Transitory Noise in Reported Earnings: Implications for
Forecasting and Valuation

by

James Ohlson¹ & Kenton K. Yee²

April 2009

¹ Professor, Stern School of Business, New York University, New York, N.Y., 10027
² Mellon Capital Management, San Francisco, CA 94105

The statements and opinions expressed in this article are those of the authors and do not necessarily
represent the view of the Bank of New York Mellon, Mellon Capital Management, or any of their affiliates.
This article does not offer investment advice.
Abstract

We consider a setting where the present value of expected dividends determines price and the information to forecast the future includes reported earnings. In the model reported earnings have been garbled by transitory noise, which cannot be inferred. "True", but now unobservable, earnings are permanent as in Ohlson (1992). We first show that capitalized expected reported earnings for the next period equals price regardless of the degree of noise. More subtle is the influence of current reported earnings on the forecast of future expected earnings. Because of the noise term, Bayesian updating implies that the investor uses the entire earnings history to learn about permanent earnings and to forecast future expected reported earnings. Specifically, we show that the next period's expected earnings equals a weighted average of (i) current reported earnings and (ii) beginning-of-the-period expected earnings for the current period. (This framework is often referred to as adaptive expectations.) Using this result we show that the weight on current earnings (term (i)) decreases as the noise increases. The model has testable implications for returns on earnings regressions and how one conceptualizes value-relevance.
I. Introduction

Graham and Dodd (1940), in their classic "Security Analysis," argued that successful equity valuation should focus on assessing a firm’s "long run sustainable earnings" or what is nowadays referred to as permanent earnings. The framework intrigues because it relies on an ideal earnings construct to evaluate whether a stock is over or under-priced. An ideal earnings construct can of course never be observed directly. At best it can be estimated. To convert the idea into practice, guiding principles in Graham and Dodd suggest that reported earnings need to be adjusted for more or less transitory items to approximate permanent earnings. A further refinement of the process considers past changes in reported earnings to reflect that such earnings generally evolve less smoothly than permanent earnings. Hence one needs to smooth past earnings to find the trend. An earnings history can thereby inform an investor concerned with approximating permanent earnings.

This paper deals with questions related to Graham and Dodd’s approach to equity valuation. How does the presence of (unobservable) transitory noise in reported earnings influence forecasting and valuation? How does this setting compare with the special case when there is no noise at all? Reported earnings may be worthwhile to forecast, and they may be usefully informative but something needs to be done to deal with the noise. These issues naturally arise in practice and empirical research, and thus they should be of interest.

Two critical assumptions underpin our model. First, there is an ideal but unobservable construct of earnings, namely permanent earnings (Ohlson (1995)). Second, the observable reported earnings equal permanent earnings plus unobservable transitory
noise. Combined with a PVED requirement, the two assumptions are shown to imply that price equals next period’s expected reported earnings capitalized. We further show how an investor learns from reported earnings to forecast future earnings: the next period’s expected reported earnings are essentially determined by a weighted average of (i) current (reported) earnings and (ii) the expected earnings for the current(!) period that prevailed at the beginning of the current period. There is also a relatively minor third term that adjusts for an effect due to retained earnings. A full payout eliminates the last term: If $e_{t+1}$ denotes expected earnings at date $t$ for period $t+1$ and $x_t$ denotes period $t$ earnings then $e_{t+1} = w \cdot x_t + (1 - w) \cdot e_t$. This recursive forecasting scheme derives from Bayesian updating in light of new information. Kalman Filter techniques supply the analytical foundation to derive the result. We then show that $w$ is small when the variance of the transitory component of $x_t$ is large. Noisy earnings therefore means that the current earnings are relatively less informative as compared to the history of earnings prior to the current period.

Dividends are present in the model in addition to the (unobservable) permanent earnings and reported earnings. The structure of the basic model, however, ensures that the price, PVED, does not depend on the dividend policy. The model exploits this dividend policy irrelevance property to shift the analysis away from future dividends to future permanent earnings. Thus the model distinguishes the distribution of wealth from its creation, though the latter variable cannot be observed.

The theoretical analysis helps to frame certain empirical questions. Consider a regression where analysts’ (consensus) expected earnings for the forthcoming year is the dependent variable. How should it be explained? In other words, how should one
conceptualize the RHS of the regression? Given the dependent variable one naturally introduces accounting variables observed at the end of the current period, such as reported earnings. But the theory here also suggests that it makes sense to add the expected value of the current period's earnings at the beginning of the current period. Thus the RHS of the regression is split into "new" information and "old" -- but still relevant -- information. Moreover, in this regression relatively more weight will be placed on "old" information for firms with relatively "noisy" new information. This kind of analysis is clearly doable, and it offers a new paradigm on how one can assess the value relevance of accounting data.

The analysis also bears on the traditional value relevance studies where the market return is the dependent variable. We show that the more noise in earnings, the lower is the earnings response coefficient (ERC). Thus it follows that the modeling here build in the classical errors-in-variables econometrics. We believe that no paper has shown how this conclusion derives from a setting in which the valuation rests on the concept that values is determined by the present value of expected dividends.

II. Setup

Our model rests on a multiple periods valuation framework with well-defined information and objective probabilities. Like most valuation models, we deal with how one represents price as it relates to earnings, when the present value of expected dividends (PVED) determines the price. But contrary to the prior literature, in our modeling the investor visualizes the future by looking at the history of reported earnings
and dividends: Prior periods’ information remains relevant and current realizations of earnings/dividends update an investor’s expectations.

We use the following notation:

\[ d_t = \text{dividends, date } t \]
\[ x_t = \text{earnings, period } t \]
\[ p_t = \text{price, date } t \]

\[ R = 1 + r = \text{the discount factor or one plus cost of capital, an exogenous constant} \]

\[ \Omega_t \equiv \{d_0, \ldots, d_t; x_0, \ldots, x_t\} = \text{the information at date } t \]

The symbol \( \Omega_t \) underscores that the relevant information refers to the history of realized dividends and earnings. This feature differs from models like Ohlson (1995), Feltham and Ohlson (1995) and the valuation literature discussed in Feltham and Christensen (2003). (These models do not depend on the entire history being relevant).

Unless we indicate otherwise, the information is understood to be \( \Omega_t \). To simplify the notation, we therefore write \( E[\tilde{x}_{s+t} | \Omega_t] = E_t[\tilde{x}_{s+t}] \) and \( E[\tilde{d}_{s+t} | \Omega_t] = E_t[\tilde{d}_{s+t}] \) and similarly for the expectation of any other random variable (like \( E[\tilde{p}_{r+1} + \tilde{d}_{r+1} | \Omega_r] = E_t[\tilde{p}_{r+1} + \tilde{d}_{r+1}] \)).

Our first assumption is standard. Price equals the present value of expected dividends:

\[ p_t = \sum_{r=1}^{\infty} R^{-r} E_t[\tilde{d}_{s+t}] \quad \text{[PVED]} \]

Our second assumption specifies the dynamic for dividends. The dividend-forecast is a function of the dividend policy and the most recent information:
\[ \tilde{d}_{t+1} = \pi_1 x_t + \pi_2 d_t + \tilde{u}_{t+1} \]  \hspace{1cm} \text{[DD]}

where \( \pi_1 (\pi_1 \neq 0) \) and \( \pi_2 \) are the fixed dividend policy parameters. The \( u_t \) are zero mean unpredictable disturbance terms. We place no other restrictions on the disturbance terms’ distributions (variances, for example, may depend on earnings or dividends).

The DD assumption can be generalized to allow for non-linearity and other complexities on the RHS. But these generalizations would not be worth it insofar they merely introduce “house-keeping” issues. The totality of assumptions -- which includes a third one stated below -- will ensure that the class of dividend policies (DD) work such that neither \( \pi_1 \) nor \( \pi_2 \) influence conclusions. That said, note that the forecasting of dividends must depend on reported earnings; thus we maintain the regularity condition \( \pi_1 > 0 \).

Our third assumption, which specifies the earnings’ dynamic, introduces the model’s main innovative concept. This dynamic revolves around the permanent earnings construct \( x^* \). The innovation is that investors cannot observe \( x^* \); that is \( \Omega_t \) excludes \( x^* \). Reported earnings – which investors do observe -- evolve stochastically according to two dynamic equations:

\begin{align*}
\tilde{x}_t &= \tilde{x}^* + \tilde{\delta}_t \\
\tilde{x}^*_{t+1} &= R \tilde{x}^* - \delta_t + \tilde{\epsilon}_{t+1} \\
\end{align*}

\hspace{1cm} \text{[ED]}

where \( \tilde{\delta}_t \sim N(0, \sigma_{\delta}^2) \), \( \tilde{\epsilon}_t \sim N(0, \sigma_{\epsilon}^2) \) are i.i.d. normally distributed disturbance terms and \( \text{cov}(\tilde{\delta}_t, \tilde{\epsilon}_t) = 0 \). We also assume \( \text{cov}(\tilde{u}_t, \tilde{\epsilon}_t) = 0 \).

To appreciate the assumption ED, a number of observations help.
(i) Using traditional terminology, one refers to $x_t^*$ as a latent, or “hidden”, variable because it cannot (in general) be observed. In our model, $x_t^*$ defines permanent earnings. This terminology reflects that the retention of earnings alone accounts for the growth 

\[
( E_t[\tilde{\delta}_{t+1}] / x_t^* = 1 + r \cdot (1 - d_t / x_t^*)), \]

which equals to one when there are no retained earnings.

(ii) Combined with DD, ED implies a sequence of $\{E_t[d_{t,r}]\}_t$ conditional on the information $\Omega_t$. This setup means that PVED, ED, DD, and $\Omega_t$ yield some reduced form valuation function $p_t = p(\Omega_t)$.

(iii) The setting when there is no noise in earnings has been previously developed by Ohlson (1995). It serves as a benchmark for the more general setting (which turns out to be substantially more complicated). If $\delta_t = 0$ for sure, then, trivially, $x_t^* = x_t$ and $x_t^*$ becomes observable. Combined with PVED and DD, this observable permanent earnings setting implies

\[
p_t = E_t[\tilde{\delta}_{t+1}] / r
\]

and

\[
p_t = (R / r)x_t^* - d_t.
\]

A strictly positive variance term, $\text{var} (\tilde{\delta}_t) > 0$, or $x_t^* \neq x_t$, changes the model radically. Though the capitalized expected earnings relation perhaps remains valid (the issue is not obvious), the second equation cannot be true since the price, $p_t$, depends solely on information that excludes $x_t^*$ (and $x_t^*$ cannot be inferred from $\Omega_t$). Hence the question arises as to how one identifies the function $p(\Omega_t)$ which replaces $(R / r)x_t^* - d_t$. 


(iv) The i.i.d. random variable \( \delta \) specifies the transitory part of earnings. This component of earnings is of course unobservable; otherwise one could infer permanent earnings, \( x^* \). An investor accordingly forecasts and infers value in a setting with incomplete information due to the transitory noise component in earnings. It makes sense to hypothesize that the information content of (reported) earnings should increase as the noise decreases. One should further expect the earnings response coefficient (i.e., the usual ERC as it shows up in returns on earnings regressions) to increase as the noise decreases. These issues will indeed be addressed.

(v) Value creation uncertainty shows up because of the second source of uncertainty, \( \epsilon \). We assume that the \( \epsilon \) have a strictly positive variance. It avoids the boundary case when \( x^* \) can be inferred from the history of dividends and the initialization condition \( x_0 = x_0^* \). (An assumption of \( x_0 = x_0^* \) is not necessary in our model. Still one can argue that at the firm’s inception \( x_0 = x_0^* = 0 \) and \( -d_0 > 0 \) specifies the start-up capital contribution).

(vi) The normality assumption on the two noise terms in ED leads to Bayesian revisions that are linear in the observables.

(vii) The requirement that the two noise terms in ED do not correlate could be relaxed. However, doing so would have led to a more elaborate analysis with the thrust of the conclusions being the same.

(viii) It may be tempting to hypothesize that DD can be modified by letting \( x^* \), no less than \( x_\alpha \), appear on the RHS in DD. But this generalization changes the model non-trivially. (This is yet another point we will return to.)
The three assumptions – PVED, DD, and ED -- provide all of the model’s ingredients. We now turn to the insights they yield.

III. Results

We first show that reported earnings works as an *ex ante* valuation attribute: the capitalization of expected earnings equals price. The result relies on a subtle property of \( x_t \). Given a forward-looking perspective, in expectation the \( x_t \) behave as if they are permanent. Further, because \( E_t[\tilde{x}_{t+1}] = 0 \), the capitalization extends to permanent earnings as well as reported earnings.

**Proposition 1:** Consider PVED, DD, and ED combined with the mild regularity condition

\[
\max \text{ root of the matrix } \begin{pmatrix} R & -r \\ \pi_1 & \pi_2 \end{pmatrix} < R. \text{ Then}
\]

\[
p_t = E_t[\tilde{x}_{t+1}] / r = E_t[\tilde{x}_{t+1}] / r.
\]

*Proof:* See appendix.

The proof exploits the expectational equivalence of \( x_t \) and \( x_t^* \) in that

\[
E_t[\Delta \tilde{x}_{t+\tau}^* - r \cdot (\tilde{x}_{t+\tau-1}^* - \tilde{d}_{t+\tau-1})] = E_t[\Delta x_{t+\tau} - r \cdot (\tilde{x}_{t+\tau-1} - \tilde{d}_{t+\tau-1})] = 0
\]

as long as \( \tau \geq 2 \). In the expression the “= 0” part reflects the permanence of the \( x_t^* \), and thus the \( x_t \) are no less permanent in expectation. \(^4\) (For \( \tau = 1 \) the first equality breaks down because \( x_t \) is observable while \( x_t^* \) is not, given the conditioning information \( \Omega_t \).

Though this aspect is irrelevant in Proposition 1, it plays a role later.)
The parameters $\pi_1$ and $\pi_2$ do not influence Proposition 1 (setting aside the regularity condition). A certain robustness in the conclusion is therefore present. That said, we have not yet proved that $p(\Omega_t)$ is independent of $\pi_1$ or $\pi_2$.

The next result articulates how the information $\Omega_t$ forecasts next period’s expected earnings. With this conclusion in place, one also obtains $p(\Omega_t)$ via Proposition 1.

**Proposition 2:** The earnings and dividends dynamics ED and DD with $\sigma_\delta, \sigma_\epsilon > 0$ imply

$$E_t[\tilde{x}_{t+1}] = R(w_t x_t + (1 - w_t) E_{t-1}[\tilde{x}_t]) - rd_t,$$

where $1 > w_t > 0$. The sequence $\{w_t\}_{t=1}^\infty$ evolves according to a difference equation

$$w_t = h(w_{t-1}; R, \sigma_\delta, \sigma_\epsilon)$$

independent of the dividend policy parameters $\pi_1$ and $\pi_2$.

**Proof:** See appendix.

An application of Kalman Filter techniques proves the proposition. Such techniques derive Bayesian revisions about future expected outcomes in light of the recent information. Though elaborate, these revisions can be worked out in a ($\Omega_t$-conditioned) linear framework because the $\epsilon$ and $\delta$ terms satisfy normality.

To interpret the earnings forecasting dynamic, recall that for observable $x^*_t$ the permanent earnings dynamic satisfies

$$E_t[\tilde{x}_{t+1} | x^*_t, d_t] = Rx^*_t - rd_t.$$ If the information $\Omega_t$ -- which excludes $x^*_t$ -- replaces the ideal information $x^*_t$ and $d_t$, then the last proposition shows that one replaces $x^*_t$ with a weighted average of (i) the current earnings, $x_t$, and (ii) those earnings that the investor *expected at the beginning of the period*, $E_{t-1}[\tilde{x}_t]$. In
this way the investor also updates her best assessment of permanent earnings since

\[ E_t[\tilde{x}_{t+1}] = E_t[\tilde{x}_{t+1}] . \]

A particularly simple expression obtains if the payout satisfies

\[ d_t = w x_t + (1 - w) E_{t-1}[\tilde{x}_t] . \]

On average, the scheme is a 100% payout ( \( E_{t-1}[\tilde{d}_t] = E_{t-1}[\tilde{x}_t] \)).

Now

\[ E_t[\tilde{x}_{t+1}] = w_t \cdot x_t + (1 - w_t) \cdot E_{t-1}[\tilde{x}_t] . \]

For any set of weights \( \{w_t\}_{t=1}^\infty \), a direct backward recursive substitution shows that

\[ E_t[\tilde{x}_{t+1}] \]

depends on the entire history of past earnings.

We next consider the properties of the weights \( w_t \). These require a time-subscript because Bayesian updating of \( E_t[\tilde{x}_{t+1}] \) includes an updating of the weight \( w_t \). Implicit in the analysis is some unspecified (exogenous) prior distribution for \( x_t^\ast \). However, the prior distribution ceases to be relevant as \( t \) becomes large. This observation conforms to dynamic Bayesian models in which, as times passes, one learns more and more about the true structure that generates the data. In analytical terms, the sequence generated by \( h(w_t) = w_{t+1} \) approaches a stationary point \( \lim_{t \to \infty} h(w_t) = w \). It turns out that the properties of \( h(w_t) \) are such that one solves for the stationary point \( w = h(w) \) via a quadratic equation.

Thus one derives \( w = f(\theta, R) \) where \( \theta = \frac{\sigma_\delta^2}{\sigma_\varepsilon^2} \) so that only the variance ratio affects \( w \).

(The quadratic solution equals \( w = 0.5 R^2 \left[ (R^2 + 1 + \theta) - \sqrt{(R^2 + 1 + \theta)^2 - 4R^2} \right] \). This equation reveals how the weight \( w \) depends on the exogenous parameters, \( \sigma_\varepsilon \) and \( \sigma_\delta \).
Corollary 1: The solution \( \lim_{t \to \infty} h(w_t) = w \) exists for any \( 0 < w_0 < 1 \). Moreover, \( w \) decreases as \( \sigma_\delta^2 \) increases and it increases as \( \sigma_\epsilon^2 \) increases.

The result is intuitive. It says that the relative importance of updating for the most recently observed earnings depends upon the underlying noise in the earnings; the less the noise, the greater the weight assigned to current earnings (keeping the economic uncertainty fixed). In a similar vein, \( w \) increases as the economic uncertainty increases because it makes the variances in the noise relatively unimportant.

As to the dividend policy, it cannot affect the weight because it is unrelated to the resolution of the economic uncertainty (\( \epsilon \)) or the noise (\( \delta \)). Indeed, DPI obtains because of not only the assumption on earnings, ED, but also the assumption on dividends, DD. The latter does not include \( x^* \) on the RHS of the dynamic. (The next section shows why putting \( x^* \) on the RHS changes the problem).

Two additional properties of \( w \) are worthwhile to discuss. Both are subtle and perhaps not so apparent.

First, the weight on earnings satisfies the lower bound \( 1 - 1/R^2 > 0 \) (given \( R > 1 \)) as the variance ratio \( \frac{\sigma_\delta^2}{\sigma_\epsilon^2} \) approaches infinity. In effect, regardless of the noise, earnings are always informative. This result is by no means obvious because at the other extreme \( w \) approaches one as \( \frac{\sigma_\delta^2}{\sigma_\epsilon^2} \) approaches zero. But \( R > 1 \) implies that on average value is created as time passes, and the only way to learn about it is through earnings, no matter how noisy.
Second, the weight \( w \) on earnings increases as the discount factor \( R \) increases (which can be conjectured on the basis of the previous observation). This makes sense if one invokes the following heuristic argument. An increase in economic uncertainty, \( \sigma_e \), increases the weight \( w \) (as noted earlier). But an increase in economic uncertainty should also increase the risk, which is part of what should go into \( R \). Hence, one can expect the signs of \( \partial w / \partial R \) and \( \partial w / \partial \sigma_e \) to be the same, which is the case. Though this line of reasoning goes beyond the formal model, it has some economic plausibility by connecting the discount factor to the economic risk in earnings.

Proposition 2 combined with PVED, i.e., the two propositions taken together, lead to the valuation function \( p(\Omega_t) \):

\[
p_t = (R / r)[w \cdot x_t + (1 - w) \cdot E_{t-1}[\tilde{x}_t]] - d_t.
\]

The expression generalizes the permanent earnings case, \( p_t = (R / r)x^*_t - d_t \) by replacing \( x^*_t \) with the estimate \( E_t[\tilde{x}^*_t] = w \cdot x_t + (1 - w) \cdot E_{t-1}[\tilde{x}_t] \). In other words, for the general setting with \( \Omega_t \),

\[
p_t = \frac{E_t[\tilde{x}^*_t]}{r} = \frac{R}{r} E_t[\tilde{x}^*_t] - d_t.
\]

Though the variable \( x_t \) is not explicit in the last expression, it is part of the information used to determine \( E_t[\tilde{x}^*_t] \) and \( E_t[\tilde{x}^*_t] \). A cohesive set of ideas thereby show how an investor learns about permanent earnings from \( \Omega_t \), which in turn determines the estimate of the present value of subsequent expected dividends. The details of the sequence \( E_t[\tilde{d}^*_t], E_t[\tilde{d}^*_{t+1}], \ldots \) need not concern an investor because of the
model’s DPI property, precisely because permanent earnings is an ideal earnings construct.

Given the expression for value we next derive the regression that explains the market return, \( ret_t = (p_t + d_t) / p_t \), in terms of unexpected earnings (normalized by start-of-period price). While the result is interesting in its own right, it also bears on the large number of empirical studies of the returns on earnings regression.

**Corollary 2:** The assumptions ED, DD and PVED imply

\[
\tilde{ret}_t = \alpha + \beta [\tilde{x}_t - E_{t-1}[\tilde{x}_t]] / p_{t-1}
\]

where \( \alpha = R \) and \( \beta = (R / r) \cdot w \).

The fact that the intercept equals \( R \) is unsurprising since PVED implies that the expected return equals \( R \) (and unexpected earnings has zero mean, of course).

The intuition behind the slope-coefficient, in contrast, depends directly on the transitory noise in earnings. The dynamic ED can be thought of as an error-in-variables model, where the relative variance \( \frac{\sigma_x^2}{\sigma_e^2} = \theta \) determines the extent of the error. Hence, the last corollary shows that, since \( w \) decreases as \( \theta \) increases, the slope coefficient \( \beta \) decreases as the relative error in earnings increases.

Standard econometrics teaches that the slope-coefficient in a univariate regression decreases if the RHS variable is measured with error. Corollary 2 conforms to this idea if one keeps in mind that \( x_t - E_{t-1}[\tilde{x}_t] \) acts as an imperfect measure of the “ideal” measure \( x_t^* - E_{t-1}[\tilde{x}_t^*] \). Only for the latter case will the slope coefficient equal to \( R/r \), which (for reasonable values of \( R \)) is far greater than that obtained in actual regressions (as has been
long known). Thus the discussion brings out that the error-of-measurement issue on the RHS may not only depend on an erroneous measurement of the “market’s” expectation of the forthcoming earnings, as is commonly argued, but instead on the earnings construct itself. The response coefficient $R/r$ depends on earnings being perfectly aligned with *permanent* earnings. If there is an error in this regard --- due to transitory effects in the accounting earnings measure -- then there is another source of error that reduces the slope coefficient. The point is not new, but we do not believe it has been formalized this way previously.\footnote{\textsuperscript{6}}\footnote{\textsuperscript{7}}

Our final comment in this section bears on the distributional assumption on $\epsilon_t$ and $\delta_t$ in the earnings dynamics ED. These random variables have a normal distribution to ensure analytical tractability. Bayesian analysis now implies that the forecast $E_t[\tilde{x}_{t+1}]$ is linear in the information $\Omega_t$. Thus normality should be viewed as a convenience assumption. That said, one can also motivate the linear forecast by stipulating that the forecast must be the minimum variance linear forecast (which is the property that text-books discussing Kalman filter techniques most often stress).

**IV. The Dividend Distribution Policy and Value Creation**

Our model sheds light on value creation as opposed to its distribution in a Bayesian framework. As noted previously, the DD assumption on the forecasting of dividends aligns with the DPI property, i.e., the policy parameters $\pi_1$ and $\pi_2$ do not influence the weights in the value function. Given this fact, they have no influence either on the forecast $E_t[\tilde{x}_{t+1}]$ or the value function $\rho(\Omega_t)$. There is an aesthetic neatness because noise in earnings does not perturb the DPI inherent in the permanent earnings model.
Noise in earnings causes no problems because, whatever the parameters of a dividend policy are, dividends do not bring to bear on the creation of value: Nothing can be learned about the $\epsilon_i$ from knowing the policy. After all, DD ultimately represents past earnings and pure noise. Thus one sees that DPI ceases if the investor can learn about the $\epsilon_i$ from dividends. As an extreme example, suppose that $d_i = \text{const} \cdot x^*_i + \text{noise}$; clearly, any evaluation of the weight $w_i$ depends no less on the $\text{const}$ and the variance of the noise. More generally, DPI does not apply if $x^*_i$ had appeared on the RHS in the dividend dynamic or if the noise in the DD correlates with the disturbance terms in the ED.

Because all of $\Omega_i$ influences the forecasting, it is of course true that the entire history of dividends influences the value function. The relevance of dividends in $\Omega_i$ reflects that the effect of retained earnings on the forecasting of (permanent) earnings cannot be neglected. Stated somewhat differently, if one forecasts $x^*_{t+1}$ plus earnings foregone due to past dividends, $r \sum_{\tau=1}^{t} d_{t-\tau} R^\tau$, then neither the policy parameters nor the past dividends will be relevant as long as one cannot learn about $\epsilon_i$ from the dividend dynamic.

The above discussion of the absence/presence of DPI is not new. In many ways it can be traced to 1960’s ideas that managers use dividends to convey information about the underlying "true" or future reported earnings (e.g., see Pettit (1972) or finance textbooks like Brealey and Myers (2002)). And now it helps to understand the dividend policy to explain returns or value. The modeling here thus formalizes what the literature has long hypothesized as a possibility.
V. Concluding Remarks

From a conceptual perspective, the results in this paper try to provide insights related to the construct of permanent earnings. The construct poses problems, it can be argued, because permanent earnings are unobservable no matter how one modifies GAAP, i.e., the construct exists only in the realm of “theory” with its negative connotation. The developments here modify this argument: Reported earnings can take on a central role as information to approximate permanent earnings. In this framework, transitory items act as noise in financial statement analysis, and the name of the game becomes one of trying to find some earnings number which can serve as a starting point to forecast next period’s earnings. Moreover, the reasoning is consistent with analysts’ perception that the forecasting of earnings is a focal point of equity valuation. Yet such focus on reported and ideal earnings is fully consistent with the concept that PVED determines value at all point in time. Hence the model distinguishes value creation from its distribution, and it even recognizes the importance of dividends not providing information on the noise in reported earnings.

We believe that viewing analysts’ forthcoming expected earnings as a dependent variable provides a paradigm of how to study value relevance empirically. The modeling here further suggests that it makes sense to put the dependent variable lagged by one year on the right hand side in the regression, and then ask what current period accounting data also explains (analysts’) expected earnings. A focus on explaining expected forthcoming earnings leads to the role of special items, indicators of the quality of earnings, explaining cash flows versus accruals, and a whole slew of issues familiar from recent accounting
research. And the current analysis brings out that the results from such regressions bear on the returns to earnings relation. Specifically, if some measures of reported earnings play a modest incremental role in explaining the next period’s expected earnings, then these earnings should not contribute much to the returns to earnings regression either. In sum, one can ask questions about what accounting information -- with specific focus on various measure of earnings -- analysts use to update their forecasts and how this information must bear on value and returns in a consistent fashion.\textsuperscript{10}
Appendix:

Proof of Proposition 1: Let $y_t = x_t^* / r$ and $z_t = y_{t+1} + d_t - R y_t$. Then

$$p_t = E_t[\tilde{y}_{t+1}] + \sum_{\tau} R^{-\tau} E_t[\tilde{z}_{t+\tau}] .$$

The result follows because ED implies $E_t[\tilde{z}_{t+\tau}] = 0$ and the regularity condition guarantees that no variable grows at rate faster than $R^\tau$. Also,

$$E_t[\tilde{x}_{t+1}] = E_t[\tilde{x}_{t+1}] .$$

Proof of Proposition 2:

The proof relies on Sargent (1987), pages 230 and 231. The only difference pertains to how to deal with dividends.

Let $y_t = r \sum_{\tau=1}^t R^{-\tau} d_{\tau}$; thus $y_t$ defines the implicit earnings due to past dividends. Let $x_t^\omega = x_t^* + y_t$ and $x_t = x_t + y_t$. The model $(x_t^*, x_t^\omega)$ satisfies the one in Sargent (1987), i.e., his equation (16) and (17), with $c = 1$, $\rho = R$, $\theta_t = x_t^\omega$, and $z_t = x_t$. Due to DD, there is no information in $d_t, d_{t-1}$ as one determines the best linear projection $E_t[\tilde{\theta}_{t+1}]$ as a function of $\Omega_t$. Now it follows from Sargent, expression (21) and (22) where $K_t = w_t / R$, that $E_t[\tilde{x}_{t+1}] = R [w_t x_t^\omega + (1 - w_t) E_{t-1}[\tilde{x}_{t}] ]$. Noting that $E_t[\tilde{x}_{t+1}] = E_t[\tilde{x}_{t+1}]$ and plugging in the definitions for $x_t^\omega$ and $x_t$ the result follows, including $1 > w_0 > 0$ from (21).
References


Ohlson, J. A. (1992), ‘The Theory of Value and Earnings, and an Introduction to the Ball-

*Contemporary Accounting Research* 11 (2), 661-687.

145-162.

Hanover.


Errors in Accounting Earnings’. Doctoral Dissertation, Graduate School of
Business, Stanford University.

edition.

Accounting Research* 12 (1), 41-55.

Research* 29 (1), 125-149.
Endnotes


2 The model here differs from Ohlson (1999). This model allows for (idiosyncratic) transitory noise that informs if observed, but this noise is never part of reported earnings.

3 The literature often refers to this weighted average scheme as "adaptive expectations" or as "exponential smoothing". In spite of the vast accounting literature on earnings forecasting, it seems to never have been applied. However, 35 years ago Elton and Gruber (1972) published an empirical paper in Management Science comparing this approach to competing "mechanical" forecasting models. They suggest adaptive expectations worked the best. It may further be noted that Elton and Gruber refer to "normal" earnings rather than to "permanent" earnings. See also Elton, Gruber, Brown and Goetzman (2007) pp 477-482 for a general discussion.

4 As an alternative way of expressing the permanence of earnings in expectation, note that

\[ E_t[\tilde{x}_{t+\tau}] = R \cdot E_t[\tilde{x}_{t}] - r \cdot E_t[\tilde{d}_{t}] \]

for all \( \tau \geq 1 \). However, in case \( \tau = 0 \)

\[ E_t[\tilde{x}_{t+1}] = R \cdot x_t - r \cdot d_t \]

unless \( x_t^* = x_t \).
Note that because \( x_i^* \notin \Omega \), \( E_i[\bar{x}_i^*] \neq x_i^* \).

Easton et al (2000) evaluates how the splitting of earnings into permanent and transitory parts influences the returns on earnings regression.

It may be noted that the modeling here has not allowed for growth and conservative accounting (including future positive NPV projects). This claim follows from Zhang (2000), which shows that \( P > E_i[\bar{x}_i^*]|r \) is necessary and sufficient for conservative accounting given growth. However, we can work with a more general assumption to avoid this problem; yet the thrust of all conclusions will remain. One can use the following assumption, which generalizes permanent earnings:

\[
\bar{x}_{i,t+1} = R x_i^* - rd_i + v_{1,t} + \bar{\varepsilon}_{i,t+1},
\]

where \( \bar{v}_{i,t+1} = \gamma v_{1,t} + \bar{\varepsilon}_{i,t+1} \) specifies the growth concept and \( v_{1,t} \) is observable.

To be sure, a dividend policy like \( \bar{d}_i = K \bar{x}_{i,t+1} + \bar{\alpha}_i \) also eliminates DPI. Now one learns about lagged \( x_{t-1}^* \), which in turn means that \( K \) and \( d_i \) bear on \( E_i[\bar{x}_i^*] \) as well as \( E_i[\bar{x}_{i,t-1}^*] \).

Easterwood and Nutt (1999) evaluates the contributing factors of current accounting data in the revision of a FY2 forecast which a year later turns into a FY1 forecast. This approach differs from ours since we consider FY1 at the end and start of the current period.

Shroff (1995) shows how one identifies firms that are likely to contribute to a relatively high \( R^2 \) when one regresses returns on earnings in the cross-section. Shroff’s screen relies on basic ratios like P/E, ROE and M/B. In light of this paper’s second corollary, one can hypothesize that such firms should also have a relatively high weight on earnings and a relatively low weight on start-of-period expected earnings.