Crimes Against Campbell-Shiller (Internet Appendix)

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

April 17, 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 Price_t, collect a dividend of Div_{t+1}, and sell at Price_{t+1}. Your gross return will be

$$1 + \text{R}_{t+1} = \frac{\text{Price}_{t+1} + \text{Div}_{t+1}}{\text{Price}_t} \tag{A.1}$$

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, $\text{PD}_t = \frac{\text{Price}_t}{\text{Div}_t}$,

$$\begin{aligned}
1 + \text{R}_{t+1} &= \left(\frac{\text{Price}_{t+1} + \text{Div}_{t+1}}{\text{Price}_t}\right) \times \left(\frac{\text{Div}_{t+1}/\text{Div}_{t+1}}{\text{Div}_t/\text{Div}_t}\right) && \text{(A.2a)} \\
&= \left(\frac{\text{Price}_{t+1}/\text{Div}_{t+1} + 1}{\text{Price}_t/\text{Div}_t}\right) \times \left(\frac{\text{Div}_{t+1}}{\text{Div}_t}\right) && \text{(A.2b)} \\
&= \left(\frac{\text{PD}_{t+1} + 1}{\text{PD}_t}\right) \times \left(\frac{\text{Div}_{t+1}}{\text{Div}_t}\right) && \text{(A.2c)}
\end{aligned}$$

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

Take the natural logarithm of both sides. Write R_{t+1} for the log gross return, $\log \mathrm{PD}_t$ for the log price-dividend ratio, and $\Delta \log \mathrm{Div}_{t+1}$ for log dividend growth. This gives

$$\mathrm{R}_{t+1} \;=\; \log(\mathrm{PD}_{t+1}{+}1) \;-\; \log \mathrm{PD}_t \;+\; \Delta \log \mathrm{Div}_{t+1} \tag{A.3}$$

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 $\log \overline{\mathrm{PD}} = \mathbb{E}[\log \mathrm{PD}_t]$, the mean log PD ratio. If we define $\overline{\mathrm{PD}} = e^{\log \overline{\mathrm{PD}}}$ and $\overline{\mathrm{DY}} = 1/\overline{\mathrm{PD}}$, then the expansion gives

$$\log(\mathrm{PD}_{t+1}{+}1) \;\approx\; \log(\overline{\mathrm{PD}}{+}1) \;+\; \left(\frac{1}{1+\overline{\mathrm{DY}}}\right) \times (\log \mathrm{PD}_{t+1} - \log \overline{\mathrm{PD}}) \tag{A.4}$$

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

$$\rho = \frac{1}{1+\overline{\mathrm{DY}}} \qquad\qquad \kappa = \log(\overline{\mathrm{PD}}{+}1) - \rho \cdot \overline{\log \mathrm{PD}} \tag{A.5}$$

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.4) and (A.5) into Equation (A.3) to arrive at the one-period version of the Campbell-Shiller formula

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

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

$$\log \mathrm{PD}_t \;\approx\; \kappa \;+\; \Delta \log \mathrm{Div}_{t+1} \;-\; \mathrm{R}_{t+1} \;+\; \rho \cdot \log \mathrm{PD}_{t+1} \tag{A.7}$$

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.7) 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

$$\begin{aligned} \log \mathrm{PD}_t \;\approx\; & \kappa \cdot (1{+}\rho) \;+\; \{\Delta \log \mathrm{Div}_{t+1} - \mathrm{R}_{t+1}\} \\ & \;+\; \rho \cdot \{\Delta \log \mathrm{Div}_{t+2} - \mathrm{R}_{t+2}\} \;+\; \rho^2 \cdot \log \mathrm{PD}_{t+2} \end{aligned} \tag{A.8}$$

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

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

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.9). 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

$$\begin{aligned}\log \mathrm{PD}_t \;\approx\; -\sum_{h=1}^{H} \rho^{h-1} \cdot \left\{ \mathbb{E}_t[\mathrm{R}_{t+h}] - \mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}] \right\} \\ +\; \rho^H \cdot \mathbb{E}_t[\log \mathrm{PD}_{t+H}]\end{aligned} \tag{A.10}$$

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

$$\lim_{H \to \infty} \rho^H \cdot \mathbb{E}_t[\log \mathrm{PD}_{t+H}] \;=\; 0 \tag{A.11}$$

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

$$\mathrm{PD} \;=\; \left(\frac{1+\mathrm{G}}{\mathrm{R}-\mathrm{G}}\right) \tag{A.12}$$

so $\log \mathrm{PD} = \log(1+\mathrm{G}) - \log(\mathrm{R}-\mathrm{G})$. The Campbell-Shiller sum on the right-hand side of Equation (1) captures the $-\log(\mathrm{R}-\mathrm{G})$ component. The $\log(1+\mathrm{G})$ term, which accounts for the one-period growth between the trailing and forward dividend, is absorbed into the leading constant.

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 number of shares outstanding at the start of the period.

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 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, Div_{t+1} = \$0.25/sh and 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 $\text{\#Shares}_{t-1}$ = 100M 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 market cap, since $98.43\text{M} \times \$3.175/\text{sh} = \$312.5\text{M}$.

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 $\#\text{Shares}_t = 100\text{M}$ 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 $\text{Div}_t = \$0.25/\text{sh}$, the same $\text{Price}_t = \$3.125/\text{sh}$, 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 look identical to Baseline Inc to begin with. Each starts out earning \$25M per period, distributes all of it as dividends, has an 8% discount rate, and trades at $\text{Price}_t = \$3.125/\text{sh}$ on $\#\text{Shares}_t = 100\text{M}$ shares. However, between t and $(t+1)$, both firms invest in the same project, which earns 12%, generating \$9.375M per year in additional earnings. They fund the project's \$78.125M upfront cost by issuing 25M new shares at \$3.125/sh. The two firms differ only in how quickly the board raises the dividend.

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

$$\mathrm{PD}_t = \frac{\$3.125}{\$0.25} = 12.5\times \tag{B.11a}$$
$$\mathrm{R}_{t+1} = \frac{\$3.4375 + \$0.275}{\$3.125} - 1 = 18.8\% \tag{B.11b}$$
$$\Delta \log \mathrm{Div}_{t+1} = \log\left(\frac{\$0.275}{\$0.25}\right) = +9.5\% \tag{B.11c}$$

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 $\#\mathrm{Shares}_t$ = 125M shares:

$$\mathrm{Div}_{t+1} = \frac{\$25\mathrm{M}}{125\mathrm{M}} = \$0.20/\mathrm{sh} \tag{B.12a}$$
$$\mathrm{Price}_{t+1} = \frac{\$439.0625\mathrm{M}}{125\mathrm{M}} = \$3.5125/\mathrm{sh} \tag{B.12b}$$

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

$$\mathrm{PD}_t = \frac{\$3.125}{\$0.25} = 12.5\times \tag{B.13a}$$
$$\mathrm{R}_{t+1} = \frac{\$3.5125 + \$0.20}{\$3.125} - 1 = 18.8\% \tag{B.13b}$$
$$\Delta \log \mathrm{Div}_{t+1} = \log\left(\frac{\$0.20}{\$0.25}\right) = -22.3\% \tag{B.13c}$$

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 exactly under correct expectations. Investors can see the current share price and the per-share dividend over the past twelve months

$$\text{Price}_t = \$100/\text{sh} \tag{C.1a}$$

$$\text{Div}_t = \$2/\text{sh} \tag{C.1b}$$

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

$$\text{R} = 8\% \tag{C.2a}$$

$$\text{G} = 6\% \tag{C.2b}$$

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

Let $\tilde{\text{R}}$ and $\tilde{\text{G}}$ 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

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

This is the exact constraint.

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

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

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 Gordon model. Moreover, her subjective beliefs $\tilde{\mathrm{R}}$ and $\tilde{\mathrm{G}}$ stem from a subjective probability measure, $\tilde{\mathbb{Q}}$, which is consistent with her model.

Eve's primitives must satisfy Equation (C.4): $\tilde{\mathrm{R}} - \tilde{\mathrm{G}} = \mathrm{DY} = 2\%$. She has one degree of freedom, not two. Suppose Eve picks $\tilde{\mathrm{R}} = 9\%$. Then $\tilde{\mathrm{G}} = 7\%$. 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

$$\tilde{\mathrm{R}} - \mathrm{R} = +1\%\mathrm{pt} \tag{C.5a}$$

$$\tilde{\mathrm{G}} - \mathrm{G} = +1\%\mathrm{pt} \tag{C.5b}$$

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

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

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

$$\mathrm{RHS} = \sum_{h=1}^{\infty} \frac{0.01}{(1.02)^{h-1}} = 0.51 \tag{C.6b}$$

No calibration was required. All the work was done by the assumption that $\tilde{\mathbb{Q}}$ 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 $\tilde{\mathbb{Q}}$. 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 $\tilde{\mathrm{R}}$ and $\tilde{\mathrm{G}}$ 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 ε_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 $\tilde{\mathrm{R}}$ collapses the last degree of freedom onto $\tilde{\mathrm{G}}$. Richer models with time-varying parameters multiply both the primitives and the constraints. Making a belief-biased investor's $\tilde{\mathbb{Q}}$ 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 Adding-Up Condition

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

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

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

The following geometric sum appears in every case below

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

The main text uses the approximation, $\sum_h \frac{1}{(1+\overline{\text{DY}})^h} = \frac{1}{\overline{\text{DY}}}$. We will analyze both forms in the derivations below, noting where the approximation differs.

D.1 Perpetually Wrong

Suppose that investors have correct dividend-growth forecasts, but they overestimate future returns by $\delta > 0$ at each horizon $h \geq 1$. We can write the resulting forecast errors as

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

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

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

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

A 1%pt perpetual error on returns looks trivial at any single horizon, but when capitalized at the long-run average dividend year, it can produce a sizeable adding-up violation.

For $\overline{\text{DY}} = 2\%$, we get a multiple of $\left(\frac{1+2\%}{2\%}\right) = 51\times$. So a +1%pt perpetual mistake about future returns would translate to an adding-up violation of

$$\text{Violation} = +1\%\text{pt} \times \left(\frac{1+2\%}{2\%}\right) = 51\% \tag{D.3}$$

The logic is identical to that of the Gordon pricing rule. This is because the adding-up condition in Equation (3) has the same infinite discounted sum structure as the dividend discount model (DDM). And the Gordon model is a special case of the DDM. Note that, in a world governed by the Gordon model, the forward dividend yield is equal to the firm's cap rate, $\mathrm{DY} = (\mathrm{R}-\mathrm{G})$. So the parallel is exact.

D.2 Moving Target

Again, assume that investors' dividend-growth forecasts are correct, but they over-estimate future returns at every horizon. Now these mistakes decay geometrically with horizon at rate $\phi \in [0, 1)$

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

The factor of ϕ^{h-1} reflects AR(1) dynamics. This structure shows up in every AR(1)-based bias model discussed in Section 3.2.

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

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

The $\left(\frac{1}{1-\rho\cdot\phi}\right)$ multiple shows up throughout the belief-based asset-pricing literature. Section **??** shows how it plays a starring role in papers that use variance decompositions.

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, $\frac{1}{\overline{\mathrm{DY}}} = \left(\frac{1}{\mathrm{R}-\mathrm{G}}\right)$, modified by a mean-reversion term $(1-\phi)$ in the denominator. Three limit cases make the formula concrete:

1. <u>Perpetual error ($\phi = 1$).</u> The correction vanishes, $\frac{1}{\overline{\mathrm{DY}}+(1-1)} = \frac{1}{\overline{\mathrm{DY}}}$. We recover the perpetually wrong case from above.
2. <u>One-off error ($\phi = 0$).</u> Denominator is $\overline{\mathrm{DY}}+1 \approx 1$. The multiplier is approximately 1, and a $\delta = +1\%$pt error produces a 1% violation. No amplification.
3. <u>Goldilocks zone ($\phi = 0.6$).</u> With $\overline{\mathrm{DY}} = 2\%$, the denominator is $2\% + 40\% = 42\%$, giving a multiplier of $\sim 2.4\times$. A $\delta = +1\%$pt error produces a 2.4% violation. Much smaller than the $\phi = 1$ case, but still an order of magnitude larger than the $\phi = 0$ case. Since dividend yields are small to begin with, small changes in persistence matter a lot.

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. Suppose a researcher who plugs sample estimates of $\mathbb{E}_t[\Delta \log \mathrm{Div}_{t+h}]$ and $\mathbb{E}_t[\mathrm{R}_{t+h}]$ into Equation (1). Errors are IID white noise with zero mean and positive variance

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

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

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

$$\mathrm{Gap} \;=\; \mathrm{LHS} - \mathrm{RHS} \;=\; \sum_{h=1}^{\infty} \frac{\varepsilon_h^{\mathrm{Div}} - \varepsilon_h^{\mathrm{R}}}{(1 + \overline{\mathrm{DY}})^{h-1}} \tag{D.6}$$

Adding-up holds if and only if $\mathrm{Gap} = 0$.

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

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

However, the variance is a different story

$$\mathbb{V}\mathrm{ar}[\mathrm{Gap}] \;=\; \sum_{h=1}^{\infty} \frac{\sigma_{\mathrm{Div}}^2 + \sigma_{\mathrm{R}}^2}{(1 + \overline{\mathrm{DY}})^{2(h-1)}} \tag{D.8a}$$

$$=\; (\sigma_{\mathrm{Div}}^2 + \sigma_{\mathrm{R}}^2) \times \left[\frac{(1 + \overline{\mathrm{DY}})^2}{(1 + \overline{\mathrm{DY}})^2 - 1} \right] \tag{D.8b}$$

$$=\; \frac{\sigma_{\mathrm{Div}}^2 + \sigma_{\mathrm{R}}^2}{1 - \rho^2} \tag{21}$$

The size of the adding-up gap is a continuous random variable with positive variance, so Equation (3) is violated almost surely, $\Pr[\mathrm{Gap} = 0] = 0$.

For a numerical anchor, take $\overline{\mathrm{DY}} = 2\%$. This choice implies $\rho \approx 0.980$ and $\frac{1}{1-\rho^2} \approx 25.8$. Suppose $\sigma_{\mathrm{Div}} = \sigma_{\mathrm{R}} = 1\%\mathrm{pt}$, so $\sigma_{\mathrm{Div}}^2 + \sigma_{\mathrm{R}}^2 = 0.0002$. In this scenario, the adding-up condition would be regularly off by around $\pm 7\%\mathrm{pt}$

$$\mathbb{S}\mathrm{d}[\mathrm{Gap}] \approx \sqrt{0.0002 \times 25.8} \;=\; \sqrt{0.0052} = 7.2\%\mathrm{pt} \tag{D.9}$$

Tiny, unbiased, independent measurement errors of 1%pt produce meaningful violations of the key condition required for Campbell-Shiller to hold.

E Popular Models of Biased Beliefs

This appendix provides the full derivations behind the four applications discussed in Section 3.2. Each application specifies three things: a law of motion for the variable being forecasted (R or $\Delta \log \text{Div}$), a law of motion for the anchor K_t, and a subjective forecast rule. The adding-up violation is then computed by summing the forecast errors with weights ρ^{h-1} where $\rho = \left(\frac{1}{1+\overline{\text{DY}}}\right)$.

E.1 Extrapolation

Law of motion. Demeaned returns follow an AR(1) with $\phi \in (0, 1)$

$$\text{R}_t \;=\; \mu \;+\; \phi \cdot (\text{R}_{t-1} - \mu) \;+\; \varepsilon_t \tag{E.1}$$

The anchor is $K_t = \varepsilon_t$, the most recent return innovation. K_t is IID. Each period brings a fresh anchor.

Forecast rule. Extrapolators put positive weight $\theta > 0$ on recent performance. The subjective forecast at horizon h is

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

K_t is the most recent return shock. Because returns are persistent, K_t still predicts returns h periods ahead with coefficient ϕ^h. The forecast error at horizon h is therefore

$$(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\text{R}_{t+h}] \;=\; \phi^h \cdot (\theta \cdot K_t) \tag{E.2}$$

We assume no offsetting dividend-growth error.

Violation. The adding-up violation lands on the return side

$$\sum_{h=1}^{\infty} \rho^{h-1} \cdot \phi^h \cdot (\theta \cdot K_t) \;=\; \phi \cdot (\theta \cdot K_t) \times \left(\frac{1}{1 - \rho \cdot \phi}\right) \tag{E.3}$$

As a numerical example, let $\overline{\text{DY}} = 2\%$ so $\rho \approx 0.980$. Assume $\phi = 0.8$, $\theta = 0.5$, and $K_t = \varepsilon_t = +10\%$pt. The forecast error is $+4.0\%$pt at $h = 1$, $+3.2\%$pt at $h = 2$, and $+0.5\%$pt at $h = 10$. The total violation is $0.8 \cdot (0.5 \cdot 10\%) \times \left(\frac{1}{1-0.980 \cdot 0.8}\right) = 18.5\%$.

Restoring consistency. The violation above has errors only on the return side of the adding-up condition. The dividend-growth side is zero. To restore balance, we need either offsetting dividend-growth errors or a joint model that pins down both sides simultaneously.

Barberis, Greenwood, Jin, and Shleifer (2015) take the second route. They embed extrapolation in a two-agent economy where extrapolators trade against rational investors. Market clearing determines the equilibrium price, which simultaneously pins down $\widetilde{\mathbb{E}}_t[\mathrm{R}_{t+h}]$, $\widetilde{\mathbb{E}}_t[\Delta \log \mathrm{Div}_{t+h}]$, and $\widetilde{\mathbb{E}}_t[\log \mathrm{PD}_{t+h}]$ under a single coherent probability measure. Jin and Sui (2022) do the same with a representative agent and a subjective Euler equation. In both cases, the equilibrium apparatus fills in both sides of the adding-up condition jointly. The extrapolation alone only fills in one side.

E.2 Diagnostic Expectations

Law of motion. Demeaned dividend growth follows an AR(1) with $\phi \in (0, 1)$

$$\Delta \log \mathrm{Div}_t \;=\; \mu \;+\; \phi \cdot (\Delta \log \mathrm{Div}_{t-1} - \mu) \;+\; \varepsilon_t \tag{E.4}$$

The anchor is $K_t = \varepsilon_t$, the most recent dividend-growth innovation. ε_t is IID.

Forecast rule. When new information arrives at t, the rational forecast revision at horizon h is $\phi^h \cdot K_t$, decaying at rate ϕ due to the AR(1) assumption. Diagnostic expectations amplify this revision by $\theta > 0$. The subjective forecast at horizon h is

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

The forecast error at horizon h is therefore

$$(\widetilde{\mathbb{F}}_t - \mathbb{E}_t)[\Delta \log \mathrm{Div}_{t+h}] \;=\; \theta \cdot (\phi^h \cdot K_t) \tag{E.5}$$

There is no offsetting error in investors' return forecasts.

Violation. The adding-up violation lands on the dividend-growth side

$$\sum_{h=1}^{\infty} \rho^{h-1} \cdot \theta \cdot \phi^h \cdot K_t \;=\; \theta \cdot (\phi \cdot K_t) \times \left(\frac{1}{1 - \rho \cdot \phi} \right) \tag{E.6}$$

Suppose $\overline{\mathrm{DY}} = 2\%$, $\phi = 0.5$, $\theta = 1$, and $K_t = +10\%$pt. The forecast error is +5.0%pt at $h = 1$, +2.5%pt at $h = 2$, and approximately zero by $h = 10$. The total violation is $1 \cdot (0.5 \cdot 10\%) \times \left(\frac{1}{1-0.980 \cdot 0.5}\right) = +9.8\%$pt. Smaller than the extrapolation example because dividend growth is less persistent, $\phi = 0.5$ vs. 0.8.

Restoring consistency. The violation above has errors only on the dividend-growth side. The return side is zero. Because the bias applies to dividend growth, a simpler fix is available: the price can be computed directly from the discounted sum of biased dividend forecasts, and the return forecast can be defined as whatever value makes the one-period Campbell-Shiller identity hold.

Bordalo, Gennaioli, La Porta, and Shleifer (2019) take this route. They derive the price from diagnostically expected dividends, then define the return forecast as the residual from the Campbell-Shiller identity. This forces the return side of Equation (3) to match the dividend-growth side by construction. The consistency comes from treating one forecast as a residual of the other, not from diagnostic expectations per se.

Bordalo, Gennaioli, La Porta, and Shleifer (2024) do not restore consistency. They apply the diagnostic distortion only to expected dividend growth while setting required returns to a constant unrelated to the belief distortion. There is no offsetting return error. This is inconsistent with the adding-up condition, and the resulting variance decomposition inherits the violation.

E.3 Personal Experience

Law of motion. Returns are approximately unpredictable: $\mathrm{R}_t = \mu + \varepsilon_t$. The anchor is the investor's lifetime experienced return

$$K_t = \theta \cdot K_{t-1} + (1{-}\theta) \cdot \varepsilon_t \tag{E.7}$$

$\theta \in (0,1)$ close to unity means old experiences fade slowly. K_t is an exponentially-weighted average of the investor's lifetime-return history.

Forecast rule. The investor simply thinks returns are K_t higher at every horizon. The subjective forecast at horizon h is

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

There is no h-dependence. This is precisely the perpetual error δ from Equation (17), with K_t playing the role of δ.

Violation. The adding-up violation lands on the return side

$$\sum_{h=1}^{\infty} \rho^{h-1} \cdot K_t \;=\; K_t \times \left(\frac{1}{1-\rho}\right) \tag{E.8}$$

As a numerical example, let $\overline{\mathrm{DY}} = 2\%$, $\theta = 0.97$, and $K_t = +2\%\mathrm{pt}$. The total violation is $2\% \times \left(\frac{1}{1-0.980}\right) \approx 100\%$. A modest +2%pt annual bias creates a 50× larger violation because the errors get discounted at the dividend yield.

Restoring consistency. The violation has errors only on the return side. The dividend-growth side is zero. The same structure as extrapolation, the same fix needed. Malmendier and Nagel (2011) measure experience effects from the UBS/Gallup survey, where cohort-level return expectations are correlated with cohort-specific experienced returns. Those survey-measured expectations are $\widetilde{\mathbb{F}}_t[\cdot]$. A researcher who combined them with separately sourced dividend-growth expectations in the forward-looking Campbell-Shiller formula would violate adding-up. Restoring consistency requires an equilibrium model in which agents learn from a finite history and price assets via an Euler equation given their posterior beliefs. The Euler equation determines the price under a single coherent probability measure, which simultaneously pins down $\widetilde{\mathbb{E}}_t[\mathrm{R}_{t+h}]$ and $\widetilde{\mathbb{E}}_t[\Delta \log \mathrm{Div}_{t+h}]$. As with extrapolation, the one-period Campbell-Shiller identity holds for every realization, and linearity of $\widetilde{\mathbb{E}}_t[\cdot]$ preserves it.

E.4 Natural Expectations

Law of motion. Demeaned dividend growth, $x_t = \Delta \log \mathrm{Div}_t - \mu$, follows an AR(2) process

$$x_t \;=\; \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} \;+\; \varepsilon_t \tag{E.9}$$

The agent fits 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} alone. 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 attributes predictable mean reversion to noise.

Forecast rule. The agent forecasts using her misspecified AR(1). The subjective forecast at horizon h is

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

The true forecast involves the AR(2) eigenvalues λ_1,λ_2

$$\mathbb{E}_t[x_{t+h}] \;=\; y_t \cdot \lambda_1^h + z_t \cdot \lambda_2^h \tag{E.10}$$

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

$$y_t = \frac{\lambda_1 \cdot (x_t - \lambda_2 \cdot x_{t-1})}{\lambda_1 - \lambda_2} \qquad z_t = \frac{-\lambda_2 \cdot (x_t - \lambda_1 \cdot x_{t-1})}{\lambda_1 - \lambda_2} \tag{E.11}$$

The forecast error at horizon h is $\hat{\phi}_1^h \cdot x_t - y_t \cdot \lambda_1^h - z_t \cdot \lambda_2^h$. The formula involves three exponential terms with different bases.

Violation. The adding-up violation lands on the dividend-growth side

$$x_t \times \left(\frac{\hat{\phi}_1}{1 - \rho \cdot \hat{\phi}_1}\right) - y_t \times \left(\frac{\lambda_1}{1 - \rho \cdot \lambda_1}\right) - z_t \times \left(\frac{\lambda_2}{1 - \rho \cdot \lambda_2}\right) \tag{E.12}$$

As a numerical example, let $\overline{\text{DY}} = 2\%$, $\phi_1 = 1.2$, and $\phi_2 = -0.3$. The AR(2) eigenvalues are $\lambda_1 \approx 0.845$ and $\lambda_2 \approx 0.355$. The agent fits $\hat{\phi}_1 = \frac{\phi_1}{1-\phi_2} \approx 0.923$, which is more persistent than both true roots. Let $x_t = +5\%\text{pt}$ and $x_{t-1} = +3\%\text{pt}$. At short horizons the agent slightly underforecasts. The AR(2) has more short-run momentum than the AR(1) captures. At longer horizons the agent overforecasts because the AR(1) misses the mean reversion from $\phi_2 = -0.3$. Summing numerically, the total violation is approximately 16%.

Restoring consistency. The violation has errors only on the dividend-growth side. The return side is zero. The same structure as diagnostic expectations: unmatched errors on one side. Fuster, Hebert, and Laibson (2012) restore 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 determines the price under the agent's coherent (but wrong) probability measure. This simultaneously pins down $\widetilde{\mathbb{E}}_t[\text{R}_{t+h}]$ and $\widetilde{\mathbb{E}}_t[\Delta \log \text{Div}_{t+h}]$. A researcher who instead took the natural-expectations forecast for dividend growth and combined it with separately measured return expectations would be working with $\widetilde{\mathbb{F}}_t[\cdot]$, and the adding-up condition would not hold.

E.5 The Asymmetry

The four applications split into two groups based on which side of the adding-up condition the errors land on. This determines what it takes to restore consistency.

Return-side biases (extrapolation, personal experience). The errors are on the return side. The dividend-growth side is zero. To close the gap, we need dividend-growth errors that match. But the bias is a story about returns, not dividends. There is no natural way to generate dividend-growth errors from a return-only bias. The only available fix is a general-equilibrium model, like in Barberis et al. (2015) and Jin and Sui (2022). The equilibrium determines the price, which simultaneously pins down expected returns, expected dividend growth, and the PD ratio under a single coherent measure. The dividend-growth side gets filled in as a byproduct of the equilibrium, not as a separate assumption.

Dividend-growth-side biases (diagnostic expectations, natural expectations). The errors are on the dividend-growth side. The return side is zero. Here a simpler fix is available because prices follow directly from dividend forecasts. The price is the present value of expected future dividends. If you have biased dividend forecasts, you get a biased price. Once you have the price, returns are mechanically determined. So the return forecast can be defined as whatever value makes the one-period Campbell-Shiller identity hold. This is the route Bordalo et al. (2019) take. Alternatively, one can solve the full optimization problem under the misspecified model, as Fuster et al. (2012) do. Both routes work, but the first is available only for dividend-growth-side biases.

This creates an asymmetry in modeling overhead. Return-side biases require a full equilibrium to restore consistency. Dividend-growth-side biases can be fixed with a one-line identity. The asymmetry exists because the present-value relationship runs from dividends to prices, not from returns to prices.

F Variance Decomposition

This appendix provides the full calculations behind the variance-decomposition discussion in Section 3.3. 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.

F.1 Shared Framework

The Campbell-Shiller identity in realized values is

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

$\overline{\mathrm{DY}} \approx 2\%$ 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

$$1 \;=\; \beta_{\mathrm{Div}}(\infty) \;-\; \beta_{\mathrm{R}}(\infty) \tag{27}$$

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

For dividend growth, the coefficient formula is

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

$\beta_{\mathrm{R}}(\infty)$ is defined analogously for returns.

F.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 is an infinite sum over future periods. Nobody can compute one 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 ρ-weighted sum of one-period slope coefficients

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

$b_{\text{Div}}(h) = \frac{\mathbb{C}\text{ov}[\Delta \log \text{Div}_{t+h},\, \log \text{PD}_t]}{\mathbb{V}\text{ar}[\log \text{PD}_t]}$ is the one-period coefficient at horizon h. Note that $b_{\text{Div}}(1) = \beta_{\text{Div}}(1)$.

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

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

The infinite sum becomes a geometric series with ratio $\rho \cdot \phi$. The total can be written as

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

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.

F.3 Cochrane (2008)

Cochrane (2008) regresses next year's realized dividend growth on the S&P 500's current log PD ratio

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

using aggregate S&P 500 data from roughly 1926 onward. The paper finds $\beta_{\text{Div}}(1) \approx 0.8\%$. i.e., one-year-ahead realized dividend growth barely responds to the log PD ratio. 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.

F.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

$$\widetilde{\mathbb{F}}_t[\Delta \log \text{Div}_{t+1}] \overset{\text{OLS}}{\sim} \tilde{\alpha}_{\text{Div}} + \tilde{\beta}_{\text{Div}}(1) \cdot \log \text{PD}_t + \tilde{\varepsilon}_{t+1} \tag{F.7}$$

The paper finds that subjective dividend-growth forecasts respond substantially to the current log PD ratio, $\tilde{\beta}_{\text{Div}}(1) \approx 39\%$. 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 $\tilde{\beta}_{\text{Div}}(\infty) \approx 93\%$.

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

$$\tilde{\beta}_{\text{Div}}(\infty) \;=\; \underbrace{\tilde{\beta}_{\text{Div}}(1)}_{39\%} \times \left(\frac{1}{1 - \underbrace{\rho}_{0.98} \cdot \underbrace{\hat{\phi}}_{0.60}} \right) \;\approx\; 39\% \times 2.4 \;\approx\; 93\% \tag{28}$$

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

F.5 Side-by-Side Comparison

	Cochrane	De la O and Myers
Data	Realized growth	Survey forecasts
$\beta_{\text{Div}}(1)$	0.8%	39.0%
ϕ	0.92	0.60
ρ	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.

F.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.