Available online at www.sciencedirect.com

ScienceDirect

Review of Development Finance 5 (2015) 24–33

Review of DEVELOPMENT FINANCE

www.elsevier.com/locate/rdf

# Cash flow forecast for South African firms

Yun Li [a], Luiz Moutinho [b], Kwaku K. Opong [b], Yang Pang [b,*]

[a] *School of Engineering, University of Glasgow, University Avenue, G12 8QQ, United Kingdom*
[b] *Adam Smith Business School, University of Glasgow, University Avenue, G12 8QQ, United Kingdom*

**Abstract**

This paper applies models in the extant literature that have been used to forecast operating cash flows to predict the cash flows of South African firms listed on the Johannesburg Stock Exchange. Out-of-sample performance is examined for each model and compared between them. The reported results show that some accrual terms, i.e. depreciation and changes in inventory do not enhance cash flow prediction for the average South African firm in contrast to the reported results of studies in USA and Australia. Inclusion of more explanatory variables does not necessarily improve the models, according to the out-of-sample results. The paper proposes the application of moving average model in panel data, and vector regressive model for multi-period-ahead prediction of cash flows for South Africa firms.


*JEL classification:* G300; M410; M490

*Keywords:* Cash flow; Moving average; Predictive models; Accruals; Vector autoregressive model; Depreciation

## 1. Introduction

Given that cash flow is the life-blood of a firm, accurate determination of cash flows enables firms to make important financial decisions that relate to whether the firm survives or goes bankrupt. As a measure of a firm's profitability and financial health, cash flows could provide potential clues about the source company's ability to pay divided and thus attract investors' interest too. There are three categories of cash flows recorded in statement of cash flow, i.e. cash flows from operation, financing and investment, of which operating cash flow, reflecting the ability of the firm to engage in day-to-day operations and its continuity in business, is of the most importance. For the managers of firms, investors or analysts, prediction of future cash flows are of extreme usefulness and value.

There are two issues to be considered when attempting to predict a firm's cash flows. First, the variables those are useful and informative to cash forecast need to be identified and incorporated into the forecast model. Secondly, the type and structure of models to be employed in the forecast should be carefully chosen to provide a more accurate prediction. This study shed light on both issues, intending to demonstrate the procedure of choosing variables and models for more accurate prediction. There are a number of difficulties with cash flow prediction. Generally speaking, cash flow is more volatile than earnings and thus harder to predict. There is no uniform cash flow generating process for the whole business world and different companies provide distinct patterns of cash flows. Besides, due to the popularity of credit trade, a firm's revenue and expenses are not equal to cash inflow and outflow and this compounds the problem of accurate cash flow prediction. Academic studies on cash flow prediction rely on public information as reflected in a firm's financial statement for cash flow data. Among the variables that have been found usefulness in cash flow

* Corresponding author.
*E-mail addresses:* Yun.Li@glasgow.ac.uk (Y. Li), Luiz.Moutinho@glasgow.ac.uk (L. Moutinho), Kwaku.Opong@Glasgow.ac.uk (K.K. Opong), Y.Pang.1@research.gla.ac.uk (Y. Pang).
Peer review under responsibility of Africagrowth Institute.

http://dx.doi.org/10.1016/j.rdf.2014.11.001
1879-9337/

prediction include earnings, accrual terms such as changes in account receivable and payable and disaggregated cash components. Empirical studies suggest that these variables are useful and informative in predicting cash flows. In this paper, a comparison is made between different sets of variables as predictors. It is expected that the more predictors that are included, the better a model will perform because more inclusion of variables often means richer exploitation of information. A second comparison is made between models which include explanatory variables with different lags. It is expected that models with more lagged explanatory variables could provide more accurate prediction whereas the reported results in this paper suggest otherwise. This paper proposes two types of cash flow modeling, i.e. moving average model and vector autoregressive (VAR) model for cash flow prediction. Moving average model also makes economic sense as it measures how an unexpected cash flow shock could influence people making future prediction. The moving average model is applied to one-period-ahead prediction and VAR model is proposed for multi-period-ahead prediction. For this purpose VAR is more powerful and relies less on data availability than linear regression. These models are applied empirically on data of South Africa firms.

This paper is organized as follows: Section 1 provides the introduction to this paper. Section 2 reviews the literature and discusses factors that influence the prediction of cash flows and the prediction models utilized in the study. Section 3 describes the data for South African firms. Section 4 reports the results of the empirical analysis and the conclusion of the study is provided in Section 5.

## 2. Literature review

Cash flow forecast is of interest to investors, creditors, employees and rating agencies among others. Investors are interested in cash flows as input into their investment models to enable them to decide on payoff relating to dividends and capital appreciation of their investments. Creditors are interested in solvency decisions relating to the firms they transact business with and employees are interested in job security and going-concern issues relating to firms they work for. Rating agencies are also interested in going-concern and a firm's ability to pay its debts when they are due.

Cash flow can be considered as complimentary information to earnings since combinative analysis of both quantities might bring better results than analyzing earnings on its own. Earnings, also sometimes referred to as net income, are the summation of net cash income and net credit income, the latter of which is based on credit trades with customers and is not yet but expected to be settled by cash in a later period. The amount of credit given to customers could potentially be overlooked without cash flow information and this may mislead investors about the risk relating to shortage of cash in the firm. In addition, cash flow directly measures the operational ability of the firm to meet its day-to-day financial commitments. In conventional finance theory, the worth of a firm is theoretically equal to the discounted value of all cash flows generated during the firm's life assuming that all the cash flows are paid out as dividend. As a result, news about cash flow can potentially have significant impact on a firm's market price. Along with earnings forecast, analysts are increasingly including cash flow forecast into their analysis and reports. Dechow, Kothari, and Watts (1998) [DWK], proposed a model of cash flow which they derived from sales and reached a conclusion that current earnings are the best forecast of future cash flow. Earnings equal to cash flow plus accruals that include changes in account payable, changes in account receivable, changes in inventory, depreciation and amortization and others. In the DKW model, accruals, for simplicity, include only changes in account receivable, changes in inventory and changes in account payable, which are equivalent to changes in working capital while long term accruals such as depreciation are not considered. The DKW model makes several strict assumptions about sales process and working capital components and their derived model relies heavily on those assumptions. Barth, Cram, and Nelson (2001) [BCN], proposed a modified version of the DKW model. They disaggregate the accruals into components, anticipating them to have different persistence in predicting future cash flow. Lorek and Willinger (2010) compared the predictive accuracy of the BCN model, in time-series and cross-sectional analysis respectively, and found that time-series model generates more accurate result. This result is not surprising since cross-sectional estimation treats all firms as homogeneous, which is hardly true in reality. The DKW and BCN models use the indirect method to measure cash flow, i.e. they calculate cash flow component from net income and adjust the results with accrual terms. In the USA, statement of Financial Accounting Standards (SFAS) No. 95 issued in 1988 allowed the disclosure of direct method cash flow statement. Therefore, cash flows after 1988 are directly available from the cash flow statement. Cheng and Hollie (2008) partition cash flow components into core and non-core ones, and analyzed their persistence for future cash flow determination. The study defines core cash flow components as cash flows from: sales, cost of goods sold, and operating and administrative expenses. The non-core cash flow components are interest, taxes, and others. When these regressors are applied in a prediction model, the adjusted $R^2$ is slightly greater than the BCN version. Similarly, in Orpurt and Zang (2009), the issue of whether direct method cash flow statement enhances cash flow modeling is examined. Their reported cash flow forecast model had adjusted $R^2$ of 43%, although their sample was smaller (compared to pre-1989 studies when SFAS No.95 had not been published). Orpurt and Zang's paper examined whether it makes a difference in estimating cash components using indirect method compared to using disclosed items directly from the statement of cash flow. They do not directly compare the accuracy of forecast models. Instead, they examine the statistical significance of articulation error that is defined as the difference between estimated cash components and disclosed ones in their regression model. Their reported results suggest that the coefficients of articulation error terms are statistically significant and thus that articulation errors have incremental information for cash flow forecast. It hence implies that the direct method for cash flows disclosure is more informative in predicting future cash flow than indirect method statement of cash flow. The direct method is not applicable for South African firms; therefore in this paper cash disaggregation will be based on the indirect method described in Orpurt and Zang's paper.

Several researchers have examined factors that influence the accuracy of cash flow prediction. For example, Brochet et al. (2008) and Ogneva (2012) suggest that accrual quality relates to cash flow shocks and discretionary accruals might reduce accruals contribution to cash flow prediction. They argue that positive accruals and/or high cash flow volatility are associated with higher predictive ability of accruals. Some researchers suggest a firms' past poor performance tends to be improved by managers, which could lead to higher predicted future cash flow (Francis and Eason, 2012) and that bold forecasters might have better private information, so they can provide more accurate prediction than others (Pae and Yoon, 2012).

Discretionary accruals may be income increasing or income decreasing, depending on managers motivations. Badertscher et al. (2009) argue in detail about the impact of earning management, including discretionary accruals and real operation management on cash flow prediction. They argue that opportunistic behavior on the part of managers can lead to the use of discretionary accruals to hide the firm's true situation as a means for management to maintain their equity value. Therefore, manipulated statements used as inputs would reduce their ability for cash flow prediction. On the other hand, management may use discretionary accruals to signal their true view of the firm's future. In such a situation, a management's statements of discretionary accruals could have a better predictive ability of future cash flow. Therefore, whether discretionary accruals can enhance or adversely affect cash flow forecast depends on the motivation behind the manipulation, if any, by managers.

When managers are motivated to manipulate non-discretionary information to suit their own selfish ends, non-discretionary information could be revealing or masked, making the variables more or less informative in forecasting. A number of researchers have suggested methods to identify discretionary accruals [see Jones (1991), Dechow et al. (1995) and Kothari et al. (2005) among others]. Apart from such discretionary problem, accrual would result in varying coefficient on relevant variables. Dechow and Ge (2006) show that under different accrual level, firms cash flows have different persistence and the correlations between cash flows and accruals also depends on accrual level. Another option for Managers to distort financial statements in order to meet reporting goals is through real activities such as those discussed in Roychowdhury (2006). Such operations include price discounts, reduction of discretionary expenditures and so on.

This study makes a number of contributions. First, the scant number of studies on cash flow studies in developing economies, suggest the need for more studies in the area as the results elsewhere are not directly transferable due to different business and operational environments. Secondly, application of models based on studies elsewhere may not suit a developing country setting. Thirdly, given the nature and importance of cash flows for firms, investors, creditors and policy makers, among others, the current study is a major addition to the extant literature.

### *2.1. Review of cash flow models and discussion of models used in this study*

Dechow et al. (1998) derived cash flow forecast from the basic relationship:

$$CF_t = (SALE_t - \Delta AR_t) - (P_t - \Delta AP_t), \quad (1)$$

where *CF* denotes cash flow from operation, *SALE* denotes sales, $\Delta AR$ denotes changes in account receivable, *P* denotes purchase and $\Delta AP$ denotes changes in account payable. Terms in the first parenthesis represents cash inflow while the second parenthesis representing cash outflow. With several assumptions on firm's operating activities, DKW model suggests that the expectation of future cash flows is current earnings:

$$E_t[CF_{t+1}] = EARN_t, \quad (2)$$

Barth et al. (2001) argued that earnings, instead of being used directly as a predictor, should be disaggregated into cash flow and several accrual terms as the components provide different predictive power for future cash flow suggested by their theoretical model below:

$$E_t[CF_{t+1}] = CF_t + [1 - (1-\beta)\gamma_1\gamma_2(1-\pi)\alpha^{-1}]\Delta AR_t + (1-\beta)\Delta INV_t - \Delta AP_t, \quad (3)$$

where $\Delta INV$ denotes changes in inventory and other terms are as defined previously. Compared to DKW model, the parameters on the components of earnings are allowed to differ. As BCN model is less restrictive, it outperforms the DKW model in cash flow prediction (see Barth et al., 2001).

Beyond these accrual terms, Cheng and Hollie (2008) and Orpurt and Zang (2009) disaggregate cash flow into core and non-core components in order to examine the persistence of these components in cash flow prediction. In Cheng and Hollie (2008), the model is specified as:

$$\begin{aligned} CF_{t+1} = \alpha &+ \beta(C\_SALE_t + C\_COG_t + C\_OE_t + C\_INT_t \\ &+ C\_TAX_t + C\_OTHERS_t) + \beta(\Delta AR_t + \Delta INV_t \\ &+ \Delta AP_t + DEPR_t + AMORT_t \\ &+ OTHERS_t) + \varepsilon_{t+1}, \end{aligned} \quad (4)$$

where the terms in the first parenthesis are cash received from sales, cash from cost of goods sold, cash related to operating and administrative expenses, cash used to pay tax, cash for interest payment, and other cash components, respectively. The first three are defined as core cash components and the latter three are non-core components. The cash flow disaggregation explores additional information that could be obtained from cash flow statement under direct method disclosure. Cheng and Hollie thus use disaggregated cash components and also accrual terms defined in BCN model as predictors for cash flow. The empirical results suggest that such cash flow disaggregation does improve the accuracy of BCN model, however the effect is minor. In Cheng and Hollie's paper, model (4) has a reported $R^2$ of 39.83% and BCN model (with cash flow un-disaggregated) has reported $R^2$ of 38.49%. In U.S., before the publication of SFAS No.95, statement of cash flow was not compulsory, so the cash flow from operation needed to be estimated using balance sheet and income statement. Orpurt and Zang (2009) also suggest that direct disclosure of cash components provides incremental effect on cash flow prediction. They introduce a concept of articulation error, which is defined as the difference between cash components estimated using balance sheet and income statement and ones that are actually disclosed (direct method) in statement of cash flow. Some studies report that the direct method cash flow statement is more informative in predicting cash flow (see Arthur and Chuang, 2006 and Farshadfar and Monem, 2013 in Australian setting). For firms that do not provide direct method cash disclosure, such incremental benefit cannot be exploited but cash disaggregation can still be done. In this paper, we model South African cash flows by disaggregating cash flow following Orpurt and Zang (2009):

$$\begin{aligned} CF_t &= CFIN_t - CFOUT_t - INT_t - TAX_t \\ &= (SALE_t - \Delta AR_t) - (COG_t + OE_t + \Delta INV_t - \Delta AP_t) \\ &\quad - INT_t - TAX_t + AE_t, \end{aligned} \tag{5}$$

where net operating cash flow is equal to cash inflow (*CFIN)* minus cash outflow (*CFOUT* which is paid to suppliers and employees) and interest and tax cash payment (*INT* and *TAX* respectively). The first and second brackets in the second line of Eq. (5) measures the estimated cash inflow and outflow respectively, using information in balance sheet and income statement; interest and tax payment in cash are also included, which are directly disclosed independent of the report format. The difference between disclosed actual cash flow from operation and these four components is termed as articulation error (*AE*), shown as the final term in the equation.

This paper starts with a comparison of two empirical cash flow prediction models following Orpurt and Zang (2009):

$$\begin{aligned} \text{Model I}: \quad CF_{t+1} &= \beta_0 + \beta_1 CF_t + \beta_2 \Delta AR_t + \beta_3 \Delta INV_t \\ &\quad + \beta_4 \Delta AP_t + \beta_5 DDA_t \\ &\quad + \beta_6 INT_t + \beta_7 TAX_t + \varepsilon_{t+1}, \end{aligned} \tag{6}$$

$$\begin{aligned} \text{Model II}: \quad CF_{t+1} &= \beta_0 + \beta_1 CFIN_t + \beta_2 CFOUT_t + \beta_3 AE_t \\ &\quad + \beta_4 DDA_t + \beta_5 \Delta AR_t + \beta_6 \Delta INV_t \\ &\quad + \beta_7 \Delta AP_t + \beta_8 INT_t + \beta_9 TAX_t + \varepsilon_{t+1}, \end{aligned} \tag{7}$$

All the variables are the same as previously defined. $\beta$s are parameters to be estimated. Model I is the empirical BCN model with the addition of interest and tax payment as predictors. Model II further disaggregates the cash flow term in model I. Cash inflow and cash outflow are estimated based on Eq. (5). Model II in principle is superior to model I in two ways: disaggregating cash flow incorporates information of other accounting items such as sales and allows cash components to have different persistence.

Apart from above mentioned linear regression models, this paper also introduce autoregressive moving average model with exogenous variables (ARMAX) developed by Whittle (1951) (see Box et al., 2008 for details). This is a time series model and this paper attempt to adapt it to panel data application. The general model ARMA(1,1) with exogenous variables is written as:

$$CF_{t+1} = \beta_0 + \beta_1 CF_t + \mathbf{BX_t} + \beta_2 \varepsilon_t + \varepsilon_{t+1}, \tag{8}$$

where $\varepsilon$ denotes the moving average (MA) terms and $\mathbf{X}$ denotes exogenous variables. The moving average terms depends on the model parameters, therefore traditional OLS method does not apply and we use maximum likelihood estimators (MLE) for parameter estimation.

In practical application, a multi-period-ahead prediction of cash flow may be of greater importance than a single period. The above mentioned models are simple to adapt to such requirement. The usual way is to adjust the lag length between target variable and its predictive variables. For one- period-ahead prediction, as discussed above, the explanatory variables take one lag values. Thus, considering two-period-ahead prediction for example, the explanatory variables should take values of their second lag. The disadvantage of such method appears critical for long term forecast where data covering a long period is demanded but not always available. Moreover, when the historical data is too outdated, the information they contain for prediction purposes may be limited. To deal with this problem, we apply vector autoregressive (VAR) model (Sims, 1980). VAR model is presented as (take original BCN model as example):

$$\begin{bmatrix} CF_t \\ \Delta AR_t \\ \Delta AP_t \\ \Delta INV_t \end{bmatrix} = \mathbf{A} + \mathbf{B} \begin{bmatrix} CF_{t-1} \\ \Delta AR_{t-1} \\ \Delta AP_{t-1} \\ \Delta INV_{t-1} \end{bmatrix} + \mathbf{e}_t, \tag{9}$$

where $\mathbf{A}$ and $\mathbf{e}$ are $4 \times 1$ vectors of constants and error terms respectively, $\mathbf{B}$ is a $4 \times 4$ parameter matrix. Therefore VAR could capture the evolvement of not only cash flow but also other predictors through time. Once the parameters $\mathbf{A}$ and $\mathbf{B}$ are estimated, we could apply the VAR model of lag one to forecast as many periods ahead as needed. The forecast equation is written as:

$$\begin{bmatrix} 1 \\ \widehat{CF}_{t+p} \\ \Delta \widehat{AR}_{t+p} \\ \Delta \widehat{AP}_{t+p} \\ \Delta \widehat{INV}_{t+p} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{A} & \mathbf{B} \end{bmatrix}^p \begin{bmatrix} 1 \\ CF_t \\ \Delta AR_t \\ \Delta AP_t \\ \Delta INV_t \end{bmatrix}, \tag{10}$$

where $\mathbf{0}$ is a $1 \times 4$ vector of zeros; p is the number of periods ahead to be forecast.

In the later empirical section, we will start from comparison between model I and model II examining the incremental power of cash flow disaggregation in cash flow prediction. Also examined is the effect of including two lags of predictors on models with inclusion of only one lag of predictors. Then we introduce ARMAX model. With inclusion of moving average (MA) term, we could in the first place see whether this model provide more accurate forecast and secondly show the direction of cash flow shock's effect on next period cash flow prediction. Finally, we apply VAR model to predict two-year-ahead cash flow, in comparison with regression model with cash flow as the only dependent variable.

## 3. Data

This study uses data for South African firms collected from Datastream. All firms listed on Johannesburg Stock Exchange are included in the study. Cash flow data before 1994 is not available, so the sample spans from 1994 to 2012. The following items are gathered for each firm: revenue from sales (*SALE*), cost of goods sold (*COST*), selling, general and administrative expenses (*SGA*), depreciation, depletion and amortization (*DDA*), account payable (*AP*), account receivable (*AR*), inventory (*INV*), net operating cash flow (*CF*), interest paid in cash (*INT*), and tax paid in cash (*TAX*), 11 variables in total. Changes in account payables, receivables and inventory, cash inflow, cash outflow and articulation error are not provided directly so they are calculated manually. Selling, general and administrative expenses (SGA) are used to represent other expenses (*OE*) in Eq. (5). All variables are deflated by average total assets for each firm in order to smooth the potential effect of firm sizes. Firms that do not have available data for all these variables are excluded. The whole sample contains 192 firms with 1021 firm-year observations. The panel is unbalanced, i.e. firms have unequal number of observations. For regression analysis, all the explanatory variables are lagged; therefore the final sample is further reduced. For models with one lag period of explanatory variables, the sample contains 791 firm-year observations in total and there are 623 firm-year observations for model estimation with two lag periods of explanatory variables.

## 4. Empirical results

### 4.1. Descriptive statistics

Table 1 provides the descriptive statistics for the sample data. It is worth noting that average sales is 1.857 times average total assets. The standard deviation of sales is 1.748, suggesting that the sample is highly dispersed. Net operating cash flow deflated by average total assets has a mean of 0.127. Eq. (5) is used to estimate cash inflow and outflow and articulation error. The estimated net operating cash flow has a relatively low correlation with the disclosed actual value and thus cash flow estimated indirectly might cause some noise in cash flow forecast.

Table 1
Descriptive statistics.

| Variables | Mean | Median | Std. dev. |
|---|---|---|---|
| *SALE* | 1.857 | 1.279 | 1.748 |
| *COST* | 1.361 | 0.834 | 1.545 |
| *DDA* | 0.053 | 0.044 | 0.044 |
| $\Delta AP$ | 0.015 | 0.011 | 0.109 |
| $\Delta AR$ | 0.022 | 0.015 | 0.143 |
| $\Delta INV$ | 0.019 | 0.010 | 0.076 |
| *INT* | 0.026 | 0.019 | 0.027 |
| *TAX* | 0.046 | 0.032 | 0.052 |
| *SGA* | 0.253 | 0.193 | 0.234 |
| *CF* | 0.127 | 0.115 | 0.175 |
| *CFIN* | 1.835 | 1.280 | 1.741 |
| *CFOUT* | 1.598 | 1.056 | 1.641 |
| *AE* | −0.019 | 0.002 | 0.220 |
| *AE*1 | −0.091 | −0.058 | 0.227 |
| *AE*2 | −0.065 | −0.032 | 0.228 |
| *AE*3 | −0.045 | −0.017 | 0.219 |

*Note*: All variables are deflated by average total assets of each firm. *CFIN* denotes Cash inflow, *CFOUT* for cash outflow, and AE is short for articulation error. These three terms are calculated by Eq. (6). Note there are three extra '*AE*'s. Due to the fact that not all firms report interest and/or tax paid in cash in their operative cash flow section, Eq. (6) is not accurate for some firms. I examined three more calculations of net operating cash flow: 1. *CFIN–CFOUT*; 2. *CFIN–CFOUT–INT*; 3. *CFIN–CFOUT*–TAX, assuming every firm must lie in one of the four groups. Estimated cash flow using method 1 has the highest correlation with actual value (0.69) whereas the one deducting both interest and tax cash payment has the lowest correlation among the four (0.65). However, note that the mean of articulation error is much lower when Eq. (6) is adopted. Therefore, Eq. (6) is still the default formula for articulation error.

### 4.2. Linear regression models

In this study, all firms are treated as homogeneous and the models described in previous section are estimated using pooled methods. The sample is partitioned into estimation period from 1995 to 2009, and out-of-sample test period from 2010 to 2012. The estimation sample contains 579 firm-year observations and the test sample 212. The parameters estimated in-sample are applied out-of-sample for performance comparison. $R^2$ of different models are then compared to determine the best predicting model.

Model I is set as the bench mark model. Model II disaggregates cash into different components. The results are shown in Table 2. Parameters with asterisk indicate a 5% level of significance. Numbers in parenthesis are *t*-statistics for each parameter. With regards to the bench mark model I, lagged cash flow, changes in account receivable, changes in account payable, and tax cash payment as well as the intercept terms are statistically significant, whereas lagged depreciation, depletion and amortization (DDA), changes in inventory, and interest cash payment are not statistically significant. This is not consistent with the reported results of BCN model using U.S data and Farshadfar and Monem (2013) of Australian data. As non-cash item, there is no direct explanation why DDA should be useful in cash flow forecast. Changes in inventory and interest payment are items that directly relate to cash flow because they measure proportions of cash outflow. For these three items, South African data shows much weaker predictive power than studies for other countries. $R^2$ is 0.54 for model I, which is much higher than that reported in studies on U.S data. For example, in Cheng and Hollie (2008), none of the empirical models report $R^2$ of more than 40%. It therefore implies that South African firms have less variable cash flows than U.S. firms. Model II has higher adjusted $R^2$ as cash disaggregation incorporates extra information from balance sheet and income statement. However, the out-of-sample test shows that model II provides less accurate prediction than model I. The additional information does not

Table 2
Regression results for one period lag models.

| | | Model I | Model II | Model III | Model IV | Model V | Model VI | Model VII |
|---|---|---|---|---|---|---|---|---|
| *Lagged variables* | | | | | | | | |
| Estimation period | | | | | | | | |
| $CF_{t-1}$ | | 0.5805* | | 0.3213* | 0.3579* | 0.3400* | 0.3121* | 0.3761* |
| | | (17.53) | | (6.24) | (7.29) | (6.66) | (6) | (7.74) |
| $CFIN_{t-1}$ | | | 0.5932* | | | | | |
| | | | (18.5) | | | | | |
| $CFOUT_{t-1}$ | | | −0.5892* | | | | | |
| | | | (−18.37) | | | | | |
| $AE_{t-1}$ | | | 0.3254* | | | | | |
| | | | (6.33) | | | | | |
| $SALE_{t-1}$ | | | | 0.2824* | 0.2574* | 0.2626* | 0.3471* | 0.2380* |
| | | | | (6.51) | (6.11) | (6.15) | (8.64) | (5.75) |
| $COST_{t-1}$ | | | | −0.2820* | −0.2574* | −0.2617* | −0.3471* | −0.2376* |
| | | | | (−6.34) | (−5.95) | (−5.99) | (−8.41) | (−5.59) |
| $SGA_{t-1}$ | | | | −0.2454* | −0.2258* | −0.2277* | −0.2867* | −0.2084* |
| | | | | (−5.51) | (−5.15) | (−5.17) | (−6.58) | (−4.81) |
| $DDA_{t-1}$ | | 0.0211 | −0.1075 | −0.127 | −0.1133 | −0.1789 | −0.1145 | −0.1647 |
| | | (0.17) | (−0.86) | (−1.01) | (−0.90) | (−1.45) | (−0.90) | (−1.33) |
| $\Delta AP_{t-1}$ | | −0.6398* | −0.6943* | −0.4222* | −0.5061* | −0.4350* | −0.4086* | −0.5179* |
| | | (−12.50) | (−13.85) | (−7.01) | (−10.61) | (−7.23) | (−6.72) | (−10.89) |
| $\Delta AR_{t-1}$ | | 0.4471* | 0.5854* | 0.3154* | 0.3208* | 0.3176* | 0.3019* | 0.3230* |
| | | (12.14) | (14.11) | (7.66) | (7.78) | (7.69) | (7.28) | (7.8) |
| $\Delta INV_{t-1}$ | | −0.0321 | −0.1680* | −0.1757* | | −0.1739* | −0.1463 | |
| | | (−0.42) | (−2.17) | (−2.27) | | (−2.24) | (−1.88) | |
| $INT_{t-1}$ | | −0.1981 | −0.7766* | −0.4616* | −0.4570* | | −0.5593* | |
| | | (−0.96) | (−3.85) | (−2.27) | (−2.24) | | (−2.74) | |
| $TAX_{t-1}$ | | 0.8715* | 0.2076 | 0.4943* | 0.4631* | 0.5339* | | 0.5026* |
| | | (7.03) | (1.5) | (3.68) | (3.46) | (4) | | (3.77) |
| Constant | | 0.0305* | 0.0260* | 0.0227* | 0.0235* | 0.0153 | 0.0264* | 0.0162 |
| | | (3.53) | (2.92) | (2.45) | (2.53) | (1.76) | (2.84) | (1.85) |
| NoB | 579 | | | | | | | |
| Adj. $R^2$ | | 0.5407 | 0.5715 | 0.572 | 0.5688 | 0.5688 | 0.5625 | 0.5658 |
| Test period | | | | | | | | |
| NoB | 212 | | | | | | | |
| $R^2$ | | 0.5476 | 0.4593 | 0.4594 | 0.4896 | 0.4638 | 0.4028 | 0.4937 |

*Note*: Asterisks indicates significance at levels of 5%. Numbers in parentheses are corresponding t statistics.
Model I: $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 \Delta AR_{t-1} + \beta_3 \Delta INV_{t-1} + \beta_4 \Delta AP_{t-1} + \beta_5 DDA_{t-1} + \beta_6 INX_{t-1} + \beta_7 TAX_{t-1} + \varepsilon_t$.
Model II: $CF_t = \beta_0 + \beta_1 CFIN_{t-1} + \beta_2 CFOUT_{t-1} + \beta_3 AE_{t-1} + \beta_4 DDA_{t-1} + \beta_5 \Delta AR_{t-1} + \beta_6 \Delta INV_{t-1} + \beta_7 \Delta AP_{t-1} + \beta_8 INX_{t-1} + \beta_9 TAX_{t-1} + \varepsilon_t$.
Model III: $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 SALE_{t-1} + \beta_3 COST_{t-1} + \beta_4 SGA_{t-1} + \beta_5 DDA_{t-1} + \beta_6 \Delta AR_{t-1} + \beta_7 \Delta INV_{t-1} + \beta_8 \Delta AP_{t-1} + \beta_9 INX_{t-1} + \beta_{10} TAX_{t-1} + \varepsilon_t$.
Model IV: $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 SALE_{t-1} + \beta_3 COST_{t-1} + \beta_4 SGA_{t-1} + \beta_5 DDA_{t-1} + \beta_6 \Delta AR_{t-1} + \beta_7 \Delta AP_{t-1} + \beta_8 INX_{t-1} + \beta_9 TAX_{t-1} + \varepsilon_t$.
Model V: $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 SALE_{t-1} + \beta_3 COST_{t-1} + \beta_4 SGA_{t-1} + \beta_5 DDA_{t-1} + \beta_6 \Delta AR_{t-1} + \beta_7 \Delta AP_{t-1} + \beta_8 INV_{t-1} + \beta_9 TAX_{t-1} + \varepsilon_t$.
Model VI: $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 SALE_{t-1} + \beta_3 COST_{t-1} + \beta_4 SGA_{t-1} + \beta_5 DDA_{t-1} + \beta_6 \Delta AR_{t-1} + \beta_7 \Delta AP_{t-1} + \beta_8 INV_{t-1} + \beta_9 INX_{t-1} + \varepsilon_t$.
Model VII: $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 SALE_{t-1} + \beta_3 COST_{t-1} + \beta_4 SGA_{t-1} + \beta_5 DDA_{t-1} + \beta_6 \Delta AR_{t-1} + \beta_7 \Delta AP_{t-1} + \beta_8 TAX_{t-1} + \varepsilon_t$.

improve the model's predictive power. Secondly, it is noteworthy that changes in inventory and interest payment are significant while tax payment is insignificant in model II. Model II appears to provide different inferences than model I. To further examine this problem, we develop model III by replacing the cash disaggregated components with revenue from sales, cost of goods sold and selling, general and administrative expenses (SGA):

$$\text{Model III}: \quad CF_{t+1} = \beta_0 + \beta_1 CF_t + \beta_2 SALE_t + \beta_3 COST_t + \beta_4 SGA_t + \beta_5 DDA_t + \beta_6 \Delta AR_t + \beta_7 \Delta INV_t + \beta_8 \Delta AP_t + \beta_9 INT_t + \beta_{10} TAX_t + \varepsilon_{t+1}. \quad (11)$$

Models III uses the same information as model II, so their in-sample fitness and out-of-sample performance are almost the same. Models III nonetheless suggests that tax payment is significant, with changes in inventory and interest payment also significant. From model I to III, we see that parameters of changes in inventory, interest payment and tax payment are not consistent. This could be due to collinearity between the explanatory variables included in the models. Models IV to VI that in turn remove changes in inventory, interest payment, tax payment from model III one at a time are then built to examine the effects of these excluded variables. When changes in inventory are removed as in model IV, the other two variables are both significant. The out-of-sample performance is highly enhanced. Therefore, changes in inventory actually deteriorate the models when we include variables from income statement. The case

Table 3
Comparison of sum squared errors of model I and VII with different division of sample.

| | Model I | Model VII |
|---|---|---|
| *a* | | |
| 2010 | 1.3278 | **1.2382** |
| 2011 | **1.8676** | 1.8734 |
| 2012 | **0.7927** | 1.3513 |
| *b* | | |
| 2011 | 1.8449 | **1.8136** |
| 2012 | **0.8023** | 1.3532 |

*Note*: the numbers in the table are sum squared errors of out-of-sample prediction for model I and VII. Bold numbers denotes they are the smaller of the two models during different test year, suggesting the model is better than the other in that particular test year.

is similar for interest payment, which is removed in model V. The $R^2$ in test period also increase, though not as remarkably as model IV. When tax payment is removed, model VI reports declined out-of-sample $R^2$, suggesting that tax payment could be actually providing incremental information, without which the model's predictive performance is harmed. In addition, changes in inventory in model VI become insignificant being consistent with model I. From the comparison of these 6 models, it can be concluded that changes in inventory and interest payment are not very informative in cash flow forecasting, and even worse, we may significantly reduce the power of the models by adding extra variables from income statement. As a result, for models including sales, cost, and other expenses, it is beneficial to remove changes in inventory and interest payment from the set of explanatory variables. This gives model VII:

$$\text{Model VII}: \quad CF_{t+1} = \beta_0 + \beta_1 CF_t + \beta_2 SALE_t + \beta_3 COST_t + \beta_4 SGA_t + \beta_5 DDA_t + \beta_6 \Delta AR_t + \beta_7 \Delta AP_t + \beta_8 TAX_t + \varepsilon_{t+1}. \qquad (12)$$

The estimation results are also reported in Table 2. The signs of parameters are consistent with expectation for cost, expenses, and changes in account payables as they are negatively related with net cash flow. Note that selling, general and administrative expenses' effect on one-period-ahead cash flow is of slightly lower magnitude than cost of goods sold. Model VII has higher out-of-sample $R^2$ than model II to VI, validating the point that removing changes in inventory and interest payment when sales, cost, and SGA are present provides superior model. However the $R^2$ is still lower than that of model I, despite the addition of extra information. The test period sample contains three years of data, 2010, 2011 and 2012. Model VII, due to its inclusion of more variables, should have stronger predictive power, at least during the year that immediately follows the end of estimation sample. Therefore, model VII is expected to outperform model I at least during 2010. To examine this point, sum squared errors (SSE) are calculated for each year in test period, which are shown in Table 3. In Table 3 section *a* are the sum of squared errors for out-of-sample prediction. Lower SSE means higher accuracy, and are displayed in bold. The results show that model VII provides a better performance at the start of the prediction period but perform less well in later years. This phenomenon could be due to the changing environment through time, which may result in little persistence of the estimated parameters. If this were the case, the parameters would be better to be updated year by year. In attempt to argue this point, the sample is then further partitioned with the period 1995–2010 as estimation period and 2011 and 2012 as test period. The SSE for 2011 and 2012 are recalculated and the results are shown in Table 3 section *b*. Similar to the reported results in panel *a*, model VII does perform better in 2011 but worse in 2012. The benchmark model I is simpler and more persistently effective than model VII while model VII could provide better one-period-ahead prediction if estimated with updated data.

Prior studies focus mainly on one lag period of predictors, but it is natural to ask whether further past explanatory variables contain potential information that are incremental to cash flow prediction. Models I and VII are separately estimated twice, the first time with one lag of predictors and the second time with two lags, so we could compare the difference made by inclusion of one more lag period of predictors. It is expected that models with two lags of independent variables should perform no worse than their one-lag counterparts. The results are shown in Table 4. The sample size is further reduced because two lags of data are required. Table 4 shows some significant differences to the reported results in Table 2. For example, for one lag version of model I, changes in inventory and interest payment become significant. It can be seen that model VII that excludes the two variables has lower in-sample fit based on $R^2$ for the one lag version. Additional variables added to model VII from the income statement become insignificant. The out-of-sample $R^2$ for the four models are close. Model VII underperforms model I, no matter how many lags are included. However, it is noteworthy that for model I, out-of-sample performance of two-lag model is slightly worse than one-lag model while for model VII, two-lag model is marginally better than one-lag model. Sum squared errors for different years are also compared in Table 5. The difference between model I and VII has declined remarkably. Model VII outperforms model I for 2010, the start of the prediction period, which is consistent with the results reported in Table 3 but under-performs for the following two years. In addition, for model I, two-lag model is better than one-lag model in 2010, but it is the opposite for 2011 and 2012, while for model VII, the two-lag model is generally better than one-lag model except in 2012. It can be told from these comparisons that for near future prediction, model VII is a better model, especially when two lags of explanatory variables are included.

In summary, this section examines different combinations of predictors in one-period-ahead cash flow forecasting. The benchmark model uses least variables and its performance is consistent in sample and out of sample. DDA, changes in inventory and interest payment are insignificant, which is different from studies in other countries. Models incorporating income statement information seem to make worse out-of-sample prediction. However, when two lags of independent variables are included in prediction models, results differ in several ways which suggests that variable inclusion in cash flow modeling should be treated with care.

Table 4
Comparison of model I and VII with different lags of independent variables.

| | | Model I, one lag | | Model I, two lags | | Model VII, one lag | | Model VII, two lags | |
|---|---|---|---|---|---|---|---|---|---|
| *Lagged variables* | | | | | | | | | |
| Estimation period | | | | | | | | | |
| $CF_{t-1}$ | | 0.5341* | (12.38) | 0.6017* | (12.37) | 0.4535* | (8.95) | 0.5082* | (9.9) |
| $t-2$ | | | | −0.0776 | (−1.68) | | | 0.1187* | (2.15) |
| $SALE_{t-1}$ | | | | | | 0.0924 | (1.95) | 0.2673* | (5.13) |
| $t-2$ | | | | | | | | −0.2894* | (−6.01) |
| $COST_{t-1}$ | | | | | | −0.0878 | (−1.81) | −0.2914* | (−5.22) |
| $t-2$ | | | | | | | | 0.3204* | (6.01) |
| $SGA_{t-1}$ | | | | | | −0.0678 | (−1.38) | −0.2599* | (−4.12) |
| $t-2$ | | | | | | | | 0.2798* | (4.51) |
| $DDA_{t-1}$ | | 0.1573 | (1.31) | 0.3249 | (1.44) | 0.0618 | (0.5) | 0.4529* | (1.98) |
| $t-2$ | | | | −0.1407 | (−0.51) | | | −0.2853 | (−1.05) |
| $\Delta AP_{t-1}$ | | −0.5009* | (−7.59) | −0.4177* | (−5.95) | −0.4241* | (−7.20) | −0.2869* | (−4.12) |
| $t-2$ | | | | 0.1586* | (2.72) | | | 0.1208* | (2.23) |
| $\Delta AR_{t-1}$ | | 0.2518* | (4.03) | 0.2433* | (3.9) | 0.2500* | (3.93) | 0.1468* | (2.3) |
| $t-2$ | | | | −0.0516 | (−0.73) | | | −0.0095 | (−0.14) |
| $\Delta INV_{t-1}$ | | 0.1903* | (2.24) | 0.1122 | (1.26) | | | | |
| $t-2$ | | | | 0.093 | (1.07) | | | | |
| $INT_{t-1}$ | | −0.4123* | (−2.08) | −0.5607 | (−1.63) | | | | |
| $t-2$ | | | | 0.2443 | (0.61) | | | | |
| $TAX_{t-1}$ | | 0.7451* | (5.6) | 0.6377* | (3.69) | 0.6240* | (3.99) | 0.3524* | (1.98) |
| $t-2$ | | | | 0.1084 | (0.66) | | | 0.2516 | (1.59) |
| Intercept | | 0.0438* | (4.98) | 0.0393* | (4.36) | 0.0250* | (2.67) | 0.0252* | (2.74) |
| NOB | 429 | | | | | | | | |
| $R^2$ | | 0.5321 | | 0.549 | | 0.5308 | | 0.586 | |
| Test period | | | | | | | | | |
| NoB | 194 | | | | | | | | |
| $R^2$ | | 0.5595 | | 0.556 | | 0.5416 | | 0.5456 | |

*Note*: Asterisks indicates significance at levels of 5%. Numbers in parentheses are corresponding $t$ statistics.
L1 and L2 denotes the variable of first lag (time $t-1$) and second lag (time $t-2$) respectively. Model I and model VII are compared when different periods of lagged predictors are applied.

Table 5
Sum squared errors for model I and VII with different lags of independent variables.

| | Model ITwo lags | Model IOne lag | Model VIITwo lags | Model VIIOne lag |
|---|---|---|---|---|
| 2010 | 1.2208 | 1.2287 | **1.2149** | 1.2183 |
| 2011 | 1.8515 | **1.8291** | 1.9303 | 1.9698 |
| 2012 | 0.7527 | **0.7367** | 0.7695 | 0.7608 |

*Note*: the numbers in the table are sum squared errors of out-of-sample prediction for model I and VII. Bold numbers denotes they are the smaller of the two models during different test year, suggesting the model is better than the other in that particular test year.

### 4.3. ARMAX models

ARMAX model explores the persistence of both the main variable and the error of the model over time. It seems that the linear regression model described above can be treated as first order autoregressive model with exogenous variables, or simply ARX(1,1). One-period-lagged cash flow is the first order autoregressive variable and the other accounting variables are treated as exogenous variables that take a first order lagged value. For model I and model VII, we include a first order moving average term. The two models for comparison are specified as below:

where $\varepsilon_t$ denotes the moving average term, and its coefficient will measure its relationship with future cash flow. Intuitively speaking, the moving average term is the unexpected shock for last period cash flow. Therefore, its parameters measures how shocks influence next period forecast. We use maximum likelihood estimation (MLE) to estimate the parameters. This method calculates the joint likelihood for all observations, and the estimator finds the optimum values for the coefficients that can maximize the likelihood. Assume the conditional expectation

$$\text{Model I with MA(1)}: \quad CF_{t+1} = \beta_0 + \beta_1 CF_t + \beta_2 \Delta AR_t + \beta_3 \Delta INV_t + \beta_4 \Delta AP_t + \beta_5 DDA_t + \beta_6 INT_t + \beta_7 TAX_t + \beta_8 \varepsilon_t + \varepsilon_{t+1}, \tag{13}$$

$$\text{Model VII with MA(1)}: \quad CF_{t+1} = \beta_0 + \beta_1 CF_t + \beta_2 SALE_t + \beta_3 COST_t + \beta_4 SGA_t + \beta_5 DDA_t + \beta_6 \Delta AR_t + \beta_7 \Delta AP_t + \beta_8 TAX_t + \beta_9 \varepsilon_t + \varepsilon_{t+1}, \tag{14}$$

Table 6
Comparison of linear regression model and ARMAX(1,1) model for model I and model VII.

| | | Model I | Model I with MA | Model VII | Model VII with MA |
|---|---|---|---|---|---|
| *Lagged variables* | | | | | |
| Estimation period | | | | | |
| $CF_{t-1}$ | | 0.5805* | 0.5491* | 0.3761* | 0.2132* |
| $SALE_{t-1}$ | | | | 0.2380* | 0.3235* |
| | | | | (5.75) | (6.93) |
| $COST_{t-1}$ | | | | −0.2376* | −0.3242* |
| | | | | (−5.59) | (−6.76) |
| $SGA_{t-1}$ | | | | −0.2084* | −0.2911* |
| | | | | (−4.81) | (−5.96) |
| $DDA_{t-1}$ | | 0.0211 | 0.0561 | −0.1647 | −0.116 |
| | | (0.17) | (0.42) | (−1.33) | (−0.84) |
| $\Delta AP_{t-1}$ | | −0.6398* | −0.6513* | −0.5179* | −0.4874* |
| | | (−12.50) | (−13.03) | (−10.89) | (−10.70) |
| $\Delta AR_{t-1}$ | | 0.4471* | 0.4340* | 0.3230* | 0.2444* |
| | | (12.14) | (11.51) | (7.8) | (5.47) |
| $\Delta INV_{t-1}$ | | −0.0321 | −0.0317 | | |
| | | (−0.42) | (−0.42) | | |
| $INT_{t-1}$ | | −0.1981 | −0.1896 | | |
| | | (−0.96) | (−0.89) | | |
| $TAX_{t-1}$ | | 0.8715* | 0.8956* | 0.5026* | 0.4677* |
| | | (7.03) | (7.04) | (3.77) | (3.39) |
| $\varepsilon_{t-1}$ | | | 0.0903 | | 0.2558* |
| | | | (1.58) | | (4.39) |
| Constant | | 0.0305* | 0.0313* | 0.0162 | 0.0167 |
| | | (3.53) | (3.46) | (1.85) | (1.71) |
| NoB | 579 | | | | |
| $R^2$ | | 0.5462 | 0.5482 | 0.5718 | 0.5854 |
| Test period | | | | | |
| NoB | 212 | | | | |
| $R^2$ | | 0.5476 | 0.5413 | 0.4937 | 0.4344 |

*Note*: Asterisks indicates significance at levels of 5%. Numbers in parentheses are corresponding *t* statistics.
Model I with MA(1): $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 \Delta AR_{t-1} + \beta_3 \Delta INV_{t-1} + \beta_4 \Delta AP_{t-1} + \beta_5 DDA_{t-1} + \beta_6 INX_{t-1} + \beta_7 TAX_{t-1} + \beta_8 \varepsilon_{t-1} + \varepsilon_t$.
Model VII with MA(1): $CF_t = \beta_0 + \beta_1 CF_{t-1} + \beta_2 SALE_{t-1} + \beta_3 COST_{t-1} + \beta_4 SGA_{t-1} + \beta_5 DDA_{t-1} + \beta_6 \Delta AR_{t-1} + \beta_7 \Delta AP_{t-1} + \beta_8 TAX_{t-1} + \beta_9 \varepsilon_{t-1} + \varepsilon_t$.

of cash flow is i.i.d by Gaussian distribution with constant variance $\sigma^2$. The coefficients obtained are reported in Table 6. When the moving average (MA) term is introduced, parameters in model I have not changed much, and the MA term is of low value and insignificant. Out-of-sample test for model I shows that the addition of MA term has done no good to the predictive power and has rather had a negative effect. The MA term in model VII is statistically significant. The positive sign implies that cash flow shocks of one period tend to persist in the cash flow of the next period. However the addition of MA term to model VII provides a worse out-of-sample performance $R^2$ of 0.43 compared to 0.49 without MA term.

### 4.4. VAR model for two-period-ahead prediction

In the previous two sections, models are developed for one-period-ahead cash flow prediction. The fundamental idea is to find the relationship between cash flow and one-period-lagged explanatory variables. Model I and model VII, as stated previously, can be treated as autoregressive model with exogenous variable. VAR model says that all explanatory variables can be considered as a vector of endogenous variables that depend on vectors of their lagged values. In this way, the whole vector can be forecast recursively. In this section, lag-two regression and VAR form of model I and model VII are examined in two-period-ahead prediction. The estimation period is from 1996 to 2009, and test period is 2011 and 2012. There is a reason for not testing performance in 2010. When 2010 cash is to be forecast, either lag 2 model or VAR model requires data in 2008 as input. For VAR model, 2008 data first is used to predict 2009, and then we use the 2009 prediction to further predict 2010. Recall that data from 2008 to 2009 is already used in parameters estimation hence VAR model for 2010 prediction is not purely out of the sample, and the comparison would have misleading results. With this concern, year 2010 are excluded from comparison. The main concern of this section is to compare out-of-sample predictive power of different models, so only the comparison of sum squared errors for each year are listed whereas regression results are not reported.

Table 7 shows SSE generated by each model during 2011 and 2012 separately. The result is obvious that VAR model for model I has outperformed the other three models in both years. For both model I and model VII, regression model and VAR model have very similar results in 2011, but VAR form is much better in year 2012 prediction. It is simple and straightforward

Table 7
Comparison of sum squared errors for model I and VII in two- period-ahead prediction.

| | Model ILag 2 | Model IVAR | Model VIILag 2 | Model VIIVAR |
|---|---|---|---|---|
| 2011 | 3.0268 | **3.0037** | 3.0891 | 3.0936 |
| 2012 | 1.3632 | **1.2168** | 1.4668 | 1.2258 |

*Note*: the numbers in the table are sum squared errors of out-of-sample prediction for model I and VII. Bold numbers denotes they are the smaller of the two models during different test year, suggesting the model is better than the other in that particular test year.

to extend the prediction to longer periods, and the result from this test is very encouraging for further studies.

## 5. Conclusion

Previous studies on cash flow prediction mainly focus on the information provided in financial statements. Studies such as Barth et al. (2001) discussed the power of accrual terms and how disaggregating accruals into its major components enhances cash flow prediction. This study applies a number of models to cash flow data for South African firms. Three main cash flow models are investigated in this study, mainly, linear regression, which has been widely adopted in similar and relevant studies, moving average model, which is mostly applied in time-series analysis, and vector autoregressive model, that has been widely applied in macroeconomics and finance. The latter two types of models are applied for the first time to cash flow prediction.

The results reported in this study contrast that reported elsewhere. Disaggregating cash flows into its major components does not appear to enhance cash flow prediction for the average South African firms compared to results reported by Barth et al. (2001) for USA and Farshadfar and Monem (2013) for Australian firms. The results suggest that implications of studies conducted elsewhere cannot be extrapolated across other countries without taking into account country context and differences. The reported results in this study show that models incorporating income statement information seem to result in worse out-of-sample prediction. However, when two lags of independent variables are included in prediction models, results differ in several ways which suggests that variable inclusion in cash flow modeling in South Africa should be treated with caution. In addition, prediction accuracy, as measured by $R^2$ for South African firms are high compared to extant studies elsewhere. Studies on cash flow prediction pool all firms' data together and ignore heterogeneity that exists among firms and industry. There is therefore the need for studies on cash flow prediction that focus on industry and individual firms.

## Acknowledgement

The authors appreciate the comments by the participants at African Accounting and Finance Association 3rd Annual Conference, 2013, Uganda and the financial support by Kelvin Smith scholarship from University of Glasgow.

## References

Arthur, N., Chuang, C., 2006. Ias 7 Alternative Methods of Disclosing Cash Flow from Operations: Evidence on the Usefulness of Direct Method Cash Flow Disclosures. Working Paper. The University of Sydney.

Badertscher, B.A., Phillips, J.D., Pincus, M., Rego, S.O., 2009. Earnings management strategies and the trade-off between tax benefits and detection risk: to conform or not to conform? Account. Rev. 84, 63–97.

Barth, M., Cram, D., Nelson, K., 2001. Accruals and the prediction of future cash flows. Account. Rev. 76, 27–58.

Box, G.E., Jenkins, G.M., Reinsel, G.C., 2008. Time Series Analysis: Forecasting and Control. John Wiley & Sons, Hoboken, New Jersey.

Brochet, F., Nam, S., Ronen, J., 2008. The Role of Accruals in Predicting Future Cash Flows and Stock Returns. Working Paper, Available at SSRN 1126022.

Cheng, C., Hollie, D., 2008. Do core and non-core cash flows from operations persist differentially in predicting future cash flows? Rev. Quant. Finance Account. 31, 29–53.

Dechow, P., Kothari, S., Watts, R.L., 1998. The relation between earnings and cash flows. J. Account. Econ. 25, 133–168.

Dechow, P.M., Ge, W., 2006. The persistence of earnings and cash flows and the role of special items: implications for the accrual anomaly. Rev. Account. Stud. 11, 253–296.

Dechow, P.M., Sloan, R.G., Sweeney, A.P., 1995. Detecting earnings management. Account. Rev. 70, 193–225.

Farshadfar, S., Monem, R., 2013. Further evidence on the usefulness of direct method cash flow components for forecasting future cash flows. Int. J. Account. 48, 111–133.

Francis, R.N., Eason, P., 2012. Accruals and the naïve out-of-sample prediction of operating cash flow. Adv. Account. 28, 226–234.

Jones, J.J., 1991. Earnings management during import relief investigations. J. Account. Res. 29, 193–228.

Kothari, S.P., Leone, A.J., Wasley, C.E., 2005. Performance matched discretionary accrual measures. J. Account. Econ. 39, 163–197.

Lorek, K.S., Willinger, G.L., 2010. Time series versus cross-sectionally derived predictions of future cash flows. Adv. Account. 26, 29–36.

Ogneva, M., 2012. Accrual quality, realized returns, and expected returns: the importance of controlling for cash flow shocks. Account. Rev. 87, 1415–1444.

Orpurt, S., Zang, Y., 2009. Do direct cash flow disclosures help predict future operating cash flows and earnings? Account. Rev. 84, 893–935.

Pae, J., Yoon, S.-S., 2012. Determinants of analysts cash flow forecast accuracy. J. Account. Audit. Finance 27, 123–144.

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

Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48, 1–48.

Whittle, P., 1951. Hypothesis testing in time series analysis. Hafner Publishing Company, New York.