Review of Quantitative Finance and Accounting
https://doi.org/10.1007/s11156-022-01048-w

ORIGINAL RESEARCH

A two-step quantile regression method for discretionary accounting

May Huaxi Zhang[1] · Stanley Iat-Meng Ko[2] · Andreas Karathanasopoulos[3] · Chia Chun Lo[4]

Accepted: 28 November 2021

Abstract

This paper proposes an analytical approach that complements the traditional two-step linear regression and one-single step linear regression suggested by Chen et al. (J Account Res 56:751–796, 2018). Using the regression residual as the dependent variable in a second regression is a procedure commonly used in studying discretionary accounting. Chen et al. (J Account Res 56:751–796, 2018) propose to adopt one-step regression to avoid estimation bias and inference error. However, the mean level effect estimated by one-step OLS regression is not sufficient to capture the overall spectrum of discretionary accounting behaviors and thus may mislead its user in drawing implications. We use two-stage quantile regression to examine determinants of discretionary accounting such as discretionary accruals, discretionary expense, discretionary book-tax differences, and abnormal investment in different quantiles. We illustrate the differences between the one-step regression and our two-step quantile regression using four common discretionary accounting studies. Our results and implications reconcile, to some extent, the contradictory findings between results of the one-step OLS regression and the previous established works based on two-step regression.

Keywords Two-stage · Residuals · Coefficient bias · Quantile regression · Discretionary accruals

JEL Classification C18 · G10 · G30 · M40 · M41

1 Introduction

Using the linear regression residual as the dependent variable in a second regression is a procedure commonly used in empirical accounting and finance research. In a review paper about earnings quality, Dechow et al. (2010) document that "almost one hundred

We are indebted to the two anonymous reviewers and the editor for providing insightful comments and directions for improvement which has resulted in the current draft.

✉ Stanley Iat-Meng Ko
stanleyko@tohoku.ac.jp

Extended author information available on the last page of the article

Published online: 28 February 2022

papers in [their] database use abnormal accruals as a measure of earnings quality and test predicted determinants or consequences. These studies test the joint hypothesis that the residual from an accruals model reflects earnings management and that the predicted determinant induces earnings management or that earnings management has a predicted consequence".[1] Specifically, in the first-step regression, the researcher decomposes a dependent variable into its predicted and residual components, and the residuals are considered as the discretionary component or abnormality of the variable in question. Further examinations of the determinants of the discretionary component are then conducted. Such a procedure has become one of the standard approaches in accounting studies and is widely applied in research on discretionary accruals, real activities management, unexplained audit fees, and discretionary book-tax difference, among others. A thorough review of the popularity of the residual approach in accounting can be found in Chen et al. (2018).

Despite the wide adoption of the procedure, Chen et al. (2018) and Christodoulou et al. (2018) recently point out the fundamental problem in the two-step regression. Chen et al. (2018) use the Frisch-Waugh-Lovell Theorem and simulations to demonstrate biases in coefficient estimates and their t-statistics. Consistent with Chen et al. (2018), Christodoulou et al. (2018) also show systematic biases in inference-in-residuals which could render the results an invalid interpretation. An earlier review and critic of the two-step regression method can be found in Beaver (1987), who shows that the conventional method can lead to downward biased estimates in studies of residual security returns. Both Chen et al. (2018) and Christodoulou et al. (2018) suggest a combined single-step instead of two-step regression. However, to the best of our knowledge, there is no satisfactory replacement of the long existing procedure in studying discretionary accounting. In this paper, we aim to fill this void by proposing a new approach. We suggest a novel two-step regression with the first-step as the OLS regression and the second-step using quantile regression.

We hypothesize that the variable in question, $TOTAL\_Y$, is composed of non-discretionary, discretionary, and random residual components,

$$TOTAL\_Y = NON\_DISCRETIONARY\_Y + DISCRETIONARY\_Y + RESIDUAL. \quad (1)$$

Comparing with the traditional two-step approach, the non-discretionary part contains the regressors in the first-step regression, and the discretionary part includes all the discretionary determining variables in the second-step regression. Hence, Eq. (1) is the same as the combined single-step full model proposed in Chen et al. (2018) and Christodoulou et al. (2018). As reviewed in Chen et al. (2018), a majority of related studies do not include the independent variables from the first-step regression as additional regressors in the second-step regression.[2] Thus, we assume that the independent variables in $NON\_DISCRETIONARY\_Y$ and $DISCRETIONARY\_Y$ do not overlap. We also do not consider other variants in the traditional two-step approach, such as transformed or partitioned variables, in calculating the first-step regression residuals.

Before analyzing the discretionary accounting, we need to obtain a consistent estimates of the non-discretionary part, which will then be purged out from $TOTAL\_Y$. We follow the suggestion in Chen et al. (2018) and Christodoulou et al. (2018) and estimate (1) by a single-step OLS regression, including both the $NON\_DISCRETIONARY\_Y$ and $DISCRETIONARY\_Y$. Once we obtain the consistent estimate, the discretionary

[1] Cited from the working paper version of Dechow et al. (2010).

[2] See page 752 in Chen et al. (2018).

value can be extracted by removing the non-discretionary part from *TOTAL_Y*; i.e., $TOTAL\_Y - \widehat{NON\_DISCRETIONARY\_Y}$. The consistency is theoretically and numerically shown in Chen et al. (2018).

As pointed out by the referee, the genuine reason behind the failure of the traditional two-step regression approach is the endogeneity problem of omitted variables. The endogeneity occurs in the first step regression when *DISCRETIONARY_Y* is correlated with *NON_DISCRETIONARY_Y* and at the same time excluded from (1). The single-step method proposed by Chen et al. (2018) address this omitted-variable endogenous problem by including all regressors in one regression equation. In general, there are other possible sources of endogeneity even we adopt the single-step regression. Reverse causality and measurement errors are common when we use non-experimental data, which is the case in most of the accounting studies. Endogeneity problem and instrumental variables (IV) method have become a popular topic in recent accounting study. Du and Shen (2018) study how peer performance affects firms' earnings management decisions. They propose the IV method to address the possible endogenous problem from measurement errors. Huang et al. (2017) investigate the effect of stock liquidity on accrual-based earnings management. They concern about the reverse causality between earnings management and stock liquidity and conduct the difference-in-differences approach with a quasi-natural experimental design. Accounting studies concerning endogeneity problem can also be found in Shi et al. (2015) and Horrace and Reddic (2014), among others. In this paper, we focus on the general methodological framework of discretionary accounting study. Other endogeneity problems and IV methods are on a case-by-case basis depending on the topics in question.

The key assumption of our approach is that the coefficients (β's) in *NON_DISCRETIONARY_Y* are fixed across quantiles while the other coefficients in *DISCRETIONARY_Y* vary across quantiles, reflecting the discretionary behavior. To motivate our assumption, let us reconsider the traditional setup. Firstly in step one, we have,

$$TOTAL\_Y = NON\_DISCRETIONARY\_Y + RESIDUAL. \tag{2}$$

Then after estimating the *NON_DISCRETIONARY_Y*, the *RESIDUAL* is extracted as the discretionary part and the researcher further studies the determinants effects by running the regression in step-two

$$RESIDUAL \quad \text{on} \quad DISCRETIONARY\_DETERMINANTS. \tag{3}$$

Hence, we can see that in the second regression above it is the "variation" of the *RESIDUAL* reflecting the discretionary behavior which is to be explained by certain determinants. However, Chen et al. (2018) point out that the above two-step estimation is inconsistent, and propose a one-step regression combining the *NON_DISCRETIONARY_Y* factors and *DISCRETIONARY_DETERMINANTS* in a single equation. The one-step estimation, on the one hand, addresses the inconsistent problem. But, on the other hand, it conceals the discretionary variation. The single equation regression only estimates the marginal effects of all the regressors on the original dependent variable *TOTAL_Y* in which the discretionary and non-discretionary parts entangled. We would like to point out that the traditional method is "statistically" problematic but "theoretically" feasible in studying discretionary accounting. However, the Chen et al. (2018) one-step approach, though "statistically" correct, cannot achieve the original purpose in studying discretionary accounting. Therefore, we compromise the two seemingly conflicting "statistical" and "theoretical" issues by using a novel model which contains the quantile variant (discretionary) and quantile

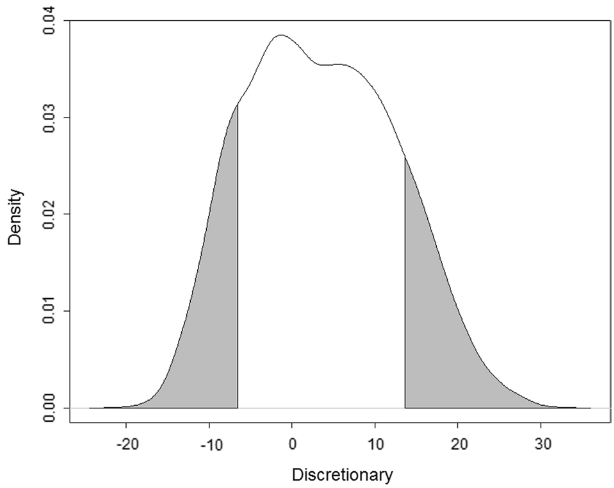

Fig. 1 Discretionary unconditional distribution where discretionary behavior is most likely taking place at the shaded area corresponding to the tail areas far away from the mean

invariant (non-discretionary) parts in one equation. To be specific, let us consider the following setup

$$y = \underbrace{\beta_0 + \beta_1 x_1}_{\text{non-discretionary}} + \underbrace{\beta_2 x_2 + \beta_3 x_3}_{\text{discretionary}} + \epsilon,$$

where β_0, β_1 are fixed accross the conditional quantiles of y, and β_2, β_3 are varying accross quantiles. In the non-discretionary part, we are trying to imitate the traditional method in (2) which only reflects the mean level by usual regression. In order to allow the model to capture the "variation" of the "*RESIDUAL*" in (2) or (3), we adopt the quantile setup in the discretionary part above which captures the variation through the varying quantile parameters β_2 and β_3.[3] Further more, we prove in Sects. 2 and 3 that the above model is consistently estimable. Therefore, our method addresses the inconsistent problem and, at the same time, restore the original design in discretionary accounting study.

After extracting the discretionary part, $TOTAL\_Y - \widehat{NON\_DISCRETIONARY\_Y}$, from (1), we proceed to study the discretionary accounting in different quantiles using quantile regression. According to Moore (1973), discretionary accounting, such as a write-down, write-off, or provision for future cost or loss, can be regarded as the departure or discrepancy from the normal level, which can be supported by normal transaction. The larger the discrepancy, the more severe the discretionary accounting behavior. Figure 1 illustrates a hypothetical density plot of the discretionary component. It is obvious that suspect discretionary accounting would be more likely to occur in the tail regions (the shaded area). From this perspective, we propose to study discretionary behavior at different quantiles through quantile regressions. Figure 2 demonstrates the regression lines obtained from regressing the discretionary component on its determinant. The plot shows that the discretionary determinant is effective at the tail quantiles; i.e. the larger the value of the determinant factor, the larger the discrepancy from the average. However, for the OLS regression line, which is relatively flat, only a small (or even insignificant) effect can be detected.

[3] See Sect. 2 for a more detailed introduction of the model.

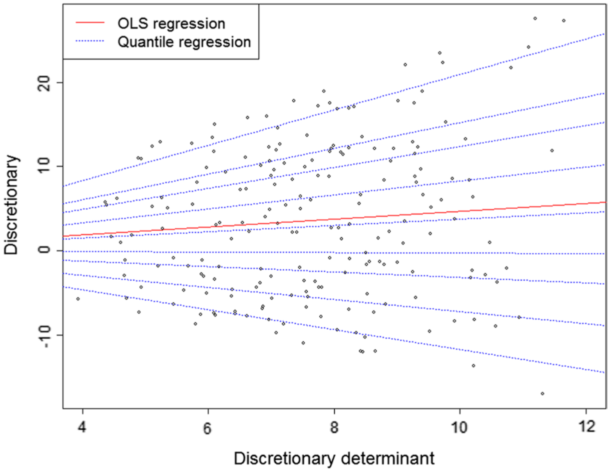

Fig. 2 Discretionary quantile regression demonstrating the possibility of heterogeneous responses with respect to the discretionary determinant across quantiles

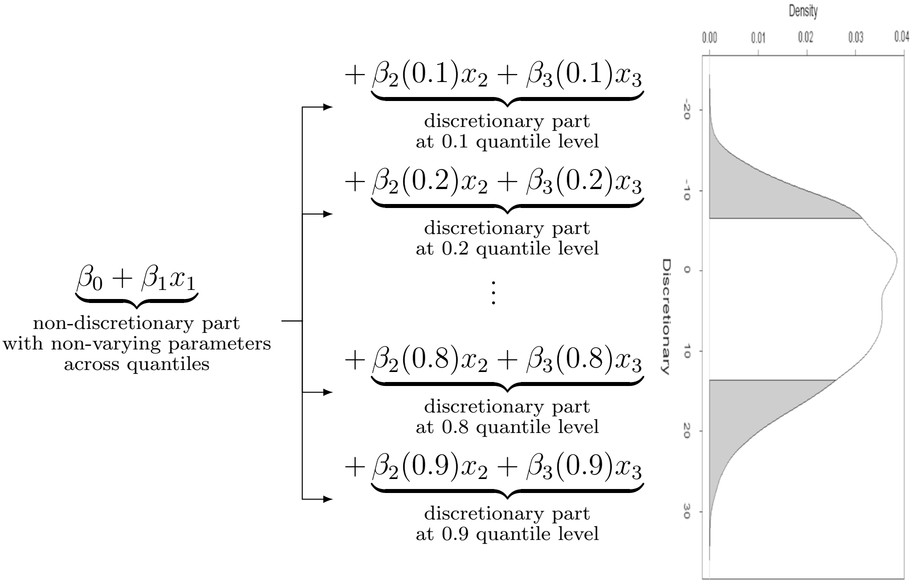

Fig. 3 The proposed two-step approach where the non-discretionary part is assumed with non-varying parameters across quantiles, whereas the discretionary part can be varying in different quantiles

Figure 3 shows an overview of our model structure. The coefficients in the non-discretionary part are assumed to be constant across qunatiles whereas the coefficients in the discretionary part are allowed to be varying.

Quantile regression, which is proposed in Koenker and Bassett (1978), has been regarded as an extension of classical OLS estimation of conditional mean models. A review can be found in Koenker and Hallock (2001). The traditional OLS regression method examines the conditional mean of the dependent variable and thus focuses on the central location of the conditional distribution. In contrast, the quantile regression estimates the effect of the regressors on the dependent variable at different quantiles of the conditional distribution. Hence, quantile regression provides a more complete picture of the marginal effects of determinants. It is typically useful when the research questions look for heterogeneous behaviors at non-central location of the response distribution. Du et al. (2013) use quantile regression to study currency exposure. They show that the quantile regression is useful in exploring heterogeneous behaviors across different quantiles. Huang (2013) proposes a process in VaR estimation with methods of quantile regression and kernel estimator to realize a tail distribution and locate the VaR estimates, see also the application of quantile regression for CoVaR in Su (2021). Du and Zhao (2017) demonstrate that the OLS regressions are incapable of revealing heterogeneity in oil price changes. They identify the differential effects by quantile regression. Ahmed and Doukas (2021) use quantile regression to study behavioral finance disposition effect and momentum. Moreover, quantile regression is known to be more robust than OLS regression which is sensitive to extreme values in the response measurements. McKee and Kagan (2016) adopt quantile regression to generate outlier-free robust estimates of cost efficiency in small asset U.S. credit unions. Accounting studies using the quantile regression method can also be found in Armstrong et al. (2015), Lai et al. (2018), and Chen et al. (2019), among others. See also a recent comprehensive review of research methodologies in Lee (2020).

Our method consists of the first-step OLS regression and the second-step quantile regression. The first-step aims at consistently estimating the discretionary component by purging out the non-discretionary part from the total values. The second-step examines the effects of the discretionary determinant at different quantiles. If discretionary behavior exist, we expect the determinants to show significant impact at the tail quantile levels, while this effect may or may not be identified at the mean level; i.e. from the OLS regression line.

The contributions of our paper are two-fold. First, we provide a solution to accounting researchers who aim at identifying the determinants of the discretionary, abnormal, or unexplained components of variables. We provide explicit theoretical proof of the consistent estimate of the discretionary component from the total value in question. We also demonstrate through a simulation experiment the differentiation of various discretionary behaviors at different quantiles. Second, we complement the quantile regression application, which has grown in recent accounting literature. We show that the quantile regression turns out to be a useful and reasonable method in studying accounting discretionary behavior.

In the subsequent sections, we specifically state all the assumptions and provide the econometric justification of our two-step approach, followed by a detail simulation study. We further replicate the applications in Chen et al. (2018) to demonstrate the usefulness our approach. Our results show that in some cases the OLS regression captures no effect of the determinants on the discretionary component, while the effects take place at the tail quantiles, as illustrated in Fig. 2. The remainder of the paper is organized as follows. Section 2 introduces the details and theoretical background of our proposed method in studying discretionary accounting determinants. Section 3 conducts a simulation study. Section 4 provides empirical applications in various accounting questions, and Sect. 5 concludes.

2 The two-step quantile regression framework

In this section, we state the assumptions for our two-step method, from which we derive the consistency proposition of the estimation. Without loss of generality, we consider the model

$$y = \underbrace{\beta_0 + \beta_1 x_1}_{\text{non-discretionary}} + \underbrace{\beta_2 x_2 + \beta_3 x_3}_{\text{discretionary}} + \epsilon. \tag{4}$$

Assumption 1 The regression model (4) is correctly specified, with $E(\epsilon|x_1, x_2, x_3) = 0$.

Assumption 2 The coefficients in the non-discretionary part in model (4); i.e. β_0 and β_1; are fixed across the conditional quantiles of y.

Assumption 3 The coefficients in the discretionary part in model (4), are varying across the conditional quantiles of y, with values $\beta_2(\tau)$ and $\beta_3(\tau)$ corresponding to the quantile level $\tau \in (0, 1)$.

Remark 1 Correct model specification is guided by accounting theory and depends on the studying subject. In this paper, we only focus on the general methodology and leave the modeling justification to researchers.

Remark 2 Assumptions 2 and 3 imply that the τth conditional quantile of y is

$$q_\tau(y|x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2(\tau)x_2 + \beta_3(\tau)x_3. \tag{5}$$

See also Figure 3 for reference.

From the above assumptions, we can derive the following consistency proposition for our first-step OLS regression.

Proposition 1 *Under Assumptions 1, 2, and 3, for OLS regression (4), the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent with respect to β_0 and β_1, and the estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ are consistent with respect to*

$$\beta_k = \int_0^1 \beta_k(\tau)d\tau \quad k = 2, 3. \tag{6}$$

Proof The OLS consistency follows Assumption 1, while equation (6) follows the random coefficient interpretation of quantile regression (see section 2.6 in Koenker 2005). □

Corollary 1 *Under Assumptions 1 to 3, the discretionary part is consistently estimated by $\tilde{y} \equiv y - \hat{\beta}_0 - \hat{\beta}_1 x_1$.*

Hence, with Proposition 1 and Corollary 1, we propose the following two-step approach for studying discretionary accounting.

Step 1 Run the OLS regression of y on x_1, x_2, x_3. Estimate the discretionary component by $\tilde{y} = y - \hat{\beta}_0 - \hat{\beta}_1 x_1$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained from the one-step OLS regression of y on x_1, x_2, and x_3.

Step 2 Conduct quantile regression of $\tilde{y}$ on x_2 and x_3 to study the discretionary behavior at different quantiles.

Note that our Step 1 estimation is exactly the same as the one-step OLS regression in Chen et al. (2018).

As discussed in Sect. 1, the discretionary variation can be explored from $\tilde{y}$ and the discretionary behavior is regarded as the departure or discrepancy from the normal level (around the median level). Suppose the accounting theory conjectures that, say, x_2 is mitigating the discretionary behavior. For lower quantile level (e.g. $\tau = 0.1$ and 0.2) when $\tilde{y}$ value is negative, the expected sign of $\beta_2(\tau)$ should be positive. This is, the effect of x_2 at the lower quantile level tends to pull back the values of $\tilde{y}$ from very negative to the central. On the other side at the upper quantiles (e.g. $\tau = 0.8$ and 0.9) when $\tilde{y}$ value is positive, the expected sign of $\beta_2(\tau)$ should be negative. Thus, for a discretion mitigating determinant, the overall effect is to shrink extreme quantile level $\tilde{y}$ towards the central. In the same vein, for a discretion exaggerating determinant, the overall effect is to push extreme quantile level $\tilde{y}$ farther away from the central, i.e., an expected negative (positive) sign of $\beta_2(\tau)$ at the lower (upper) quantile level.

3 Simulation study

Note that the model (4) or (5) and the corresponding estimation procedure are new in the related literature. The novelty of the model comes from that the variant $\beta_2(\tau), \beta_3(\tau)$ and invariant β_0, β_1 simultaneously exist in the linear model. In this section, we conduct a detailed simulation study for our proposed two-step method. We simulate the data from equation (4) with

$$(x_1, x_2, x_3)' \sim N(\mu, \Sigma), \quad \mu = (1.0, 6.0, 0.0)', \quad \Sigma = \begin{bmatrix} 1.0 & -1.0 & 0.5 \\ -1.0 & 4.0 & 0.3 \\ 0.5 & 0.3 & 1.0 \end{bmatrix}$$

which implies $Var(x_1) = 1$, $Var(x_2) = 2$, $Var(x_3) = 1$, and $corr(x_1, x_2) = -0.5$, $corr(x_1, x_3) = 0.5$, $corr(x_2, x_3) = 0.15$. The non-discretionary coefficients are fixed across quantiles such that $\beta_0 = 0.5$ and $\beta_1 = 1.5$. In order to conduct a more accurate simulation, we consider a fine quantile grid, $0 < \tau_1 < \cdots < \tau_{99} < 1$, such that $\tau_j - \tau_{j-1} = 0.01$. We set $-1.5 = \beta_2(\tau_0) < \cdots < \beta_2(\tau_{99}) = 2.5$ with $\beta_2(\tau_j) - \beta_2(\tau_{j-1}) = 0.04082$, and $-1.0 = \beta_3(\tau_0) < \cdots < \beta_3(\tau_{99}) = 1.0$ with $\beta_3(\tau_j) - \beta_2(\tau_{j-1}) = 0.02041$. Thus, the simulated y satisfies the specification in (5) such that

$$q_{\tau_j}(y|x_1, x_2, x_3) = 0.5 + 1.5x_1 + \beta_2(\tau_j)x_2 + \beta_3(\tau_j)x_3, \quad j = 1, \ldots, 99.$$

Note that the above simulation design, though arbitrary, tries to capture different specifications such as regressor correlations and variations in quantile coefficients. We further consider four sample sizes: $n = 200, 500, 1000$, and 5000. For each n, we repeat the data sampling and the two-step estimation procedure 10,000 times. Tables 1, 2, 3 summarize the simulation results. Values shown for each parameter are the average and standard deviation (in parenthesis) across the 10,000 repetitions. For ease of presentation, we only report a coarser grid of nine quantile results from 0.1 to 0.9.

Table 1 shows the simulation results of Step 1 OLS regression. We can see that β_0 and β_1 are consistently estimated with respect to their true values, with shrinking standard deviations when sample size grows. We can also see that the OLS estimates of β_2 and β_3 are consistent with respect to $\sum_j \beta_2(\tau_j)/99 = 0.5$ and $\sum_j \beta_3(\tau_j)/99 = 0.0$ (see Proposition 1).

Table 1 Simulation results of coefficient estimates of the non-discretionary part (OLS regression in Step 1)

Size	β_0	β_1	β_2	β_3
$n = 200$	0.4984	1.5026	0.5002	−0.0001
	(2.7723)	(0.8328)	(0.3815)	(0.7270)
$n = 500$	0.5187	1.4921	0.4976	0.0020
	(1.7283)	(0.5177)	(0.2391)	(0.4589)
$n = 1000$	0.4891	1.4996	0.5017	0.0003
	(1.2450)	(0.3708)	(0.1720)	(0.3225)
$n = 5000$	0.5055	1.4977	0.4993	−0.0003
	(0.5427)	(0.1643)	(0.0750)	(0.1450)
True value	0.5000	1.5000	0.5000	0.0000

The above simulation results are based on 10,000 repetitions. The values shown in the table are average and standard deviation (in parenthesis) of the point estimates across repetitions

Table 2 Simulation results of $\beta_2(\tau)$ in the discretionary part (Quantile regression in Step 2)

Quantiles	True values	$\beta_2(\tau)$			
		$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
0.1	−1.1327	−1.0883	−1.1064	−1.1087	−1.1157
		(0.3736)	(0.2344)	(0.1689)	(0.0737)
0.2	−0.7245	−0.6899	−0.7049	−0.7058	−0.7119
		(0.3896)	(0.2447)	(0.1755)	(0.0772)
0.3	−0.3163	−0.2929	−0.3042	−0.3033	−0.3081
		(0.3988)	(0.2516)	(0.1805)	(0.0790)
0.4	0.0918	0.1031	0.0966	0.0994	0.0953
		(0.4061)	(0.2549)	(0.1833)	(0.0802)
0.5	0.5000	0.5004	0.4977	0.5016	0.4990
		(0.4079)	(0.2555)	(0.1839)	(0.0805)
0.6	0.9082	0.8967	0.8983	(0.9045	0.9031
		(0.4052)	(0.2535)	(0.1829)	(0.0799)
0.7	1.3163	1.2928	1.2992	(1.3066	1.3064
		(0.3991)	(0.2511)	(0.1804)	(0.0787)
0.8	1.7245	1.6904	1.6995	(1.7092	1.7104
		(0.3894)	(0.2437)	(0.1757)	(0.0765)
0.9	2.1327	2.0892	2.1020	2.1120	2.1147
		(0.3728)	(0.2346)	(0.1684)	(0.0736)
Average	0.5000	0.4501	0.4478	0.4515	0.4493
		(0.3460)	(0.2171)	(0.1562)	(0.0682)

The above simulation results are based on 10,000 repetitions. The values shown in the table are average and standard deviation (in parenthesis) of the point estimates across repetitions

Tables 2 and 3 show the Step 2 quantile regression of $\tilde{y} = y - \hat{\beta}_0 - \hat{\beta}_1 x_1$, which is the discretionary part after removing the estimated non-discretionary part from y, on x_2 and x_3. From Table 2, we can see that the average point estimates of $\beta_2(\tau)$ converge to their true values at

Table 3 Simulation results of $\beta_3(\tau)$ in the discretionary part (Quantile regression in Step 2)

Quantiles	True values	$\beta_3(\tau)$			
		$n = 200$	$n = 500$	$n = 1000$	$n = 5000$
0.1	−0.8163	−0.7564	−0.7805	−0.7958	−0.8031
		(0.6811)	(0.4164)	(0.2972)	(0.1296)
0.2	−0.6122	−0.5757	−0.5881	−0.6009	−0.6028
		(0.7664)	(0.4725)	(0.3357)	(0.1477)
0.3	−0.4082	−0.3890	−0.3899	−0.4027	−0.4024
		(0.8181)	(0.5096)	(0.3579)	(0.1598)
0.4	−0.2041	−0.2033	−0.1976	−0.2011	−0.2016
		(0.8500)	(0.5307)	(0.3711)	(0.1648)
0.5	0.0000	−0.0116	0.0007	−0.0012	0.0007
		(0.8583)	(0.5370)	(0.3772)	(0.1655)
0.6	0.2041	0.1842	0.1970	0.2013	0.2014
		(0.8447)	(0.5336)	(0.3748)	(0.1634)
0.7	0.4082	0.3907	0.3993	0.4022	0.4020
		(0.8108)	(0.5118)	(0.3608)	(0.1571)
0.8	0.6122	0.5888	0.5922	0.5998	0.6035
		(0.7593)	(0.4763)	(0.3352)	(0.1454)
0.9	0.8163	0.7691	0.7855	0.7965	0.8055
		(0.6800)	(0.4244)	(0.2962)	(0.1275)
Average	0.0000	-0.0003	0.0019	−0.0002	0.0003
		(0.6098)	(0.3825)	(0.2698)	(0.1189)

The above simulation results are based on 10,000 repetitions. The values shown in the table are average and standard deviation (in parenthesis) of the point estimates across repetitions

different quantile levels when the sample size increases from $n = 200$ to 5000, along with increasing precision (shrinking standard deviations). Intriguingly, the average of $\beta_3(\tau_j)$ across τ_j also consistently estimates the average of true values across quantiles (see the last two rows). Table 3 shows similar simulation results of $\beta_3(\tau)$.

We can see from the above results that our two-step approach works well with the simulated data. The asymptotic behavior of the estimates are in line with Proposition 1. In the next section, we conduct various applications of discretionary accounting studies to further demonstrate the usefulness of our estimating method.

4 Applications

In this section, we apply our proposed method to several existing studies that were conducted based on the traditional method, which uses residual as dependent variable. Specifically, we reconsider all the applications in Chen et al. (2018) and assume the following quantile structure

$$q_\tau(y|X_{non-dis}, X_{dis}, VOI) = \beta_0 + \beta_1 X_{non-dis} + \gamma_0(\tau) + \gamma_1(\tau)VOI + \gamma_2(\tau)X_{dis}, \tag{7}$$

Table 4 Application models

y	Non-discretionary $X_{non-dis}$	Discretionary X_{dis}	Variable of interest $VOI_{i,t}$
Discretionary Accruals			
$TA_{i,t}$	$(1/ASSETS_{i,t-1})$, $(\Delta SALES_{i,t} - \Delta AR_{i,t})$, $PPE_{i,t}$, $INTERACTIONS$	$VOI_{i,t}$, $ROA_{i,t}$, $CFO_{i,t}$, $SIZE_{i,t-1}$, $LEV_{i,t-1}$, $MTB_{i,t-1}$	$BIG\_N_{i,t}$, $SMALL\_PROFIT_{i,t}$, $NOA\_LAG_{i,t}$
Discretionary Expenditures			
$DISC\_EXP_{i,t}$	$(1/ASSETS_{i,t-1})$, $(SALES_{i,t-1}/ASSETS_{i,t-1})$, $INTERACTIONS$	$VOI_{i,t}$, $BIG\_N_{i,t}$, $ROA_{i,t}$, $SIZE_{i,t-1}$, $MTB_{i,t-1}$, $\lvert DISC\_ACC\rvert_{i,t}$	$MTR_{i,t}$, $ZSCORE_{i,t}$
Discretionary Permanent Book-Tax Differences			
$PERMDIFF_{i,t}$	$INTANG_{i,t}$, $UNCON_{i,t}$, $MI_{i,t}$, $CSTE_{i,t}$, $\Delta NOL_{i,t}$, $LAGPERM_{i,t}$, $INTERACTIONS$	$VOI_{i,t}$, $SIZE_{i,t}$, $LEV_{i,t}$, $R\&D_{i,t}$, $CAPEX_{i,t}$, $NOL_{i,t}$	$MNC_{i,t}$, $\lvert DISC\_ACC\rvert_{i,t}$
Abnormal Investment			
$INVEST_{i,t}$	$NEG_{i,t-1}$, $\%REV\_GROWTH_{i,t-1}$, $NEG_{i,t-1} \times \%REV\_GROWTH_{i,t-1}$, $INTERACTIONS$	$VOI_{i,t}$, $SLACK_{i,t}$, $LOSS_{i,t}$, $LEV_{i,t}$, $SIZE_{i,t}$, $TANGIBILITY_{i,t}$	$\lvert DISC\_ACC\rvert_{i,t}$

where $X_{non-dis}$ and X_{dis} are the vectors of non-discretionary part and discretionary part regressors, respectively. Note that the above model is different from the setting in (5) in that we allow varying quantile intercept $\gamma_0(\tau)$ in (7). The varying quantile intercept makes the model more flexible in capturing quantile behaviors. Hence, for the first-step OLS regression, we estimate

$$\hat{y} = (\hat{\beta}_0 + \hat{\gamma}_0) + \hat{\beta}_1 X_{non-dis} + \hat{\gamma}_1 VOI + \hat{\gamma}_2 X_{dis},$$

where $\hat{\beta}_0$ and $\hat{\gamma}_0$ cannot be separately identified. After extracting discretionary $\tilde{y} = y - (\hat{\beta}_0 + \hat{\gamma}_0) - \hat{\beta}_1 X_{non-dis}$, the resulting second-step quantile regression equation is

$$q_\tau(\tilde{y}|X_{dis}, VOI) = (\gamma_0(\tau) - \hat{\gamma}_0) + \gamma_1(\tau)VOI + \gamma_2(\tau)X_{dis},$$

where $\gamma_0(\tau)$ still cannot be separately identified from the random effect of quantile. However, the other parameters $\gamma_1(\tau)$ and $\gamma_2(\tau)$ are consistently estimated.

Our main interest is the effect of the determinant VOI, which is captured by $\gamma_1(\tau)$. For comparison purpose, we follow the data preparation procedure in Chen et al. (2018) and use the same firm-year observations from the Compustat database with fiscal years between 1996 and 2015. Specifically, we exclude financial firms (single-digit SIC code equal to 6) and observations without sufficient data to calculate all the regression variables, and winsorize all variables at the 1st and 99th percentiles by year. Table 4 summarizes all the variables in (7) (i.e. y, $X_{non-dis}$, X_{dis} and, VOI) utilized within the four applications; namely discretionary accruals, discretionary expenditures, discretionary permanent book-tax differences, and abnormal investments. Detailed definitions of all the variables can be found in the Appendix Table 7.

4.1 Discretionary accruals

We consider the three variables of interest; *BIG_N*, *SMALL_PROFIT*, and *NOA_LAG*; that prior literature suggest as the determinants of discretionary accruals. Chen et al. (2018) compare the signs of coefficients and *t*-statistics of the variables of interest between single and traditional two-step regression methods. They find conclusions that are inconsistent with existing studies. We provide here alternative findings with our method in order to have a better understanding of discretionary accruals.

The first variable, *BIG_N*, indicates whether the firm employs a "Big-N" auditor. Employing a "Big-N" auditing firm is widely considered to be a mitigating factor in firms' use of discretionary accruals (see Becker et al. 1998; Francis et al. 1999; Chung et al. 2003; and Dechow et al. 2010). The second variable of interest is an indicator of small profit firms, *SMALL_PROFIT*. Firms with small positive profits are considered to have larger discretionary accruals (see Dechow et al. 2003). The third variable of interest whose association with discretionary accruals has been proved is the lagged value of net operating assets, *NOA_LAG*. *NOA_LAG* proxies for balance sheet slack and is shown as an earnings management constraint (see Barton and Simko 2002; and Zang 2011).

Table 5 reports the estimates of the effects (both single-step OLS estimates and $\gamma_1(\tau)$ in (7)) of the three variables of interests on discretionary accruals. The unconditional quantiles of discretionary accruals $\tilde{y}$ are also reported for reference. For *BIG_N*, the OLS estimate is similar to that in Chen et al. (2018) showing a positive effect on discretionary accruals rather than the expected negative effect. However, the quantile regression results show different directions in the tail quantiles. For the 90% quantile, the estimated effect is −0.008 and significant at 0.01 level. The unconditional quantile of discretionary accrual at the 90% quantile is positive. This implies that at the uppermost quantile level, those firms deviate from the mean on the positive side and *BIG_N* is significantly reducing firms' discretionary accruals, pulling the discretionary accrual back to the mean level. Interestingly, for the 10% quantile, the effect of *BIG_N* on discretionary accruals turns positive, which implies *BIG_N* is getting firms' discretionary accruals less negative at the lowest quantile level where the unconditional $\tilde{y}$ is −0.267. This implies that firms with big auditors are engaging in less income-decreasing discretionary behavior at the 10% level, suggesting that big auditor firms also help to prevent income-decreasing earnings management. The significant positive OLS estimate is due to the impact of the extreme values at the 10% quantile level which shift the OLS estimate to positive. Using quantile regression helps to discern discretionary behavior at different quantile levels. The results above show that it is useful to inspect the effects of *BIG_N* at different quantiles, especially the tail levels where the discretionary behavior is with higher chance to take place.

For *SMALL_PROFIT*, the OLS estimates show no effect on discretionary accruals. If we consider the quantile regression at the 10% level, the effect is significantly negative with value −0.001 while the unconditional quantile of discretionary accrual is negative. This means that at the lowest 10% quantile of discretionary accruals, the discretionary accruals depart further from central quantile if the firm is making a small profit. This is in line with the conventional conjecture of the discretionary behavior in small profit firms. The significant negative effects are further detected at the 20%, 30%, and 40% quantiles where the unconditional quantiles of $\tilde{y}$ are all negative. Results in the lower quantiles from 10% to

Table 5 Determinants of Discretionary Accruals – A Comparison of One-step OLS regression and Two-step Quantile regression (Chen et al. 2018, Table 5)

Variable of interest (expected effect)	OLS	Quantiles								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BIG_N (−)	0.009***	0.001***	−0.000***	−0.001***	−0.002***	−0.001***	−0.002***	−0.002***	−0.003***	−0.008***
t-stat	(5.458)	(7.039)	(−3.390)	(−13.050)	(−17.221)	(−19.211)	(−20.175)	(−22.650)	(−28.842)	(−30.982)
Uncon-Q of $\tilde{y}$		−0.267	−0.139	−0.090	−0.061	−0.039	−0.021	−0.003	0.020	0.070
SMALL_PROFIT (+)	−0.001	−0.001**	−0.001***	−0.001***	−0.001***	−0.000*	−0.000	0.000	0.000	−0.003***
t-stat	(−0.498)	(−2.147)	(−5.228)	(−3.711)	(−2.982)	(−1.910)	(−0.783)	(0.747)	(1.112)	(−4.720)
Uncon-Q of $\tilde{y}$		−0.269	−0.142	−0.093	−0.063	−0.042	−0.024	−0.006	0.018	0.068
NOA_LAG (−)	0.000	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	−0.000***
t-stat	(0.350)	(26.552)	(29.134)	(25.357)	(21.729)	(19.232)	(16.114)	(11.849)	(4.641)	(−5.771)
Uncon-Q of $\tilde{y}$		−0.269	−0.142	−0.092	−0.063	−0.041	−0.023	−0.006	0.018	0.068

This table is a replication of Chen et al. (2018, Table 11) using both single step OLS regression and two-step quantile regression. The following model is estimated: $q_\tau(y|X_{non-dis}, X_{dis}, VOI) = \beta_0 + \beta_1 X_{non-dis} + \gamma_0(\tau) + \gamma_1(\tau)VOI + \gamma_2(\tau)X_{dis}$, where X_{nondis} and X_{dis} are the vectors of non-discretionary part and discretionary part regressors, respectively. *VOI* is the determinant of discretionary accrual; namely, *BIG_N*, *SMALL_PROFIT*, and *NOA_LAG*, repectively. This table only reports coefficient estimates of *VOI*s with *t*-statistics in parentheses. Uncon-Q of $\tilde{y}$ indicates the unconditional quantile of the estimated discretionary accrual. Statistical significance (two-sided) at the 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively

40% show the “cookies jar reserve” behavior of firms.[4] However, the effects disappear from 50% to 80%. Yet at the uppermost 90% level, the effect is significant with a value of −0.003 implying reduction of discretionary accruals. The results of *SMALL_PROFIT* clearly show heterogeneous behaviors of firms at different quantiles of discretionary accruals.

Finally, for *NOA_LAG*, the quantile regressions show significant results in all quantiles but with very small values, whereas the OLS estimate displays no effect. Though the estimated values are very small in magnitude, but the signs in the extreme quantiles, i.e. 10% and 90%, are in line with mitigation effect that shrinks extreme tails to the central. The estimated small values may due to the highly nonlinear partial relation between discretionary accruals and *NOA_LAG*. The correct inference may require a more elaborate nonlinear (or nonparametric) model.

4.2 Discretionary expenditures

Following Chen et al. (2018), we also consider accounting discretionary expenditures; see, for example, Cohen and Zarowin (2010), Cheng et al. (2015), and Lo et al. (2017), among others. We measure real earnings management following the measure developed by Roychowdhury (2006). We adopt two determinants of discretionary expenditures: marginal tax rates (*MTR*) and financial distress indicator (*ZSCORE*) (see Abernathy et al. 2014; Zang 2011).

The marginal tax rate (*MTR*) is defined as firm’s tax loss carryforward, which captures differences in tax burdens across firms (Plesko 2003). Zang (2011) shows that *MTR* has a negative effect on real earnings management. Since an abnormally low level of discretionary expenditure is a sign of real activity manipulations (see Cohen et al. 2008; Cohen and Zarowin 2010; Cheng et al. 2015; Lo et al. 2017), *MTR* is expected to positively affect the firm’s discretionary expenses. The second variable, *ZSCORE*, is proposed in Altman (1968) and measures a firm’s financial distress and health (Oluwo 2007). *ZSCORE* predicts a high chance of bankruptcy when its value is below 1.8, whereas firms with a *ZSCORE* value above 3 are considered unlikely to go bankrupt. If a firms is financially healthy, it is less motivated to cut discretionary expense that benefit the firm from long term. Therefore, we predict that *ZSCORE* should have a positive effect on discretionary expenditures (see Sakaki et al. 2017).

The upper panel in Table 6 shows the coefficient estimates and *t*-statistics of both the one-step single regression and our two-step quantile regression of the two variables of interests on discretionary expense. For *MTR*, the one-step OLS estimate is consistent with our estimation approach in all quantiles. All show a positive and significant effect of MTR on discretionary expense. Moreover, the quantile regression displays an increasing effect from low to high quantile level. Specifically, $\gamma_1(\tau)$ increases from 0.006 at the 10% quantile to 0.057 at 90% quantile. At the upper quantiles from 60 to 90%, the unconditional quantiles of discretionary expenditures are positive. This implies that firms with lower levels of MTR cut discretionary expense more severely, since the cost of engaging in real earnings management is lower. The results above show that it is useful to inspect the *MTR* effect at different quantiles, since the economic magnitude is varying across quantiles.

For *ZSCORE*, we find no significant relationship with discretionary expenditures in the one-step OLS estimation. However, most of the quantile regression results show positive signs for the effect of *ZSCORE* on discretionary expense. The significant positive

[4] Moore (1973) studies the income-reducing discretionary accounting decisions that are often made by a new CEO after a change in management. This is because the reported low earnings may be blamed on the old manager, and the historical bases for future comparison will be reduced. More importantly, the new manager could “release” the income reserve in the future in order to smooth earnings and report a robust trend of increasing earnings.

Table 6 Five researches related to discretionary accounting – A comparison of one-step OLS regression and two-step quantile Regression

Variable of interest (expected effect)	OLS	Quantiles								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Discretionary expenditures										
MTR (+)	0.024***	0.006***	0.019***	0.024***	0.028***	0.030***	0.033***	0.037***	0.043***	0.057***
t-stat	(8.786)	(16.187)	(71.456)	(95.656)	(100.297)	(86.819)	(79.876)	(72.502)	(73.403)	(67.652)
Uncon-Q of $\tilde{y}$		−0.267	−0.196	−0.146	−0.099	−0.037	0.044	0.146	0.294	0.603
ZSCORE (+)	0.000	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.002***	0.000
t-stat	(−0.413)	(36.470)	(46.586)	(49.141)	(54.409)	(43.077)	(36.910)	(26.009)	(14.711)	(1.272)
Uncon-Q of $\tilde{y}$		−0.293	−0.222	−0.173	−0.126	−0.064	0.016	0.119	0.266	0.573
Discretionary permanent book-tax differences										
MNC (+)	0.023***	0.024***	0.016***	0.013***	0.010***	0.010***	0.010***	0.008***	0.004***	−0.006***
t-stat	(3.843)	(9.725)	(10.553)	(13.339)	(13.389)	(14.060)	(14.718)	(10.264)	(4.426)	(−3.191)
Uncon-Q of $\tilde{y}$		−0.368	−0.149	−0.069	−0.039	−0.021	−0.007	0.008	0.030	0.078
$\lvert DISC\_ACC \rvert$ (+)	−1.072***	−1.979***	−1.454***	−1.174***	−0.942***	−0.663***	−0.370***	−0.150***	0.018***	0.331***
t-stat	(−25.409)	(−670.165)	(−665.948)	(−627.414)	(−556.361)	(−479.288)	(−306.299)	(−137.175)	(14.733)	(148.983)
Uncon-Q of $\tilde{y}$		−0.590	−0.350	−0.266	−0.230	−0.210	−0.197	−0.184	−0.167	−0.127
Abnormal investment										
$\lvert DISC\_ACC \rvert$ (+)	0.084***	−0.040***	−0.025***	−0.018***	−0.009***	0.006***	0.031***	0.096***	0.224***	0.392***
t-stat	(7.790)	(−28.521)	(−24.493)	(−16.060)	(−7.820)	(5.034)	(19.269)	(41.338)	(66.140)	(58.877)
Uncon-Q of $\tilde{y}$		−0.020	0.006	0.025	0.044	0.066	0.093	0.132	0.192	0.326

The upper panel reports the estimation results of the effects of *MTR* and *ZSCORE* on *Discretionary Expenditures*. The second panel shows the estimation results of the effects of *MNC* and $|DISC\_ACC|$ on *Discretionary Permanent Book-Tax Differences*. The last panel examines the determinants of *Abnormal Investment*, with $|DISC\_ACC|$ as the variable of interest. This table only reports coefficient estimates of *VOI*s with *t*-statistics in parentheses; other control variable estimates are omitted. Uncon-Q of $\tilde{y}$ indicates the unconditional quantile of the estimated discretionary expenditures, discretionary permanent book-tax differences, and abnormal investment in each panels. Statistical significance (two-sided) at 10%, 5%, and 1% levels are denoted by *, **, and ***, respectively relationships are estimated in most of the quantile regressions, except at 90% quantile. Hence, our findings conform to the expected effect. Overall, the inferences drawn from our proposed two-step approach reveal different discretionary behaviors, which are similar to the results for discretionary accruals in Sect. 4.1.

4.3 Discretionary permanent book-tax differences

In this section, we consider the study of discretionary permanent book-tax differences, which is measured as unexplained permanent book-tax differences (Desai and Dharmapala 2006; and Frank et al. 2009). We adopt two determinants in studying discretionary permanent book-tax differences: (1) an indicator for multinational corporations, *MNC*, which captures foreign operations; and (2) the absolute value of discretionary accruals, |*DISC_ACC*|, which measures financial reporting aggressiveness.

Leblang (1998) asserts that multinational corporations "may have significantly greater opportunities to escape tax with respect to cross-border investments than with respect to strictly domestic investments." Consistently, Rego (2003) finds that multinational firms have superior positions relative to domestic firms in avoiding income tax. Thus, the predicted relationship between *MNC* and discretionary permanent book-tax differences is positive. Our second variable of interest, |*DISC_ACC*|, is expected to positively affect the discretionary permanent book-tax differences (Cazier et al. 2009 and Rego and Wilson 2012).

The inferences of the single-step and our two-step quantile regression approaches are reported in middle panel in Table 6. For *MNC*, we can see that the OLS and quantile regressions all estimate the positive effects of variables of interest on discretionary permanent book-tax differences, except the quantile regression at the 90% level gives a negative effect. The unconditional quantiles of discretionary permanent book-tax differences turn positive only at the 70 to 90% levels. Hence, the *MNC* effects are in line with expected sign only at the 70% and 80% levels, and there are obvious decreasing magnitudes of effects from lower to higher quantile level. Therefore, our method provides significant incremental information about the characteristics of accounting discretionary behavior.

As for |*DISC_ACC*|, the expected effect is positive on discretionary permanent book-tax differences. The OLS estimate shows an opposite sign at the mean level compared to conventional findings. However, the quantile regression results show different effects in different quantile levels. For lower quantiles, the negative estimated effects show that |*DISC_ACC*| makes the discretionary permanent book-tax differences further deviate from the mid-quantile level, as the unconditional quantile of $\tilde{y}$ is negative. The negative effects continue until reaching 70%. However, for the 80% and 90% quantiles, the effects turn positive. Again, our two stage quantile regression method demonstrate its advantage in analyzing the discretionary permanent book-tax differences.

4.4 Abnormal investment

Our final examination is on abnormal investment. Firms with higher financial reporting quality are found to deviate less from expected investment levels (Biddle et al. 2009). Hence, we adopt the absolute value of discretionary accruals (|*DISC_ACC*|) as the measure of financial reporting quality. We predict a positive effect for absolute discretionary accruals on abnormal investment.

In the last panel of Table 6, we see that the OLS estimate shows a significant positive effect of |*DISC_ACC*| on abnormal investment, which is consistent with prior findings.

Moreover, our quantile regression results further support the expected sign. At the lowest quantile of 10%, where the unconditional quantile of abnormal investment is also negative, the significant estimate of −0.04 confirms that $|DISC\_ACC|$ makes the abnormal investment more negative. However, for upper quantiles of 70 to 90% levels, the estimated effects of $|DISC\_ACC|$ are increasing with positive values of 0.096, 0.224, and 0.392, respectively. The dramatic increment demonstrates that severe abnormal investment takes place at the upper quantile levels.

In summary, the results presented in Table 5 and Table 6 clearly support that the discretionary and abnormal behavior varies across quantile levels and, with high probability, occurs in the tail quantile levels. The mean level effect estimated by OLS regression is insufficient to capture the overall quantile spectrum.

5 Conclusion

The study of discretionary behavior, such as discretionary accruals, is one of the most important topics in accounting. However, Chen et al. (2018) and Christodoulou et al. (2018) point out that the traditional two-step approach provides biased estimates and incorrect inferences for discretionary accounting study. The single-step regression proposed by Chen et al. (2018) is a way to eliminate biases and restore correct inference. However, it also shows limitation in revealing the full picture of accounting behavior, since the linear regression only shows the conditional mean relation.

In this paper, we propose a novel two-step approach in which we conduct linear regression in the first step and quantile regression in the second step. The main assumption is that the coefficients of the non-discretionary part of the variable in question are invariant across different quantiles, whereas the coefficients on the discretionary part vary to reflect discretionary behavior.

This method contributes to the measurements of accounting discretions that may be applied in accounting empirical studies ranging from earnings management literature, investment efficiency literature, to tax avoidance literature, etc. For example, a determinant that is positively related to earnings management, on average, as revealed by the OLS method may show heterogeneous effects at different quantile levels. By revealing extreme heterogeneity in the relationship between the accounting discretions and some of its determinants, we demonstrate the inadequacy of the OLS estimates, which capture only the mean relationship between the dependent variable and explanatory variables. Compared to the method proposed by Chen et al. (2018), our estimates reveal the heterogeneous behavior of accounting discretions and reconcile some conflicting findings in prior literature.

We show by simulation that under the above assumption, our two-step approach successfully extracts discretionary part in the first step, and consistently estimates coefficients for the discretionary part through the second step quantile regressions. The empirical applications further support our approach in real accounting data analysis. Overall, our two-step approach provides an alternative to the traditional two-steps linear regression method and the suggested single-step linear regression in Chen et al. (2018).

Appendix

See Tables 7, 8, 9, 10, 11.

Table 7 Key variable used in applications

Variable	Definition
%REV_GROWTH	Lagged revenue (revt) growth as a percentage. The lagged value of %REV_GROWTH is one of the determinants of normal level of investment.
DISC_ACC	Absolute value of discretionary accruals derived using the modified Jones model developed by Dechow, Sloan, and Sweeney (1995). It is used as a control variable when discretionary expenditure is the residual measure.
ΔAR	Total receivables (rect) scaled by lagged total assets (at). It is one of the determinants of non-discretionary accrual and non-discretionary expenditure.
ΔSALES	Total revenues (revt) scaled by lagged total assets (at). It is one of the determinants of non-discretionary accrual.
ASSETS	Total assets (at). The inverse value of ASSET is used as one of the determinants of non-discretionary accrual.
BIG_N	A dummy variable equal to one for firm-years with a Big-N auditor (i.e., if the value of Compustat variable "au" is between 01 and 08.
CAPEX	Capital expenditures (capx) scaled by lagged total assets (at). It is used as control variable when discretionary permanent book-tax difference is the residual measure.
CFO	Operating cash flows (oancf) scaled by lagged total assets (at). It is used as control variable when discretionary accrual is the residual measure.
CSTE	State income taxes (txs) scaled by lagged total assets (at). It is one of the determinants of non-discretionary permanent book-tax differences.
DISC_EXP	Discretionary expenditure is defined as research and development expense (xrd) plus advertising expense (xad) plus selling, general, and administrative expense (xsga) all scaled by lagged total assets (at).
INTANG	Intangible assets (intan) scaled by lagged total assets (at). It is used as control variable when abnormal investment is the residual measure.
INTERACTIONS	A set of year indicator variables and their interactions with each of the independent variables in regression specifications. They are used as the determinants of non-discretionary estimation.
INVEST	Research and development expenses (xrd) plus capital expenditures (capx) plus acquisitions (aqc) minus sales of property, plant, and equipment (sppe) all scaled by lagged total assets (at).
LAGPERM	Lagged value of permanent book-tax differences.
LEV	Leverage is defined as total long-term debt (dltt) plus total debt in current liabilities (dlc) scaled by total assets (at). It is used as a control variable when discretionary permanent book-tax differences and abnormal investment are the residual measure.
LOSS	A dummy variable equal to one for firm-years with negative income before extraordinary items (ib). It is one of the control variables for abnormal level of investment.
MI	Minority interest from the income statement (mii) scaled by lagged total assets (at). It is one of the determinants of non-discretionary permanent book-tax differences.
MNC	A dummy variable equal to one for firm-years with either foreign pretax income (pifo) or foreign income taxes (txfo) greater than zero.
MTB	Market value of equity (csho × prcc_f) divided by total stockholder's equity (seq). It is used as control variable when discretionary accrual and discretionary expenditure are the residual measure.
MTR	Proxy for marginal tax rate, defined as tax loss carryforward (tlcf) scaled by lagged total assets (at).
NEG	An indicator variable equal to one for firm-years with lagged revenue (revt) growth less than zero. The lagged value of NEG is one of the determinants of normal level of investment.
NOA_LAG	Lagged net operating assets is defined as the total stockholder's equity (seq) minus cash and short-term investments (che) plus total long-term debt (dltt) plus total debt in current liabilities (dlc) all scaled by lagged total revenues (revt).

Table 7 (continued)

Variable	Definition
NOL	Tax loss carryforward (tlcf) scaled by lagged total assets (at). The change value of NOL is one of the determinants of non-discretionary permanent book-tax differences.
PERMDIFF	Permanent book-tax differences is defined as pre-tax income (pi) minus federal (txfed) and foreign income taxes (txfo) divided by 0.35 minus deferred income taxes (txdi) divided by 0.35 all scaled by lagged total assets (at).
PPE	Total (net) property, plant, and equipment (ppent) scaled by lagged total assets (at). It is one of the determinants of non-discretionary accrual.
R&D	Research and development expense (xrd) scaled by lagged total assets (at). It is used as control variable when discretionary permanent book-tax difference is the residual measure.
ROA	Income before extraordinary items (ib) scaled by lagged total assets (at). It is used as control variable when discretionary accrual and discretionary expenditure are the residual measure.
SIZE	Natural log of the market value of equity ($csho \times prcc$_f). It is used as control variable when discretionary accrual and discretionary expenditure are the residual measure.
SLACK	Lagged cash and short-term investments (che) scaled by lagged total assets (at). It is used as control variable when abnormal investment is the residual measure.
SMALL_PROFIT	A dummy variable equal to one for firm-years with income before extraordinary items (ib) scaled by lagged market value of equity ($csho \times prcc$_f) between 0 and 0.01.
TA	Income before extraordinary items from the statement of cash flows (ibc) minus operating cash flows (oancf) scaled by lagged total assets (at).
TANGIBILITY	Lagged net property, plant, and equipment (ppent) scaled by lagged total assets (at). It is one of the control variables for abnormal level of investment.
TAX_LOSS	Tax loss carryforward (tlcf) scaled by lagged total assets (at). It is used as control variable when discretionary permanent book-tax difference is the residual measure.
UNCON	Equity in earnings of unconsolidated subsidiaries (esub) scaled by lagged total assets (at). It is one of the determinants of non-discretionary permanent book-tax differences.
ZSCORE	Altman's z-score, a proxy for financial health, defined as $0.3 \times ni$/at+1.0×revt/at+1.4×re/at+1.2×(act - lct)/AT+0.6×(csho×prcc_f)/lt, with all variables be lagged one year.

Table 8 Summary statistics in discretionary accruals study

	Obs.	Mean	SD	Min.	P25	Median	P75	Max.
TA	113404	−0.15	0.53	−7.75	−0.12	−0.06	−0.02	0.62
1/ASSETS_LAG	113404	0.12	0.56	0.00	0.00	0.01	0.03	7.35
ΔSALES - ΔAR	113404	0.08	0.37	−1.99	−0.04	0.04	0.17	2.53
PPE	113404	0.32	0.32	0.00	0.08	0.21	0.47	1.95
BIG_N	113404	0.72	0.45	0.00	0.00	1.00	1.00	1.00
SMALL_PORFIT	113404	0.03	0.17	0.00	0.00	0.00	0.00	1.00
NOA_LAG	113404	1.38	5.56	−38.06	0.28	0.56	1.12	106.56
ROA	113404	−0.18	0.90	−12.57	−0.12	0.02	0.07	0.60
CFO	113404	−0.03	0.43	−4.90	−0.03	0.07	0.13	0.56
SIZE_LAG	113404	5.18	2.50	−1.38	3.39	5.15	6.93	11.56
LEV_LAG	113404	0.29	0.40	0.00	0.02	0.19	0.37	5.82
MTB_LAG	113404	2.80	7.31	−51.54	0.98	1.84	3.43	72.61

Table 9 Summary statistics in discretionary expenditures Study

	Obs.	Mean	SD	Min.	P25	Median	P75	Max.
DISC_EXP	96126	0.53	0.82	0.01	0.14	0.32	0.61	9.91
1/ASSETS_LAG	96126	0.10	0.46	0.00	0.00	0.01	0.03	6.17
SALES_LAG	96126	1.13	0.88	0.01	0.53	0.94	1.48	5.83
MTR	96126	0.75	3.30	0.00	0.00	0.00	0.14	43.92
ZSCORE	96126	2.69	14.94	−140.69	1.21	2.74	5.02	132.18
BIG_N	96126	0.71	0.45	0.00	0.00	1.00	1.00	1.00
ROA	96126	−0.15	0.77	−9.78	−0.10	0.02	0.08	0.61
SIZE_LAG	96126	5.10	2.48	−1.39	3.32	5.06	6.80	11.56
MTB_LAG	96126	2.76	6.85	−50.69	0.96	1.83	3.41	66.22
DISC_ACC	96126	0.15	0.28	0.00	0.03	0.07	0.14	3.32

Table 10 Summary statistics in discretionary permanent book-tax differences study

	Obs.	Mean	SD	Min.	P25	Median	P75	Max.
PERMDIFF	75525	−0.25	1.05	−15.05	−0.14	−0.00	0.02	0.68
INTANG	75525	0.16	0.25	0.00	0.00	0.05	0.23	2.41
UNCON	75525	0.00	0.00	−0.05	0.00	0.00	0.00	0.03
MI	75525	0.00	0.00	−0.06	0.00	0.00	0.00	0.03
CSTE	75525	0.00	0.00	−0.01	0.00	0.00	0.00	0.03
ΔNOL	75525	0.14	0.88	−3.61	0.00	0.00	0.02	13.28
PERMDIFF_LAG	75525	−0.31	1.38	−23.41	−0.14	0.00	0.02	0.69
MNC	75525	0.44	0.50	0.00	0.00	0.00	1.00	1.00
DISC_ACC	75525	0.16	0.34	0.00	0.03	0.07	0.14	3.89
SIZE	75525	5.10	2.58	−1.87	3.25	5.17	6.94	11.43
LEV	75525	0.36	0.77	0.00	0.02	0.20	0.41	9.37
R&D	75525	0.07	0.16	0.00	0.00	0.00	0.08	1.40
CAPEX	75525	0.06	0.09	0.00	0.01	0.03	0.07	0.77
NOL	75525	1.31	6.05	0.00	0.00	0.00	0.25	93.55

Table 11 Summary statistics in abnormal investment study

	Obs.	Mean	SD	Min.	P25	Median	P75	Max.
INVEST	97592	0.16	0.22	−0.09	0.04	0.09	0.19	1.79
NEG_LAG	97592	0.33	0.47	0.00	0.00	0.00	1.00	1.00
%REV_GROWTH_LAG	97592	−0.01	0.66	−7.58	−0.05	0.07	0.20	0.92
DISC_ACC	97592	0.14	0.27	0.00	0.03	0.07	0.13	3.23
SLACK	97592	0.19	0.22	0.00	0.03	0.10	0.27	0.93
LOSS	97592	0.41	0.49	0.00	0.00	0.00	1.00	1.00
LEV	97592	0.28	0.46	0.00	0.03	0.20	0.37	5.75
SIZE	97592	5.28	2.55	−1.39	3.44	5.27	7.06	11.60
TANGIBILITY	97592	0.29	0.25	0.00	0.08	0.21	0.45	0.95

References

Abernathy JL, Beyer B, Rapley ET (2014) Earnings management constraints and classification shifting. J Bus Financ Account 41:600–626
Ahmed MS, Doukas JA (2021) Revisiting disposition effect and momentum: a quantile regression perspective. Rev Quant Financ Acc 56:1087–1128
Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Financ 23:589–609
Armstrong CS, Blouin JL, Jagolinzer AD, Larcker DF (2015) Corporate governance, incentives, and tax avoidance. J Account Econ 60:1–17
Barton J, Simko PJ (2002) The balance sheet as an earnings management constraint. Account Rev 77:1–27
Beaver WH (1987) The properties of sequential regressions with multiple explanatory variables. Account Rev 62:137–144
Becker CL, DeFond ML, Jiambalvo J, Subramanyam KR (1998) The effect of audit quality on earnings management. Contemp Account Res 15:1–24
Biddle GC, Hilary G, Verdi RS (2009) How does financial reporting quality relate to investment efficiency? J Account Econ 48:112–131
Cazier RA, Rego SO, Tian XS, Wilson RJ (2009) Early evidence on the determinants of unrecognized tax benefits. Ryan J., Early Evidence on the Determinants of Unrecognized Tax Benefits (September 14, 2009)
Chen W, Hribar P, Melessa S (2018) Incorrect inferences when using residuals as dependent variables. J Acc Res 56:751–796
Chen Y, Ge R, Louis H, Zolotoy L (2019) Stock liquidity and corporate tax avoidance. Rev Acc Stud 24:309–340
Cheng Q, Lee J, Shevlin T (2015) Internal governance and real earnings management. Account Rev 91:1051–1085
Christodoulou D, Ma L, Vasnev A (2018) Inference-in-residuals as an estimation method for earnings management. Abacus 54:154–180
Chung R, Firth M, Kim J-B (2003) Auditor conservatism and reported earnings. Account Bus Res 33:19–32
Cohen DA, Dey A, Lys TZ (2008) Real and accrual-based earnings management in the pre-and post-Sarbanes-Oxley periods. Account Rev 83:757–787
Cohen DA, Zarowin P (2010) Accrual-based and real earnings management activities around seasoned equity offerings. J Account Econ 50:2–19
Dechow P, Ge W, Schrand C (2010) Understanding earnings quality: a review of the proxies, their determinants and their consequences. J Account Econ 50:344–401
Dechow PM, Richardson SA, Tuna I (2003) Why are earnings kinky? An examination of the earnings management explanation. Rev Acc Stud 8:355–384
Desai MA, Dharmapala D (2006) Corporate tax avoidance and high-powered incentives. J Financ Econ 79:145–179
Du D, Ng P, Zhao X (2013) Measuring currency exposure with quantile regression. Rev Quant Financ Acc 41:549–566
Du D, Zhao X (2017) Financial investor sentiment and the boom/bust in oil prices during 2003–2008. Rev Quant Financ Acc 48:331–361
Du Q, Shen R (2018) Peer performance and earnings management. J Bank Financ 89:125–137
Francis JR, Maydew EL, Sparks HC (1999) The role of Big 6 auditors in the credible reporting of accruals. Auditing J Pract Theory 18:17–34
Frank MM, Lynch LJ, Rego SO (2009) Tax reporting aggressiveness and its relation to aggressive financial reporting. Account Rev 84:467–496
Horrace WC, Reddic WD (2014) "Managerial liability coverage and earnings manipulation: exploiting the Sarbanes-Oxley Act," Working Paper
Huang AYH (2013) Value at risk estimation by quantile regression and kernel estimator. Rev Quant Financ Acc 41:225–251
Huang K, Lao B, McPhee G (2017) Does stock liquidity affect accrual-based earnings management? J Bus Financ Account 44:417–447
Koenker R (2005) Quantile regression. Econometric society monographs. Cambridge University Press, Cambridge
Koenker R, Bassett G (1978) Regression quantiles. Econometrica pp 33–50
Koenker R, Hallock KF (2001) Quantile regression. J Econ Perspect 15:143–156

Lai Y, Lin W, Kuo L (2018) Forestalling capital regulation or masking financial weakness? Evidence from loss reserve management in the property-liability insurance industry. Rev Quant Financ Acc 50:481–518

Leblang S (1998) International double nontaxation. Tax Notes Int 134:181–3

Lee CF (2020) Financial econometrics, mathematics, statistics, and financial technology: an overall view. Rev Quant Financ Acc 54:1529–1578

Lo K, Ramos F, Rogo R (2017) Earnings management and annual report readability. J Account Econ 63:1–25

McKee G, Kagan A (2016) Determinants of recent structural change for small asset US credit unions. Rev Quant Financ Acc 47:775–795

Moore ML (1973) Management changes and discretionary accounting decisions. J Account Res 11:100–107

Oluwo M (2007) Strategic use of financial ratio to prevent bankruptcy: a study of opportunity for business enterprises. Capella University, Minneapolis

Plesko GA (2003) An evaluation of alternative measures of corporate tax rates. J Account Econ 35:201–226

Rego SO (2003) Tax-avoidance activities of US multinational corporations. Contemp Account Res 20:805–833

Rego SO, Wilson R (2012) Equity risk incentives and corporate tax aggressiveness. J Account Res 50:775–810

Roychowdhury S (2006) Earnings management through real activities manipulation. J Account Econ 42:335–370

Sakaki H, Jackson D, Jory S (2017) Institutional ownership stability and real earnings management. Rev Quant Financ Acc 49:227–244

Shi G, Sun J, Luo R (2015) Geographic dispersion and earnings management. J Acc Public Policy 34:490–508

Su ED (2021) Testing stock market contagion properties between large and small stock markets. Rev Quant Financ Acc 57:147–202

Zang AY (2011) Evidence on the trade-off between real activities manipulation and accrual-based earnings management. Account Rev 87:675–703

Authors and Affiliations

May Huaxi Zhang[1] · Stanley Iat-Meng Ko[2] · Andreas Karathanasopoulos[3] · Chia Chun Lo[4]

May Huaxi Zhang
zhanghuaxi@bit.edu.cn

Andreas Karathanasopoulos
akarathanasopoulos@ud.ac.ae

Chia Chun Lo
clo5@gsu.edu

[1] Department of Accounting, School of Management and Economics, Beijing Institute of Technology, Beijing, China

[2] Graduate School of Economics and Management, Tohoku University, Sendai, Japan

[3] Dubai Business School, University of Dubai, Dubai, UAE

[4] J. Mack Robinson College of Business, Georgia State University, Atlanta, GA, US