SAMUELSON’S DICTUM: EVIDENCE, THEORY AND IMPLICATIONS

Yongok Choi∗ Giacomo Rondina† Todd B. Walker‡

January 2020

ABSTRACT

The idea that stock prices can simultaneously display “micro efficiency” and “macro inefficiency” is accredited to Paul Samuelson and has been labeled “Samuelson’s Dictum.” Using firm-level data from the Center for Research in Security Prices (CRSP) we present empirical evidence consistent with Samuelson’s dictum. Specifically, we conduct two standard tests of the linear present value model of stock prices: a regression of future dividend changes on the dividend-price ratio, and a test for excess volatility. We show that the dividend price ratio forecasts the future growth in dividends at the firm level as predicted by the present value model, and that excess volatility can be rejected for most firms. When the same firms are aggregated into equal-weighted or cap-weighted portfolios, the estimated coefficients are no longer consistent with the present value model and “excess” volatility is observed. To investigate the source of our empirical findings, we propose a theory of aggregation bias based on incomplete information and segmented markets. Traders specializing in individual stocks conflate idiosyncratic and aggregate shocks to dividends. To an econometrician using aggregate data, these assumptions generate a rejection of the present value model even though individual traders are efficiently using their available information. We connect our results to a broader literature and discuss the implications associated with concepts such as micro-foundations of macroeconomic modeling.

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 letter from Paul Samuelson to John Campbell and Robert Shiller. It appears in Jung and Shiller (2005), who refer to it as “Samuelson’s Dictum.” The quote articulates Samuelson’s belief that market efficiency applies to individual firms but is vitiated when stock prices are aggregated to form indexes like the S&P 500. This paper provides empirical evidence consistent with Samuelson’s Dictum. To investigate the source of our empirical findings, we propose a theory of aggregation bias based on incomplete information and segmented markets.

We work within the context of a standard present value model of stock prices to document that the model cannot be rejected when tested with firm-level data.¹ That a simple present value model can fit asset price data along any dimension is a remarkable finding given that an overwhelming majority of the empirical finance literature has reached the exact opposite conclusion. We reconcile our findings with the broader literature by demonstrating that aggregating the individual stocks into equal-weighted or cap-weighted portfolios delivers the well-known rejections of the model. Moreover tests that use an aggregate index (e.g., S&P 500), as is common in the literature, lead to even more resounding rejections. Apparently the more one aggregates, the poorer the fit of the present value model. Thus we interpret our findings as an interesting result with respect to aggregation as opposed to market efficiency per se.

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 a test for excess volatility. Using data on all firms that paid a consistent dividend in the Center for Research in Security Prices (CRSP), 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. The model cannot be rejected at the individual firm level. When the same firms are aggregated into equal-weighted or cap-weighted portfolios, the estimated coefficients are no longer consistent with the present value model. Running the same regressions with data from aggregate indexes does not even yield coefficients of the correct sign, which led Campbell and Shiller (1987) to conclude that in “the entire history of the U.S. stock market, the

¹We occasionally use the terms “micro efficient” and “macro inefficient” in describing results. To be clear, these terms pertain to the fit of the present value model when analyzed with firm-level data (micro efficient) and the lack of fit when estimated with aggregate data (macro inefficient). Our analysis of Samuelson’s Dictum takes advantage of the long history of testing the efficient markets hypothesis (EMH) while not subscribing to the conclusions of that literature. That is, we are not interested in asking whether a stock price is “efficient” as labeled in the above quote but rely on the well-established econometric techniques for testing the EMH to conduct our thought experiments.
dividend-price ratio has never predicted dividend growth in accordance with the simple efficient markets theory.” A statement that we argue should only be reserved for market aggregates.

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 present value model [Cochrane (1991), Shiller (2002)]. We conduct the “second-generation” volatility test of West (1988), which is based on an inequality of the variance of the innovation in the expected present value of dividends.\(^2\) The theory states that this variance should be smaller when expectations condition on market information (current and past prices and dividends) relative to a subset of that information. We again find evidence that stock prices of individual firms are not excessively volatile. Over 90% of firms in our sample do not exhibit excess volatility at a 1% significance level. However, portfolios of these same firms exhibit excess volatility, and aggregate indexes commonly used in asset pricing studies exhibit excess volatility on a grander scale. As with our regression results, the more one aggregates, the more significant the rejection of the model.

To reconcile our findings, we propose a theory of aggregation bias based on incomplete information and segmented markets in Section 3. We assume firm dividends are subject to two types of shocks—idiosyncratic and aggregate—and that traders cannot perfectly distinguish between them. If markets are segmented in the sense that traders specialize in trading individual firms and incomplete information persists, aggregate prices and dividends differ from their complete information counterpart. To the econometrician using aggregate data, such differences would materialize as a rejection of the present value model, even though individual traders are pricing stocks using this exact model and are efficiently using all available information. Conversely, tests of individual firms do not suffer from the same bias as the econometric tests properly account for the incomplete information. Monte Carlo investigations show that this theory is a plausible explanation of the results of Section 2.

The final section of the paper, Section 4, relates our findings to a broader literature. The underlying message of our paper is closely related to the pioneering works of Goodfriend (1992) and Pischke (1995), who pointed to the risks of information aggregation bias when conducting tests of the life-cycle permanent income theory. Analogously, the asset pricing literature has documented “micro efficiency” by showing that variation in individual firms’ stock prices are largely driven by fundamentals [e.g., Campbell (1991), Vuolteenaho (2002), Cohen et al. (2003)]. Conversely, “macro inefficiency” has been documented on a much larger scale. Many of the classic empirical studies were conducted with S&P 500 data or an alternative index (CRSP Index). The overwhelming conclusion was that the present-value model did not fit data well [e.g., Shiller (1981), Campbell

\(^2\)West (1988) is a good example of how we rely on the econometrics developed to test the EMH in our analysis. West’s test of excess volatility does not rely on stationarity of the time series, a constraint that is crucial to relax. Flavin (1983) argued 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 in aggregate data.
Choi, Rondina & Walker: Samuelson’s Dictum

and Shiller (1987), West (1988)]. In response, the asset pricing literature took a notable shift toward relaxing the defining assumptions of the linear present value model (e.g. rational expectations). Nearly all of the subsequent literature framed the debate through the lens of an aggregate index. Our results suggest that the linear present value model may indeed be a good approximation for the behavior of individual firms, and that aggregation holds the clues for understanding anomalous behavior at the aggregate level.

We believe Samuelson’s Dictum has implications that extend beyond the asset pricing example studied here. Given the ubiquitous nature of the present value model, our results highlight properties of aggregation that may appear in many other contexts. If micro-foundations are the cornerstone of macroeconomic modeling, then aggregation is of first-order importance and asset pricing data provides the perfect laboratory to test the aggregation hypothesis. In this sense, we follow Cochrane and Hansen (1992) in arguing that “security market data are among the most sensitive and hence attractive proving grounds for models of the aggregate economy.” Thus, we view this paper’s relative contributions as: [i.] documenting an important stylized fact with respect to aggregation; and [ii.] proposing a plausible theory to explain the stylized fact.

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 pricing equation

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

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

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

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 (1988a) and Jung and Shiller (2005) in proxying for the future dividend growth \( g_t^D \) by truncating the summation after \( K \) years:

\[
\hat{g}_t^D = \frac{\sum_{k=1}^{K} (\Delta D_{t+k}/P_t)/(1 + r_t)^{k-1}}{(1 + r_t)^{K-1}} \quad (5)
\]

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

\[
g_t^D = \alpha_t + \beta (D_t/P_t) + \epsilon_t \quad (6)
\]

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 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,

\[
\overline{(D_t/P_t)} = \frac{1}{T} \sum_{i=1}^{T} \left( \frac{D_{i,t}}{P_{i,t}} \right), \quad \quad (D_t/P_t)_{\text{CAP}} = \sum_{i=1}^{T} \xi_i \left( \frac{D_{i,t}}{P_{i,t}} \right) \quad (7)
\]

where \( \xi_i \) denotes the market-capitalization weight assigned to firm \( i \) for time period \( t \).\(^3\) 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,

\[
\overline{D_t/P_t} = \frac{\sum_{i=1}^{T} D_{i,t}}{\sum_{i=1}^{T} P_{i,t}}, \quad \quad (\overline{D_t/P_t})_{\text{CAP}} = \frac{\sum_{i=1}^{T} \xi_i D_{i,t}}{\sum_{i=1}^{T} \xi_i P_{i,t}} \quad (8)
\]

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 \( (D_t/P_t) \) to be positive—a higher dividend-price ratio counter-intuitively portends a higher expected growth rate.\(^4\) The result was interpreted as “indicating that in the entire history of the US stock market, the dividend-price

\(^3\)The cap weights are calculated as follows: abs(PRCC\*SHROUT) where PRC: Price or Bid/Ask Average, SHROUT: Shares Outstanding.

\(^4\)Specifically, Campbell and Shiller (2001) regressed ten-year log dividend growth rates \( \ln(D_{t+10}/D_t) \) onto \( \ln(D_t/P_t) \).
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.”

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-month observations whose Cumulative Factor to Adjust Prices over a Date Range \((\text{cumfacpr})\) are missing or negative. As indicated in Data Descriptions Guide (CRSP, 2010), in order to compare the stock prices across time, we divide each series by \text{cumfacpr}. We only keep observations generated by ordinary dividends data, excluding observations due to liquidation, acquisition, reorganization, and issuances.

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

---

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>
<thead>
<tr>
<th>Number of Firms (Firm Years)</th>
<th>Firm Age</th>
<th>Non-Dividend</th>
<th>Dividend</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discontinuous</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>613 (7,506)</td>
<td></td>
<td>394 (7,486)</td>
<td></td>
<td>1,007 (14,954)</td>
</tr>
<tr>
<td>Continuous</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10 12,290 (53,773)</td>
<td></td>
<td>7,449 (35,549)</td>
<td></td>
<td>19,739 (89,332)</td>
</tr>
<tr>
<td>11-20 2,640 (38,143)</td>
<td></td>
<td>4,482 (58,739)</td>
<td></td>
<td>7,122 (96,882)</td>
</tr>
<tr>
<td>21-40 803 (15,063)</td>
<td></td>
<td>3,113 (74,673)</td>
<td></td>
<td>3,916 (89,736)</td>
</tr>
<tr>
<td>41-60 16 (378)</td>
<td></td>
<td>897 (30,837)</td>
<td></td>
<td>913 (31,215)</td>
</tr>
<tr>
<td>61-80 1 (69)</td>
<td></td>
<td>193 (12,823)</td>
<td></td>
<td>194 (12,892)</td>
</tr>
<tr>
<td>81-90 0</td>
<td></td>
<td>48 (7,027)</td>
<td></td>
<td>48 (7,027)</td>
</tr>
<tr>
<td>90+ 0</td>
<td></td>
<td>46 (4,042)</td>
<td></td>
<td>46 (4,042)</td>
</tr>
<tr>
<td>Total</td>
<td>16,363 (114,932)</td>
<td>16,584 (227,136)</td>
<td>32,947 (342,068)</td>
<td></td>
</tr>
</tbody>
</table>

---

5 We confirm the results of Jung and Shiller (2005) continue to hold with updated data through 2018. Assuming a 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 \((K = 10)\).
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 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, $D_t/P_t = r_t - E_tg^D_t$, 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

$$g_{it} = \alpha_t + \beta(D_{it}/P_{it}) + \gamma_i + \epsilon_{i,t} \tag{9}$$

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
Responding to 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 β that were further away from -1 relative to the unbalanced panel. We also report results for the S&P 500 Index, examining both the same time horizon

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. Standard errors are clustered at the firm level.

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.6

For each K, the first three rows of Table 2 report the β estimate and 95% confidence interval (clustered at the firm level) 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 β 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 β that were further away from -1 relative to the unbalanced panel. We also report results for the S&P 500 Index, examining both the same time horizon

---

6Increasing this value to 66% did not substantially alter results for small K. For K = 50, the β estimate falls to -0.66 with 95% confidence interval [-0.80, -0.51].
as our sample (1926) and the full sample which dates back to 1872.

All aggregate measures lead to obvious rejections of the constant discount rate model. For many of the aggregate measures, the estimated coefficient is positive. This is true for all values of $K$ for the price-dividend averages and nearly all of the S&P 500 samples. The aggregate measure with an estimated coefficient that is closest to the theoretical counterpart of -1 is “Equal Weight 1”, which is estimated to be of the correct sign. However, it remains significantly different from the individual-firm regressions as there is no overlap in the 95% confidence intervals. The present value model for individual stock prices and dividends appears consistent with the data, while any form of aggregation leads to an outright rejection.

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.\(^7\) 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%).

There is a clear relationship between the extent of aggregation and deviation from the hypothesized value for $\beta$. The parameter estimates are ordered vertically in Table 2 according to the extent of aggregation. The loss of information due to moving from “Equal Weight 1” to “Equal Weight 2” is substantial. Similarly, the estimates using the S&P 500 data are clearly inconsistent with the present value model. Conversely, the estimates derived from “Equal Weight 1” and “Cap Weight 1”, while inconsistent with the theory, are not substantial deviations.

To further investigate the phenomena, 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 consistent with the theory ($\beta$ closest to -1) to least consistent. 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 consistent firms leads to a statistically significant deviation from $\beta = -1$. This is true even though the ninth most consistent firm has a $\beta$ that is not statistically different from -1. A portfolio

\(^7\)Additional results available upon request.
Figure 1: Aggregation Bias. The figure reports the estimates for $\beta$ in regression (6) with $K = 10$ and 95% confidence interval in dashed lines, for equal-weighted portfolios as in (7) of an expanding number of firms. Firms are sequentially added to the portfolio from the most to the least consistent with the theory, i.e., based on the point estimate of $\beta$ being closest to -1 when regression (6) is ran at the individual firm level. The number of firms in the portfolio is indicated in the horizontal axis. Of the 35 most consistent 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. It is also worth noting that 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),

$$P_t = \theta E_t [P_{t+1} + D_{t+1}]$$  \hspace{1cm} (10)

$$D_t = A(L)v_t$$  \hspace{1cm} (11)

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 (see Appendix A):

$$P_t = P(L)v_t = \left( \frac{\theta A(L)}{L - \theta} - \frac{\theta A(\theta)}{L - \theta} \right) v_t$$  \hspace{1cm} (12)

$$= \sum_{j=1}^{\infty} \theta^j D_{t+j} - \left( \frac{\theta A(\theta)}{L - \theta} \right) v_t$$

$$= P^*_t - P^*_R$$  \hspace{1cm} (13)

The equilibrium price $P_t$ can be decomposed into two components—the perfect foresight price $P^*_t$ and a remainder term $P^*_R$. 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, $P^*_R$ represents a conditioning down term that is orthogonal to information known at $t$. This orthogonality condition implies the following well-known inequality for the variance of the price:

$$\text{var}(P^*_t) > \text{var}(P_t)$$  \hspace{1cm} (14)

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 solid line represents Shiller’s perfect foresight price using an interest rate of $r = 0.048$ and the dashed line is S&P 500 price data. As the figure strongly suggests, the inequality in (14) is 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 actual prices 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 PDV 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 $s,q,r \geq 0$. This $s$’th difference is assumed without loss of generality to have a zero mean. All variables are assumed to be identically zero for $t \leq q$. 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 $\theta$ be a positive constant, $0 \leq \theta < 1$. Let $\Pi(\cdot|\cdot)$ denote linear projections, calculated for $s > 0$ as in Hansen and Sargent

\(^8\)Flavin (1983) and Kleidon (1986) document that the perfect foresight calculations of Shiller lead to small sample bias in tests of excess volatility.
(1980). Let
\[ x_{t,t} = \lim_{k \to \infty} \Pi \left( \sum_{j=0}^{k} \theta^j d_{t+j} | I_t \right), \quad x_{t,H} = \lim_{k \to \infty} \Pi \left( \sum_{j=0}^{k} \theta^j d_{t+j} | H_t \right) \]

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 \( P_{tR}^I \) term represents the conditioning down and will be larger for the smaller information set, \( H_t \). From (14), \( \text{var}(P_{tR}^I | I_t) < \text{var}(P_{tR}^I | H_t) \) if \( H_t \subset I_t \). Therefore any two nested information sets can be used to test for excess volatility.9 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 \( \hat{\sigma}^2_u \) where \( \hat{\sigma}^2_u \) is an unbiased estimator for the second moment of the residual in the regression (16). This follows from the expansion of (10),
\[ P_t = \theta(P_{t+1} + D_{t+1}) - \theta[P_{t+1} + D_{t+1} - E(P_{t+1} + D_{t+1} | I_t)] \]
\[ = \theta(P_{t+1} + D_{t+1}) + u_{t+1} \]
\[ \sigma^2_u = \theta^2[P_{t+1} + D_{t+1} - E(P_{t+1} + D_{t+1} | I_t)]^2 \] (18)

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

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
\[ \sqrt{V(H_{t-1})} \equiv E[P_t - \Pi(P_t | H_{t-1})]^2 = E[P(L)v_t - L^{-1}(P(L) - P(0))v_{t-1}]^2 \]
\[ = P(0)^2 \sigma^2_v = A(\theta)^2 \sigma^2_v \]
\[ = ((1 - \hat{\theta})^s (1 - \sum_{j=1}^{q} \hat{\theta}^j \phi_j))^{-2} \hat{\sigma}^2_v \] (19)

This nests the excess volatility calculation behind Figure 2. When \( I_t \) contains all past, current and future innovations \( \{v_t\}_{t=-\infty} \), the forecast error and variance is identically zero.
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

\[
E[P_t - \Pi(P_t|H_{t-1})]^2 - E[P_t - \Pi(P_t|I_{t-1})]^2,
\]

which we have established as

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

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

(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 than 60 observations and the 47 firms with continuous observations. The last three columns report the number of firms which cannot reject excess volatility \( (H_0 : \Upsilon \geq 0) \) for 1%, 5%, and 10% significance levels. The remaining rows report the equal and cap weight average using the cross-sectional average, \( \bar{D_t}/\bar{P_t} \), of both sets of individual firms in addition to the S&P 500 index for the sample beginning in 1872 and 1926.

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
Table 3: Volatility Test Results. This table reports the statistic from equations (20) (Υ) and (21) (˜Υ). For individual firms, we report the number out of the 278 and 47 firms that cannot reject the excess volatility test \( H_0 : \text{Υ} \geq 0 \) for various significance levels.

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. Similar to the regression analysis, the extent of aggregation plays an important role in excess volatility. The present value model miserably fails to account for the dynamics of broad aggregate measures such as the S&P 500 Index while failing much less spectacularly for portfolios of the individual firms.\(^\text{10}\)

3 Theory

We now propose a theory based on incomplete information and market segmentation that is able to account for the empirical results of Section 2. Our focus is on explaining the difference in results between firm-level data and aggregate measures of the same firm-level data. To that end, suppose there are \( I \) firms indexed by \( i = 1, 2, ..., I \) with stocks priced according to the linear present value model

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

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

\(^{10}\)We conducted several additional tests affirming Samuelson’s Dictum. Examples include Campbell-Shiller VAR regressions and alternative aggregation measures (e.g., mean-variance portfolios). Results available upon request.
an idiosyncratic shock ($\eta_{it}$) and is given by

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

$$a_t = \rho a_{t-1} + \varepsilon_t \quad (24)$$

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 Complete Information

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

$$P_t|\mathcal{F}_t = \left( \frac{\theta \rho}{1 - \theta \rho} \right) a_t \quad (25)$$

Now consider the implications of forming the cross-sectional average, “Equal Weight 1”, $D_t = (1/I) \sum_{i=1}^I D_{it}$, $P_t = (1/I) \sum_{i=1}^I P_{it}$. With a sufficiently large $I$, the law of large numbers is operational and the idiosyncratic shock washes out of the average dividend, $D_t \equiv (1/I) \sum_i D_{it} = a_t$.

Pricing such a dividend stream according to the linear present value model would yield the rational expectations equilibrium,

$$P_t|\mathcal{F}_t = \left( \frac{\theta \rho}{1 - \theta \rho} \right) a_t = \left( \frac{1}{I} \sum_{i=1}^I \right) P_{it}|\mathcal{F}_t \quad (26)$$

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. A full-information rational expectations equilibrium cannot explain Samuelson’s Dictum.

### 3.2 Incomplete Information & Specialization

We now impose two assumptions, which we state explicitly.

**Assumption 1: Market Segmentation.** We assume traders specialize in individual stocks. Trader $i$ follows firm $i$ and only trades firm $i$’s stock.

**Assumption 2: Incomplete Information.** Trader $i$ has incomplete information in that her information set consists of current and past dividends, $\mathcal{I}_{it} = \{D_{it}, D_{it-1}, \ldots\}$.

Assumption 2 challenges the notion that traders can perfectly distinguish shocks that affect only individual firms from shocks that impact the market more broadly. Assumption 1 challenges the
standard assumption that aggregate indices (e.g., S&P 500) are the actively traded commodity. These assumptions are stringent and are imposed in order to keep the algebra of the paper-and-pencil variety. Kasa et al. (2014) and Rondina and Walker (2017) show how to allow for stock prices to enter Trader \( i \)'s information set while preserving results. Maintaining incomplete information would require an additional noise component which would serve to only make the algebra much less tractable but would not significantly alter our theoretical results. Market segmentation can also be relaxed with traders forming optimal portfolios from multiple stocks along the lines of Glasserman and Mamaysky (2019). They show in a model of portfolio choice that traders will endogenously specialize in either macro or micro information, and derive conditions under which it takes more effort to acquire information about individual stocks than market aggregates. This asymmetry provides the incentive for traders to specialize and acquire micro-level information, and serves to justify our assumptions. Additionally, one robustness check of our empirical results described above constructed mean-variance portfolios and showed that Samuelson’s Dictum is preserved. As long as the traders are not pricing only the aggregate measure, our theoretical results will continue to hold.

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 \( \mathcal{I}_{it} = \{D_{it}, D_{it-1}, \ldots\} \) with the dividend process given by (23), then the equilibrium price is given by

\[
P_{it|\mathcal{I}_{it}} = \left( \frac{\theta(\rho - \lambda)}{(1 - \theta \rho)(1 - \lambda L)} \right) a_t + \left( \frac{\theta(\rho - \lambda)}{(1 - \lambda L)} \right) \eta_{it}
\]

where

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

**Proof:** See Appendix A.

Notice that elements of the idiosyncratic shock now bleed into the loading on the aggregate shock \( a_t \) through \( \lambda \), which is a function of the signal-to-noise ratio. Forming the cross-sectional average ("Equal Weight 1") dividend and price

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

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

The aggregate measures are now incongruent in that pricing the dividend process (28) would yield the equilibrium given by (26) and not (29). Therefore while the present value model is an accurate
depiction of individual firms, any test of the linear present value model based upon aggregate measures, (28) and (29), would incorrectly reject the theory of the linear present value model of stock prices. By aggregating the price and dividend process, the econometrician is not capturing the information set of the individual traders.

3.3 Regression Analysis Consider the correlation between the equilibrium price and the discounted expected value of dividends, \( g_t = \sum \theta^j D_{t+j} \) (a slightly modified form of the \( g \) used in Section 2). Note that these correlations drive the regression coefficients tested in Section 2.

Rational expectations implies that at the individual firm level, the covariance between the price and discounted sum of dividends is equal to the variance of the price. This result holds at the individual firm level independent of the information structure. Therefore, the regression analysis of individual firms similar to those examined in Section 2 will yield a regression coefficient of (minus) 1 in theory. However, this result is only applicable under aggregation when information is complete. When information is incomplete, the typical correlation structure generated by the rational expectations equilibrium breaks down and regression coefficients that deviate from (minus) one are possible. We show this formally as a proposition.

**Proposition 3.** Define \( g_{it} \equiv \sum_{j=1}^{\infty} \theta^j D_{it+j} \) and \( g_t \equiv \sum_{j=1}^{\infty} \theta^j D_{t+j} \), where \( D_t = (1/I) \sum_i D_{it} | I_{it}. \)

Let \( P_{it} | F_t \) and \( P_{it} | I_{it} \) be the complete and incomplete information rational expectations equilibrium, respectively; then,

\[
\begin{align*}
\text{Cov}(P_{it}, g_{it} | F_t) &= \text{Var}(P_{it} | F_t) \\
\text{Cov}(P_{it}, g_{it} | I_{it}) &= \text{Var}(P_{it} | I_{it}) \\
\text{Cov}(P_t, g_t | F_t) &= \text{Var}(P_t | F_t) \\
\text{Cov}(P_t, g_t | I_{it}) &= \text{Var}(P_t | I_{it}) \left( \frac{(1 - \lambda^2)}{(\rho - \lambda)(1 + \lambda \rho)} \right)
\end{align*}
\]

**Proof:** See Appendix A.

That the covariance of the price and discounted sum of dividends equals the variance of the price is another way of saying that the discounted sum of dividends accurately reflects the price, i.e., \( \text{Cov}(P_{it}, g_{it}) = \text{Cov}(P_{it}, P_{it}) = \text{Var}(P_{it}) \). This is true for all but the last term, which shows that the correlation between equilibrium price \( (P_t) \) and discounted sum of dividends formed by the econometrician \( (g_t | I_{it}) \) is equal to the variance of the price formed by the econometrician \( (P_t | I_{it}) \) and a correction term that accounts for the discrepancy between the econometrician’s information set and that of private agents. This correction term can be positive or negative contingent on the values of \( \rho \) and \( \lambda \).

To test the extent to which our theory can account for the type of discrepancy between the micro and macro estimates found in Table 2, we conduct the following Monte Carlo thought experiment that exactly replicates the empirical procedure of Section 2.2:

1. Simulate data for several hypothetical firms from the model under incomplete information.
Table 4: Simulated Regression Results. This table reports the slope coefficients ($\beta$) of the regression of future dividend growth on current dividend-price ratio for individual firms (averaged, “Ind.”) and for equal-weighted aggregates (“Agg.”). Results are ordered according to truncation parameter $K$ and signal-to-noise ratio (“high” or “low”).

The dividend process is drawn from (23)–(24) and the price process from (27).

2. Calculate the “Equal Weight 1” dividend-price ratio,

$$\overline{D_t}/\overline{P_t} = \frac{\sum_{i=1}^{I} D_{i,t}}{\sum_{i=1}^{I} P_{i,t}}$$

3. Calculate the truncated proxy for the dividend growth given by (5), which we repeat here for convenience,

$$\hat{g}^D_t = \sum_{k=1}^{K} (\Delta D_{t+k}/P_t)/(1 + r_t)^{k-1}$$

4. Estimate the regression, $g^D_t = \alpha + \beta(D_t/P_t) + \varepsilon_t$, for the individual firms and the equal-weighted average. Test the theoretical prediction of $\beta = -1$.

Table 4 presents the results for the simulated data. Entries are the slope coefficients of the regression of $\hat{g}^D_t$ on the dividend price ratio ($D_t/P_t$) for various truncation parameters, $K$. The dividend processes was calibrated to match the estimated persistence of the Equal-Weight 1 Balanced Panel dividend; that is, we estimated an AR(1) process for the aggregate dividend using the data from the Equal-Weight 1 Balanced Panel and used the same autocorrelation coefficient for the simulation, $\rho = 0.95$. In order to assess the importance of the signal extraction problem, we examined both a “high noise” and a “low noise” estimate. Under both calibrations, we assume the majority of the variance of the dividend process is attributable to aggregate factors. This assumption comes from a principal component analysis of the 42 dividend series of firms that survived the entire sample; only one factor was selected. The high noise calibration $\{\sigma^2_i = 0.8, \sigma^2_\varepsilon = 0.05\}$ (low noise calibration $\{\sigma^2_i = 0.2, \sigma^2_\varepsilon = 0.05\}$) implies 86% (72%) of the variance is attributable to the aggregate shock. This calibration implies a $\lambda$ of 0.77 in the low-noise case and 0.921 in the high-noise case. Following Jung and Shiller (2005), we set $\theta = 0.94$. Our simulations assume 100 firms and at least 8,000 observations.

Table 4 shows that our theory is largely able to capture the empirical results of Table 2. The aggregate estimated coefficients are nearly 50% below the true coefficient of -1 for $K = 10$. The
bias is not as severe as $K$ increases but remains significantly different from the true value, which is also consistent with Table 2. In order to match the highest values for Cap Weight 1 in Table 2, we would need to increase the variance of the idiosyncratic shock such that a majority of the total variance of the dividend process is attributable to the idiosyncratic component. The larger the firm (in terms of cap size), the more realistic is this assumption. Thus, our theory can also explain why the Cap Weight estimate is higher than the Equal Weight estimate in Table 2, and further from -1. Finally, note that Table 4 suggests the approximation error associated with the truncation parameter is not substantial.

3.4 Excess Volatility  Much like the regression analysis, the proposed theory can help explain the excess volatility apparent in the aggregate measures but absent from individual firm observations. Recall the variance bound from Section 2.3,

$$\text{Var}(P_t^*) = \text{Var}(P_t) + \text{Var}(P_t^R) \geq \text{Var}(P_t)$$  \hspace{1cm} (31)

where $P_t^*$ is defined as the perfect foresight price, $P_t^R$ is the conditioning down term and the equilibrium price is given by the difference between the two, $P_t = P_t^* - P_t^R$. The bound is driven by an informational argument. Under standard assumptions, traders use all available information to price the asset, $P_t = \mathbb{E}_t(P_t^*)$, ensuring orthogonality between $P_t$ and $P_t^R$ and therefore $\text{cov}(P_t^R, P_t) = 0$, which proves the bound (31). Under Assumptions 1-2, the variance bound applies to the pricing of individual firms, even when information is incomplete, because traders are using all available information to price the asset. Micro efficiency holds. When information is complete, the bound also holds in the aggregate. Substituting the value for the dividend process (23) and (24) into the equal-weighted aggregate price $P_t | \mathcal{F}_t = (1/I) \sum_{i=1}^I P_i | \mathcal{F}_t$ and calculating the variance yields,

$$\text{Var}(P_t^* | \mathcal{F}_t) = \text{Var}(P_t | \mathcal{F}_t) + \frac{\theta^2 \sigma_x^2}{(1 - \theta \rho)(1 - \theta^2)}$$  \hspace{1cm} (32)

However, the argument does not apply to an aggregate index in which traders have incomplete information. Under these assumptions, an econometrician would aggregate the dividend process, $D_t | \mathcal{I}_{it} = (1/I) \sum_{i=1}^I D_i | \mathcal{I}_{it}$, in positing the perfect foresight price for the index

$$P_t^* | \mathcal{I}_{it} \equiv \sum_{j=1}^{\infty} \theta^j D_{t+j} = \frac{\theta \varepsilon_t}{(1 - \rho L)(L - \theta)}$$  \hspace{1cm} (33)

This is inconsistent with the true aggregate perfect foresight price; that is, $P_t \neq \mathbb{E}_t(P_t^* | \mathcal{I}_{it})$, where the conditional expectation is taken with respect to the econometrician’s information set. By averaging over the dividends in (33), the econometrician eliminates the effect of the idiosyncratic shocks. When information is complete, this is harmless as traders can distinguish between aggregate and idiosyncratic shocks. However with incomplete information, idiosyncratic shocks influence the aggregate price and thus, the econometrician’s information set is inconsistent with the traders’.
To get a quantitative assessment of the excess volatility apparent in Figure 2 and Table 3, we follow the same simulation steps as in Section 3.3 and invoke the test of West (1988). Recall our simulation contained a high noise calibration \( \sigma_i^2 = 0.8, \sigma_e^2 = 0.05 \) and a low noise calibration \( \sigma_i^2 = 0.2, \sigma_e^2 = 0.05 \), with 100 firms and 8,000 observations per firm. Recall also that the present value model can be tested based on the sign of the statistic \( E[P_t - \Pi(P_t | H_{t-1})]^2 - E[P_t - \Pi(P_t | I_{t-1})]^2 \), which Section 2.3 establishes as

\[
\Upsilon = \mathbb{V}(H_{t-1}) - \mathbb{V}(I_{t-1}) = (1 - \hat{\theta})(1 - \sum_{j=1}^{q} \hat{\phi}_j) - \hat{\theta} - 2\hat{\sigma}_v^2 - 2\hat{\sigma}_u^2
\]

where \( H_{t-1} \) and \( I_{t-1} \) are nested information sets, \( H_t \subset I_t \). 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 average statistic for the individual firm for the high-noise (low-noise) case is \( \Upsilon = 0.1759 \) (0.125), indicating no excess volatility. Averaging across firms gives a statistic of \(-0.3034\) in the high-noise case and \(-0.2702\) under low noise. One can also calculate an approximate percentage of the variation that is excessive, \( \tilde{\Upsilon} \) defined in (21), which is 25.89% in the low noise case and 29.24% in the high-noise calibration. While our theory is able to qualitatively deliver the correct sign, we are unable to generate the magnitude of excess volatility observed in data. Under current assumptions, matching the S&P 500 data would require an idiosyncratic shock variance that is roughly 25 times larger than the aggregate shock variance. While this is not consistent with data, if we were to allow traders to observe stock prices, we would need additional noise to ensure incomplete information persists in equilibrium. Kasa et al. (2014) show that such a model is capable of generating excess volatility consistent with data.

4 Connections to the Literature and Concluding Thoughts

Stylized facts distinguishing individual stock price behavior from that of portfolios are now well established and contained in many popular textbooks [e.g., Campbell et al. (1997)]. A careful review of all papers that hint at discrepancies between firm-level data and aggregate data would take us too far afield. Indeed, subtle manifestations of Samuelson’s Dictum have appeared throughout the empirical finance literature, even in its infancy. As an example, early tests of the “multiple security expected return model” demonstrated important statistical differences between individual stock behavior and that of portfolios; Fama (1970) credits Kendall (1953) and King (1966). Kendall (1953) showed that individual stock price changes demonstrated no serial correlation (micro efficiency) but a high degree of cross-correlation existed between returns of different securities (potential macro inefficiency). Using factor analysis on sixty NYSE stocks, King (1966) found that market factors accounted for 50% of the variation and industry factors accounted for 10%, suggesting that aggregate shocks were important drivers of individual stock prices. While the literature is too voluminous to
cite every worthy example, very few papers have confronted the aggregation issue head-on which is our primary interest here. Jung and Shiller (2005) is an important exception.

One empirical section of our paper, Section 2.2, most resembles Jung and Shiller (2005). However, we extend their empirical results along several dimensions. First, we analyze much more data by including all firms that paid a dividend between 1926 and 2018. Our sample size of 227,136 firm-year observations is several orders of magnitude larger than their 3,995 observations. Second, we conduct additional tests like West’s (1988) test of excess volatility to substantially buttress Jung and Shiller’s initial findings. Related empirical papers include Campbell (1991), Vuolteenaho (2002), Cohen et al. (2003), and Startz and Tsang (2014), which we discuss in turn. Startz and Tsang (2014) is the only other empirical paper to test specifically for Samuelson’s Dictum. They use a slightly different methodology and find that the result is contingent on investment horizon. While our results also depend on horizon (e.g., Table 2), our findings are much more robust. Campbell (1991), Vuolteenaho (2002) and Cohen et al. (2003) employ vector autoregression analysis along the lines of Campbell and Shiller (1987) and Campbell and Shiller (1988b) to document the extent to which variation in individual firms’ stock prices are driven by fundamentals. Using variance decomposition analysis, Campbell (1991) attributed one third to one half of the variance of unexpected stock returns to news about future cash flows. The remainder was attributed to news about future expected returns. 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 et al. (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. Thus, these papers show that relative prices across firms are explained by cash flows news (micro efficiency). Our relative contribution to the literature—both theoretically and empirically—is to show what happens as you aggregate these individual firms. An interesting question that remains for future: At what point does information about individual firms become dominated by aggregate considerations?

Our empirical and theoretical results have much in common with a dynamic income-consumption puzzle studied several decades earlier. Early tests of Hall’s (1978) permanent income model were clearly rejected when examined with aggregate U.S. data—changes in aggregate consumption were highly correlated with lagged income changes. However, the model’s orthogonality conditions could not be rejected when tested with individual data [Deaton (1992)]. Goodfriend (1992) and Pischke (1995) hypothesized that incomplete information facing individual households could bias the econometrician’s aggregate tests.11 Similar to our Assumption 2, Pischke (1995) assumed that household income was driven by an individual and an aggregate component, and that households could not distinguish between the two. Goodfriend (1992) achieved a nearly identical result by assuming that

---

11In an earlier contribution, Deaton (1991) also showed how uncertainty of the income process, coupled with the liquidity constraints of Zeldes (1989), could cause the disparity between micro and macro data.
households could not observe aggregate data contemporaneously. Both papers demonstrate how individual tests of the linear present value model of consumption could not be rejected, while an econometrician—imposing a representative agent / complete information setup—would find substantial violations of the theory with aggregate data. Thornton (2014) provides an update of these results and connects the theory to Granger’s (1980) aggregation of long-memory processes. He shows how incomplete information can cause aggregate measures to display substantial persistence even when the micro data follow random walks.

In addition to Goodfriend (1992) and Pischke (1995), our theory has antecedents in Phelps (1969) and Lucas (1975). Our rationale for Assumption 1 (Market Segmentation) can be taken verbatim from Lucas (1975): “The problem is that in an economy in which all trading occurs in a single competitive market, there is ‘too much’ information in the hands of traders for them ever to be ‘fooled’ into altering real decision variables.” If agents in our model traded in a single market—the aggregate index of all firms—there would be too much information and macro efficiency would attain. Theoretical justification for both Assumptions 1 and 2 is provided by Glasserman and Mamaysky (2019), who study information acquisition and portfolio choice with idiosyncratic and aggregate shocks. Their primary result is that traders specialize in either macro or micro information, and several of their corollary results support Samuelson’s Dictum.

The influence of Phelps (1969) and Lucas (1975) has had a resurgence in the macroeconomics literature. The excellent handbook article by Angeletos and Lian (2016) documents how incomplete information “offers a useful method for introducing frictions in coordination and for enriching the dynamics of expectations in macroeconomic models. This enrichment leads to the questioning of existing interpretations of, and helps shed new light on, important phenomena such as business cycles and crises.” We believe our results harbor far-reaching consequences that may question existing interpretations. Many decades ago, the rational expectations, present-value model was deemed anomalous and a notable shift toward relaxing defining assumptions ensued. However, nearly all of the prominent papers in the empirical finance literature framed the debate through the lens of an aggregate index or portfolio of assets. Our results suggest that a combination of incomplete information and aggregation is worthy of further investigation.

Acknowledgments

For helpful comments, we would like to thank Marios Angeletos, Jesús Fernandez-Villaverde, Marvin Goodfriend, Grey Gordon, Jennifer La’O, Eric Leeper, Jianjun Miao, and Harald Uhlig. We also thank participants at the SED Annual Conference, the Econometric Society Meetings, Indiana University, Miami University, University of Texas, the SWUFE International Macro-Finance Conference, University of Tokyo, Osaka University and Kansai University.
References


5 Appendix A: Derivations and Proofs

5.1 Deriving (3) from (2) The present-value model of stock prices is given by

\[ P_t = \mathbb{E}_t \sum_{j=1}^{\infty} (1 + r_t)^{-j} D_{t+j} \] (34)

Multiple by \( r_t \) and add \( D_t \) to both sides of (34).

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

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

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

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

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

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

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

\[
= r_t - \mathbb{E}_t g_t \text{ } (35)
\]

5.2 Deriving the Hansen-Sargent Formula The present value model of stock prices is given by equations (10)–(11), which we replicate here for convenience,

\[ P_t = \theta \mathbb{E}_t[P_{t+1} + D_{t+1}] \] (10)

\[ D_t = A(L)v_t \] (11)

For this derivation, we assume traders have complete information and observe the sequence of shocks \( F_t = \{v_t, v_{t-1}, v_{t-2}, \ldots\} \). This suggests an equilibrium price that is a linear function of \( F_t \); namely, \( P_t = P(L)v_t \).
Expectations are taken using the Wiener-Kolmogorov optimal prediction formula,

\[ \mathbb{E}(P_{t+1}|F_t) = L^{-1}[P(L) - P_0] \varepsilon_t \]
\[ \mathbb{E}(D_{t+1}|F_t) = L^{-1}[D(L) - D_0] \varepsilon_t \]

Substituting these into the equilibrium (10) and using the z-transform

\[ P(z) = \theta(z^{-1}[P(z) - P_0] + z^{-1}[D(z) - D_0]) \]
\[ zP(z) = \theta P(z) + \theta[D(z) - D_0] - \theta P_0 \]
\[ (z - \theta)P(z) = \theta[D(z) - D_0] - \theta P_0 \]  
\[ z \mathbb{E}(P_{t+1}|F_t) = \theta \mathbb{E}(P_{t+1}|F_t) + \theta \mathbb{E}(D_{t+1}|F_t) - \theta \mathbb{E}(P_{t+1}) \]
\[ z \mathbb{E}(D_{t+1}|F_t) = \theta \mathbb{E}(D_{t+1}|F_t) - \theta \mathbb{E}(D_{t+1}) \]

The goal is to solve for \( P(z) \) as a function of the exogenous processes. Note that \( \theta \in (0, 1) \) implies that \( P(z) \) will not be analytic inside the unit circle. Analyticity in the space of z-transforms is tantamount to stationarity in the time domain. This instability can be offset by \( P_0 \), which is a free parameter. Evaluating at \( z = \theta \) gives \( P_0 = D_0 - D(\theta) \) and the solution is

\[ P(z) = \frac{\theta D(z) - \theta D(\theta)}{z - \theta} \]  
\[ \mathbb{E}(P_{t+1}|F_t) = \frac{\theta \mathbb{E}(P_{t+1}|F_t) + \theta \mathbb{E}(D_{t+1}|F_t) - \theta \mathbb{E}(P_{t+1})}{z - \theta} \]  
\[ \mathbb{E}(D_{t+1}|F_t) = \frac{\theta \mathbb{E}(D_{t+1}|F_t) - \theta \mathbb{E}(D_{t+1})}{z - \theta} \]

5.3 Proof of Proposition 2 We now prove Proposition 2, which we repeat for convenience.

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

\[ P_{it} = \left( \frac{\theta(\rho - \lambda)}{(1 - \theta \rho)(1 - \lambda L)} \right) a_{it} + \left( \frac{\theta(\rho - \lambda)}{(1 - \lambda L)} \right) \eta_{it} \]

where

\[ \lambda = \frac{1}{2} \left[ \left( \frac{\sigma^2}{\sigma_n^2 \rho} \right) + \left( \frac{1}{\rho} + \rho \right) - \left\{ \frac{\sigma^2}{\sigma_n^2 \rho} + \left( \frac{1}{\rho} + \rho \right) \right\}^2 - 4 \right]^{1/2} \]
Proof. The dividend process is a combination of an aggregate shock and idiosyncratic shock,

\[ D_{it} = a_t + \eta_{it} \quad (23) \]
\[ a_t = \rho a_{t-1} + \varepsilon_t \quad (24) \]

and trader $i$’s information set is given by $I_{it} = \{D_{it}, D_{it-1}, \ldots\}$. Sargent (1987) shows the fundamental Wold representation for (23)–(24) is given by

\[ D_{it} = \left( \frac{1 - \lambda L}{1 - \rho L} \right) \xi_{it} \quad (38) \]

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

\[ \xi_{it} \equiv \frac{\varepsilon_t}{1 - \lambda L} + \left( \frac{1 - \rho L}{1 - \lambda L} \right) \eta_{it} \]

\[ \sigma^2 = \frac{\sigma^2 + \sigma^2 (1 - \rho)^2}{(1 - \lambda)^2} \]

We can use the formulas of the previous subsection, Section 5.2, to derive the equilibrium price by substituting (38) into (37), which yields

\[ P(z) = \theta \left( \frac{1 - \lambda z}{1 - \rho z} \right) \left( \frac{\theta (\rho - \lambda)}{(1 - \rho z)(1 - \rho \theta)} \right) \]

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

\[ = \left( \frac{\theta (\rho - \lambda)}{(1 - \rho L)(1 - \rho \theta)} \right) \left( \frac{\varepsilon_t}{1 - \lambda L} + \left( \frac{1 - \rho L}{1 - \lambda L} \right) \eta_{it} \right) \]

\[ = \left( \frac{\theta (\rho - \lambda)}{(1 - \rho \theta)(1 - \rho L)(1 - \lambda L)} \right) \varepsilon_t + \left( \frac{\theta (\rho - \lambda)}{(1 - \lambda L)(1 - \rho \theta)} \right) \eta_{it} \]

which is consistent with (27).

5.4 Proof of Proposition 3. Define $g_t \equiv \sum_{j=1}^{\infty} \theta^j D_{it+j}$ and $g_t \equiv \sum_{j=1}^{\infty} \theta^j D_{t+j}$, where $D_t = (1/I) \sum_i D_{it} | I_{it}$. Let $P_t | F_t$ and $P_t | I_{it}$ be the complete and incomplete information rational expectations equilibrium, respectively; then,

\[ \text{Cov} \left( P_t, g_t | F_t \right) = \text{Var} \left( P_t | F_t \right) \]

\[ \text{Cov} \left( P_t, g_t | I_{it} \right) = \text{Var} \left( P_t | I_{it} \right) \]

\[ \text{Cov} \left( P_t, g_t | F_t \right) = \text{Var} \left( P_t | F_t \right) \]

\[ \text{Cov} \left( P_t, g_t | I_{it} \right) = \text{Var} \left( P_t | I_{it} \right) \left( \frac{\rho - \lambda}{\rho (1 - \lambda \rho)} \right) \quad (30) \]
The first three equalities are trivial to prove as $g_{it}$ and $g_t$ are accurate reflections of the equilibrium price. That is, $P_{it} = E(g_{it} | F_t)$, $P_{it} = E(g_{it} | I_{it})$, and $P_{it} = E(g_{it} | F_t)$. To show the last equality, note

$$
P_{t | I_{it}} = \left( \frac{\theta(\rho - \lambda)}{(1 - \rho\theta)(1 - \rho L)(1 - \lambda L)} \right) \varepsilon_t
$$

$$
g_{t | I_{it}} = \sum_{j=1}^{\infty} \theta^j a_{t+j} = \left( \frac{\theta}{(1 - \theta L^{-1})(1 - \rho L)} \right) \varepsilon_t
$$

The variance of $P_{t | I_{it}}$ is given by

$$
\text{Var}(P_{t | I_{it}}) = \frac{\theta^2(\rho - \lambda)^2 (1 + \lambda \rho) \sigma^2_{\varepsilon}}{(1 - \theta \rho)^2(1 - \lambda \rho)(1 - \lambda^2)(1 - \rho^2)}
$$

and the covariance term

$$
\text{Cov} (P_{t}, g_{t} | I_{it}) = \frac{\theta^2(\rho - \lambda) \sigma^2_{\varepsilon}}{(1 - \theta \rho)^2(1 - \rho^2)(1 - \lambda \rho)}
$$

A few algebraic manipulations of the above equations delivers (30).