Stock Returns Decomposition*

Preliminary draft — please do not cite without permission

Gabriel E. Cabrera Guzmán†

This revision: 19 September 2026

Abstract

These notes develop the Campbell–Shiller present-value decomposition of stock returns and its vector autoregressive implementation, following the treatment in Rapach, Ringgenberg, and Zhou (2016). I start from the exact definition of the log return, derive the log-linear approximation and the discount coefficient ρ, solve the resulting difference equation forward under a no-bubble condition, and show how the unexpected return splits into cash flow news and discount rate news. A first-order VAR then turns the two unobservable infinite sums into closed-form linear functions of the VAR residuals, and I discuss the specification choices that make or break the exercise. The final section applies the machinery to the question that motivates it: when aggregate short interest predicts market returns, which component of the return is it predicting? Table 8 of the source paper is replicated exactly, and the replication is then used to show that while the total predictive slope is invariant to ρ, its split between the two channels is not — the published cash flow result requires a ρ calibrated from an annual dividend–price ratio, and reverses under the frequency-consistent monthly calibration.

Keywords: return decomposition, cash flow news, discount rate news, vector autoregression, return predictability
JEL Classification: G12, G14, C32

*These are working notes. They restate and expand the derivations in Section 5 of Rapach, Ringgenberg, and Zhou (2016) and replicate that paper's Table 8. Replication code is available at https://github.com/gabbocg/rrz2016. Any errors are my own.

†Email: gabriel.cabreraguzman@postgrad.manchester.ac.uk. Address: The University of Manchester, Manchester, UK.

1. Introduction

A stock price is a claim on a stream of future cash flows, discounted at a rate investors require for bearing risk. If the price falls today, only two things can have happened. Either the market revised downward its forecast of the cash flows the asset will deliver, or it revised upward the rate at which it discounts them. Nothing else can move the price, because nothing else enters the present-value relation.

That observation is nearly a tautology, and its usefulness comes from the fact that the two cases carry *opposite* implications for what happens next. A price drop caused by bad news about cash flows is a permanent loss — the asset is simply worth less, and expected returns going forward are unchanged. A price drop caused by an increase in required returns is different. The investor takes a capital loss today, but is compensated by a *higher* expected return from tomorrow onward. The same realized return means two different things depending on its source.

Splitting realized returns into these two pieces is the Campbell–Shiller decomposition (Campbell and Shiller 1988), made operational by Campbell (1991) and Campbell and Ammer (1993). The difficulty is that neither piece is observed. Both are revisions in expectations of infinite sums of future quantities, and expectations are not in the data. The empirical strategy is to posit a forecasting model — a vector autoregression — for the variables investors are assumed to watch, and let the VAR stand in for the market's conditional expectation. The infinite sums then collapse into a closed-form matrix expression in the VAR coefficients and residuals.

These notes work through that construction in full, then apply it. Section 2 derives the

log-linear present-value identity and the discount coefficient ρ. Section 3 turns the identity into the news decomposition. Section 4 shows how the VAR extracts the two news series and discusses what can go wrong. Section 5 follows Rapach et al. (2016) in using the decomposition to ask *why* aggregate short interest forecasts market returns, replicates their Table 8, and examines how much of the answer rests on the calibration of ρ. Section 6 collects the remaining caveats. Section 7 concludes. Section A carries the log-linearization algebra in full, and Section B the companion-form algebra.

Throughout, lowercase letters denote natural logarithms of the corresponding uppercase quantity, so $p_t = \log P_t$ and $d_t = \log D_t$. The operator E_t denotes expectations conditional on information available at the end of month t.

2. The Log-Linear Present-Value Identity

2.1 Why an approximation is needed

Start from the definition of the gross one-period return on a stock that pays dividend D_{t+1} during period $t+1$:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}. \tag{1}$$

Eq. (1) is an accounting identity, not a model. Taking logs,

$$r_{t+1} = \log(P_{t+1} + D_{t+1}) - p_t = p_{t+1} + \log\left(1 + e^{\,d_{t+1} - p_{t+1}}\right) - p_t. \tag{2}$$

Eq. (2) is still exact, but it is useless as it stands. The log of a sum does not decompose, so r_{t+1} is a nonlinear function of p_{t+1} and d_{t+1} and cannot be solved forward into a present-value expression. The entire purpose of the Campbell–Shiller approximation is to replace

that nonlinear term with a linear one, at which point the identity becomes a linear difference equation in p_t that *can* be iterated.

2.2 The first-order expansion

Write $\delta_t \equiv d_t - p_t$ for the log dividend–price ratio and let $f(x) = \log(1+e^x)$ be the offending function. Expanding f to first order around the sample mean $\overline{d-p}$ of the log dividend–price ratio and collecting terms (Section A gives every step) yields the workhorse relation

$$r_{t+1} \approx k + \rho\, p_{t+1} + (1-\rho)\, d_{t+1} - p_t, \tag{3}$$

where

$$\rho = \frac{1}{1+\exp(\overline{d-p})}, \qquad k = -\log(\rho) - (1-\rho)\log\left[\frac{1}{\rho} - 1\right]. \tag{4}$$

Three features of Eq. (3) deserve comment.

First, the log return is now a *linear* combination of today's log price, tomorrow's log price, and tomorrow's log dividend. That linearity is the whole point.

Second, ρ is not a free parameter and not something to be estimated. It is pinned down by the average dividend–price ratio in the sample through Eq. (4). Because the average dividend–price ratio is a small number, $\overline{d-p}$ is a large negative number, $\exp(\overline{d-p})$ is close to zero, and ρ is close to but strictly below one. For monthly post-war U.S. data, ρ is typically around 0.997, and at annual frequency, it is typically around 0.96 to 0.97. The weights ρ and $1-\rho$ in Eq. (3) sum to one, so ρ is best read as the share of the asset's value coming from capital gains, with $1-\rho$ the share coming from the dividend. That ρ is close to one is not a technicality — it is why the decomposition weights the very distant future

almost as heavily as next period, and why persistent predictors matter so much.

Third, k is a constant. It carries no information and can be ignored in everything that follows, because the decomposition operates on *innovations* and constants have zero innovation.

> **Warning**
>
> The approximation error in Eq. (3) is second order in the deviation of δ_{t+1} from its mean. It is small in samples where the dividend–price ratio does not wander far, which describes aggregate U.S. data reasonably well but describes individual firms and short samples much less well. The approximation is not the fragile part of the exercise — the VAR specification in Section 4 is.

2.3 Solving forward

Rearranging Eq. (3) to isolate today's price gives a first-order stochastic difference equation:

$$p_t \approx k + \rho\, p_{t+1} + (1-\rho)\, d_{t+1} - r_{t+1}. \tag{5}$$

Now substitute the same expression for p_{t+1} into the right side, then for p_{t+2}, and so on. After J substitutions,

$$p_t = \sum_{j=0}^{J-1} \rho^j \Big[k + (1-\rho)\, d_{t+1+j} - r_{t+1+j} \Big] + \rho^{J} p_{t+J}. \tag{6}$$

Eq. (6) is still an identity for any finite J — no economics has entered. Letting $J \to \infty$ requires the final term to vanish, which is the no-bubble transversality condition

$$\lim_{j \to \infty} \rho^j p_{t+j} = 0. \tag{7}$$

Eq. (7) is the one genuinely economic assumption in this section. It rules out rational bubbles — price paths that grow fast enough forever that they are self-justifying without reference to any cash flow. Imposing it and summing the constant term gives the canonical Campbell and Shiller (1988) price decomposition:

$$p_t = \sum_{j=0}^{\infty} \rho^j (1-\rho)\, d_{t+1+j} - \sum_{j=0}^{\infty} \rho^j r_{t+1+j} + \frac{k}{1-\rho}. \tag{8}$$

Read Eq. (8) slowly, because it is the entire content of the decomposition. The log price today is a discounted sum of all future log dividends *minus* a discounted sum of all future log returns, plus a constant. The dividend sum is the "cash flow" term and enters positively. The return sum is the "discount rate" term and enters negatively: for a *given* stream of dividends, higher future returns can only be delivered by a lower price today.

Note that Eq. (8) holds *ex post*, realization by realization, not just in expectation. It is an identity linking realized paths. Taking E_t of both sides, and using $E_t p_t = p_t$, gives the expectational version stating that today's price reflects expected dividends discounted at expected returns.

3. Cash Flow News and Discount Rate News

3.1 The dividend–price form

It is convenient to restate Eq. (3) in terms of the log dividend–price ratio. Adding and subtracting d_t and d_{t+1} in Eq. (3) and collecting terms gives

$$r_{t+1} \approx k + \delta_t - \rho\, \delta_{t+1} + \Delta d_{t+1}, \tag{9}$$

where $\Delta d_{t+1} = d_{t+1} - d_t$ is log dividend growth. Eq. (9) says that the realized return

equals dividend growth, plus the change in valuation as measured by the dividend–price ratio, plus a constant. Solving Eq. (9) forward under the analogous condition $\lim_{j\to\infty} \rho^j \delta_{t+j} = 0$ gives

$$\delta_t = -\frac{k}{1-\rho} + \sum_{j=0}^{\infty} \rho^j (r_{t+1+j} - \Delta d_{t+1+j}). \tag{10}$$

Eq. (10) is Cochrane (2011)'s central point in one line. A high dividend–price ratio — a cheap market — must forecast either high future returns or low future dividend growth. There is no third possibility, and the identity gives no guidance as to which it is. That is an empirical question.

3.2 Taking innovations

Now apply the innovation operator $E_{t+1} - E_t$ to Eq. (9). Both k and δ_t are known at time t, so their innovations are zero, leaving

$$r_{t+1} - E_t r_{t+1} = (E_{t+1} - E_t)\, \Delta d_{t+1} - \rho\, (E_{t+1} - E_t)\, \delta_{t+1}. \tag{11}$$

Substituting Eq. (10) one period forward for δ_{t+1}, and re-indexing the two resulting sums so the $j = 0$ dividend term folds in (Section B), gives the Campbell (1991) decomposition of the return innovation:

$$r_{t+1} - E_t r_{t+1} = (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j\, \Delta d_{t+1+j} - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j\, r_{t+1+j}. \tag{12}$$

Name the three objects in Eq. (12):

$$\eta^r_{t+1} = r_{t+1} - E_t r_{t+1} \qquad \text{(return innovation)}, \tag{13}$$

$$\eta^{\mathrm{CF}}_{t+1} = (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \, \Delta d_{t+1+j} \qquad \text{(cash flow news)}, \tag{14}$$

$$\eta^{\mathrm{DR}}_{t+1} = (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j \, r_{t+1+j} \qquad \text{(discount rate news)}, \tag{15}$$

so that Eq. (12) reads

$$\eta^{r}_{t+1} = \eta^{\mathrm{CF}}_{t+1} - \eta^{\mathrm{DR}}_{t+1}. \tag{16}$$

3.3 Reading the signs and the indices

Two details of Eq. (16) trip people up, and both have clean explanations.

Why discount rate news enters negatively. Good news about future expected returns is bad news for the current price. If the market learns at $t+1$ that returns from $t+2$ onward will be higher, and the expected dividend stream is unchanged, the only way to deliver those higher returns is for the price to fall today. The investor holding the asset through the news takes a capital loss now in exchange for better prospects later. A positive η^{DR} therefore *reduces* the realized return η^{r}, which is exactly the minus sign in Eq. (16). Cash flow news needs no such reversal: better cash flows raise today's price and leave expected returns alone.

Why the two sums start at different indices. The cash flow sum runs from $j = 0$, so it includes the revision in expectations about the *contemporaneous* dividend Δd_{t+1}. The discount rate sum runs from $j = 1$, excluding r_{t+1}. This is not an arbitrary convention. The revision in expectations about r_{t+1} is $r_{t+1} - E_t r_{t+1}$, which is the left-hand side of Eq. (12). It cannot also appear on the right without making the equation circular. Put differently,

η^{DR} is news about returns *beyond* the current period, which is precisely the object with implications for future performance.

A useful mnemonic for persistence. Because $\rho \approx 1$, both sums are close to unweighted sums over the entire future. A small but *persistent* revision in expected returns therefore produces large discount rate news, while a large but transitory revision produces very little. This is why persistent predictors such as the dividend–price ratio load so heavily on the discount rate channel, and it is a standing reason to be careful with the persistence of whatever goes into the VAR.

3.4 The variance decomposition

Taking the variance of Eq. (16) gives the identity that names the literature:

$$\mathrm{Var}(\eta^{r}) = \mathrm{Var}(\eta^{\mathrm{CF}}) + \mathrm{Var}(\eta^{\mathrm{DR}}) - 2\,\mathrm{Cov}(\eta^{\mathrm{CF}}, \eta^{\mathrm{DR}}). \tag{17}$$

Applied to aggregate U.S. equity returns, Campbell (1991) and Campbell and Ammer (1993) find that discount rate news accounts for the bulk of the variance of unexpected market returns, with cash flow news playing a modest role. At the individual-firm level, Vuolteenaho (2002) finds the reverse, with cash flow news dominating, because idiosyncratic cash flow shocks are large and largely diversified away in the aggregate. The application in Section 5 concerns neither of these — it asks not which component is *larger*, but which component a particular predictor *forecasts*.

4. Extracting the News with a VAR

4.1 The identification problem

Eq. (14) and Eq. (15) are not estimable as written. Each is a revision in a conditional expectation of an infinite sum, and conditional expectations are not observed. What is needed is a model of how the market forms forecasts. The Campbell (1991)/Campbell and Ammer (1993) solution is to assume investors forecast using a first-order vector autoregression in a chosen state vector, and to treat the VAR's linear projections as the market's conditional expectations.

This substitution is the load-bearing assumption of the whole exercise, and it is worth being explicit that it is an assumption, not a derivation. Everything up to Eq. (16) is an identity plus a Taylor expansion. Everything from here on is conditional on the VAR being a correct description of the information investors actually use.

4.2 The VAR and its state vector

Let

$$y_{t+1} = A\, y_t + u_{t+1}, \qquad y_t = (r_t,\ d_t - p_t,\ z_t')', \tag{18}$$

where z_t is an n-vector of additional predictors, A is the $(n+2) \times (n+2)$ matrix of slope coefficients, and u_{t+1} is the $(n+2)$-vector of zero-mean innovations. The constant is suppressed, which costs nothing because the decomposition works on innovations.

The ordering matters only for notational convenience. Placing the log return first means the selector vector

$$e_1 = (1,\, 0,\, \ldots,\, 0)' \tag{19}$$

picks the return row out of any vector or matrix product.

The recursive structure of Eq. (18) gives everything else. Iteration gives $E_t y_{t+1+j} = A^{j+1} y_t$, and therefore the revision in expectations between t and $t+1$ about the state j periods after $t+1$ is

$$(E_{t+1} - E_t)\, y_{t+1+j} = A^j\, u_{t+1}, \qquad j \geq 0. \tag{20}$$

Eq. (20) is the workhorse. The entire revision in expectations, at every horizon, is driven by the single innovation vector u_{t+1}, propagated forward by powers of A.

4.3 Closed forms for the news components

The return innovation is immediate from Eq. (20) at $j = 0$:

$$\eta^r_{t+1} = e_1'\, u_{t+1}. \tag{21}$$

Discount rate news follows by substituting Eq. (20) into Eq. (15) and summing the resulting matrix geometric series, $\sum_{j=1}^{\infty} (\rho A)^j = \rho A (I - \rho A)^{-1}$:

$$\eta^{\mathrm{DR}}_{t+1} = e_1'\, \rho A\, (I - \rho A)^{-1}\, u_{t+1}. \tag{22}$$

The series converges provided every eigenvalue of ρA lies inside the unit circle. Since ρ is just below one, this is effectively a requirement that the VAR itself be stationary. With highly persistent regressors — and the dividend–price ratio is about as persistent as macro-finance data gets — the largest eigenvalue of $\hat{A}$ can sit uncomfortably close to $1/\rho$, in which

case $(I-\rho\hat{A})^{-1}$ blows up and $\hat{\eta}^{\mathrm{DR}}$ becomes enormous and unstable. Checking the eigenvalues of $\rho\hat{A}$ is a mandatory diagnostic, not an optional one.

Cash flow news is then obtained *residually* from Eq. (16):

$$\eta^{\mathrm{CF}}_{t+1} = \eta^{r}_{t+1} + \eta^{\mathrm{DR}}_{t+1} = e_1' \Big[I + \rho A (I - \rho A)^{-1} \Big] u_{t+1}. \tag{23}$$

Finally, the expected return implied by the VAR is

$$E_t\, r_{t+1} = e_1'\, A\, y_t, \tag{24}$$

and the realized return decomposes into three estimable pieces:

$$r_{t+1} = E_t r_{t+1} + \eta^{\mathrm{CF}}_{t+1} - \eta^{\mathrm{DR}}_{t+1}. \tag{25}$$

4.4 Why cash flow news is a residual

Eq. (23) backs out cash flow news as whatever is left over rather than modeling dividend growth directly. This is deliberate, and it is a genuine trade-off.

The case for it: the state vector in Eq. (18) need not contain dividend growth at all. Any variable that forecasts returns contributes to η^{DR}, and the identity Eq. (16) then delivers η^{CF} for free. Modeling cash flows directly would require taking a stand on the dividend process, which for aggregate data is badly complicated by payout policy shifts, share repurchases, and seasonality.

The case against it: every specification error in the VAR lands in η^{CF}. If the return-forecasting model is misspecified, $\hat{\eta}^{\mathrm{DR}}$ is wrong, and because $\hat{\eta}^{r}$ is essentially pinned down by the data, the error transfers one-for-one into $\hat{\eta}^{\mathrm{CF}}$. The residual component is therefore the *least* reliable of the two, and it is unfortunately often the one the researcher cares about.

Chen and Zhao (2009) press this point hard, showing that residually computed cash flow news is sensitive to the choice of state variables in ways that direct modeling is not.

4.5 Specification choices that matter

Include the log dividend–price ratio. This is not optional. Engsted et al. (2012) show that the decomposition is only internally consistent when the state vector includes the variable that appears in the log-linearization itself — here $d_t - p_t$. Omit it and the estimated $\hat{\eta}^{\text{CF}}$ and $\hat{\eta}^{\text{DR}}$ no longer correspond to the objects defined in Eq. (14) and Eq. (15), even asymptotically. Note that Rapach et al. (2016) include the log dividend–price ratio in *every* one of their VAR specifications for exactly this reason.

Calibrate, do not estimate, ρ. Compute ρ from Eq. (4) using the sample mean of $d_t - p_t$. Treating it as a free parameter is a category error — it is a linearization constant, not a preference or technology parameter. Calibrating it is not the same as its being innocuous, though: because ρ enters through $(I - \rho A)^{-1}$, small changes in it produce large changes in the news components, and it must be measured at the frequency of the data. Section 5.5 shows what is at stake empirically.

Higher-order VARs are still first-order. Eq. (18) is written as a VAR(1) with no loss of generality. A VAR(p) is rewritten in companion form, stacking p lags into an expanded state vector, and every formula above applies unchanged with A replaced by the companion matrix. Section B records the mapping.

Persistence and small-sample bias. The regressors are persistent and their innovations are strongly correlated with return innovations, which is the Stambaugh (1999) setup. The OLS estimate of A is biased in small samples, the bias propagates nonlinearly

through $(I - \rho\hat{A})^{-1}$, and standard errors on the news components computed as though $\hat{A}$ were known are too small. Inference should be bootstrapped, and Rapach et al. (2016) use heteroskedasticity- and autocorrelation-robust statistics throughout.

Keep the state vector small. Each added predictor adds $n + 2$ parameters to A. Estimating a VAR with fifteen state variables on five hundred monthly observations produces a $\hat{A}$ so imprecise that the resulting news series are mostly estimation noise. The standard workaround, used by Rapach et al. (2016) following Goyal and Welch (2008), is to run many small VARs — the return, the dividend–price ratio, and *one* predictor at a time — and separately a compact VAR using the first few principal components extracted from the full predictor set. Consistency of the answer across those specifications is the evidence, not any single one of them.

5. Application: What Does Short Interest Predict?

5.1 The question

Rapach et al. (2016) document that aggregate short interest is arguably the strongest known predictor of the market equity risk premium. They build a short interest index (SII) as the equal-weighted mean of firm-level short interest across U.S.-listed stocks, detrended to strip out the secular growth of the equity lending market and standardized to unit variance. A one-standard-deviation increase in SII forecasts roughly a six to seven percentage point decline in the annualized market excess return, with an in-sample R^2 of 12.89% at the annual horizon.

The predictive result on its own is silent about mechanism. A variable that forecasts

returns is doing one of two things. It may be tracking the time-varying risk premium, in which case it carries information about *discount rates* and implies nothing about anybody being informed. Or it may be anticipating future *cash flows*, in which case whoever generates the signal knows something about fundamentals that the market has not yet priced. The decomposition is exactly the tool that separates these.

5.2 Component predictive regressions

The strategy is to run the same predictive regression four times: once on the total return, and once on each of the three pieces into which Eq. (25) splits it. First the benchmark,

$$r_{t+1} = \alpha + \beta\, \mathrm{SII}_t + \varepsilon_{t+1}, \tag{26}$$

and then, using the fitted components from the VAR,

$$\hat{E}_t r_{t+1} = \alpha_E + \beta_E\, \mathrm{SII}_t + \varepsilon^E_{t+1}, \tag{27}$$

$$\hat{\eta}^{\mathrm{CF}}_{t+1} = \beta_{\mathrm{CF}}\, \mathrm{SII}_t + \varepsilon^{\mathrm{CF}}_{t+1}, \tag{28}$$

$$\hat{\eta}^{\mathrm{DR}}_{t+1} = \beta_{\mathrm{DR}}\, \mathrm{SII}_t + \varepsilon^{\mathrm{DR}}_{t+1}. \tag{29}$$

Eq. (28) and Eq. (29) carry no intercept because the news components have mean zero by construction — they are innovations.

5.3 The adding-up identity

The four slope estimates are not independent. They satisfy

$$\hat{\beta} = \hat{\beta}_E + \hat{\beta}_{\mathrm{CF}} - \hat{\beta}_{\mathrm{DR}} \tag{30}$$

exactly, in every sample, as a matter of arithmetic. The reason is worth stating because it makes clear what the exercise can and cannot show. Eq. (25) holds observation by observation for the fitted components, so the realized return is the exact sum $\hat{E}_t r_{t+1} + \hat{\eta}^{\mathrm{CF}}_{t+1} - \hat{\eta}^{\mathrm{DR}}_{t+1}$. The OLS slope on a fixed regressor is a linear operator on the dependent variable, $\hat{\beta}_y = \widehat{\mathrm{Cov}}(y_{t+1}, \mathrm{SII}_t)/\widehat{\mathrm{Var}}(\mathrm{SII}_t)$, and a linear operator applied to a sum returns the sum of the parts.

So Eq. (30) is not a testable restriction and its holding is not evidence of anything. What it *does* provide is a clean accounting: it tells you how the total predictive slope $\hat{\beta}$ is apportioned across channels, and the question becomes which term on the right accounts for most of the left.

> **Note**
>
> A caveat Rapach et al. (2016) flag: this clean apportionment is available for the slope coefficient but not for the R^2. Because the three components are correlated with one another, their R^2 contributions do not add up, and no analogous decomposition of predictive R^2 into expected-return, cash-flow, and discount-rate parts is available.

5.4 Replication

Rapach et al. (2016) estimate Eq. (26) through Eq. (29) on monthly data from 1973:01 to 2014:12, repeating the exercise for fifteen VAR specifications: the bivariate VAR in the log return and the log dividend–price ratio, thirteen trivariate VARs adding one of the remaining Goyal and Welch (2008) predictors, and one adding the first three principal components of the full predictor set.

Table 1 reports my replication of their Table 8. The construction follows the paper exactly. The state vector orders the S&P 500 log *total* return first, then the log dividend–price ratio, then the additional predictor. Note that the decomposed variable is the total return rather than the excess return used in the predictive-regression tables, which is why the total slope here, $\hat{\beta} = -0.51$ with a robust t-statistic of -2.53, differs slightly from the -0.50 obtained for the excess return. The VAR is estimated by OLS equation by equation on the 503 usable observations, ρ is calibrated from the sample mean log dividend–price ratio as in Eq. (4), and the news components follow from Eq. (22) and Eq. (23). All coefficients are in percent, and t-statistics are heteroskedasticity- and autocorrelation-robust.

Table 1. Predictive regressions for market return components, 1973:01–2014:12.

Note: The table replicates Table 8 of Rapach et al. (2016), decomposing the total predictive slope of the short interest index into expected-return, cash-flow-news, and discount-rate-news components. DP is the log dividend–price ratio, which appears in every specification, and the remaining labels follow Goyal and Welch (2008). PC denotes the first three principal components of the fourteen popular predictors. Robust t-statistics in brackets. Every row satisfies the adding-up identity in Eq. 30 to within 0.0001.

VAR variables	$\hat{\beta}_E$	$\hat{\beta}_{\mathrm{CF}}$	$\hat{\beta}_{\mathrm{DR}}$
r, DP	−0.07 [−3.13]	−0.35 [−2.27]	0.09 [2.03]
r, DP, DY	−0.06 [−2.53]	−0.35 [−2.27]	0.10 [2.23]
r, DP, EP	−0.08 [−3.52]	−0.40 [−2.36]	0.04 [0.98]
r, DP, DE	−0.08 [−3.52]	−0.40 [−2.36]	0.04 [0.98]
r, DP, RVOL	−0.11 [−3.64]	−0.27 [−1.95]	0.13 [1.91]
r, DP, BM	+0.01 [+0.50]	−0.33 [−2.35]	0.19 [2.89]
r, DP, NTIS	−0.05 [−2.49]	−0.36 [−2.32]	0.10 [2.15]
r, DP, TBL	−0.09 [−3.69]	−0.31 [−2.08]	0.11 [1.44]
r, DP, LTY	−0.08 [−3.43]	−0.32 [−2.20]	0.11 [1.60]
r, DP, LTR	−0.07 [−2.71]	−0.35 [−2.28]	0.09 [1.92]
r, DP, TMS	−0.08 [−3.53]	−0.34 [−2.21]	0.09 [1.64]
r, DP, DFY	−0.07 [−3.15]	−0.36 [−2.27]	0.09 [1.98]
r, DP, DFR	−0.09 [−2.74]	−0.34 [−2.20]	0.08 [1.92]
r, DP, INFL	−0.07 [−3.24]	−0.35 [−2.27]	0.09 [1.98]
r, DP, PC	−0.05 [−1.97]	−0.31 [−2.20]	0.15 [2.23]

All forty-five coefficients and t-statistics reproduce the published table. Checked against the authors' own stored output rather than the rounded printed table, the largest discrepancy across all ninety statistics is 9×10^{-11}.

Reaching that agreement required matching one estimation detail that the paper does not mention: the authors demean every column of the state vector by its *full-sample* mean and then run the VAR regressions without an intercept. This is not the same as including an intercept, which effectively demeans using the means of the subsamples actually used — observations $2, \ldots, T$ on the left and $1, \ldots, T-1$ on the right. The two differ because those subsample means are not equal. The gap is small but not invisible: estimating with an intercept moves the coefficients by about 10^{-4} and shifts four of the reported t-statistics by 0.01 in the second decimal.

Three readings follow.

The cash flow channel dominates. In every specification $\hat{\beta}_{\text{CF}}$ is the largest term in absolute value, running between roughly -0.27 and -0.40, and it is statistically significant throughout. High short interest forecasts *negative cash flow news*. Short sellers are anticipating that aggregate fundamentals will disappoint.

The expected-return channel is negligible. The $\hat{\beta}_E$ estimates are frequently significant but tiny, clustered near -0.07. SII barely moves the VAR's fitted conditional expected return, which is another way of saying SII is close to orthogonal to the popular predictors — consistent with the fact that its largest correlation with any of the fourteen is only -0.28.

The discount rate channel is present but small. The $\hat{\beta}_{\text{DR}}$ estimates are positive, around 0.04 to 0.19, and enter Eq. (30) with a minus sign, so they do push $\hat{\beta}$ in the observed

negative direction. But their magnitude is a fraction of the cash flow term. Some of what short interest forecasts is a rise in required returns, but most of it is bad news about cash flows.

The economic interpretation Rapach et al. (2016) draw is that short sellers are informed traders operating at the *macroeconomic* level. The existing literature had established that short sellers process firm-specific information skillfully. What the decomposition adds is that they also anticipate aggregate cash flows, and that this — rather than any risk-premium story — is what their aggregate positioning reveals.

5.5 The split is sensitive to ρ

The published table rests on a calibration choice for ρ that the paper does not spell out, and the replication turns out to be informative about how much that choice matters.

The dividend series in the Goyal and Welch (2008) data is `D12`, a twelve-month moving sum, so `D12`$/P$ is an *annual* dividend–price ratio even though the VAR runs at monthly frequency. The authors' replication code calibrates Eq. (4) directly from it,

```
rho_hat = 1/(1+exp(mean(log_DP)));
```

with `log_DP = log(D12./SP)`, giving $\rho = 0.9738$. The frequency-consistent alternative divides the dividend sum by twelve before forming the ratio, giving a monthly $\rho = 0.9978$ — much closer to one, as a monthly discount coefficient should be. Note that $\rho = 0.9738$ implies a half-life of about twenty-six months for the weights ρ^j, so the news sums are effectively truncated at a horizon of a few years rather than running over the indefinite future the theory in Section 3 specifies.

Table 2 reports the (r, DP) decomposition across a grid of ρ values spanning both.

Table 2. Sensitivity of the (r, DP) decomposition to ρ.

Note: The row $\rho = 0.9738$ is the published calibration, taken from the annual dividend–price ratio, and the row $\rho = 0.9978$ is the frequency-consistent monthly one. The Total column is $\hat{\beta}_E + \hat{\beta}_{\mathrm{CF}} - \hat{\beta}_{\mathrm{DR}}$.

ρ	$\hat{\beta}_E$	$\hat{\beta}_{\mathrm{CF}}$	$\hat{\beta}_{\mathrm{DR}}$	Total
0.9500	−0.07	−0.40	0.04	−0.51
0.9600	−0.07	−0.39	0.05	−0.51
0.9700	−0.07	−0.37	0.08	−0.51
0.9738	−0.07	−0.35	0.09	−0.51
0.9800	−0.07	−0.32	0.12	−0.51
0.9900	−0.07	−0.23	0.21	−0.51
0.9950	−0.07	−0.11	0.33	−0.51
0.9978	−0.07	0.02	0.46	−0.51

Three things are visible in Table 2, and each one illustrates a point made earlier on theoretical grounds.

The total is invariant. Every row sums to −0.51. This is Eq. (30) doing exactly what Section 5.3 said it would: the identity is arithmetic, holds for any ρ, and therefore carries no information. A replication that checked only the adding-up would report success at every ρ in the table.

The expected-return channel is invariant too. The $\hat{\beta}_E$ column never moves, because Eq. (24) contains no ρ. The conditional expected return depends only on the VAR, so this column is a genuine check on the VAR specification, independent of the linearization. That it matched the published values before ρ was correctly calibrated was what localized the discrepancy.

The split between the two channels is not invariant. It moves across the full range of the table, and at the monthly calibration $\hat{\beta}_{\mathrm{CF}}$ changes sign. Under $\rho = 0.9978$

the conclusion reverses: short interest would be forecasting discount rates, not cash flows. The mechanism is the one flagged in Section 4.3. The largest eigenvalue of ρA rises from 0.97 to 0.995 as ρ goes from 0.9738 to 0.9978, so $(I - \rho A)^{-1}$ grows sharply, discount rate news inflates, and because cash flow news is computed residually by Eq. (23) the inflation transfers into it one-for-one with the opposite sign.

> ⚠ Warning
>
> This is not a claim that the published result is wrong. A monthly ρ of 0.9978 pushes the spectral radius of $\rho\hat{A}$ to 0.995, close enough to unity that the news components are dominated by a near-unit root in an estimated matrix, which is precisely the configuration Section 4.5 warns is unreliable. The lower ρ effectively truncates the news sums at a horizon where the VAR is still informative. But the choice does the work, it is not innocuous, and a decomposition result of this kind should be reported alongside its sensitivity to ρ rather than at a single calibration.

6. Caveats

The decomposition is only as good as the VAR, and Rapach et al. (2016) are candid about the resulting joint hypothesis problem. Their own statement of the alternative interpretation is the sharpest one:

> An alternative explanation is that the predictive ability of SII relates to the time-varying equilibrium aggregate risk premium. Under this interpretation, the VAR in Section 5 is misspecified: by excluding SII from the variables appearing in the VAR, we effectively exclude SII from the

market information set and the VAR's estimate of the expected return.

The logic is worth spelling out. Suppose SII genuinely does track the equilibrium risk premium, and investors use it. Then the correct state vector includes SII. Because the estimated VAR omits it, the fitted $\hat{E}_t r_{t+1}$ is missing precisely the variation SII drives, and that variation has to go somewhere — it lands in the residual u_{t+1} and hence in the estimated news components. The finding that SII forecasts cash flow news could then be an artifact of having left SII out of the information set.

Note the shape of this objection. Including SII in the VAR does not resolve it. If SII is in the state vector, it becomes uncorrelated with the news components *by construction*, so the component regressions are guaranteed to return zero and the test has no content either way. There is no specification that discriminates cleanly, which is the joint hypothesis problem in its standard form: any test of what a predictor forecasts is a joint test of that and of the model of expectations.

Rapach et al. (2016) argue the informed-trading interpretation is the more plausible of the two, on the grounds that SII is largely orthogonal to the popular predictors normally thought to proxy for the risk premium, so a risk-premium story would need to explain why this particular risk-premium proxy is uncorrelated with all the others. That is an argument from plausibility, not a test, and they say so.

Two more limitations carry over from Section 4. Cash flow news is computed residually, so it absorbs all specification error, which is a particular concern when cash flow news is the headline result. And the persistence of the state variables means both the point estimates and the standard errors deserve bootstrap treatment rather than asymptotic approximations.

7. Conclusion

The chain of reasoning runs as follows. The definition of a return is an identity. Log-linearizing it around the mean dividend–price ratio produces a linear difference equation whose only free constant, ρ, is calibrated from the data rather than estimated. Solving that equation forward under a no-bubble condition gives the price as a discounted sum of future dividends minus a discounted sum of future returns. Differencing the expectation of that sum between adjacent periods splits the unexpected return into cash flow news and discount rate news, with the latter entering negatively because higher required returns are a capital loss today and a gain tomorrow. All of this is identity plus approximation.

The empirical content enters with the VAR, which supplies the conditional expectations the identity requires but does not provide. Given a VAR, both news components are closed-form linear functions of a single innovation vector, and the entire decomposition reduces to the matrix $\rho A(I - \rho A)^{-1}$. Given the components, asking what any predictor forecasts becomes four ordinary regressions whose slopes add up exactly.

Applied to aggregate short interest, the answer is that short sellers forecast cash flows. Replicating that result exactly, and then perturbing the one calibration it rests on, sharpens what the claim can bear. The total predictive slope and the expected-return channel are pinned down by the VAR alone. The division of the remainder between cash flow and discount rate news is not — it depends on ρ, and at the calibration ordinarily called for at a monthly frequency, it goes the other way. The conclusion is also conditional on the VAR representing the market's information set, which cannot be verified. The honest statement is that short sellers forecast cash flows if one accepts the standard predictors as a proxy for

what the market knows and accepts a discount coefficient that truncates the news sums at a horizon over which the VAR remains informative.

References

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

Campbell, J. Y. and Ammer, J. (1993). What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns. *Journal of Finance*, 48(1):3–37.

Campbell, J. Y. and Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies*, 1(3):195–228.

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

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4):1047–1108.

Engsted, T., Pedersen, T. Q., and Tanggaard, C. (2012). Pitfalls in VAR Based Return Decompositions: A Clarification. *Journal of Banking and Finance*, 36(5):1255–1265.

Goyal, A. and Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 21(4):1455–1508.

Rapach, D. E., Ringgenberg, M. C., and Zhou, G. (2016). Short Interest and Aggregate Stock Returns. *Journal of Financial Economics*, 121(1):46–65.

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics*, 54(3):375–421.

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

Appendix A. Deriving ρ and k

Start from the exact expression Eq. (2),

$$r_{t+1} = p_{t+1} - p_t + \log(1 + e^{\,\delta_{t+1}}), \tag{A.1}$$

where $\delta_{t+1} = d_{t+1} - p_{t+1}$. Define $f(x) = \log(1+e^x)$ and let $\bar{\delta} = \overline{d-p}$ denote the sample mean of the log dividend–price ratio. The first derivative is

$$f'(x) = \frac{e^x}{1+e^x}. \tag{A.2}$$

Now *define*

$$\rho \equiv \frac{1}{1+e^{\bar{\delta}}}, \tag{A.3}$$

which is Eq. (4). Two consequences follow immediately. From Eq. (A.2) evaluated at $\bar{\delta}$,

$$f'(\bar{\delta}) = \frac{e^{\bar{\delta}}}{1+e^{\bar{\delta}}} = 1 - \frac{1}{1+e^{\bar{\delta}}} = 1 - \rho, \tag{A.4}$$

and from Eq. (A.3), $1 + e^{\bar{\delta}} = 1/\rho$, so the level is

$$f(\bar{\delta}) = \log(1 + e^{\bar{\delta}}) = \log(1/\rho) = -\log\rho. \tag{A.5}$$

The first-order Taylor expansion of f about $\bar{\delta}$ is therefore

$$f(\delta_{t+1}) \approx -\log\rho + (1-\rho)(\delta_{t+1} - \bar{\delta}). \tag{A.6}$$

Substituting Eq. (A.6) into Eq. (A.1) and expanding $\delta_{t+1} = d_{t+1} - p_{t+1}$:

$$\begin{aligned} r_{t+1} &\approx p_{t+1} - p_t - \log\rho + (1-\rho)(d_{t+1} - p_{t+1}) - (1-\rho)\bar{\delta} \\ &= \underbrace{-\log\rho - (1-\rho)\bar{\delta}}_{k} + \underbrace{p_{t+1} - (1-\rho)p_{t+1}}_{\rho\, p_{t+1}} + (1-\rho)\, d_{t+1} - p_t, \end{aligned} \tag{A.7}$$

which is Eq. (3) with $k = -\log\rho - (1-\rho)\bar{\delta}$. Finally, inverting Eq. (A.3) gives $e^{\bar{\delta}} = 1/\rho - 1$, hence $\bar{\delta} = \log[1/\rho - 1]$, and substituting this into the expression for k recovers the form printed in Eq. (4):

$$k = -\log(\rho) - (1-\rho)\log\left[\frac{1}{\rho} - 1\right]. \tag{A.8}$$

Note that k depends only on ρ, which in turn depends only on the sample mean dividend–price ratio. Neither is estimated.

Appendix B. Companion Form and the Geometric Sum

Appendix B.1 Re-indexing the innovation sums

Lead Eq. (10) one period and take innovations. Since the constant drops out,

$$(E_{t+1} - E_t)\,\delta_{t+1} = (E_{t+1} - E_t)\sum_{j=0}^{\infty}\rho^j(r_{t+2+j} - \Delta d_{t+2+j}). \tag{B.1}$$

Substituting into Eq. (11),

$$\begin{aligned}\eta^r_{t+1} &= (E_{t+1} - E_t)\Delta d_{t+1} + \rho\,(E_{t+1} - E_t)\sum_{j=0}^{\infty}\rho^j\,\Delta d_{t+2+j} \\ &\quad - \rho\,(E_{t+1} - E_t)\sum_{j=0}^{\infty}\rho^j\,r_{t+2+j}.\end{aligned} \tag{B.2}$$

For the dividend terms, shifting the summation index by $i = j + 1$ gives $\rho\sum_{j=0}^{\infty}\rho^j\Delta d_{t+2+j} = \sum_{i=1}^{\infty}\rho^i\Delta d_{t+1+i}$, which combines with the standalone $j = 0$ term $(E_{t+1} - E_t)\Delta d_{t+1}$ to give the single sum $\sum_{j=0}^{\infty}\rho^j\Delta d_{t+1+j}$. The same index shift on the return terms gives $\rho\sum_{j=0}^{\infty}\rho^j r_{t+2+j} = \sum_{j=1}^{\infty}\rho^j r_{t+1+j}$, which starts at $j = 1$ because there is

no standalone $j = 0$ return term to absorb. This is the origin of the index asymmetry discussed in Section 3.3, and together the two give Eq. (12).

Appendix B.2 The matrix geometric series

Substituting Eq. (20) into Eq. (15),

$$\eta_{t+1}^{\mathrm{DR}} = e_1' \sum_{j=1}^{\infty} \rho^j A^j \, u_{t+1} = e_1' \left[\sum_{j=1}^{\infty} (\rho A)^j \right] u_{t+1}. \tag{B.3}$$

Let $S = \sum_{j=1}^{\infty} (\rho A)^j$. Then $S = \rho A + \rho A \, S$, so $(I - \rho A) S = \rho A$ and

$$S = \rho A \, (I - \rho A)^{-1}, \tag{B.4}$$

which is Eq. (22). The manipulation requires $\lim_{j \to \infty} (\rho A)^j = 0$, equivalently that the spectral radius of ρA be strictly less than one.

Appendix B.3 Higher-order VARs

A VAR(p), $x_{t+1} = A_1 x_t + \cdots + A_p x_{t-p+1} + v_{t+1}$ in the m-vector x_t, is written in companion form by stacking lags into $y_t = (x_t', x_{t-1}', \ldots, x_{t-p+1}')'$ of dimension mp, giving $y_{t+1} = A y_t + u_{t+1}$ with

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{pmatrix}, \qquad u_{t+1} = \begin{pmatrix} v_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{B.5}$$

Every expression in Section 4 applies verbatim with this A and u_{t+1}, provided e_1 is

the mp-vector selecting the log return, which by construction sits in the first block. The convergence requirement becomes a condition on the eigenvalues of the companion matrix scaled by ρ — the same stationarity check, applied to a larger matrix.