A Generalised Dynamic Model of Accounting Earnings on Stock Market Returns

January 2010

Please do not cite – preliminary and incomplete

Abstract

This paper examines the dynamic relationship between accounting earnings and stock market returns, including the speed with which market returns are reflected in earnings within the current period (timeliness) and in the long run (persistence), as well as the existence of a long run equilibrium. We provide a unified analytical framework that examines these properties of earnings in a dynamic panel data context and expresses previous specifications as special-case models.

Key Words: earnings properties, market returns, dynamic panel data.

*Corresponding Author: E-Mail: Demetris.Christodoulou@sydney.edu.au. Tel: +61.2.9036.7815. Fax: +61.2.9351.6638. Address: Discipline of Accounting, Economics and Business Building H69, The University of Sydney, NSW 2006, Sydney, Australia.
1. Introduction

This paper examines the dynamic relationship between accounting earnings and stock market returns, on the basis that market expectations that are reflected immediately tend to be transitory, in contrast to those that are reflected more gradually and must have a more permanent effect. The empirical timeliness and permanence of expected economic returns, as proxied by stock market changes in equity prices, is widely accepted to be asymmetric and defined about the origin. Specifically, the often complete and immediate recognition of current negative returns and most of future expected unrealised losses in current earnings implies a transitory abrupt shock with no lasting effects. On the other hand, the more cautious and gradual recognition of current positive returns and future expected unrealised gains in current earnings implies a more permanent presence in the future stream of earnings.

We navigate through literature to offer a review of analytical developments that led to the examination of the relationship between earnings and returns. Given this knowledge, we describe a unified framework of analysis that shows how previous specifications that have looked at earnings timeliness and persistence in separate specifications are in fact merely special-case models. The proposed generalised dynamic panel data model has the additional advantage of examining the potential existence of long run equilibrium, as well as the degree and direction of panel data causality between earnings and returns.

2. The association between accounting earnings and stock market returns

One of the most significant contributions of capital market research is that the movement in stock market returns is directly associated with changes in accounting earnings (Ball and Brown 1968; Beaver 1968). The natural question that follows is how earnings grow over time. For firm $i$ and year $t$, using autoregressive processes with or without a drift, it has been shown that the annual series of earnings per share $E_{it}$ does not follow a random walk and so
the first differences \( \Delta E_{it} \) are not pure noise, hence an integrated moving average process of \( \Delta E_{it} \) may capture a significant unexpected component of earnings (Beaver 1970; Ball and Watts 1972; Albrecht, Lookabill and McKeown 1977; Watts and Leftwich 1977).

In addition, the movement in \( \Delta E_{it} \) is better explained by an expectation function containing not only information on past changes in earnings, but also on current annual buy-and-hold market returns \( R_{it} \) (Beaver, Lambert and Morse 1980; Collins, Kothari and Rayburn 1987), where \( \tau \) indicates the order of lag. Indeed, it is now a standard result that prices lead earnings, in the sense that \( R_{it} \) conveys expectations of future earnings releases. It follows that a regression of \( R_{it} \) on \( \Delta E_{it} \) is likely to bring biased results since the independent variable and the error term are linked contemporaneously (Beaver, Lambert and Ryan 1987). A reverse regression of earnings on returns appears to be more appropriate.

Since prices anticipate earnings many periods before their release and at the same time some information of current earnings is already impounded in past prices and so it would not appear again in current prices, then past returns \( R_{it-\tau} \) should also be very useful in explaining \( E_{it} \) (Beaver et al. 1987; Collins et al. 1987; Collins and Kothari 1989). However, a puzzling observation at the time was the weak contemporaneous correlation between \( R_{it} \) and \( E_{it} \), which raised the question of what is the optimal memory length required for \( E_{it-\tau} \) and \( R_{it-\tau} \) in order to explain the current level of \( E_{it} \) (Collins and Kothari 1989; Lev 1989).

---

1 Bottom-line accounting income is most likely generated by a non-stationary process, which is why the time series of earnings alone has proven to be very difficult to model and it is sometimes found to be at best a martingale with a drift. Freeman, Ohlson and Penman (1982) explain that one of the difficulties faced is the differential degree of persistence amongst earnings components as well as the lack of structure in the forecasting equations (Ohlson 1995; Penman 2009).

2 Beaver et al. (1980) offer a valuation model that directly links the demand of accounting earnings to its equity market, and explore the contemporaneous relationship between earnings and the implied stock market returns using a model that utilises the growth of prices as a characterisation for the stochastic generating process of earnings growth. It is commonly accepted that the market price of equity conveys significant added information to the book value of equity especially because of expectations for future earnings. Hence, it can be said that earnings is a lagged function of prices in the sense that the variation in current and future accounting returns may be partially explained by the variation in contemporaneous market returns.
Moreover, the earnings yield $E_{it}/P_{it-1}$ (i.e. the current level of earnings per share $E_{it}$ divided by the beginning of the period price $P_{it-1}$) was found to be a more informative covariate to $R_{it}$ than the change in earnings per share $\Delta E_{it}$, because $E_{it}/P_{it-1}$ approximates the proportion of current earnings to the earnings expected to be realised (Easton and Harris 1991; Ohlson 1991; Ohlson and Shroff 1992; Easton, Harris and Ohlson 1992; Kothari 1992). Combining the propositions of lengthy memory and the use of the earnings yield gives the potential for increased power in empirical specifications. Specifically, it is found that the higher the cumulative order of lagged returns, $\sum_{\tau=1}^{T-1} R_{it-\tau}$, the greater the degree of association between $E_{it}/P_{it-1}$ and $R_{it}$ because the permanent value-relevant component of earnings increases by comparison to the transient value-irrelevant component which is largely random and may cancel-out (Easton et al. 1992; Kothari 1992; Kothari and Sloan 1992).

Further research found the bivariate relationship to be governed by a nonlinear tendency and to approximate an S-shaped curve that is convex for negative returns and concave for positive returns. This may be partly due to extreme $\Delta E_{it}$ or $E_{it}/P_{it-1}$ that are seen as largely transitory and receive little weight by the market, implying that the tails of the distributions are largely populated by value-irrelevant information (Freeman and Tse 1992), or even because of omitted value-relevant information that is accounted for by the market but not found in booked earnings (Lev and Thiagarajan 1993; Antle, Demski and Ryan 1994). Furthermore, the uncertainty in the precision of disclosed earnings (especially under conservative accounting) may cause a non-monotonic reaction of prices to earnings, so that as the

3 The earnings yield has an intuitive economic interpretation. It is the inverse of the price-to-earnings ratio and is useful for the valuation of equity as it proxies for the unexpected portion of earnings, i.e. the ratio of current realised earnings to expected period earnings (Ohlson 1991; Ohlson and Shroff 1992; Easton et al. 1992). Another attractive feature of the earnings yield is that it avoids misspecification in regressions that employ observables in levels, because the deflation by lagged price is a direct function of the independent variable (Christie 1987).

4 Easton et al. (1992) extend the returns window to ten years and achieve very high levels of correlation between cumulative returns and earnings. They further show that for adequately large return intervals and under certain restrictions on goodwill, earnings and returns would be perfectly correlated.
magnitude of earnings surprise increases the uncertainty over the precision of news becomes greater and the price reaction decreases (Subramanyam 1996). General nonlinearity has also been found to prevail in aggregate cross-sectional and pooled studies that overlook certain statistical properties of the data, including heteroscedasticity, residual non-normality and firm-specific fixed effects (Beneish and Harvey 1998; Lipe, Bryant and Widener 1998).

A defining characteristic of this nonlinearity is the asymmetry about the origin of expected economic gains, i.e. of materialised stock market returns. This may be because investors do not expect losses to carry on in perpetuity by contrast to profits which are expected to persist in the long run (Hayn 1995), or due to traditionally prudent financial reporting that applies more sensitivity to market expectations of earnings downgrades than market expectations of earnings upgrades. The asymmetry can be approximated through the use of linear splines, with the simplest case comprising of only one turning point at $R_t = 0$ (e.g. Basu 1997; Beneish and Harvey 1998; Ball, Kothari, and Robin 2000). This implies a differential timeliness in earnings that reflects net gains more gradually and with more persistence by comparison to net losses that are reflected with a steeper gradient and are more transient.

Another robust result is that the asymmetry in contemporaneous timeliness and aggregate long run persistence increases with higher cumulative orders of lagged returns (Pope and

---

5 A number of nonlinear models are employed for fitting the asymmetry observed between earnings and returns, such as the parametric inverse tangent function and the sigmoid function (Freeman and Tse 1992) and the nonparametric locally weighted regression (Das and Lev 1994). These attempts confirm the expected S-shaped asymmetry described as inherent to the relationship that is irrespective of transitory components in earnings. It is worth noting that their analysis is in accord with the Prospect Theory of Kahneman and Tversky (1979) under which risk aversion defines point zero as a ‘reflection effect’ (p.279), which allocates asymmetric expectations of realised income with an S-shape function that is steeper for losses than for profits.

6 The asymmetry in how earnings reflect market returns is often attributed to prudential accounting requiring a higher degree of verification when recognizing positive returns than negative returns so that assets are not overstated, liabilities are not understated and uncertainties are accounted for. Basu (1997) adopts this definition for ‘earnings conservatism’, later re-labelled as ‘market-based’, ‘news-depended’ or ex post conservatism (Richardson and Tinakar, 2004; Beaver and Ryan 2005; Pae, Thornton and Welker, 2005; Ryan 2006; Givoly et al. 2007; Roychowdhury and Watts 2007). However, the notion of conservatism is no longer relevant and is now viewed as “biased and unacceptable” (FASB, 2006, p.43), and it is formally substituted by the principle of neutrality as the only acceptable qualitative characteristic for faithful representation (IAS Plus, 2006). In any case, this paper is not concerned with conservatism or neutrality per se since such effects are likely to cancel-out in the long run.
Walker 1999; Giner and Rees 2001; Ryan and Zarowin 2003). In any case, for efficient and rationale markets, it is reasonable to expect that long run earnings will eventually reflect the consensus market expectations as expressed by negative and positive returns.

3. A generalised framework of analysis

Building on the above knowledge, this study provides a structural framework for modelling the dynamic properties of earnings with respect to its market, in the short term and in the long run. Consider the following generalised specification of an autoregressive model with distributed lags, denoted as $\text{ARDL}(p,q)$, of the earnings yield $E'_{it} = E_{it}/P_{it-1}$ on stock market returns $R_{it}$, for firm $i = 1, 2, \ldots, I$ and year $t_i = 1, 2, \ldots, T_i$.

$$E'_{it} = \alpha_0 + \alpha_1 \mu_{it} + \sum_{\tau=1}^{p} \rho_{\tau} E'_{it-\tau} + \sum_{\tau=0}^{q} \beta_{\tau} R_{it-\tau} + \epsilon_{it}$$  \hspace{1cm} (1)

where $\alpha_0$ indicates the average drift and $\alpha_1$ the coefficient on the deterministic trend $\mu_{it}$. $\rho_{\tau}$ indicates the autoregressive coefficient on $E'_{it-\tau}$ with order $\tau = 1, 2, \ldots, p \leq T_i - 1$, and $\beta_{\tau}$ the distributed lag coefficient on $R_{it-\tau}$ with order $\tau = 0, 1, 2, \ldots, q \leq T_i - 1$. Also, $R_{it-\tau}$ is unlikely to be strictly exogenous and could be predetermined because the error term at time $t$ may affect the subsequent realisations of returns, so that $\text{E}[R_{it}, \epsilon_{it}] \neq 0$. Prior research has even argued that $R_{it-1}$ may also be correlated with the error term and this is a stronger definition of predeterminedness, $\text{E}[R_{it-1}, \epsilon_{it}] \neq 0$ for $\tau \geq 0$. We allow for these possibilities when estimating equation (1).

---

7 The time index $t$ is formally subscripted by firm $i$, as $t_i$, with time period $T_i$ unique to the firm and total number of observations equal to $N = \sum_{i=1}^{I} I \times T_i$. The subscripted time is intended to emphasise that firms are characterised by specific time-series, so that when pooled together they may result in an unbalanced weighted sample. This has important implications in empirical analysis, as we explain later in the paper. However, to simplify notation, in all equations firm-specific time steps are denoted by $t$ and the overall time period by $T_i$. 
The ARDL($p,q$) model describes an expectation function where the effect of returns on earnings is distributed over time so that it is felt gradually in future realised earnings. This is a different definition that the one usually employed in the literature where market returns are aggregated over the years and the estimation recovers a single aggregate parameter $\beta$ from $\beta \sum_{\tau=0}^{q} R_{t-\tau}$ (hence assuming that $\beta_0 = \beta_1 = \ldots = \beta_q$). As discussed above, the explained variation in $E'_u$ increases with higher orders of $q$, but given the concurrent causal relationship between earnings and returns, then as $q$ increases so should $p$ at the same rate to ensure that this relation is maintained also for lagged periods.

Equation (1) provides the means for measuring two key properties of earnings. First, the instantaneous multiplier of returns $\beta_0$ captures the transient component of earnings that is expected by its market. Second, the long run multiplier $\lambda_{pq}$, which is a function of the estimated parameters from equation (1) and includes $\beta_0$, captures the more persistent component of earnings as reflected by the aggregate change in materialised returns, as follows:

$$\lambda_{pq} = \frac{\sum_{\tau=0}^{q} \beta_\tau}{1 - \sum_{\tau=1}^{p} \rho_\tau} \quad (2)$$

Given the orders of $p$ and $q$, the long run multiplier has a bounded range, $0 \leq \lambda_{pq} \leq 1$, and measures how a change in $R_t$ explains a change in $E'_u$ that is distributed over time. For instance, if returns change in the same direction by $\$1$ for all $q$ then the mean expected value of $E'_u$ will change by $\lambda_{pq}$. Hence, it can be said that $\lambda_{pq}$ measures the degree of earnings persistence due to the changes in long run market expectations of earnings, whereas the
contemporaneous impact that is captured by $\beta_0$ measures transient expectations as reflected in current earnings.\(^8\)

In addition, following the documented asymmetry that is defined about the origin of $R_u$, we also expect a differential degree of earnings timeliness and persistence. This question can be investigated by partitioning samples on the sign of returns.\(^9\) Specifically, since positive market returns reflect expectations of economic gains from future earnings releases, the long run multiplier $\lambda_{pq}$ from a sample of $R_{it} \geq 0$ should approximate unity as the order of distributed lags increases, $q \to T - 1$, while the instantaneous multiplier $\beta_0$ should be kept relatively low at all orders. On the other hand, given the generally transient nature of economic losses, a sample of $R_{it} < 0$ should yield much smaller $\lambda_{pq}$ but larger $\beta_0$ irrespective of memory length. We examine these empirical questions using standard statistical procedures that test whether $\beta_0 \to 1$ or $\sum_{\tau=1}^{p} \rho_{\tau} + \sum_{\tau=0}^{q} \beta_{\tau} \to 1$.\(^{10}\)

\(^8\) A key assumption for equation (1) is that stock prices lead accounting earnings so that prices are efficient in reflecting publicly available information and clean surplus violations are not systematic. Moreover, to ensure the existence of a long run relationship between earnings and returns, an implicit assumption is that the earnings series does not contain a unit root so that it holds $(1 - \sum \rho_{\tau}) \neq 0$.

\(^9\) The differential degree of empirical earnings timeliness on positive and negative returns can be examined using a regression analysis of earnings on positive and negative returns using a dummy variable (e.g. Basu 1997; Pope and Walker 1999; Ball et al. 2000; Giner and Rees 2001; Ryan and Zarowin 2003). This approach approximates the observed nonlinearity between the two variables using the simplest form of linear spline with one turning point at zero. Instead, the approach adopted in this paper is the partition of the sample using the sign of returns, which however yields exactly the same result. For instance, it can be shown that the parameters recovered from two stratified ARDL(0,1) regressions on a sample of negative returns $E_{it} = \alpha_{0} + \beta_{0} R_{it} + \epsilon_{it}$ and on a sample of positive returns $E_{it} = \alpha_{0} + \beta_{0} R_{it} + \epsilon_{it}$ are equivalent to those recovered from the single regression $E_{it} = \alpha_{0} D_{it}^{+} + \beta_{0}^{+} R_{it}^{+} + \epsilon_{it}^{+}$, where $D_{it}$ is a dummy variable that takes the value of one for $R_{it} < 0$ and zero otherwise, and $D_{it}^{+}$ is a dummy variable that takes the value of one for $R_{it} \geq 0$ and zero otherwise. Within the same lines, Dietrich, Muller and Riedl (2007) show that the single equation above can be reduced to $E_{it} = \alpha_{0} + \beta_{0} R_{it} + \gamma_{0} D_{it}^{+} + \delta_{0} D_{it}^{-} R_{it} + \epsilon_{it}$, with the direct parameter equivalence of $\alpha_{0} = \alpha_{0}$, $\beta_{0}^{+} = \beta_{0}$, $\gamma_{0} = \gamma_{0}$ and $\delta_{0} = \delta_{0}$.

\(^{10}\) Ryan and Zarowin (2003) attempt the derivation of a similar long run effect by adding together all estimated coefficients from an OLS regression of earnings on a number of lagged positive and negative price changes, i.e., using an ARDL(0,q) model. This specification assumes $\sum \rho_{\tau} = 0$ for all $\tau$ and ignores the relative effect of a highly likely dynamic structure within earnings that by assumption are causally related to stock returns.
The drift $\alpha_0$ is expected to be positive and highly significant for samples of $R_t \geq 0$ as it captures the part of permanent earnings that is proportional to past positive market expectations, the realisation of which was deferred to future earnings in a gradual manner.\footnote{Pope and Walker (1999) explain that if permanent earnings are defined as price times the cost of capital, dividends are equal to permanent earnings and prices follow a random walk, then if prior period’s realised gains were to be modelled separately the remainder of $\alpha_0$ would be reduced to the cost of capital.} For samples of $R_t < 0$, the drift is expected to lose statistical relevance and approach zero, $\alpha_0 \to 0$, since the recognition of the expectation of future losses is less likely to be postponed in future periods.

The model described in equations (1) and (2) is a generalised specification that can be reduced to a number of special-case models. To give some examples, Beaver (1970), Ball and Watts (1972), Albrecht et al. (1977) and Watts and Leftwich (1977) examine the univariate time series process of $\text{ARDL}(1,\cdot)$ with no distributed lag terms ($\beta_t = 0$), ignoring therefore any useful information that may be conveyed through market returns. Easton and Harris (1991) and Easton et al. (1992) examine a $\text{ARDL}(1,0)$ process, also known as the ‘partial adjustment model’, where the long run multiplier $\lambda_{10} = \beta_0/(1 - \rho_1)$ defines the degree of adjustment over one period only. That is, $\lambda_{10} = 1$ indicates complete adjustment and $\lambda_{10} = 0$ indicates no adjustment, where any intermediate state $0 > \lambda_{10} > 1$ implies partial or gradual adjustment from the beginning to the end of the financial year. If $\rho_1 = 1$ then there is a unit root in earnings, which means that the firm is out-of-equilibrium and it is very likely that it will stay there (Heij, de Boer, Frances, Kloek, and van Dijk, 2004). This is one of the reasons why it is useful to consider higher orders of $p$ and $q$ as there is more potential in capturing the degree of adjustment. For instance, Collins and Kothari (1989) propose the $\text{ARDL}(1,1)$ model where Easton et al. (1992) include up to nine distributed lags, $\text{ARDL}(1,9)$. Contrary to $\text{ARDL}(1,0)$, the $\text{ARDL}(1,1)$ model takes into account that earnings and returns move together at time $t$,
whereas ARDL(1,9) recognises that it takes many periods for market expectations to be reflected in earnings, but it is somewhat under-specified as it overlooks their contemporaneous relation at all $t$. Furthermore, by partitioning samples on the sign of market returns, Basu (1997) estimates two static processes of ARDL(.0) with no autoregressive terms ($\rho_1 = 0$), and Pope and Walker (1999) estimate two ARDL(.3). Instead, Giner and Rees (2001) partition a sample on the basis of the sign of earnings and estimate two dynamic partial adjustment models ARDL(1,0).

3.1. Allowing for periodic correction to the long run equilibrium

There is plenty of evidence documenting a significant moving average component of order one in the annual time series of earnings (e.g. Watts and Leftwich 1977; Beaver et al. 1980). The inclusion of the first order moving average (i.e. the past period error term) relies on the premise that the surprising element of earnings is expected to average over adjacent periods, and its estimated parameter is expected to capture the average gradient of mean reversion. The generalised form of equation (1) with a moving average of order one $\varepsilon_{it-1}$ with coefficient $\theta_1$ is given by $E_{it} = \alpha_0 + \alpha_1 \mu_{it} + \sum_{\tau=1}^{p} \rho_{\tau} E_{it-\tau} + \sum_{\tau=0}^{q} \beta_{\tau} R_{it-\tau} - \theta_1 \varepsilon_{it-1} + \varepsilon_{it}$. For example, for $p = q = 1$, the model yields a correction mechanism known as the ‘adaptive expectations model’ that describes how firms may decide on the next level of $E_{it+1}$ on the basis of mean expected changes of the next period’s $R_{it+1}$, so that the expectation of $E_{it+1}$ is adjusted on the basis of current realised $E_{it}^r$ (Finger 1994).12

---

12 The ‘adaptive expectations model’ with $\theta_1 = 0$ reduces to the ARDL(1,1) process, but with $\theta_1 = 1$ the model turns into a complete mean reverting process where the impact of a shock from lagged returns $R_{it-1}$ to the earnings yield $E_{it}$ takes no more than one period to fade away. When realised $E_{it}$ is larger than expected then there is an instantaneous upwards correction in future expectations by $\beta_0 > 0$ that is however anticipated by $R_{it-1}$, whereas when $E_{it}$ is smaller than expected then the correction is downwards by $\beta_0 < 0$ (Heij, de Boer, Frances, Kloek, and van Dijk, 2004). Nevertheless, past empirical exercises report a moving average of $0 < \theta_1 < 1$, hence verifying that earnings only partially revert to their mean in adjacent periods (Finger 1994).
However, the focus of this paper is on the longer run relationship between earnings and returns, and moving averages are only useful in describing the behaviour surrounding adjacent periods. A more informative long-run correction mechanism is the ‘error correction model’ that allows for the adjustment of $E_{it\tau}$ towards a long-run equilibrium. By defining $E_{it\tau} = \Delta E_{it\tau} + E_{it\tau-1}$ and $R_{it\tau} = \Delta R_{it\tau} + R_{it\tau-1}$ for successive $\tau$ values, we rewrite equation (1) and derive the generalised form of the error correction model:

$$\Delta E_{it} = \alpha_0 + \alpha_1 u_{it} + \sum_{\tau=1}^{p} \rho_{\tau} \Delta E_{it-\tau} + \sum_{\tau=0}^{q} \beta_{\tau} \Delta R_{it-\tau} - \left(1 - \sum_{\tau=1}^{p} \rho_{\tau}\right) \left(E_{it-1} + \lambda_{pq} R_{it-1}\right) + \epsilon_{it}$$

(3)

where $\lambda_{pq}$ is again given by equation (2). Equation (3) captures the variation in earnings around the long-run equilibrium $\Delta E_{it} = \alpha_0 + \alpha_1 u_{it} + \sum_{\tau=1}^{p} \rho_{\tau} \Delta E_{it-\tau} + \sum_{\tau=0}^{q} \beta_{\tau} \Delta R_{it-\tau} + \epsilon_{it}$, given the correction in deviations from the equilibrium within each successive period $- \left(1 - \sum_{\tau=1}^{p} \rho_{\tau}\right) \left(E_{it-1} + \lambda_{pq} R_{it-1}\right)$.

Note that equation (3) assumes that $E_{it}$ is stationary because if $\sum_{\tau=1}^{p} \rho_{\tau} = 1$ the $\lambda_{pq}$ is infinitely large, the error correction component vanishes and there is no long-run equilibrium. In the other extreme, if $\sum_{\tau=1}^{p} \rho_{\tau} = 0$ the adjustment is instantaneous and complete. Hence, of most interest are the intermediary states that trigger the error correction mechanism, so that when lagged values of $E_{it\tau}$ are below [above] the long-run equilibrium the current level of earnings will be corrected upwards [downwards] by $- \left(1 - \sum_{\tau=1}^{p} \rho_{\tau}\right)$ and the aggregate effect on $\Delta E_{it}$ is positive [negative]. Hence, equation (3) captures the long-run equilibrium through successive between-period temporary adjustments.

---

13 The error correction model is a conceptual model for cointegrating variables. Specifically, it is assumed that earnings and returns cointegrate in first order (hence the first-differenced relationship), and therefore they share a common stochastic trend that however can be eliminated by estimating $\lambda_{pq}$ within the error correcting term $\left(E_{it\tau-1} + \lambda_{pq} R_{it\tau-1}\right)$; see Greene (2008, p.689) for the derivation of the error correction model for an ARDL(1,1) process.
that are targeted to the permanent component of $E_{it}^*$, as anticipated by $\Delta R_{it}$ (that is, by the future expected change in $E_{it}^*$).

Given the available time series, the choice of memory length $p$ and $q$ in equations (1) to (3) depends on the desired degree of temporal stability and estimation constraints. We employ a variety of model diagnostics to the evaluation of informativeness, complexity and relative significance of each process, including Wald tests, $F$-tests and information criteria.

3.2. Panel data structure and estimation

The time series of earnings and returns is primarily driven by firm-specific factors (Watts and Leftwich, 1977; Kormendi and Lipe, 1987), and the bivariate relationship cannot be constant across firms in cross-sectional or pooled samples (Collins and Kothari, 1989; Cheng, Hopwood and McKeeown 1992; Beneish and Harvey, 1998; Lipe et al. 1998), especially since earnings properties are conditional on accounting choices defined at the firm level (Ryan 2006). Indeed, firm-based measurements that are repeatedly observed from successive annual periods imply a panel data structure with associated heterogeneous fixed effects to the cross-sectional dimension of the firm and the time dimension of the financial year (Teets and Wasley 1996; Callen and Segal 2005; Grambovas, Giner and Christodoulou 2006). Therefore, the error term of equation (1) and (3) is likely to be as a compound of:

$$\varepsilon_{it} = (u_i + s_t + v_{it})$$

where $u_i$ indicates firm-specific fixed effects and $s_t$ time-specific fixed effects, and only the idiosyncratic residual follows the standard normal assumptions $v_{it} \sim N(0, \sigma^2)$. $u_i$ remains invariant by time and may include observed measurements such as industry classification or country of domicile, as well as other firm-specific factors that cannot be observed or are too difficult to measure/proxy, such as invisible assets or undisclosed proprietary information that
remains fixed in time. $s_t$ remains invariant by cross-section and describes effects that are generally unobserved and commonly experienced across the market, such as macro shocks and economy-wide disruptions. These effects are assumed to be significantly correlated to the expected mean function of equations (1) and (3), which means that their omission violates the key least-squares assumption of orthogonality and the analysis is certain to yield biased and inconsistent estimates, regardless of sample size.\textsuperscript{14}

Equation (4) implies the estimation of $(I + T_i + k + 1)$ parameters, even though the analysis is more concerned with the recovery of the autoregressive and distributed lag coefficients $k$. The estimation of intercepts $I$ and $T_i$ is of less interest especially because sampling tends to the firm population $I \to \infty$, hence the targeted asymptotically consistent properties, whereas $T_i$ is kept relatively short in empirical studies for stability reasons (and due to data constraints). Thus, it is standard practice to simply control for the cross-sectional heterogeneity (Baltagi 2005; Cameron and Trivedi 2006).

Pooled OLS or GLS estimation for dynamic panel data models, such as those considered in equations (1) and (3), is largely inconsistent because the lagged dependent variables (\textit{i.e.} the autoregressive terms) are by construction correlated with the firm fixed effects and hence with the error term (Nickell 1981). It is now a standard result that the most appropriate dynamic panel data estimator especially for datasets with ‘large $I$ short $T$’ is the Arellano-Bover/Blundell-Bond ‘system-GMM’ estimator (Arellano and Bover 1995; Blundell and

\textsuperscript{14} Panel data analysis offers important advantages over cross-sectional or time-series analysis, and especially over pooled analysis. Panel data methods allow for additional variability in the model by decomposing the error term into dimension-specific components that effectively control for heterogeneity across space and time. Intuitively, the panel data analysis of a firm-year dataset recognises that a firm operates in a market and while it is characterised by certain fixed effects that distinguish it from the rest, it is still affected by shocks that are commonly felt across the market. In this paper we consider another fixed effects, that of the financial year-end date, which introduces a hierarchy in the cross-sectional dimension. For a comprehensive econometric treatment of panel data analysis, see Hsiao (2003), Baltagi (2005) and Cameron and Trivedi (2006), amongst others.
A key advantage of the system-GMM estimator is that it allows us to relax the assumption that $R_{it}$ is strictly exogenous and we can treat its current and first lagged values as predetermined covariates to be instrumented.

Estimation proceeds as follows. First, we identify the order of autocorrelation $p$ in the differenced error term and use higher order instruments of lag levels and lag differences of $E_{it-\tau}$ at $p+1$ (Roodman 2006). Given previous findings of a moving average of order one in $E_{it}$, we begin GMM-type instrumentation from $p = 3$. As for $R_{it}$, we consider its strong definition of predeterminedness and begin GMM-type instrumentation from $q = 2$ (i.e. we assume that not only $R_{it}$ is correlated with $v_{it}$, but also is $R_{it-1}$). Treating variables as predetermined increases dramatically the instrument matrix and the efficiency of estimation. Also, given that the time fixed effects $s_i$ and the linear time trend $\mu_{it}$ are strictly exogenous variables, they serve well as additional IV-type of instruments in both the levels and differences equations of the GMM-system estimator. An important advantage of this additional instrumentation is that it prevents the most likely form of cross-sectional

---

15 Dynamic panel data models with ‘large $I$ short $T_i$’ are estimated through first-differencing in order to remove the panel-specific fixed effects $u_i$. However, because first-differenced autoregressive terms are directly linked to the error term results are highly biased even in the absence of $u_i$. To escape this bias, Anderson and Hsiao (1981, 1982) proposed to use as instruments further lagged terms of the autoregressive regressors. Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991) further noted that the panel data structure provides a unique opportunity to utilise many more instruments (for each lag and time period instrumented) and use the GMM framework to identify valid instruments of lagged levels from not only the autoregressive variables, but also from other endogenous, as well as predetermined and exogenous regressors. However, Arellano and Bover (1995) and Blundell and Bond (1998) show that the instruments on lagged levels become weak as the autoregressive process strengthens or the ratio $u_{it}/v_{it}$ increases. To address this issue, they propose a system-GMM estimator where lagged differences are used as instruments for a levels equation and the lagged levels are used as instruments for a differences equation.

16 On the basis of the assumption that $v_{it}$ is i.i.d. then the first-differences transformation will impose a first-order autocorrelation in the errors, which means that the moment conditions of the system-GMM are valid only if $E[u_i, \Delta^2 v_{it}] = 0$, where $\Delta^2 v_{it}$ is the second-differenced $v_{it}$. If $v_{it}$ also contains a moving average of order one $\theta_1 v_{it-1}$ then depending on its persistence then the transformation will also impose a second-order autocorrelation, and the moment conditions are valid only if $E[u_i, \Delta^3 v_{it}] = 0$, where $\Delta^3 v_{it}$ is the third-differenced $v_{it}$. We validate these key assumptions by examining the null of no autocorrelation in the errors. Also, we examine the null of Sargan’s test that the overidentifying restrictions are valid, meaning that the selected GMM-type of instruments are uncorrelated with the transformed error term. A rejection of the null of no autocorrelation in higher orders or of valid overidentifying restrictions implies a misspecification either in the model or in the selection of instruments. Both tests are discussed in Arellano and Bond (1991).
dependence, that of contemporaneous correlation (Roodman 2006; Sarafidis and Robertson 2009). Highly collinear instruments are removed from both the levels and the differences regression of the system-GMM, and all results are corrected for short panel bias and standard errors are robust to general autocorrelation and heteroscedasticity within panels (Windmeijer 2005). Also, to examine the validity of our results we estimate the Arellano and Bond (1991) autocorrelation test in the first-differenced error term and the Sargan test for overidentifying restrictions.

Estimation is performed on the partitioned sample of ‘expected loss-making firms’ and ‘expected profit-making firms’. That is to say, for the time period examined and the available annual information per firm, we find the average market return which if positive [negative] it indicates a market expectation of profit-making [loss-making] firm, on average. This is a different treatment to sample partitioning than previous studies that we adopt for two reasons. First, it is widely accepted that market returns convey expectations for future realised earnings and not so for the concurrently released earnings hence their weak contemporaneous correlation. Therefore the partitioning of samples on the basis of concurrent returns and earnings (on the so called firm-year observation) does not assign the targeted expectations to future profitability. Our approach alleviates this problem. The second reason is more practical. The partitioning of samples on the basis of firm-year losses or profits creates discontinuous panels with gaps in the individual time-series, which makes it impossible to apply system-GMM estimation.

4. Panel data stationarity and causality

4.1. Panel data stationarity of Earnings and Returns

Work-in progress.

4.2. Panel data causality between earnings and returns
The regression of earnings on returns implies that returns cause earnings. It has been argued that the opposite also holds, that earnings may cause returns as well and therefore returns are endogenous to the regression. Ball and Kothari (2007) refute these concerns by arguing that the degree of causality in the direction of earnings to returns is negligible relative to the opposite. Specifically, they explain, since accounting is a transaction-based system then accounting earnings cannot timely proxy for economic income by contrast to stock returns that value expectations on future changes in income and therefore should incorporate forthcoming earnings releases.

This is an important empirical question that, to the best of our knowledge, it has not been addressed yet within the context of panel data analysis. To examine the direction and degree of causality between the two variables we employ the idea of Granger (1969) causality, as extended by Sims (1972) for applications in economic-based time series, and as adjusted by Hurlin and Venet (2001) for panel data structures. To test the Granger-Sims-Hurlin-Venet (GSHV) causality of returns on earnings, we modify the equation (1) to also include one forward lead $t + 1$ for returns, because next period’s returns may also affect current earnings in the sense that the impending market reaction may well affect the reporting of current earnings, as follows:

$$E_{it} = \alpha_0 + \alpha_{11}u_{it} + \sum_{\tau=1}^{g} \beta_{\tau} E_{it-\tau} + \sum_{\tau=-1}^{g} \beta_{\tau} R_{it-\tau} + u_i + s_t + v_{it}$$

(4)

The Granger (1969) causality test examines whether past values of earnings [returns] have predictive power over the current values of returns [earnings]. The test is conceptually developed around the premise that the future of one variable can affect neither the present nor the past of another variable. However, as discussed throughout the paper, it is a standard result that current prices are formed on the basis of future expected changes in accounting earnings, and accounting earnings may well reflect future expected economic losses and gains which can be approximated by market consensus expectations. Sims (1972) recognises the possibility of causality from future values in economic data and extends the idea of Granger-causality to allow for current and future values of one variable to have predictive power over the current values of another variable. Hurlin and Venet (2001) further adjust the Granger-Sims time-series causality to panel data structures and this provides several benefits to the analysis as a whole, including: (i) substantial increase to the degrees of freedom, (ii) non-dependence on lengthy time series, (iii) reduction of collinearity amongst regressors, and (iii) improvement of efficiency for testing causality (Hsiao, 2003).
where $\varepsilon_{it} = (u_i + s_t + v_{it})$. There are two differences between equation (1) and equation (4). First, the order for the distributed lag has changed to $\tau = -1, 0, 1, 2, \ldots, q \leq T_i - 1$, and second the slope parameters are now indexed by firm $i$ and indicate fixed firm-specific deterministic trends $\alpha_{1i}$, autoregressive coefficients $\rho_{it}$ and distributed lags $\beta_{it}$. Note that this modification makes redundant the recovery of time-specific effects $s_t$. Equation (4) allows for the examination of causality that may be due not only because of within-time correlation but also because of across-firm dependency (Hurlin and Venet 2001; Hurlin 2007). This approach has significant implications to our study, as we can examine whether returns [earnings] have predictive ability over earnings [returns] but also whether firms across the market share such causality patterns.  

To examine the opposite direction of causality we reverse equation (4), as follows:

$$R_{it} = \tilde{\alpha}_0 + \tilde{\alpha}_{1i} \mu_{it} + \sum_{\tau=1}^{p} \tilde{\rho}_{it} R_{it-\tau} + \sum_{\tau=-1}^{q} \tilde{\beta}_{it} E_{it-\tau} + u_i + s_t + \tilde{v}_{it}$$

(5)

where $\tilde{\alpha}_0$, $\tilde{\alpha}_{1i}$, $\tilde{\rho}_{it}$, $\tilde{\beta}_{it}$ and $\tilde{v}_{it}$ indicate estimation of reverse GSHV-causality. The estimation of equations (4) and (5) allow for the examination of four states of panel data causality (Hurlin and Venet 2001).

The first step is to assess the hypothesis of homogeneous non-causality (HNC) with no type of individual causality between the two variables for all firms $I$ and all orders $q$, as follows:

$$H_{HNC} : \beta_{it} = 0$$
$$H_{HC} : \beta_{it} \neq 0$$

$$F_{HNC} = \frac{(RSS_{HNC} - RSS_{HC})/I_q}{RSS_{HC} / (IT_i - I - I_q - q)} \sim F_{Iq,(IT_i-I-I_q-q)}$$

(6)

$\rho_{it}$ and $\beta_{it}$ are known as fixed-coefficients whereas $u_i$ is known as fixed-effects. Hurlin and Venet (2001) and Hurlin (2007) prove that the inclusion of fixed-coefficients is a sufficient condition for escaping potential econometric biases that may arise from examining panel data causality especially in panel datasets with ‘large $I$ short $T_i$’, as it is the case in our study. Also, it is assumed that earnings and returns are covariance stationary variables, that orders $p$ and $q$ are identical for all firms, and that the idiosyncratic residuals $v_{it}$ follow the standard normal assumptions for each firm individually as well as across all combinations of firms.
for $\forall i=1,2,\ldots, I$ and $\forall \tau = -1,0,\ldots,q \leq T_i - 1$, where testing is repeated for $\hat{\beta}_{i\tau}$ in the same way. The null $H_{\text{HNC}}$ is tested against its alternative $H_{\text{HC}}$ of homogeneous causality (HC) using an $F_{\text{HNC}}$ test for linear restrictions, with $I_q$ number of parameter constraints. The residual sum of squares $RSS_{\text{HNC}} = \sum_{i=1}^{I} RSS_{\text{HNC},i}$ is obtained by summing up all firm-specific residual sum of squares from $I$ number of regressions under the restriction $\beta_{i\tau} = 0$, where $RSS_{\text{HC}}$ is obtained through a within-effects estimator. If the null $H_{\text{HNC}}$ is accepted then the distributed lag variable does not GSHV-cause the depended variable for all $i$ and $q$ and testing stops here. However, if the alternative $H_{\text{HC}}$ is accepted then there is evidence of firm causality in the panel data structure.

The next step is to accept the existence of firm-specific causality and to examine the hypothesis whether firms share the same type of homogeneous causality (HC) against the alternative hypothesis of heterogeneous causality (HEC), as follows:

$$
H_{\text{HC}} : \beta_{i\tau} = \beta_{j\tau} \neq 0 \\
H_{\text{HEC}} : \beta_{i\tau} \neq \beta_{j\tau}
$$

$$
F_{\text{HC}} = \frac{(RSS_{\text{HC}} - RSS_{\text{U}})/(I-1)q}{RSS_{\text{U}}/(IT_i - 2I_q - I)} \sim F_{(I-1)q,(IT_i-2I_q-1)}
$$

for $\forall i,j = 1,2,\ldots,I$ and $\forall \tau = -1,0,\ldots,q \leq T_i - 1$. The null $H_{\text{HC}}$ of equation (8) is the same as the alternative hypothesis of equation (7), and can be assessed through an $F_{\text{HC}}$ test with $(I-1)q$ number of parameter constraints. $RSS_{\text{U}} = \sum_{i,j=1}^{I} RSS_{\text{U},ij}$ sums up all unrestricted residual sum of squares from $I$ individual estimations. If $H_{\text{HC}}$ is accepted then testing stops here, but if it is rejected then this is an indication of the existence of more than one types of

---

19 The within effects (WE) estimator transforms the variables into deviations from panel-specific means and purges the cross-sectional fixed effects $u_t$ that imply recovery of $I$ number of intercepts. The Frisch-Waugh-Lovell theorem states that the WE estimator is equivalent to the least-squares dummy variable (LSDV) estimator with $I$ number of individual intercepts (Frisch and Waugh 1933; Lovell 1963), but given the short panel structure of ‘large $I$ and small $T_i$’ with $I \rightarrow \infty$, feasibility of estimation is ensured only through a WE transformation.
heterogeneous causalities, hence implying that the depended variable’s dynamics are completely defined by firm-specific factors.

The final step is concerned with verifying whether all relationships are heterogeneous as implied by $H_{HEC}$, or if heterogeneous causality exists only for an arbitrarily chosen subgroup of firms $n < I$ while the rest of the firms $I - n$ share similar dynamic structure. To examine this hypothesis, we test the existence of homogeneous non-causality (HNC) against the hypothesis of heterogeneous non-causality (HNEC), as follows:

$$
\begin{align*}
H_{HNC} : & \beta_{it} = \beta_{jt} = 0 \\
H_{HEC} : & \beta_{it} = 0, \beta_{jt} \neq 0
\end{align*}
$$

$$Z_{HNC} = \sqrt{T} \left[ \overline{W}_{HNC} - \frac{1}{I} \sum_{i=1}^{I} q \left( \frac{T_i - 2q - 1}{T_i - 2q - 3} \right) \right] \left[ \frac{1}{I} \sum_{i=1}^{I} 2q \left( \frac{T_i - 2q - 1}{T_i - 2q - 3} \right)^2 \left( \frac{T_i - 2q - 3}{T_i - 2q - 5} \right) \right]^{1/2},
$$

for $\forall i = 1, 2, \ldots, n$, $\forall j = n + 1, n + 2, \ldots, I$ and $\forall \tau = -1, 0, \ldots, q \leq T_i - 1$.

$\overline{W}_{HNC} = \frac{1}{I} \sum_{i=1, j=n+1}^{I} W_{HNC_i}$ indicates the arithmetic average of $I$ number of individual Wald tests that examine the null $H_{HNC_i} : \beta_{it} = \beta_{jt} = 0$.\(^{20}\) Hurlin (2007) proves that the $Z_{HNC}$ test of equation (9) approximates standard normality as $I \rightarrow \infty$ irrespective of time length $T_i$.

Also, as it can be seen from the denominator of equation (9) the test is only feasible only for firms with $T_i \geq 2q + 6$. That is, each firm-panel must have minimum $T_i = 8$ observations with no gaps in the individual time-series. In addition, the estimation of firm-specific least squares

\(^{20}\) The test outlined by equation (9) seeks the existence of heterogeneous causality for only subgroup $n$, but without identifying its size. Intuitively, if $n = 0$ then causality is of type HEC whereas if $n = I$ then it is of type HNC. Hence, the acceptance of $H_{HEC}$ implies $n < I$ and therefore rejection of $H_{HEC}$ and of $H_{HNC}$. If equation (9) rejects $H_{HENC}$ and equation (8) accepts $H_{HEC}$ then causality is of type HEC. This explains the rationale of Hurlin and Venet (2001) and Hurlin (2007) for drawing a clear distinction between heterogeneity that is introduced into panel datasets because of fixed differences amongst individuals and heterogeneity in the type of causality between two variables and across individuals. This distinction is aptly captured in their own words: “In the HNC hypothesis, there does not exist any individual causality from $x$ to $y$. On the contrary, in the HC and HEC cases, there is a causality relationship for each individual of the sample. In the HC case, the data generating process is homogenous, whereas it is not the case in the HEC hypothesis. Finally in the HENC hypothesis, there is an heterogeneity of the causality relationships since there is a subgroup of $n$ units for which the variable $x$ does not cause $y$” (Hurlin 2007, p.3).
and standard errors is feasible only for firms with $T_i \geq 2 + k + 1$, where $k$ indicates the total number of slope coefficients. Thus, given $p$ and $q$, the restrictions $T_i \geq 2q + 6$ and $T_i \geq 2 + k + 1$ serve as sample screens for identifying the firm-panels with sufficient non-discontinuous length in their individual time-series.

Generally, if only one variable causes the other variable then there is unidirectional causality, whereas if both variables cause each other then we have bidirectional causality. If the latter case is verified then we examine the degree of causality using a number of selection criteria. Whatever the result, it should be noted that this does not affect the consistency of estimation of equations (1) and (3) since the applied GMM estimator relaxes the assumption of strict exogeneity between the regressors and the error term. Hence, the examination of GSHV casualty is a test for examining the direction and type of causality in a panel data context. To estimate equations (5) and (6) we extend $\tau$ to the maximum length permitted by the dataset, and examine the tests of equations (7), (8) and (9) at the 99% confidence level.

5. Analysis

Data is obtained from Compustat North for the US (NYSE, NASDAQ, AMEX) and from Compustat Global for the UK (LSE), for listed non-financial and non-public administration equities (SIC codes 0000-5999 and 7000-8999) that were active or inactive for the financial period from 1989 to 2008. Buy-and-hold discretely compounded annual stock returns $R_{it}$ are calculated at fiscal year-end using closing prices, and the earnings yield $E_{it}$ is computed as bottom line net income per ordinary share divided by the beginning of the year closing price.\textsuperscript{21} The sample is reduced to common pair-wise non-missing values for $E_{it}$ and $R_{it}$, and

\textsuperscript{21} There are mixed evidence on whether earnings should include extraordinary items or not in earnings-returns regressions (Pope and Walker 1999; Garcia Lara, Garcia Osma and Mora 2005). For this paper, following Giner and Rees (2001) and Grambovas et al. (2006) we choose to employ the ‘all-inclusive’ figure of earnings after extraordinary items. Also, we do not include dividends to the calculation of aggregate returns since dividends have an asymmetric effect on returns and earnings, in the sense that current earnings do not reflect earnings on
then filtered for values with nonsensical signs and duplicate data points. The estimation sample is further reduced because of the computation of first differences, and given the requirements of the panel data causality tests it is restricted to continuous firm-panels with minimum eight years and no gaps in the individual time series. This gives the final estimation sample of 66,214 observations that is collected over 4,932 firms (10,027 observations over 858 UK-based firms plus 51,255 observations over 4,074 US-based firms). It covers reporting dates from 31 December 1990 to 31 May 2009 for annual statements released at financial year-end for the financial period of 1991 to 2008.

5.1 Extreme values

To evaluate the benefits of the proposed generalised model over the more traditional methods, it is best to employ a sample that is robust to extreme values. This will also ensure comparability across the two jurisdictions considered. Prior literature has mostly dealt with extreme values either through winsorising (e.g., by replacing the upper and lower 1% extremes of each distribution with the respective values of the 2nd and 98th percentile), or through truncation (e.g., by discarding the upper and lower 1% extremes of each distribution). Grambovas et al. (2006) show that truncation gives rise to univariate past dividend distribution, where returns assumes the constant reinvestment of dividends and this causes severely biased coefficients (Kothari and Sloan, 1992).

The initial downloaded sample comprised of 108,718 firm-year observations. To estimate the multivariate models considered in this study the sample is reduced by 14,870 observations for pair-wise missing values for the following Compustat (CS) items: CS North – net income $ni; CS Global – income before extraordinary items $ib and extraordinary items and discontinuous operations $xido; CS North – annual fiscal price close $prcc_f; CS Global – close daily price $prcd; CS North – common shares outstanding $csho; CS Global – current common shares outstanding $cshoc, further scaled by millions. Moreover, we drop 4 observations with zero number of shares outstanding (items $csho or $cshoc) and another 9 observations with less than 1,000 shares outstanding, as well as 15 observations with zero total assets (item $at); a closer examination reveals that for these firm-years most financial statement variables are coded with zero values or with error and double-entry does not hold. We also discard 32 duplicate observations at the combined level of firm, year and stock exchange (i.e., these are excess duplicate firm-years that appear with the same listing under the same stock exchange). Finally, 10,677 observations are dropped due to the computation of first differences, and another 16,897 observations to ensure continuous panels with minimum eight years per firm-panel time series.

Christodoulou and Bradbury (2009) show that for panel data specifications that take into account the implied fixed firm-specificity it is not necessary to restrict samples of firm-year observations on the basis of financial year-end.

Extreme values (outliers) are a great concern to accounting research. Hoping to obtain robust samples, many empirical studies trim the upper and lower 1% extremes of each univariate distribution (e.g., Kothari and
frequencies that are severely distorted and fails to achieve multivariate robustness, and Christodoulou and Bradbury (2009) demonstrate that winsorising is also inappropriate for economic ratio variables, such as $E_{it}$ and $R_{it}$, because it may exacerbate the problem of extremities by creating influential modes at the imposed bounds of each univariate distribution. This may result in fat-tailedness and biased estimation. Instead, we apply an outlier filter developed by Hadi (1992, 1994) for detecting multiple extreme data points in multivariate relationships, which is known to provide considerable robustness in regression analysis (Hadi and Simonoff 1993; Durnev, Morck, Yeung and Zarrowin 2003; Grambovases et al. 2006; Christodoulou and Bradbury 2009), and described by Bradbury (2006) as “an important innovation on prior research that contributes to the literature”.

Due to panel data structure considered, Hadi’s filter is applied by jurisdiction to the multivariate relationship prior to estimation on level of firm-means, and outliers are detected at the 5% level of statistical significance. If mean annual return, $\bar{R}_{i}$, and mean earnings yield, $\bar{E}_{it}$, for firm $i$ over the firm-specific time period $T_{i}$ appears as a multivariate outlier, then the entire firm-panel is excluded. Therefore, the analysis of extreme values is performed on the basis of ‘extreme profit-making’ or ‘extreme loss-making’ firms for the time period examined (where profit and loss is perceived as either realised $E_{it}$ or expected $R_{it}$). In order to ensure robust regression results, Hadi’s filter is applied prior to estimation for each ARDL$(p,q)$ model separately, given the orders of $p = q$. For instance, a multivariate outlier for the static

---

Zimmerman 1995; Fama and French 1998; Ball, Kothari and Robin 2000; Cook, Huston and Omer 2008), or winsorise the upper and lower 1% univariate extremes (e.g., Liu and Thomas 2000; Jones et al. 2008). Others apply a combination of the two methods, e.g., Hogan and Wilkins (2008) trim the 1% extremes and then winsorise the remaining 1% extremes. These ad hoc univariate conventions are not robust against ‘masking’ and ‘swamping’ problems. That is, when one extreme value may mask the appearance of another, or a small cluster of outliers may swamp mean and inflate the variance in such a manner that another observation will appear as outlier when in fact it is not. More importantly, trimming and winsorising are univariate approaches that fail to address the problem of multivariate robustness that is of main concern to regression analysis. In contrast, the filter developed by Hadi (1992, 1994) is specifically designed for multivariate analysis (Hadi and Simonoff 1993; Hadi 2006).
ARDL(1,0) model, \( E_{it} = \alpha_0 + \beta_0 R_{it} + \epsilon_{it} \), is an extreme data point over the combination of the two variables employed as calculated on the firm-mean level.

Figure 1 demonstrates the superior robustness attained by Hadi’s multivariate filter, by contrasting the bivariate relationship \( E_{it} \) and \( R_{it} \), over the pooled sample, the univariate winsorised sample at 98% and the truncated sample to the 98% of variation. The total pooled sample of 66,214 firm-year observations is observed over 4,932 firms across the two jurisdictions, and the linear fits clearly indicate the presence of highly influential observations that leverage estimation in either direction (negative slopes for the UK -0.3017 and an inflated positive slope for the US 2.5358). The winsorised approach employs the same number of data points but for each distribution separately and per jurisdiction replaces the upper 1% with the value of the 99th percentile and the lower 1% with the value of the 1st percentile, therefore giving rise to two trimodal frequencies that ‘box-in’ the relationship. The truncated distributions eliminate the upper and lower percentiles of each univariate frequency leaving a sample of 63,955 observations observed over the same number of firms as the pooled samples. This means that truncation imposes ad hoc selected bounds of variation, ignores the bivariate relationship (as in winsorising), but more importantly it creates a number of discontinuous panels compromising therefore the estimation of the dynamic panel data models.

In contrast, Hadi’s filter for multivariate outliers identifies those firms that are extreme jointly in their current earnings yield and current market returns. Hadi’s filter leaves a sample of 61,210 data points that is observed over 4,516 firms, and does not impose any bounds to either distribution that by construction are assumed to be unbounded. Indeed, it only removes those multivariate outliers of the extreme profit-making or loss-making firms that abnormally leverage estimation and does not discard observations that lie on the path of the robust linear
fit. Hadi’s filter provides the more robust sample for the purposes of our study with slopes that now make sense (UK 0.0717, USA 0.0289).

5.2 Panel data analysis of variance

Table 1 gives the estimation samples \( N \) and the total number of firms \( I \) that are absent of multivariate outliers, for different orders of \( 0 \leq p, q \leq T - 1 \) that are used in the estimation of equations (1) and (3). Table 1 also provides a statistical summary of the variability explored through the proposed panel data specification. The Pooled standard deviation of each variable is disaggregated into the standard deviation that exists Between panel averages (i.e. variability due to differences across panels) and the standard deviation that is found Within each panel (i.e. due to average intra-panel variability). In all cases, the added variability introduced through the decomposition of variance exceeds the original pooled variability. This is strong evidence that support the existence of the compound error term described in equation (4), with error components of firm-specific and time-specific effects (Wooldridge 2008).

5.3. Estimation

Table 2 provides the system-GMM estimates for different orders of \( p = q \) for the ARDL\((p,q)\) model of equation (1). The attention is focused on the estimates of the drift \( \alpha_0 \), the instantaneous multiplier \( \beta_0 \) and the long run multiplier \( \lambda_{pq} \). Given the large number of observations \( N \) and number of firms \( I \) and in some cases quite substantial length of individual time-periods \( \max\{T_i\} \), system-GMM produces highly efficient estimates and all coefficients are significant at the 1% level. Also, note that although the GMM conditions use first-differenced errors all output is at levels, and interpretation follows from the specification of equation (1). The chi-square Wald test verifies the overall significance of all models and the residual sum of squares is minimised as the order of \( p = q \) is increased. At all instances, the
null of the Sargan test of valid overidentifying restrictions is accepted. Also, the null of no autocorrelation is either accepted at the second-order, $AR(2)$, or the third-order differences residuals, $AR(3)$, and this is an indication that the original residuals $v_i$ were $i.i.d.$ and, as expected, there is likely to be a moving average process or order one.

Findings confirm that, for firms domiciled in the UK and the US, for negative expectations on earnings ($i.e.$ negative returns conveying expectations on loss-making firms) show up in reported earnings much quicker than positive expectations on earnings ($i.e.$ positive returns conveying expectations on profit-making firms). This seems to be the case for both the concurrent returns, as indicated by the instantaneous multiplier, as well as for aggregate long-run returns, as indicated by the long run multiplier. Moreover, the reflection of negative returns shows up in an accelerated fashion as we increase the memory length reaching to a level of 73.51% for the UK and 83.31% for the US once we expand the window to 10 years in advance. These are similar results to what Easton $et al.$ (1992) with the difference that they examined they persistence of overall market returns regardless their sign.

However, once we look at positive market expectations on expected profit-making firms very little information from positive returns seems to show up in earnings. In fact, in most instances, the level of the long run multiplier (which is an indication of persistence) is not much higher than the level of the instantaneous multiplier (which indicates contemporaneous timeliness). Our results suggest an extreme form of asymmetry between positive and negative expectations for earnings but also in their final recognition in the reported book value of earnings.
References


Table 1: Panel Structure

<table>
<thead>
<tr>
<th>1990-2008</th>
<th>UK (London Stock Exchange)</th>
<th>USA (NYSE, AMEX, NASDAQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>I</td>
</tr>
<tr>
<td>$E_{it}$</td>
<td>9,714</td>
<td>756</td>
</tr>
<tr>
<td>$E_{it-1}$</td>
<td>8,554</td>
<td>718</td>
</tr>
<tr>
<td>$E_{it-2}$</td>
<td>7,595</td>
<td>689</td>
</tr>
<tr>
<td>$E_{it-3}$</td>
<td>6,369</td>
<td>626</td>
</tr>
<tr>
<td>$E_{it-4}$</td>
<td>5,318</td>
<td>567</td>
</tr>
<tr>
<td>$E_{it-5}$</td>
<td>4,350</td>
<td>499</td>
</tr>
<tr>
<td>$E_{it-6}$</td>
<td>3,676</td>
<td>466</td>
</tr>
<tr>
<td>$E_{it-7}$</td>
<td>3,162</td>
<td>450</td>
</tr>
<tr>
<td>$E_{it-8}$</td>
<td>2,626</td>
<td>394</td>
</tr>
<tr>
<td>$E_{it-9}$</td>
<td>2,148</td>
<td>346</td>
</tr>
</tbody>
</table>

Note: $N$ indicates the total number of firm-year observations and $I$ the number of firm-panels. $E_{it-\tau}$ is the earnings yield and $R_{it-\tau}$ is stock market returns, where $\tau = 0,1,2,...,T_i - 1$. Mean indicates the arithmetic average, and the Pooled standard deviation is disaggregated into Between (variability across firms) and Within (intra-firm variability).
Table 2: System-GMM estimation of the ARDL($p,q$) model of equation (1)

<table>
<thead>
<tr>
<th>$p = q$</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>0.0205</td>
<td>-0.1472</td>
<td>-0.0596</td>
<td>0.025</td>
<td>-0.0827</td>
<td>0.0169</td>
<td>0.0476</td>
<td>0.0119</td>
<td>0.0751</td>
<td>0.2012</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.0151</td>
<td>0.0246</td>
<td>0.0284</td>
<td>0.0335</td>
<td>0.0467</td>
<td>0.0725</td>
<td>0.0814</td>
<td>0.0112</td>
<td>0.1079</td>
<td>0.0939</td>
</tr>
<tr>
<td>$\lambda_{pq}$</td>
<td>0.0347</td>
<td>0.0471</td>
<td>0.0510</td>
<td>0.0528</td>
<td>0.1099</td>
<td>0.1059</td>
<td>0.1405</td>
<td>0.1795</td>
<td>0.1434</td>
<td>0.1823</td>
</tr>
<tr>
<td>$N$</td>
<td>6,900</td>
<td>6,227</td>
<td>5,145</td>
<td>4,482</td>
<td>3,965</td>
<td>3,538</td>
<td>2,932</td>
<td>2,585</td>
<td>2,166</td>
<td>1,789</td>
</tr>
<tr>
<td>$I$</td>
<td>564</td>
<td>552</td>
<td>490</td>
<td>442</td>
<td>392</td>
<td>354</td>
<td>312</td>
<td>278</td>
<td>253</td>
<td>228</td>
</tr>
<tr>
<td>$Z$-rank</td>
<td>302</td>
<td>301</td>
<td>296</td>
<td>290</td>
<td>279</td>
<td>267</td>
<td>253</td>
<td>238</td>
<td>220</td>
<td>200</td>
</tr>
<tr>
<td>min($T_i$)</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>max($T_i$)</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Wald</td>
<td>3120.6</td>
<td>3779.4</td>
<td>6632.8</td>
<td>9299.2</td>
<td>13890.0</td>
<td>15555.0</td>
<td>9991.4</td>
<td>13794.1</td>
<td>19300.7</td>
<td>18556.5</td>
</tr>
<tr>
<td>RSS</td>
<td>296.56</td>
<td>252.80</td>
<td>177.70</td>
<td>165.00</td>
<td>128.78</td>
<td>64.06</td>
<td>37.39</td>
<td>27.98</td>
<td>13.84</td>
<td>9.77</td>
</tr>
<tr>
<td>AR(2)</td>
<td>2.1167</td>
<td>1.2642</td>
<td>0.2577</td>
<td>0.2779</td>
<td>0.4728</td>
<td>0.9709</td>
<td>1.2518</td>
<td>0.5107</td>
<td>0.2459</td>
<td>0.0412</td>
</tr>
<tr>
<td>$R$</td>
<td>1402.77</td>
<td>1452.4</td>
<td>17,971</td>
<td>0.0576</td>
<td>2.999</td>
<td>2,999</td>
<td>2,999</td>
<td>2,999</td>
<td>2,999</td>
<td>2,999</td>
</tr>
<tr>
<td>$T$-rank</td>
<td>347.20</td>
<td>347.54</td>
<td>325.58</td>
<td>296.10</td>
<td>294.64</td>
<td>268.87</td>
<td>252.45</td>
<td>226.76</td>
<td>201.15</td>
<td>150.96</td>
</tr>
<tr>
<td>min($T_i$)</td>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>max($T_i$)</td>
<td>3279.12</td>
<td>1979.80</td>
<td>1400.72</td>
<td>1007.11</td>
<td>568.35</td>
<td>264.02</td>
<td>118.34</td>
<td>76.99</td>
<td>52.40</td>
<td>32.96</td>
</tr>
<tr>
<td>Wald</td>
<td>3790.5</td>
<td>1452.4</td>
<td>1377.16</td>
<td>1377.36</td>
<td>1579.6</td>
<td>1102.7</td>
<td>1413.8</td>
<td>1213.7</td>
<td>26105.3</td>
<td>24004.37</td>
</tr>
<tr>
<td>RSS</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
<td>6995.81</td>
</tr>
<tr>
<td>AR(2)</td>
<td>0.3838</td>
<td>-1.5436</td>
<td>-1.5964</td>
<td>-0.3159</td>
<td>0.6204</td>
<td>2.1922</td>
<td>-4.8914</td>
<td>-0.8230</td>
<td>1.7384</td>
<td>0.0430</td>
</tr>
<tr>
<td>AR(3)</td>
<td>1.5425</td>
<td>2.0232</td>
<td>1.0973</td>
<td>0.0166</td>
<td>-1.5321</td>
<td>-2.8830</td>
<td>-0.5438</td>
<td>-1.1397</td>
<td>-0.8663</td>
<td>-1.0468</td>
</tr>
<tr>
<td>Sargan</td>
<td>622.11</td>
<td>0.0068</td>
<td>3,205</td>
<td>0.0178</td>
<td>326</td>
<td>326</td>
<td>326</td>
<td>326</td>
<td>326</td>
<td>326</td>
</tr>
</tbody>
</table>

Note: $\alpha_0$ indicates the drift, $\beta_0$ the instantaneous multiplier and $\lambda_{pq}$ the long run multiplier per ARDL($p,q$) model. $N$ is the total number of firm-year observations and $I$ the number of firm-panels. min($T_i$) and max($T_i$) indicates the minimum and maximum length of continuous years for a firm-panel. Wald indicates a chi-square test of whether all $p$ parameters are equal to zero. RSS is the residual sum of squares. AR(1), AR(2), AR(3) indicate the $z$-values of the Arelano and Bond (1991) autocorrelation tests for orders 1, 2 and 3 ($z$-values $> 2.58$) indicate rejection of
Figure 1: Extreme values and robust samples

Note: The scatter graphs plot the multivariate data points between the $E_{it}$ and $R_{it}$, i.e. the two variables employed by the static ARDL(0,0) model, over the Pooled sample, the Univariate winsorised sample at 98%, the Univariate truncated sample to 98%, and the sample remaining following Hadi’s filter on multivariate outliers applied at the 5% of statistical significance. To assist visual representation, for the pooled sample the $y$-axis is restricted to $-2 \leq E_{it} \leq 2$ and the $x$-axis to $-1 \leq R_{it} \leq 6.5$. The vertical and horizontal dashed lines in the scatter graph of the Univariate truncated sample to 98% (bottom-left graph) indicate the bounds imposed by truncation. The superimposed lines in all graphs indicate linear fits an OLS regression $E_{it} = a + bR_{it} + e_{it}$ for the UK (GBR) and the USA. $b$ indicates the slope of the linear fit, $N$ the total number of firm-year observations and $I$ the total number of firms.
Figure 2: System-GMM estimation of instantaneous and long run multipliers

Note: The x-axis indicates the ARDL($p,q$) model estimated through the system-GMM estimator with GMM-instruments, and the y-axis indicates the estimated coefficient for the instantaneous multiplier and the long run multipliers. The left hand side graph indicates estimation on firm-panels with negative returns on average and the right hand side graph indicates estimation firm-panels with positive returns on average. All coefficients are significant at the 1% level.