A PANEL DATA ANALYSIS OF THE RELATIONSHIP BETWEEN CORPORATE SOCIAL
RESPONSIBILITY AND EARNINGS MANAGEMENT: EVIDENCE FROM IRAN

(Recibido el 05-07-2017. Aprobado el 06-09-2017)

Ashkan Khosropour
Department of Mathematic,
Firoozabad Branch, Islamic
Azad University, Firoozabad,
Iran.

Abstract. In this paper we present the results of a panel data analysis of the relationship between corporate social responsibility and earnings management. We used a fixed effect regression model with Driscoll and Kraay standard errors to account for the possible problem of heteroskedastic and autocorrelated error structure. The main conclusion is that in the studied period the Romanian economy was characterized by a strong correlation between productivity and remuneration, which indicates the economic efficiency.

Keywords: corporate social responsibility, earnings management, econometric model, panel data, fixed effects, robust standard errors

1. INTRODUCTION

Earnings management has recently received considerable attention both from regulators and the popular press. Earnings management can be defined as the alteration of firms’ reported economics performance by insiders to either mislead some stakeholders or to influence contractual outcomes (Healy and Wahlen, 1999; Leuz et al., 2003). Managers may be inclined to manage earnings due to the existence of explicit and implicit contracts, the firm’s relation with capital markets, the need for external financing, the political and regulatory environment or several other specific circumstances (Vander Bauwhede, 2001). These deliberate managerial actions, contrived to disguise the real value of a firm’s assets, transactions, or financial position, have negative consequences for shareholders, employees, the communities in which firms work, society at large, and managers’ reputations, job security and careers (Zahra et al., 2005). Accounting earnings are more reliable and more informative when managers’ opportunistic behavior is controlled through a variety of monitoring systems (Dechow et al., 1996). After several recent financial scandals, there has been an international trend toward developing and implementing corporate governance mechanisms to fight against the opportunistic behaviors that have undermined investors’ credibility in financial information. Corporate governance attributes help investors by aligning the interests of managers with the interests of shareholders and by enhancing the reliability of financial information and the integrity of the financial reporting process (Watts and Zimmerman, 1986).

Decades of empirical research have focused on the factors influencing the quality of earnings, specifically the accruals. However, there is also increasing attention being paid to the managerial activities which can lead to the manipulation of earnings. Previous studies document mixed results regarding the association between corporate social responsibility (CSR) and transparent financial reporting. In this article, we examine the relationship between CSR and earnings quality by using Spanish firms from 2005 to 2012. The earnings quality is measured by using the absolute value of abnormal discretionary accruals from the modified Jones model.

CSR is related to ethical and moral aspects about corporate decision-making and behavior and, as such, addresses complex issues like environmental protection, human resources management, health and safety at work, local community relations, and relationships with suppliers and customers (Castelo and Lima, 2006). Engaging in socially responsible activities not only improves stakeholder satisfaction, but also has a positive effect on corporate reputation (Orlitzky et al., 2003) and reduces the financial risk incurred by the firm (Orlitzky and Benjamin, 2001).

CSR research has employed a variety of theories and methodologies to study the potential relationship between CSR activities and other traditional measures of a firm’s success (Mahoney and Roberts, 2007, p. 234). Previous studies focus on the link between CSR and economic or financial firm’s performance (Moore, 2001; Orlitzky et al., 2003; Brammer et al., 2007) and the evidence is mixed. As Jorgensen and Knudsen (2006) note, this relationship represents the most questioned area of CSR (Angelidis et al., 2008); while a lot of research points in favor of a mild positive relationship (Aupperle et al., 1985; McGuire et al., 1988; Orlitzky et al., 2003; Maron, 2006; Wu, 2006; Rodgers et al., 2013) this connection has not been fully established (Neville et al., 2005; Prado-Lorenzo et al., 2008; Park and Lee, 2009) and the mechanisms through which financial performance is enhanced by CSR is not well understood (Jawahar and McLaughlin, 2001; Doh et al., 2009).

The literature review suggests there remains a lack of understanding about how CSR initiatives can influence on accounting quality by reducing earnings management. Accordingly, our study fills this gap by studying the effects of CSR practices on discretionary accruals. The country’s legal system, economic development, the importance of stock markets and ownership concentration all affect the country’s accounting standards, which in turn affect the country’s quality of financial reporting. We use Spanish data because they generally reflect an institutional setting similar to most continental
A panel data regression differs from a regular time-series or cross-section regression in that it has a double subscript on its variables: $y_{it} = a + X_{it}'b + u_{it}$, $i = 1, ..., N$; $t = 1, ..., T$ (1) The $i$ subscript denotes the cross-section dimension and $t$ denotes the time-series dimension. Most of the panel data application utilize a one-way error component model for the disturbances, with: $u_{it} = \alpha_i + \varepsilon_{it}$ (Badi H. Baltagi, 2008).

There are several different linear models for panel data. The fundamental distinction is that between fixed-effects and random-effects models. In the fixed-effects (FE) model, the $\alpha_i$ are permitted to be correlated with the regressors $x_{it}$, while continuing to assume that $x_{it}$ is uncorrelated with the idiosyncratic error $\varepsilon_{it}$. In the random-effects (RE) model, it is assumed that $\alpha_i$ is purely random, a stronger assumption implying that $\alpha_i$ is uncorrelated with the regressors (A. Colin Cameron and P.K. Trivedi, 2009).

2.1 Test for poolability of the data

One of the main motivations behind pooling a time series of cross-sections is to widen the database in order to get better and more reliable estimates of the parameters of the model. The question is to pool or not to pool the data. The simplest poolability test has its null hypothesis the OLS model: $y_{it} = a + b'X_{it} + \varepsilon_{it}$ and as its alternative the FE model: $y_{it} = a + b'X_{it} + \alpha_i + \varepsilon_{it}$ (Robert M. Kunst, 2009).

In other words, we test for the presence of individual effects. Formally, we write $H_0: \alpha_i = 0$, $i = 1, ..., N$. We consider the $F$ statistics according to the construction principle:

$$F_{1\text{-way}} = \frac{(ESS_U - ESS_R)/(N-K)}{ESS_R/((T-1)N-K)}$$

(2)

where $ESS_R$ denotes the residual sum of squares under the null hypothesis, $ESS_U$ the residual sum of squares under the alternative. Under $H_0$, the statistic $F_{1\text{-way}}$ is distributed as $F$ with $(N-1, (T-1)N-K)$ degrees of freedom. The two sums of squares evolve as intermediate results from OLS and from FE estimation.

2.2 The Hausman test

The Hausman principle can be applied to all hypothesis testing problems, in which two different estimators are available, the first of which $\hat{b}$ is
efficient under the null hypothesis, however inconsistent under the alternative, while the other estimator $\tilde{b}$ is consistent under both hypotheses, possibly without attaining efficiency under any hypothesis. Hausman had the intuitive idea to construct a test statistic based on $q = b - \tilde{b}$. Because of the consistency of both estimators under the null, this difference will converge to zero, while it fails to converge under the alternative. Hausman suggested the statistic $m = q^\prime (\text{var } q)^{-1} q$, where $\text{var } q = \text{var } \tilde{b} - \text{var } b$ follows from the known properties of both estimators under the null hypothesis and from uncorrelatedness. The statistic $m$ is distributed as $\chi^2$ under the null hypothesis, with degrees of freedom corresponding to the dimension of $b$. In the concrete case of panel models, we know that the FE estimator is consistent in the RE model as well as in the FE model. In the FE model it is even efficient, in the RE model it has good asymptotic properties. By contrast, the RE-GLS estimator cannot be used in the FE model, while it is efficient by construction in the RE model. The inconsistency of the RE estimator in the FE model follows from the fact that, as $T \rightarrow \infty$, the individual fixed effects $a_i$ are not estimated but are viewed as realizations of random variables with mean zero. The violation of the assumption $Ea = 0$ for the regression model leads to an inconsistency (Robert M. Kunst, 2009).

2.3 Estimators for the fixed-effects model

Estimators of the parameters $b$ of the FE model must remove the fixed-effects $a_i$.

2.4 Within estimator

The within estimator eliminates the fixed-effect by mean-differencing. It performs OLS on the mean-differenced data. Because all the observations of the mean-difference of a time-invariant variable are zero, we cannot estimate the coefficient on a time-invariant variable. Because the within estimator provides a consistent estimate of the FE model, it is often called the FE estimator. It is also consistent under the RE model, but alternative estimators are more efficient.

The fixed-effects $a_i$ can be eliminated by subtraction of the corresponding model for individual means $\bar{y}_i = \bar{x}_i b + \bar{\epsilon}_i$ = leading to the within model or mean-difference model:

$$ (y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i) b + (\epsilon_{it} - \bar{\epsilon}_i) $$  \hspace{1cm} (2)

The within estimator is the OLS estimator of this model. Because $ai$ has been eliminated, OLS leads to consistent estimates of $b$ even if $ai$ is correlated with $x_{it}$ as is the case in the FE model. This result is a great advantage of panel data. Consistent estimation is possible even with endogenous regressors, provided that $x_{it}$ is correlated only with the time-invariant component of the error, $ai$, and not with the time-varying component of the error, $\epsilon_{it}$. Stata fits the model:

$$ (y_{it} - \bar{y}_i + \bar{\bar{y}}) = a + (x_{it} - \bar{x}_i + \bar{\bar{x}}) b + (\epsilon_{it} - \bar{\epsilon}_i + \bar{\bar{\epsilon}}) $$ \hspace{1cm} (3)

where, for example, $\bar{\bar{\epsilon}} = (\frac{1}{T}) \bar{\epsilon}_i$ is the grand mean of $\epsilon_{it}$.

This parameterization has the advantage of providing an intercept estimate, the average of the individual effects $a_i$, while yielding the same slope estimate $b$ as that from the within model. The default standard errors assume that after controlling for $a_i$ the error $\epsilon_{it}$ is independent and identically distributed (i.i.d) (A. Colin Cameron and P.K. Trivedi, 2009).

2.5 First-difference estimator (FD)

The first-difference estimator is obtained by performing OLS on the first-differenced variables:

$$ (y_{it} - y_{it-1}) = (x_{it} - x_{it-1}) b + (\epsilon_{it} - \epsilon_{it-1}) $$ \hspace{1cm} (4)

First-differencing has eliminated $ai$, so OLS estimation of this model leads to consistent estimates of $b$ in the FE model. The coefficients of time-invariant regressors are not identified. The FD estimator is relying on weaker exogeneity assumptions that become important in dynamic panels. For the static FE models, the within estimator is traditionally favored as it is more efficient estimator if the $\epsilon_{it}$ are i.i.d. (A. Colin Cameron and P.K. Trivedi, 2009).

2.6 Heteroskedasticity

The standard error component given by equation (1) assumes that the regression disturbances are homoskedastic with the same variance across time and individuals. This may be a restrictive assumption for panels. When heteroskedasticity is present the standard errors of the estimates will be biased and we should compute robust standard
errors correcting for the possible presence of heteroskedasticity. The fixed-effects regression model estimated by stata software invokes the OLS estimator under the classical assumptions that the error process is independently and identically distributed (Christopher F. Baum, 2001). Also, the stata software estimates this model assuming homoskedasticity. The most likely deviation from homoskedastic errors in the context of pooled cross-section time-series data (or panel data) is likely to be error variances specific to the cross-sectional unit.

When the error process is homoskedastic within crosssectional units, but its variance differs across units we have so called groupwise heteroskedasticity.

The stata software calculates a modified Wald statistic for groupwise heteroskedasticity in the residuals of a fixed-effect regression model. The null hypothesis specifies that \( \sigma_i^2 = \sigma^2 \) for \( i = 1, \ldots, N_p \), where \( N_p \) is the number of cross-sectional units. The resulting test statistic is distributed Chi-squared under the null hypothesis of homoskedasticity.

### 2.7 Serial correlation

Because serial correlation in linear panel-data models biases the standard errors and causes the results to be less efficient, researchers need to identify serial correlation in the idiosyncratic error term in a panel-data model. While a number of tests for serial correlation in panel-data models have been proposed, a new test discussed by Wooldridge (2002) is very attractive because it requires relatively few assumptions and is easy to implement (David M. Drukker, 2003).

Wooldridge’s method uses the residuals from a regression in first-differences. Note that first-differencing the data removes the individual-level effect, the term based on the time-invariant covariates and the constant,

\[
(y_{it} - y_{i(t-1)}) = (X_{it} - X_{it-1})'b1 + (\varepsilon_{it} - \varepsilon_{it-1})
\]

(5)

\[
\Delta y_{it} = \Delta X_{it}b_t + \Delta \varepsilon_{it}
\]

(6)

where \( \Delta \) is the first-difference operator.

Wooldridge’s procedure begins by estimating the parameters \( b1 \) by regressing \( \Delta y_{it} \) on \( \Delta X_{it} \) and obtaining the residuals \( \hat{\varepsilon}_{it} \). Central to this procedure is Wooldridge’s observation that, if the \( \varepsilon_{it} \) are not serially correlated, then \( \text{Corr}(\Delta \varepsilon_{it} - \Delta \varepsilon_{i(t-1)}) = -0.5 \). Given this observation, the procedure regresses the residuals \( \hat{\varepsilon}_{it} \) from the regression with first-differenced variables on their lags and tests that the coefficient on the lagged residuals is equal to -.5. To account for the within-panel correlation in the regression of \( \hat{\varepsilon}_{it} \) on \( \hat{\varepsilon}_{i(t-1)} \), the VCE is adjusted for clustering at the panel level. Since \textit{cluster ()} implies \textit{robust}, this test is also robust to conditional heteroskedasticity. This test is implemented in Stata by David Drukker that performs a Wald test, where the null hypothesis is no first order autocorrelation.

### 2.8 Driscoll and Kraay estimator

Standard error estimates of commonly applied covariance matrix estimation techniques are biased and hence statistical inference that is based on such standard errors is invalid. Fortunately, Driscoll and Kraay (1998) propose a nonparametric covariance matrix estimator which produces heteroskedasticity consistent standard errors that are robust to very general forms of spatial and temporal dependence. Stata has a long tradition of providing the option to estimate standard errors that are “robust” to certain violations of the underlying econometric model. The Stata program, implemented by Daniel Hoehle, estimates pooled OLS and fixed effects (within) regression models with Driscoll and Kraay standard errors. The error structure is assumed to be heteroskedastic, autocorrelated up to some lag, and possibly correlated between the groups (panels) (Daniel Hoehle, 2007).

### 2.9 The econometric results

We consider a linear regression model with ABS(DACC) the dependent variable and CSR and SIZE and LEV and ROA as regressors. Descriptive statistics for the variables are reported in Table 1. The mean value of absolute earnings management moves around 0.114. This value is higher in our sample in comparison to that of Choi et al. (2013) and Prior et al. (2008).

The descriptive analysis shows that mean of CSR variable is 2.346 while the mean of SIZE variable is 6.067. The mean of LEV variable is 0.094 and the mean of ROA variable is 0.127.
A starting point for estimating the model is a pooled OLS regression. But we must know if pooling the data is the solution in our case. So, a poolability test is needed.

The results obtained in Stata tells us to reject the null hypothesis that all $\alpha_i$ are zero. This also means that the OLS estimator is biased and inconsistent and we accept the presence of the individual effects.

Next, we run a Hausman test to decide whether we have a random-effects model or a fixed-effects one.

**Table 3: Hausman test**

<table>
<thead>
<tr>
<th>CHI-SQ Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Statistic</td>
</tr>
<tr>
<td>Degree of freedom</td>
</tr>
<tr>
<td>Probability of Test statistic</td>
</tr>
</tbody>
</table>

The probability is 0.000, less than 0.05, so we reject the null hypothesis that individual effect is random and that RE provides consistent estimates. Concluding that we have a fixed-effects model, we continue with the estimation of our model using the within estimator.

**Table 4: Fixed-effect (within) regression**

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>T-statistic</th>
<th>significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSR</td>
<td>-0.0164</td>
<td>0.028</td>
<td>-5.87</td>
<td>0.000</td>
</tr>
<tr>
<td>SIZE</td>
<td>-0.0002</td>
<td>0.0046</td>
<td>-0.05</td>
<td>0.956</td>
</tr>
<tr>
<td>Lev</td>
<td>0.2450</td>
<td>0.0205</td>
<td>11.95</td>
<td>0.000</td>
</tr>
<tr>
<td>ROA</td>
<td>0.0889</td>
<td>0.0257</td>
<td>3.45</td>
<td>0.001</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>0.0984</td>
<td>0.0293</td>
<td>3.35</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**F-statistic**

| 108.23 |

**R-squared**

| within | 0.3412 |

| SIGMA_U | 0.0608 |

| SIGMA_E | 0.0250 |

The estimated standard deviation of $\alpha_i$ (sigma_u) is 0.060, much bigger than the standard deviation of $\epsilon_i$ (sigma_e) which is 0.025, suggesting that the individual-specific component of the error is much more important than the idiosyncratic error.

The standard error component model assumes that the regression disturbances are homoskedastic.

After the estimation of our model we can perform a
modified Wald test for groupwise heteroskedasticity in the fixed effect model, implemented in Stata by Christopher Baum, using the stata software.

**Table 5: Heteroskedasticity test**

<table>
<thead>
<tr>
<th>Exploratory variables</th>
<th>Coefficient Standard Error</th>
<th>T-statistic</th>
<th>significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSR</td>
<td>-0.0164</td>
<td>0.016</td>
<td>-10.08</td>
</tr>
<tr>
<td>SIZE</td>
<td>-0.0002</td>
<td>0.004</td>
<td>-0.04</td>
</tr>
<tr>
<td>Lev</td>
<td>0.0345</td>
<td>0.0203</td>
<td>12.03</td>
</tr>
<tr>
<td>ROA</td>
<td>0.0899</td>
<td>0.0162</td>
<td>5.46</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>0.0984</td>
<td>0.0463</td>
<td>2.12</td>
</tr>
</tbody>
</table>

The results (Probability of Test statistic < 0.05) indicate that we must reject the null hypothesis of homoskedasticity.

We also need to test for serial correlation which is very likely to appear in an individual-effects model. We do so with the Stata software, implemented by David Drukker.

**Table 6: serial correlation test**

<table>
<thead>
<tr>
<th>Exploratory variables</th>
<th>Coefficient Standard Error</th>
<th>T-statistic</th>
<th>significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSR</td>
<td>-0.0164</td>
<td>0.016</td>
<td>-10.08</td>
</tr>
<tr>
<td>SIZE</td>
<td>-0.0002</td>
<td>0.004</td>
<td>-0.04</td>
</tr>
<tr>
<td>Lev</td>
<td>0.0345</td>
<td>0.0203</td>
<td>12.03</td>
</tr>
<tr>
<td>ROA</td>
<td>0.0899</td>
<td>0.0162</td>
<td>5.46</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>0.0984</td>
<td>0.0463</td>
<td>2.12</td>
</tr>
</tbody>
</table>

The probability obtained for our model is 0.000. This indicates that the errors are autocorrelated.

We are now facing two problems with our model: heteroskedasticity and serial correlation. Assuming homoskedastic disturbances when heteroskedasticity is present or ignoring correlation in the estimation of panel models can lead to biased statistical results. To ensure validity of the statistical results, most recent studies which include a regression on panel data therefore adjust the standard errors of the coefficient estimates for possible dependence in the residuals (Daniel Hoechle, 2007).

The stata software performs fixed-effects (within) regression with Driscoll and Kraay standard errors. The error structure is assumed to be heteroskedastic, autocorrelated up to some lag and possibly correlated between the groups. The author of this Stata command is Daniel Hoechle.

**Table 7: Fixed-effect (within) regression with Driscoll and Kraay standard errors**

The resulted econometric model is:

\[ \text{ABS(DACC)} = 0.984 - 0.0164 \text{CSR} + 0.245 \text{LEV} + 0.0889 \text{ROA} \]

We use t-statistics based on Driscoll and Kraay standard errors, which are robust both to heteroskedasticity and within-firm serial correlation. The results show a consistently significant negative relationship between corporate social responsibility practices (CSR) and absolute discretionary accruals. The discretionary accruals are substantially lower to more socially responsible firms, and this difference is statistically significant at the 0.05 level. These results provide strong evidence for the effect of corporate social responsibility practices in reducing earnings management.

Also, the variable Leverage (Lev) is found to be significantly and positively related to absolute discretionary accruals. And the variable net income divided by end-of-year total assets (ROA) is found to be significantly and positively related to absolute discretionary accruals.

### 3 CONCLUSIONS

In this article, we examine the relationship between CSR and earnings quality by using IRANIAN companies from 2008 to 2015. To explain this connection, we used a fixed-effects regression model, and t-statistics based on Driscoll and Kraay standard errors, which are robust both to heteroskedasticity and within-firm serial correlation.

For the analysis, the MERCO index is employed as a proxy for the CSR ratings of Spanish firms. The
earnings quality is measured by using the absolute value of abnormal discretionary accruals from the modified Jones model. The empirical results confirm to our theoretical contention. In particular, we find a negative impact of CSR practices on earnings management, so firms that are more committed to CSR engage less in earnings management. These results are consistent with Chih et al. (2008). They find that companies with higher social responsibility engage in less earnings smoothing and less earnings decrease/loss avoidance. Similarly, Shleifer (2004) interprets that earnings manipulation, which many people findethically objectionable, occurs less often in corporations with a strong commitment to social responsibility. As suggested by Hong and Andersen (2011), more socially responsible firms have higher quality accruals and less activity-based earnings management, both of which impact financial reporting quality.

REFERENCES


A. Colin Cameron, P.K. Trivedi, Microeconometrics Using Stata, Stata Press, 2009


Badi H. Baltagi, Econometric Analysis of Panel Data, John Wiley & Sons Ltd, 2008


La Porta, R., Lopez de Silanes, F., Shleifer, A.,


