NBER WORKING PAPER SERIES

# CRIMES AGAINST CAMPBELL-SHILLER

Itzhak Ben-David
Alex Chinco

Working Paper 35189
http://www.nber.org/papers/w35189

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2026

This paper has benefited from careful feedback by John Campbell on earlier drafts. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Itzhak Ben-David and Alex Chinco. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Crimes Against Campbell-Shiller
Itzhak Ben-David and Alex Chinco
NBER Working Paper No. 35189
May 2026
JEL No. G12, G14, G35, G40, G41

**ABSTRACT**

The Campbell and Shiller (1988) log-linear approximation is widely viewed as a model-free accounting identity that always holds: in sample, in expectation, and under arbitrary subjective beliefs. None of these claims is true. The formula is far from automatic even in realized data. Many companies do not pay dividends, making the calculation ill-defined. For dividend payers, the results are not always what they seem. The formula registers buybacks and new issuance as phantom cash-flow shocks. Taking expectations comes with its own complications. Researchers see the forward-looking version of Campbell-Shiller as a dynamic Gordon model, but this interpretation requires investors to consistently think in present-value terms and to know the cap rate with basis-point precision. Both are strong assumptions that do not always hold in the data. Finally, the formula generically fails under arbitrary subjective beliefs. The exceptions represent knife-edge cases where forecast errors obey a precise adding-up condition. Insisting that Campbell-Shiller always holds makes it harder to learn about the true pricing model in the many cases where it does not.

Itzhak Ben-David
The Ohio State University
and NBER
ben-david.1@osu.edu

Alex Chinco
Michigan State University
AlexChinco@gmail.com

# Introduction

Campbell and Shiller (1988) is one of the most influential articles in empirical asset pricing. The paper approximates a stock's current price-to-dividend (PD) ratio by log-linearizing the definition of realized returns

$$\log \mathrm{PD}_t \;=\; \text{constant} - \sum_{h=1}^{\infty} \rho^{h-1} \cdot \left\{ \mathbb{E}_t[\mathrm{R}_{t+h}] - \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] \right\} \tag{1}$$

$\mathbb{E}_t[\mathrm{R}_{t+h}]$ is the stock's expected return $h$ years from now, and $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}]$ is its expected dividend growth. $\rho = 1/(1 + e^{-\overline{\log \mathrm{PD}}})$ stems from centering the approximation around the long-run average log PD ratio, $\overline{\log \mathrm{PD}}$.

The original analysis is sharp, self-aware, and careful about assumptions. This paper is not a critique of Campbell-Shiller. It is an inventory of the many qualifications needed to apply the formula. The derivation of Campbell-Shiller starts with the definition of realized returns. Since it is always possible to calculate a return, researchers have come to believe that Campbell-Shiller inherits the same sort of universality. It does not. We focus on three misconceptions.

First, researchers believe it is straightforward to apply Campbell-Shiller in realized data. Bernanke and Kuttner (2005) describes the formula as "nothing more than a dynamic accounting identity" that "contains no real economic content." Greenwood and Shleifer (2014) emphasizes the same talking points, calling Campbell-Shiller "an accounting identity from the viewpoint of the econometrician." Gourinchas and Rey (2007) writes that, if you remove the two expectation operators, Equation (1) "must hold along every sample path."

This is not true. To start with, 20% of S&P 500 companies do not pay dividends. Equation (1) is undefined for these companies. Researchers could apply Campbell-Shiller at the firm level using total corporate payouts. Instead, they typically look at per-share dividends. This seemingly innocuous change introduces two sources of confounding variation: composition (investors receive payouts via buybacks rather than dividends) and dilution (changes in the share count from repurchases and issuance). Both register as phantom shocks in Campbell-Shiller. Neither has anything to do with fundamentals.

Consider a company that increases buybacks while maintaining the same total payout. Campbell-Shiller sees this change as a negative cash-flow shock because the firm's dividend has dropped. Yet, shareholders are receiving the same cash flow, just in a slightly different form. New issuance mechanically dilutes per-share dividends. Suppose a company issues equity to fund a positive-NPV project. If the board immediately raises the firm's total dividend to reflect the project's earnings boost, then the increase offsets the dilution, so Equation (1) sees a positive cash-flow shock. But, if the board delays for any reason, the same positive-NPV project will show up as a negative cash-flow shock. What matters is the timing of board approval, not the present value.

Second, researchers believe that Campbell-Shiller must hold in expectation. Cochrane (2011) writes that "movements in the price-dividend ratio must reflect changes in expected future dividends or changes in discount rates." Lettau and Ludvigson (2001) argues that when the dividend yield is high, "agents must be expecting either high returns on assets in the future or low dividend growth rates." Cohen, Polk, and Vuolteenaho (2003) insists that all differences in the current multiple "must necessarily be accounted for by cross-sectional variation in expected long-horizon returns and/or growth."

This is not true. The move from realized data to expected outcomes is a massive leap. From 1963 through 1982, Exxon Mobil paid a regular dividend, did not repurchase shares, and maintained a stable share count. It would be reasonable to think the in-sample version of Campbell-Shiller might hold for Exxon Mobil during this period. But this does not imply that Exxon Mobil's $29.75/sh price on December 31st 1982 was determined by the forward-looking version of Campbell-Shiller. These are two very different claims.

The expectation operators in Equation (1) are doing real work. They completely change the economic interpretation of the formula. The in-sample version of Campbell-Shiller summarizes observed prices. The forward-looking version is a statement about how investors determine prices. Researchers often use "ex post versus ex ante" language that makes the difference sound like a technicality. It is not. One version of Campbell-Shiller describes a price path. The other makes a claim about the mechanism that produced it.

Researchers call Campbell-Shiller an “accounting identity”. Take that at face value. The present-value framework counts expected future dividends. Each one, discounted back to today, is an entry in the ledger. The price is the sum of those entries. For the relationship to be an identity, every entry must be accounted for, starting with the largest: next year’s expected dividend.

It is hard to see how real-world investors could know the forecasts on the right-hand side of Equation (1) with enough precision to close these books. To make things easy on investors, let’s assume expected returns and dividend growth do not vary. The present value of a stock’s future dividend stream simplifies to

$$\mathrm{Price}_t \;=\; \mathbb{E}_t[\mathrm{Div}_{t+1}] \times \left(\frac{1}{\mathrm{R} - \mathrm{G}}\right) \tag{2}$$

This is the Gordon model—i.e., the thing Campbell-Shiller is approximating.

Even in this simple setting, we show that an investor would have to know $(\mathrm{R}-\mathrm{G})$ to an accuracy of $\pm\mathrm{DY}^2$. Otherwise, noise in the cap rate would have a bigger effect on the price than next year’s expected dividend, the single largest entry in the ledger. For the S&P 500, with $\overline{\mathrm{DY}} \approx 2\%$, the required tolerance is $\overline{\mathrm{DY}}^2 \approx 0.04$%pt. No investor can hit this precision. Standard estimates of the equity risk premium span 4% to 8%, a range 100× wider. You cannot call something an “accounting identity” when you cannot count the biggest term.

Third, researchers believe that Campbell-Shiller holds under arbitrary subjective beliefs. Conventional wisdom says that the two expectation operators in Equation (1) do not have to be rational. For example, De la O and Myers (2024) writes that, “because the relationship comes from an accounting identity, it holds even if subjective expectations are not rational.”

This is not true. Here is a counterexample. For the sake of argument, imagine that Campbell-Shiller holds in-sample and under correct expectations. Now suppose that investors unilaterally increase their forecast for the S&P 500’s return over the next twelve months. Investors add 1%pt to the objectively correct forecast. The change in next year’s return forecast has no impact on realized data, so the left-hand side of Equation (1) is the same. But the right-hand side is now 1%pt lower. Equality has been violated.

If it is acceptable for investors to hold objectively correct beliefs, then it must also be permissible to hold objectively correct views about everything except for next year's return. You might think this is a silly mistake to make, and we can have that debate. But it will be a debate about what researchers think, not about what the math permits. When researchers use Campbell-Shiller to study subjective expectations, they are only considering a small subset of all possible subjective beliefs. We characterize this subset.

Suppose that Campbell-Shiller holds under the correct expectation operator, $\mathbb{E}_t[\cdot]$. For Equation (1) to remain valid under an alternative subjective forecasting rule, $\widetilde{\mathbb{F}}_t[\cdot]$, the resulting errors must satisfy the following condition

$$\underbrace{\sum_{h=1}^{\infty} \frac{(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \mathrm{Div}_{t+h}]}{(1+\overline{\mathrm{DY}})^{h-1}}}_{\substack{\text{Present value of errors about future} \\ \text{dividend growth discounted at } \overline{\mathrm{DY}}}} = \underbrace{\sum_{h=1}^{\infty} \frac{(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\mathrm{R}_{t+h}]}{(1+\overline{\mathrm{DY}})^{h-1}}}_{\substack{\text{Present value of return} \\ \text{errors discounted at } \overline{\mathrm{DY}}}} \tag{3}$$

For Campbell-Shiller to hold, the present value of mistakes about future dividend growth and mistakes about future returns must cancel out when discounted at $\overline{\mathrm{DY}}$. The time-cost of money is R, but mistakes need to get discounted at $\overline{\mathrm{DY}}$.

The adding-up condition clearly rules out unilateral changes. If investors increase their next-twelve-month return forecast for the S&P 500 by 1%pt, then Campbell-Shiller requires some offsetting mistake. Investors might overestimate dividend growth. Or they could underestimate returns at some longer horizon. Either way, the magnitude of this mistake needs to be carefully calibrated. A 1%pt error about expected dividend growth next year would net out one for one. But to accomplish the same thing with a mistake about the S&P 500's dividend growth three years out, the error would have to be $(1+0.02)^2 \cdot 1\%\text{pt} \approx 1.04\%\text{pt}$.

Consistency with Campbell-Shiller is not a property of beliefs. It is a property of beliefs together with the way that they are modeled. Notice that the unilateral 1%pt error in next year's S&P 500 return forecast can be brought back in line with Campbell-Shiller by adding additional assumptions. Investors could overestimate the S&P 500's dividend growth next year by 1%pt or underestimate its returns in three years by 1.04%pt.

We work through numerous examples from the literature. For instance, investors with extrapolative beliefs overweight recent price changes when forecasting future returns, $(\widetilde{\mathbb{F}} - \mathbb{E})[R] \neq 0$. On its own, this univariate bias is inconsistent with Campbell-Shiller. But it is possible to restore consistency. Barberis, Greenwood, Jin, and Shleifer (2015) embeds this belief specification in an equilibrium model with two investor types and a price determined by market-clearing. When modeled this way, extrapolative beliefs do satisfy the adding-up condition in Equation (3).

The modeling overhead needed to bring subjective beliefs in line with Campbell-Shiller can be considerable. Jin and Sui (2022) provides a case in point. The paper's title is "Asset Pricing with Return Extrapolation," but the description of return extrapolation takes up less than a page. The remaining 22 pages build the apparatus required to maintain consistency with Campbell-Shiller: a Lucas-tree economy, Epstein-Zin preferences, two coupled differential equations for the equilibrium price path, 11 calibrated parameters, etc. The authors' purpose is to study return extrapolation. The Campbell-Shiller framework forces them to install 22 pages of additional machinery.

This has consequences for the belief-based asset-pricing literature. In the past, researchers have treated Campbell-Shiller as a model-free way to learn about how subjective beliefs affect asset prices. Instead, they have been getting a signal that was roughly $\frac{22}{23} \approx 96\%$ model overhead and only $\frac{1}{23} \approx 4\%$ subjective beliefs. These are page counts from Jin and Sui (2022), but the fact generalizes. The infrastructure needed to satisfy Equation (3) shapes every conclusion.

Our point is not that Campbell-Shiller never holds. Our point is that the formula does not have to hold. It is not an "identity." The equation "contains real economic content." When a researcher uses Equation (1), they are making a strong claim. Multiples can move for reasons unrelated to "changes in expected future dividends or changes in discount rates." Campbell-Shiller is not automatically satisfied "even if subjective expectations are not rational."

All these misconceptions stem from the same root cause. Equation (1) generalizes the Gordon model to allow for time-varying parameters. This is a useful result, but only in situations where prices come from a time-varying Gordon model. If investors are not applying present-value logic, then allowing R and G to change over time does not help. If it is too hard to estimate this pair of parameters when constant, then what hope is there for the dynamic extension? It does not make sense to approximate a special case of the dividend discount model in a world where the majority of corporate payouts are buybacks.

Calling Campbell-Shiller an "identity" is not innocuous. The label tells researchers the formula is a logical truth, so they stop asking what is actually pricing stocks. There are specific situations where it does make sense to apply Campbell-Shiller. However, when researchers insist that Campbell-Shiller always holds, they make it harder to learn about the true pricing model in the many situations where it does not. Variance decompositions provide a clear illustration of the problem.

For example, Cochrane (2008) wants to understand why the S&P 500's log PD ratio varies over time. The paper answers this question by leaning on the "fact" that all such variation must come from changes in future returns or future dividend growth. This is the common move in the literature. No one directly tests Campbell-Shiller. Researchers assume the formula holds and use that restriction to estimate something else of interest.

Here is how the ideal variance decomposition would work in this scenario. First, you would regress $\sum_{h=1}^{\infty} \rho^{h-1} \cdot \Delta \log \mathrm{Div}_{t+h}$ on $\log \mathrm{PD}_t$ to get the cash-flow variance share, $\beta_{\mathrm{Div}}(\infty)$. Then, you would regress $-\sum_{h=1}^{\infty} \rho^{h-1} \cdot \mathrm{R}_{t+h}$ on $\log \mathrm{PD}_t$ to get the discount-rate share, $\beta_{\mathrm{R}}(\infty)$. If Campbell-Shiller holds, then it stands to reason that these two shares would sum to 100%.

Of course, it is not possible to run either regression. They both require summing an infinite series at each date $t$. Here is the calculation Cochrane does instead. First, he gets a short-term slope of $\beta_{\mathrm{Div}}(1) \approx 0.8\%$ by regressing next year's realized dividend growth on $\log \mathrm{PD}_t$. Then, he assumes that $\log \mathrm{PD}_t$ follows an AR(1) process, which allows him to calculate

$$\beta_{\mathrm{Div}}(\infty) \;=\; \overset{0.8\%}{\beta_{\mathrm{Div}}(1)} \times \overbrace{\left(\frac{1}{1 - \underset{0.98}{\rho} \cdot \underset{0.92}{\phi}}\right)}^{\sim 10\times} \;\approx\; 8\% \tag{4}$$

Cochrane attributes all residual variation in the S&P 500's log PD ratio to discount-rate news, $100\% - \beta_{\text{Div}}(\infty) = \beta_{\text{R}}(\infty) \approx 92\%$. This is not a model-free measurement. It is a model-implied output. The $10\times$ multiplier is nowhere in the data. It comes from assuming that the persistence parameter stays constant forever. That assumption explains why Cochrane combined the S&P 500's $\rho = \left(\frac{1}{1+2\%}\right) \approx 0.98$ and its persistence parameter of $\phi \approx 0.92$ in the way that he did. Strip this assumption and the headline result disappears. The scaling-up step is doing 90% of the work, and the scaling-up step is a modeling choice.

Notice that Equation (4) is Gordon's move. The dividend discount model (DDM) says to price a stock by discounting the company's expected dividend at every future horizon. It is not possible to perform this calculation directly. As a work-around, the Gordon model assumes constant dividend growth, making it possible to calculate the infinite sum by multiplying next year's expected dividend by $\left(\frac{1}{\text{R}-\text{G}}\right)$. Cochrane faced an analogous infinite-sum problem and solved it in a similar way, which explains why his headline result came from multiplying a short-term estimate by $\left(\frac{1}{1-\rho\cdot\phi}\right)$. Both formulas require a key parameter to stay constant forever. Both stem from models.

Campbell and Shiller (1988) set out to relax the Gordon model's assumptions "that dividends will grow at a constant rate forever, and that the discount rate will never change." This paper exists because researchers saw the Gordon model as a model. Variance decompositions use the same constant-parameter assumption that Campbell-Shiller set out to relax. The "identity" label obscures this fact. Today's researchers should be afforded the same freedoms that Campbell and Shiller enjoyed 40 years ago. Doing otherwise makes it harder to discover how assets are actually priced in the many situations where Campbell-Shiller does not hold. It is not always appropriate to use Gordon logic. Often, the reason has nothing to do with time-varying parameters.

**Paper Outline.** Each section builds on the last. Section 1 shows that Campbell-Shiller does not automatically hold in-sample. Section 2 shows that, even in cases where the formula holds in sample, it need not hold in expectation. In Section 3, we assume Campbell-Shiller holds in-sample and under correct expectations. We then show it is generically violated under arbitrary subjective beliefs.

**Related Work.** Campbell and Shiller (1988) derives the log-linear approximation and applies it to aggregate US stock-market data. Campbell (1991) uses the framework to decompose unexpected returns into cash-flow news and discount-rate news, finding that most aggregate return variation comes from discount-rate revisions. Vuolteenaho (2002) finds the opposite at the firm level. Larrain and Yogo (2008) and Boudoukh, Michaely, Richardson, and Roberts (2007) both show that Campbell-Shiller holds at the firm level using total payouts rather than per-share dividends, for companies that have payouts.

A large empirical literature uses Campbell-Shiller to decompose the sources of variation in price-dividend ratios. Cochrane (2011) argues that virtually all variation reflects changing discount rates. Lettau, Ludvigson, and Manoel (2018) revisits the finding with improved econometric methods. We do not take a position in this debate. We argue that the framework within which it takes place rests on assumptions that are routinely violated.

Nagel and Xu (2022) surveys the broader literature on expectations in asset pricing. See Bordalo, Gennaioli, and Shleifer (2018), Greenwood and Shleifer (2014), Barberis et al. (2015), Fuster, Laibson, and Mendel (2010), and Adam, Marcet, and Nicolini (2016); Adam, Marcet, and Beutel (2017). Our paper does not argue against any of these models. We show that using them with Campbell-Shiller requires more infrastructure than is typically acknowledged. Much of the belief-based asset-pricing literature assumes that Campbell-Shiller holds under arbitrary subjective expectations. We derive the conditions under which Campbell-Shiller holds for biased beliefs. Meeting these conditions requires substantial modeling overhead.

Other papers have investigated whether Campbell-Shiller's approximation is accurate. Engsted, Pedersen, and Tanggaard (2012) puts an upper bound on the average error. Cho, Kremens, Lee, and Polk (2024) proposes an alternative present-value formula with smaller approximation errors. Chen and Zhao (2009) argues that VAR-based return decompositions are sensitive to the choice of state variables. Our question is logically prior: a perfect approximation of the wrong model is not helpful.

# 1 Realized Data

It is widely believed that Campbell-Shiller holds as an "accounting identity" in-sample. Remove the expectation operators from Equation (1), and there is no way the formula can fail. It has to be correct in-sample. That is the conventional wisdom. This section shows why it is wrong.

Even if you remove the expectation operators, the left-hand side of Equation (1) must be well-defined. Subsection 1.1 notes that roughly 20% of S&P 500 constituents pay no cash dividend, making the formula undefined for these firms. Researchers typically apply Campbell-Shiller to per-share data rather than firm-level quantities, and Subsection 1.2 characterizes the errors introduced by this substitution. Subsections 1.3 and 1.4 provide concrete examples showing how these errors show up in response to repurchases and new issuance.

The bottom line is simple: Campbell-Shiller is not an accounting identity in realized data. There is nothing easy or automatic about applying Equation (1). It requires numerous assumptions even before taking expectations.

## 1.1 Zero Dividends

Campbell-Shiller cannot be applied to stocks with $\text{Div}_t = \$0/\text{sh}$. This number is the denominator in the price-to-dividend (PD) ratio, $\text{PD}_t = \left(\frac{\text{Price}_t}{\text{Div}_t}\right)$. Log dividend growth is also undefined when there is no current or future dividend, $\Delta \log \text{Div}_{t+1} = \log \text{Div}_{t+1} - \log \text{Div}_t$. These problems are not edge cases. Roughly 20% of S&P 500 constituents pay no cash dividend at all. Alphabet, Amazon, Meta, and Berkshire Hathaway all paid no dividend for years or decades. Microsoft went public in 1986 and paid its first dividend in January 2003. In the intervening years, its equity valuation exceeded \$500B while the company returned little cash to shareholders in any form.

On some level, researchers know all this. They use the original dividend-based Campbell-Shiller formula when studying the S&P 500 at the index level. But it is common to replace $\mathbb{E}[\text{Div}]$ with $\mathbb{E}[\text{EPS}]$ when studying individual stocks. Vuolteenaho (2002) showed how to make the jump from dividends to earnings per share. The paper has been cited over 1,000 times. It must be adding something to the original formula to warrant all these citations.

We are not claiming that Campbell-Shiller cannot be modified to handle non-dividend payers. It clearly can. Our point is that the necessity of such modifications tells us something important: there is nothing automatic about applying Campbell-Shiller in-sample. It is not an accounting identity that always holds. Barring further assumptions, it is literally undefined for 20% of stocks in the S&P 500 even before worrying about expectations.

## 1.2 Per-Share Data

"Asset-pricing theory all stems from one simple concept: price equals expected discounted payoff." (Cochrane, 2001, page 1) It should not matter how the payoff is subdivided. Hence, the right realized return to look at is the firm-level calculation

$$1 + \mathrm{R}_{t+1} \quad = \quad \frac{\mathrm{MCap}_{t+1} + \mathrm{TotPmt}_{t+1}}{\mathrm{MCap}_t} \tag{5}$$

$\mathrm{MCap}_t$ is the ex-payout firm value and $\mathrm{TotPmt}_{t+1}$ is the total cash paid to shareholders between $t$ and $(t+1)$, which includes both dividends and repurchases.

When applied to Equation (5), Campbell-Shiller would relate the total-payout multiple $(\frac{\mathrm{MCap}_t}{\mathrm{TotPmt}_t})$ to the present-value-weighted sums of total-payout growth $\Delta \log \mathrm{TotPmt}_{t+h}$ and returns $\mathrm{R}_{t+h}$. The correct growth-rate variable is $\Delta \log \mathrm{TotPmt}$. The left-hand side should have $\log(\frac{\mathrm{MCap}}{\mathrm{TotPmt}})$ as the valuation ratio.

But researchers do not typically use these inputs. Instead, the standard empirical convention replaces this firm-level calculation with the per-share version

$$1 + \mathrm{R}_{t+1} \quad = \quad \frac{\mathrm{Price}_{t+1} + \mathrm{Div}_{t+1}}{\mathrm{Price}_t} \tag{6}$$

$\mathrm{Price}_t$ is the price per share, and $\mathrm{Div}_{t+1}$ is the per-share dividend.

What are the consequences of this substitution? We give the conventional approach the benefit of the doubt. The remainder of this section assumes that $\mathrm{R}_{t+1}$ is still correct. Modigliani-Miller guarantees that per-share returns are unaffected by payout policy when total payout and investment are the same.

Even so, researchers who plug per-share dividend growth into Campbell-Shiller are using a distorted version of total-payout growth.

Assume that a stock paid a dividend this year and last year. Let $\mathrm{TotPmt}_{t+1}$ denote the firm's total payout next year, and let $\mathrm{TotDiv}_{t+1}$ denote the total dividend payout next year. By contrast, $\mathrm{Div}_{t+1} = \frac{\mathrm{TotDiv}_{t+1}}{\#\mathrm{Shares}_t}$ denotes the firm's dividend on a per-share basis. $\#\mathrm{Shares}_t$ represents the share count at the end of period $t$ after repurchases and issuances have occurred.

**Proposition 1.2.** *Suppose that Equations* (5) *and* (6) *yield the same* $\mathrm{R}_{t+1}$*. The gap between per-share dividend growth and total-payout growth decomposes as*

$$\begin{aligned}\Delta \log \mathrm{Div}_{t+1} - \Delta \log \mathrm{TotPmt}_{t+1} \;=\; & \overbrace{\big(\Delta \log \mathrm{TotDiv}_{t+1} - \Delta \log \mathrm{TotPmt}_{t+1}\big)}^{\text{Payout composition}} \\ & - \underbrace{\Delta \log \#\mathrm{Shares}_t}_{\text{Share dilution}}\end{aligned} \tag{7}$$

*The first term is non-zero whenever the dividend share of total payout changes. The second term is non-zero whenever the share count changes.*

The first term is payout composition. It captures the gap between dividend growth and total-payout growth. When a firm shifts cash from dividends to repurchases, $\Delta \log \mathrm{TotDiv}$ falls even though $\Delta \log \mathrm{TotPmt}$ is unchanged. The second term is about changing share counts. Repurchases and new issuance change the denominator of $\mathrm{Div}_{t+1} = \frac{\mathrm{TotDiv}_{t+1}}{\#\mathrm{Shares}_t}$, leading to mechanical dilution and accretion effects even if the numerator remains constant.

By assuming that the switch from total firm-level payouts to per-share dividends has no impact on returns, $\mathrm{R}_{t+1}$, we are giving the conventional wisdom about in-sample Campbell-Shiller the benefit of the doubt. Nevertheless, problems remain. The confounds enter through $\Delta \log \mathrm{Div}_{t+h}$ and through $\mathrm{PD}_t$, which absorbs whatever the growth term does not. The "cash-flow news" and "discount-rate news" that researchers typically extract from the Campbell-Shiller formula are not pure-tone signals. Even under ideal conditions, both signals get blended with payout-policy changes in non-trivial ways.

## 1.3 Repurchases

Researchers typically reason that all variation in the price-dividend ratio must reflect either future dividend growth or future returns. Proposition 1.2 shows that the switch from firm-level total payouts to per-share dividends introduces two additional suspects: payout composition and share dilution. To see how payout composition works in practice, we look at three otherwise identical firms that deliver the same total payout in different ways.

Each firm has $\#\text{Shares}_{t-1} = 100\text{M}$ shares outstanding and a share price of $\text{Price}_{t-1} = \$3.125/\text{sh}$. Each period, every firm earns \$25M and distributes all of it to shareholders. All three firms have the same 8% discount rate, and since there is no investment, the ex-dividend firm value is a flat perpetuity, $\$25\text{M} \times \left(\frac{1}{8\%}\right) = \$312.5\text{M}$. The three firms differ only in how they split the \$25M payout between dividends and repurchases:

1. Baseline Inc. This company always pays \$25M in dividends and repurchases \$0 (100/0 split). With 100M shares at time $t$, the firm has $\text{Div}_t = \$0.25/\text{sh}$ and $\text{Price}_t = \$3.125/\text{sh}$. At time $(t{+}1)$, the numbers are the same: $\text{Div}_{t+1} = \$0.25/\text{sh}$, $\text{Price}_{t+1} = \$3.125/\text{sh}$, $\text{R}_{t+1} = 8\%$, and $\Delta \log \text{Div}_{t+1} = 0\%$.

2. Always Co. This company maintains a steady 80/20 split, paying out \$20M in dividends and repurchasing \$5M every year. At time $t$, the firm has $\text{Div}_t = \frac{\$20\text{M}}{100\text{M}} = \$0.20/\text{sh}$. The \$5M repurchase happens at the ex-dividend price. After dividends, the firm holds \$5M in cash earmarked for repurchase, so the ex-dividend equity value is \$312.5M + \$5M = \$317.5M across 100M shares, giving $\text{Price}_t = \$3.175/\text{sh}$. The firm buys back $\frac{\$5\text{M}}{\$3.175/\text{sh}} \approx 1.575\text{M}$ shares, leaving $\#\text{Shares}_t \approx 98.43\text{M}$.

   At time $(t{+}1)$, Always Co maintains the same 80/20 split on a smaller base, leading to $\text{Div}_{t+1} \approx \$0.2032/\text{sh}$, $\text{Price}_{t+1} \approx \$3.226/\text{sh}$, $\text{R}_{t+1} = 8\%$, and $\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.2032/\text{sh}}{\$0.20/\text{sh}}\right) \approx 1.6\%$. The dividend per share grows by 1.6% not because the firm's fundamentals improved, but because the shrinking share count mechanically boosts the firm's dividend payout on a per-share basis. The PD ratio is 15.9× versus Baseline Inc's 12.5×, a 27% relative difference caused entirely by the lower per-share dividend.

| | Baseline Inc | Always Co | Switch Ltd |
|---|---|---|---|
| Total earnings | \$25.0M | \$25.0M | \$25.0M |
| Total payout | \$25.0M | \$25.0M | \$25.0M |
| Total dividend$_t$ | \$25.0M | \$20.0M | \$25.0M |
| Buyback amount$_t$ | \$0.0M | \$5.0M | \$0.0M |
| Total dividend$_{t+1}$ | \$25.0M | \$20.0M | \$20.0M |
| Buyback amount$_{t+1}$ | \$0.0M | \$5.0M | \$5.0M |
| Payout split$_t$ | 100/0 | 80/20 | 100/0 |
| Payout split$_{t+1}$ | 100/0 | 80/20 | 80/20 |
| #Shares$_t$ | 100.0M | 98.4M | 100.0M |
| #Shares$_{t+1}$ | 100.0M | 96.9M | 98.4M |
| PD$_t$ | 12.5× | 15.9× | 12.5× |
| R$_{t+1}$ | 8.0% | 8.0% | 8.0% |
| $\Delta \log \mathrm{Div}_{t+1}$ | 0.0% | +1.6% | −22.3% |

**Table 1.** Three firms with the same earnings, total payout, and return. Baseline Inc and Switch Ltd are indistinguishable at time $t$. Both firms have the same PD$_t$, Div$_t$, Price$_t$, and R$_t$. Yet one delivers $\Delta \log \mathrm{Div}_{t+1} = 0\%$ while the other has −22.3%. Always Co maintains a stable 80/20 payout split. It shows a mechanical 1.6% growth in dividends per-share due to its shrinking share count.

3. Switch Ltd. This company switches payout policies. At time $t$, Switch Ltd only pays \$25M in dividends (100/0 split). At time $(t{+}1)$, the firm switches to an 80/20 split (\$20M dividends, \$5M repurchases). The firm's numbers at time $t$ match Baseline Inc: $\mathrm{Div}_t = \$0.25/\mathrm{sh}$, and $\mathrm{Price}_t = \$3.125/\mathrm{sh}$. Its new numbers at time $(t{+}1)$ are: $\mathrm{Div}_{t+1} = \$0.20/\mathrm{sh}$, $\mathrm{Price}_{t+1} = \$3.175/\mathrm{sh}$, $\mathrm{R}_{t+1} = 8\%$, and $\Delta \log \mathrm{Div}_{t+1} = \log\left(\frac{\$0.20/\mathrm{sh}}{\$0.25/\mathrm{sh}}\right) = -22.3\%$.

Table 1 summarizes the results. Notice that Baseline Inc and Switch Ltd are indistinguishable at time $t$. They have the same $\mathrm{PD}_t = 12.5\times$, the same $\mathrm{Div}_t = \$0.25/\mathrm{sh}$, the same $\mathrm{Price}_t = \$3.125/\mathrm{sh}$, and the same return $\mathrm{R}_t = 8\%$. A forecasting model estimated on per-share data must produce the same forecast for both firms. Yet one delivers $\Delta \log \mathrm{Div}_{t+1} = 0\%$ and the other has −22.3%.

This −22.3% shock in year $(t{+}1)$ is not a cash-flow shock. Switch Ltd has the same earnings, the same total payout, and the same return as Baseline

Inc. Nothing fundamental has changed. The firm simply redirected \$5M from dividends to buybacks. But Campbell-Shiller has no notion of share count or payout policy. The formula does not distinguish between a 22.3% decline in $\Delta \log \mathrm{Div}$ caused by deteriorating earnings and a 22.3% decline caused by a shift toward buybacks. This is a distinction that makes a difference to investors.

Even the steady-state case is affected. Always Co maintains the same 80/20 payout policy throughout. This firm sees $\Delta \log \mathrm{Div} \approx 1.6\%$ purely from the shrinking share count, and a PD ratio of 15.9× versus Baseline Inc's 12.5×. Campbell-Shiller is internally consistent for Always Co. But the numbers encode payout-policy decisions, not firm fundamentals. To a researcher, Always Co looks like a firm with G = 1.6%. To know otherwise requires data on buybacks, which the Campbell-Shiller framework does not use.

Researchers currently believe that all variation in $\mathrm{PD}_t$ must reflect either future returns or future dividend growth. In this example, none of it does. All three firms earn the same amount, pay out the same amount, and deliver the same 8% return. Total-payout growth is zero every period. The 27% gap in PD ratios between Always Co and Baseline Inc comes from the share-dilution term in Proposition 1.2. The −22.3% per-share dividend shock at Switch Ltd comes from the payout-composition term. A researcher using the Campbell-Shiller formula with per-share dividends would attribute these differences to fundamentals. They are not. Getting from the formula to an empirical implementation requires choices about how to measure payouts. These choices have first-order consequences. Calling the formula an "identity" obscures this fact.

## 1.4 New Issuance

The previous subsection isolated payout composition. This subsection isolates share dilution, the second channel from Proposition 1.2. All payouts are dividends, so the composition term drops out. We now consider two new firms that issue equity to fund the same positive-NPV project. The new earnings are real. So you would think the Campbell-Shiller formula would register a positive cash-flow shock next year. That is not always what happens.

The two firms in this example look identical to Baseline Inc at the start of period $t$. Each starts out earning \$25M per period, distributes all of it as dividends, has an 8% discount rate, and trades at $\text{Price}_{t-1} = \$3.125/\text{sh}$ on $\#\text{Shares}_{t-1} = 100\text{M}$ shares. During period $t$, Now Corp and Delay PLC invest in the same positive-NPV project that earns 12% and generates \$9.375M per year in additional earnings starting in period $(t{+}1)$. Both firms cover the up-front cost by issuing 25M new shares at \$3.125/sh, so that by the end of period $t$ they each have $\#\text{Shares}_t = 125\text{M}$ shares outstanding. The two firms differ only in how quickly they raise their dividend.

1. Now Corp. This company immediately raises its total dividend to \$25M + \$9.375M = \$34.375M. The ex-dividend firm value is $\$34.375\text{M} \times \left(\frac{1}{8\%}\right) = \$429.6875\text{M}$. With $\#\text{Shares}_t = 125\text{M}$, the company has $\text{Div}_{t+1} = \$0.275/\text{sh}$ and $\text{Price}_{t+1} = \frac{\$429.6875\text{M}}{125\text{M}} = \$3.4375/\text{sh}$. The firm's other Campbell-Shiller inputs are $\text{R}_{t+1} = 18.8\%$ and $\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.275/\text{sh}}{\$0.25/\text{sh}}\right) = 9.5\%$.
2. Delay PLC. This firm issues the same number of new shares and invests in the same project as Now Corp. However, Delay PLC keeps its total dividend at \$25M at time $(t{+}1)$ and retains \$9.375M in earnings. The ex-dividend value consists of operating assets (\$429.6875M) plus retained cash (\$9.375M). With $\#\text{Shares}_t = 125\text{M}$, Delay PLC has $\text{Div}_{t+1} = \frac{\$25\text{M}}{125\text{M}} = \$0.20/\text{sh}$ and $\text{Price}_{t+1} = \frac{\$429.6875\text{M}+\$9.375\text{M}}{125\text{M}} = \$3.5125/\text{sh}$. The cum-dividend value per share is \$3.7125/sh. The other Campbell-Shiller inputs are $\text{R}_{t+1} = 18.8\%$ and $\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.20/\text{sh}}{\$0.25/\text{sh}}\right) = -22.3\%$.

Table 2 summarizes the results. Now Corp and Delay PLC did the same thing: they issued the same number of shares, invested in the same project, earned the same total earnings, and delivered the same return to existing shareholders. The cum-dividend value per share is \$3.7125/sh in both cases. The only difference is a board-level decision about when to raise the dividend. Yet the Campbell-Shiller formula registers Now Corp as a 9.5% cash-flow shock (positive) and Delay PLC as a −22.3% cash-flow shock (negative). That is a 32%pt swing from dividend timing alone. This is not what researchers have in mind when they say "cash-flow shock."

| | Baseline Inc | Now Corp | Delay PLC |
|---|---|---|---|
| New shares issued | 0.0M | 25.0M | 25.0M |
| $\#\text{Shares}_{t-1}$ | 100.0M | 100.0M | 100.0M |
| $\#\text{Shares}_{t}$ | 100.0M | 125.0M | 125.0M |
| Project earnings | | \$9.4M | \$9.4M |
| Total earnings | \$25.0M | \$34.4M | \$34.4M |
| Total payout | \$25.0M | \$34.4M | \$25.0M |
| Retained cash | \$0.0M | \$0.0M | \$9.4M |
| $\text{Div}_{t+1}$ | \$0.25/sh | \$0.27/sh | \$0.20/sh |
| $\text{Price}_{t+1}$ | \$3.13/sh | \$3.44/sh | \$3.51/sh |
| $\text{Div}_{t+1} + \text{Price}_{t+1}$ | \$3.38/sh | \$3.71/sh | \$3.71/sh |
| $\text{PD}_{t}$ | 12.5× | 12.5× | 12.5× |
| $\text{R}_{t+1}$ | 8.0% | 18.8% | 18.8% |
| $\Delta \log \text{Div}_{t+1}$ | 0.0% | +9.5% | −22.3% |

**Table 2.** Now Corp and Delay PLC issue the same shares, invest in the same project, earn the same total earnings, and deliver the same return to shareholders. The only difference is a board-level decision about when to raise the dividend. Yet the Campbell-Shiller formula registers one as a 9.5% positive cash-flow shock and the other as a −22.3% negative cash-flow shock. That is a 32%pt swing from dividend timing alone.

Firms typically adjust their dividend slowly like Delay PLC (Lintner, 1956). The market prices the same project the same way regardless of the timing. Now Corp and Delay PLC both have cum-dividend prices of \$3.7125/sh. Yet, in Campbell-Shiller, one company invests in a positive-NPV project and has a positive cash-flow shock. The other company makes the same investment and records a negative shock. The only difference is the timing.

Researchers currently believe that all variation in $\text{PD}_t$ must reflect either changes in future returns or changes in future dividend growth. This is not true when looking at per-share data. We just saw two different ways that payout-policy decisions can lead to phantom dividend-growth shocks. Getting from the textbook formula to an empirical implementation requires choices about how to measure payouts, when to recognize them, and how to handle share-count changes. Each choice has first-order consequences. Campbell-Shiller cannot see the difference. It is misleading to call that blindness an “identity”.

# 2 Correct Expectations

The previous section gave a number of reasons why Campbell-Shiller can fail to hold in-sample. The formula is clearly not an identity that always has to hold. There are, however, empirical settings in which the conditions needed to apply the formula in-sample are satisfied. Suppose we restrict attention to one of these cases. Researchers currently believe that, whenever Campbell-Shiller holds in sample, it must also hold in expectation. This is not true.

Taking expectations changes the economics. The in-sample version of Equation (1) describes observed prices. The forward-looking version says why those prices were observed. Moreover, it is not just any claim. Campbell-Shiller extends the Gordon model to allow for time-varying parameters. This interpretation requires the other assumptions behind the Gordon model to remain valid. Subsection 2.1 catalogs these assumptions. The first two require investors to consistently apply present-value logic. Subsection 2.2 shows that many do not. Subsection 2.3 shows that, even if investors did think this way, calling Campbell-Shiller an "accounting identity" requires an implausible level of precision.

## 2.1 Gordon Model

Campbell-Shiller extends the Gordon model to allow for time-varying coefficients. Equation (1) is a log-linear approximation to Equation (2). Researchers understand this. Campbell and Shiller (1988) describes their own model as "a dynamic version of the Gordon (1962) model." Gao and Martin (2021) writes that "Campbell and Shiller (1988) generalizes the Gordon growth model to the empirically relevant case in which these quantities are time-varying."

What's more, the Gordon model (Equation 2) is a special case of the Dividend Discount Model (DDM). It offers a simpler implementation in a world where the expected dividend-growth rate is constant. Campbell-Shiller relaxes this constant-parameter assumption. But to interpret the formula as a statement about how prices are set, the remaining steps in the derivation of the DDM must still work. We now review what those steps are.

Step 1. Assume current price obeys one-period-ahead present-value logic.

Campbell-Shiller may start with the definition of a realized return, but the model it approximates starts with the one-period-ahead present-value formula

$$\text{Price}_t = \frac{\mathbb{E}_t[\text{Div}_{t+1}]}{1+\text{R}} + \frac{\mathbb{E}_t[\text{Price}_{t+1}]}{1+\text{R}} \tag{8}$$

The DDM assumes that a stock's current price will reflect the discounted value of its total payout next year. Part of that payout will come in the form of a cash dividend. Part will come from the resale price.

Step 2. Assume investors think about future prices in present-value terms.

Suppose you buy a 1-year \$100 bond with a 2% coupon payment. Next year you will receive a 2% × \$100 = \$2 coupon and a \$100 face-value payment. You do not need to have a theory of how large the face-value payment will be. The \$100 amount is stated on the bond at the time of purchase. Things are different for stocks. The anticipated resale price is analogous to the bond's \$100 face value. You do not know $\text{Price}_{t+1}$ ahead of time.

Hence, a theory of current stock prices requires a theory of future stock prices. The DDM assumes that they are the same theory. The model says that investors reason about future prices at every horizon $h = 1, 2, 3, \ldots$ in the same way as the current price level

$$\mathbb{E}_t[\text{Price}_{t+h}] = \frac{\mathbb{E}_t[\text{Div}_{(t+h)+1}]}{1+\text{R}} + \frac{\mathbb{E}_t[\text{Price}_{(t+h)+1}]}{1+\text{R}} \tag{9}$$

This second present-value assumption is much stronger than the first. $\text{Price}_t$ is an observed outcome that is subject to market forces. One could argue that the invisible hand somehow enforces Equation (8) even if no individual investor uses it. The same argument does not work for Equation (9). $\mathbb{E}_t[\text{Price}_{t+h}]$ is a number that exists in an investor's head, not on a Bloomberg Terminal. Prognostications and guesses do not have to satisfy budget constraints or clear markets.

Step 3. Iterate forward, pushing unknown resale price far into the future.

The DDM assumes that investors think about the current price level and future resale prices using the same present-value logic. This self-similarity makes it possible to iterate forward, recursively replacing the future resale price with its one-step-ahead present-value equivalent

$$\mathrm{Price}_t = \sum_{h=1}^{H} \frac{\mathbb{E}_t[\mathrm{Div}_{t+h}]}{(1+\mathrm{R})^h} + \frac{\mathbb{E}_t[\mathrm{Price}_{t+H}]}{(1+\mathrm{R})^H} \tag{10}$$

The first term is the present value of the stock's dividends over the next $H$ years. The second is the present value of its resale price in year $H$.

Step 4. Use transversality condition to eliminate resale price completely.

If $H$ is large enough, discounting takes over and the present value of the future resale price can be ignored. Researchers call this a "transversality condition"

$$\lim_{H\to\infty} \frac{\mathbb{E}_t[\mathrm{Price}_{t+H}]}{(1+\mathrm{R})^H} = \$0/\mathrm{sh} \tag{11}$$

Romans did not care about prices today when trading stocks in 1AD.

If we can ignore the second term in Equation (10) when taking the limit $H \to \infty$, then we arrive at the Dividend Discount Model (DDM)

$$\mathrm{Price}_t = \sum_{h=1}^{\infty} \frac{\mathbb{E}_t[\mathrm{Div}_{t+h}]}{(1+\mathrm{R})^h} \tag{12}$$

This classic model says that a stock's current price should reflect the discounted value of its expected future dividend stream.

Recall that pricing a stock required a theory of the current price level and a theory of the future resale price. The DDM assumes that both theories are one and the same. As a result, it is possible to iterate forward and remove the future resale price from Equation (10). The DDM effectively prices stocks by pretending they are fixed-income products. This is why the discount rate in the DDM behaves like a bond yield, moving in the opposite direction as the current price.

Step 5. Assume expected annual dividend-growth rate is constant.

The Gordon model is a special case of the DDM. To operationalize Equation (12), you must make some sort of assumption about a stock's future dividend growth. It is not physically possible to sum an infinite series of arbitrary forecasts. Gordon (1962) points out a particularly convenient special case. Suppose that, in addition to R being constant, expected dividend growth is also constant, $\mathbb{E}_t[\mathrm{Div}_{t+h}] = (1{+}\mathrm{G})^h \cdot \mathrm{Div}_t$. It is then possible to price a stock by multiplying next year's dividend forecast times

$$\left(\frac{1}{\mathrm{R}-\mathrm{G}}\right) = \left(\sum_{h=1}^{\infty}\frac{\mathbb{E}_t[\mathrm{Div}_{t+h}]}{(1+\mathrm{R})^h}\right) \Big/ \ \mathbb{E}_t[\mathrm{Div}_{t+1}] \tag{13}$$

This is the pricing formula for a growing perpetuity. Gordon says to price stocks by pretending they are really simple bonds.

There is a lot more to the Gordon model than assuming constant parameters. Step 5 is just the icing on the cake. Before you get there, investors must set today's price using one-step-ahead present-value logic. They must believe every future price also obeys the same present-value relationship. You must iterate this logic forward indefinitely. The present value of the resale price must vanish in the limit. Only after all of this does the constant-parameter assumption enter, converting the general DDM into a tractable special case.

Campbell-Shiller is meant to relax the constant-parameter assumption. The log of Equation (13) is directly comparable to the Campbell-Shiller formula

$$\log(\mathrm{R}-\mathrm{G}) \ \sim \ \sum_{h=1}^{\infty} \rho^{h-1}\cdot \left\{ \mathbb{E}_t[\mathrm{R}_{t+h}] - \mathbb{E}_t[\Delta\log \mathrm{Div}_{t+h}] \right\} \tag{14}$$

The formula allows R and G to vary over time, replacing the constant (R–G) with a weighted sum of future expected values. This is a useful extension in situations where prices are governed by an approximate Gordon model. However, serious problems arise when researchers try to interpret Equation (1) as a generalized Gordon multiple outside of the original model's applicable domain.

## 2.2 Present-Value Logic

The forward-looking version of Campbell-Shiller is meant to be a generalized Gordon model. For that interpretation to hold, Steps 1 through 4 of the previous subsection still have to hold. Researchers sometimes flag Step 4 as the substantive assumption. It is not. Transversality is a mopping-up condition. The heavy lifting is done by the forward iteration in Step 3, and this step relies on the prior two. Step 1 says investors price today's share using one-period-ahead present-value logic. Step 2 says they forecast every future resale price the same way. For Equation (1) to always hold under correct expectations, investors must reliably think in present-value terms. They do not.

Sell-side equity analysts are professional forecasters whose price targets move billions of dollars of trading volume. Ben-David and Chinco (2025) documents that they typically forecast a company's share price next year using a trailing multiple, not a forward-looking one like in the Gordon model. To set a one-year-ahead price target, analysts start by making a next-twelve-month EPS forecast, $\mathbb{E}[\text{EPS}]$. Then, they multiply this short-term estimate times the company's trailing PE ratio, $\text{TrailingPE} = \text{Price}/\text{EPS}$

$$\text{PriceTarget} \;=\; \mathbb{E}[\text{EPS}] \times \text{TrailingPE} \tag{15}$$

Textbook finance theory takes it for granted that analysts set price targets by asking: "How much is the company's expected future earnings stream worth to me in today's dollars?" Instead, they ask themselves: "How would the company's next-twelve-month earnings forecast have been priced in today's market or last year?" A PE ratio is a characteristic of the firm, not a property of expected future cash flows. Analysts rely on the same trailing multiple even in situations where they anticipate changes to the company's future earnings stream. With this sort of dynamic inconsistency, it is not possible to do the forward iteration needed in Step 3 of the Gordon model.

Sell-side analysts are not the only investors who think this way, either. In February 2013, David Einhorn of Greenlight Capital Management gave a public presentation to fellow Apple shareholders in an effort to convince the

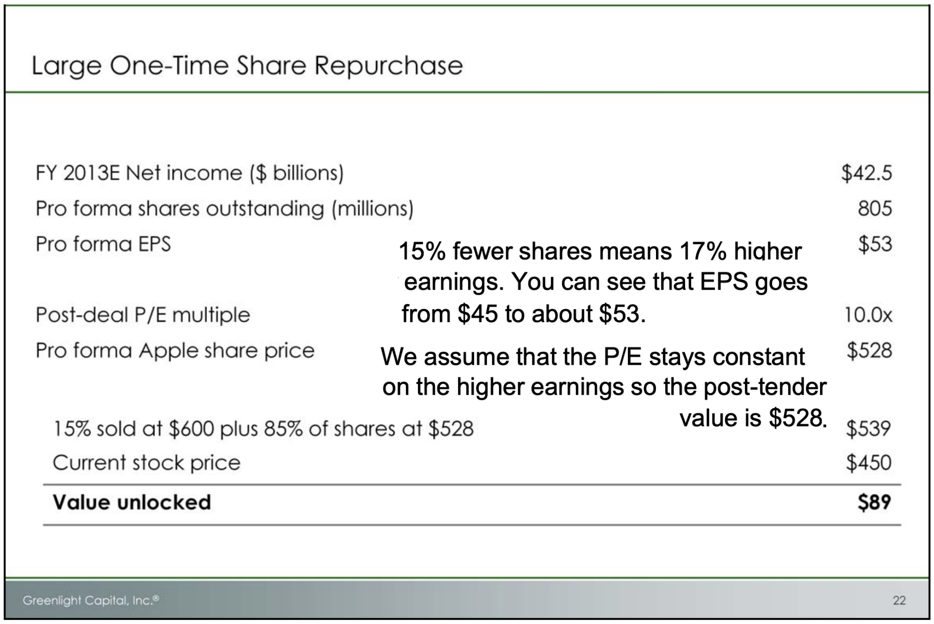


**Figure 1.** Slide from a February 2013 presentation by David Einhorn of Greenlight Capital Management to Apple (AAPL) shareholders. Einhorn calculates that an $84B cash-financed repurchase program would generate a 7.80/sh EPS pop and a share price increase of $\$528/\mathrm{sh} - \$450/\mathrm{sh} = \$7.80/\mathrm{sh} \times 10 = \$78/\mathrm{sh}$.

company to distribute its cash hoard.[1] Figure 1 shows a key slide from this presentation. Einhorn's argument was that an $84B cash-financed repurchase program would mechanically boost Apple's per-share earnings from $45/sh to $52.80/sh, and that this $7.80/sh EPS gain should be valued at Apple's existing 10× PE ratio. His pitch was that repurchases would increase share prices by $\$528/\mathrm{sh} - \$450/\mathrm{sh} = \$7.80/\mathrm{sh} \times 10 = \$78/\mathrm{sh}$.

Einhorn was a sophisticated investor making a public case to the shareholders of one of the world's largest companies. The valuation framework he reached for was $\mathbb{E}[\mathrm{EPS}] \times \mathrm{TrailingPE}$, not a discounted cash-flow model. Einhorn knew that removing $84B of cash from Apple's balance sheet would radically change the risk profile and growth rate of future earnings. He still used a trailing PE to determine the price impact: post-deal multiple = pre-deal multiple = 10×. This is not dynamically consistent. Einhorn did it anyway.

[1] David Einhorn. "iPrefs: Unlocking Value." Conference Call. February 21, 2013. [link]

Hartzmark and Sussman (2025) provides perhaps the most direct test of whether investors apply present-value logic. If people reliably set prices by discounting expected future payoffs, then you should be able to hand someone information about a company's fundamentals and have them give you back the correct price. This is exactly what the authors do. First, they provide survey participants with information about fundamentals. Then, the authors show participants an incorrect price and ask whether it is too high or too low. Subjects cannot reliably classify even extreme deviations from the true price. In their exit interview, many participants explained that it is just not possible to accurately determine the level of the stock market using present-value logic alone.

One reaction to this evidence is that a lot of people are making dumb mistakes. That could be. But it also might not be unreasonable to deviate from present-value logic. Consider the housing market. If you own a home, which discount rate did you use when deciding how much to pay for it? We are guessing you did not calculate a discounted cash flow model. This does not make you a bad economist. Almost no one does.

Homebuyers look at comparable sales, recent transaction prices for similar homes in similar neighborhoods. The pricing rule is backward-looking: what did this kind of house sell for recently? Not what is the discounted present value of the rents this house will generate over the next thirty years? This backward-looking approach is so deeply embedded that the dominant methodology for measuring home prices, the Case-Shiller repeat-sales index, is explicitly built on it (Case and Shiller, 1987, 1989). Every market does not operate according to present-value principles. Bond markets do. Housing markets clearly do not. The evidence above suggests that, in many cases, neither do equity markets.

Gao and Martin (2021) calls time-varying R and G "the empirically relevant case." That is a claim about Step 5. The question is whether Steps 1 through 4 are empirically relevant, too. The examples above demonstrate that Steps 1 and 2 are not always justified. Without them it is not possible to perform the forward iteration in Step 3, rendering any notion of transversality in Step 4 a moot point. In these scenarios, the Gordon model does not describe how stocks are priced, and the reason has nothing to do with time-varying parameters.

## 2.3 Required Precision

Suppose, for the sake of argument, that Steps 1 through 4 hold. Investors reliably apply one-period-ahead present-value logic at $t$ and at every future date. The forward iteration runs. The transversality condition clears the tail. On the surface, a time-varying extension of the Gordon model now looks reasonable. But could an investor actually use such a model to set prices? The answer is no. Even before allowing R and G to vary over time, the Gordon formula requires investors to know both parameters with an implausible level of precision.

We use the "accounting identity" language to quantify how much precision is needed. For something to qualify as an accounting identity, you have to be able to reliably count each term. In the Dividend Discount Model (Equation 12), the single largest term is next year's expected dividend, $\frac{\mathbb{E}_t[\text{Div}_{t+1}]}{1+\text{R}}$. Every subsequent term is smaller by a factor of $\left(\frac{1+\text{G}}{1+\text{R}}\right) < 1$. This puts an upper bound on the amount of noise the Gordon model can tolerate. If errors have a larger price impact than the present value of next year's dividend forecast, we cannot reliably tell whether the larger term has been "counted" in the price.

**Proposition 2.3.** *Suppose prices obey the Gordon model in Equation (2) with* $\text{G} \geq 0\%$. *If noise distorts the cap rate by* $|\Delta(\text{R}-\text{G})| = \delta > 0$, *then the resulting price change will exceed the present value of next year's dividend forecast whenever*

$$|\delta| > \text{DY}^2 \qquad \rightsquigarrow \qquad |\Delta\text{Price}_t| > \frac{\mathbb{E}_t[\text{Div}_{t+1}]}{1+\text{R}} \tag{16}$$

The S&P 500 has an average dividend yield of $\overline{\text{DY}} \approx 2\%$, so the threshold is $\overline{\text{DY}}^2 \approx 0.04\%$. This is preposterously small. Standard estimates of the equity risk premium range from 4% to 8%, a spread that is 100× larger. The variance decomposition literature treats Equation (1) as an accounting identity. Researchers attribute all fluctuations in log PD to revisions in expected future returns or dividend growth. This is wrong. If investors were actually using the kind of model that Campbell-Shiller is meant to approximate, then log PD would have to move for lots of other random reasons. Allowing R and G to change over time does not produce a better model. It makes the problem worse.

# 3 Subjective Beliefs

Assume that Campbell-Shiller holds both in-sample and under correct expectations. When will the formula remain valid under alternative subjective beliefs? Conventional wisdom says: always. The correct answer is: rarely. Subsection 3.1 shows that the formula requires forecast errors to satisfy an infinite-horizon adding-up condition involving a precise weighting scheme. Subsection 3.2 looks at various belief specifications. Each one can be both consistent and inconsistent with Campbell-Shiller. The outcome depends on the additional modeling assumptions that a researcher makes. Subsection 3.3 generalizes this point by proving that consistency with Campbell-Shiller is not a model-free property of subjective beliefs. Calling the formula an "identity" makes it harder to understand how assets are actually priced. Subsection 3.4 underscores this conclusion by digging into the details of one popular application: variance decompositions.

## 3.1 Adding-Up Condition

Assume that Campbell-Shiller holds in-sample and under correct expectations. Equation (1) describes the market price of some stock. Suppose a researcher surveys investors to find their subjective beliefs about this stock's future dividend growth and returns, $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}]$ and $\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}]$. Then, she plugs this survey data into Equation (1) in place of the objectively correct predictions, $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}]$ and $\mathbb{E}_t[\mathrm{R}_{t+h}]$. The resulting formula will not always remain valid. Equality only persists if the resulting forecast errors satisfy a precise adding-up condition. This subsection characterizes that condition.

Researchers currently think of "subjective expectations" as a synonym for "subjective beliefs." De la O and Myers (2024) writes that Campbell-Shiller "holds even if subjective expectations are not rational." Adam and Nagel (2023)'s survey chapter on subjective beliefs is titled "Expectations data in asset pricing." As things stand, researchers take it for granted that every prediction, forecast, and prognostication can be written as the output of a (potentially biased) subjective expectation operator, $\widetilde{\mathbb{E}}_t[\cdot]$.

Not so. It is possible to hold views about the future that do not satisfy the rules of probability. Let $\widetilde{\mathbb{F}}_t[\cdot]$ denote an arbitrary forecasting rule, a mapping from random variables to real numbers. This is a more permissive object than a subjective expectation. Every $\widetilde{\mathbb{E}}_t[\cdot]$ is a legal $\widetilde{\mathbb{F}}_t[\cdot]$. The converse is not true. $\widetilde{\mathbb{E}}_t[\cdot]$ is linear. $\widetilde{\mathbb{F}}_t[\cdot]$ need not be. $\widetilde{\mathbb{E}}_t[\cdot]$ must satisfy the law of iterated expectations. $\widetilde{\mathbb{F}}_t[\cdot]$ need not.

The distinction between $\widetilde{\mathbb{F}}_t[\cdot]$ and $\widetilde{\mathbb{E}}_t[\cdot]$ makes a big difference. A forecasting rule can be arbitrarily wrong. A subjective expectation operator cannot. Its predictions must be coherent and respect well-known rules: linearity, iterated expectations, and respect for almost-sure identities all follow from having one. When researchers assume that investors' subjective beliefs come from $\widetilde{\mathbb{E}}_t[\cdot]$ rather than $\widetilde{\mathbb{F}}_t[\cdot]$, they are imposing this extra structure on their analysis.

Consider an investor anchored in Gordon logic. She believes that future returns and dividend growth are both constant. Rearranging Equation (2) gives a single pricing constraint, $(\mathrm{R}-\mathrm{G}) = \mathrm{DY}_t$. Once the investor observes a stock's current dividend yield, $\mathrm{DY}_t$, she is no longer free to believe anything she wants about R and G. If the investor's subjective beliefs satisfy the laws of probability given her model, then there is only one degree of freedom, not two. If her return forecast is $\widetilde{\mathbb{E}}_t[\mathrm{R}]-\mathrm{R} = 1$%pt too high, then her anticipated dividend-growth rate must be $\widetilde{\mathbb{E}}_t[\Delta \log \mathrm{Div}]-\mathrm{G} = 1$%pt too high, as well. Given the investor's model, the two errors are locked by the requirement that she must update her beliefs in response to $\mathrm{DY}_t$ in accordance with the laws of probability.

This internal-consistency requirement disappears when we replace subjective expectations, $\widetilde{\mathbb{E}}_t[\cdot]$, with a more general forecasting rule, $\widetilde{\mathbb{F}}_t[\cdot]$. If we relax the assumption that every subjective belief can be written as a subjective expectation, then the investor is no longer required to reconcile her views about R and G with the observed data given her model. Her return forecasts can move independently from her dividend-growth forecasts.

The following proposition shows that changes in $\widetilde{\mathbb{F}}_t[\mathrm{R}]$ must be carefully choreographed with changes in $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}]$ to preserve equality in Campbell-Shiller. Among all possible subjective forecasting rules, the fraction that perform this dance is essentially zero. They are knife-edge cases.

**Proposition 3.1.** *Suppose that Campbell-Shiller holds in-sample and under correct expectations as shown in Equation (1). If we replace $\mathbb{E}_t[\cdot]$ with $\widetilde{\mathbb{F}}_t[\cdot]$, the formula continues to hold if and only if the following adding-up condition is satisfied*

$$\underbrace{\sum_{h=1}^{\infty} \frac{(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \mathrm{Div}_{t+h}]}{(1+\overline{\mathrm{DY}})^{h-1}}}_{\substack{\text{Present value of errors about future} \\ \text{dividend growth discounted at } \overline{\mathrm{DY}}}} = \underbrace{\sum_{h=1}^{\infty} \frac{(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\mathrm{R}_{t+h}]}{(1+\overline{\mathrm{DY}})^{h-1}}}_{\substack{\text{Present value of return} \\ \text{errors discounted at } \overline{\mathrm{DY}}}} \tag{3}$$

The size of the mismatch between the left- and right-hand side of Equation (3) has a direct economic interpretation. It equals the difference between the log PD ratio implied by investors' subjective beliefs and the log PD ratio observed in the data. A value of 0.1 log units means that investors' beliefs predict a PD ratio that is $e^{0.1}-1 \approx 11\%$ higher than the one in the data.

Notice how the errors are discounted at $\overline{\mathrm{DY}}$, not at R. This has nothing to do with risk preferences or optimization. It comes from centering the log-linear approximation around a stock's long-run average PD ratio, $\overline{\log \mathrm{PD}}$. The S&P 500 has $\overline{\mathrm{DY}} \approx 2\%$, so an error in next year's forecast, $h = 1$, matters $(1{+}2\%)^9 \approx 19.5\%$ more than an error about what will happen in $h = 10$ years.

An arbitrarily chosen $\widetilde{\mathbb{F}}_t[\cdot]$ will generically violate the adding-up condition in Equation (3). Satisfying it requires restricting attention to some small subset. The subset the literature has chosen is $\widetilde{\mathbb{E}}_t[\cdot]$. To see what "generically" means, note that a subjective forecasting rule $\widetilde{\mathbb{F}}_t[\cdot]$ is a point in an infinite-dimensional space, with one coordinate per horizon for each of future returns and dividend growth. Adding-up is a single linear restriction on this space. The set of $\widetilde{\mathbb{F}}_t[\cdot]$ that satisfies it is a thin slice of the whole space. Under any reasonable way of putting a probability measure on forecasting rules, that slice has measure zero.

$\widetilde{\mathbb{E}}_t[\cdot]$ stems from a coherent probability measure. This is why it is linear, why it obeys the law of iterated expectations, and why it respects almost-sure equalities. These three properties imply that the adding-up condition holds. The derivation of the Gordon model relies heavily on the law of iterated expectations. The preservation of almost-sure relationships is relevant for transversality. An arbitrary forecasting rule, $\widetilde{\mathbb{F}}_t[\cdot]$, is under no obligation to follow these rules.

## 3.2 Belief Specifications

Proposition 3.1 provides a mathematical requirement in the abstract. We now fill in the details by looking at specific examples. We start simple and gradually build up to popular models of subjective belief formation. The examples highlight a recurring pattern. In its simplest form, each error is incompatible with Campbell-Shiller. To restore consistency, a researcher must introduce further assumptions.

**Perpetually Wrong.** Suppose investors have correct dividend-growth forecasts, $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] = \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}]$. However, they overestimate future returns at every horizon by the same amount, $\delta > 0$

$$\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] = \mathbb{E}_t[\mathrm{R}_{t+h}] + \delta \tag{17}$$

Investors are perpetually wrong about annual returns.

Since there are no dividend-growth mistakes, the left-hand side of Equation (3) is zero. The right-hand side is a geometric sum

$$\sum_{h=1}^{\infty} \frac{\delta}{(1+\overline{\mathrm{DY}})^{h-1}} \approx \delta \times \left(\frac{1}{\overline{\mathrm{DY}}}\right) \tag{18}$$

For the S&P 500 with $\overline{\mathrm{DY}} \approx 2\%$, a perpetual $\delta = 1$%pt return overestimate generates an adding-up violation of 0.5 log units. Investors' beliefs imply a PD ratio that is $1-e^{-0.5} \approx 39\%$ lower than the one in the data.

Notice that, if prices come from a Gordon model, then the multiplier $\left(\frac{1}{\overline{\mathrm{DY}}}\right) = \left(\frac{1}{\mathrm{R-G}}\right)$. In the Gordon model, a stock's current price is $\mathbb{E}_t[\mathrm{Div}_{t+1}] \times \left(\frac{1}{\mathrm{R-G}}\right)$, with $(\mathrm{R-G}) = \overline{\mathrm{DY}}$ in steady state. A perpetual $\delta$ error on future returns is capitalized at exactly the rate Gordon uses to capitalize cash flows. That is why a small bias turns into a large inconsistency: $\overline{\mathrm{DY}}$ is small.

**Moving Target.** Now suppose the investor's initial $\delta > 0$ overestimate about next year's return decays geometrically at rate $\phi \in [0, 1)$

$$\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] = \mathbb{E}_t[\mathrm{R}_{t+h}] + \phi^{h-1} \cdot \delta \tag{19}$$

The $\phi^{h-1}$ reflects AR(1) dynamics with persistence $\phi$. Dividend-growth forecasts are correct, just like before.

The adding-up violation is now a slightly different geometric sum

$$\sum_{h=1}^{\infty} \frac{\phi^{h-1} \cdot \delta}{(1+\overline{\mathrm{DY}})^{h-1}} \;=\; \delta \times \left(\frac{1}{1-\rho \cdot \phi}\right) \tag{20}$$

The multiplier $\left(\frac{1}{1-\rho \cdot \phi}\right)$ recurs throughout the literature. For small $\overline{\mathrm{DY}}$, it is the Gordon multiple with a mean-reversion correction, $\left(\frac{1}{1-\rho \cdot \phi}\right) \approx \frac{1}{\overline{\mathrm{DY}}+(1-\phi)}$.

It is helpful to think about some limiting cases. When $\phi = 1$, the error persists indefinitely. We are back in the perpetually wrong case from before. At $\overline{\mathrm{DY}} \approx 2\%$, a $\delta = 1\%$pt error gets transformed into a $1-e^{-0.5} \approx 39\%$ PD mismatch. When $\phi = 0$, the multiplier collapses to 1. A $\delta = 1\%$pt error generates an adding-up violation of 0.01 log units and implies a PD mismatch of $1-e^{-0.01} \approx 1\%$. For an interior case with $\overline{\mathrm{DY}} \approx 2\%$ and $\phi = 0.6$, the multiplier is roughly 2.4×. A $\delta = 1\%$pt error generates a violation of 0.024 log units, implying a PD ratio that is $1-e^{-0.024} \approx 2.4\%$ lower than the observed data.

**Measurement Error.** No one knows $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}]$ and $\mathbb{E}_t[\mathrm{R}_{t+h}]$ with infinite precision. Suppose a researcher plugs her best guesses into Equation (1). In an ideal world, her estimation errors on all inputs would be IID white noise. This is an unrealistically optimistic view of measurement error. In general, mistakes will not be mutually independent across horizons and between variables.

Nevertheless, the resulting mistakes still break the adding-up condition in Equation (3) almost surely. The violation is a mean-zero random variable with volatility

$$\sqrt{\sigma_{\mathrm{Div}}^2 + \sigma_{\mathrm{R}}^2} \times \left(\frac{1}{1-\rho^2}\right)^{0.5} \tag{21}$$

$\sigma_{\mathrm{Div}}$ and $\sigma_{\mathrm{R}}$ are the standard deviations of the forecast errors about future dividend growth and returns. Equality will be violated if either is positive.

For the S&P 500, $\overline{\mathrm{DY}} \approx 2\%$, so $\rho = \left(\frac{1}{1+\overline{\mathrm{DY}}}\right) \approx 0.98$ and $\left(\frac{1}{1-0.98^2}\right)^{0.5} \approx 5$. Suppose researchers knew R and G to within ±1%pt. In that case, the adding-up violations would have a standard deviation of $\sqrt{0.0002} \times 5 \approx 0.07$ log units, corresponding to a PD mismatch of $e^{0.07}-1 \approx 7\%$. Researchers do not know R and G to anything close to that level of accuracy. Standard estimates of the equity risk premium range from 4% to 8%. If we use $\sigma_{\text{Div}} = \sigma_{\text{R}} = 4\%\text{pt}$, then the volatility balloons to $\sqrt{0.0032}\times 5 \approx 0.28$ log units and the implied PD mismatch rises to $e^{0.28}-1 \approx 33\%$. This calculation rhymes with Proposition 2.3. Two routes, one destination: plausible levels of noise create real problems for Campbell-Shiller.

**Return Extrapolation.** Investors who extrapolate returns wind up overweighting recent price increases when forecasting future returns. Suppose demeaned returns follow an AR(1) process, $\text{R}_t = \phi \cdot \text{R}_{t-1} + \varepsilon_t$. The residual represents the gap between a stock's realized return over the past year and the return investors were anticipating. The subjective forecasting rule for returns is given by

$$\widetilde{\mathbb{F}}_t[\text{R}_{t+h}] \;=\; \mathbb{E}_t[\text{R}_{t+h}] \;+\; \phi^h \cdot (\theta \cdot \varepsilon_t) \tag{22}$$

$\theta > 0$ controls the bias strength. $\phi^h$ comes from the rational forecast revision. The adding-up violation sums to $\phi \cdot (\theta \cdot \varepsilon_t) \times \left(\frac{1}{1-\rho\cdot\phi}\right)$.

The S&P 500 has $\overline{\text{DY}} \approx 2\%$, $\rho = \left(\frac{1}{1+2\%}\right) \approx 0.98$, and persistence $\phi = 0.1$. Given an extrapolative strength of $\theta = 1$, a one-year surprise of $\varepsilon_t = 10\%\text{pt}$ would translate into an adding-up violation of roughly 0.01 log units, giving a PD mismatch of just $1-e^{-0.01} \approx 1\%$. The total violation is small because $\phi = 0.1$ is small. Longer-term forecast errors have little effect since $0.1^h$ gets small fast. The same small persistence also keeps the multiple small, $\left(\frac{1}{1-0.98\cdot 0.1}\right) \approx 1.1$. As a result, the total adding-up violation in response to a 10%pt forecast error is dominated by the first term, $0.1 \times 10\%\text{pt} = 1\%$.

Going from 1% to 0% requires an equilibrium model. Barberis et al. (2015) embeds the bias in a two-agent economy where extrapolators trade against rational investors. Market clearing determines the equilibrium price process, which satisfies the adding-up condition under return extrapolators' subjective measure. Jin and Sui (2022) does the same with a representative agent and a subjective Euler equation. Both papers assume that investors' views about future dividend growth are correct, which is what forces them to build out the equilibrium structure.

**Diagnostic Beliefs.** Bordalo et al. (2018) proposes diagnostic beliefs: investors overweight recent cash-flow surprises rather than recent unexpected price movements. Suppose a stock's demeaned dividend growth follows an AR(1) process, $\Delta \log \mathrm{Div}_t = \phi \cdot \Delta \log \mathrm{Div}_{t-1} + \varepsilon_t$. When news arrives at time $t$, a rational investor would revise her $h$-period forecast by $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] - \mathbb{E}_{t-1}[\Delta \log \mathrm{Div}_{t+h}] = \phi^h \cdot \varepsilon_t$. Investors with diagnostic beliefs overweight this change by $\theta > 0$

$$\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] \;=\; \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] \;+\; \theta \cdot (\phi^h \cdot \varepsilon_t) \tag{23}$$

When paired with correct return forecasts, diagnostic beliefs predict a current log PD ratio that is off by $\theta \cdot (\phi \cdot \varepsilon_t) \times \big(\frac{1}{1-\rho \cdot \phi}\big)$.

The S&P 500 has $\rho \approx 0.98$. Suppose that dividend-growth has persistence of $\phi = 0.5$ and investors have a diagnostic strength of $\theta = 1$. Under these choices, a cash-flow surprise of $\varepsilon_t = 10\%\text{pt}$ would translate into an adding-up violation of 0.1 log units. In other words, the model-implied PD ratio would be $e^{0.10}-1 \approx 10\%$ higher than the observed value. While the mathematics is analogous to extrapolation applied to dividend growth instead of returns, the underlying economics is different. Dividend growth is much more persistent than returns. As a result, the same initial surprise leads to an adding-up violation that is roughly 10× larger.

Bordalo, Gennaioli, La Porta, and Shleifer (2019) shows how to restore consistency with Campbell-Shiller. The paper satisfies the adding-up condition by treating the price as a residual. The authors model diagnostic investors who anticipate a constant annual return, say, $\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] = 10\%$ for all $h \geq 1$. The paper then calculates a stock's current price by discounting investors' diagnostic dividend forecasts at 10% per year.

**Personal Experience.** Malmendier and Nagel (2011) proposes that agents overweight personal experience when forecasting returns. Persistence adds little in this application, so we assume demeaned returns are approximately unpredictable, $\mathrm{R}_t = \varepsilon_t$. Let $\mu_t$ denote the investor's lifetime experience, $\mu_t = \theta \cdot \mu_{t-1} + (1-\theta) \cdot \varepsilon_t$, which updates with weight $\theta \in (0, 1)$. Larger values of $\theta$

imply old surprises stay in memory longer. The investor's subjective return forecasts are

$$\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] \;=\; \mathbb{E}_t[\mathrm{R}_{t+h}] \;+\; \mu_t \tag{24}$$

The forecast error is the same at every horizon. The investor simply thinks returns are $\mu_t$ higher, indefinitely. The adding-up violation is $\mu_t \times \left(\frac{1}{1-\rho}\right)$ log units.

For the S&P 500 with $\rho \approx 0.98$ and an experience-driven bias of $\mu_t = 1\%\text{pt}$, the violation is 0.5 log units, implying a PD ratio that is $1-e^{-0.5} \approx 39\%$ lower than the observed value. This is the perpetually-wrong case from earlier.

Malmendier and Nagel (2011) measures these experience effects from the UBS/Gallup survey, where cohort-level return forecasts are correlated with cohort-specific experienced returns. The survey data comes from $\widetilde{\mathbb{F}}_t[\cdot]$, not $\widetilde{\mathbb{E}}_t[\cdot]$. A researcher who combined these forecasts with separately sourced dividend-growth data in a forward-looking Campbell-Shiller framework would violate adding-up. Restoring consistency requires an equilibrium model.

**Natural Expectations.** Fuster et al. (2010) proposes that agents fit a parsimonious model to data generated by a higher-order process. For simplicity, assume $\Delta \log \mathrm{Div}_t$ has been demeaned and follows an AR(2), $\Delta \log \mathrm{Div}_t = \phi_1 \cdot \Delta \log \mathrm{Div}_{t-1} + \phi_2 \cdot \Delta \log \mathrm{Div}_{t-2} + \varepsilon_t$, but investors fit an AR(1), $\Delta \log \mathrm{Div}_t = \hat{\phi}_1 \cdot \Delta \log \mathrm{Div}_{t-1} + \hat{\varepsilon}_t$. The agent's subjective forecast is

$$\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] \;=\; \hat{\phi}_1^h \cdot \Delta \log \mathrm{Div}_t \tag{25}$$

The agent winds up overestimating persistence, capturing short-run momentum but missing long-run mean reversion.

The adding-up condition is generically violated. Fuster, Hebert, and Laibson (2012) solves a consumption-based asset-pricing model in which the agent prices assets via an Euler equation under the misspecified model. The Euler equation simultaneously pins down expected returns and expected dividend growth under the agent's coherent (but wrong) probability measure. The beliefs become $\widetilde{\mathbb{E}}_t[\cdot]$, and the adding-up condition is satisfied by construction.

## 3.3 Model Overhead

We have just seen that, when considered in isolation, popular models of biased beliefs violate the adding-up condition needed by Campbell-Shiller. However, in each example, it was possible to restore consistency by layering on extra assumptions. What exactly do these assumptions accomplish? This subsection answers that question.

Let's start with the Gordon model. Suppose a stock is currently trading at $\text{Price}_t = \$100/\text{sh}$. Over the past twelve months, the company has distributed a per-share dividend of $\text{Div}_t = \$2/\text{sh}$, giving the stock a dividend yield of $\text{DY}_t = 2\%$. The Gordon pricing rule in Equation (2) requires an investor to forecast two key parameters: $\widetilde{\mathbb{F}}_t[\text{R}_{t+h}] = \tilde{\text{R}}$ and $\widetilde{\mathbb{F}}_t[\Delta \log \text{Div}_{t+h}] = \tilde{\text{G}}$. The internal logic of the model ties these two subjective forecasts, which exist in the investor's head, to the stock's dividend yield, which can be viewed on a Bloomberg terminal.

This model-implied connection allows the investor to cross-validate her subjective beliefs about future returns and dividend growth with the observed data. The dividend yield implied by her subjective beliefs ought to match the yield seen in the data

$$\underbrace{\left(\tfrac{\$2/\text{sh}}{\$100/\text{sh}}\right) = \text{DY}_t}_{\text{In data, 2\%}} \quad = \quad \underbrace{\tilde{\text{R}} - \tilde{\text{G}}}_{\text{In head}} \tag{26}$$

Under rational expectations, the investor's subjective forecasts are objectively correct. Suppose that the company's per-share dividend grows at $\text{G} = 3\%$ per year, $\mathbb{E}_t[\text{Div}_{t+h}] = (1{+}3\%)^h \cdot \$2/\text{sh}$, and the investor's forecast is spot on, $\tilde{\text{G}} = 3\%$. The stock's price comes from discounting these expected future dividends at $\text{R} = 5\%$ per year. In a full-information rational-expectations setup, the investor knows this and anticipates future returns of $\tilde{\text{R}} = 5\%$ each year.

Now let's think about what would happen if the investor thinks returns will be 1%pt higher at every horizon. Instead of forecasting $\text{R} = 5\%$, she anticipates $\tilde{\text{R}} = 6\%$. If the investor still holds correct views about the future dividend-growth rate, $\tilde{\text{G}} = \text{G} = 3\%$, then the adding-up condition in Equation (3) will be violated. If the stock's long-run average dividend yield reflects correct pricing, $\overline{\text{DY}} = \text{R}{-}\text{G} = 2\%$, the size of the violation will be $1\%\text{pt} \times \left(\tfrac{1}{2\%}\right) = 0.5$ log units.

To make the Gordon model internally consistent with this perpetual mistake, a researcher would need to introduce further assumptions. On its own, an investor would immediately recognize this 1%pt-too-high return-forecast mistake. Her subjective beliefs imply a PD ratio of $\left(\frac{1}{6\%-3\%}\right) \approx 33\times$. The PD ratio in the data is $\mathrm{PD}_t = \left(\frac{\$100/\mathrm{sh}}{\$2/\mathrm{sh}}\right) = 50\times$, which is $\left(\frac{50\times - 33\times}{33\times}\right) \approx 50\%$ larger.

One way to restore consistency would be to assume that the investor also overestimates the dividend-growth rate at every horizon by 1%pt. Instead of $\mathrm{G} = 3\%$, suppose that the investor forecasts $\tilde{\mathrm{G}} = 4\%$. In that case, the two forecasting errors would cancel out, $\tilde{\mathrm{R}} - \tilde{\mathrm{G}} = \mathrm{R} - \mathrm{G} = 2\%$. The investor would be wrong about both forecasts in a way that does not affect the dividend yield she anticipates seeing in the data. The model-implied PD ratio matches the observed value, and her beliefs are cross-validated.

There are many ways to pull off this same trick. The Diagnostic Expectations example has the investor overweighting recent cash-flow surprises when forecasting future dividend growth. Bordalo et al. (2019) restores adding-up by also assuming that the investor anticipates constant annual returns. The stock's price gets computed by discounting at this same rate. If all investors think like this, then the new equilibrium price will work like a residual, adjusting to ensure there is no adding-up violation.

The pattern repeats across both fixes. The investor must be able to cross-validate her subjective forecasts against the observed data. Otherwise her subjective beliefs are not consistent with what the market shows. Any arbitrary set of beliefs will not work. This requirement imposes a lot of additional structure on investors' subjective forecasts. In fact, it imposes so much structure that her subjective forecasts must be conditional expectations under a coherent probability measure that respects her asset-pricing model.

**Proposition 3.3.** *Suppose the investor's subjective forecasts cross-validate: when plugged into her asset-pricing model, they are consistent with the observed market data. Then:*

a) *The investor's subjective beliefs are conditional expectations under a coherent probability measure that respects her asset-pricing model.*
b) *The adding-up condition in Equation (3) is satisfied.*

In the pre-proposition discussion, the additional structure showed up as a single equation: $\mathrm{DY}_t = \tilde{\mathrm{R}} - \tilde{\mathrm{G}}$. The Gordon model tied the investor's two subjective forecasts together with the observed dividend yield in a specific way. Proposition 3.3 generalizes this idea. The investor's subjective forecasting rule must make predictions consistent with all possible events that the data-generating process can produce given her asset-pricing model.

This level of consistency points to a well-posed probability measure. To be consistent across every event her asset-pricing model can produce, the investor's predictions cannot be arbitrary. They must be conditional expectations under a coherent probability measure that respects her asset-pricing model. In that case, her subjective forecasting rule, $\widetilde{\mathbb{F}}_t[\cdot]$, must be a subjective expectation operator $\widetilde{\mathbb{E}}_t[\cdot]$, with all the helpful structure that entails: linearity, the law of iterated expectations, respect for almost-sure identities.

Once the investor's forecasts are subjective expectations, the adding-up condition follows immediately. Every expectation operator is linear. So apply $\widetilde{\mathbb{F}}_t[\cdot]$ to the in-sample version of the Campbell-Shiller formula. Do the same thing with $\mathbb{E}_t[\cdot]$. Then subtract. The difference is Equation (3). The adding-up condition is a direct consequence of assuming that investors have well-formed subjective expectations. If investors' subjective beliefs are truly arbitrary, then all bets are off. There is no guarantee that Campbell-Shiller will hold.

"Model overhead" is the structure a theorist builds into his model to ensure adding-up parity. Without this structure, the investor's subjective forecasts are arbitrary point predictions. With it, they become subjective expectations and adding-up holds. The theorist supplies the structure. An investor cannot. Why would someone who understood the structure of their own mistakes well enough to enforce Equation (3) continue to make these errors?

## 3.4 Variance Decomposition

Researchers currently view the Campbell-Shiller formula as a relatively model-free way to study subjective beliefs. It is not. For Campbell-Shiller to hold, the resulting forecast errors must obey a precise adding-up condition. Beliefs can be wrong, but they cannot be arbitrarily wrong. Hence, papers that use Equation (1) are smuggling a lot of structure in through the back door. Variance decompositions offer a clear example of the problems this creates.

Campbell-Shiller variance decompositions are designed to answer the following question: Which variable is the main driver of market fluctuations? Researchers currently assume that the formula always holds: in-sample, in expectation, and under arbitrary subjective beliefs. Hence, they take it for granted that all variation in the S&P 500's log PD ratio must be due to changes in future dividend growth or returns. Variance decompositions operationalize this assumption, giving researchers a way to infer which input matters more.

Here is the logic. At each time $t$, we can observe the S&P 500's log PD ratio. Suppose we could also compute both of the infinite sums on the right-hand side of Equation (1). In that case, we could regress each infinite sum on the S&P 500's current log PD ratio. For dividend growth, the time-series regression would look like

$$\underbrace{\textstyle\sum_{h=1}^{\infty} \rho^{h-1} \cdot \Delta \log \mathrm{Div}_{t+h}}_{\text{One number at each time } t} \overset{\text{OLS}}{\sim} \alpha_{\mathrm{Div}} + \beta_{\mathrm{Div}}(\infty) \cdot \log \mathrm{PD}_t \tag{27}$$

We could run the same sort of regression for returns to get an analogous infinite-horizon slope coefficient, $\beta_{\mathrm{R}}(\infty)$.

Why is this helpful? Because both slope coefficients take the form $\frac{\mathbb{C}\mathrm{ov}[S,\, \log \mathrm{PD}]}{\mathbb{V}\mathrm{ar}[\log \mathrm{PD}]}$ where $S$ represents an infinite sum involving either dividend growth or returns. Hence, if all variation in the S&P 500's log PD ratio came from one of these two sources, we could conclude that

$$100\% = \beta_{\mathrm{Div}}(\infty) + \beta_{\mathrm{R}}(\infty) \tag{28}$$

This restriction is the key implication of a Campbell-Shiller variance decomposition. The whole approach boils down to this equation.

Notice how the logic behind this calculation differs from the way that researchers use other asset-pricing models. For example, when you deploy the CAPM, you fit the model to the data and then evaluate the size of the resulting errors. A researcher does the opposite when performing a variance decomposition. She assumes Equation (28) holds and uses it to estimate one of the two components. Violations of Campbell-Shiller do not show up as errors. They get absorbed into the parameter being estimated.

Moreover, it is not physically possible to run the time-series regression in Equation (27). Computing the dependent variable involves summing an infinite series at each point in time. Researchers use theory to get around this problem. First, they estimate a one-period analog to the coefficient of interest, $\beta_{\text{Div}}(1)$. Researchers replace the infinite sum on the left-hand side of Equation (27) with data on the S&P 500 the following year. Then, they scale up this short-term estimate using a theory-implied multiple to get $\beta_{\text{Div}}(\infty)$. All remaining variation in the S&P 500's log PD ratio gets labeled as $\beta_{\text{R}}(\infty)$.

Cochrane (2008) performs this exercise using realized data. First, he regresses next year's S&P 500 dividend growth on the index's current log PD ratio. This regression yields $\beta_{\text{Div}}(1) \approx 0.8\%$. He then estimates a persistence of $\phi \approx 0.92$, giving a multiplier of roughly 10×. The full-horizon estimate is thus

$$\beta_{\text{Div}}(\infty) \;=\; \overset{0.8\%}{\beta_{\text{Div}}(1)} \times \overbrace{\left(\frac{1}{1 - \underset{0.98}{\rho} \cdot \underset{0.92}{\phi}}\right)}^{\sim 10\times} \;\approx\; 8\% \tag{4}$$

The paper attributes the remaining 92% of variance to discount-rate changes.

De la O and Myers (2021) runs the same calculations using data from surveys. This approach yields a much larger short-term slope coefficient, $\beta_{\text{Div}}(1) \approx 39\%$. Survey respondents also think future dividend growth will be less persistent, $\phi \approx 0.60$, leading to a smaller multiplier, 2.4×. Putting the pieces together, the authors arrive at

$$\beta_{\text{Div}}(\infty) \;=\; \overset{39\%}{\beta_{\text{Div}}(1)} \times \overbrace{\left(\frac{1}{1 - \underset{0.98}{\rho} \cdot \underset{0.60}{\phi}}\right)}^{\sim 2.4\times} \;\approx\; 93\% \tag{29}$$

Apparently, discount-rate news only accounts for 7%.

Which number is correct? Both cannot be right. Yet both papers claim to be performing a model-free exercise based on an accounting identity that always holds. Cochrane (2008) describes the regression coefficients as "linked by an identity" implied by the Campbell-Shiller linearization. De la O and Myers (2021) writes that they "can calculate the variance decomposition for the price-dividend ratio without making any assumptions about the objective distribution for earnings and dividends."

These two papers are performing calculations that are structurally identical to the Gordon pricing rule. Gordon solves an infinite present-value sum by assuming constant dividend growth, which collapses the infinite sum to a more manageable calculation: a short-term estimate, $\mathbb{E}_t[\text{Div}_{t+1}]$, times a theory-implied multiple, $\left(\frac{1}{\text{R}-\text{G}}\right)$. Cochrane (2008) and De la O and Myers (2021) side-step the infinite sum required by the variance decomposition in Equation (28) by assuming AR(1) persistence in the predictor. This choice is what collapses the infinite sum to a more manageable calculation: a short-term estimate, $\beta_{\text{Div}}(1)$, times a theory-implied multiple, $\left(\frac{1}{1-\rho\cdot\phi}\right)$. Same operational problem. Same structural solution.

Campbell-Shiller is a model in all but name. It is imperative that researchers treat it as such. Calling Equation (1) an "identity" masks all the underlying assumptions. The variance-decomposition exercise in Cochrane (2008) and De la O and Myers (2021) is not model free. It comes from imposing an analogous constant-parameter restriction to the one Campbell-Shiller wanted to relax. The original result got published because everyone understood that the Gordon model was a model. The variance decompositions discussed above treat Equation (1) as a fact, making it harder to decipher what is actually going on.

# Conclusion

Our goal is not to discredit Campbell and Shiller (1988). There is nothing incorrect about the original paper. The formula does sometimes hold, and it is a useful extension of the Gordon model. The log-linear approximation is exactly the right tool for the job in situations where prices come from a time-varying Gordon model. We are making a specific claim about how the literature has come to view Campbell-Shiller.

Researchers currently believe that Equation (1) is an accounting identity that always holds no matter what. Not true. It fails in realized data whenever per-share dividends diverge from total payouts. It fails under correct expectations whenever investors do not apply present-value logic, or when they cannot estimate the cap rate to basis-point precision. And it fails under subjective beliefs whenever the resulting forecast errors do not satisfy the adding-up condition in Equation (3).

We did not write this paper to scold researchers for past mistakes. Every example and application we use comes from leading researchers that do careful work. The takeaway is not: "These people did something wrong." It is: "Wow, there is a lot more to using Campbell-Shiller than I realized."

The Gordon model does not describe how every asset gets priced. Often, the reason has nothing to do with constant parameters. A homebuyer who looks at comparable recent sales in the same neighborhood is not approximating a time-varying Gordon model. A sell-side analyst who sets a price target by multiplying his EPS forecast for next year times a trailing PE is not approximating Gordon logic, either. Both of these pricing rules anchor on a recently observed number rather than discounting future cash flows. When researchers insist that it is possible to interpret every asset's price through the lens of Campbell-Shiller, they blind themselves to the true pricing mechanism in these scenarios.

Our goal is to open up a broader range of modeling possibilities. Time-varying Gordon is such a narrow lens. Markets can be so much more interesting. Prices can anchor on comparable transactions, as in the housing market. Price targets can come from a backward-looking multiple times a short-term earnings forecast, as in sell-side equity research. Valuations can reflect payout-policy choices that Campbell-Shiller does not see. None of these possibilities are available to a researcher who treats Equation (1) as an accounting identity. To appreciate them, you have to be open to the possibility that the formula does not hold. This paper is an argument for that openness.

# References

Adam, K., A. Marcet, and J. Beutel (2017). Stock price booms and expected capital gains. American Economic Review 107(8), 2352–2408.

Adam, K., A. Marcet, and J. Nicolini (2016). Stock-market volatility and learning. Journal of Finance 71(1), 33–82.

Adam, K. and S. Nagel (2023). Expectations data in asset pricing. In Handbook of Economic Expectations, pp. 477–506. Elsevier.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-capm: An extrapolative capital asset pricing model. Journal of Financial Economics 115(1), 1–24.

Ben-David, I. and A. Chinco (2025). Expected EPS × Trailing PE: Pricing without discounting. Working paper.

Bernanke, B. and K. Kuttner (2005). What explains the stock market's reaction to Federal Reserve policy? Journal of Finance 60(3), 1221–1257.

Billingsley, P. (1995). Probability and Measure (3 ed.). Wiley.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer (2019). Diagnostic expectations and stock returns. Journal of Finance 74(6), 2839–2874.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. Journal of Finance 73(1), 199–227.

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. Journal of Finance 62(2), 877–915.

Campbell, J. (1991). A variance decomposition for stock returns. The Economic Journal 101(405), 157–179.

Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1(3), 195–228.

Case, K. E. and R. J. Shiller (1987). Prices of single-family homes since 1970: New indexes for four cities. New England Economic Review, 45–56.

Case, K. E. and R. J. Shiller (1989). The efficiency of the market for single-family homes. American Economic Review 79(1), 125–137.

Chen, L. and X. Zhao (2009). Return decomposition. Review of Financial Studies 22(12), 5213–5249.

Cho, T., L. Kremens, D. Lee, and C. Polk (2024). Scale or yield? A present-value identity. Review of Financial Studies 37(3), 950–988.

Cochrane, J. (2001). Asset Pricing. Princeton University Press.

Cochrane, J. (2011). Presidential address: Discount rates. Journal of finance 66(4), 1047–1108.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. Review of Financial Studies 21(4), 1533–1575.
Cohen, R., C. Polk, and T. Vuolteenaho (2003). The value spread. Journal of Finance 58(2), 609–641.
De la O, R. and S. Myers (2021). Subjective cash flow and discount rate expectations. Journal of Finance 76(3), 1339–1387.
De la O, R. and S. Myers (2024). Which subjective expectations explain asset prices? Review of Financial Studies 37(6), 1829–1867.
Engsted, T., T. Q. Pedersen, and C. Tanggaard (2012). The log-linear return approximation, bubbles, and predictability. Journal of Financial and Quantitative Analysis 47(3), 643–665.
Fuster, A., B. Hebert, and D. Laibson (2012). Natural expectations, macroeconomic dynamics, and asset pricing. NBER Macroeconomics Annual 26(1), 1–48.
Fuster, A., D. Laibson, and B. Mendel (2010). Natural expectations and macroeconomic fluctuations. Journal of Economic Perspectives 24(4), 67–84.
Gao, C. and I. Martin (2021). Volatility, valuation ratios, and bubbles: An empirical measure of market sentiment. Journal of Finance 76(6), 3211–3254.
Gordon, M. (1962). The investment, financing, and valuation of the corporation. RD Irwin.
Gourinchas, P. and H. Rey (2007). International financial adjustment. Journal of Political Economy 115(4), 665–703.
Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. Review of Financial Studies 27(3), 714–746.
Hartzmark, S. and A. Sussman (2025). Price agnostic demand. Working paper.
Jin, L. J. and P. Sui (2022). Asset pricing with return extrapolation. Journal of Financial Economics 145(2), 273–295.
Larrain, B. and M. Yogo (2008). Does firm value move too much to be justified by subsequent changes in cash flow? Journal of Financial Economics 87(1), 200–226.
Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56(3), 815–849.
Lettau, M., S. Ludvigson, and P. Manoel (2018). Characteristics of mutual-fund portfolios: Where are the value funds? Working paper.
Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. American Economic Review 46(2), 97–113.
Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? Quarterly Journal of Economics 126(1), 373–416.
Nagel, S. and Z. Xu (2022). Asset pricing with fading memory. Review of Financial Studies 35(5), 2190–2245.
Vuolteenaho, T. (2002). What drives firm-level stock returns? Journal of Finance 57(1), 233–264.

# Technical Appendix

*Proof.* **(Proposition 1.2)**

*By definition, the per-share dividend is* $\mathrm{Div}_{t+h} = \frac{\mathrm{TotDiv}_{t+h}}{\#\mathrm{Shares}_{t+h-1}}$*, where* $\#\mathrm{Shares}_{t+h-1}$ *is the share count at the end of period* $(t+h-1)$ *on which the* $(t+h)$*-period dividend is paid. Taking logs and first-differencing gives*

$$\Delta \log \mathrm{Div}_{t+h} \;=\; \Delta \log \mathrm{TotDiv}_{t+h} \;-\; \Delta \log \#\mathrm{Shares}_{t+h-1} \tag{30}$$

*Subtracting* $\Delta \log \mathrm{TotPmt}_{t+h}$ *from both sides gives Equation* (7)*.*

*Remark. The proposition assumes* $\mathrm{R}_{t+1}$ *remains constant. This leaves only one channel through which the per-share substitution can distort Campbell-Shiller:* $\Delta \log \mathrm{Div}_{t+1}$*. Allowing* $\mathrm{R}_{t+1}$ *to change would introduce additional distortions, making it even harder to defend the claim that Campbell-Shiller is a mere accounting identity. Holding* $\mathrm{R}_{t+1}$ *fixed thus amounts to studying the best-case scenario for the conventional wisdom.* □

*Proof.* **(Proposition 2.3)**

*Under the Gordon model in Equation* (2)*, the forward dividend yield equals the cap rate,* $\mathrm{DY} = (\mathrm{R}-\mathrm{G})$*.*

*Cap-rate decrease. Suppose noise lowers the cap rate by* $\delta > 0$*. The price moves from* $\mathrm{Price}_t = \mathbb{E}_t[\mathrm{Div}_{t+1}] \times \left(\frac{1}{\mathrm{R}-\mathrm{G}}\right)$ *to* $\mathbb{E}_t[\mathrm{Div}_{t+1}] \times \left(\frac{1}{[\mathrm{R}-\mathrm{G}]-\delta}\right)$*, so*

$$\Delta \mathrm{Price}_t \;=\; \mathbb{E}_t[\mathrm{Div}_{t+1}] \times \left(\frac{1}{\mathrm{R}-\mathrm{G}}\right) \cdot \left[\frac{\delta}{(\mathrm{R}-\mathrm{G})-\delta}\right] \tag{31}$$

*Requiring* $|\Delta \mathrm{Price}_t| > \frac{\mathbb{E}_t[\mathrm{Div}_{t+1}]}{1+\mathrm{R}}$ *and cancelling out* $\mathbb{E}_t[\mathrm{Div}_{t+1}]$ *gives*

$$\delta \;>\; \frac{(\mathrm{R}-\mathrm{G})^2}{(1+\mathrm{R})+(\mathrm{R}-\mathrm{G})} \;=\; \mathrm{DY}^2 \cdot \left(\frac{1}{1+\mathrm{R}+\mathrm{DY}}\right) \tag{32}$$

*The second equality uses* $\mathrm{DY} = (\mathrm{R}-\mathrm{G})$*.*

*Cap-rate increase. Suppose instead that the cap rate increases by* $\delta > 0$*. The analogous calculation yields*

$$\delta \;>\; \frac{(\mathrm{R}-\mathrm{G})^2}{(1+\mathrm{R})-(\mathrm{R}-\mathrm{G})} \;=\; \mathrm{DY}^2 \times \left(\frac{1}{1+\mathrm{G}}\right) \tag{33}$$

*The two thresholds differ because* $\left(\frac{1}{\mathrm{R}-\mathrm{G}}\right)$ *is convex. A drop in the cap rate moves the price by more than an equal-sized rise, so the decrease case requires a smaller* $\delta$ *to exceed a given price change.*

*Sufficient condition. Both exact thresholds in Equations* (32) *and* (33) *lie strictly below* $\mathrm{DY}^2$ *whenever* $\mathrm{G} > 0$*. Hence* $|\delta| > \mathrm{DY}^2$ *is sufficient regardless of the direction of the perturbation.*

*For the S&P 500,* $\mathrm{DY} \approx 2\%$*, so the threshold is* $\mathrm{DY}^2 \approx 0.04\%\mathrm{pt}$. □

*Proof.* **(Proposition 3.1)**

*Campbell-Shiller holds under correct expectations by assumption. Thus, replacing* $\mathbb{E}_t[\cdot]$ *with* $\widetilde{\mathbb{F}}_t[\cdot]$ *will preserve equality if and only if the following difference vanishes*

$$\sum_{h=1}^{\infty} \rho^{h-1} \cdot (\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \mathrm{Div}_{t+h}] = \sum_{h=1}^{\infty} \rho^{h-1} \cdot (\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\mathrm{R}_{t+h}] \tag{34}$$

*Rewriting* $\rho = \left(\frac{1}{1+\overline{\mathrm{DY}}}\right)$ *gives the desired result.*

*Remark. Transversality plays no role in this argument. The infinite-horizon Campbell-Shiller formula under* $\mathbb{E}_t[\cdot]$ *is taken as the starting assumption, so the transversality condition is already baked in. A separate "subjective transversality" condition on* $\widetilde{\mathbb{F}}_t[\cdot]$ *is not well-defined.* $\widetilde{\mathbb{F}}_t[\cdot]$ *need not correspond to any probability measure, so the tail expectation* $\lim_{H\to\infty} \rho^H \cdot \widetilde{\mathbb{F}}_t[\log \mathrm{PD}_{t+H}]$ *has no coherent interpretation.*

*Remark. The proposition asks when the observed* $\log \mathrm{PD}_t$ *coincides with the subjective-forecast-based right-hand side of Equation* (1)*. Both sides of the subtraction above share the same observed left-hand side. Neither* $\mathbb{E}_t[\cdot]$ *nor* $\widetilde{\mathbb{F}}_t[\cdot]$ *affect* $\log \mathrm{PD}_t = \log\left(\frac{\mathrm{Price}_t}{\mathrm{Div}_t}\right)$*. Campbell-Shiller approximates the multiple when writing Gordon as* $\mathrm{Div}_t \times \left(\frac{1+\mathrm{G}}{\mathrm{R}-\mathrm{G}}\right)$ *rather than* $\mathbb{E}_t[\mathrm{Div}_{t+1}] \times \left(\frac{1}{\mathrm{R}-\mathrm{G}}\right)$. □

*Proof.* **(Proposition 3.3)**

*A Gordon investor observes* $(\mathrm{Price}_t, \mathrm{Div}_t)$ *and forecasts a pair of constants* $(\tilde{\mathrm{R}}, \tilde{\mathrm{G}})$*. The proposition considers an investor in a more general model. Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be the probability space implied by the true data-generating process. Let* $\mathrm{data}_t$ *denote the data observed up until time* $t$*. This object is the generalized* *version of* $(\text{Price}_t, \text{Div}_t)$ *in the Gordon model. Let* $X = (X_n)_{n\geq 1}$ *be the future outcomes that the investor forecasts. Let* $\tilde{\text{f}}$ *denote the investor's subjective forecast rule, with* $\widetilde{\mathbb{F}}_t[X_n] = \tilde{\text{f}}_n(\text{data}_t)$. *The Gordon constants* $(\tilde{\text{R}}, \tilde{\text{G}})$ *become* $\tilde{\text{f}}(\text{data}_t)$.

*The Gordon model requires that* $\text{DY}_t = \tilde{\text{R}} - \tilde{\text{G}}$. *Let* $\varphi(\text{data}_t, X) = 0$ *denote the analogous system of equations in a more general model. We write the set of subjective beliefs that satisfy this system of equations given the observed data as*

$$\mathcal{X}_t^\star \;=\; \{\, x : \varphi(\text{data}_t, x) = 0 \,\} \tag{35}$$

*The proposition assumes that subjective forecasts can be cross-validated:* $\tilde{\text{f}}(\text{data}_t) \in \mathcal{X}_t^\star$ *almost surely.*

*<u>Part (a).</u> A coherent subjective probability measure* $\widetilde{\mathbb{Q}}$ *must respect the investor's asset-pricing model. That is, it must satisfy* $\widetilde{\mathbb{Q}}[\,\varphi(\text{data}_t, X) = 0\,] = 1$. *We would like to reverse-engineer the* $\widetilde{\mathbb{Q}}$ *implied by the investor's forecasting rule,* $\tilde{\text{f}}(\cdot)$. *If successful, then* $\widetilde{\mathbb{F}}_t[\cdot]$ *must be a subjective expectation operator under* $\widetilde{\mathbb{Q}}$

$$\widetilde{\mathbb{F}}_t[X] \;=\; \tilde{\text{f}}(\text{data}_t) \;=\; \mathbb{E}^{\widetilde{\mathbb{Q}}}[\, X \,|\, \text{data}_t \,] \tag{36}$$

*Construct* $\widetilde{\mathbb{Q}}$ *on* $(\Omega, \mathcal{F})$ *as follows. Set the marginal of* $\widetilde{\mathbb{Q}}$ *on* $\text{data}_t$ *equal to the marginal of* $\mathbb{P}$. *Set the conditional distribution of* $X$ *given* $\text{data}_t$ *equal to the Dirac point mass at* $\tilde{\text{f}}(\text{data}_t)$. *The regular conditional probability theorem (e.g., see Billingsley, 1995) says these two choices define a unique measure on* $(\Omega, \mathcal{F})$.

*Cross-validation places the Dirac mass inside* $\mathcal{X}_t^\star$ *almost surely. This is what makes* $\widetilde{\mathbb{Q}}$ *a valid subjective measure for the investor*

$$\widetilde{\mathbb{Q}}\big[\,\varphi(\text{data}_t, X) = 0\,\big] \;=\; 1 \tag{37}$$

*If cross-validation is not possible, then* $\widetilde{\mathbb{Q}}$ *must assign positive probability to events that the investor's model rules out. If cross-validation is guaranteed, then a coherent* $\widetilde{\mathbb{Q}}$ *must exist, and* $\widetilde{\mathbb{F}}_t[\cdot]$ *is a subjective expectation operator under it.*

*<u>Part (b).</u> The adding-up condition in Equation (3) immediately follows from Part (a). If* $\widetilde{\mathbb{F}}_t[\cdot]$ *is an expectation operator, then it is linear. Apply* $\widetilde{\mathbb{F}}_t[\cdot]$ *to in-sample Campbell-Shiller. Linearity lets* $\widetilde{\mathbb{F}}_t[\cdot]$ *distribute through the infinite sum. Apply* $\mathbb{E}_t[\cdot]$ *to the same formula and subtract. The left-hand side is the same, giving*

$$0 \;=\; \sum_{h=1}^{\infty} \rho^{h-1} \cdot \big\{\, (\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \text{Div}_{t+h}] - (\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\text{R}_{t+h}] \,\big\} \tag{38}$$

*Using* $\rho = \big(\tfrac{1}{1+\overline{\text{DY}}}\big)$ *and rearranging gives the desired result.* □

# Crimes Against Campbell-Shiller (Internet Appendix)

Itzhak Ben-David[*] and Alex Chinco[†]

April 30, 2026
[Click here for the latest version]

## A Campbell-Shiller Derivation

We now walk through the derivation of Equation (1). The starting point is the definition of a one-period realized return. Suppose you buy a share at time $t$ for $\text{Price}_t$, collect a dividend of $\text{Div}_{t+1}$, and sell at $\text{Price}_{t+1}$. Your gross return will be

$$1 + \mathrm{R}_{t+1} \;=\; \frac{\text{Price}_{t+1} + \text{Div}_{t+1}}{\text{Price}_t} \tag{6}$$

This is a definition, not an approximation. It holds for any stock in any time period, regardless of what investors believe or how they form expectations. Note that shares outstanding does not appear in this formula. The return is defined for a single share that is held from $t$ to $(t{+}1)$.

Next, multiply the right-hand side of Equation (6) by $1 = \left(\frac{1}{1}\right) = \left(\frac{\text{Div}_{t+1}/\text{Div}_{t+1}}{\text{Div}_t/\text{Div}_t}\right)$ to express the gross return in terms of the price-to-dividend ratio, $\text{PD}_t = \frac{\text{Price}_t}{\text{Div}_t}$,

$$\begin{aligned}
1 + \mathrm{R}_{t+1} &= \left(\frac{\text{Price}_{t+1} + \text{Div}_{t+1}}{\text{Price}_t}\right) \times \left(\frac{\text{Div}_{t+1}/\text{Div}_{t+1}}{\text{Div}_t/\text{Div}_t}\right) && \text{(A.1a)}\\
&= \left(\frac{\text{Price}_{t+1}/\text{Div}_{t+1} + 1}{\text{Price}_t/\text{Div}_t}\right) \times \left(\frac{\text{Div}_{t+1}}{\text{Div}_t}\right) && \text{(A.1b)}\\
&= \left(\frac{\text{PD}_{t+1} + 1}{\text{PD}_t}\right) \times \left(\frac{\text{Div}_{t+1}}{\text{Div}_t}\right) && \text{(A.1c)}
\end{aligned}$$

[*]The Ohio State University and NBER. ben-david.1@osu.edu
[†]Michigan State University. alexchinco@gmail.com

Take the natural logarithm of both sides. For the remainder of this derivation, we slightly abuse notation and use $\mathrm{R}_{t+1}$ to denote the log gross return, $\mathrm{R}_{t+1} \equiv \log\left(\frac{\mathrm{Price}_{t+1}+\mathrm{Div}_{t+1}}{\mathrm{Price}_t}\right)$, rather than the simple net return. This matches the convention used in the paper, including Equation (1). We also write $\log \mathrm{PD}_t$ for the log price-dividend ratio, and $\Delta \log \mathrm{Div}_{t+1}$ for log dividend growth. This gives

$$\mathrm{R}_{t+1} \;=\; \log(\mathrm{PD}_{t+1}{+}1) \;-\; \log \mathrm{PD}_t \;+\; \Delta \log \mathrm{Div}_{t+1} \tag{A.2}$$

So far, everything is exact. The only nonlinearity is the $\log(\mathrm{PD}_{t+1}{+}1)$ term, which is a function of the future PD ratio.

The key step in the Campbell-Shiller derivation is to replace $\log(\mathrm{PD}_{t+1}{+}1)$ with a first-order Taylor approximation around $\overline{\log \mathrm{PD}} = \mathbb{E}[\log \mathrm{PD}_t]$, the mean log PD ratio. If we define $\overline{\mathrm{PD}} = e^{\overline{\log \mathrm{PD}}}$ and $\overline{\mathrm{DY}} = 1/\overline{\mathrm{PD}}$, then the expansion gives

$$\log(\mathrm{PD}_{t+1}{+}1) \;\approx\; \log(\overline{\mathrm{PD}}{+}1) \;+\; \left(\frac{1}{1+\overline{\mathrm{DY}}}\right) \times (\log \mathrm{PD}_{t+1} - \overline{\log \mathrm{PD}}) \tag{A.3}$$

It is helpful to tidy things up by introducing two new constants

$$\rho = \frac{1}{1+\overline{\mathrm{DY}}} \qquad\qquad \kappa = \log(\overline{\mathrm{PD}}{+}1) - \rho \cdot \overline{\log \mathrm{PD}} \tag{A.4}$$

For the aggregate US stock market, the S&P 500's average dividend yield of $\overline{\mathrm{DY}} \approx 2\%$ implies $\rho \approx 0.98$ at an annual frequency.

The next step is to substitute Equations (A.3) and (A.4) into Equation (A.2) to arrive at the one-period version of the Campbell-Shiller formula

$$\mathrm{R}_{t+1} \;\approx\; \kappa \;+\; \Delta \log \mathrm{Div}_{t+1} \;+\; \rho \cdot \log \mathrm{PD}_{t+1} \;-\; \log \mathrm{PD}_t \tag{A.5}$$

Equivalently, if we put the current log PD ratio on the left-hand side, we get

$$\log \mathrm{PD}_t \;\approx\; \kappa \;+\; \Delta \log \mathrm{Div}_{t+1} \;-\; \mathrm{R}_{t+1} \;+\; \rho \cdot \log \mathrm{PD}_{t+1} \tag{A.6}$$

Today's log PD ratio is approximately equal to next year's dividend growth, minus next year's return, plus a discounted version of next year's log PD ratio. The power of Equation (A.6) is that it can be applied recursively.

We now face a problem: the stock's current log PD ratio is defined in terms of its future log PD ratio. We need to get rid of the future multiple. Campbell-Shiller does this by replacing $\log \mathrm{PD}_{t+1}$ on the right-hand side with its own one-period expansion, $\log \mathrm{PD}_{t+1} \approx \kappa + \Delta \log \mathrm{Div}_{t+2} - \mathrm{R}_{t+2} + \rho \cdot \log \mathrm{PD}_{t+2}$. Doing this gives

$$\begin{aligned} \log \mathrm{PD}_t \;\approx\; \kappa \cdot (1{+}\rho) \;+\; \{\Delta \log \mathrm{Div}_{t+1} - \mathrm{R}_{t+1}\} \\ +\; \rho \cdot \{\Delta \log \mathrm{Div}_{t+2} - \mathrm{R}_{t+2}\} \;+\; \rho^2 \cdot \log \mathrm{PD}_{t+2} \end{aligned} \tag{A.7}$$

Repeating the same forward-substitution procedure $H$ times gives something that looks a lot like Equation (1) from the introduction

$$\log \mathrm{PD}_t \;\approx\; -\sum_{h=1}^{H} \rho^{h-1} \cdot \left\{ \mathrm{R}_{t+h} - \Delta \log \mathrm{Div}_{t+h} \right\} \;+\; \rho^{H} \cdot \log \mathrm{PD}_{t+H} \tag{A.8}$$

Note that we have suppressed the leading $\kappa \cdot \left(\frac{1-\rho^H}{1-\rho}\right)$ term for simplicity's sake.

To get rid of the $\rho^H \cdot \log \mathrm{PD}_{t+H}$ term, we take the conditional expectation of both sides of Equation (A.8). Since $\log \mathrm{PD}_t$ is known at time $t$, the left-hand side does not change. On the right-hand side, the objectively correct conditional expectation passes through the sum

$$\begin{aligned} \log \mathrm{PD}_t \;\approx\; -\sum_{h=1}^{H} \rho^{h-1} \cdot \left\{ \mathbb{E}_t[\mathrm{R}_{t+h}] - \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] \right\} \\ +\; \rho^{H} \cdot \mathbb{E}_t[\log \mathrm{PD}_{t+H}] \end{aligned} \tag{A.9}$$

We need the terminal term, $\rho^H \cdot \mathbb{E}_t[\log \mathrm{PD}_{t+H}]$, to vanish as $H \longrightarrow \infty$. This is known as a transversality condition

$$\lim_{H \to \infty} \rho^H \cdot \mathbb{E}_t[\log \mathrm{PD}_{t+H}] \;=\; 0 \tag{A.10}$$

It says that the expected log PD ratio does not grow so fast that its discounted value explodes.

Under this assumption, we arrive at the forward-looking version of Campbell-Shiller that we saw in the first paragraph of the introduction

$$\log \mathrm{PD}_t \;\approx\; -\sum_{h=1}^{\infty} \rho^{h-1} \cdot \left\{ \mathbb{E}_t[\mathrm{R}_{t+h}] - \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] \right\} \tag{1}$$

It says that a stock's log PD ratio is approximately equal to a constant plus the discounted sum of expected future dividend growth net of returns.

One subtlety is worth noting. The Gordon multiple in Equation (2) is the forward PD ratio, $\left(\frac{1}{\mathrm{R}-\mathrm{G}}\right) = \frac{\mathrm{Price}_t}{\mathbb{E}_t[\mathrm{Div}_{t+1}]}$. The left-hand side of Equation (1) is the trailing PD ratio, $\log \mathrm{PD}_t = \log\left(\frac{\mathrm{Price}_t}{\mathrm{Div}_t}\right)$. The direct analog in Gordon's constant-parameter world would be

$$\mathrm{PD} \;=\; \frac{\mathrm{Price}_t}{\mathrm{Div}_t} \;=\; \left(\frac{1+\mathrm{G}}{\mathrm{R}-\mathrm{G}}\right) \tag{A.11}$$

so $\log \mathrm{PD} = \log(1+\mathrm{G}) - \log(\mathrm{R}-\mathrm{G})$.

In Gordon's never-changing world, both of these terms are numerical constants. The Campbell-Shiller approximation linearizes around this steady state. The sum on the right-hand side of Equation (1) captures the dynamic variation in the $-\log(R-G)$ component as R and G fluctuate over time. The leading constant $\kappa$ absorbs the steady-state levels of both the $\log(1+G)$ term (the one-period growth between the trailing and forward dividend) and the $-\log(R-G)$ term (the Gordon multiplier).

# B Repurchases and New Issuance

This appendix provides the full calculations behind the examples in Sections 1.3 and 1.4. Throughout, per-share quantities are defined as $\text{Div}_{t+1} = \frac{\text{Dividend Amount}_{t+1}}{\#\text{Shares}_t}$ and $\text{Price}_{t+1} = \frac{\text{Equity Value}_{t+1}}{\#\text{Shares}_t}$, where $\#\text{Shares}_t$ is the share count at the end of period $t$. Any repurchase or issuance that takes place during period $t$ is already reflected in $\#\text{Shares}_t$.

## B.1 Repurchases

Consider three firms that are identical at time $(t-1)$. Each has $\#\text{Shares}_{t-1}$ = 100M shares outstanding and a share price of $\text{Price}_{t-1}$ = \$3.125/sh. Each period, every firm earns \$25M and distributes all of it to shareholders. All three firms have the same 8% discount rate, and since there is no investment, the ex-dividend firm value is a flat perpetuity: $\$25\text{M} \times \left(\frac{1}{8\%}\right) = \$312.5\text{M}$ every period. The three firms differ only in how they split the \$25M payout between dividends and repurchases, and in whether this split has changed.

**Baseline Inc.** This company always pays \$25M in dividends and repurchases \$0 (100/0 split).

At time $t$, dividends are paid on $\#\text{Shares}_{t-1}$ = 100M shares

$$\text{Div}_t = \frac{\$25\text{M}}{100\text{M}} = \$0.25/\text{sh} \tag{B.1a}$$

$$\text{Price}_t = \frac{\$312.5\text{M}}{100\text{M}} = \$3.125/\text{sh} \tag{B.1b}$$

No shares are repurchased, so $\#\text{Shares}_t$ = 100M.

At time $(t+1)$, the numbers are identical, $\text{Div}_{t+1}$ = \$0.25/sh and $\text{Price}_{t+1}$ = \$3.125/sh. The Campbell-Shiller inputs are

$$\text{PD}_t = \frac{\$3.125}{\$0.25} = 12.5\times \tag{B.2a}$$

$$\text{R}_{t+1} = \frac{\$3.125 + \$0.25}{\$3.125} - 1 = 8\% \tag{B.2b}$$

$$\Delta \log \text{Div}_{t+1} = 0\% \tag{B.2c}$$

**Always Co.** This company maintains a steady 80/20 split, paying \$20M in dividends and repurchasing \$5M every year.

At time $t$, dividends are paid on $\#\text{Shares}_{t-1} = 100\text{M}$ shares

$$\text{Div}_t = \frac{\$20\text{M}}{100\text{M}} = \$0.20/\text{sh} \tag{B.3}$$

The \$5M repurchase happens at the ex-dividend price. After dividends, the firm holds \$5M in cash earmarked for repurchase, so the ex-dividend equity value is \$312.5M + \$5M = \$317.5M across 100M shares

$$\text{Price}_t = \frac{\$317.5\text{M}}{100\text{M}} = \$3.175/\text{sh} \tag{B.4}$$

The firm buys back $\frac{\$5\text{M}}{\$3.175/\text{sh}} \approx 1.575\text{M}$ shares, leaving $\#\text{Shares}_t \approx 98.43\text{M}$. This is consistent with Modigliani-Miller. The repurchase does not change the price per share; the market cap mechanically drops by the \$5M paid out, from \$317.5M (pre-repurchase) to $98.43\text{M} \times \$3.175/\text{sh} = \$312.5\text{M}$ (post-repurchase).

At time $(t+1)$, the firm earns another \$25M and again splits it 80/20. Dividends are now paid on $\#\text{Shares}_t \approx 98.43\text{M}$ shares

$$\text{Div}_{t+1} = \frac{\$20\text{M}}{98.43\text{M}} \approx \$0.2032/\text{sh} \tag{B.5a}$$

$$\text{Price}_{t+1} = \frac{\$317.5\text{M}}{98.43\text{M}} \approx \$3.226/\text{sh} \tag{B.5b}$$

The Campbell-Shiller inputs are

$$\text{PD}_t = \frac{\$3.175}{\$0.20} = 15.9\times \tag{B.6a}$$

$$\text{R}_{t+1} = \frac{\$3.226 + \$0.2032}{\$3.175} - 1 = 8\% \tag{B.6b}$$

$$\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.2032}{\$0.20}\right) \approx +1.6\% \tag{B.6c}$$

The dividend per share grows by 1.6%, not because the firm's fundamentals improved, but because the shrinking share count mechanically boosts the per-share dividend. The PD ratio is 15.9× versus Baseline Inc's 12.5×, a 27% relative difference caused entirely by the lower per-share dividend.

**Switch Ltd.** This company switches payout policies. At time $t$, Switch Ltd pays all \$25M as dividends and repurchases nothing (100/0 split). At time $(t+1)$, the firm switches to an 80/20 split.

Its time-$t$ numbers match Baseline Inc exactly

$$\text{Div}_t = \frac{\$25\text{M}}{100\text{M}} = \$0.25/\text{sh} \tag{B.7a}$$

$$\text{Price}_t = \frac{\$312.5\text{M}}{100\text{M}} = \$3.125/\text{sh} \tag{B.7b}$$

At time $(t+1)$, the firm pays \$20M in dividends on $\#\text{Shares}_t$ = 100M shares and spends \$5M on repurchases

$$\text{Div}_{t+1} = \frac{\$20\text{M}}{100\text{M}} = \$0.20/\text{sh} \tag{B.8a}$$

$$\text{Price}_{t+1} = \frac{\$317.5\text{M}}{100\text{M}} = \$3.175/\text{sh} \tag{B.8b}$$

The Campbell-Shiller inputs are

$$\text{PD}_t = \frac{\$3.125}{\$0.25} = 12.5\times \tag{B.9a}$$

$$\text{R}_{t+1} = \frac{\$3.175 + \$0.20}{\$3.125} - 1 = 8\% \tag{B.9b}$$

$$\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.20}{\$0.25}\right) = -22.3\% \tag{B.9c}$$

Baseline Inc and Switch Ltd are indistinguishable at time $t$. They have the same $\text{PD}_t = 12.5\times$, the same $\text{Div}_t$ = \$0.25/sh, the same $\text{Price}_t$ = \$3.125/sh, and the same return $\text{R}_t = 8\%$. A forecasting model estimated on per-share data must produce the same forecast for both firms. Yet one delivers $\Delta \log \text{Div}_{t+1} = 0\%$ and the other delivers $\Delta \log \text{Div}_{t+1} = -22.3\%$.

## B.2 New Issuance

The repurchase example isolated payout composition. This example isolates share dilution. All payouts are dividends, so the composition channel drops out.

The two firms in this example, Now Corp and Delay PLC, look identical to Baseline Inc at the start of period $t$. Each starts out trading at $\text{Price}_{t-1}$ = \$3.125/sh on $\#\text{Shares}_{t-1}$ = 100M shares. But, during period $t$, Now Corp and Delay PLC both invest \$78.125M in a project that will generate \$9.375M per year in additional earnings starting in period $(t+1)$. Both firms fund the project by issuing 25M new shares at \$3.125/sh. By the end of period $t$, each has $\#\text{Shares}_t$ = 125M shares. The two firms differ only in how quickly they raise dividends.

**Now Corp.** This company immediately raises its total dividend to \$25M+\$9.375M = \$34.375M. Its ex-dividend value is \$34.375M $\times \left(\frac{1}{8\%}\right)$ = \$429.6875M. With $\#\text{Shares}_t$ = 125M shares

$$\text{Div}_{t+1} = \frac{\$34.375\text{M}}{125\text{M}} = \$0.275/\text{sh} \tag{B.10a}$$

$$\text{Price}_{t+1} = \frac{\$429.6875\text{M}}{125\text{M}} = \$3.4375/\text{sh} \tag{B.10b}$$

The Campbell-Shiller inputs are

$$\text{PD}_t = \frac{\$3.125}{\$0.25} = 12.5\times \tag{B.11a}$$

$$\text{R}_{t+1} = \frac{\$3.4375 + \$0.275}{\$3.125} - 1 = 18.8\% \tag{B.11b}$$

$$\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.275}{\$0.25}\right) = +9.5\% \tag{B.11c}$$

**Delay PLC.** This firm issues the same shares and invests in the same project as Now Corp. However, Delay PLC keeps its total dividend at \$25M at time $(t+1)$ and retains \$9.375M in earnings. The ex-dividend value consists of operating assets, \$429.6875M, plus retained cash, \$9.375M. With $\#\text{Shares}_t$ = 125M shares:

$$\text{Div}_{t+1} = \frac{\$25\text{M}}{125\text{M}} = \$0.20/\text{sh} \tag{B.12a}$$

$$\text{Price}_{t+1} = \frac{\$439.0625\text{M}}{125\text{M}} = \$3.5125/\text{sh} \tag{B.12b}$$

The cum-dividend value per share is \$0.20+\$3.5125 = \$3.7125/sh, identical to Now Corp. The market correctly prices the same project regardless of the board's dividend decision. The Campbell-Shiller inputs are

$$\text{PD}_t = \frac{\$3.125}{\$0.25} = 12.5\times \tag{B.13a}$$

$$\text{R}_{t+1} = \frac{\$3.5125 + \$0.20}{\$3.125} - 1 = 18.8\% \tag{B.13b}$$

$$\Delta \log \text{Div}_{t+1} = \log\left(\frac{\$0.20}{\$0.25}\right) = -22.3\% \tag{B.13c}$$

Now Corp and Delay PLC issued the same number of shares, invested in the same project, earned the same total earnings, and delivered the same 18.8% return to existing shareholders. The only difference is a board-level decision about when to raise the dividend. Yet the Campbell-Shiller formula registers one as a +9.5% positive cash-flow shock and the other as a −22.3% negative cash-flow shock, a 32%pt swing from dividend timing alone. Firms typically adjust their dividend slowly (Lintner, 1956). In practice, a firm that issues equity to fund a positive-NPV project is far more likely to resemble Delay PLC.

# C Forecasts vs. Expectations

This appendix works through a single example to make the distinction between $\widetilde{\mathbb{F}}_t[\cdot]$ (arbitrary subjective forecast), $\widetilde{\mathbb{E}}_t[\cdot]$ (coherent subjective expectation), and $\mathbb{E}_t[\cdot]$ (correct expectation) concrete. The goal is to show what researchers rule out when they move from $\widetilde{\mathbb{F}}_t[\cdot]$ to $\widetilde{\mathbb{E}}_t[\cdot]$.

## C.1 Asset-Pricing Model

Consider a stock where the Gordon model holds under correct expectations. Investors can see the current share price and the per-share dividend over the past twelve months

$$\begin{aligned} \mathrm{Price}_t &= \$100/\mathrm{sh} && \text{(C.1a)} \\ \mathrm{Div}_t &= \$2/\mathrm{sh} && \text{(C.1b)} \end{aligned}$$

Under correct beliefs, the annual return and dividend-growth rate are both constant

$$\begin{aligned} \mathrm{R} &= 8\% && \text{(C.2a)} \\ \mathrm{G} &= 6\% && \text{(C.2b)} \end{aligned}$$

$\mathbb{E}_t[\mathrm{R}_{t+h}] = 8\%$ and $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] = 6\%$ for all $h$.

Let $\tilde{\mathrm{R}}$ and $\tilde{\mathrm{G}}$ denote subjective beliefs about future returns and dividend growth. To be consistent with the observed share price and dividend given the Gordon model, these two parameters must satisfy

$$\mathrm{PD}_t = \frac{\mathrm{Price}_t}{\mathrm{Div}_t} = \left(\frac{1+\tilde{\mathrm{G}}}{\tilde{\mathrm{R}}-\tilde{\mathrm{G}}}\right) \tag{C.3}$$

This is the exact constraint.

As noted in Section A, the $(1+\tilde{\mathrm{G}})$ term in the numerator gets absorbed into the constant $\kappa$ in the Campbell-Shiller approximation. What matters for the log-linear formula is $\left(\frac{1}{\tilde{\mathrm{R}}-\tilde{\mathrm{G}}}\right)$, which corresponds to the forward PD ratio in the Gordon model. In the CS approximation, the constraint therefore simplifies to

$$\tilde{\mathrm{R}} - \tilde{\mathrm{G}} = \mathrm{DY} = 2\% \tag{C.4}$$

We work with this approximate constraint below. It is the same approximation that underlies the Campbell-Shiller formula and gets used in Proposition 3.1.

## C.2 Subjective Expectations

Eve is an investor who views the world through the lens of the Gordon model. Moreover, her subjective beliefs $\tilde{\text{R}}$ and $\tilde{\text{G}}$ stem from a subjective probability measure, $\widetilde{\mathbb{Q}}$, which is consistent with her model.

Eve's primitives must satisfy Equation (C.4): $\tilde{\text{R}}-\tilde{\text{G}} = \text{DY} = 2\%$. She has one degree of freedom, not two. Suppose Eve picks $\tilde{\text{R}} = 9\%$. Then $\tilde{\text{G}} = 7\%$. Not 6%, not 8%. Exactly 7%. Her dividend-growth belief is pinned down by her return belief and the observed price.

Eve's errors relative to the truth are

$$\tilde{\text{R}} - \text{R} \;=\; +1\%\text{pt} \tag{C.5a}$$
$$\tilde{\text{G}} - \text{G} \;=\; +1\%\text{pt} \tag{C.5b}$$

The two errors are of equal magnitude. Not because Eve chose it that way. Instead, her model's structure, $\tilde{\text{R}}-\tilde{\text{G}} = 2\%$, forces her beliefs in lockstep with the true relationship $\text{R}-\text{G} = 2\%$. A single primitive choice, $\tilde{\text{R}} = 9\%$, pinned down both errors.

The adding-up condition holds automatically. Under Eve's constant parameters, $\widetilde{\mathbb{E}}_t[\text{R}_{t+h}] - \mathbb{E}_t[\text{R}_{t+h}] = +1\%\text{pt}$ and $\widetilde{\mathbb{E}}_t[\Delta \log \text{Div}_{t+h}] - \mathbb{E}_t[\Delta \log \text{Div}_{t+h}] = +1\%\text{pt}$ at every horizon. The two sides of Equation (3) become

$$\text{LHS} \;=\; \sum_{h=1}^{\infty} \frac{0.01}{(1.02)^{h-1}} \;=\; 0.01 \times \left(\frac{1.02}{0.02}\right) \;=\; 0.51 \tag{C.6a}$$
$$\text{RHS} \;=\; \sum_{h=1}^{\infty} \frac{0.01}{(1.02)^{h-1}} \;=\; 0.51 \tag{C.6b}$$

No calibration was required. All the work was done by the assumption that $\widetilde{\mathbb{Q}}$ represents a coherent probability measure given Eve's asset-pricing model. Eve is allowed to be wrong about the true parameters. She is not allowed to be wrong in a way that is internally inconsistent. That second restriction is what delivers consistency with Campbell-Shiller.

## C.3 Subjective Forecasts

Now drop the restriction that beliefs come from a coherent probability measure. Freddy's beliefs are described by $\widetilde{\mathbb{F}}_t[\cdot]$, an arbitrary mapping from random variables to real numbers. Freddy has no $\widetilde{\mathbb{Q}}$. He might use Gordon to assign prices to assets, but it is a one-way street. The formula does not constrain his beliefs. Freddy does not cross-validate his views about $\tilde{R}$ and $\tilde{G}$ given observed prices.

Suppose Freddy also has return forecasts that are +1%pt too high at every horizon, $\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] = 9\%$ for all $h$. What does this imply about Freddy's views about future dividend growth, $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}]$? Nothing. The argument that worked for Eve relied on a pricing equation inside her model. $\widetilde{\mathbb{F}}_t[\cdot]$ has no such equation. Freddy does not update his beliefs after observing $\mathrm{Price}_t = \$100/\mathrm{sh}$ and $\mathrm{Div}_t = \$2/\mathrm{sh}$. As a result, Freddy's subjective beliefs are free to violate the adding-up condition in Proposition 3.1.

Consider five different examples where Freddy always makes the same perpetual return-forecasting error, $(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\mathrm{R}_{t+h}] = +1\%\mathrm{pt}$ for all $h$:

1. $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] = 7\%$ for all $h$. Freddy picks the same dividend-growth forecast as Eve. The left-hand side equals 0.51. Adding-up is satisfied.
2. $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] = 6\%$ for all $h$ (the correct forecast). The left-hand side equals 0. Adding-up fails by 0.51.
3. $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] = 8\%$ for all $h$. The left-hand side equals $\sum_h \frac{0.02}{(1.02)^{h-1}} = 1.02$. Adding-up fails by 0.51 in the other direction.
4. $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+1}] = 9\%$ and $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] = 6\%$ for $h \geq 2$. The left-hand side equals 0.03. Adding-up fails by 0.48.
5. $\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] = 6\% + \varepsilon_h$ with $\varepsilon_h$ IID mean-zero. The left-hand side is a random variable whose probability of equaling exactly 0.51 is zero.

Only Scenario 1 satisfies the adding-up condition in Proposition 3.1. It is a knife-edge case. Freddy happened to make the exact dividend-growth forecasting error that would offset his perpetual return-forecasting error. Nothing in $\widetilde{\mathbb{F}}_t[\cdot]$ told him to. He got lucky. The remaining examples show how things could have been different. They are all equally valid subjective beliefs that Freddy could hold. Subjective beliefs that satisfy Campbell-Shiller are the exception, not the rule. They are a special case, not the universe.

## C.4 Key Takeaway

When researchers use $\widetilde{\mathbb{E}}_t[\cdot]$ in place of $\widetilde{\mathbb{F}}_t[\cdot]$, they restrict attention to the first scenario above. The Gordon model is unusually simple: two primitives, one pricing equation, one constraint from the observed price. Choosing $\tilde{R}$ collapses the last degree of freedom onto $\tilde{G}$. Richer models with time-varying parameters multiply both the primitives and the constraints. Making a belief-biased investor's $\widetilde{\mathbb{Q}}$ coherent in these settings becomes a whack-a-mole exercise. Every fix in the belief-based asset-pricing literature is tailored to its specific model-plus-bias combination because the problems are different in each one.

# D Belief Specifications

This appendix provides the full derivations behind the three case studies in Section 3.2. Throughout, the adding-up condition from Proposition 3.1 requires

$$\underbrace{\sum_{h=1}^{\infty} \frac{(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \mathrm{Div}_{t+h}]}{(1+\overline{\mathrm{DY}})^{h-1}}}_{\mathrm{LHS}} = \underbrace{\sum_{h=1}^{\infty} \frac{(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\mathrm{R}_{t+h}]}{(1+\overline{\mathrm{DY}})^{h-1}}}_{\mathrm{RHS}} \tag{3}$$

Both sides are present values of forecast errors, discounted at the stock's long-run average dividend-yield, $\overline{\mathrm{DY}}$, rather than its risk-adjusted R.

The following geometric sum appears in every case below

$$\sum_{h=1}^{\infty} \frac{1}{(1+\overline{\mathrm{DY}})^{h-1}} = \frac{1+\overline{\mathrm{DY}}}{\overline{\mathrm{DY}}} \approx \frac{1}{\overline{\mathrm{DY}}} \quad \text{for small } \overline{\mathrm{DY}} \tag{D.1}$$

The main text uses the approximation, $\frac{1}{\overline{\mathrm{DY}}}$. We will analyze both forms in the derivations below, noting where the approximation differs.

## D.1 Perpetually Wrong

Suppose investors have correct dividend-growth forecasts but overestimate future returns by $\delta > 0$ at each horizon $h \geq 1$. The forecast errors are

$$(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\mathrm{R}_{t+h}] = \delta \quad \text{for all } h \tag{D.2a}$$

$$(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \mathrm{Div}_{t+h}] = 0 \quad \text{for all } h \tag{D.2b}$$

The left-hand side of the adding-up condition in Equation (3) is zero. The right-hand side is the geometric sum

$$\sum_{h=1}^{\infty} \frac{\delta}{(1+\overline{\mathrm{DY}})^{h-1}} = \delta \times \left(\frac{1+\overline{\mathrm{DY}}}{\overline{\mathrm{DY}}}\right) \approx \delta \times \left(\frac{1}{\overline{\mathrm{DY}}}\right) \tag{18}$$

A 1%pt error on returns looks trivial at any single horizon. But when capitalized at the long-run average dividend yield, it produces a sizeable violation.

For $\overline{\mathrm{DY}} \approx 2\%$, the approximate multiplier is $\left(\frac{1}{2\%}\right) = 50\times$. A $\delta = 1$%pt perpetual mistake about future returns translates to an adding-up violation of

$$\text{Violation} = +1\%\text{pt} \times \left(\tfrac{1}{\overline{\mathrm{DY}}}\right) \approx 0.5 \text{ log units} \tag{D.3}$$

Investors' beliefs imply a PD ratio that is $1-e^{-0.50} \approx 39\%$ lower than the one in the data. The exact multiplier $(\frac{1+2\%}{2\%}) = 51$ gives a violation of 0.51 log units, corresponding to a PD mismatch of $1-e^{-0.51} \approx 40\%$. The two versions agree to within a percentage point or two.

The logic mirrors the Gordon pricing rule. The adding-up condition in Equation (3) has the same infinite-discounted-sum structure as the dividend-discount model, and the Gordon model is a special case of the DDM. If prices come from a Gordon model, then the multiplier $(\frac{1}{\overline{\mathrm{DY}}}) = (\frac{1}{\mathrm{R-G}})$. A stock's current price is $\mathbb{E}_t[\mathrm{Div}_{t+1}] \times (\frac{1}{\mathrm{R-G}})$, with $(\mathrm{R-G}) = \overline{\mathrm{DY}}$ in steady state. A perpetual $\delta$ error on future returns is capitalized at exactly the rate Gordon uses to capitalize cash flows. That is why a small bias turns into a large inconsistency: $\overline{\mathrm{DY}}$ is small.

## D.2 Moving Target

Again, assume investors' dividend-growth forecasts are correct, but they overestimate future returns at every horizon. The initial $\delta > 0$ overestimate about next year's return now decays geometrically with persistence $\phi \in [0, 1)$

$$\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] \;=\; \mathbb{E}_t[\mathrm{R}_{t+h}] \;+\; \phi^{h-1} \cdot \delta \qquad \text{for all } h = 1, 2, 3, \ldots \tag{19}$$

The factor of $\phi^{h-1}$ reflects AR(1) dynamics with persistence $\phi$.

The left-hand side of Equation (3) is again zero. The right-hand side is a slightly different geometric sum

$$\sum_{h=1}^{\infty} \frac{\phi^{h-1} \cdot \delta}{(1+\overline{\mathrm{DY}})^{h-1}} \;=\; \delta \times \left(\frac{1}{1-\rho\cdot\phi}\right) \tag{20}$$

The $(\frac{1}{1-\rho\cdot\phi})$ multiplier recurs throughout the literature. For small $\overline{\mathrm{DY}}$, the same quantity is approximately

$$\frac{1}{1-\rho\cdot\phi} \;\approx\; \frac{1}{\overline{\mathrm{DY}} + (1-\phi)} \tag{D.4}$$

This is the Gordon multiple, $\frac{1}{\overline{\mathrm{DY}}} = (\frac{1}{\mathrm{R-G}})$, modified by a mean-reversion term $(1-\phi)$ in the denominator.

Three limiting cases make the formula concrete:

- Perpetual error ($\phi = 1$). The correction vanishes, $\frac{1}{\overline{\mathrm{DY}}+(1-1)} = \frac{1}{\overline{\mathrm{DY}}}$. We recover the perpetually wrong case from above. With $\overline{\mathrm{DY}} \approx 2\%$ and $\delta = 1\%$pt, the violation is 0.5 log units, corresponding to a PD mismatch of $1-e^{-0.5} \approx 39\%$.

- One-off error ($\phi = 0$). The denominator is $\overline{\text{DY}}+1 \approx 1$, so the multiplier is approximately 1. A $\delta = 1$%pt error produces a violation of 0.01 log units, corresponding to a PD mismatch of $1-e^{-0.01} \approx 1\%$. No amplification.

- Interior case ($\phi = 0.6$). With $\overline{\text{DY}} = 2\%$, the denominator is $2\%+40\% = 42\%$, giving a multiplier of ~2.4×. A $\delta = 1$%pt error produces a violation of 0.024 log units, corresponding to a PD mismatch of $1-e^{-0.024} \approx 2.4\%$. Much smaller than the $\phi = 1$ case, but still an order of magnitude larger than the $\phi = 0$ case. Small changes in persistence matter a lot.

The numbers above use the small-$\overline{\text{DY}}$ approximation $\left(\frac{1}{1-\rho\cdot\phi}\right) \approx \frac{1}{\overline{\text{DY}}+(1-\phi)}$. The exact multiplier $\left(\frac{1}{1-\rho\cdot\phi}\right)$ is slightly larger in each case. At $\overline{\text{DY}} = 2\%$, the exact and approximate multipliers are 51× vs. 50× at $\phi = 1$, 1.00× vs. 0.98× at $\phi = 0$, and 2.43× vs. 2.38× at $\phi = 0.6$. The two versions agree to within a percentage point of PD mismatch.

## D.3 Measurement Error

Both of the preceding examples involved biased beliefs. The same gap between $\widetilde{\mathbb{F}}_t[\cdot]$ and $\mathbb{E}_t[\cdot]$ can also arise from estimation error. No one knows $\mathbb{E}_t[\Delta \log \text{Div}_{t+h}]$ and $\mathbb{E}_t[\text{R}_{t+h}]$ with infinite precision. Suppose a researcher plugs her best guesses into Equation (1). In an ideal world, her estimation errors on all inputs would be IID white noise. This is an unrealistically optimistic view of measurement error. In general, mistakes will not be mutually independent across horizons or between variables. We focus on the IID case below to make the point in its cleanest form.

Suppose the per-horizon errors are

$$(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \text{Div}_{t+h}] = \varepsilon_h^{\text{Div}} \overset{\text{IID}}{\sim} \text{Normal}(0, \sigma^2_{\text{Div}}) \tag{D.5a}$$

$$(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\text{R}_{t+h}] = \varepsilon_h^{\text{R}} \overset{\text{IID}}{\sim} \text{Normal}(0, \sigma^2_{\text{R}}) \tag{D.5b}$$

Define the gap between the two sides of Equation (3) as

$$\text{Gap} = \text{LHS} - \text{RHS} \tag{D.6a}$$

$$= \sum_{h=1}^{\infty} \frac{\varepsilon_h^{\text{Div}} - \varepsilon_h^{\text{R}}}{(1+\overline{\text{DY}})^{h-1}} \tag{D.6b}$$

Adding-up holds if and only if Gap = 0.

Each term in the sum is a weighted difference of independent mean-zero random variables, so the gap is zero on average

$$\mathbb{E}[\text{Gap}] \;=\; \sum_{h=1}^{\infty} \frac{\mathbb{E}[\varepsilon_h^{\text{Div}}] - \mathbb{E}[\varepsilon_h^{\text{R}}]}{(1+\overline{\text{DY}})^{h-1}} \;=\; 0 \tag{D.7}$$

However, volatility is a different story. The size of the adding-up gap is a continuous random variable with positive standard deviation, so Equation (3) is violated almost surely, $\Pr[\text{Gap} = 0] = 0$

$$\mathbb{S}\text{d}[\text{Gap}] \;=\; \left( \sum_{h=1}^{\infty} \frac{\sigma_{\text{Div}}^2 + \sigma_{\text{R}}^2}{(1+\overline{\text{DY}})^{2\cdot(h-1)}} \right)^{0.5} \tag{D.8a}$$

$$= \sqrt{\sigma_{\text{Div}}^2 + \sigma_{\text{R}}^2} \times \left( \frac{(1+\overline{\text{DY}})^2}{(1+\overline{\text{DY}})^2 - 1} \right)^{0.5} \tag{D.8b}$$

$$= \sqrt{\sigma_{\text{Div}}^2 + \sigma_{\text{R}}^2} \times \left( \frac{1}{1-\rho^2} \right)^{0.5} \tag{21}$$

For a numerical anchor, take $\overline{\text{DY}} \approx 2\%$. This implies $\rho \approx 0.980$ and $\sqrt{\frac{1}{1-\rho^2}} \approx 5$. Suppose researchers could predict future values of R and $\Delta \log \text{Div}$ to within $\pm 1$%pt each year. In that case, we would have $\sigma_{\text{Div}} = \sigma_{\text{R}} = 1$%pt and $\sigma_{\text{Div}}^2 + \sigma_{\text{R}}^2 = 0.0002$. The adding-up gap would then have a volatility of

$$\mathbb{S}\text{d}[\text{Gap}] \;\approx\; \sqrt{0.0002} \times 5 \tag{D.9a}$$

$$= \; 0.071 \text{ log units} \tag{D.9b}$$

These numerical inputs yield a PD mismatch of $e^{0.072}-1 \approx 7\%$.

Researchers do not know R and $\Delta \log \text{Div}$ to anything close to that level of accuracy. Standard estimates of the equity risk premium range from 4% to 8%. Take $\sigma_{\text{Div}} = \sigma_{\text{R}} = 4$%pt to match this level of uncertainty. Then $\sigma_{\text{Div}}^2 + \sigma_{\text{R}}^2 = 0.0032$ and

$$\mathbb{S}\text{d}[\text{Gap}] \;\approx\; \sqrt{0.0032} \times 5 \tag{D.10a}$$

$$= \; 0.283 \text{ log units} \tag{D.10b}$$

More realistic numbers produce a PD mismatch of $e^{0.283}-1 \approx 33\%$. This last calculation rhymes with Proposition 2.3. Two routes, one destination. Plausible levels of noise create real problems for Campbell-Shiller.

## D.4 Return Extrapolation

Investors who extrapolate returns wind up overweighting recent price increases when forecasting future returns. Suppose demeaned returns follow an AR(1) process with $\phi \in (0, 1)$

$$\mathrm{R}_t \;=\; \phi \cdot \mathrm{R}_{t-1} \;+\; \varepsilon_t \tag{D.11}$$

The residual $\varepsilon_t$ is IID and represents the gap between the stock's realized return over the past year and the return investors were anticipating.

Extrapolators put positive weight $\theta > 0$ on the most recent return surprise, so the subjective forecast at horizon $h$ is

$$\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] \;=\; \mathbb{E}_t[\mathrm{R}_{t+h}] \;+\; \phi^h \cdot \{\theta \cdot \varepsilon_t\} \tag{22}$$

The factor $\phi^h$ comes from the rational forecast revision. Because returns are persistent, $\varepsilon_t$ predicts returns $h$ periods ahead with coefficient $\phi^h$. The forecast error at horizon $h$ is $\phi^h \cdot (\theta \cdot \varepsilon_t)$.

We assume no offsetting dividend-growth error. Hence, the adding-up violation lands on the return side

$$\sum_{h=1}^{\infty} \frac{\phi^h \cdot \{\theta \cdot \varepsilon_t\}}{(1 + \overline{\mathrm{DY}})^{h-1}} \;=\; \phi \cdot \{\theta \cdot \varepsilon_t\} \times \left(\frac{1}{1 - \rho \cdot \phi}\right) \tag{D.12}$$

As a numerical example, take $\overline{\mathrm{DY}} = 2\%$ so $\rho \approx 0.98$. Set $\phi = 0.1$, $\theta = 1$, and $\varepsilon_t = 10\%\mathrm{pt}$. The total violation is

$$0.1 \cdot \{1 \cdot 10\%\mathrm{pt}\} \times \left(\frac{1}{1 - 0.98 \cdot 0.1}\right) \;\approx\; 0.011 \text{ log units} \tag{D.13}$$

The PD mismatch is $1 - e^{-0.011} \approx 1\%$. The violation is small because $\phi = 0.1$ is small. Longer-term forecast errors have little effect since $(0.1)^h$ gets small fast. The same small persistence also keeps the multiplier small, $\left(\frac{1}{1-0.98\cdot 0.1}\right) \approx 1.1$. The total violation is dominated by the first term, $0.1 \times 10\%\mathrm{pt} = 1\%$.

Going from 1% to 0% requires an equilibrium model. Barberis, Greenwood, Jin, and Shleifer (2015) embeds extrapolation in a two-agent economy where extrapolators trade against rational investors. Dividends follow an exogenous endowment process, and the extrapolators' beliefs about future dividend growth are objectively correct. The bias is confined to returns. Market-clearing determines the equilibrium price, under which the extrapolator's coherent subjective measure satisfies Campbell-Shiller. The subjective return errors at different horizons offset each other when weighted by $\rho^{h-1}$. Jin and Sui (2022) accomplishes the same thing using a representative agent and a subjective Euler equation. Both papers assume investors' views about future dividend growth are correct, which is what forces them to build out the equilibrium structure.

## D.5 Diagnostic Beliefs

Bordalo, Gennaioli, and Shleifer (2018) proposes diagnostic beliefs: investors overweight recent cash-flow surprises rather than recent unexpected price movements. Suppose demeaned dividend growth follows an AR(1) process with $\phi \in (0, 1)$

$$\Delta \log \mathrm{Div}_t \;=\; \phi \cdot \Delta \log \mathrm{Div}_{t-1} \;+\; \varepsilon_t \tag{D.14}$$

$\varepsilon_t$ is IID. When news arrives at time $t$, a rational investor would revise her $h$-period forecast by $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] - \mathbb{E}_{t-1}[\Delta \log \mathrm{Div}_{t+h}] = \phi^h \cdot \varepsilon_t$.

Investors with diagnostic beliefs overweight this change by $\theta > 0$

$$\widetilde{\mathbb{F}}_t[\Delta \log \mathrm{Div}_{t+h}] \;=\; \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] \;+\; \theta \cdot \{\phi^h \cdot \varepsilon_t\} \tag{23}$$

The forecast error at horizon $h$ is $\theta \cdot \{\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] - \mathbb{E}_{t-1}[\Delta \log \mathrm{Div}_{t+h}]\}$.

We assume no offsetting error in investors' return forecasts. The adding-up violation lands on the dividend-growth side

$$\sum_{h=1}^{\infty} \frac{\theta \cdot \{\phi^h \cdot \varepsilon_t\}}{(1+\overline{\mathrm{DY}})^{h-1}} \;=\; \theta \cdot \{\phi \cdot \varepsilon_t\} \times \left(\frac{1}{1-\rho \cdot \phi}\right) \tag{D.15}$$

As a numerical example, take $\overline{\mathrm{DY}} = 2\%$ so $\rho \approx 0.98$. Set $\phi = 0.5$, $\theta = 1$, and $\varepsilon_t = 10\%\mathrm{pt}$. The total violation is

$$1 \cdot \{0.5 \cdot 10\%\mathrm{pt}\} \times \left(\frac{1}{1-0.98 \cdot 0.5}\right) \;\approx\; 0.10 \text{ log units} \tag{D.16}$$

which corresponds to a PD mismatch of $e^{0.10}-1 \;\approx\; 10\%$. The mathematics is analogous to return extrapolation applied to dividend growth instead of returns. The economics differs. Dividend growth is much more persistent than returns. The same initial 10%pt surprise that produced a $\approx 1\%$ violation in the extrapolation case now produces a violation roughly 10× larger.

The violation has errors only on the dividend-growth side. The return side is zero. Bordalo, Gennaioli, La Porta, and Shleifer (2019) restores consistency with Campbell-Shiller by treating the current price as the residual. The authors model diagnostic investors who anticipate a constant annual return, $\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] = 10\%$ for all $h \geq 1$. The price is then defined as the discounted sum of investors' diagnostic dividend forecasts at 10% per year. The adding-up condition in Equation (3) holds by construction. The consistency comes from how the price is defined, not from diagnostic expectations per se.

## D.6 Personal Experience

Malmendier and Nagel (2011) proposes that agents overweight personal experience when forecasting returns. Persistence adds little in this application, so we assume demeaned returns are approximately unpredictable

$$\mathrm{R}_t \;=\; \varepsilon_t \tag{D.17}$$

Let $\mu_t$ denote the investor's lifetime experience, which updates with weight $\theta \in (0, 1)$

$$\mu_t \;=\; \theta \cdot \mu_{t-1} \;+\; (1-\theta) \cdot \varepsilon_t \tag{D.18}$$

Larger values of $\theta$ mean old surprises stay in memory longer. $\mu_t$ is an exponentially-weighted average of the investor's lifetime-return history.

The investor's subjective return forecast at horizon $h$ is

$$\widetilde{\mathbb{F}}_t[\mathrm{R}_{t+h}] \;=\; \mathbb{E}_t[\mathrm{R}_{t+h}] \;+\; \mu_t \tag{24}$$

The forecast error is the same at every horizon. The investor simply thinks returns are $\mu_t$ higher, indefinitely. This is the perpetual error $\delta$ from Equation (17), with $\mu_t$ playing the role of $\delta$.

We assume no offsetting error in investors' dividend-growth forecasts. The adding-up violation lands on the return side

$$\sum_{h=1}^{\infty} \frac{\mu_t}{(1+\overline{\mathrm{DY}})^{h-1}} \;=\; \mu_t \times \left(\frac{1}{1-\rho}\right) \tag{D.19}$$

As a numerical example, take $\overline{\mathrm{DY}} = 2\%$ so $\rho \approx 0.98$. With an experience-driven bias of $\mu_t = 1\%\text{pt}$, the violation is

$$1\%\text{pt} \times \left(\frac{1}{1-0.98}\right) \;=\; 0.5 \text{ log units} \tag{D.20}$$

which corresponds to a PD mismatch of $1-e^{-0.5} \approx 39\%$. This is the perpetually-wrong case from earlier.

Malmendier and Nagel (2011) measures these experience effects from the UBS/Gallup survey, where cohort-level return forecasts are correlated with cohort-specific experienced returns. Survey responses come from $\widetilde{\mathbb{F}}_t[\cdot]$, not $\widetilde{\mathbb{E}}_t[\cdot]$. A researcher who combined them with separately sourced dividend-growth data in a forward-looking Campbell-Shiller framework would violate adding-up. Restoring consistency requires more assumptions.

## D.7 Natural Expectations

Fuster, Laibson, and Mendel (2010) proposes that agents fit a parsimonious model to data generated by a higher-order process. For simplicity, assume demeaned dividend growth, $x_t = \Delta \log \text{Div}_t - \overline{\Delta \log \text{Div}}$, follows an AR(2) process

$$x_t \;=\; \phi_1 \cdot x_{t-1} \;+\; \phi_2 \cdot x_{t-2} \;+\; \varepsilon_t \tag{D.21}$$

Investors fit an AR(1), $x_t = \hat{\phi}_1 \cdot x_{t-1} + \hat{\varepsilon}_t$. The parameter $\hat{\phi}_1$ represents the OLS coefficient from projecting $x_t$ on $x_{t-1}$. The perceived innovation $\hat{\varepsilon}_t = x_t - \hat{\phi}_1 \cdot x_{t-1}$ differs from the true innovation $\varepsilon_t = x_t - [\phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2}]$.

The agent's subjective forecast at horizon $h$ is

$$\widetilde{\mathbb{F}}_t[x_{t+h}] \;=\; \hat{\phi}_1^h \cdot x_t \tag{25}$$

The agent winds up overestimating persistence, capturing short-run momentum but missing long-run mean reversion. The true forecast involves the AR(2) eigenvalues, $\lambda_1$ and $\lambda_2$

$$\mathbb{E}_t[x_{t+h}] \;=\; y_t \cdot \lambda_1^h \;+\; z_t \cdot \lambda_2^h \tag{D.22}$$

$y_t$ and $z_t$ are linear functions of the current state $(x_t, x_{t-1})$

$$y_t \;=\; \left[\frac{\lambda_1 \cdot (x_t - \lambda_2 \cdot x_{t-1})}{\lambda_1 - \lambda_2}\right] \tag{D.23a}$$

$$z_t \;=\; -\left[\frac{\lambda_2 \cdot (x_t - \lambda_1 \cdot x_{t-1})}{\lambda_1 - \lambda_2}\right] \tag{D.23b}$$

The forecast error at horizon $h$ is $\hat{\phi}_1^h \cdot x_t - y_t \cdot \lambda_1^h - z_t \cdot \lambda_2^h$.

We assume no offsetting error in investors' return forecasts. The adding-up violation lands on the dividend-growth side

$$x_t \times \left(\frac{\hat{\phi}_1}{1 - \rho \cdot \hat{\phi}_1}\right) \;-\; y_t \times \left(\frac{\lambda_1}{1 - \rho \cdot \lambda_1}\right) \;-\; z_t \times \left(\frac{\lambda_2}{1 - \rho \cdot \lambda_2}\right) \tag{D.24}$$

Take $\overline{\text{DY}} = 2\%$ so $\rho \approx 0.98$. Set $\phi_1 = 1.2$ and $\phi_2 = -0.3$. The AR(2) eigenvalues are $\lambda_1 \approx 0.845$ and $\lambda_2 \approx 0.355$. The agent fits $\hat{\phi}_1 = \left(\frac{\phi_1}{1-\phi_2}\right) \approx 0.923$, which is more persistent than both true roots. Let $x_t = 5\%$pt and $x_{t-1} = 3\%$pt. Summing numerically, the total violation is approximately 0.16 log units, corresponding to a PD mismatch of $e^{0.16}-1 \approx 17\%$. At short horizons the agent slightly underforecasts. At longer horizons the agent overforecasts because the AR(1) misses the mean reversion from $\phi_2 = -0.3$.

The violation has errors only on the dividend-growth side. The return side is zero. The adding-up condition is generically violated. Fuster, Hebert, and Laibson (2012) restores consistency by solving a consumption-based asset-pricing model in which the agent prices assets via an Euler equation under the misspecified AR(1) model. The Euler equation simultaneously pins down expected returns and expected dividend growth under the agent's coherent (but wrong) probability measure. The beliefs become $\widetilde{\mathbb{E}}_t[\cdot]$, and the adding-up condition is satisfied by construction.

# E Variance Decomposition

This appendix provides the full calculations behind the variance-decomposition discussion in Section 3.4. Both Cochrane (2008) and De la O and Myers (2021) use the same Gordon-style formula: a short-term estimate multiplied by a theory-implied $\left(\frac{1}{1-\rho\cdot\phi}\right)$ factor. The only difference between them is which data goes into the short-term estimate.

## E.1 Shared Framework

Campbell-Shiller in realized values is

$$\log \mathrm{PD}_t \;\approx\; \text{const} \;-\; \sum_{h=1}^{\infty} \rho^{h-1} \cdot \left\{ \mathrm{R}_{t+h} - \Delta \log \mathrm{Div}_{t+h} \right\} \tag{E.1}$$

$\overline{\mathrm{DY}} \approx 2\%$ for the S&P 500 and $\rho = \left(\frac{1}{1+\overline{\mathrm{DY}}}\right)$, so $\rho \approx 0.98$.

Projecting both sides on $\log \mathrm{PD}_t$ and using the linearity of covariance gives

$$100\% \;=\; \beta_{\mathrm{Div}}(\infty) \;+\; \beta_{\mathrm{R}}(\infty) \tag{28}$$

This condition says that 100% of the variation in the S&P 500's log PD ratio must come from dividend growth or returns.

For dividend growth, the coefficient formula is

$$\beta_{\mathrm{Div}}(\infty) \;=\; \frac{\mathbb{C}\mathrm{ov}\Big[\, \sum_{h=1}^{\infty} \rho^{h-1} \cdot \Delta \log \mathrm{Div}_{t+h},\; \log \mathrm{PD}_t \Big]}{\mathbb{V}\mathrm{ar}[\log \mathrm{PD}_t]} \tag{E.2}$$

$\beta_{\mathrm{R}}(\infty)$ is defined analogously for returns. We use $-\left( \sum_{h=1}^{\infty} \rho^{h-1} \cdot \mathrm{R}_{t+h} \right)$ so that the two slope coefficients in Equation (28) sum to 100%.

## E.2 Infinite-Sum Problem

Both the dividend discount model (DDM) and the Campbell-Shiller variance decomposition face the same operational problem. The object of interest in each model is an infinite sum over future periods. Nobody can compute the quantity directly.

For the DDM, the object is the price

$$\mathrm{Price}_t \;=\; \sum_{h=1}^{\infty} \frac{\mathbb{E}_t[\mathrm{Div}_{t+h}]}{(1+\mathrm{R})^h} \tag{12}$$

For the variance decomposition, the object is $\beta_{\text{Div}}(\infty)$. In neither case can the researcher observe the infinite future.

Both formulas require further assumptions to become operational. Gordon's solution for the DDM is to assume constant dividend growth. The infinite sum collapses to $\text{Price}_t = \mathbb{E}_t[\text{Div}_{t+1}] \times \left(\frac{1}{\text{R}-\text{G}}\right)$. The constant-growth assumption is where the infinity gets tamed.

Variance decompositions assume the predictor is AR(1) with persistence $\phi \in (0, 1)$. By linearity of covariance, the full-horizon slope is a $\rho$-weighted sum of one-period slope coefficients

$$\beta_{\text{Div}}(\infty) \;=\; \sum_{h=1}^{\infty} \rho^{h-1} \cdot b_{\text{Div}}(h) \tag{E.3}$$

$b_{\text{Div}}(h) = \frac{\mathbb{C}\text{ov}[\Delta \log \text{Div}_{t+h},\, \log \text{PD}_t]}{\mathbb{V}\text{ar}[\log \text{PD}_t]}$ is the one-period coefficient at horizon $h$.

Under the AR(1) assumption, each horizon coefficient is a geometrically decaying version of the $h$=1 coefficient

$$b_{\text{Div}}(h) \;=\; \phi^{h-1} \cdot b_{\text{Div}}(1) \tag{E.4}$$

The infinite sum becomes a geometric series with ratio $\rho \cdot \phi$. Since $b_{\text{Div}}(1) = \beta_{\text{Div}}(1)$, we can write the sum in Equation (E.3) as

$$\beta_{\text{Div}}(\infty) \;=\; \beta_{\text{Div}}(1) \times \left(\frac{1}{1 - \rho \cdot \phi}\right) \tag{E.5}$$

The AR(1) assumption is where the infinity gets tamed. This is the same [short-term estimate] × [theory-implied multiple] structure as the Gordon model. The difference between Cochrane (2008) and De la O and Myers (2021) is what $\beta_{\text{Div}}(1)$ measures: realized dividend growth or subjective forecasts.

## E.3 Cochrane (2008)

Cochrane (2008) regresses next year's realized dividend growth on the S&P 500's current log dividend yield, $\log \text{DY}_t = -\log \text{PD}_t$,

$$\Delta \log \text{Div}_{t+1} \;\overset{\text{OLS}}{\sim}\; \alpha_{\text{Div}} \;+\; \beta_{\text{Div}}(1) \cdot \log \text{DY}_t \;+\; \varepsilon_{t+1}^{\text{Div}} \tag{E.6}$$

using aggregate S&P 500 data from roughly 1926 onward. The paper finds $\beta_{\text{Div}}(1) \approx 0.8\%$. The coefficient is economically small and often statistically insignificant.

We will use $\phi \approx 0.92$ in the calculations below. Cochrane (2008) works with $\rho \approx 0.96$ and $\phi \approx 0.94$, reflecting a higher average dividend yield over his sample than the 2% anchor this paper uses. The main text uses $\rho = 0.98$ and $\phi = 0.92$ for consistency with $\overline{\text{DY}} \approx 2\%$. The multiplier is the same within rounding: $\frac{1}{1-0.98\cdot 0.92} \approx 10.2$ vs. $\frac{1}{1-0.96\cdot 0.94} \approx 10.3$. The argument does not depend on the exact values.

The key thing is that Cochrane (2008) uses a theory-implied factor to scale up his short-run $\beta_{\text{Div}}(1) = 0.8\%$ to arrive at a final answer of

$$\beta_{\text{Div}}(\infty) \;=\; \underbrace{\beta_{\text{Div}}(1)}_{0.8\%} \times \left( \frac{1}{1 - \underbrace{\rho}_{0.98} \cdot \underbrace{\phi}_{0.92}} \right) \;\approx\; 0.8\% \times 10.2 \;\approx\; 8\% \tag{4}$$

Since $\beta_{\text{Div}}(\infty) + \beta_{\text{R}}(\infty) = 1$, the discount-rate variance share is $1 - 0.08 = 0.92$. Cochrane (2008) claims ~92% of log PD variance comes from discount rates.

## E.4 De la O and Myers (2021)

De la O and Myers (2021) replace realized future dividend growth with analysts' survey forecasts from IBES. They regress the one-year-ahead subjective forecast on the current log PD ratio

$$\widetilde{\mathbb{F}}_t[\Delta \log \text{Div}_{t+1}] \overset{\text{OLS}}{\sim} \alpha_{\text{Div}} + \beta_{\text{Div}}(1) \cdot \log \text{PD}_t + \varepsilon_{t+1} \tag{E.7}$$

The paper finds that subjective dividend-growth forecasts respond substantially to the current log PD ratio, $\beta_{\text{Div}}(1) \approx 39\%$. This point estimate is roughly fifty times larger than Cochrane (2008)'s value. The persistence of the subjective forecasts is lower than the persistence of realized data, $\phi \approx 0.60$.

De la O and Myers (2021) work in quarterly data with $\rho \approx 0.996$ and estimate $\phi \approx 0.58$ at quarterly frequency. The main text uses $\rho = 0.98$ and $\phi = 0.60$ for consistency with the $\overline{\text{DY}} \approx 2\%$ annual anchor. The multiplier is the same within rounding: $\frac{1}{1-0.98\cdot 0.60} \approx 2.43\times$ vs. $\frac{1}{1-0.996\cdot 0.58} \approx 2.37\times$. Both give the same headline $\beta_{\text{Div}}(\infty) \approx 93\%$.

This paper also uses a theory-implied multiple to scale up their short-run $\beta_{\text{Div}}(1) = 39\%$. But the result of this calculation could not be more different

$$\beta_{\text{Div}}(\infty) \;=\; \underbrace{\beta_{\text{Div}}(1)}_{39\%} \times \left( \frac{1}{1 - \underbrace{\rho}_{0.98} \cdot \underbrace{\phi}_{0.60}} \right) \;\approx\; 39\% \times 2.4 \;\approx\; 93\% \tag{29}$$

Subjective dividend-growth revisions now explain 93% of log PD variance, leaving 7% for subjective returns. The opposite answer from Cochrane (2008).

## E.5 Side-by-Side Comparison

| | Cochrane | De la O and Myers |
|---|---|---|
| Data | Realized growth | Survey forecasts |
| $\beta_{\text{Div}}(1)$ | 0.8% | 39.0% |
| $\phi$ | 0.92 | 0.60 |
| $\rho$ | 0.98 | 0.98 |
| Multiplier $\left(\frac{1}{1-\rho\cdot\phi}\right)$ | 10.2× | 2.4× |
| $\beta_{\text{Div}}(\infty)$ | 0.08 | 0.93 |
| $\beta_{\text{R}}(\infty)$ | 0.92 | 0.07 |

Same arithmetic. Different data. Opposite conclusions.

## E.6 The Gordon Connection

The Gordon pricing formula and the variance decomposition share the same analytical structure. Both compute a large, policy-relevant quantity by estimating something small and easy to measure at one horizon, then multiplying by a theory-implied factor. The theory-implied multiple is doing most of the heavy lifting. The two multiples share inputs on the S&P 500, since $\rho = \frac{1}{1+\overline{\text{DY}}}$ and the long-run Gordon relationship gives $\overline{\text{DY}} \approx (\text{R}-\text{G})$. But the structural parallel would hold even if the inputs were different. The point is that both calculations have the form: short-term estimate times theory-implied multiple.