# INFORMATION AGGREGATION BIAS AND SAMUELSON'S DICTUM[*]

Yongok Choi[†]       Giacomo Rondina[‡]       Todd B. Walker[§]

September 2022

## ABSTRACT

Under the assumption of incomplete information, idiosyncratic shocks may not dissipate in the aggregate. An econometrician who incorrectly imposes complete information and applies the law of large numbers may be susceptible to information aggregation bias. Tests of aggregate economic theory will be misspecified even though tests of the same theory at the micro level deliver the correct inference. A testable implication of information aggregation bias is "Samuelson's Dictum" or the idea that stock prices can simultaneously display "micro efficiency" and "macro inefficiency;" an idea accredited to Paul Samuelson. 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 much more accurately 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 typically deviate from the present value model and "excess" volatility is observed; this is especially true for aggregates (e.g., S&P 500) that are used in most asset pricing studies. 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.

Keywords: Aggregation, Incomplete Information
JEL Codes: C43, C33, G14

---

[†]Chung-Ang University, choiyongok@cau.ac.kr,
[‡]UCSD, grondina@ucsd.edu
[§]Indiana University, walkertb@indiana.edu

# 1 INTRODUCTION

Early tests of the Friedman (1957)–Hall (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) proposed "information aggregation bias" as a possible explanation for the discrepancy. They showed how incomplete information facing individual households could bias the econometrician's full-information aggregate tests, while leaving the household-level econometrics correctly specified. The linear present value model of consumption was a good depiction of economic behavior and yet an econometrician—imposing a representative agent / complete information setup—would find substantial violations of the theory with aggregate data.

This "information aggregation bias" result has taken on renewed significance given the recent interest in mapping microeconomic shocks into macroeconomic outcomes under incomplete information [see Angeletos and Lian (2016b) and Section 1.1]. Similar to Goodfriend and Pischke, many recent papers feature models in which idiosyncratic shocks do not succumb to the law of large numbers but instead complicate the aggregation of economic time series. We contribute to this line of work by studying the aggregate implications of incomplete information within the context of an asset pricing model. We believe this setting has the following advantages relative to the more familiar macroeconomic framework. While significant strides have been made over the last three decades in data collection and analysis of household- and firm-level data, testing basic economic theory at both the micro and macro level *simultaneously* remains nontrivial [Meghir and Pistaferri (2011)]. The richness of stock price data, on the other hand, allows us to show definitively how aggregation alters time series properties. 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." Secondly, the theory we use to test our hypotheses is well established as it relies on the long history of testing the efficient markets hypothesis (EMH). While we are not interested in asking whether a stock price is "efficient," we can apply the econometric techniques for testing the EMH to conduct our thought experiments.[1]

In the asset pricing context, testing for the aggregation bias of Goodfriend and Pischke is akin to testing what Jung and Shiller (2005) dub *Samuelson's Dictum*, which articulates Paul Samuelson's belief that market efficiency applies to individual firms (micro efficiency) but is vitiated when stock prices are aggregated to form indexes like the S&P 500 (macro inefficiency).[2] In Section 2, we conduct two familiar tests

---

[1]Our application of West (1988b) 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. These overstatements were shown not to be 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 (1988b)], which do not rely on stationarity, also find evidence for excess volatility in aggregate data.

[2]In a private letter to John Campbell and Robert Shiller, Paul Samuelson wrote: "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," Jung and Shiller (2005).

of the present value model of stock prices: a regression of future dividend changes on the dividend-price ratio, and a test for excess volatility. Both tests confirm Samuelson's Dictum in the follow sense. 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 does a better job of forecasting the future growth in dividends and that the deviation in prices relative to dividends is mostly in line with the discounted present value model. For a significant number of firms, 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. Thus, applying the same econometric methodology to aggregate data delivers sound rejections of a model that cannot rejected at the individual firm level.

We then conduct the "second-generation" volatility test of West (1988b), which is based on an inequality of the variance of the innovation in the expected present value of dividends. The theory states that this variance should be smaller when conditioning 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. Nearly 80% of firms in our sample do not exhibit excess volatility at a 1% significance level. However, portfolios of these same firms almost always 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 linear present value model.

To reconcile our findings, we propose a theory of aggregation bias based on incomplete information and segmented markets in Section 3 that is a generalization of Goodfriend (1992) and Pischke (1995). 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.

### 1.1 CONNECTION TO LITERATURE

The insights of Goodfriend (1992) Pischke (1995) were put on more solid footing by Reis (2006), Luo (2008) and Thornton (2014). Goodfriend (1992) hypothesized that individuals processed aggregate information with a one-period lag and Reis (2006) showed how this "inattentiveness" implied that the sluggishness in aggregate consumption was due to the aggregation step, which proved an important distinguishing feature relative to other modeling assumptions (e.g., habit formation in utility). Pischke (1995) went one step further arguing that when agents faced idiosyncratic and aggregate shocks, ignoring contemporaneous aggregate data may be the optimal choice for households. Specifically referencing Pischke's information structure, Luo (2008) conjectured that, if rational

inattention could be modeled endogenously, "it would be optimal for them to devote low attention to monitoring the aggregate component because the aggregate component is less important for individuals' optimal decisions." Thornton (2014) provides an explicit characterization of the econometrics underlying the aggregation problem 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. Our contribution to this literature is in arguing that these insights and techniques (appropriately modified) can be applied to asset pricing data. As mentioned in the introduction, asset pricing data does not suffer from the same challenges of aggregate and household-level income and consumption; our panel of observations is significantly more than any test of the permanent income hypothesis. Our main result—that informational frictions appear larger at the macro level—is also related to Mackowiak and Wiederholt (2009) where agents rationally choose to pay more attention to idiosyncratic fundamentals, with Angeletos and Lian (2016a) and Angeletos and Huo (2018) where higher-order uncertainty is more pronounced at the aggregate level. However, we contribute to this literature by showing how this result can hold theoretically even when not deviating from rational expectations. Our theoretical interpretation relies simply on a signal extraction problem at the individual trader level, which propagates into aggregate dynamics. The econometrician and private agent are acting optimally, and yet Samuelson's Dictum manifests.

Stylized facts distinguishing individual stock price behavior from that of portfolios or indices are now well established and contained in many popular textbooks [e.g., Campbell et al. (1997), Campbell (2017)]. Manifestations of Samuelson's Dictum can be found in Campbell (1991), Vuolteenaho (2002), Cohen et al. (2003) and Cohen et al. (2009), who 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 idiosyncratic versus aggregate factors. Our econometric approach is different from the papers cited above[3] because we are explicitly testing for aggregation bias, which is not of interest in these papers. Jung and Shiller (2005) is an important exception but we view the evidence provided therein as suggestive, at best. Section 2.2 extends their 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 is several orders of magnitude larger than theirs. Second, we conduct several additional tests like West's (1988b) test of excess volatility, panel regressions, numerous and various forms of aggregation. Additionally, we offer a theoretical explanation for our empirical findings. 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 more robust.

Our theory in Section 3 relies on two assumptions to make information aggregation bias operable. First, traders have incomplete information in that they cannot disentangle idiosyncratic from aggregate shocks, similar to Pischke (1995). An econometrician—imposing complete information—will assume

---

[3]We rely on panel data methods and formal tests of the linear present value model. Our interest in how well the the model fits the data for various levels of aggregation (e.g., Table 2) and how quickly aggregation changes the nature of these results (e.g., Figure 1) clearly distinguishes our empirical work from the papers cited above. Our theoretical explanation also differs from this literature (see, Section 4)

the idiosyncratic shocks "wash out" in the aggregation step and will attribute the idiosyncratic influence to a rejection of the linear present value model. In this sense, our paper contributes to the broader literature on articulating the perils of assuming complete information, representative agent models and has antecedents in Phelps (1983), Forni and Lippi (1999), Blundell and Stoker (2005), and Blanchard et al. (2013), who show how incomplete information coupled with some form of heterogeneity can lead to incorrect inference. The second assumption is market segmentation or that traders specialize in individual stocks. A justification for this assumption is given by Glasserman and Mamaysky (2019), who 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. Unlike Glasserman and Mamaysky (2019), we show how these assumptions can generate time series that are consistent with our empirical findings of Section 2.

## 2 EMPIRICAL EVIDENCE

A testable implication of information aggregation bias is a change in the econometric inference as one aggregates. We provide empirical evidence in favor of information aggregation bias 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

$$P_t = (1 + r)^{-1} \mathbb{E}_t (P_{t+1} + D_{t+1}) \tag{1}$$

$$P_t = \mathbb{E}_t \sum_{j=1}^{\infty} (1 + r)^{-j} D_{t+j} \tag{2}$$

where $r$ is the time-invariant discount rate,[4] $P_t$ 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). The online appendix shows that (2) can be written in dividend-price ratio form as

$$D_t/P_t = r - \mathbb{E}_t g_t^D \tag{3}$$

$$g_t^D \equiv \sum_{k=1}^{\infty} (\Delta D_{t+k}/P_t)/(1 + r)^{k-1} \tag{4}$$

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.

---

[4]While our focus is on a constant discount rate specification, we acknowledge the importance of time-varying rates in explaining asset price behavior. Previous versions of the paper introduced a time-varying discount rate, which only reinforced our results.

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

$$\hat{g}_t^D = \sum_{k=1}^{K} (\Delta D_{t+k}/P_t)/(1+r)^{k-1} \tag{5}$$

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

$$\hat{g}_t^D = \alpha + \beta(D_t/P_t) + \epsilon_t \tag{6}$$

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 intercept term should coincide with the discount rate ($\alpha = r$).

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,

$$\overline{(D_t/P_t)} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{D_{i,t}}{P_{i,t}}\right), \qquad \left(D_t/P_t\right)_{CAP} = \sum_{i=1}^{I}\xi_i\left(\frac{D_{i,t}}{P_{i,t}}\right) \tag{7}$$

where $\xi_i$ denotes the market-capitalization weight assigned to firm $i$ for time period $t$.[5] 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. Therefore in addition to the Jung-Shiller average (7), we construct the cross-sectional averages separately for prices and dividends,

$$\overline{D_t}/\overline{P_t} = \frac{\sum_{i=1}^{I} D_{i,t}}{\sum_{i=1}^{I} P_{i,t}}, \qquad (\overline{D_t}/\overline{P_t})_{CAP} = \frac{\sum_{i=1}^{I} \xi_i D_{i,t}}{\sum_{i=1}^{I} \xi_i P_{i,t}} \tag{8}$$

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 1871 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. 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, $\bar{r} = 0.054$ (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

---

[5]The cap weights are calculated as follows: abs(`PRC`*`SHROUT`) where `PRC`: Price or Bid/Ask Average, `SHROUT`: Shares Outstanding.

Number of Firms (Firm Years)

|  | Firm Age | Non-Dividend | Dividend | Total |
|---|---|---|---|---|
| Discontinuous |  | 612 (8,922) | 394 (8,707) | 1,006 (17,629) |
| Continuous | 1-10 | 12,297 (58,121) | 7,453 (40,965) | 19,750 (99,086) |
|  | 11-20 | 2,641 (37,739) | 4,485 (65,973) | 7,126 (103,712) |
|  | 21-40 | 802 (20,847) | 3,112 (87,200) | 3,914 (108,047) |
|  | 41-60 | 16 (752) | 896 (43,764) | 912 (44,516) |
|  | 61-80 | 1 (69) | 193 (13,223) | 194 (13,292) |
|  | 81-90 | 0 | 48 (4,154) | 48 (4,154) |
|  | 90+ | 0 | 46 (4,266) | 46 (4,266) |
| Total |  | 16,369 (126,450) | 16,627 (268,252) | 32,996 (349,702) |

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.

year dividends were recorded (1926) through 2001. They found that the average estimate for $\beta$ was of the correct sign (negative) and significantly different from zero. 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 ($K = 10$) 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-month observations whose Cumulative Factor to Adjust Prices over a Date Range (`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 `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 32,996 firms, roughly half of these firms have issued ordinary dividends (16,369). A small percentage, 1,006 (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. Table 1 shows the summary statistics for the number of firm-year observations. There are 349,702 firm-year observations in total, and 268,252 (76.7%) are dividend paying firms.

---

[6]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 ($K = 10$).

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 examined firms that had price and dividend data available for their entire sample. This amounts to 47 firms or roughly 1% of our sample when $K = 10$.

Following standard procedure in the asset pricing literature, 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 (0.054) is the average annual return over all firms and dates in the sample. All 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, $D_t/P_t = r - \mathbb{E}_t g_t^D$, where $g_t^D \approx \sum_{k=1}^{K} (\Delta D_{t+k}/P_t)/(1+r)^{k-1}$ for nine different specifications and for four truncation parameters, $K = 10, 20, 30, 50$.[7] Our specifications include balanced and unbalanced panels for all dividend-paying 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, $g_{it} = \alpha + \beta(D_{it}/P_{it}) + \gamma_i + \epsilon_{i,t}$, for firm observations $i = 1, \cdots, I$, annual observations $t = 1, \cdots, T$, and where $\gamma_i$ captures firm-specific effects.

Table 2 reports the results, organized by truncation parameter $K$. 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.[8]

For each $K$, the first three rows of Table 2 report the $\beta$ estimate and 95% confidence interval (calculated using the Newey and West (1987) HAC estimator with data-dependent, Bartlett bandwidth $0.75 T^{1/3} + 0.5$) 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. All estimates for $\beta$ in the first three rows are of the correct sign with the unbalanced panel very close to the hypothesized theoretical value of minus one for smaller values of $K$. For $K = 10$, one cannot reject the two-sided hypothesis $H_0 : \beta = -1$ for the unbalanced panel, even at the 1% significance level. As $K$ increases, the number of observations are substantially reduced and the confidence intervals widen as a result.

---

[7]We set the interest rate to 0.054, which is the average return on assets in our sample. We also examined values for the interest rate as low as 0.045 and as high as 0.064, with little to no change in our results.

[8]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].

| K=10 | # obs. | β | 95% CI | K=20 | # obs. | β | 95% CI |
|---|---|---|---|---|---|---|---|
| Bal. Panel | 3,154 | −0.616 | −0.805, −0.428 | Bal. Panel | 2,774 | −0.778 | −0.959, −0.598 |
| Unbal. Panel | 112,637 | −0.996 | −1.002, −0.990 | Unbal. Panel | 57,467 | −0.864 | −0.909, −0.818 |
| Unbal. ($D > 0$) | 92,253 | −0.996 | −1.002, −0.989 | Unbal. ($D > 0$) | 49,388 | −0.835 | −0.892, −0.778 |
| Equal Weight 1 | 83 | −1.516 | −3.311, 0.280 | Equal Weight 1 | 73 | −1.034 | −1.870, −0.198 |
| Cap Weight 1 | 83 | −0.370 | −0.591, −0.149 | Cap Weight 1 | 73 | −0.332 | −0.521, −0.144 |
| Equal Weight 2 | 83 | 0.257 | −0.043, 0.556 | Equal Weight 2 | 73 | 0.400 | 0.063, 0.738 |
| Cap Weight 2 | 83 | 0.018 | −0.257, 0.292 | Cap Weight 2 | 73 | 0.151 | −0.370, 0.672 |
| S&P 500 (1926) | 83 | 0.023 | −0.274, 0.320 | S&P 500 (1926) | 73 | 0.178 | −0.258, 0.613 |
| S&P 500 (1871) | 138 | 0.036 | −0.181, 0.253 | S&P 500 (1871) | 128 | 0.046 | −0.197, 0.290 |

| K=30 | # obs. | β | 95% CI | K=50 | # obs. | β | 95% CI |
|---|---|---|---|---|---|---|---|
| Bal. Panel | 2,394 | −0.860 | −1.005, −0.668 | Bal. Panel | 1,634 | −0.757 | −1.065, −0.448 |
| Unbal. Panel | 30,427 | −0.804 | −0.880, −0.728 | Unbal. Panel | 7,483 | −0.749 | −0.869, −0.629 |
| Unbal. ($D > 0$) | 27,418 | −0.788 | −0.874, −0.702 | Unbal. ($D > 0$) | 7,148 | −0.736 | −0.862, −0.610 |
| Equal Weight 1 | 63 | −0.771 | −1.224, −0.318 | Equal Weight 1 | 43 | −0.610 | −0.898, −0.323 |
| Cap Weight 1 | 63 | −0.228 | −0.419, −0.038 | Cap Weight 1 | 43 | −0.321 | −0.516, −0.127 |
| Equal Weight 2 | 63 | 0.535 | 0.034, 0.910 | Equal Weight 2 | 43 | 0.745 | 0.382, 1.108 |
| Cap Weight 2 | 63 | 0.259 | −0.258, 0.777 | Cap Weight 2 | 43 | 0.119 | −0.283, 0.520 |
| S&P 500 (1926) | 63 | 0.222 | −0.219, 0.662 | S&P 500 (1926) | 43 | 0.205 | −0.242, 0.652 |
| S&P 500 (1871) | 118 | 0.062 | −0.226, 0.350 | S&P 500 (1871) | 98 | 0.102 | −0.161, 0.364 |

Table 2: Regression Results. This table reports the results of the regression of future dividend growth on current dividend-price ratio for individual firms and 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 corresponding to aggregate measures. Standard errors are calculated using a Newey-West HAC estimator with data-dependent, Bartlett bandwidth.

However, these results 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 unbalanced panel. We also report results for the S&P 500 Index, examining both the same time horizon as our sample (1926) and the full sample which dates back to 1871.

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 is not tightly estimated with relatively large 95% confidence intervals. Moreover, the estimate for the S&P 500 Index is not even of the correct sign for any value of $K$. The present value model for individual stock prices and dividends is more consistent using the disaggregated data, while any form of aggregation leads to a 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;

| | | $\beta$ | | $H_0 : \beta = -1$ $H_1 : \beta \neq -1$ | | | $\alpha$ | |
|---|---|---|---|---|---|---|---|---|
| $K$ | # obs. | Mean | Std | 1% | 5% | 10% | Mean | Std |
| 10 | 3,858 | -0.642 | 0.651 | 33.9% | 45.9% | 53.3% | 0.028 | 0.026 |
| 20 | 1,936 | -0.515 | 0.932 | 37.3% | 48.7% | 55.7% | 0.037 | 0.040 |
| 30 | 1,075 | -0.317 | 1.127 | 41.8% | 55.3% | 62.2% | 0.042 | 0.047 |
| 50 | 277 | -0.326 | 0.977 | 41.9% | 52.7% | 59.2% | 0.057 | 0.056 |

Table 3: Individual Firm Regression Results. This table reports the results of the regression of future dividend growth on current dividend-price ratio as given by (6) for individual firms. Standard errors are calculated using a Newey-West HAC estimator with data-dependent, Bartlett bandwidth with 5% winsorization.

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.[9]

Table 3 estimates equation (6), $g_t^D = \alpha + \beta(D_t/P_t) + \epsilon_t$, individually for each firm and then aggregates the results, as opposed to the panel regression of Table 2. In order to mitigate the influence of outliers, we apply a 5% winsorization.[10] While not as close to the theoretical value of -1 displayed by the panel regression estimates, well over half of all firms cannot reject the null hypothesis $H_0 : \beta = -1$ at the 10% significant level. Similar to previous findings, the estimates are less consistent with the theory and more dispersed as the horizon $K$ increases. Here we also report the estimates for the intercept term, $\alpha$. While the average roughly doubles from 0.028 ($K = 10$) to 0.057 ($K = 50$), the estimates are in-line with the average 10-year yield on US Treasuries over this time period, 0.0362.

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, 36 firms who survived the entire sample from 1926 through 2018 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 $K = 10$ (solid line with dashed 95% confidence intervals). The individual estimates for the firms are plotted using dot-whiskers, which represents the $\beta$ estimate and 95% confidence interval. The horizontal axis indicates the number of firms combined to form the equal-weighted portfolio. As the figure shows, the aggregate estimate turns positive after 13 firms, while the individual firm regressions do not turn positive until the 33rd firm. A portfolio of the 18 most consistent firms has a $\beta$ that is of the wrong sign and a statistically significant *positive* value, despite the fact that nearly all of these firms cannot reject the null $H_0 : \beta = -1$ at the 10% significant level when estimated individually. This analysis suggests that aggregation leads to a quick and decisive deviation from the constant discount rate present value model, even when the individual regressions are not substantially deviating from the theory.

**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:

---

[9]Additional results available upon request.

[10]A majority of these outliers are firms that left the sample after only a short period of time and hence, may not be indicative of a "representative firm." Moreover, these firms almost never appear in the S&P 500 or other aggregates used to test asset pricing models, and their influence on aggregate results in Table 2 are negligible given the firm-year weighting.
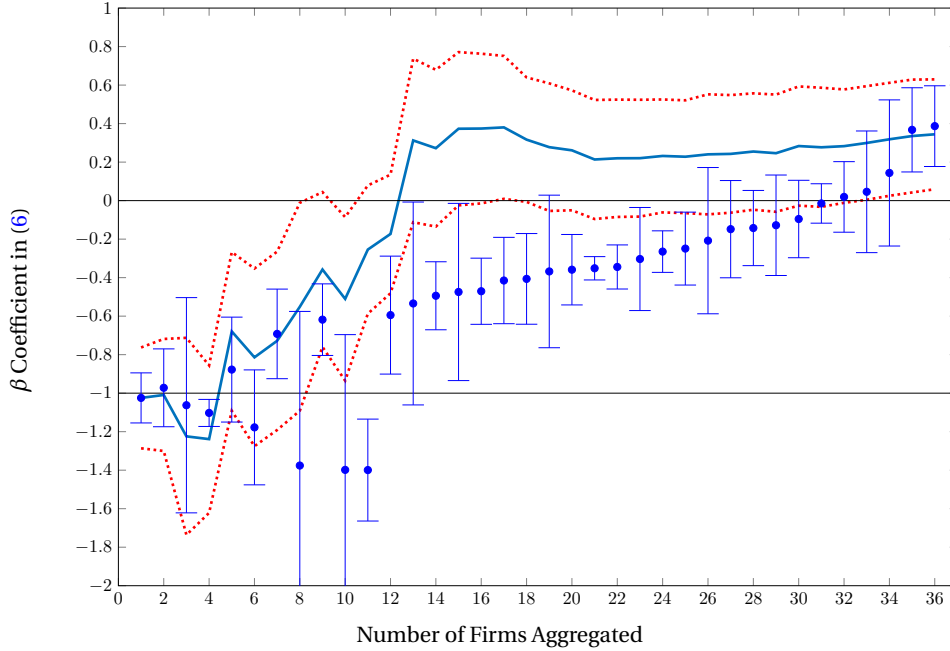
Figure 1: Aggregation Bias. The figure reports the estimates for $\beta$ in regression (6) with $K = 10$ and 95% confidence interval in dashed lines, for equal-weighted portfolios as in (7) of an expanding number of firms. The dot-whiskers plot is the regression coefficient for the individual firm with 95% confidence intervals. 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 run at the individual firm level. The number of firms in the portfolio is indicated on the horizontal axis.

> "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 pric-

ing model of (1),

$$P_t = \theta \mathbb{E}_t[P_{t+1} + D_{t+1}] \tag{9}$$

$$D_t = A(L) v_t \tag{10}$$

where the discount factor $\theta = (1 + r)^{-1}$ is assumed to be constant, $\mathbb{E}$ 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 (10). 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 Online Appendix):

$$P_t = P(L) v_t = \left( \frac{\theta A(L)}{L - \theta} - \frac{\theta A(\theta)}{L - \theta} \right) v_t \tag{11}$$

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

$$= P_t^{\star} - P_t^R \tag{12}$$

The equilibrium price $P_t$ can be decomposed into two components—the perfect foresight price $P_t^{\star}$ 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 is orthogonal to information known at $t$. This orthogonality condition implies the following well-known inequality for the variance of the price:

$$\text{var}(P_t^{\star}) > \text{var}(P_t) \tag{13}$$

The variance of the perfect foresight price is greater than the variance of the equilibrium (or observed) price.

Much of the debate immediately following Shiller (1981) centered around the calculation of $P_t^{\star}$ and the appropriate way to test for excess volatility.[11] A series of papers, including that of West (1988b), corrected for the statistical and measurement issues and all concluded that excess volatility was a robust feature of stock price data. We implement West (1988b) 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 (13), 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 $H_t$ (i.e., $\text{var}(P_t^R | I_t) < \text{var}(P_t^R | H_t)$ if $H_t \subset I_t$). Therefore any two nested information sets can be used to test for excess volatility.

---

[11]Flavin (1983) and Kleidon (1986) document that the perfect foresight calculations of Shiller lead to small sample bias in tests of excess volatility.

In what follows, we use the notation of and implementation strategy of West (1988b) and refer readers to that paper for a full derivation of results discussed below. We construct nested information sets by estimating the following bivariate system

$$P_t = \theta(P_{t+1} + D_{t+1}) + u_{t+1} \tag{14}$$

$$\Delta^s D_{t+1} = \mu + \phi_1 \Delta^s D_t + \cdots + \phi_q \Delta^s D_{t-q+1} + v_{t+1} \tag{15}$$

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 information set $I_t$ 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 (14). Thus, the innovation $u_{t+1}$ contains the price history plus other terms. We label this forecast error as $\mathbb{V}(I_{t-1}) \equiv \mathbb{E}[P_t - \Pi(P_t|I_{t-1})]^2$.

Conversely, using the Hansen and Sargent formula (11), an inferior information set assumes $H_t$ contains only the history of dividends and knowledge of the model structure. The variance of the one-step-ahead forecast error is given by

$$\begin{aligned}
\mathbb{V}(H_{t-1}) \equiv \mathbb{E}[P_t - \Pi(P_t|H_{t-1})]^2 &= \mathbb{E}[P(L)v_t - L^{-1}(P(L) - P(0))v_{t-1}]^2 \\
&= P(0)^2 \sigma_v^2 = A(\theta)^2 \sigma_v^2 \\
&= ((1-\hat{\theta})^s (1 - \sum_1^q \hat{\theta}^j \hat{\phi}_j))^{-2} \hat{\sigma}_v^2
\end{aligned} \tag{16}$$

The first equality of (16) follows from the Wiener-Kolmogorov optimal prediction formula. The final equality comes from (11) and substituting in the stochastic process for the dividends (15). The $\hat{\sigma}_v^2$ is an unbiased estimator for the second moment of the residual in the regression (15), and therefore 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

$$\begin{aligned}
\Upsilon &= \mathbb{V}(H_{t-1}) - \mathbb{V}(I_{t-1}) \\
&= ((1-\hat{\theta})^s (1 - \sum_1^q \hat{\theta}^j \hat{\phi}_j))^{-2} \hat{\sigma}_v^2 - \hat{\theta}^{-2} \hat{\sigma}_u^2
\end{aligned} \tag{17}$$

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

$$\tilde{\Upsilon} = -100 \left( \frac{\mathbb{V}(H_{t-1}) - \mathbb{V}(I_{t-1})}{\mathbb{V}(I_{t-1})} \right) \tag{18}$$

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

|  | $\Upsilon$ | $\tilde{\Upsilon}$ | 1% | 5% | 10% |
|---|---|---|---|---|---|
| Unbalanced Panel (287 firms) | 513.72 | — | 224 | 147 | 100 |
| Balanced Panel (47 firms) | 3,825.64 | — | 33 | 25 | 21 |
| Equal Weight 1 (287 firms) | -12.55 | 68.8 | | | |
| Cap Weight 1 (287 firms) | -12.45 | 44.9 | | | |
| Equal Weight 1 (47 firms) | -7.05 | 18.0 | | | |
| Cap Weight 1 (47 firms) | 73.38 | | | | |
| S&P 500 (1926) | -25,291.15 | 91.6 | | | |
| S&P 500 (1871) | -15,994.52 | 90.9 | | | |

Table 4: Volatility Test Results. This table reports the statistic from equations (17) ($\Upsilon$) and (18) ($\tilde{\Upsilon}$). For individual firms, we report the number out of the 287 and 47 firms that cannot reject the excess volatility test $H_0 : \Upsilon \geq 0$ for various significance levels.

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 an unbalanced panel of 287 firms with more than 60 years of observations. For each firm, we estimate the bivariate system (15) to obtain the corresponding statistics for formula (17). Equation (14) is estimated by Hansen's (1982) two-step, two-stage least squares, using current and lagged values of dividends as instrumental variables. We use the augmented Dickey-Fuller (ADF) test for each dividend series to determine $s$ in the Equation (15). 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 (1988b) and use the delta method to get the variance of the statistic (17) and (18). Moreover, since $P_t$ and $D_t$ arguably have unit roots, standard distribution theory is applicable as shown in West (1988a).

The first two rows of Table 4 report the average West statistic ($\Upsilon$) for the individual firms, while 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, $\overline{D_t}/\overline{P_t}$, of both sets of individual firms in addition to the S&P 500 index for the sample beginning in 1871 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 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 1% significance level. Conversely the statistic for every aggregated measure—with the exception of the cap-weighted portfolio—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 4 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 mea-

sures such as the S&P 500 Index while failing much less spectacularly for portfolios of the individual firms.

# 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

$$P_{it} = \mathbb{E}_t \sum_{j=1}^{\infty} \theta^j D_{it+j} \tag{19}$$

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

$$D_{it} = a_t + \eta_{it} \tag{20}$$

$$a_t = \rho a_{t-1} + \varepsilon_t \tag{21}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\eta_{it} \sim N(0, \sigma_\eta^2)$, 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, $\mathscr{F}_t = \{\varepsilon_t, \eta_{it}, \varepsilon_{t-1}, \eta_{it-1}, ....\}$. The rational expectations equilibrium price for firm $i$ follows from the Hansen-Sargent optimal prediction formula (substituting (20) into (11)),

$$P_{it} | \mathscr{F}_t = \left( \frac{\theta \rho}{1 - \theta \rho} \right) a_t \tag{22}$$

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, $D_t \equiv (1/I) \sum_i D_{it} \approx a_t$. Pricing such a dividend stream according to the linear present value model would yield the rational expectations equilibrium,

$$P_t | \mathscr{F}_t = \left( \frac{\theta \rho}{1 - \theta \rho} \right) a_t = (1/I) \sum_{i=1}^{I} P_{it} | \mathscr{F}_t \tag{23}$$

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, $\mathscr{I}_{it} = \{D_{it}, D_{it-1}, ....\}$.

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.[12] 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 (2021) 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 (20), she cannot distinguish between idiosyncratic and aggregate shocks. The following proposition derives the rational expectations equilibrium price.

**Proposition 1.** *Let Trader $i$'s information set be current and past dividends $\mathscr{I}_{it} = \{D_{it}, D_{it-1}, ....\}$ with the dividend process given by* (20)*, then the equilibrium price is given by*

$$P_{it}|\mathscr{I}_{it} = \left(\frac{\theta(\rho-\lambda)}{(1-\theta\rho)(1-\lambda L)}\right) a_t + \left(\frac{\theta(\rho-\lambda)}{(1-\lambda L)}\right) \eta_{it} \tag{24}$$

---

[12]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.

*where*

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

*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

$$D_t|\mathcal{I}_{it} \equiv (1/I)\sum_{i=1}^{I} D_{it}|\mathcal{I}_{it} \approx a_t \tag{25}$$

$$P_t|\mathcal{I}_{it} \equiv (1/I)\sum_{i=1}^{I} P_{it}|\mathcal{I}_{it} = \left(\frac{\theta(\rho-\lambda)}{(1-\theta\rho)(1-\lambda L)}\right) a_t \tag{26}$$

The aggregate measures are now incongruent in that pricing the dividend process (25) would yield the equilibrium given by (23) and not (26). 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, (25) and (26), 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.

The intuition for this result follows that of Goodfriend (1992) and Pischke (1995), who stress how an information wedge between the behavior of individual consumption dynamics and their aggregate counterparts may arise because individuals fail to perfectly disentangle the idiosyncratic component of income from the aggregate component.[13] This interpretation—applied to asset prices—says that idiosyncratic components should not be incorporated into stock prices, if perfectly observed, (22). Yet, under incomplete information, this component affects individual stock prices because traders use it to predict the aggregate component (via $\lambda$ in (24)), which is not perfectly observed. Therefore, traders do not react to aggregate shocks as they would in a complete information framework where idiosyncratic shocks wash out, and this causes the failure of the present value model.

### 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

---

[13]We give credit to an anonymous referee for suggesting some of the language included in this paragraph.

down and regression coefficients that deviate from (minus) one are possible. We show this formally as a proposition.

**Proposition 2.** *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 $D_t = (1/I) \sum_i D_{it} | \mathscr{I}_{it}$. Let $P_{it} | \mathscr{F}_t$ and $P_{it} | \mathscr{I}_{it}$ be the complete and incomplete information rational expectations equilibrium, respectively; then,*

$$Cov\,(P_{it}, g_{it} | \mathscr{F}_t) = Var\,(P_{it} | \mathscr{F}_t)$$

$$Cov\,(P_{it}, g_{it} | \mathscr{I}_{it}) = Var\,(P_{it} | \mathscr{I}_{it})$$

$$Cov\,(P_t, g_t | \mathscr{F}_t) = Var\,(P_t | \mathscr{F}_t)$$

$$Cov\,(P_t, g_t | \mathscr{I}_{it}) = Var\,(P_t | \mathscr{I}_{it}) \left( \frac{(1 - \lambda^2)}{(\rho - \lambda)(1 + \lambda\rho)} \right) \tag{27}$$

*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 ($g_t | \mathscr{I}_{it}$) is equal to the variance of the price formed by the econometrician ($P_t | \mathscr{I}_{it}$) 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. The dividend process is drawn from (20)–(21) and the price process from (24).

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,

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

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

|  | K=10 | K=20 | K=30 | K=40 | K=50 | K=200 |
|---|---|---|---|---|---|---|
| Ind. (high noise) | −0.993 | −0.998 | −1.001 | −1.008 | −1.005 | −1.000 |
| Agg. (high noise) | −0.508 | −0.611 | −0.658 | −0.645 | −0.649 | −0.659 |
| Ind. (low noise) | −0.956 | −0.996 | −0.997 | −1.004 | −0.999 | −1.008 |
| Agg. (low noise) | −0.519 | −0.692 | −0.716 | −0.741 | −0.725 | −0.746 |

Table 5: 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").

Table 5 presents the results for the simulated data. Entries are the slope coefficients of the regression of $\hat{g}_t^D$ 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 46 dividend series of firms that survived the entire sample; only one factor was selected. The low-noise calibration, $\{\sigma_i^2 = 0.2, \sigma_\varepsilon^2 = 0.05\}$, implies 72% of the variance of the dividend process is attributable to the aggregate shock; while the high-noise calibration, $\{\sigma_i^2 = 0.8, \sigma_\varepsilon^2 = 0.05\}$, attributes 40% 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 $\theta = 0.95$, which is consistent with the average return of the 47 firms that survived the entire sample (inclusive of dividends). Our simulations assume 100 firms and at least 8,000 observations.

We make two points about our calibration. First, note that we attribute a significant amount of the variation in the dividend to the aggregate shock. While this is consistent with the principal component analysis of the 46 dividend series of firms that survived the sample, one can argue that it is too high for the firms that did not persist for the entire sample (see, e.g., Vuolteenaho (2002)). Results presented below improve monotonically as we attribute more of the variance to the individual component, so one can argue that our results are a lower bound. Second, as long as the autocorrelation coefficient for the dividend process remains positive, our qualitative results continue to hold. Moreover, autocorrelation coefficients as low as 0.75 do not substantially alter our quantitative message, which is well within several standard errors of the 0.95 estimate discussed above.

Table 5 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 5 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,

$$\text{Var}(P_t^\star) = \text{Var}(P_t) + \text{Var}(P_t^R) \geq \text{Var}(P_t) \tag{28}$$

where $P_t^\star$ 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, $P_t = P_t^\star - P_t^R$. The bound is driven by an informational argument. Under standard assumptions, traders use all available information to price the asset, $P_t = \mathbb{E}_t(P_t^\star)$, ensuring orthogonality between $P_t$ and $P_t^R$ and therefore $\text{cov}(P_t^R, P_t) = 0$, which proves the bound (28). 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 (20) and (21) into the equal-weighted aggregate price $P_t|\mathscr{F}_t = (1/I)\sum_{i=1}^I P_{it}|\mathscr{F}_t$ and calculating the variance yields,

$$\text{Var}(P_t^\star|\mathscr{F}_t) = \text{Var}(P_t|\mathscr{F}_t) + \frac{\theta^2 \sigma_\varepsilon^2}{(1-\theta\rho)^2(1-\theta^2)} \tag{29}$$

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, $D_t|\mathscr{I}_{it} = (1/I)\sum_i D_{it}|\mathscr{I}_{it}$, in positing the perfect foresight price for the index

$$P_t^\star|\mathscr{I}_{it} \equiv \sum_{j=1}^\infty \theta^j D_{t+j} = \frac{\theta\varepsilon_t}{(1-\rho L)(L-\theta)} \tag{30}$$

This is inconsistent with the true aggregate perfect foresight price; that is, $P_t|\mathscr{I}_{it} \neq \mathbb{E}_t(P_t^\star|\mathscr{I}_{it})$, where the conditional expectation is taken with respect to the econometrician's information set. By averaging over the dividends in (30), the econometrician eliminates the effect of the idiosyncratic shocks entirely. When information is complete, this is harmless as traders can distinguish between aggregate and idiosyncratic shocks. However with incomplete information, the reaction of the aggregate price is different from the complete information benchmark because traders react also to individual shocks when they price individual shares.

To get a quantitative assessment of the excess volatility of Table 4, we follow the same simulation steps as in Section 3.3 and invoke the test of West (1988b). Recall our simulation contained a high noise calibration $\{\sigma_i^2 = 0.8,\ \sigma_\varepsilon^2 = 0.05\}$ and a low noise calibration $\{\sigma_i^2 = 0.2,\ \sigma_\varepsilon^2 = 0.05\}$, 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

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

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

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 (18), 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  CONCLUDING THOUGHTS

The influence of Phelps (1969) and Lucas (1975) has had a resurgence in the macroeconomics literature. The excellent handbook article by Angeletos and Lian (2016b) 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. The conventional view of Samuelson's Dictum (if one exists) is nicely captured by Campbell (2017) (p. 144), which we paraphrase here: Cash-flow news drives stock-level return variation and can be completely diversified away in the aggregate, allowing discount-rate news to determine the lion's share of aggregate stock returns. This result relies on being able to perfectly disentangle idiosyncratic and aggregate shocks, which we view as a strong assumption. Our results suggest that a combination of incomplete information and aggregation offer an alternative interpretation that is not mutually exclusive. Indeed, several recent papers document how incomplete information can help reconcile theory and macroeconomic data [e.g., Lorenzoni (2009), Angeletos and Lian (2016b), Wu et al. (2020)] and we have shown a similar explanation can be applied to asset pricing data.

# References

Angeletos, George-Marios and Chen Lian, "Dampening General Equilibrium: From Micro Elasticities to Macro Effects," 2016. Working Paper, MIT.

__ and __ , "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," *Handbook of Macroeconomics*, 2016, *2*, 1065–1240.

__ and Zhen Huo, "Myopia and Anchoring," 2018. Working Paper, MIT.

Blanchard, Olivier J., Jean-Paul L'Huillier, and Guido Lorenzoni, "News, Noise, and Fluctuations: An Empirical Exploration," 2013, *103* (7), 3045–3070. Forthcoming in *American Economic Review*.

Blundell, Richard and Thomas M. Stoker, "Heterogeneity and Aggregation," *Journal of Economic Literature*, 2005, *XLII*, 347–391.

Campbell, John Y., "A Variance Decomposition for Stock Returns," *The Economic Journal*, 1991, *101* (405), 157–179.

__ , *Financial Decisions and Markets: A Course in Asset Pricing*, Princeton University Press, 2017.

__ and Robert J. Shiller, "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 1987, *95* (5), 1062–1088.

__ and __ , "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1988, *1*, 195–228.

__ and __ , "Stock Price, Earnings, and Expected Dividends," *Journal of Finance*, 1988, *43*, 661–676.

__ and __ , "Valuation Ratios and the Long-Run Stock Market Outlook," *The Journal of Portfolio Management*, 1998, *24* (2), 11–26.

__ and __ , "Valuation Ratios and the Long-Run Stock Market Outlook: An Update," 2001. NBER Working Paper No. 8221.

__ , Andrew W. Lo, and A. Craig MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press, 1997.

Cochrane, John H. and Lars Peter Hansen, "Asset Pricing Explorations for Macroeconomics," *NBER Macroeconomics Annual*, 1992, *7*, 115–165.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, "The Value Spread," *The Journal of Finance*, 2003, *58* (2), 609–641.

__ , __ , and __ , "The Price is (Almost) Right," *The Journal of Finance*, 2009, *LXIV*, 2739–2782.

Deaton, Angus, *Understanding Consumption*, Oxford, U.K.: Oxford University Press, 1992.

Flavin, M., "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence," *Journal of Political Economy*, 1983, *91*, 929–956.

Forni, Mario and Marco Lippi, "Aggregation of Linear Dynamic Microeconomic Models," *Journal of Mathematical Economics*, 1999, *31* (1), 131–158.

Friedman, Milton, "The Permanent Income Hypothesis," in "A Theory of the Consumption Function," Princeton University Press, 1957, pp. 20–37.

Glasserman, Paul and Harry Mamaysky, "Investor Information Choice with Macro and Micro Information," Technical Report, Columbia Business School Research Paper No. 16-45 2019.

Goodfriend, Marvin, "Information-Aggregation Bias," *American Economic Review*, 1992, *82* (3), 508–519.

Granger, Clive W. J., "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 1980, *14* (2), 227–238.

Hall, Robert E., "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 1978, *86* (6), 971–988.

Hannan, E. J. and B. G. Quinn, "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society. Series B.*, 1979, *41* (2), 190–195.

Hansen, Lars P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (4), 1029–1054.

_ and Thomas J. Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 1980, *2*, 7–46.

Jung, Jeeman and Robert J. Shiller, "Samuelson's Dictum and the Stock Market," *Economic Inquiry*, 2005, *43 (2)*, 221–228.

Kasa, Kenneth, Todd B. Walker, and Charles H. Whiteman, "Heterogeneous Beliefs and Tests of Present Value Models," *Review of Economic Studies*, 2014, *81* (3), 1137–1163.

Kleidon, Allan W., "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, 1986, *94*, 953–1001.

LeRoy, Stephen and Richard D. Porter, "The Present Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, 1981, *64*, 555–574.

Lorenzoni, Guido, "A Theory of Demand Shocks," *American Economic Review*, 2009, *99* (5), 2050–84.

Lucas, Jr. Robert E., "An Equilibrium Model of the Business Cycle," *Journal of Political Economy*, 1975, *83*, 1113–1144.

Luo, Yulei, "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics*, 2008, *11* (2), 366–385.

Mackowiak, Bartosz and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 2009, *99* (3), 769–803.

Marsh, T. and R. Merton, "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, 1986, *76*, 483–498.

Meghir, Costas and Luigi Pistaferri, "Earnings, Consumption and Life Cycle Choices," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4 2011, pp. 773 – 854.

Newey, W.K. and K.D. West, "A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 1987, *55*, 703–708.

Phelps, Edmund S., "The New Microeconomics in Inflation and Employment Theory," *American Economic Review*, 1969, *59* (2), 147–160.

—, "The Trouble with 'Rational Expectations' and the Problem of Inflation Stabilization," in R. Frydman and E.S. Phelps, eds., *Individual Forecasting and Aggregate Outcomes*, Cambridge: Cambridge University Press, 1983, pp. 31–45.

Pischke, Jorn-Steffen, "Individual Income, Incomplete Information, and Aggregate Consumption," *Econometrica*, 1995, *63* (4), 805–840.

Reis, Ricardo, "Inattentive consumers," *Journal of Monetary Economics*, 2006, *53* (8), 1761 – 1800.

Rondina, Giacomo and Todd B. Walker, "Confounding Dynamics," *Journal of Economic Theory*, 2021, *196* (September).

Sargent, Thomas J., *Macroeconomic Theory*, second ed., London, U.K.: Academic Press, 1987.

Shiller, Robert J., "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 1981, *71* (3), 421–436.

—, "From Efficient Market Theory to Behavioral Finance," Cowles Foundation Discussion Papers 1385, Cowles Foundation for Research in Economics, Yale University October 2002.

Startz, Richard and Kwok Ping Tsang, "On the Present Value Model in a Cross Section of Stocks," 2014. Manuscript.

Thornton, Michael A., "The Aggregation of Dynamic Relationships Caused by Incomplete Information," *Journal of Econometrics*, 2014, *178*, 342–351.

Vuolteenaho, Tuomo, "What Drives Firm-Level Stock Returns?," *The Journal of Finance*, 2002, *57* (1), 233–264.

West, Kenneth D., "Asymptotic Normality, When Regressors Have a Unit Root," *Econometrica*, 1988, *56*, 1397–1417.

＿ , "Dividend Innovations and Stock Price Volatility," *Econometrica*, 1988, *56*, 37–61.

Whiteman, Charles, *Linear Rational Expectations Models: A User's Guide*, Minneapolis: University of Minnesota Press, 1983.

Wu, Jieran, Jianjun Mao, and Eric R. Young, "Macro-Financial Volatility under Dispersed Information," 2020. forthcoming in *Theoretical Economics*.

# 5 APPENDIX A: DERIVATIONS AND PROOFS

## 5.1 DERIVING (3) FROM (2)    The present-value model of stock prices is given by

$$P_t = \mathbb{E}_t \sum_{j=1}^{\infty} (1+r)^{-j} D_{t+j} \tag{31}$$

Multiple by $r$ and add $D_t$ to both sides of (31).

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

Dividing by $P_t$ and re-arranging yields

$$\begin{aligned}
\frac{D_t}{P_t} &= r + \frac{D_t}{P_t} - \frac{r}{P_t} \mathbb{E}_t \sum_{j=1}^{\infty} (1+r)^{-j} D_{t+j} \\
&= r + \frac{D_t}{P_t} - \frac{r}{P_t} \mathbb{E}_t \Big( \frac{D_{t+1}}{1+r} + \frac{D_{t+2}}{(1+r)^2} + \frac{D_{t+3}}{(1+r)^3} + \cdots \Big) \\
&= r + \frac{1}{P_t} \mathbb{E}_t \Big( D_t - \frac{r D_{t+1}}{1+r} - \frac{r D_{t+2}}{(1+r)^2} - \frac{r D_{t+3}}{(1+r)^3} + \cdots \Big)
\end{aligned}$$

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

$$\begin{aligned}
\frac{D_t}{P_t} &= r - \frac{1}{P_t} \mathbb{E}_t \Big( -D_t + D_{t+1}\Big(1 - \frac{1}{1+r}\Big) + D_{t+2}\Big(\frac{1}{1+r} - \frac{1}{(1+r)^2}\Big) \\
&\quad + D_{t+3}\Big(\frac{1}{(1+r)^2} - \frac{1}{(1+r)^3}\Big) + \cdots \Big)
\end{aligned}$$

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

$$\begin{aligned}
\frac{D_t}{P_t} &= r - \frac{1}{P_t} \mathbb{E}_t \Big( \Delta D_{t+1} + \frac{\Delta D_{t+2}}{1+r} + \frac{\Delta D_{t+3}}{(1+r)^2} + \cdots \Big) \\
&= r - \mathbb{E}_t g_t^D \tag{32}
\end{aligned}$$

## 5.2 DERIVING THE HANSEN-SARGENT FORMULA    The present value model of stock prices is given by equations (9)–(10), which we replicate here for convenience,

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

$$D_t = A(L) v_t \tag{11}$$

For this derivation, we assume traders have complete information and observe the sequence of shocks $\mathcal{F}_t = \{v_t, v_{t-1}, v_{t-2}, ...\}$. This suggests an equilibrium price that is a linear function of $\mathcal{F}_t$; namely, $P_t = P(L) v_t$.

Expectations are taken using the Wiener-Kolmogorov optimal prediction formula,

$$\mathbb{E}(P_{t+1}|\mathscr{F}_t) = L^{-1}[P(L) - P_0]\varepsilon_t$$

$$\mathbb{E}(D_{t+1}|\mathscr{F}_t) = L^{-1}[D(L) - D_0]\varepsilon_t$$

Substituting these into the equilibrium (9) and using the $z$-transform

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

$$zP(z) = \theta P(z) + \theta[D(z) - D_0] - \theta P_0$$

$$(z - \theta)P(z) = \theta[D(z) - D_0] - \theta P_0 \tag{33}$$

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

$$P(z) = \frac{\theta D(z) - \theta D(\theta)}{z - \theta} \tag{34}$$

See Whiteman (1983) for further exposition.

As a specific example, suppose $D(z) = 1/(1 - \rho z)$, then

$$P(z) = \theta \left( \frac{1}{1 - \rho z} - \frac{1}{1 - \rho \theta} \right) \Big/ (z - \theta)$$

$$= \theta \left( \frac{1 - \rho \theta - (1 - \rho z)}{(1 - \rho \theta)(1 - \rho z)} \right) \Big/ (z - \theta)$$

$$= \frac{\theta \rho}{(1 - \rho \theta)(1 - \rho z)}$$

## 5.3 PROOF OF PROPOSITION 1    We now prove Proposition 1, which we repeat for convenience.

**Proposition 2.** *Let Trader $i$'s information set be current and past dividends $\mathscr{I}_{it} = \{D_{it}, D_{it-1}, ....\}$ with the dividend process given by* (20), *then the equilibrium price is given by*

$$P_{it} = \left( \frac{\theta(\rho - \lambda)}{(1 - \theta \rho)(1 - \lambda L)} \right) a_t + \left( \frac{\theta(\rho - \lambda)}{(1 - \lambda L)} \right) \eta_{it}$$

where

$$\lambda = \frac{1}{2} \left[ \left( \frac{\sigma_{\varepsilon_1}^2}{\sigma_\eta^2 \rho} \right) + \left( \frac{1}{\rho} + \rho \right) - \left\{ \left[ \frac{\sigma_{\varepsilon_1}^2}{\sigma_\eta^2 \rho} + \left( \frac{1}{\rho} + \rho \right) \right]^2 - 4 \right\}^{1/2} \right]$$

*Proof.* The dividend process is a combination of an aggregate shock and idiosyncratic shock,

$$D_{it} = a_t + \eta_{it} \tag{23}$$

$$a_t = \rho a_{t-1} + \varepsilon_t \tag{24}$$

and trader $i$'s information set is given by $\mathscr{I}_{it} = \{D_{it}, D_{it-1}, ...\}$. Sargent (1987) shows the fundamental Wold representation for (20)–(21) is given by

$$D_{it} = \left( \frac{1 - \lambda L}{1 - \rho L} \right) \xi_{it} \tag{35}$$

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

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

$$\sigma_\xi^2 = \frac{\sigma_\varepsilon^2 + \sigma_\eta^2 (1 - \rho)^2}{(1 - \lambda)^2}$$

We can use the formulas of the previous subsection, Section 5.2, to derive the equilibrium price by substituting (35) into (34), which yields

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

$$P_{it} = \left( \frac{\theta(\rho - \lambda)}{(1 - \rho L)(1 - \rho \theta)} \right) \xi_t$$

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

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

which is consistent with (24).

### 5.4 PROOF OF PROPOSITION 2

**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 $D_t = (1/I) \sum_i D_{it} | \mathscr{I}_{it}$. Let $P_{it} | \mathscr{F}_t$ and $P_{it} | \mathscr{I}_{it}$ be the complete and incomplete information rational expectations equilibrium, respectively; then,*

$$\text{Cov}\,(P_{it}, g_{it} | \mathscr{F}_t) = \text{Var}\,(P_{it} | \mathscr{F}_t)$$

$$\text{Cov}\,(P_{it}, g_{it} | \mathscr{I}_{it}) = \text{Var}\,(P_{it} | \mathscr{I}_{it})$$

$$\text{Cov}\,(P_t, g_t | \mathscr{F}_t) = \text{Var}\,(P_t | \mathscr{F}_t)$$

$$\text{Cov}\,(P_t, g_t | \mathscr{I}_{it}) = \text{Var}\,(P_t | \mathscr{I}_{it}) \left( \frac{\rho - \lambda}{\rho(1 - \lambda \rho)} \right) \tag{30}$$

The first three equalities are trivial to prove as $g_{it}$ and $g_t$ are accurate reflections of the equilibrium price. That is, $P_{it} = E(g_{it}|\mathscr{F}_t)$, $P_{it} = E(g_{it}|\mathscr{I}_{it})$, and $P_{it} = E(g_{it}|\mathscr{F}_t)$. To show the last equality, note

$$P_t|\mathscr{I}_{it} = \left( \frac{\theta(\rho - \lambda)}{(1 - \rho\theta)(1 - \rho L)(1 - \lambda L)} \right) \varepsilon_t$$

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

The variance of $P_t|\mathscr{I}_{it}$ is given by

$$\mathrm{Var}(P_t|\mathscr{I}_{it}) = \frac{\theta^2 (\rho - \lambda)^2 (1 + \lambda\rho)\sigma_\varepsilon^2}{(1 - \theta\rho)^2 (1 - \lambda\rho)(1 - \lambda^2)(1 - \rho^2)}$$

and the covariance term

$$\mathrm{Cov}\,(P_t, g_t|\mathscr{I}_{it}) = \frac{\theta^2 (\rho - \lambda)\sigma_\varepsilon^2}{(1 - \rho\theta)^2 (1 - \rho^2)(1 - \lambda\rho)}$$

A few algebraic manipulations of the above equations delivers (27).