptcy; GM followed on June 1.¹ Supported by \$62 billion of Treasury loans, the two were ushered through an extremely quick bankruptcy process. Though the process was painful, the two companies were on a path to profitability by the end of the year (The Economist, 2009). In hindsight, the process appears to have worked well; however, at the time, a successful outcome was anything but guaranteed.

The automotive industry is massive. In 2008, the Big 3 alone employed over 239,000 people. Moreover, their suppliers counted a payroll over twice as large (McAlinden, Dziczek, & Menk [2008]). With such an industry structure, the ripple effects of what happens at the top may be momentous downstream. Therefore, this paper examines the impact of the economic difficulties of the Big 3 on analysts following firms in the automotive industry.

2.2 Literature Review

Financial analysts have long been studied for the important role they play in capital allocation decisions. As the success of their career is closely linked to how well they are able to obtain, process, and interpret information about firms' prospects, they have a large incentive to provide accurate forecasts about various measures. Investors, who typically have much less expertise in this area, may allocate their investment capital based on analysts' forecasts.² Thus, it

¹ Source: <http://www.msnbc.msn.com/id/31030038/ns/business-autos/t/humbled-gm-files-bankruptcy-protection/>

² It is well established that analysts outperform time-series predictions of earnings, and that they do so based on making full use of available information and using information that becomes known after the end of the prior quarter (Brown, Hagerman, Griffin, & Zmijewski, 1987).

is useful to determine in which circumstances analysts may perform better or worse. Armed with this information, investors can effectively use forecasts generated through analyst activity.

A major thrust of Hettler (2012) is an examination of analyst forecast accuracy after the bailout. His study hypothesizes and finds an increase in forecast accuracy following the bailout, as measured by a variation of the Jacob, Lys, & Neale (1999) accuracy model.

In addition, Hettler (2012) attempts to investigate analysts' forecast optimism. The literature finds conflicting results regarding analyst optimism and economically distressed firms. Some studies find that analysts tend to be overly optimistic prior to bankruptcy filings. For example, Moses (1990) shows that analysts' bias is significantly more positive toward failing firms than healthy firms, even up to the point of bankruptcy (perhaps because firms have an incentive to not fully disclose all relevant information).

However, other studies find decreased optimism prior to bankruptcy filings and greater optimism after financially distressing events. Clarke, Ferris, Jayaraman, & Lee (2006) find that analyst recommendations decline monotonically over the eight quarters preceding bankruptcy. Similarly, Espahbodi, Dugar, & Tehranian (2001) examine both firms that declare bankruptcy and those that recover ("turnaround firms"). For firms that declare bankruptcy, they find that optimism disappears by the year of bankruptcy because analysts no longer feel a need to appease current management. The authors also find that analysts return to being optimistic about turnaround firms after two to three years.

In general, previous research on optimism and bankruptcies has studied periods of relative economic calm. Indeed, modern literature has not investigated a time as tumultuous as that experienced between 2007 and 2009. In this study, we expect to observe an increase in analysts' optimism following the automotive industry bailout. An increase in optimism would be consistent with prior findings that suggest a low level of optimism during crisis periods for firms and increasing optimism thereafter. Greater optimism may also result from a sense of an improvement in the overall economic environment in general.

Prior finance and accounting work has explicitly examined the existence and impact of outliers. A handful of such studies exist (Booth, 1982; Hauser & Booth, 2011a; Hauser & Booth, 2011b; Kimmel, Booth, & Booth, 2010; Bhattacharyya, Datta, & Booth, 2011; Sorokina, Booth, & Thornton 2013). These studies find that robust (outlier-resistant) regression methods significantly improve results and highlight relationships in samples with outliers that otherwise are not detected with the use of traditional (i.e., non-robust) regression methods. We approach our problem in a manner similar to Sorokina, Booth, & Thornton (2013), and show that robust regression methods are useful in the analysts' optimism analysis of contaminated samples.

3. Data, Method, and Hypotheses

The optimism model introduced by Hettler (2012) and employed in this study is developed from a measure of relative forecast bias utilized to measure analyst optimism. We measure the relative optimism of financial analysts' earnings forecasts using an approach similar to that of Groyberg, Healy, Chapman, & Shanthikumar (2007) and Cowen, Groyberg, & Healy (2006). This definition of relative optimism controls for company- or time-specific factors that can affect forecast performance. The relative forecast optimism/pessimism (RFOPT) of each analyst's earnings forecast is investigated by estimating the following model:

$$RFOPT_{k,j,t} = \beta_0 + \beta_1 PostBailout + \beta_2 Horizon_{k,j,t} + \beta_3 AExp_{k,j,t} + \varepsilon_{k,j,t} \quad (1)$$

for analyst k , company j , and year t ,

where:

$$RFOPT_{k,j,t} = \frac{FORECAST_{k,j,t} - \overline{FORECAST}_{k,j,t}}{STDEV(FORECAST_{j,t})} \quad (1a)$$

<i>FORECAST</i>	analyst k 's forecast of company j 's earnings for year t ;
$\overline{FORECAST}$	average forecast for all analysts making forecasts for company j 's earnings in year t ;
<i>STDEV(FORECAST)</i>	the standard deviation of forecasts across all analysts forecasting earnings for company j in year t ;
<i>PostBailout</i>	dummy variable that takes the value 1 when the forecast period ends on or after 6/30/09;
<i>Horizon</i>	the number of calendar days between the forecast issue date and the earnings announcement date;
<i>AExp</i>	the number of years analyst k has worked as an analyst.

The sample of publicly-traded automotive industry firms in the United States is identified using the Compustat database. First, we select all firms with a North American Industry Classification System (NAICS) code of 336111 through 336999. Next, we pull IBES annual forecasts for the entire IBES universe for forecast period end dates between January 1999 and December 2011. We then eliminate 252,409 observations with incomplete forecast information, missing actual data, or unreasonable forecasts (defined as forecasts with a negative horizon or a horizon greater than 365 days). Further, we keep only the most recent analyst-firm-period forecast, reducing the sample to 431,060 observations. Because prior literature (Altman, 1968; Zmijewski, 1984; Espahbodi, Dugar & Tehranian, 2001) finds that symptoms of financial distress typically become evident four years before filing, with the most noticeable change in financial condition often between the third and second years, the sample period begins five years before the 2009 bankruptcy filings.

After filtering for these years and the automotive industry NAICS codes referenced above, we are left with a total of 1,916 observations (see Table 1). From these observations and for the purpose of estimating the forecast optimism model, we exclude 82 observations for which there is only one analyst observation for a given period (thus preventing the calculation of a standard deviation). The final sample size is 1,834 for regression purposes. Descriptive statistics are shown in Table 2.

[Insert Table 1 about here]

[Insert Table 2 about here]

Hettler (2012) estimates the model using ordinary least squares regression. Column A in Table 3 displays the results of that model. Contrary to initial expectations, the coefficient on the post-bailout dummy is not significantly different from zero ($p = 0.875$). Thus, it initially appears as though the bailout had no effect on analyst optimism. However, it is worth noting that *Horiz* is also insignificant. This is in stark contrast to Groyberg, Healy, Chapman, & Shanthikumar (2007) and Cowen, Groyberg, & Healy (2006), who find that forecast horizon is extremely significant. Finally, the model as a whole is not significant ($F(3,1830) = 1.45$, $p = 0.227$).

[Insert Table 3 about here]

One reason why our results do not reflect those in prior literature may be the sample period studied (2004–2011). To test this, we run the model on the entire IBES sample for the same period as the auto industry sample; this sample results in 257,007 observations (Column B, Table 3). Consistent with prior work, these results show a positive significant ($p < 0.001$) relationship between forecast horizon and relative forecast optimism. The marginally significant coefficient on *AExp* is also consistent with the significance seen in Groyberg, Healy, Chapman, & Shanthikumar (2007). The p -value for the omnibus F is < 0.001 . Therefore, the time period examined is, not driving the unexpected results.

In an attempt to determine the root of the disparity between the observed model results, we analyse the pattern of the absolute value of residuals against the predicted values of our dependent variable (relative forecast optimism, RFOPT). As can be seen in Figure 1, the plot of the absolute value of the residuals from the OLS regression model against the predicted values shows that the variance is non-constant, and that an increasing-function transformation may be appropriate to treat the observed heteroskedasticity.

[Insert Figure 1 about here]

However, the values of relative forecast optimism can be both negative and 0. As such, before a square root or natural log transformation can be applied to the dependent variable, it must first be made positive. Thus, we add a constant (ten) to all relative forecast optimism numbers. The plot resulting from the natural log-transformed dependent variable is shown in Figure 2. With the exception of apparent outliers, the transformation seems to have improved the overall constancy of variance. As shown in Column C of Table 3, the natural logarithm transformation has a positive effect on the overall model. Compared to the results of the untransformed RFOPT variable, the overall model is nearly significant ($F(3,1830) = 2.00$, $p = 0.112$). The R^2 also improves when the transformed dependent variable is used, increasing from 0.24% to 0.33%. However, the significance of the variables follows that of the first OLS automotive industry run, and so more work must be done to investigate the unexpected results. Two remaining explanations for these results are that the automotive industry differs from other industries or that the relative forecast optimism model used in this paper is not appropriate for relatively small sample sizes. Moreover, a third explanation is that outliers in the automotive industry dataset are causing the unexpected results.

[Insert Figure 2 about here]

Outliers pose a problem in OLS regressions due to the assumption of normally distributed variables. Even a single outlier may have a large effect on coefficient estimates (Yohai, 1987). As illustrated by Yohai (1987), outliers pull the entire regression line towards their values. Thus, outliers may cause regression estimates to be less effective at explaining a statistically significant relationship, or may camouflage a relationship entirely. Although there are a number of suboptimal ways to deal with outliers in a dataset (e.g., deleting, trimming, Winsorizing) – see Sorokina, Booth, & Thornton (2013) for details – the approach taken in this paper is to use robust regression methods that allow for appropriate treatment of the outliers. The first robust regression method utilizes the M-estimator (Huber, 1973). Instead of merely minimizing the squares of the residuals, as OLS does, M-estimation re-weights observations according to a predefined influence function in an iterative fashion until a solution converges. The second robust method is MM-estimation, developed by Yohai (1987). This method builds upon earlier work done by Rousseeuw (1984) and Rousseeuw & Yohai (1984). MM-estimation is robust to both outliers and leverage points, and benefits from a relatively high breakdown point (i.e., the maximum fraction of outliers that a given sample may contain without spoiling the estimate completely); it is also highly efficient when error terms are normally distributed.³

³ We acknowledge there are other robust regression methods to deal with outliers, for example Theil-Sen (Theil, 1950; Sen, 1968) and Koenker & Bassett (Koenker & Bassett, 1978; Koenker & Hallock, 2001)

Thus, if outliers are driving the unusual results observed when OLS regression is applied to the 2004–2011 automotive industry sample, the robust estimation methods may produce more accurate coefficient estimates. Moreover, given the inherent differences between the M- and MM- robust estimation methods, the coefficients may not be the same between the two. Our three hypotheses, with H2 and H3 stated in null form, are as follows:

H1: At least one outlier exists in the sample.

H2: There is no difference in the optimism model obtained using OLS and M-estimation.

$$\begin{aligned} \beta_{1\text{PostBailoutOLS}} = \beta_{1\text{PostbailoutM}} & \quad p(\hat{\beta}_{1\text{PostBailoutOLS}} = \hat{\beta}_{1\text{PostbailoutM}}) = \\ \beta_{2\text{HorizOLS}} = \beta_{2\text{HorizM}} & \quad p(\hat{\beta}_{2\text{HorizOLS}}) = p(\hat{\beta}_{2\text{HorizM}}) \\ \beta_{3\text{AExpOLS}} = \beta_{3\text{AExpM}} & \quad p(\hat{\beta}_{3\text{AExpOLS}}) = p(\hat{\beta}_{3\text{AExpM}}) \end{aligned}$$

H3: There is no difference in the optimism model obtained using M-estimation and MM-estimation.

$$\begin{aligned} \beta_{1\text{PostBailoutM}} = \beta_{1\text{PostbailoutMM}} & \quad p(\hat{\beta}_{1\text{PostBailoutM}} = \hat{\beta}_{1\text{PostbailoutMM}}) = \\ \beta_{2\text{HorizM}} = \beta_{2\text{HorizMM}} & \quad p(\hat{\beta}_{2\text{HorizM}}) = p(\hat{\beta}_{2\text{HorizMM}}) \\ \beta_{3\text{AExpM}} = \beta_{3\text{AExpMM}} & \quad p(\hat{\beta}_{3\text{AExpM}}) = p(\hat{\beta}_{3\text{AExpMM}}) \end{aligned}$$

4. Results

As discussed in the prior section, using OLS to estimate the optimism model developed by Cowen, Groyberg, & Healy (2006) and Groyberg, Healy, Chapman, & Shanthikumar (2007) for the automobile industry around the 2009 bailout (2004–2011) produces counterintuitive results: even after attempting to correct for the apparent heteroscedasticity, the overall model is

estimators; however, we believe M- and MM-estimators are the most appropriate for our application. First, the maximum breakdown point for the Theil-Sen estimator is known to be 29.3%, which occurs for the simple linear regression model. As more predictors are added to the model, the breakdown point decreases. The maximum breakdown possible is 50% and can be achieved by the M- and MM-estimators; they are thus more satisfactory in this regard. For several predictors, the maximum breakdown point is very low relative to M- and MM-estimators. Although the Siegel (1982) modification to the Theil-Sen estimator has a breakdown of 50%, which is comparable to the M- and MM-estimators, this approach does not allow for differentiation between outliers and leverage points. Concerning the Koenker and Bassett quantile regression estimator, we first note that the algorithm is based on linear programming and, hence, similar to the Theil-Sen estimator, cannot distinguish between outliers and leverage points. It is often desirable to make this distinction (Sorokina, Booth, & Thornton, 2013). Examples of problems where the Koenker-Bassett estimator is useful are given in Cade & Noon (2003).

still not significant ($F = 2.00$, $p = 0.1120$, untabulated); β_2 (Horiz) is not significant ($p = 0.368$), though it has been shown to be highly significant in prior work; and β_3 (AExp) ($p = 0.027$) is significant, whereas previously it has generally been found to be insignificant or marginally significant (see Table 3).

We next apply the M- and MM-estimator robust estimation methods. We run data in SAS using ProcRobustReg with both a square root and natural log transformation to correct for heteroscedasticity. ProcRobustReg identifies 39 outliers (2.13% of the sample) in the robust regression model with the square root transformation applied to relative forecast optimism, and 44 outliers (2.40% of the sample) in the robust regression model with a natural log transformation applied. Of the 44 outliers resulting from the natural log transformation, 25 come from six companies with more than two outlier observations. There are a total of 67 companies represented in the auto industry sample between 2004 and 2011; thus, firm-specific characteristics may be factors in the observed outliers. These results support H1. We, therefore, have a reason to believe that the M- and MM-estimation methods may produce superior results relative to OLS. Even though the relative number of outliers appears to be small, just one outlier can throw off data analysis (Yohai, 1987).

Table 4 presents the results of the optimism model when M-estimation is applied to both the square root and natural log transformations of the dependent variable. The findings using robust regression are striking: the results mirror those of prior research when the optimism model is applied to the entire automobile industry 2004–2011 sample. The effect of the transformations is very similar for both the square root and natural log functions. However, as the natural log-transformed dependent variable results in a higher R^2 (0.70% vs. 0.54% for the square root transformation), we focus our analysis on the results from this regression. The first difference noted between M-estimation and OLS regression for the models is the R^2 . Using M-estimation more than doubled the observed R^2 , which increased from 0.33% to 0.70%. Clearly, robust regression has helped the model fit our data better.

[Insert Table 4 about here]

Next, the coefficient on the primary variable of interest from Hettler (2012), PostBailout, is significant at the 5% level; the p-value decreased from 0.643 to 0.038 between OLS and M-estimator robust regression. The actual coefficient estimate changes from 0.002 to 0.0089, an over-fourfold increase in economic significance. In other words, financial analysts became more optimistic about automotive industry firms following the mid-2009 bailouts. The coefficient on Horiz is significant at the 1% level when M-estimation is applied, and increased from -0.0000 to 0.0001 between OLS and M-estimation. The sign-flip is also in the appropriate direction based on the IBES 2004–2011 data and prior work (e.g., Groyberg, Healy, Chapman, & Shanthikumar [2007]). The coefficient on AExp is not significant when M-estimation regression is applied (p-value increases from 0.027 to 0.575). However, this follows previous literature that typically finds only a weak or no association between analyst experience and forecast optimism. Based on these results, we can reject the null hypothesis H2 in favour of the alternative that coefficient estimates and significance levels change between OLS and M-estimation robust regression.

Next, We run MM-estimator robust regression. As seen in Table 5, MM-estimation improves the fit of the optimism model with the natural log of forecast optimism as the dependent

variable (results are provided for the square root transformation as well). Because the natural log transformation of forecast optimism is the model we concentrate on, this discussion is primarily limited to that model.⁴

[Insert Table 5 about here]

The R^2 improves from 0.70% to 0.95% when MM-estimation is applied. Although not as dramatic as the jump from OLS to M-estimation, the increase nonetheless suggests that MM-estimation regression is better able to explain the relationships present in our data. It is also worthwhile to note that the model's R^2 increases nearly fourfold from the model's original R^2 with a non-transformed dependent variable and OLS regression. MM-estimation results in coefficients and significance levels that differ from those obtained using M-estimation. The coefficient on PostBailout increases to 0.0125, and the p-value falls to 0.004 from 0.038. Thus, the findings are even stronger (now at the 1% significance level) that forecast optimism increased following the automotive industry bailout. The coefficients and p-values for forecast horizon did not change between M- and MM-estimation (0.0001 and $<.0001$, respectively). The coefficient for AExp changes from -0.0002 to 0.0001, and the p-value for the variable remains similar. Overall, we are largely able to reject the null hypothesis H_3 that coefficients and p-values do not differ between M- and MM-robust estimation techniques. This event is likely due to the presence of leverage points in the dataset.

It is interesting to note that the R^2 from our final model, MM-estimation with the natural log of forecast optimism as the dependent variable, is higher when using the untransformed number of years as an independent variable than it is when using the natural log of the number of years of analyst experience (0.95% vs. 0.92%, with results for the regression with natural log of years of experience untabulated). For this reason, our final optimism model differs from that used in prior literature. Furthermore, in another illustration of the effect of transformations, we run the forecast optimism model with an untransformed dependent variable using MM-estimation (results untabulated). We find that the model's R^2 is only 0.55%, just over half of that obtained when using the natural log-transformed dependent variable.

As a robustness check, all OLS, robust M, and robust MM regressions are run in SAS and R. The results are identical between the two programs for OLS. Moreover, the results do not qualitatively differ between SAS and R for the robust estimation methods. At most, coefficients are off by a few thousandths, and all significance levels are identical. The slight differences observed are likely a result of the rounding used in R function `rlm` or settings not being identical between the two software programs (for example, the Ψ -function used for weighting observations). The methods for `rlm` are described in Venables & Ripley (2002) section 6.5.

In additional analyses, we test whether the financial crisis period itself (defined as 1/1/2008 through 6/30/2009, modelled with a Crisis dummy variable that takes a value of 1 for fiscal year ends falling in this range and 0 otherwise) is associated with lower analyst optimism (untabulated). Using MM-estimation to regress `ln_RFOPT` on Crisis and the same controls as used in Model 1,

⁴ However, it is worth noting that the R^2 increases for the square root transformation model as well.

Furthermore, the post-bailout dummy variable increases in significance when MM-estimation is applied in place of M-estimation for the square root-transformed dependent variable.

we find a negative and significant coefficient on Crisis (coefficient estimate = -0.019 , p-value = 0.0018). Therefore, we find that financial analyst optimism is significantly lower for the automotive industry during the immediate pre-bailout period. This result provides additional support that the bailout is associated with increased analyst optimism.

5. Simulation

5.1 Simulation Design

In this section, we perform a Monte-Carlo simulation in order to validate and extend our empirical results. We generate only dependent variables, maintaining the independent variables from the empirical data set. We further use the results of MM-estimation of Eq.(1) to simulate dependent variable values. Based on MM-estimation coefficients Eq.(1) can be re-written as:

$$\ln_RFOP_{k,j,t} = 2.2863 + \beta_1 Postbailout + 0.0001 Horiz + 0.0001 Genex + \varepsilon \quad (2)$$

First, we generate 10,000 normal random variates using the mean (-0.008) and the standard deviation (0.0964) of the MM-estimation residuals. We identify outliers using Cook's distance. Cook's distance is the most commonly used method to estimate the influence of a data point in a least-squares regression. It identifies both outliers and leverage points as well as the effect of deleting a given observation. The Cook's distance cut-off for an outlier detection is defined as $\frac{4}{n-k-1}$, where n is the number of observations, and k is the number of independent variables (Sorokina, Booth, & Thornton [2013]). We contaminate the simulated sample errors with outliers using the number of outliers from our empirical sample. We position the outliers within simulated samples as described in Table 6.

Second, we simulate the dependent variable to capture the post-bailout change in the levels of analyst optimism, as specified by Eq.(2). We develop five simulation specifications that differ by the level of post-bailout effect and outliers' presence/positions. In Specification 1, we set the mid-level post-bailout effect to be equal to the empirical mean observed using MM-estimation, which is $\beta_1 = 0.0125$. In Specifications 2–3, we set the low-level post-bailout effect at the lower bound of the 95% confidence interval ($\beta_1 = 0.0039$) and the high-level at the upper bound ($\beta_1 = 0.0211$). For specifications 1–3, we replace observations from the normally distributed simulated sample of errors with outliers from a double exponential (Laplace) distribution. We position the outliers randomly and obtain a new set of random positions for each sample.

In Specifications 4–5, we set post-bailout effects to be the same as in Specification 1. However, in Specification 4, when we replace normally distributed simulated sample errors with outliers, we position them identically to those in the original empirical sample. We do not contaminate simulated sample errors in Specification 5. Table 6 summarizes each of the discussed specifications.

[Insert Table 6 about here]

5.2 Simulation Results

The results of the simulation-based tests are shown in Table 7. For each specification, we report the percentage of correct recognitions of the simulated effect at the 90%, 95%, and 99% confidence levels. We limit our discussion to the results found statistically significant at the 95% level. The results significant at the 90% and 99% levels hold similar patterns.

Additionally, we compare the estimated post-bailout coefficients resulting from simulation to the true specified values, and identify the most appropriate estimation method by computing the absolute percentage error (APE) defined by Eq.(3). Table 8 shows the percentage of correct recognitions as well as APE values at the 95% confidence level:

$$APE = \left| \frac{TrueValue - PostSimulationValue}{TrueValue} \right| \times 100\% \quad (3)$$

In Specifications 1–4, we observe that percentages of correct recognitions (or statistically significant post-bailout slopes/effects) of the robust estimators are much higher than with OLS (see Table 8). These results are expected due to the presence of outliers in each of the specifications. Because the M-estimator is robust to outliers, the MM-estimator is robust to both outliers and leverage points, but OLS is robust to neither, the two robust estimators perform better than OLS in the above four specifications. For instance, M- and MM-estimation methods capture and correctly recognize significance about 66% of the time on average from the simulated data; however, OLS captures the correct, significant value only 10% of the time. Furthermore, APE values show that OLS does not perform well in estimating the true value of the post-bailout slope in the presence of outliers. Particularly, for Specification 1, the APE value for OLS is about 254%, whereas it is about 22–23% for both M and MM-estimations.

[Insert Table 7 about here]

As mentioned earlier, Specifications 1–3 test for the recognition of a change in the slope of the post-bailout variable (see Table 6). In addition to robust estimators being better than OLS at detecting the change at all three levels, we note that as the simulated change in slopes increases in magnitude (from Specifications 2 through 1 to 3), the number of successful recognitions increases, and APE values tend to decline for all three estimation methods (see Table 8). For example, overall, both the M- and MM-estimations results show a percentage increase in correct recognitions, starting from 11% in Specification 2 (lower bound) to 66% in Specification 1 (mid-level) to 97% in Specification 3 (upper bound). Moreover, a decline in APE values is apparent for these estimators as the change in slope increases.

[Insert Table 8 about here]

In Specification 5, OLS performs better than do the two robust estimators. This result is also expected. Because there are no outliers in the simulated data, the simulated sample shows a simple linear relationship between analysts' optimism and post-bailout (see Eq. (2)). Hence, in this case, OLS is the best linear unbiased estimator (Cook and Weisberg, 1982).

6. Conclusion

This paper is an extension of Hettler (2012), where the forecast optimism model developed by Cowen, Groysberg, & Healy (2006) and Groysberg, Healy, Chapman, & Shanthikumar (2007) produces results inconsistent with prior literature. Noting apparent heteroskedasticity, we first transform the dependent variable, relative forecast optimism, with increasing functions (square root and natural log). We then apply M- and MM-estimation robust regression to the models. M-estimation greatly improves the OLS results: the model, as a whole, becomes significant; the control variables, horizon and analyst experience, become consistent with the values found in prior literature; and the main independent variable of interest (a dummy variable that takes the value of 1 for observations after June 30, 2009, representing the date of the bailouts of GM and Chrysler) becomes positive and significant, as expected. This relationship is not observed in OLS regression estimates. When MM-estimation is used, the overall model fit, as measured by R², increases further, and the post-bailout dummy becomes significant at the 1% level. Thus, by transforming the dependent variable and utilizing robust regression in the apparently outlier-contaminated sample, we can conclude that financial analysts became more optimistic following the bailouts. Furthermore, we support our empirical results by simulations and find significant improvements in describing the impact of the bailout on levels of analyst optimism.

A reviewer pointed out the possibility of using a structural break model as an alternative to an event study methodology (Liao, 2010; Liu, Li, & Wang, 2011). We agree there are possible advantages to this approach and suggest this as an avenue for future research. This paper contributes to the literature by demonstrating the efficacy of robust regression in dealing with outliers in analyst optimism models and the importance of proper outlier treatments in general. It serves as a reminder to check for outliers and leverage points, especially when unexpected or seemingly nonsensical results are obtained.

We would like to thank the editor and editorial referees for their valuable suggestions.

References:

- [1] Altman, E. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance*, **23**, 589-609.
- [2] Barniv, R., Hope, O. K., Myring, M. J., & Thomas, W. B. (2009). Do analysts practice what they preach and should investors listen? Effects of recent regulations. *The Accounting Review*, **84**, 1015-1039.
- [3] Bhattacharyya, K., Datta, P., & Booth, D. (2011). To do or not to do business with a country: A robust classification approach. *Journal of Data Science*, **9**, 607-623.
- [4] Booth, D. (1982). The analysis of outlying data points by robust regression: A multivariate problem bank identification model. *Decision Sciences*, **13**, 71-81.
- [5] Brown, L., Hagerman, R., Griffin, P., and Zmijewski, M. (1987). Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of Accounting and Economics*, **9**, 61-87.

- [6] Cade, B. & Noon, B. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, **1**, 412-420.
- [7] Clarke, J., Ferris, S., Jayaraman, N., and Lee, J. (2006). Are Analyst Recommendations Biased? Evidence from Corporate Bankruptcies. *Journal of Financial and Quantitative Analysis*, **41**, 169-195.
- [8] Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression (Vol. 5). New York: Chapman and Hall.
- [9] Cowen, A., Groyberg, B. and Healy, P. (2006). Which types of analyst firms make more optimistic forecasts? *Journal of Accounting & Economics*, **41**, 119-146.
- [10] Espahbodi, R., Dugar, A., and Tehranian, H. (2001). Further evidence on optimism and under-reaction in analysts' forecasts. *Review of Financial Economics*, **10**, 1-21.
- [11] Groyberg, B., Healy, P., Chapman, C., and Shanthikumar, D. (2007). Do buy-side analysts out-perform the sell-side? Working paper. Retrieved from <http://ssrn.com/abstract=806264>.
- [12] Hauser, R.P., & Booth, D. (2011a). CEO bonuses as studied by robust logistic regression. *Journal of Data Science*, **9**, 293-310.
- [13] Hauser, R.P., & Booth, D. (2011b). Predicting bankruptcy with robust logistic regression. *Journal of Data Science*, **9**, 565-584.
- [14] Hettler, B. (2012). A Bailout for Auto Industry Analysts? An analysis of the bankruptcies of General Motors and Chrysler. Kent State University, Unpublished manuscript.
- [15] Huber, P.J. (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo. *Annals of Statistics*, **1**, 799-821.
- [16] Jacob, J., Lys, T., and Neale, M. (1999). "Expertise in forecasting performance of security analysts." *Journal of Accounting and Economics*, **28**, 51-82.
- [17] Kimmel, R. K., Booth, D. E., & Booth, S. E. (2010). The analysis of outlying data points by robust Locally Weighted Scatter Plot Smooth: a model for the identification of problem banks. *International Journal of Operational Research*, **7**, 1-15.
- [18] Koenker, R. & Bassett, Jr, G. (1978) Regression Quantiles. *Econometrica*, **46**, 33-50.
- [19] Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *The Journal of Economic Perspectives*, **15**, 143-156.
- [20] Liao, W. (2010). Structural Breaks in Panel Data Models: A New Approach. Retrieved from http://hkimr.org/uploads/seminars/139/sem_paper_0_346_wei-liao.pdf
- [21] Liu, D., Li, R., & Wang, Z. (2011). Testing for structural breaks in panel varying coefficient models: With an application to OECD health expenditure. *Empirical Economics*, **40**, 95-118.
- [22] McAlinden, S., Dziczek, K., Menk, D. (2008). CAR Research Memorandum: The Impact on the U.S. Economy of a Major Contraction of the Detroit Three Automakers. *The Center for Automotive Research*. Retrieved from http://www.cargroup.org/documents/Detroit_Three_Contraction_Impact.pdf.
- [23] "Rinsed and raring to go." (2009). November 5. *The Economist*.
- [24] Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871-880.
- [25] Rousseeuw, P.J. & Yohai, V.J. (1984). Robust regression by means of S-estimators. In: Franke, J., Hardle, W., & Martin, R.D. (ed) *Robust and nonlinear time series (Lecture note in statistics No. 26)*. New York, NY: Springer-Verlag.
- [26] Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, **63**, 1379-1389.

- [27] Siegal, A., (1982). Robust Regression Using Repeated Medians. *Biometrika*, **69**, 242-244.
- [28] Sorokina, N.Y., Booth, D.E., & Thornton, Jr., J.H. (2013). Robust Methods in Event Studies: empirical evidence and theoretical implications. *Journal of Data Science*, **11**, 607-621
- [29] Sorokina, N., & Thornton, J.H. (2013). Cross-Industry and Cross-Country Reactions of Equity Markets to Recent Financial Reforms. Kent State University, Working paper.
- [30] Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I, II, III, *Nederl. Akad. Wetensch Proc.*, 53, 386–392, 512–525, 1397–1412.
- [31] Venables, W.N. & Ripley, B.D. (2002). *Modern applied statistics with S*. 4th ed. New York: Springer-Verlag.
- [32] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, **15**, 642-656.
- [33] Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* (Supplement). **22**, 59-82.

Received March 15, 2013; accepted November 10, 2013.

Barry Hettler
School of Business Administration and Economics,
The College at Brockport – SUNY, 350 New Campus Drive, Brockport, NY 14420
Corresponding author: Phone: 585-395-2623; Fax: 585-395-2542;
Email: bhettler@brockport.edu

Nonna Sorokina
School of Business,
Wake Forest University, 1834 Wake Forest Rd, Winston-Salem, NC 27106
Phone: 336-758-6177;
Email: sorokiny@wfu.edu

Yertai Tanai
Department of Management and Information Systems,
College of Business Administration, Kent State University, P.O. Box 5190, Kent, OH 44242
Phone: 330-672-1164;
Email: ytanai@kent.edu

David Booth
Department of Management and Information Systems,
College of Business Administration, Kent State University, P.O. Box 5190, Kent, OH 44242
Phone: 330-805-0239;
Email: dvdbooth8@gmail.com

Table 1: Sample Selection Process (1999-2011)

Table 1	
Sample Selection Process (1999-2011)	
IBES Universe (Forecast Period End Date Jan. 1999 – Dec. 2011)	1,834,535
Less:	
Missing actual information, activation dates, or horizon < 0 or > 365	<u>252,409</u>
	1,582,126
Less:	
Earlier analyst-firm forecasts (keep most recent)	<u>1,150,966</u>
	431,060
Less:	
Non-Auto Industry and before 2004	<u>429,144</u>
Observations in Sample (2004 – 2011)	<u><u>1,916</u></u>

Subsample Detail:

Pre-Bailout (Period End Date Before June 30, 2009)	1,220
Post-Bailout (Period End Date After June 30, 2009)	696

Table 2: Descriptive Statistics: Means (Medians) by Sample

Table 2				
Descriptive Statistics: Means (Medians) by Sample				
Variable	All IBES Firms	All Auto Industry	Pre-Bailout	Post-Bailout
<i>RFOPT</i>	0.00 (0.04)	0.00 (0.00)	0.00 (-0.09)	0.01 (0.10)
<i>HORIZ</i>	109.53 (96.00)	103.13 (92.00)	110.55 (95.00)	90.36 (89.00)
<i>AExp</i>	11.68 (11.00)	11.38 (11.00)	12.39 (12.00)	9.66 (8.00)
Period	2004 - 2011	2004 - 2011	2004 - 2009	2009 - 2011
n	257,007	1,916	1,220	696

Table 3: OLS Regression Results Coefficient Estimates (p-values)

Table 3						
OLS Regression Results Coefficient Estimates (p-values)						
Column	A		B		C	
Sample	Auto	Industry	IBES	Universe	Auto	Industry
DV	<i>RFOPT</i>		<i>RFOPT</i>		<i>ln_RFOPT</i>	
Intercept	0.0682		-0.0278	***	2.3113	***
	(0.251)		(<.001)		(<.001)	
<i>Post_Bailout</i>	-0.0073				-0.002	
	(0.875)				(0.643)	
<i>Horiz</i>	0.0002		0.0002	***	-0.0000	
	(0.447)		(<.001)		(0.369)	
<i>AExp</i>	-0.0073	**	0.0006	*	-0.0008	**
	(0.048)		(0.071)		(0.027)	
R-squared	0.24%		0.03%		0.33%	
Period	2004-2011		2004-2011		2004-2011	
n	1,834		257,007		1,834	

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. The dependent variable is Relative Forecast Optimism (natural log of Relative Forecast Optimism in Column C). *Post_Bailout* is a dummy variable that takes the value 1 when the forecast period ends on or after 6/30/09. *Horiz* is the number of calendar days between the forecast issue date and the earnings announcement date. *AExp* is the number of years analyst k has worked as an analyst.

Table 4: M-estimation Robust Regression Results Coefficient Estimates (p-values)

Table 4				
M-estimation Robust Regression Results Coefficient Estimates (p-values)				
Sample	Auto Industry		Auto Industry	
DV	<i>sqrt_RFOPT</i>		<i>ln_RFOPT</i>	
Intercept	3.1492	***	2.2917	***
	(<.001)		(<.001)	
<i>Post_Bailout</i>	0.0129	*	0.0089	**
	(0.059)		(0.038)	
<i>Horiz</i>	0.0002	***	0.0001	***
	(<.001)		(<.001)	
<i>AExp</i>	-0.0004		-0.0002	
	(0.488)		(0.575)	
R-squared	0.54%		0.70%	
Period	2004-2011		2004-2011	
n	1,834		1,834	

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. The dependent variable is the square root (natural log) of Relative Forecast Optimism in the left (right) column. *Post_Bailout* is a dummy variable that takes the value 1 when the forecast period ends on or after 6/30/09. *Horiz* is the number of calendar days between the forecast issue date and the earnings announcement date. *AExp* is the number of years analyst *k* has worked as an analyst.

Table 5: MM-estimation Robust Regression Results Coefficient Estimates (p-values)

Table 5				
MM-estimation Robust Regression Results Coefficient Estimates (p-values)				
Sample	Auto Industry		Auto Industry	
DV	<i>sqrt_RFOPT</i>		<i>ln_RFOPT</i>	
Intercept	3.1401	***	2.2863	***
	(<.001)		(<.001)	
<i>Post_Bailout</i>	0.0190	***	0.0125	***
	(0.006)		(0.004)	
<i>Horiz</i>	0.0002	***	0.0001	***
	(<.001)		(<.001)	
<i>AExp</i>	-0.0001		0.0001	
	(0.916)		(0.855)	
R-squared	0.73%		0.95%	
Period	2004-2011		2004-2011	
n	1,834		1,834	

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. The dependent variable is the square root (natural log) of Relative Forecast Optimism in the left (right) column. *Post_Bailout* is a dummy variable that takes the value 1 when the forecast period ends on or after 6/30/09. *Horiz* is the number of calendar days between the forecast issue date and the earnings announcement date. *AExp* is the number of years analyst *k* has worked as an analyst.

Table 6: Simulation specifications

Table 6		
Simulation specifications		
	<i>Post-bailout effect</i>	<i>Outliers' position</i>
Specification 1	$\beta_1 = 0.0125$	Randomly using double exponential distribution
Specification 2	$\beta_1 = 0.0039$	Randomly using double exponential distribution
Specification 3	$\beta_1 = 0.0211$	Randomly using double exponential distribution
Specification 4	$\beta_1 = 0.0125$	Identical to original sample data
Specification 5	$\beta_1 = 0.0125$	No outliers

Table 7: Simulation results

Table 7											
Simulation results											
<i>Spec</i>	<i>Bailout</i>		90%			95%			99%		
			OLS	M	MM	OLS	M	MM	OLS	M	MM
1	0.0125	Co	0.0371	0.0142	0.0144	0.0436	0.015	0.0152	0.0552	0.0169	0.0172
		PC	16.95	77.81	76.4	10.22	67.57	65.46	2.81	43.2	40.9
2	0.0039	Co	0.0193	0.0102	0.0104	0.0246	0.0118	0.012	0.0425	0.0147	0.015
		PC	10.6	19.3	18.87	5.54	11.79	11.39	1.24	3.5	3.37
3	0.0211	Co	0.0426	0.0208	0.0208	0.0474	0.021	0.021	0.0568	0.0214	0.0215
		PC	28.84	99.07	98.85	19.3	98.17	97.62	7.29	93.08	91.63
4	0.0125	Co	0.0440	0.0152	0.0151	0.0469	0.0158	0.0158	0.0539	0.0175	0.0176
		PC	73.76	85.61	82.32	59.43	77.47	73.56	29.71	55.04	49.78
5	0.0125	Co	0.0137	0.0139	0.0143	0.0144	0.0145	0.0150	0.0160	0.0163	0.0170
		PC	82.82	81.21	76.97	73.83	71.9	67.14	51.48	47.97	42.19

Note: Co = Post simulation estimated coefficient, PC = Percentage of correct recognitions

Table 8: Percentage of correct recognitions and APE values at 95% confidence level

Table 8						
Percentage of correct recognitions and APE values at 95% confidence level						
Specifications	<i>Percentage of recognition</i>			<i>APE</i>		
	OLS	M	MM	OLS	M	MM
1	10.22%	67.57%	65.46%	254%	22.0%	23.6%
2	5.54%	11.79%	11.39%	530.8%	202.6%	207.7%
3	19.30%	98.17%	97.62%	129.0%	1.4%	1.4%
4	59.43%	77.47%	73.56%	281.3%	28.5%	28.5%
5	73.83%	71.90%	67.14%	17.1%	17.9%	22.0%

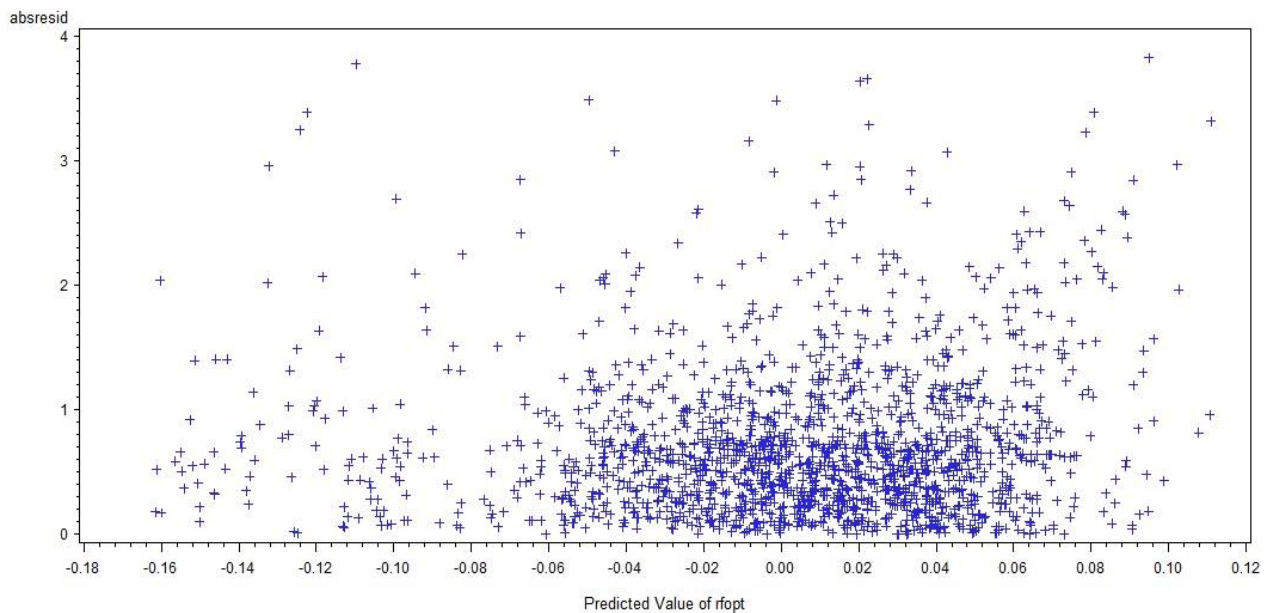


Figure 1. Plot of absolute value of OLS residuals against predicted RFOPT

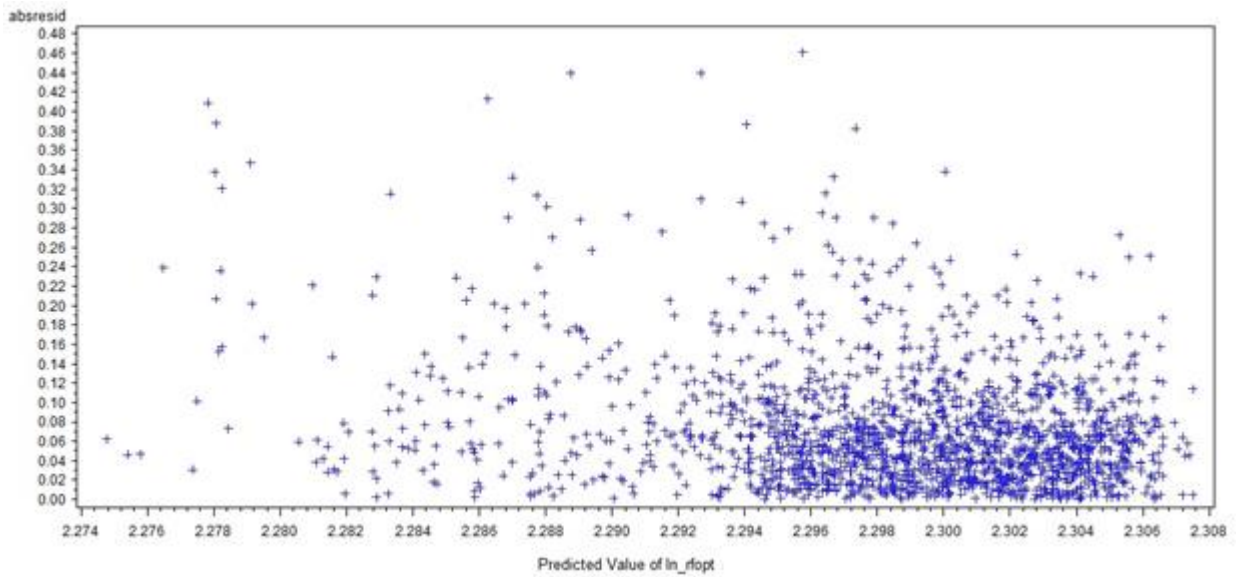


Figure 2: Plot of absolute value of OLS residuals against predictedLog(RFOPT+10)