Beating a Random Walk*

September 6, 2017

Peter Easton

Peter Kelly

Andreas Neuhierl

Abstract

Until recently, the accounting and finance literatures generally used either a random walk or analysts' estimates as forecasts of accounting earnings. Hou et al. [2012] suggested the use of a cross-sectional forecasting model and several studies immediately used this model because of the obvious advantage that forecasts can be formed for a sample that is much greater than the sample of firms covered by analysts. Unfortunately, subsequent studies have shown that the Hou et al. [2012] forecasts are not significantly better than random walk forecasts. We present a simple modification of Hou et al. [2012] – the use of quantile rather than OLS regressions in the prediction model - that produces forecasts significantly better than random walk forecasts. Quantile regressions are intuitively appealing because these regressions minimize the sum of the absolute deviations, which is consistent with the use of the absolute difference between forecast and actual as the indicator of accuracy. OLS regressions, on the other hand, minimize the sum of squared deviations, which is not consistent with this indicator.

*All authors are from the Mendoza College of Business, University of Notre Dame. Correspondence: Peter Easton, (Email) peaston@nd.edu (Tel) 574-631-6096. Peter Kelly, (Email) Peter.W.Kelly.198@nd.edu (Tel) 574-631-0933. Andreas Neuhierl, (Email) aneuhier@nd.edu (Tel) 574-631-1068.

1 Introduction

We show that the use of quantile regressions rather than ordinary least squares regressions (OLS) in the cross-sectional earnings forecasting model proposed by Hou et al. [2012] results in forecasts that are significantly more accurate. We measure accuracy, the same way as Hou et al. [2012], as the absolute difference between the forecast and actual. Use of quantile regressions rather than OLS (as in Hou et al. [2012]) in the estimation of the prediction model minimizes the sum of absolute deviations, which is consistent with this measure of accuracy. We provide evidence of other advantages of the use of quantile regressions including stability of the prediction model coefficient estimates over time and the requirement of fewer years of data for the estimation of the prediction models.

Until recently, the accounting and finance literatures generally used either a random walk or analysts' estimates as forecasts of accounting earnings. The observation that the best model of the time-series of earnings is a random walk was established by Watts and Leftwich [1977] following a considerable literature on the time-series properties of earnings.[1] In their 2012 study, Hou et al. [2012] build on models suggested by Fama and French [2000] and Fama and French [2006] to develop a cross-sectional prediction model. Several studies immediately used these models because they offer the obvious advantage that forecasts can be formed for a sample that is much greater than the sample of firms covered by analysts.[2]

Unfortunately, subsequent studies have shown that the Hou et al. [2012] forecasts are not significantly better than random walk forecasts.[3] Our simple modification of Hou et al. – the use of quantile rather than OLS regressions in the prediction model leads to forecasts of earnings that are significantly more accurate than a random walk and superior to those of

[1] See Brealey [1967], Ball and Brown [1968], Fama and Babiak [1968], Lintner and Glauber [1972], Ball and Watts [1972], Griffin [1977], and Albrecht et al. [1977].

[2] See, for example, Chang and Monahan [2012], Jones and Tuzel [2012], Lee et al. [2017], and Patatoukas [2011].

[3] See Gerakos and Gramacy [2013] and Li and Mohanram [2014]

Hou et al. in almost all of the years that we analyse (1970 to 2015).

There are a number of advantages to using quantile regressions in the prediction models. Again, and perhaps most importantly, a quantile regression methodology offers higher accuracy. Using this methodology not only improves the Hou et al. [2012] original predictions, but offers predictions that are more accurate than a random walk. Second, quantile regression predictions are not influenced by the presence of outliers. Thus, our results do not change with winsorization. This allows for a more general approach than the OLS methodology. Third, our parameters are far more stable over time and it follows that we can use shorter prediction windows, and capture more short-term dynamics. Fourth, quantile regressions are also better at predicting directional changes. In particular, quantile regressions are better at predicting movements across the profitability line. Finally, our predictions produce earnings-to-price ratios that have stronger partial correlations with ex-post returns.

The rest of the paper is organized as follows. Section 2 details the data used. In Section 3, we describe the quantile regression methodology and explain why it offers more accurate forecasts. Next, in Section 4, we show how a simple modification of the Hou et al. [2012] methodology offers forecasts that outperform a random walk. Section 5 concludes.

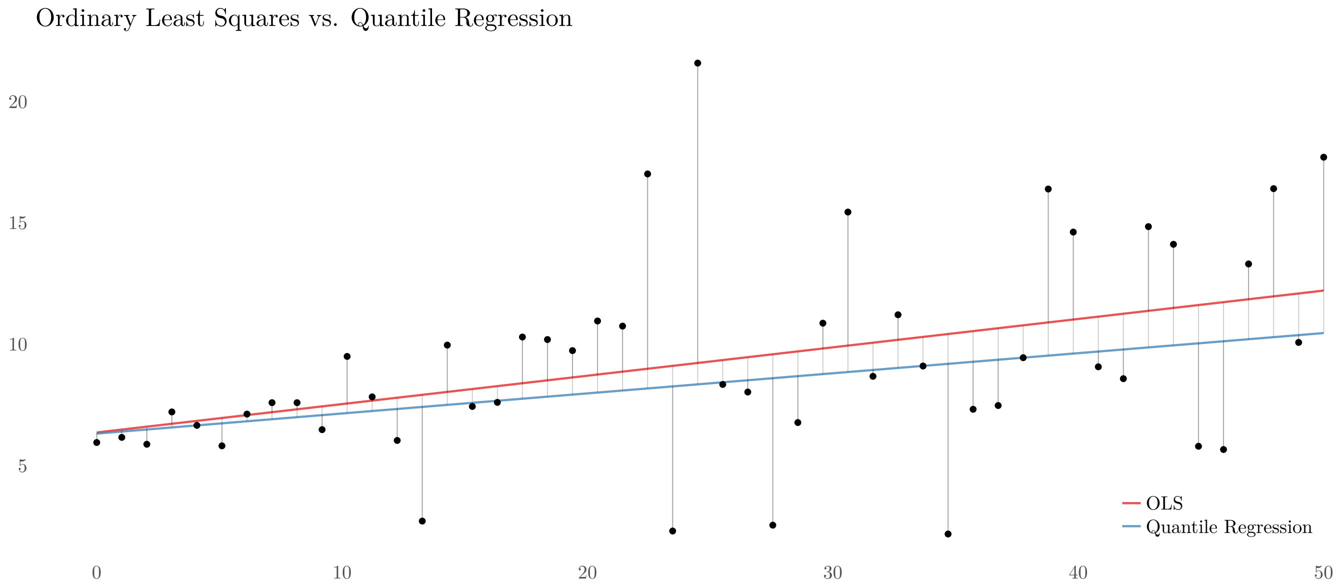
2 Data

We use CRSP/Compustat for a collection of firm-level information regarding accounting information and returns. Our sample runs from the fiscal year 1961 to the fiscal year 2016. We drop observations that are duplicates by permno and fiscal year. We exclude ADRs, closed-end funds, and REITs (i.e. we keep an observation if its share code equals 10 or 11). Earnings is equal to income before extraordinary items. We create a dummy, Negative Earnings, that equals one if income before extraordinary items is negative. Following Hou et al. [2012], we use a balance sheet method to define accruals. Accruals are equal to the change in non-cash current assets minus the change in current liabilities excluding the change in short-term debt and the change in taxes payable minus the depreciation and amortization expense. In some of our regression specifications, we scale the earnings variable by market equity, which is equal to the number of shares outstanding times the fiscal year-end closing price.

3 Quantile Regression Methodology

Our empirical work surrounds the use of a quantile regression methodology. Of course, the linear OLS model is the standard model for most empirical work in accounting and finance.[4] As such, before advancing to a more formal discussion of quantile regression, we briefly compare it to the least squares regression to build some intuition. In OLS, we try to find the line that minimizes the sum of squared deviations between an observed outcome and and a prediction – the prediction is modelled as a linear function of some set of observed predictors. Median regression (a special case of quantile regression), minimizes the sum of absolute deviations. The figure below illustrates the differences. The red line is the least squares fit and the blue line is the median regression line.

[4] We are aware of two recent accounting papers that make use of the quantile regression methodology. Konstantinidi and Pope [2016] and Chang et al. [2017] make use of the quantile regression methodology to understand the higher moments of earnings.



Large deviations have a stronger influence on the slope of the blue line compared to the red line. If both methods of regression are readily available, which one should we use? It depends on the purpose and how deviations between the prediction and the observed outcome are evaluated, i.e. which loss function is used. OLS may be better in some circumstances, whereas median regression (or other quantiles) might be more desirable in other cases.

If we are interested in making predictions and our prediction error is evaluated as squared loss, then least squares is indeed optimal. More formally, we can state that if we are looking to minimize the following loss function,

$$\min_g E[(y - g(x))^2],$$

then choosing $g(x) = E[y|x]$ is the best choice. In the evaluation of earnings forecasts, however, the common practice (see Bradshaw [2011] for a summary) is to evaluate the quality of a prediction by looking at the average absolute deviation, i.e. we typically look at averages (over firms and years) of

$$|Earnings - Predicted\ Earnings|.$$

In other words, we typically care about the mean (or median) absolute deviation between realized and predicted earnings. If this is our objective, it seems natural to incorporate this into the prediction directly. Therefore, it seems prudent to use a prediction model that directly minimizes mean absolute deviations. More formally, we could choose g to minimize:

$$\min_{g} E[|y - g(x)|].$$

That is, rather than minimizing squared deviations, we try to find a $g(X)$ that minimizes the absolute deviations. The median minimizes (expected) absolute deviations. Since we always concern ourselves with linear conditional mean and quantile functions, we aim to find b's to minimize:

$$\min_{b_0, b_1, \ldots, b_n} \sum_{i=1}^{i=N} |y_i - b_o - b_1 * x_{i,1} - b_2 * x_{i,2} - \ldots - b_n * x_{i,n}|.$$

At first, this objective function may look somewhat clumsy because it is not differentiable everywhere, but the important property is its convexity which makes numerical solutions very efficient (the complexity of the computation is actually less for this regression than it is for OLS!). [5]

Since our objective is to obtain the best possible earnings forecast in terms of absolute deviation from realized earnings, it seems quite natural to make forecasts using median regression. Similar to OLS regression, inference for estimated parameters in the case of quantile regression is usually based on asymptotic approximations, i.e. an application of a central limit theorem. In the case of linear quantile regression, we have the following result (Koenker [2005]):

[5] Standard statistical software packages such as R or Stata have readily available routines for quantile regression. In R, an implementation of estimation and inference for quantile regression is available in the "rq" package. Stata provides quantile regression functionality through the "qreg" command - one could run a simple quantile regression as "qreg y x".

$$\sqrt{N}(\hat{\beta}_T - \beta_T) \rightarrow^d \ N(0, V_T)$$

where $V_T = T(1-T)(E(x_i x_i' f(0|x_i)))^{-1}(E(x_i x_i')(E(x_i x_i' f(0|x_i)))^{-1}$ and $f(e|x_i)$ is the conditional density of e_i given $x_i = x$. In empirical work, V_T has to be estimated. Koenker [2005] shows how to estimate this matrix directly. Alternatively, inference can also be obtained through an application of the bootstrap.[6] Overall, statistical inference for the estimated parameters of a quantile regression model is readily available.

4 Empirical Results

In this section, we document how the use of quantile regressions in the estimation window when following the Hou et al. [2012] methodology produces better earnings estimates than the use of OLS in the estimation window when following the Hou et al. [2012] methodology.

4.1 Forecast Accuracy

4.1.1 Standard Estimation Window

In this subsection, we document how a modification to the Hou et al. [2012] methodology produces more accurate earnings estimates. Specifically, the use of quantile regressions in the estimation window produces earnings estimates that are more accurate than the ones produced using an OLS methodology in the estimation window.[7]

Hou et al. [2012] predict future earnings based on firm characteristics. Specifically, using the past ten years of data, they estimate coefficients using OLS from the following regression model for firm i in year t:

[6] Both a direct estimate of the asymptotic covariance matrix and inference by bootstrapping is readily available in standard statistical packages such as R or Stata.

[7] For the purposes of this section, we consider an OLS methodology with winsorization.

$$E_{i,t+1} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 ACC_{i,t} + \epsilon_{i,t+1},$$

where $E_{i,t}$ is the earnings of firm i during year t, $A_{i,t}$ is the total assets of firm i during year t, $D_{i,t}$ is the total dividends paid by firm i during year t, $DD_{i,t}$ is a dividend dummy that equals one if firm i paid dividends in year t, $NegE_{i,t}$ is a dummy that equals one if firm i had negative earnings in year t, and $ACC_{i,t}$ is the accruals of firm i over year t.

We use the same model and the same ten year estimation window, but estimate coefficients using quantile regressions. We then examine the difference in forecast accuracy from using a quantile regression methodology relative to an OLS methodology. Again, forecast accuracy is defined as the absolute value of the difference between forecasted earnings and actual earnings. Without winsorization, the results are striking. The results are presented in Table 1. First, we see the effect documented by Gerakos and Gramacy [2013] and Li and Mohanram [2014]; the mean absolute forecast error for one-year ahead earnings forecasts that are based on the OLS prediction model is significantly less accurate than forecasts based on a random walk (mean difference of \$4.52 million with a t-statistic of 8.19). Further, we note that the OLS-based forecasts for two and three years ahead are not significantly different from those based on a random walk. Second, we observe, however, that the forecasts of earnings based on the quantile regression prediction model are significantly more accurate than those based on a random walk for all three of the next three years (t-statistics of -3.44, -5.35 and -6.90). Finally, we see that the forecasts based on the quantile regression prediction model are significantly more accurate than those based on the OLS regression model for all three of the next three years (t-statistics of -5.32, -3.70, and -3.85).

Our approach offers an advantage over an approach that uses winsorization as it does not require ad-hoc specifications of when/how to cut the distribution, and allows us to consider the entire sample. Still, it is interesting to see whether a quantile regression approach is superior to an OLS approach after winsorization. We follow the approach in Hou et al. [2012] and winsorize all level variables each year at the 1 percent and 99 percent levels; we find significant evidence that a quantile regression approach is still superior to an OLS approach. The results are presented in Table 2, Panel A. Interestingly, we see that the forecasts based on the OLS prediction model are now significantly superior to a random walk forecast of earnings for years 2 and 3 (t-statistics of -3.55 and -4.87), but, importantly, the forecasts of next-year earnings based on the OLS prediction model fail to beat random walk predictions. The significance of the superior accuracy of the quantile regression-based prediction over the random walk prediction and over the OLS-based predictions observed in Table 1 (before winsorizing) is still apparent.

Next, we examine how scaling affects the results in a winsorized sample. We scale earnings by market equity. The results are presented in Table 2, Panel B. We find that a quantile regression approach relative to an OLS approach improves earnings forecasts by about 3 percent of the firm's market equity at the one-year horizon and by about 4 percent of the firm's market equity at the two-year and three-year horizons. These differences are each statistically significant at the one-percent level. They are also economically significant - the median earnings scaled by market equity in our sample is about 5 percent. Again, when the Hou et al. [2012] model is estimated using OLS, their predictions fail to statistically beat a random walk at the one-year horizon. However, when their model is estimated using a quantile regression approach, their predictions beat a random walk at the one-percent significance level.

In Figure 1, we highlight how a quantile regression approach produces more accurate forecasts on a year-by-year basis. We present box plots of the forecast errors from regressions that predict future earnings scaled by market equity. The average absolute forecast error is significantly smaller (at the one-percent level) in 36 out of the 44 years we estimate. A quantile regression approach produces far more accurate results in the vast majority of years. Overall, these results suggest that a quantile regression approach should be used to produce accurate estimates.

4.1.2 Shorter Estimation Window

We currently follow Hou et al. [2012] and use ten-year estimation windows. There are many reasons why it might be desirable to have shorter estimation windows - a lot of new listings, accounting rule changes, etc.. In this section, we analyze how shorter estimation windows affect the accuracy of our earnings predictions. When we use a quantile regression methodology, we see similar levels of accuracy for forecasts produced from a short estimation window and forecasts produced from a long estimation window. When we use an OLS regression methodology, we see that forecasts produced from a short estimation window are far less accurate than forecasts produced from a long estimation window.

The motivation for shorter estimation windows comes from the stability of the parameters. In Figure 2, we see that the estimated coefficients are far more stable when they are estimated via quantile regressions than via OLS. In particular, Table 3 lists the standard deviation of the parameters - the standard deviation is much lower for coefficients estimated via quantile regressions relative to coefficients estimated via OLS. This suggests that shortening the estimation window length may not affect the reliability of estimates produced using quantile regressions. In Table 4, we present how accurate our forecasts are when we use a one-year estimation window. Forecasts produced via a quantile regression methodology offer a similar level of accuracy to what we saw when we used a ten-year estimation window, while forecasts produced via an OLS regression methodology offer a far worse level of accuracy than what we saw when we used a ten-year estimation window. Specifically, when a one-year estimation window is used, a quantile regression approach relative to an OLS approach improves earnings forecast accuracy at the one-year horizon by about six percent of market equity. When a ten-year estimation window is used, this improvement in accuracy is only about three percent.[8] As such, it seems far more reasonable to use shorter estimation windows when following a quantile regression approach than when following an OLS approach.

4.2 Predicting Directional Changes

Practictioners and academics are concerned not only with the accuracy of earnings predictions, but also with the accuracy of earnings direction predictions. Will the firm's earnings grow? In this subsection, we document how a modification of the Hou et al. [2012] methodology - specifically, the use of quantile regressions in the estimation window - produces estimates that more accurately predict the direction of changes in earnings.

We present summary statistics in Table 5. When OLS is used in the estimation window, the right direction is predicted about 56.1 percent of the time. When quantile regressions are used, the right direction is predcited about 57.6 percent of the time. This difference in means is significant at the one-percent level.

Academics and practictioners are likely most concerned whether a firm will become profitable/unprofitable. As such, it is important to be able to predict movement across the zero earnings line. We find that the use of quantile regressions offers better predictions of whether earnings will cross the zero line than the use of OLS. The results are presented in Table 6. When OLS is used in the estimation window of the Hou et al. [2012] methodology, movement across/not across the zero earnings line is accurately predicted 78.9 percent of the time. When quantile regressions are used in the estimation window, movement across/not across the zero earnings line is accurately predicted 82 percent of the time. This difference

[8] Unsurprisingly, we also see that the coefficient estimates for the OLS regressions are less stable when a one-year estimation window is used. (Figure 3)

in means is statistically significant at the one-percent level.

4.3 Correlation of E/P with Realized Returns

Basu [1977] suggests that there is a positive relationship between earnings-to-price (E/P) ratios and future returns. We use this as motivation to examine the relationship between forecasted E/P ratios and future returns. We look at the partial correlations between our E/P ratios (one based on quantile regressions and one based on OLS regressions) and ex-post returns. The partial correlation measures the strength of the correlation between future returns and the E/P ratio controlling for the effect of the other E/P ratio. We find that the E/P ratio constructed using a quantile regression methodology produces stronger partial correlations with ex-post returns than the partial correlations between the E/P ratio constructed using an OLS regression methodology and ex-post returns.

We create E/P ratios by summing the next two years of forecasted earnings and dividing that number by the market value at the end of the fiscal year. We measure ex-post returns by looking at the return starting at the beginning of the third month after the end of the fiscal year.[9] The results are presented in Table 7. We first highlight one of the advantages of the original Hou et al. [2012] model; there is a much stronger partial correlation with future returns for an E/P measure based on the Hou et al. [2012] model than for the E/P measure based on a random walk. We then compare the original Hou et al. [2012] model with our modified Hou et al. [2012] model; an E/P measure based on quantile regressions has a stronger partial correlation with future returns than an E/P measure based on OLS regressions. Specifically, we find that the partial correlation between one-year future returns and the E/P measure based on quantile regressions is 0.0529, while the partial correlation between one-year future returns and the E/P measure based on OLS regressions is 0.0377. At the two-year horizon, we find that the partial correlation between future returns and the

[9] We find stronger results if we look at the return period starting in the month after the end of the fiscal year.

E/P measure based on quantile regressions is 0.0539, while the partial correlation between future returns and the E/P measure based on OLS regressions is 0.0487. If we look at the correlation between ranked observations, the results are even more striking. Specifically, the partial correlation between the future one-year return decile and the E/P decile based on quantile regressions is 0.0915, while the partial correlation between the future one-year return decile and the E/P decile based on OLS regressions is 0.0052. At the two-year horizon, we find that the partial correlation between the future return decile and the E/P decile based on quantile regressions is 0.1102 while the partial correlation between the future return decile and the E/P decile based on OLS regressions is 0.0242.

5 Conclusion

Earnings forecasts form the foundation for valuation models and, as such, are an important issue to accountants and financial economists. The Hou et al. [2012] methodology produces earnings forecasts that offer a number of advantages over analyst forecasts. Unfortunately, the resulting earnings forecasts are less accurate than a random walk. We show that an intuitive and simple modification to the Hou et al. [2012] methodology offers a number of improvements. Specifically, the use of quantile regressions in the estimation procedure produces earnings forecasts that are more accurate than a random walk, make better directional predictions, and allow for shorter estimation windows. As such, the Hou et al. [2012] methodology - with a slight modification - provides reliable earnings forecasts that can be used in valuation models. Future valuation models that need long-term earnings forecasts for all firms should use this modified version of the Hou et al. [2012] methodology.

References

S. Albrecht, L. Lookabill, and J. McKeown. The time-series properties of annual earnings. *Journal of Accounting Research*, pages 226–244, 1977.

R. Ball and P. Brown. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, pages 159–178, 1968.

R. Ball and R. Watts. Some time series properties of accounting income. *Journal of Finance*, 27(3):663–681, 1972.

S. Basu. Investment performance of common stocks in relation to their price - earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, 32(3):663–682, 1977.

M. Bradshaw. Analysts' forecasts: What do we know after decades of work? *Working paper*, 2011.

R. Brealey. Statistical properties of successive changes in earnings. *Unpublished paper*, 1967.

W. Chang, W.and Landsman and S. Monahan. Selecting an accounting based valuation model. *Working paper*, 2012.

W.J. Chang, S. Monahan, A. Ouazad, and F. Vasvari. The higher moments of future earnings. *Working paper*, 2017.

E. Fama and H. Babiak. Dividend policy: An empirical analysis. *Journal of American Statistical Association*, pages 1132–1161, 1968.

E. Fama and K. French. Forecasting profitability and earnings. *Journal of Business*, 73: 161–175, 2000.

E. Fama and K. French. Profitability, investment and average returns. *Journal of Financial Economics*, 82:491–518, 2006.

J. Gerakos and R. Gramacy. Regression-based earnings forecasts. *Working paper*, 2013.

P. Griffin. The time series properties of annual quarterly earnings: Preliminary evidence. *Journal of Accounting Research*, pages 71–83, 1977.

K. Hou, M. van Dijk, and Y. Zhang. The implied cost of capital: A new approach. *Journal of Accounting and Economics*, 53(3):504–526, 2012.

C. Jones and S. Tuzel. Inventory investment and the cost of capital. *Journal of Financial Economics*, 107(3):557–579, 2012.

R. Koenker. *Quantile regression.* Cambridge University Press, 2005.

T. Konstantinidi and P.F. Pope. Forecasting risk in earnings. *Contemporary Accounting Review*, 33(2):487–525, 2016.

C. Lee, E. So, and C. Wang. Evaluating firm-level expected return proxies. *Working paper*, 2017.

K. Li and P. Mohanram. Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies*, 19(3):1152–1185, 2014.

J. Lintner and R. Glauber. *Modern Development in Investment Management: A Book of Readings*, chapter Higgledy Piggledy Gowth in America. Praeger Publishers, 1972.

P. Patatoukas. Customer-base concentration: Implications for firm performance and capital markets. *The Accounting Review*, 87(2):363–392, 2011.

R. Watts and R. Leftwich. The time series of annual accounting earnings. *Journal of Accounting Research*, pages 253–271, 1977.

Table 1: Forecast Accuracy - Main Result

This table highlights how quantile regressions substantially improve the accuracy of forecasts as measured by mean absolute deviation of the forecast from the actual. HVZ = the Hou et al. [2012] methodology and RW = Random walk. The HVZ methodology involves predictions based on estimating a model for predicting earnings based on the past ten years of data. Specifically, to forecast next year's earnings, they estimate: $E_{i,t+1} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 ACC_{i,t} + \epsilon_{i,t+1}$, where $E_{i,t}$ is the earnings of firm i at time t, $A_{i,t}$is the total assets of firm i at time t, $D_{i,t}$is the total dividends paid of firm i at time t, $DD_{i,t}$is a dividend dummy for firm i at time t that equals one if the firm is a dividend payer, $NegE_{i,t}$is a dummy that equals one if firm i had negative earnings in time t, and $ACC_{i,t}$is the accruals of firm i during time t. Earnings is defined as income before extraordinary items from Compustat. LS indicates the model is estimated using least squares. RQ indicates that the model is estimated using a quantile regression. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

	Earnings in year (t+1)	Earnings in year (t+2)	Earnings in year (t+3)
Accuracy of HVZ (LS) - Accuracy of RW	4.52***	1.13	-0.26
t-statistic	(8.19)	(1.49)	(-0.33)
Number of Observations	147,062	134,846	123,841
Accuracy of HVZ (RQ) - Accuracy of RW	-0.80***	-2.57***	-4.11***
t-statistic	(-3.44)	(-5.35)	(-6.90)
Number of Observations	147,062	134,846	123,841
Accuracy of HVZ (RQ) - Accuracy of HVZ(LS)	-5.32***	-3.70***	-3.85***
t-statistic	(-13.95)	(-11.03)	(-11.71)
Number of Observations	147,062	134,846	123,841

Table 2: Forecast Accuracy (After Winsorization and Scaling)

This table highlights how quantile regressions still substantially improve the accuracy of forecasts as measured after winsorizing and after scaling by market equity. HVZ = the Hou et al. [2012] methodology and RW = Random walk. The HVZ methodology involves predictions based on estimating a model for predicting earnings based on the past ten years of data. Specifically, to forecast next year's earnings, they estimate: $E_{i,t+1} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 ACC_{i,t} + \epsilon_{i,t+1}$, where $E_{i,t}$ is the earnings of firm i at time t, $A_{i,t}$is the total assets of firm i at time t, $D_{i,t}$is the total dividends paid of firm i at time t, $DD_{i,t}$is a dividend dummy for firm i at time t that equals one if the firm is a dividend payer, $NegE_{i,t}$is a dummy that equals one if firm i had negative earnings in time t, and $ACC_{i,t}$is the accruals of firm i during time t. Earnings is defined as income before extraordinary items from Compustat. LS indicates the model is estimated using least squares. RQ indicates that the model is estimated using a quantile regression. ME indicates that earnings are scaled by market equity in the prediction model and in the predictions. To scale by market equity, we divide by the quantity of the number of common shares outsanding times the price at the close of the fiscal year. We winsorize each level variable at the 1% level and the 99% level each year. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

	Earnings in year (t+1)	Earnings in year (t+2)	Earnings in year (t+3)
Panel A: No Scaling			
Accuracy of HVZ (LS) - Accuracy of RW	0.21	-1.94***	-2.89***
t-statistic	(0.60)	(-3.55)	(-4.87)
Number of observations	147,062	134,846	123,841
Accuracy of HVZ (RQ) - Accuracy of RW	-0.80***	-2.57***	-4.11***
t-statistic	(-3.44)	(-5.35)	(-6.90)
Number of observations	147,062	134,846	123,841
Accuracy of HVZ (RQ) - Accuracy of HVZ(LS)	-1.01***	-0.64***	-1.22***
t-statistic	(-5.51)	(-3.85)	(-7.74)
Number of observations	147,062	134,846	123,841
Panel B: Scaling by Market Value of Equity			
Accuracy of HVZ (LS) - Accuracy of RW - ME	-0.01	-0.02***	-0.03***
t-statistic	(-1.55)	(-3.82)	(-4.53)
Number of observations	144,696	132,390	121,444
Accuracy of HVZ (RQ) - Accuracy of RW - ME	-0.04***	-0.06***	-0.07***
t-statistic	(-7.35)	(-8.52)	(-8.64)
Number of observations	144,696	132,390	121,44
Accuracy of HVZ (RQ) - Accuracy of HVZ(LS) - ME	-0.03***	-0.04***	-0.04***
t-statistic	(-18.24)	(-33.04)	(-31.90)
Number of observations	144,696	132,390	121,444

Figure 1: Forecast Accuracy

The graph below shows how the forecasts of earnings scaled by market equity - from the Hou et al. [2012] model - are far more accurate in the vast majority of years if we use a quantile regression prediction model (RQ) rather than a least squares prediction model (OLS). We present box plots representing the mean, interquartile range, min and max forecast error.

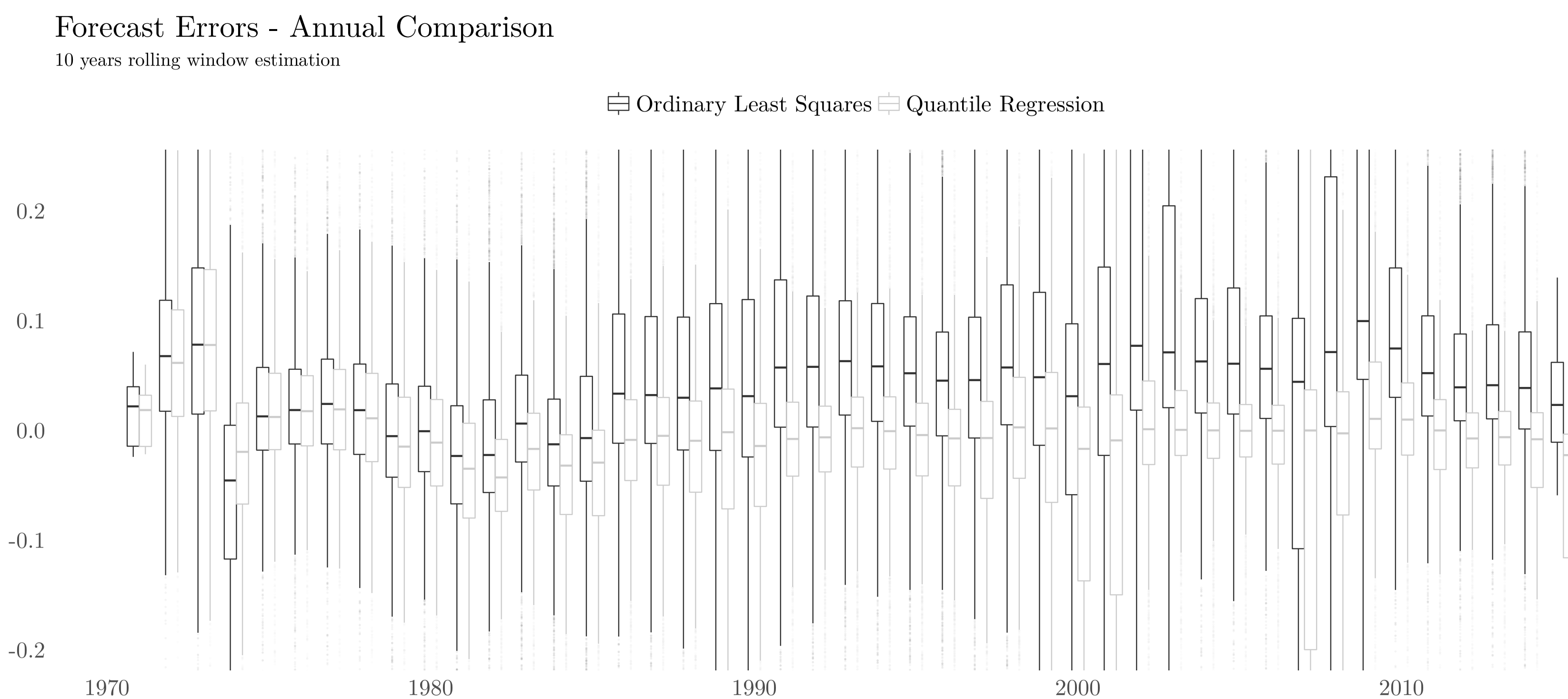


Figure 2: Parameter Stability

The graph below shows how the parameters - from the Hou et al. [2012] model - are much more stable (we normalize each parameter to have mean zero) over time from using a quantile regression prediction model (RQ) rather than a least squares prediction model (OLS).

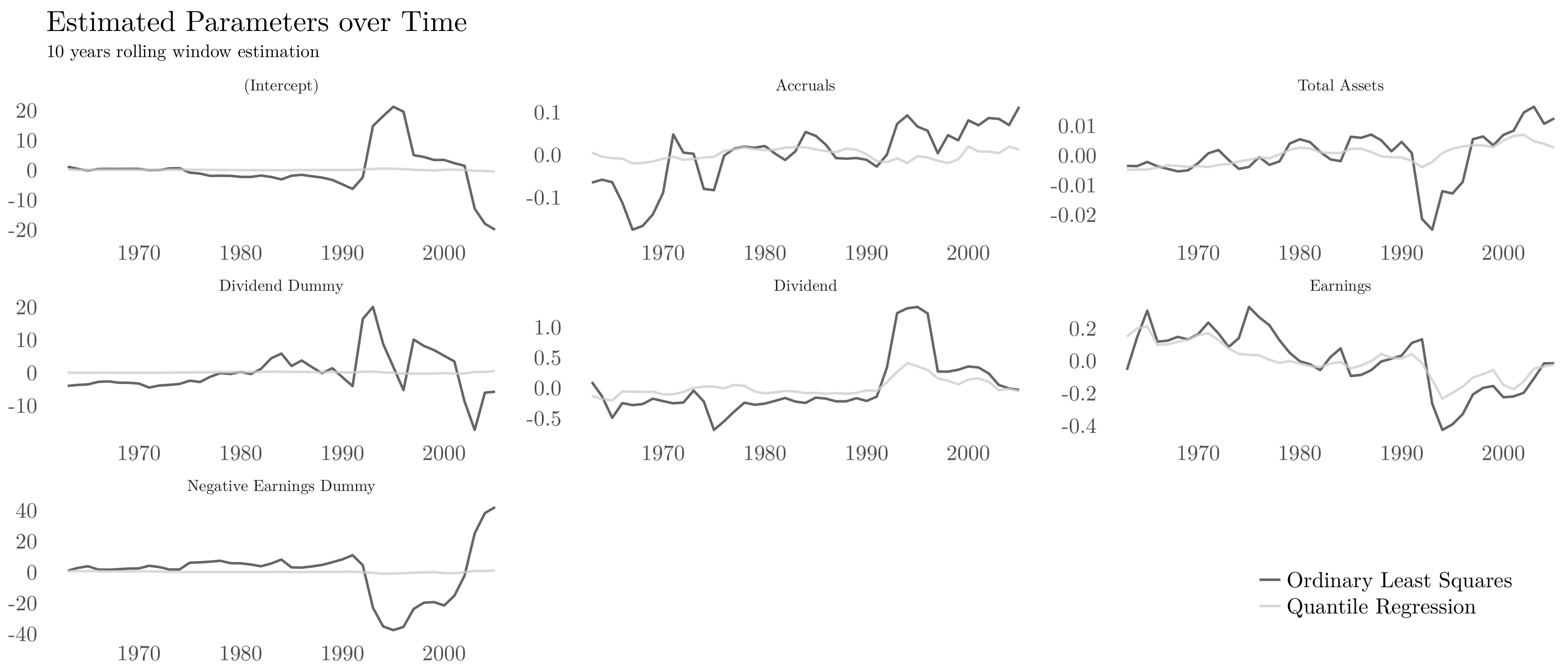


Table 3: Summary of the Parameter Estimates

This table presents the mean, median, and standard deviation of the estimated coefficients from the Hou et al. [2012] model. Specifically, we estimate: $E_{i,t+1} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 ACC_{i,t} + \epsilon_{i,t+1}$, where $E_{i,t}$ is the earnings of firm i at time t, $A_{i,t}$is the total assets of firm i at time t, $D_{i,t}$is the total dividends paid of firm i at time t, $DD_{i,t}$is a dividend dummy for firm i at time t that equals one if the firm is a dividend payer, $NegE_{i,t}$is a dummy that equals one if firm i had negative earnings in time t, and $ACC_{i,t}$is the accruals of firm i during time t. OLS indicates that we used an ordinary least squares model and RQ indicates that we used a quantile regression method.

Variable	Method	Mean	Median	Standard Deviation
Intercept	OLS	-0.1823	-0.4164	7.6923
Accruals	OLS	-0.0280	-0.0235	0.0695
Total Assets	OLS	0.0076	0.0082	0.0082
Dividend Dummy	OLS	5.6309	4.3001	6.4385
Dividends	OLS	0.2723	0.0925	0.4753
Earnings	OLS	0.7305	0.7422	0.1823
Negative Earnings	OLS	-1.3238	1.8597	16.2610
Intercept	RQ	-0.0003	0.0273	0.1745
Accruals	RQ	-0.0218	-0.0211	0.0128
Total Assets	RQ	0.0067	0.0070	0.0032
Dividend Dummy	RQ	0.2059	0.2166	0.1891
Dividends	RQ	0.1076	0.0521	0.1358
Earnings	RQ	0.9055	0.8970	0.1065
Negative Earnings	RQ	0.2691	0.2690	0.4630

Table 4: Forecast Accuracy with Shorter Estimation Windows

This table documents forecast accuracy when the estimation window is one-year long. HVZ = the Hou et al. [2012] methodology and RW = Random walk. The HVZ methodology involves predictions based on estimating a model for predicting earnings based on the past ten years of data. Specifically, to forecast next year's earnings, they estimate: $E_{i,t+1} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 ACC_{i,t} + \epsilon_{i,t+1}$, where $E_{i,t}$ is the earnings of firm i at time t, $A_{i,t}$is the total assets of firm i at time t, $D_{i,t}$is the total dividends paid of firm i at time t, $DD_{i,t}$is a dividend dummy for firm i at time t that equals one if the firm is a dividend payer, $NegE_{i,t}$is a dummy that equals one if firm i had negative earnings in time t, and $ACC_{i,t}$is the accruals of firm i during time t.Earnings is defined as income before extraordinary items from Compustat. LS indicates the model is estimated using least squares. RQ indicates that the model is estimated using a quantile regression. ME indicates that earnings are scaled by market equity in the prediction model and in the predictions. To scale by market equity, we divide by the quantity of the number of common shares outsanding times the price at the close of the fiscal year. We winsorize each level variable at the 1% level and the 99% level each year. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

	Earnings in year (t+1)	Earnings in year (t+2)	Earnings in year (t+3)
Accuracy of HVZ (LS) - Accuracy of RW - ME	0.03***	-0.01**	-0.02***
t-statistic	(12.59)	(-2.47)	(-5.48)
Number of observations	144,696	132,390	121,444
Accuracy of HVZ (RQ) - Accuracy of RW - ME	-0.03***	-0.06***	-0.06***
t-statistic	(-6.68)	(-9.96)	(-10.97)
Number of observations	144,696	132,390	121,444
Accuracy of HVZ (RQ) - Accuracy of HVZ(LS) - ME	-0.06***	-0.05***	-0.05***
t-statistic	(-14.58)	(-12.98)	(-17.93)
Number of observations	144,696	132,390	121,444

Figure 3: Parameter Stability with a One-Year Estimation Period

The graph below shows how the parameters - from the Hou et al. [2012] model - are much more stable (we normalize each parameter to have mean zero) over time from using a quantile regression prediction model (RQ) rather than a least squares prediction model (OLS).

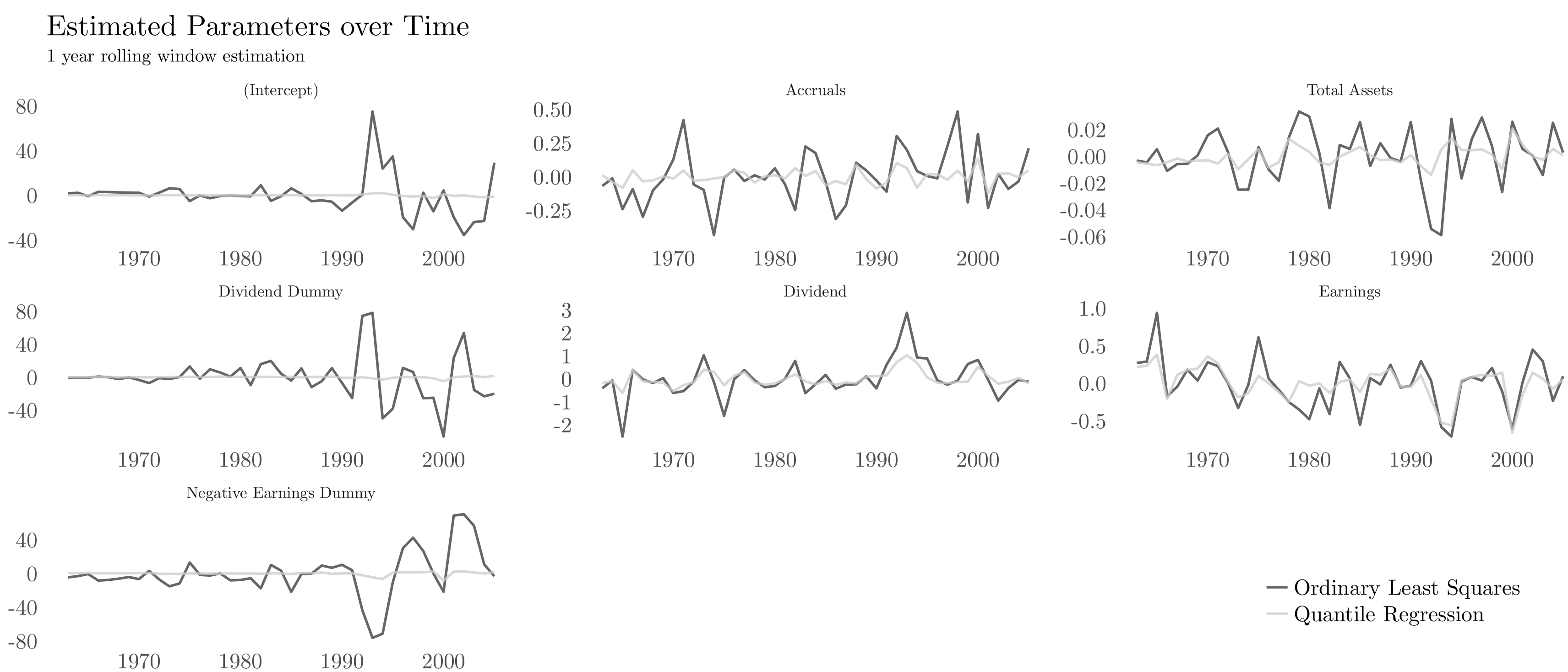


Table 5: Predicting Increases and Decreases in Earnings

This table demonstrates how a quantile regression methodology is better than a least squares methodology to predict directional movements in earnings.

	Least Squares	Quantile Regression
Direction is Accurately Predicted	82,554	84,743
Direction is Inaccurately Predicted	64,508	62,319
Fraction of Time Prediction is Right	0.561	0.576

Table 6: Predicting Movements Across Zero Earnings

This table demonstrates how the use of quantile regressions to predict movements across zero earnings is better than a least squares prediction methodology. The first two rows indicate the number of times when there is a predicted cross that agrees with the actual. The next two rows indicates the number of times there was a cross that was not predicted. The fifth and sixth rows indicate the number of times a cross was predicted and did not happen. The final row indicates the fraction of times that the prediction agreed with the actual.

	Least Squares	Quantile Regression
Crossing Zero Earnings is Accurately Predicted	116,046	120,640
Crossing Zero Earnings is Not Accurately Predicted	31,016	26,422
Fraction of Time Prediction is Right	0.789	0.820

Table 7: Partial Correlations

This table documents the partial correlation between future returns and the sum of next two years forecasted earnings scaled by market equity (E/P). Market equity is as of the end of the fiscal year. The return period starts at the beginning of the third month after the end of the fiscal year. LS indicates the use of earnings forecasts derived from a prediction model that uses least squares. RQ indicates the use of earnings forecasts derived from a prediction model that uses quantile regressions. RW indicates the use of earnings forecasts based on a random walk. The second column indicates what variable we are controlling for when we estimate the partial correlation. The third and fourth columns indicate the partial correlation with the one-year return and two-year return, respectively. Panel A documents the partial correlation among the variables. Panel B documents the partial correlation among the variables after sorting the variables into deciles. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

	Controlling for	One-year Return	Two-year Return
Panel A: Variable Correlations			
E/P (LS)	E/P (RW)	0.1001***	0.1144***
E/P (RW)	E/P (LS)	-0.0069**	-0.0054*
E/P (LS)	E/P (RQ)	0.0377***	0.0487***
E/P(RQ)	E/P(LS)	0.0529***	0.0539***
Panel B: Decile Correlations			
E/P (LS)	E/P (RQ)	0.0052	0.0242***
E/P(RQ)	E/P(LS)	0.0915***	0.1102***