Samuelson’s Dictum: Evidence, Theory, and Implications

Yongok Choi∗ Giacomo Rondina† Todd B. Walker‡

May 2019

Abstract

Using data on all firms that paid a dividend in the Center for Research in Security Prices (CRSP) data set, we show that for individual firms the dividend price ratio accurately forecasts the future growth in dividends, and that the deviation in prices relative to fundamentals is in line with the discounted present value model of stock prices at the individual firm level. This micro efficiency stands in stark contrast to the macro inefficiency that we document. Aggregating individual stocks into equal-weighted, cap-weighted portfolios or an aggregate index delivers resounding rejections of the present value model. We conduct two standard tests of the linear present value model of stock prices: a simple regression of future dividend changes on the dividend-price ratio and a test for excess volatility (i.e., price volatility above and beyond that accounted for by the expected present value of dividends). Both tests deliver micro efficiency and macro inefficiency.

Keywords: Aggregation, Asset Pricing

JEL Codes: G14, G10

∗Chung-Ang University, choiyongok@cau.ac.kr,
†UCSD, grondina@ucsd.edu
‡Indiana University, walkertb@indiana.edu
“Modern markets show considerable micro efficiency (for the reason that the minority who spot aberrations from micro efficiency can make money from those occurrences and, in doing so, they tend to wipe out any persistent inefficiencies). In no contradiction to the previous sentence, I had hypothesized considerable macro inefficiency, in the sense of long waves in the time series of aggregate indexes of security prices below and above various definitions of fundamental values.”

– Paul Samuelson

1 Introduction

The above quote is taken from a private letter from Paul Samuelson to John Campbell and Robert Shiller. The quote appears in Jung and Shiller (2005). It articulates Samuelson’s belief that aggregating individual stocks to form indexes vitiates market efficiency.1 This paper provides empirical evidence in support of Samuelson’s dictum by conducting standard tests of the linear discounted present value model of stock prices. Consistent with Samuelson’s dictum, we find that market efficiency cannot be rejected at the individual stock level. This micro efficiency stands in stark contrast to the macro inefficiency. Aggregating individual stocks into equal-weighted and cap-weighted portfolios delivers resounding rejections of the linear, constant discount rate model of stock prices. Conducting tests using well-known aggregate indexes (e.g., S&P 500)—as is standard in the literature—also leads to rejections of the model.

In Section 2, we conduct two familiar tests of the present value model of stock prices: a simple regression of future dividend changes on the dividend-price ratio, and the test for excess volatility of West (1988). Using data on all firms that paid a dividend in the Center for Research in Security Prices (CRSP) data set, we show that for individual firms, the dividend price ratio accurately forecasts the future growth in dividends and that the deviation in prices relative to dividends is in line with the discounted present value model of stock prices at the individual firm level. The constant discount rate model cannot be rejected. When the same firms are aggregated into equal-weighted or cap-weighted portfolios, the dividend price ratio is no longer a statistically significant predictor of the dividend growth rate. Moreover, the same regressions conducted on aggregate indexes typically do not even yield coefficients of the correct sign. This led Campbell and Shiller (1987) to conclude that in “the entire history of the U.S. stock market, the dividend-price ratio has never predicted dividend growth in accordance with the simple efficient markets theory.”

Price volatility, above and beyond that accounted for by the expected present value of dividends, is claimed to be the most damning piece of evidence against simple versions of the efficient markets hypothesis (EMH) [Shiller (1989), Cochrane (1991)]. We conduct the “second-generation” volatility test of West (1988), which does not rely on stationarity assumptions. The test is based on an inequality on the variance of the innovation in the expected present value of dividends. Assuming constant discount rates, the EMH states that this variance should be smaller when expectations condition on market information (current and past prices and dividends) relative to a subset of

1He articulated this belief well before 2005 [see, Samuelson (1998)].
information. We again find evidence of micro efficiency in that stock prices of individual firms tend not to be excessively volatile, while portfolios of the firms and aggregate indexes do exhibit excess volatility on a grand scale.

Many of the classic empirical studies of the present value model were conducted with the S&P 500 or an alternative index (e.g., CRSP Index) and concluded that the present-value model did not fit data [e.g., Shiller (1981), Campbell and Shiller (1987), West (1988)]. We update these studies and continue to document the inability of the model to explain the behavior of a broad index. We also examine different types of aggregation by forming equal-weighted and cap-weighted portfolios. These aggregates continue to deliver a negative result. It is only for individual firms for which the model cannot be rejected.

We take advantage of the long history of testing the efficient markets hypothesis, while not subscribing to the conclusions of that literature. That is, we are not particularly interested in whether the stock market is “efficient”, but interpret our findings as a test of a particular model (the constant discount rate, present-value model), and document an interesting result with respect to aggregation. The well-established econometric techniques for testing the EMH provide an avenue to conduct our thought experiments.²

Our results have implications for how the asset pricing literature has sought to explain financial anomalies. There has been a notable shift toward relaxing the defining assumptions of rational expectations [Brav and Heaton (2002)]. Namely, perfect information concerning the fundamental structure of the economy and that of optimal decision makers. Behavioral finance papers relax the latter assumption [e.g. Daniel, Hirshleifer, and Subrahmanyam (1998)], while learning models (e.g., Timmerman (1996)) relax the former. Nearly all of these studies have framed the anomaly debate through the lens of an aggregate index or portfolio of assets.³ Our results suggest that a full-information, linear rational expectations model may indeed be a good approximation for the individual firms studied here, while rational expectations (or linearity) breaks down once firms are aggregated.

2 Empirical Evidence

We provide empirical evidence in favor of Samuelson’s dictum by testing various forms of the present value model at the individual firm and aggregate levels. Our tests begin with the standard asset

²Specifically, our test of excess volatility does not rely on stationarity, an important assumption in the EMH literature. Flavin (1983) argued that the procedures used to measure volatility may in fact overstate it due to non-stationarity. However, the overstatement was not sufficient to account for S&P 500 data. Thus, several authors [e.g., Marsh and Merton (1986); Kleidon (1986)] extended this notion by suggesting that the trend-stationarity assumption made about dividends in the original studies was questionable, and that upon adopting the alternative (difference-stationarity), prices did not appear excessively volatile. However, “second-generation” volatility tests [e.g. Campbell and Shiller (1987) West (1988)], which do not rely on stationarity, also find evidence for excess volatility.

³Even papers that examine cross-sectional anomalies have yet to focus on aggregation [see Nagel (2013) for a nice review.]
Pricing equation

\[ P_t = (1 + r_t)^{-1} \mathbb{E}_t(P_{t+1} + D_{t+1}) \]  

(1)

\[ P_t = \mathbb{E}_t \sum_{j=1}^{\infty} (1 + r_t)^{-j} D_{t+j} \]  

(2)

where \( r_t \) is the discount rate known at date \( t \) and assumed to be constant henceforth \((r_t = r_{t+j}, \forall j > 0)\), \( P_t \) is the real stock price at the end of year \( t \), \( D_t \) is the real dividend paid throughout year \( t \), and \( \mathbb{E}_t \) denotes the time \( t \) conditional expectation. The top equation is the standard asset pricing equation in which the price today is equal to the discounted expected payout (price plus dividend) tomorrow. Invoking a no-bubbles condition and the law of iterated expectations yields (2). Appendix A shows that (2) can be written in dividend-price ratio form as

\[ \frac{D_t}{P_t} = r_t - \mathbb{E}_t g_t^D \]  

(3)

\[ g_t^D \equiv \sum_{k=1}^{\infty} (\Delta D_{t+k}/P_t)/(1 + r_t)^{k-1} \]  

(4)

where \( \Delta D_t = D_t - D_{t-1} \). The dividend growth rate is expressed as the sum of discounted future dividend changes for a \$1 investment at \( t \). That is, growth rates are computed relative to the price of the stock as opposed to changes in \( D \). This permits a continuous value for dividend growth even when no dividends were paid.

Given that we do not observe an infinite amount of data, we cannot test (3)–(4) directly. We follow Campbell and Shiller (1988) and Jung and Shiller (2005) in proxying for the future dividend growth \( g_t^D \) by truncating the summation after \( K \) years:

\[ g_t^D = \sum_{k=1}^{K} (\Delta D_{t+k}/P_t)/(1 + r_t)^{k-1} \]  

(5)

Rearranging (3) according to \( \mathbb{E}_t g_t^D = r_t - D_t/P_t \) yields a test of the linear model by running the following regression,

\[ g_t^D = \alpha_t + \beta(D_t/P_t) + \epsilon_t \]  

(6)

If the present value model (3)–(4) is the correct model, the current dividend-price ratio should be negatively correlated with the expected growth rate \((\beta = -1)\), and the time-varying intercept term should coincide with the constant discount rate \((\alpha_t = r_t)\).

In order to understand the effects of aggregation, we follow Jung and Shiller (2005) and construct portfolios by taking the cross-sectional average of the price-dividend ratios. We also control for market size by constructing cap-weighted, cross-sectional averages. Specifically, we use the following
dividend-price averages,

\[
\overline{\frac{D_t}{P_t}} = \frac{1}{L} \sum_{i=1}^{L} \left( \frac{D_{i,t}}{P_{i,t}} \right), \quad \left( \frac{D_t}{P_t} \right)_{CAP} = \sum_{i=1}^{L} \xi_i \left( \frac{D_{i,t}}{P_{i,t}} \right)
\]  

(7)

where \( \xi_i \) denotes the market-capitalization weight assigned to firm \( i \) for time period \( t \).\(^4\) In reporting results, we refer to this cross-sectional average as “Equal Weight 2” and “Cap Weight 2”, respectively. Averaging price-dividend ratios as in (7), as opposed to averaging prices and dividends separately to form the portfolio, could lead to a loss of information. We therefore construct the cross-sectional averages separately for prices and dividends,

\[
\overline{\frac{D_t}{P_t}} = \frac{\sum_{i=1}^{L} D_{i,t}}{\sum_{i=1}^{L} P_{i,t}}, \quad \left( \frac{D_t}{P_t} \right)_{CAP} = \frac{\sum_{i=1}^{L} \xi_i D_{i,t}}{\sum_{i=1}^{L} \xi_i P_{i,t}}
\]  

(8)

and refer to these averages as “Equal Weight 1” and “Cap Weight 1.”

Campbell and Shiller (1998, 2001) tested various versions of (6) using Standard & Poor Composite stock price dating back to 1872 and found the coefficient on \( \left( \frac{D_t}{P_t} \right) \) to be positive—a higher dividend-price ratio counter-intuitively portends a higher expected growth rate.\(^5\) The result was interpreted as “indicating that in the entire history of the US stock market, the dividend-price ratio has never predicted dividend growth in accordance with the simple efficient markets theory.” Jung and Shiller (2005) estimated (6) assuming a constant discount rate, \( \bar{r} = 0.064 \) (annual average return over all firms and dates in the sample), for each of the 49 individual stocks that survived the entire CRSP sample from the first year dividends were recorded (1926) through 2001. They found that the average estimate for \( \beta \) was of the correct sign (negative) and significant. The average \( \beta \) ranged from -0.499 (\( K = 25 \)) to -0.44 (\( K = 10 \)). Aggregating the individual stocks using Equal Weight 2 given by (7) delivered a \( \beta \) range of 0.336 (\( K = 20 \)) to 0.697 (\( K = 25 \)). These results led Jung and Shiller to conclude that “there is now substantial evidence supporting Samuelson’s dictum where market inefficiency is defined as predictability of future (excess) returns.”\(^6\)

2.1 DATA  We test the present value model with stock and dividend data from the Center for Research on Security Prices (CRSP). The data cover the period from January 1926 to December 2018. Firms are grouped based on the continuity of stock price and dividend data. We label firms as discontinuous if they temporarily leave the sample. If a firm has paid an ordinary dividend at any point in the sample, we count them as a dividend-paying firm. We delete firm-month observations whose stock prices are missing or whose dividends are negative for that month. We also delete firm-}

\(^4\)The cap weights are calculated as follows: \( \text{abs}(\text{PRC} * \text{SHROUT}) \) where PRC: Price or Bid/Ask Average, SHROUT: Shares Outstanding.

\(^5\)Specifically, Campbell and Shiller (2001) regressed ten-year log dividend growth rates \( \ln(D_{t+10}/D_t) \) onto \( \ln(D_t/P_t) \).

\(^6\)We confirm the results of Jung and Shiller (2005) continue to hold with updated data through 2018. Assuming the constant discount rate of 0.064, the average \( \beta \) for the individual firms that survived the entire sample (\( n = 46 \)) ranged from -0.58 (\( K = 25 \)) to -0.42 (\( K = 10 \)); while the coefficient on the equal-weighted portfolio given by (7) was positive and ranged from 0.33 (\( K = 20 \)) to 0.25 (\( K = 10 \)).
Table 1: Total number of firms and firm-year observations in the CRSP dataset from 1926 to 2018 decomposed by age and dividend / non-dividend paying.

Table 1 summarizes the number of firms, firm-year observations, and age of the firms in the CRSP dataset. Of the 32,947 firms, roughly half of these firms have issued ordinary dividends (16,584). A small percentage, 1,007 (3.05%), left the sample and reappeared at a later date. The dividend paying firms have a more dispersed distribution in terms of firm age. A majority of dividend-paying firms have survived more than 11 years, which is not true for the non-dividend paying firms. Panel B of Table 1 shows the summary statistics for the number of firm-year observations. There are 342,068 firm-year observations in total, and 227,136 (66%) are dividend paying firms. Table 1 suggests that our analysis consists of firms that are representative of the sample contained in the CRSP dataset.

Forming the approximation (5) requires dividends $K$ periods into the future. Therefore, the number of observations is decreasing in $K$. We only use firms that have at least ten or more observations. For example, in the case of $K = 10$, our data consists of firms with stock price and dividend data with more than 20 annual observations. For $K = 20$, we require 30 annual observations, and so on. Our sample size is vastly larger than previous studies. For example, Jung and Shiller (2005) only examine firms that have price and dividend data available for the entire sample. This amounts to 47 firms or roughly 1% of our sample when $K = 10$.

We create the yearly series of stock prices by selecting the last available observation for each firm and year combination. We the divide the stock price series by the corresponding year’s December Consumer Price Index (CPI) from the Bureau of Labor Statistics to get the inflation-adjusted values. We create the real annual dividend series by summing up 12 monthly adjusted dividends from January to December and dividing by that year’s December CPI. The discount rate in year $t$ ($r_t$) is the December value of the 10-year constant maturity rate [WGS10YR]. Prior to
1962, the data are from Homer and Sylla (1996). All interest rate and CPI data are available at http://www.econ.yale.edu/~shiller/data.htm

2.2 Regression Analysis  We test Samuelson’s Dictum by estimating the linear present value model, $D_t/P_t = r_t - E_t g^D_t$, where $g^D_t \approx \sum_{k=1}^{K} ((\Delta D_{t+k}/P_t)/(1+r_t)^{k-1}$ for nine different specifications and for four truncation parameters, $K = 10, 20, 30, 50$. Our specifications include balanced and unbalanced panels for all firms in CRSP; portfolios of these firms formed by the weighting schemes discussed in the previous section; and the S&P 500 Index. For the individual firms, we estimate the following panel regression

$$g_{it} = \alpha_t + \beta(D_{it}/P_{it}) + \gamma_i + \epsilon_{i,t}$$

(9)

for firm observations $i = 1, \cdots, I$, annual observations $t = 1, \cdots, T$, and where $\gamma_i$ and $\alpha_t$ capture individual and time effects, respectively.

Table 2 reports the results, organized by truncation parameter $K$. We follow previous studies in focusing on the slope coefficient, $\beta$, as opposed to the intercept term, $\alpha$. We discuss this in more detail below. For the balanced panel regressions, we use the 46 firms whose data are continuously observed for the entire sample period of 1926 to 2018. The unbalanced panel contains a maximum number of 3,764 firms and 112,684 firm-year observations when $K = 10$. As mentioned above, our approach permits continuous value for dividend growth even when no dividends are paid. As discussed in Jung and Shiller (2005), zero-dividend observations are informative (e.g., many firms did not pay a dividend during the Great Depression). However, as a check on the influence of zero-dividend observations, the entry labeled “Unbal. (D > 0)” eliminates all firms that have more than 50% of annual dividend entries equal to zero.\footnote{Increasing this value to 66% did not substantially alter results for small $K$. For $K = 50$, the $\beta$ estimate falls to -0.66 with 95% confidence interval [-0.80, -0.51].}

For each $K$, the first three rows of Table 2 report the $\beta$ estimate and 95% confidence interval for the balanced panel (“Bal. Panel”), unbalanced panel (“Unbal. Panel”), and “Unbal. Panel (D > 0)”. These entries constitute the test of Samuelson’s Dictum at the individual firm level. Nearly all estimates for $\beta$ in the first three rows are statistically significant and very close to the hypothesized theoretical value of minus one—even more so for the unbalanced panel regression. As $K$ increases, the number of observations are substantially reduced and the confidence intervals widen as a result. However, these results strongly confirm that at the individual-firm level, the present-value model with constant discount rate is an adequate description of the data.

The remaining entries of Table 2 report the results for aggregate measures of prices and dividends. Recall that “Equal Weight 1” and “Cap Weight 1” refer to the cross-sectional averages constructed separately for prices and dividends (8), while “Equal Weight 2” and “Cap Weight 2” refer to the cross-sectional average of the price-dividend ratio (7). These cross-sectional averages are constructed using the entire unbalanced panel, as opposed to the balanced panel. Aggregating the balanced panel resulted in estimates for $\beta$ that were further away from -1 relative to the unbal-
aggregate the (rows 4–9) assume a time-invariant intercept term when estimating (Table 2: Regression Results. This table reports the results of the regression of future dividend growth on current dividend-price ratio as given by (9) for individual firms and (6) for several aggregate measures. Results are ordered according to truncation parameter $K$ with the first three rows representing individual-firm regressions and the last six are corresponding aggregate measures.

Our results are robust to several variations of the data. Rows 4–9 in Table 2 aggregate the entire sample of firms. We also examined aggregates formed from subsets of these firms based upon longevity; we employed “optimal” portfolio weights constructed from mean-variance optimality conditions; we eliminated outliers caused by the Great Depression and other recessions. Slicing the data along these dimensions did not substantially alter our findings. The time effects and fixed effects are important for generating our result. Removing either effect from the panel regression (9) leads to estimated coefficients that are below one, substantially so for $K = 50$. However, adding time variation to the intercept term of the aggregate regression (6) did not improve the overall fit or push the $\beta$ estimates substantially closer to one. For this reason, all aggregated results reported in Table 2 (rows 4–9) assume a time-invariant intercept term when estimating (6). Our analysis
follows that of Jung and Shiller (2005) in focusing on the slope coefficient but our results extend to the intercept term as well, albeit with much less accuracy. Between 1926 and 2018, the mean (min, max) 10-year constant maturity yield was 4.52% (1.72%, 13.72%). The mean time-fixed effect was 4.25% (-0.02%, 10.25%), while the average estimated intercept term for the aggregated data was 0.83% (0.01%, 1.71%).

Figure 1 shows how quickly aggregation leads to a deviation of the regression coefficient away from -1. In this figure, the 47 firms who survived the entire sample from 1926 through 2015 are ordered from most efficient ($\beta$ closest to -1) to least efficient. We then form sequential, equal-weighted portfolios across the firms following (7) and run the regression (6) for $K = 10$. The horizontal axis indicates the number of firms combined to form the equal-weighted portfolio. As the figure shows, forming a portfolio of the nine most efficient firms leads to a statistically significant deviation from $\beta = -1$. This is true even though the ninth most efficient firm has a $\beta$ that is not statistically different from -1. A portfolio of the 35 most efficient firms has a $\beta$ that is of the wrong sign and aggregating all 47 firms leads to a statistically significant positive value for $\beta$. This analysis suggests that aggregation leads to a quick and decisive deviation from the constant discount rate present value model.

2.3 Excess Volatility  Shiller (2002) has made the case that excess volatility is the one anomaly that is most troubling from the perspective of the efficient markets hypothesis:

“The anomaly represented by the notion of excess volatility seems to be much more troubling for efficient markets theory than some other financial anomalies, such as the January effect or the day-of-the-week effect. The volatility anomaly is much deeper
than those represented by price stickiness or tatonnement or even by exchange-rate overshooting. The evidence regarding excess volatility seems, to some observers at least, to imply that changes in prices occur for no fundamental reason at all, that they occur because of such things as ‘sunspots’ or ‘animal spirits’ or just mass psychology.”

Not only is excess volatility theoretically damaging, it is pervasive in financial data. First introduced by Shiller (1981) and LeRoy and Porter (1981), there is now over four decades of documentation supporting excess volatility in stocks, bonds, foreign exchange and other financial markets. Nearly all of the stock market studies focus exclusively on aggregate indices when testing for excess volatility (e.g. S&P 500, Dow Jones Industrial).

To develop some intuition for this anomaly, we start again with the constant discount rate asset pricing model of (1),

\[ P_t = \theta E_t[P_{t+1} + D_{t+1}] \]  
\[ D_t = A(L)v_t \]

where the discount factor \( \theta = (1 + r_t)^{-1} \) is assumed to be constant henceforth, \( E \) is the expectation operator, \( P_t \) is the price of the individual stock, \( D_t \) the dividend, and \( A(L) \) is a square-summable polynomial in the lag operator \( L \) with \( v \sim N(0, \sigma_v^2) \). The Wold representation theorem permits the use of such a general specification for the exogenous dividend process (11). We use this generalization to show that the results are not unique to the assumed exogenous process for dividends. Solving for the unique rational expectations equilibrium delivers the well-known Hansen-Sargent (1980) optimal prediction formula:

\[ P_t = P(L)v_t = \left( \frac{LA(L)}{L-\theta} - \frac{\theta A(\theta)}{L-\theta} \right) v_t \]

\[ = \sum_{j=0}^{\infty} \theta^j d_{t+j} - \left( \frac{\theta A(\theta)}{L-\theta} \right) v_t \]

\[ = P_t^* - P_t^R \]

The equilibrium price \( P_t \) can be decomposed into two components—the perfect foresight price \( P_t^* \) and a remainder term \( P_t^R \). The perfect foresight price is the price that would prevail if the agent knew past, current, and future values of \( v \). Because the information set of the agent only contains current and past \( v \)’s, \( P_t^R \) represents a conditioning down term that must be orthogonal to information known at \( t \). This orthogonality condition implies the following inequality for the variance of the price:

\[ \text{var}(P_t^*) = \text{var}(P_t) + \text{var}(P_t^R) > \text{var}(P_t) \]

The variance of the perfect foresight price is greater than the variance of the equilibrium (or observed) price.
Shiller’s (1981) classic paper constructed a measure of the perfect foresight price using realized dividends and prices from the S&P 500. He then plotted this measure of the perfect foresight price against actual prices. Figure 2 (a) updates this famous plot. The dashed line represents Shiller’s perfect foresight price using an interest rate of $r = 0.048$ and the solid line is S&P 500 price data. As the figure strongly suggests, the inequality in (14) is grossly violated with the actual price several orders of magnitude more volatile than the perfect foresight price. Figure 2 (b) plots the same measure of $P^*$ against the detrended price for a subset of the 46 firms that survived the entire sample. Note the difference in scale between the two figures.

Much of the debate immediately following Shiller (1981) centered around the calculation of $P^*$ and the appropriate way to test for excess volatility.\(^8\) A series of papers, including that of West (1988), corrected for the statistical and measurement issues and all concluded that excess volatility was a robust feature of stock price data. We implement West (1988) to test for excess volatility in individual firm and aggregated data. The main idea of West’s approach is to test for excess volatility by constructing two nested information sets $I_t$ and $H_t$ with $I_t \supset H_t$. Analogous to the calculation of (14), the present value model implies that the variance of the innovation in the expected present discounted value (PVD) of dividends when expectations are conditional on $I_t$ is smaller than that of the innovation in the expected PVD of dividends when expectations are taken with respect to $H_t$. The following proposition of West (1988) formalizes these concepts.

**Proposition 1.** Let $I_t$ be the linear space spanned by the current and past values of a finite number of random variables with $I_t$ a subset of $I_{t+1}$ for all $t$. It is assumed that after $s$ differences, all random variables in $I_t$ jointly follow a covariance stationary ARMA($q,r$) process for some finite

---

\(^8\)Flavin (1983) and Kleidon (1986) document that the perfect foresight calculations of Shiller lead to small sample bias in tests of excess volatility.
s, q, r ≥ 0. This s’th difference is assumed without loss of generality to have a zero mean. All variables are assumed to be identically zero for t ≤ q. Let d_t be one of these variables. Let H_t be a subset of I_t consisting of the space spanned by current and past values of some subset of the variables in I_t, including at a minimum current and past values of D_t. Let θ be a positive constant, 0 ≤ θ < 1. Let Π(·|·) denote linear projections, calculated for s > 0 as in Hansen and Sargent (1980). Let

\[ x_{t,I} = \lim_{k \to \infty} \Pi \left( \sum_{j=0}^{k} \theta^j d_{t+j} | I_t \right), \quad x_{t,H} = \lim_{k \to \infty} \Pi \left( \sum_{j=0}^{k} \theta^j d_{t+j} | H_t \right) \]

Let E denote mathematical expectations. Then

\[ E[x_{t,H} - \Pi(x_{t,H} | H_{t-1})]^2 \geq E[x_{t,I} - \Pi(x_{t,I} | I_{t-1})]^2 \] (15)

The intuition for West’s proposition can be seen from the Hansen-Sargent prediction formula, (13). The \( P_t^R \) term represents the conditioning down and will be larger for the smaller information set, \( H_t \). From (14), \( \text{var}(P_t^R | I_t) < \text{var}(P_t^R | H_t) \) if \( H_t \subset I_t \). Therefore any two nested information sets can be used to test for excess volatility. See West (1988) for a complete proof.

To calculate the variances of the innovation in (15), we estimate the following bivariate system

\[ P_t = \theta(P_{t+1} + D_{t+1}) + u_{t+1} \] (16)

\[ \Delta^s D_{t+1} = \mu + \phi_1 \Delta^s D_t + \cdots + \phi_q \Delta^s D_{t-q+1} + v_{t+1} \] (17)

where \( P_t \) and \( D_t \) are the stock price and dividend in real terms and \( s \) is the integrated order of the dividend process.

The variance of the innovation in the expected discounted present value conditional on \( I_t \) information can be estimated as \( \hat{\theta}^2 \hat{\sigma}_u^2 \) where \( \hat{\sigma}_u^2 \) is an unbiased estimator for the second moment of the residual in the regression (16). This follows from the expansion of (10),

\[ P_t = \theta(P_{t+1} + D_{t+1}) - \theta[P_{t+1} + D_{t+1} - E(P_{t+1} + D_{t+1} | I_t)] \]

\[ = \theta(P_{t+1} + D_{t+1}) + u_{t+1} \]

\[ \sigma_u^2 = \theta^2[P_{t+1} + D_{t+1} - E(P_{t+1} + D_{t+1} | I_t)]^2 \] (18)

Therefore, the right-hand side of (15) can be estimated by \( \hat{\theta}^2 \hat{\sigma}_u^2 \). We label this forecast error as \( \forall(I_{t-1}) \equiv E[P_t - \Pi(P_t | I_{t-1})]^2 \).

Conversely, using the Hansen and Sargent formula (12), we can derive the left-hand side of (15) by assuming \( H_t \) contains the history of dividends. The variance of the one-step-ahead forecast error...
is given by

\[ \mathbb{V}(H_{t-1}) = \mathbb{E}[P_t - \Pi(P_t|H_{t-1})] = \mathbb{E}[P(L)v_t - L^{-1}(P(L) - P(0))v_{t-1}]^2 \]

\[ = P(0)^2\sigma_v^2 = A(\theta)^2\sigma_v^2 \]

\[ = ((1 - \hat{\theta})^s(1 - \sum_1^q \hat{\theta}^j\hat{\phi}_j))^{-2}\hat{\sigma}_v^2 \] (19)

The first equality of (19) follows from the Wiener-Kolmogorov optimal prediction formula. The final equality comes from (12) and substituting in the stochastic process for the dividends (17). The \( \hat{\sigma}_v^2 \) is an unbiased estimator for the second moment of the residual in the regression (17). Then the present value model can be tested based on the sign of the statistic

\[ \mathbb{E}[P_t - \Pi(P_t|H_{t-1})] - \mathbb{E}[P_t - \Pi(P_t|I_{t-1})] \]

which we have established as

\[ \Upsilon = \mathbb{V}(H_{t-1}) - \mathbb{V}(I_{t-1}) \]

\[ = ((1 - \hat{\theta})^s(1 - \sum_1^q \hat{\theta}^j\hat{\phi}_j))^{-2}\hat{\sigma}_v^2 - \hat{\theta}^{-2}\hat{\sigma}_u^2 \] (20)

where the second equality provides an estimate of this statistic. A test of excess volatility is then

\[ H_0: \Upsilon \geq 0. \]

A negative value of \( \Upsilon \) indicates that enlarging the information set leads to the perverse result of an increase in the innovation variance. The volatility is then excessive relative to the theoretical model. Following West, when \( \Upsilon \) is negative, we also report

\[ \tilde{\Upsilon} = -100 \left( \frac{\mathbb{V}(H_{t-1}) - \mathbb{V}(I_{t-1})}{\mathbb{V}(I_{t-1})} \right) \] (21)

which gives a rough estimate of the percentage of the variance in the price process that is excessive.

West applied this method to test the present value model on S&P 500 data for 1871–1980 and the Dow Jones index from 1928–1978 for a wide variety of \( q \) and \( s \). After a battery of robustness checks, he concluded that stock prices were too volatile to be the expected discounted value of dividends. Typical values for \( \Upsilon \) (\( \tilde{\Upsilon} \)) were -230 (92.92) for S&P data and -21,545 (96.92) for Dow Jones data [Table II, pg. 51].

We update these results for aggregate indices and also apply the test at the individual firm level. We use a balanced panel of 47 firms whose data are continuously observed for the whole sample period of 1926 to 2015, and an unbalanced panel of 278 firms with more than 60 years of observations. For each observation, we estimate the bivariate system (17) to obtain the corresponding statistics for formula (20). Equation (16) is estimated by Hansen’s (1982) two-step, two-stage least squares. We use the augmented Dickey-Fuller (ADF) test for each dividend series to determine \( s \) in the Equation (17). For each regression, we set the AR order based on the Hannan-Quinn (1979) procedure. We estimate the variance-covariance matrix in accordance with West (1988) and use the delta method to get the variance of the statistic (20) and (21).

The first two rows of Table 3 report the average West statistic \( \Upsilon \) for the 278 firms with more
Table 3: Volatility Test Results. This table reports the statistic from equations (20) \((\Upsilon)\) and (21) \((\tilde{\Upsilon})\). For individual firms, we report the number out of the 278 and 47 firms that cannot reject the excess volatility test \(H_0: \Upsilon \geq 0\) for various significance levels.

Consistent with the previous section, the excess volatility tests confirm Samuelson’s Dictum—stocks appear excessively volatile only when aggregated. The average of the test statistic for the individual firms is of the correct sign, and the null of a positive value cannot be rejected for a majority of the firms at the 10% significance level for the unbalanced panel and 5% significance level for the balanced panel. Conversely the statistic for every aggregated measure is of the wrong sign and statistically different from zero, confirming excess volatility. The S&P 500 index displays substantial excess volatility relative to the aggregate measures of the individual firms. Table 3 reflects the findings of Table 2 with the present value model successfully accounting for the dynamics of individual firms while failing to account for any aggregate of the individual firms. The present value model miserably fails to account for the dynamics of broad aggregate measures such as the S&P 500 Index.

3 Theory

Suppose there are \(I\) firms indexed by \(i = 1, 2, ..., I\) with stocks priced according to the linear present value model

\[
P_{it} = E_t \sum_{j=1}^{\infty} \theta^j D_{it+j}
\]  

(22)

where \(P_{it}\) is the stock price of firm \(i\), \(\theta\) is the constant discount factor \((\theta = (1 + r)^{-1})\), and \(D_{it}\) is the dividend of firm \(i\). The firm-specific dividend process is driven by an aggregate shock \((\varepsilon_t)\) and
an idiosyncratic shock ($\eta_{it}$) and is given by

$$D_{it} = a_t + \eta_{it}$$  \hfill (23)

$$a_t = \rho a_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2_{\varepsilon})$, $\eta_{it} \sim N(0, \sigma^2_{\eta})$, with $\varepsilon_t$ and $\eta_{it}$ uncorrelated at all leads and lags. The aggregate shock follows an AR(1) specification with autocorrelation coefficient $|\rho| < 1$.

3.1 Full Information  Under full information, traders are assumed to observe the entire past history of both the aggregate and idiosyncratic shocks, $F_t = \{\varepsilon_t, \eta_{it}, \varepsilon_{t-1}, \eta_{it-1}, \ldots\}$, and the rational expectations equilibrium price for firm $i$ follows from the Hansen-Sargent optimal prediction formula (substituting (23) into (12)),

$$P_t|F_t = \left(\frac{\theta\rho}{1 - \theta\rho}\right)a_t$$  \hfill (24)

Now consider the implications of forming the cross-sectional average, “Equal Weight 1”, $D_t = (1/I)\sum_{i=1}^{I} D_{it}$, $P_t = (1/I)\sum_{i=1}^{I} P_{it}$. If we assume the law of large numbers is operational, then the idiosyncratic shock would wash out of the average dividend, $D_t \equiv (1/I)\sum_i D_{it} = a_t$. Pricing such a dividend stream would yield the rational expectations equilibrium,

$$P_t|F_t = \left(\frac{\theta\rho}{1 - \theta\rho}\right)a_t = (1/I)\sum_{i=1}^{I} P_{it}|F_t$$  \hfill (25)

The last equality emphasizes that a trader pricing the aggregate dividend stream $D_t$ would deliver the same stock price as the average of individual traders pricing each stock $i$ independently. That is, we could have market segmentation where individual traders specialize in specific stocks—an analogy we make use of shortly—and, under the assumption of perfect information, aggregating the individual prices for each firm $i$ perfectly replicates the price of the aggregate index $P_t$. Thus under full information, the empirical tests of the present value model conducted in Section 2 would be accurate (by construction) with no difference between individual and aggregate statistics. A full-information rational expectations equilibrium cannot explain Samuelson’s Dictum.

3.2 Incomplete Information  Under incomplete information, we assume traders specialize in individual stocks. Trader $i$ follows firm $i$ and her information set consists of current and past dividends, $H_{it} = \{D_{it}, D_{it-1}, \ldots\}$. That trader $i$ does not observe the stock price of firm $i$, or the dividends / stock price of firm $j$ is obviously a simplification imposed to keep the algebra of the paper-and-pencil variety. We discuss how to relax this assumption while maintaining our results in Appendix B.

Trader $i$ now has a signal extraction problem to solve. Given the stochastic process for dividends (23), she cannot distinguish between idiosyncratic and aggregate shocks. The following proposition derives the rational expectations equilibrium price.

**Proposition 2.** Let Trader $i$’s information set be current and past dividends $H_{it} = \{D_{it}, D_{it-1}, \ldots\}$
with the dividend process given by (23), then the equilibrium price is given by

\[ P_t = \left( \frac{\theta(\rho - \lambda)}{(1 - \theta \rho)(1 - \lambda L)} \right) a_t + \left( \frac{\theta(\rho - \lambda)}{(1 - \lambda L)} \right) \eta_t \]  

(26)

\[ \lambda = \frac{1}{2} \left\{ \left( \frac{\sigma^2_{\varepsilon}}{\sigma^2_{\eta \rho}} \right) + \left( \frac{1}{\rho} + \rho \right) - \left\{ \left[ \frac{\sigma^2_{\varepsilon}}{\sigma^2_{\eta \rho}} + \left( \frac{1}{\rho} + \rho \right) \right]^2 - 4 \right\}^{1/2} \right\} \]

Proof: See Appendix A.

Notice that elements of the idiosyncratic shock now bleed into the loading on the aggregate shock through \( \lambda \), which is a function of the signal-to-noise ratio. Forming the aggregate dividend and price

\[ D_t \equiv (1/I) \sum_{i=1}^{I} D_{it} = a_t \]  

(27)

\[ P_t \equiv (1/I) \sum_{i=1}^{I} P_{it} = \left( \frac{\theta(\rho - \lambda)}{(1 - \theta \rho)(1 - \lambda L)} \right) a_t \]  

(28)

The aggregate measures are incongruent in that the pricing the dividend process (27) would yield the equilibrium given by (25) and not (28). Therefore, while the micro efficiency holds by assumption, the model would display macro inefficiency. Any test of the linear present value model based upon (27) and (28) would generate biased results.

To see this more clearly, consider the Gordon growth expression \( \mathbb{E}(g_{it}|I_{it}) = rP_{it} - D_{it} \). Aggregating both sides gives \( \mathbb{E}(g_{it}) = rP_t - D_t \) where \( \mathbb{E}(g_{it}) = \sum_{i=1}^{I} \mathbb{E}(g_{it}|I_{it}) \). Now suppose following Section 2, we test this theory with the econometric specification

\[ g_t = \alpha P_t + \beta D_t + u_t \]

and test the null \( H_0 : \beta = -1 \). Under full information, the theory is clearly consistent with the hypothesized test, but under incomplete information, the theory is misspecified. Note that \( u_t \) can be written as

\[ u_t = g_t - \mathbb{E}(g_{it}) = [g_t - \mathbb{E}(g_t)] + [\mathbb{E}(g_t) - \mathbb{E}(g_{it})] = s_t + v_t \]

where the covariance between \( \text{cov}(D_t, s_t) = 0 \) but \( \text{cov}(D_t, v_t) \neq 0 \), which implies \( \text{cov}(D_t, u_t) \neq 0 \). In our simple specification

\[ v_t = \mathbb{E}(g_t) - \mathbb{E}(g_{it}) = \left( \frac{(1 - \delta)\lambda}{(1 - \rho \delta)(1 - \lambda L)} \right) \varepsilon_t \]

and hence as long as \( \lambda \neq 0 \), then \( \text{cov}(D_t, v_t) \neq 0 \). Table 4 reports the estimated \( \beta \) in the equally weighted exercise using CRSP data for \( K = 50 \), and the median estimated \( \beta \) from simulated data of the theoretical model with \( \delta = 0.94, \sigma_{\varepsilon}/\sigma_{\eta} = 0.2, \rho = 0.9 \) and \( I = 100 \). The correct value is \( \beta = -1 \) under EMH, which holds in the disaggregated model by construction.
Table 4: Simulated $\beta$ Estimates

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Simulated</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>-0.634</td>
<td>-0.6537</td>
<td>-1</td>
</tr>
</tbody>
</table>

4 Implications and Concluding Thoughts

There is a vast literature on empirical asset pricing but few have focused exclusively on the implications of aggregation. Our results and analysis most resemble that of Jung and Shiller (2005), who also focus on testing Samuelson’s dictum. We extend the results in Jung and Shiller along two dimensions. First we introduce panel data methodology, which permits the expansion of the data set to include all firms that paid a dividend between 1926 and 2014. This dramatically increases the number of observations from 3,995 to 342,068. Second, we conduct West’s (1988) test of excess volatility in addition to the tests described in Jung and Shiller (2005).

Jung and Shiller provide prima facie evidence for Samuelson’s dictum, but by increasing the number of observations by several orders of magnitude and by conducting other econometric tests, we substantially buttress their findings. Specifically, the results of Jung and Shiller are not statistically significant. A formal test of Samuelson’s dictum is easily rejected using their data and methodology. By adding observations and using panel-data techniques, we show that Samuelson’s dictum cannot be rejected by the data.

A different flavor of our findings appear in Startz and Tsang (2014), Vuolteenaho (2002) and Cohen, Polk, and Vuolteenaho (2003), who employ vector autoregression analysis along the lines of Campbell and Shiller (1987) to document that variation in individual firms’ stock prices appear to driven by fundamentals. Using a time-varying discount rate model, Startz and Tsang (2014) find that the present value model is rejected at longer horizons. Our results also show some sign of sensitivity to horizon, but not to the extent highlighted by Startz and Tsang. The difference in results is most likely attributed to the time-varying discount rate and alternative econometric methodology employed by Startz and Tsang. Using variance decomposition analysis, Vuolteenaho (2002) concluded that even though most of the variation in individual firm stock prices is due to fundamentals, a significant portion of the variation can be attributed to inefficient components. Using similar methods and data from 1937–1997, Cohen, Polk, and Vuolteenaho (2003) decompose the cross-portfolio variance of the log book value to market value into a component of future cash flows, a component due to the persistent of the value premium, and a component that generates predictable returns (an inefficient component). They conclude that for 15-year returns, only 20% of the variation is due to the inefficient component. By focusing on portfolio formation, these studies miss the link between individual firm behavior and the aggregate counterparts.

We believe our results have far-reaching consequences that extend beyond the asset pricing literature. Understanding the nature of aggregation is a crucial component of modern economics.
Tests of market efficiency provide a natural laboratory for exploring the effects of aggregation. Thus, more work needs to be done to understand why Samuelson’s dictum holds.

A few have attributed this finding to the ability of market participants to better forecast idiosyncratic changes in prices and dividends of individual firms relative to the aggregated changes of indexes [Jung and Shiller (2005), Campbell, Lo, and MacKinlay (1997)]. If individual firms’ fundamentals contain a significant amount of firm-specific or idiosyncratic noise and assuming that aggregation attenuates these idiosyncratic shocks, then one would potentially expect a different result when testing the the present value model on individual firms and a portfolio of the firms. The underlying assumptions behind this result are [i.] idiosyncratic shocks are the primary drivers of firms’ fundamentals; [ii.] investors are better able to forecast changes in idiosyncratic shocks relative to aggregate shocks.

Since the rejection of the present value model, there has also been a notable shift toward relaxing the defining assumptions of rational expectations models in explaining financial anomalies. Namely, perfect information concerning the fundamental structure of the economy and that of optimal decision makers. For example, Michener (1982) shows that the introduction of risk aversion will produce some (though not enough) volatility in asset prices. Yet tests of such models generally result in rejections. Models of time-non-separable preferences [e.g., Constantinides (1990)] are capable of explaining some stock market volatility, but are in general unwieldy and do not fit overly well. We believe our paper signals an alternative explanation is possible.
References


5 Appendix A: Derivations and Proofs

5.1 Deriving (3) from (2)  

The present-value model of stock prices is given by

\[ P_t = \mathbb{E}_t \sum_{j=1}^{\infty} (1 + r_t)^{-j} D_{t+j} \tag{29} \]

Multiple by \( r_t \) and add \( D_t \) to both sides of (29).

\[ r_t P_t + D_t = D_t + r_t \mathbb{E}_t \sum_{j=1}^{\infty} (1 + r_t)^{-j} D_{t+j} \]

Dividing by \( P_t \) and re-arranging yields

\[
\frac{D_t}{P_t} = r_t + \frac{D_t}{P_t} - \frac{r_t}{P_t} \mathbb{E}_t \sum_{j=1}^{\infty} (1 + r_t)^{-j} D_{t+j} \\
= r_t + \frac{D_t}{P_t} - \frac{r_t}{P_t} \mathbb{E}_t \left( \frac{D_{t+1}}{1 + r_t} + \frac{D_{t+2}}{(1 + r_t)^2} + \frac{D_{t+3}}{(1 + r_t)^3} + \cdots \right) \\
= r_t + \frac{1}{P_t} \mathbb{E}_t \left( D_t - \frac{r_tD_{t+1}}{1 + r_t} - \frac{r_tD_{t+2}}{(1 + r_t)^2} - \frac{r_tD_{t+3}}{(1 + r_t)^3} + \cdots \right) 
\]

Note \( \frac{r_t}{1 + r_t} = 1 - \frac{1}{1 + r_t} \), \( \frac{r_t}{(1 + r_t)^2} = \frac{1}{1 + r_t} - \frac{1}{(1 + r_t)^2} \) and therefore

\[
\frac{D_t}{P_t} = r_t - \frac{1}{P_t} \mathbb{E}_t \left( -D_t + D_{t+1} \left( 1 - \frac{1}{1 + r_t} \right) + D_{t+2} \left( \frac{1}{1 + r_t} - \frac{1}{(1 + r_t)^2} \right) \right) \\
+ D_{t+3} \left( \frac{1}{(1 + r_t)^2} - \frac{1}{(1 + r_t)^3} \right) + \cdots 
\]

Defining \( \Delta D_{t+j} \equiv D_{t+j} - D_{t+j-1} \)

\[
\frac{D_t}{P_t} = r_t - \frac{1}{P_t} \mathbb{E}_t \left( \Delta D_{t+1} + \frac{\Delta D_{t+2}}{1 + r_t} + \frac{\Delta D_{t+3}}{(1 + r_t)^2} + \cdots \right) \\
= r_t - \mathbb{E}_t g_t 
\tag{30} 
\]
5.2 Deriving Hansen-Sargent Formula

\[ p_t = \beta \mathbb{E}_t (p_{t+1} + d_{t+1}) \]

Taking expectations

\[ \mathbb{E}_t (p_{t+1}) = L^{-1} [P(L) - P_0] \varepsilon_t \]
\[ \mathbb{E}_t (d_{t+1}) = L^{-1} [D(L) - D_0] \varepsilon_t \] (31)

Substituting these into the equilibrium

\[ P(z) = \beta (z^{-1} [P(z) - P_0] + z^{-1} [D(z) - D_0]) \]
\[ z P(z) = \beta P(z) + \beta [D(z) - D_0] - \beta P_0 \]
\[ (z - \beta) P(z) = \beta [D(z) - D_0] - \beta P_0 \] (32)

Evaluating at \( z = \beta \) gives \( P_0 = D_0 - D(\beta) \) and the solution is

\[ P(z) = \frac{\beta D(z) - \beta D(\beta)}{z - \beta} \] (33)

If \( D(z) = (1 + \theta z) / (1 - \rho z) \), then

\[ P(z) = \beta \left( \frac{1 + \theta z}{1 - \rho z} - \frac{1 + \theta \beta}{1 - \rho \beta} \right) / (z - \beta) \]
\[ = \beta \left( \frac{(1 + \theta z)(1 - \rho \beta) - (1 + \theta \beta)(1 - \rho z)}{(1 - \rho \beta)(1 - \rho z)} \right) / (z - \beta) \] (34)

Simplifying the numerator

\[ 1 + \theta z - \rho \beta - \theta \rho \beta z - (1 + \theta \beta - \rho z - \theta \beta \rho z) = \theta (z - \beta) + \rho (z - \beta) \] (35)

\[ P(z) = \beta \left( \frac{\theta + \rho}{(1 - \rho \beta)(1 - \rho z)} \right) \]
Therefore the equilibrium price is
\[ p_t = \frac{\mu \beta}{1 - \beta} + \left( \frac{\beta(\theta + \rho)}{(1 - \rho \beta)(1 - \rho L)} \right) \varepsilon_t \]
\[ p_t = \frac{\mu \beta (1 - \rho)}{1 - \beta} + \rho p_{t-1} + \left( \frac{\beta(\theta + \rho)}{1 - \rho \beta} \right) \varepsilon_t \]
\[ p_t = \tilde{\mu} + \rho p_{t-1} + \tilde{\varepsilon}_t \]
\[ d_t = \mu + \left( \frac{1 + \theta L}{1 - \rho L} \right) \varepsilon_t \]  
(36)

So the price process is an AR(1) with mean \( \mu \beta / (1 - \beta) \) and variance
\[ \left( \frac{\beta(\theta + \rho)}{1 - \rho \beta} \right)^2 \left( \frac{\sigma^2}{1 - \rho^2} \right) \]  
(37)

6 Appendix B: Relaxing Assumptions