

Taming the VAR

Aizhan Anarkulova* Scott Cederburg[†] Zhi Da[‡] Yi Zhou[§]

October 15, 2026[¶]

Abstract

Researchers decomposing stock market returns into cash-flow and discount-rate news with a vector autoregression (VAR) face a dilemma. The VAR cannot be estimated precisely with many predictors as state variables; with few, its news estimates depend on which are chosen. We escape this dilemma by sorting 25 predictors into three groups by persistence and using each group’s forecast combination as a state variable. A VAR that recovers true news must produce discount-rate news that forecasts returns and cash-flow news that forecasts dividend growth, with neither adding to the other’s forecast. Ours does, and its estimates are stable across design choices.

JEL classifications: C32, C53, G10, G12, G17

Key words: Return decomposition, forecast combination, vector autoregression

*Goizueta Business School, Emory University. Email: aizhan.anarkulova@emory.edu.

[†]Eller College of Management, University of Arizona. Email: cederburg@arizona.edu.

[‡]Mendoza College of Business, University of Notre Dame. Email: zda@nd.edu.

[§]Monash Business School, Monash University. Email: yi.zhou1@monash.edu.

[¶]We thank Thummim Cho, Andrei Gonçalves, Jonathan Lewellen, Marc Lipson, Sean Myers, Michael O’Doherty, Lin Peng, Richard Sias, Roman Sigalov, Yirong Wang, and seminar participants at the 2024 Financial Management Association Annual Meeting, the 2025 Eastern Finance Association Annual Meeting, the Fourth Annual Valuation Workshop, the Junior Faculty Mentoring Program at the 2026 AFA meeting, Indiana University, and the University of Arizona for helpful comments and suggestions.

1 Introduction

Investor expectations about future cash flows and discount rates determine stock prices. In seminal studies, Campbell and Shiller (1988a, 1988b) and Campbell (1991) formalize this intuition, and their price and return decompositions are ubiquitous in the literature. Many studies use Campbell’s (1991) vector autoregression (VAR) approach to decompose unexpected stock market returns into discount-rate news and cash-flow news components. However, this approach is known to be sensitive to the choice of state variables [e.g., Chen and Zhao (2009)], leading to a crisis of confidence in the estimates. While the return decomposition implies that true news components exist, conventional VAR implementations cast doubt on whether we can accurately recover them from the data.

The sensitivity stems from a dilemma about how many predictors to include in the VAR. The literature offers dozens of return predictors. A VAR that includes many of them cannot be estimated precisely, so its news estimates are noisy. The conventional response is to include just a few. Doing so discards the rest and leaves the news estimates sensitive to the choice of state variables. We escape this dilemma with a low-dimensional VAR that retains information from many predictors. Our method is straightforward to implement. We first sort 25 predictors into three groups based on their persistence and then average their univariate return forecasts within the low-, medium-, and high-persistence groups. These three averages, together with the lagged market return, form a parsimonious VAR with only 16 coefficients, compared with 676 coefficients in a VAR that includes all 25 predictors directly.

The key design choices follow from a decomposition of the mean squared error of the discount-rate news estimate. Retaining more predictors reduces error from the unspanned part of expected returns, and keeping the VAR small limits the sampling variance of the news estimate. Using forecast combination [e.g., Rapach, Strauss, and Zhou (2010)] to distill many predictors into a few state variables accomplishes both, and this averaging also diversifies estimation error in the individual forecasts [Timmermann (2006)]. Averaging has a cost, because a combination imposes a single persistence on predictors whose persistence

differs, and persistence determines how heavily a shock enters discount-rate news. Sorting the predictors by persistence minimizes the resulting bias.

A VAR that recovers the true news components must meet four criteria, and we evaluate conventional VARs and our method against each. Three concern the series the VAR estimates: the expected excess return, discount-rate news, and cash-flow news. The expected return should forecast returns, which the return equation of the VAR tests directly. The news components cannot be tested the same way, because under Campbell’s (1991) construction they sum to the unexpected return for any set of state variables. Their definitions do, however, imply testable restrictions. Discount-rate news should forecast future excess returns and cash-flow news should forecast future dividend growth, and neither should add to the forecast of the other’s target. The fourth criterion is stability across design choices, because the true components do not depend on which reasonable predictors a researcher chooses.

Figure 1 shows how badly conventional VARs fail the stability criterion. As a baseline for comparison throughout the paper, we estimate the 2,298 conventional VARs that each include the log excess market return and three of the 25 predictors, using monthly data from January 1927 to December 2021. The figure plots the histogram of the pairwise time-series correlation of discount-rate news estimates across these models. Among the 2,639,253 model pairs, only 67% achieve positive correlation, which is the bare minimum for success. Nearly half (49%) of correlations fall between -0.2 and 0.2 , with few high correlations. The lack of confidence in conventional VARs is warranted, as two researchers making independent modeling choices are unlikely to estimate discount-rate news series that are statistically or economically similar.

Our method meets the first criterion, that the expected return should forecast returns. The return equation of our VAR has an R^2 of 4.93% for monthly excess market returns, compared with an average of 1.76% across the 2,298 conventional VARs. The R^2 is significant with a p -value of 0.000 from a bootstrap that repeats every stage of the estimation under the null of no return predictability. The gain over conventional VARs is the spanning effect of

retaining more predictors, since the combinations draw on all 25 and the conventional VARs use only three.

Our method also meets the second and third criteria. Following Campbell, Giglio, and Polk (2013), we summarize past news with a trailing moving average and regress annual excess returns and annual dividend growth on the moving averages of both news components. Discount-rate news forecasts excess returns with an R^2 of 20.22% and cash-flow news forecasts dividend growth with an R^2 of 18.13%, both significant at the 1% level. Neither adds to the forecast of the other’s target, with p -values of 0.753 for cash-flow news in the return regression and 0.807 for discount-rate news in the dividend growth regression. The cash-flow result is notable because cash-flow news is not fitted to dividend growth but estimated as a residual, the part of the unexpected return that discount-rate news does not account for.

The pattern survives in real time, with every stage of the method re-estimated each month from data through that month. Real-time discount-rate news forecasts excess returns with an R^2 of 2.92% (coefficient p -value of 0.048) and real-time cash-flow news forecasts dividend growth with an R^2 of 19.89% (coefficient p -value of 0.000), and neither adds to the forecast of the other’s target. Against the 2,298 conventional VARs, our method has a higher in-sample return R^2 and a higher real-time dividend growth R^2 than every one of them, and fewer than 4% beat it on both the in-sample dividend growth R^2 and the real-time return R^2 .

Our method meets the fourth criterion. Its estimates are stable across design choices, and the sensitivity that remains appears where our formal analysis predicts. We vary each design input and correlate the resulting discount-rate news with the base case. The first exercise varies how many predictors are in each combination and whether they are sorted by persistence or drawn at random. With one predictor per combination, the correlations with the base case are widely dispersed with or without the sort, and the random case is the set of conventional VARs in Figure 1. With five predictors per combination, the persistence-sorted models have a median correlation of 0.92 and a tenth percentile of 0.81, whereas the random

models have a median of 0.54 and a tenth percentile of 0.35. Stability requires both many predictors and the persistence sort.

The second exercise varies the number of groups. The estimates are stable for any choice of three to seven persistence groups, with correlations of 0.95 or higher. Instability with fewer groups reflects a sort too coarse to separate heterogeneous predictors, and instability beyond seven reflects the growing VAR dimension, both as predicted. The third exercise varies the persistence cutoffs. The results are insensitive to the low-persistence cutoff, with every correlation above 0.97, and sensitive to the high-persistence cutoff only when it leaves the high-persistence group with too few predictors or too much heterogeneity. That group carries much of the weight in discount-rate news, so it is the one that must be both narrow in persistence and well populated.

With these criteria met, our estimates address the two questions a return decomposition is meant to answer. Discount-rate news is the larger component, with a standard deviation of 5.90% per month (90% bootstrap confidence interval of 5.08% to 7.13%) versus 2.84% for cash-flow news (2.60% to 3.64%). The two components are positively correlated at 0.47 (0.25 to 0.70). The intervals are tight, which is the payoff of the variance reduction in our design. The dominance of discount-rate news matches Campbell (1991). The positive correlation does not, but it matches the later estimates of Vuolteenaho (2002), Campbell, Giglio, and Polk (2013), and Gonçalves (2021). Anarkulova, Cederburg, and Zhou (2026) use these news estimates to show that the positive comovement requires underreaction to cash-flow news unless the representative investor’s subjective discount rates are strongly procyclical.

Our paper contributes to the empirical literature that uses Campbell’s (1991) return decomposition. Applications include intertemporal risk [e.g., Campbell (1996) and Gonçalves (2021)], monetary policy [e.g., Patelis (1997) and Bernanke and Kuttner (2005)], foreign exchange [e.g., Bekaert and Hodrick (1992), Atanasov and Nitschka (2015), and Chiang and Mo (2026)], and sources of market return predictability [e.g., Lamont (1998) and Rapach, Ringgenberg, and Zhou (2016)], among many other topics. Each depends on estimates of news components, and Chen and Zhao (2009) show that conventional estimates depend

heavily on the researcher’s choice of state variables. Our method gives these applications stable news estimates that conventional VARs cannot deliver.

Our approach also draws on the forecast combination literature [e.g., Bates and Granger (1969), Clemen (1989), and Timmermann (2006)] and its application to return prediction by Rapach, Strauss, and Zhou (2010). That literature combines forecasts to improve accuracy for a given target. We use the combinations as VAR state variables. Combination then serves two purposes. It reduces the dimension of the VAR, so that 25 predictors enter through three state variables. It also reduces variance, because averaging diversifies estimation error in the individual forecasts. We combine within persistence groups, which limits the bias that a single persistence for all 25 predictors would impose.

The remainder of the paper is organized as follows. Section 2 presents the decomposition, the bias-variance analysis of the news estimate, our method, and the bootstrap. Section 3 describes the data. Section 4 evaluates the method against the four criteria. Section 5 concludes.

2 Methodology

2.1 Vector autoregression approach

Campbell (1991) uses the loglinear present-value identity of Campbell and Shiller (1988a, 1988b) to decompose unexpected stock market returns into cash-flow news and discount-rate news. Following common empirical practice [e.g., Campbell and Vuolteenaho (2004) and Campbell, Giglio, and Polk (2013)], we apply the decomposition to excess returns,

$$\begin{aligned} r_{t+1}^e - E_t r_{t+1}^e &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j (\Delta d_{t+1+j} - r_{f,t+1+j}) - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}^e \\ &= N_{CF,t+1} - N_{DR,t+1}, \end{aligned} \tag{1}$$

where r_{t+1}^e is the log excess market return in month $t + 1$, Δd_{t+1} is the log dividend growth,

$r_{f,t+1}$ is the log risk-free rate, and ρ is a loglinearization constant.¹ Discount-rate news $N_{DR,t+1}$ is the revision in expected future excess returns, and cash-flow news $N_{CF,t+1}$ is the revision in expected current and future dividend growth in excess of the risk-free rate. We adopt this specification as our base case, and we study an alternative that models the risk-free rate separately so that cash-flow news reflects only dividend growth (Section 4.5).

The news components in equation (1) are nearly always estimated with a vector autoregression (VAR). Let the VAR take the form

$$\begin{bmatrix} r_{t+1}^e \\ x_{t+1} \end{bmatrix} = a + A \begin{bmatrix} r_t^e \\ x_t \end{bmatrix} + \varepsilon_{t+1}, \quad (2)$$

where x_{t+1} is a vector of return predictors. The first equation of the VAR is a predictive regression that produces the one-month forecast $E_t r_{t+1}^e$. Because the system also models the dynamics of the predictors r_t^e and x_t , the VAR produces the forecasts $E_t r_{t+2}^e, E_t r_{t+3}^e, \dots$ that the infinite sum in equation (1) requires.

The VAR residual ε_{t+1} is the vector of unexpected changes in the state variables, and the implied revision in the expected excess return at horizon j is $(E_{t+1} - E_t)r_{t+1+j}^e = e_1' A^j \varepsilon_{t+1}$, where e_1 selects the first element of the vector. Summing these revisions with weights ρ^j gives discount-rate news,

$$N_{DR,t+1} = e_1' \rho A (I - \rho A)^{-1} \varepsilon_{t+1}, \quad (3)$$

where I is an identity matrix. Equation (1) then gives cash-flow news as the unexpected return plus discount-rate news,

$$N_{CF,t+1} = (e_1' + e_1' \rho A (I - \rho A)^{-1}) \varepsilon_{t+1}. \quad (4)$$

Equation (4) estimates cash-flow news as the residual, following Campbell (1991) and much of the subsequent literature. Cash-flow news can instead be estimated directly by modeling

¹We set the loglinearization constant $\rho = 1/(1 + \exp(\overline{d-p}))$ to 0.9972, which is calculated using the average log monthly dividend-price ratio $\overline{(d-p)}$ in our sample. The decomposition assumes the terminal condition $\lim_{j \rightarrow \infty} \rho^j (d_{t+j} - p_{t+j}) = 0$.

cash flows in the VAR [e.g., Chen and Zhao (2009)], and Chen (2010) and Engsted, Pedersen, and Tanggaard (2012) show that the two methods coincide when the log dividend-price ratio is included as a state variable in the VAR. We use the residual method in our base case, and we show that the direct method produces similar estimates (Section 4.5).

2.2 A bias-variance tradeoff for VAR news estimates

The news estimate $\hat{N}_{DR,t+1}$ following equation (3) inherits the sampling error in the estimated coefficients. The discounted sum $\rho\hat{A}(I - \rho\hat{A})^{-1}$ places high weights on the most persistent predictors, so errors in estimating their dynamics are particularly costly. Dimensionality compounds this weighting effect. When the VAR has many state variables, $\hat{A}$ is imprecise given empirical sample sizes, so news estimates will be noisy. Let K denote the number of available predictors. An unrestricted VAR including all $K = 25$ predictors and the lagged return, for example, carries $(K + 1)^2 = 676$ VAR coefficients against 1,140 monthly observations. To combat dimensionality, researchers commonly choose a small subset of the K available predictors to use as VAR state variables, at the cost of discarding information and making estimates dependent on the choice. Our methodology instead partitions the K predictors into k groups $G_1, \dots, G_k$ and constructs one state variable per group by averaging the univariate return forecasts within the group, which reduces the VAR to $(k + 1)^2$ coefficients (16 in the three-group case). Two design choices govern the quality of the resulting news estimate: a sorting rule that partitions the predictors and an aggregation rule that combines predictors within each group. We motivate these choices using a bias-variance tradeoff.

To characterize the tradeoff, we express discount-rate news in terms of shocks to the individual predictors. We model each demeaned predictor as a first-order autoregression,

$$x_{i,t+1} = \phi_i x_{i,t} + v_{i,t+1}, \tag{5}$$

with $|\phi_i| < 1$ and $E_t v_{i,t+1} = 0$. The expected excess return is

$$E_t r_{t+1}^e = a + \sum_{i=1}^K \beta_i x_{i,t} + q_t, \quad (6)$$

where the β_i are the multiple regression coefficients of r_{t+1}^e on $\{x_{i,t}\}_{i=1}^K$ and q_t is the part of expected returns that the predictors do not span.

Discount-rate news is the revision in the discounted sum of expected excess returns at horizons $j \geq 1$. Because equation (6) holds at every date, the revision at horizon j is

$$(E_{t+1} - E_t) r_{t+1+j}^e = \sum_{i=1}^K \beta_i (E_{t+1} - E_t) x_{i,t+j} + (E_{t+1} - E_t) q_{t+j}, \quad (7)$$

and iterating equation (5) forward gives $(E_{t+1} - E_t) x_{i,t+j} = \phi_i^{j-1} v_{i,t+1}$. A shock to predictor $x_{i,t+1}$ therefore changes the expected return at horizon j by $\beta_i \phi_i^{j-1} v_{i,t+1}$, an effect that decays geometrically at rate ϕ_i . Multiplying by ρ^j and summing over horizons gives

$$N_{DR,t+1} = \sum_{i=1}^K \omega_i \beta_i v_{i,t+1} + n_{q,t+1}, \quad \omega_i = \frac{\rho}{1 - \rho \phi_i}, \quad (8)$$

where $n_{q,t+1} = \sum_{j=1}^{\infty} \rho^j (E_{t+1} - E_t) q_{t+j}$ is the corresponding news for the unspanned component.² The amplification weight ω_i is sharply increasing in persistence ϕ_i . To illustrate, with persistence levels of 0.50, 0.90, and 0.99 the weights are 2, 10, and 78: a shock to a near-unit-root predictor with $\phi_i = 0.99$ shifts expected returns across many future periods and enters $N_{DR,t+1}$ about 40 times more heavily than a shock to a predictor with $\phi_i = 0.50$.

²The condition $E_t v_{i,t+1} = 0$, with the conditional expectation E_t of equation (1), and the condition that the predictors do not forecast q_t together constitute Assumption 1 in Internet Appendix A. The lagged excess return is a state variable in the VAR but does not appear in equation (8) because its innovation is the unexpected return, which by equation (1) already contains $N_{DR,t+1}$. Any forecasting power of the lagged return beyond the K predictors is included in q_t , and Internet Appendix A accounts for it.

The mean squared error of $\hat{N}_{DR,t+1}$ in a sample of T months decomposes into four terms,

$$\underbrace{E\left[\left(\hat{N}_{DR,t+1} - N_{DR,t+1}\right)^2\right]}_{\text{MSE}} = \underbrace{B(k, \{G_l\}, T)}_{\text{bias}} + \underbrace{V(k, \{G_l\}, T)}_{\text{variance}} + \underbrace{S(K)}_{\text{spanning}} + \underbrace{\Delta(K, k, \{G_l\}, T)}_{\text{remainder}}, \quad (9)$$

where B is a squared-bias term from collapsing multiple predictors into group-level state variables, V is the variance of $\hat{N}_{DR,t+1}$ from sampling error in the estimated coefficients, $S = E[n_{q,t+1}^2]$ is the irreducible variance of news in the part of expected returns the predictor set does not span, and Δ is a remainder from the VAR's use of the lagged return (Internet Appendix A). Both S and Δ are zero when spanning is complete, and a large predictor set limits both. The sorting and aggregation choices govern the tradeoff between B and V , as detailed in Proposition 1.

Let ϕ_l denote the first-order autocorrelation of the group- l state variable and $\omega_l = \rho/(1 - \rho\phi_l)$ denote its amplification weight.

Proposition 1. *Under the benchmark of Assumption 1 in Internet Appendix A, holding the predictor dynamics fixed and varying only the design choices:*

- (a) *Spanning. Increasing the number of predictors K weakly decreases S . Given the predictor set, S does not depend on the sorting or aggregation rules.*
- (b) *Within-group averaging and variance. For a given number of groups, adding predictors to a group leaves the VAR dimension unchanged. Averaging the univariate forecasts within the group diversifies idiosyncratic estimation error and retains only error common to the group, which tends to reduce V .*
- (c) *Persistence sorting and bias. Estimating a common persistence ϕ_l for predictors with dispersed ϕ_i produces a bias that falls to zero as the within-group dispersion shrinks to zero. For a given number of groups, sorting on persistence minimizes this dispersion, while sorting randomly or on other characteristics does not.*
- (d) *Number of groups. For a given predictor set, increasing the number of groups k*

raises the VAR dimension, which increases V due to sampling error in the coefficients. Increasing k reduces B when it separates predictors that differ in persistence or in predictive coefficients. The number of groups therefore trades B against V .

- (e) Concentration at high persistence. Because ω_l is increasing in ϕ_l , both B and V concentrate in the high-persistence groups. For a given predictor set and number of groups, minimizing the MSE therefore favors partitions that are narrow in ϕ_i and dense in predictors in these groups.

Internet Appendix A contains the proofs and also gives conditions under which the $(k+1)$ -state VAR recovers spanned discount-rate news exactly in population: equal within-group persistence, univariate coefficients proportional to multiple regression coefficients within each group, and no marginal forecasting power in the lagged return.

2.3 Our approach

The K predictors $x_{i,t}$ are observed for months $t = 1, \dots, T + 1$. For each predictor, we estimate the univariate predictive regression

$$r_{t+1}^e = a_i + b_i x_{i,t} + \nu_{i,t+1}, \quad (10)$$

over the months $t = 1, \dots, T$ and compute the fitted values $\hat{r}_{i,t+1}^e = \hat{a}_i + \hat{b}_i x_{i,t}$, for $t = 1, \dots, T + 1$. We also estimate the autoregression in equation (5) for each predictor to obtain its persistence estimate $\hat{\phi}_i$.

We consider $K = 25$ predictors, as a large predictor set reduces the unspanned part of expected returns [Proposition 1(a)]. We sort the K predictors into k groups by estimated persistence $\hat{\phi}_i$, because persistence sorting minimizes the within-group dispersion in $\hat{\phi}_i$ that drives the bias [Proposition 1(c)]. We choose $k = 3$ in our base case, giving low-, medium-, and high-persistence groups. A single group forces predictors of very different persistence to share one amplification weight, which generates a large bias. Each additional

group reduces this bias, at the cost of enlarging the VAR and the sampling error in $\hat{N}_{DR,t+1}$ [Proposition 1(d)]. We find little gain beyond three groups (Section 4.4).

We select the persistence cutoffs to deliver strong return forecasting across horizons, subject to having at least five predictors in each group. Because discount-rate news depends on forecasts at all horizons, we measure forecasting performance by regressing the excess market return h months ahead on the lagged return and the k combinations for $h = 1, \dots, 60$ and summing the Akaike information criterion across horizons.³ Predictors with half-lives under three months ($\hat{\phi}_i < 0.794$) form the low-persistence group, those with half-lives from three months to four years ($0.794 \leq \hat{\phi}_i < 0.986$) the medium group, and the remainder ($\hat{\phi}_i \geq 0.986$) the high group. Because the persistence of return predictors clusters near one, the high-persistence group is both narrow in $\hat{\phi}_i$ and well populated (Section 3), which is the partition that the bias-variance tradeoff favors [Proposition 1(e)].

Within each group, we average the univariate forecasts,

$$\tilde{r}_{l,t+1}^e = \frac{1}{K_l} \sum_{i \in G_l} \hat{r}_{i,t+1}^e, \quad l = 1, \dots, k, \quad (11)$$

where K_l is the number of predictors in group l , following the forecast combination literature [e.g., Bates and Granger (1969) and Timmermann (2006)] and its use in return prediction by Rapach, Strauss, and Zhou (2010). Averaging diversifies the idiosyncratic estimation error in the univariate forecasts [Proposition 1(b)]. We average univariate forecasts, so each predictor is weighted by the estimate of its univariate regression coefficient b_i from equation (10) rather than its multiple regression coefficient β_i from equation (6). This weighting introduces a component of the bias term B that is zero only when $b_i \propto \beta_i$ within each group (Internet Appendix A). We accept this bias component in place of the estimation error from fitting β_i directly on the full, highly collinear predictor set [the “kitchen-sink” regression of Goyal and Welch (2008)]. We use equal weights rather than estimated combination weights. With the low predictive R^2 s of monthly return forecasts, equal weights are close to optimal.⁴ They

³Internet Appendix C.1 describes the procedure.

⁴Equal weights are optimal when the forecast errors have equal variances and pairwise correlations [Tim-

also avoid the weight-estimation error [Palm and Zellner (1992)] and in-sample overfitting that optimized weights would introduce in the subsequent VAR step.

Our method uses these combination return forecasts, standardized to mean zero and standard deviation one, as state variables in the VAR,

$$\begin{bmatrix} r_{t+1}^e \\ \tilde{r}_{1,t+2}^e \\ \tilde{r}_{2,t+2}^e \\ \tilde{r}_{3,t+2}^e \end{bmatrix} = a + A \begin{bmatrix} r_t^e \\ \tilde{r}_{1,t+1}^e \\ \tilde{r}_{2,t+1}^e \\ \tilde{r}_{3,t+1}^e \end{bmatrix} + \varepsilon_{t+1}. \quad (12)$$

Discount-rate news and cash-flow news follow from equations (3) and (4). The standardization rescales the VAR coefficients for comparability but leaves the news estimates unchanged.

2.4 Statistical inference

We conduct inference for the predictability test statistics in the paper with a bootstrap that imposes the relevant null and repeats every stage of our estimation in each draw. Under the null of no return predictability, returns equal the sample mean plus resampled return innovations, and each predictor follows its autoregression in equation (5) with bias-corrected persistence coefficients [Kendall (1954)]. We resample the return and predictor innovations jointly by date to preserve their contemporaneous correlations, which are the source of small-sample distortions in predictive regressions with persistent regressors [Stambaugh (1999)]. In each draw, we re-estimate the univariate predictive regressions, re-sort the simulated predictors into persistence groups by applying the same half-life cutoffs to the draw's own persistence estimates, rebuild the forecast combinations, re-estimate the VAR, and recompute the test statistics. We compute each reported statistic identically for the data and for each draw and obtain bootstrap p -values by comparing the statistic from the data to its dis-

mermann (2006)]. Low monthly predictive R^2 s imply that $\text{Var}(\widehat{E_t r_{t+1}^e})$ is small relative to $\text{Var}(r_{t+1}^e)$, so each model's forecast error $r_{t+1}^e - \widehat{E_t r_{t+1}^e}$ is close to r_{t+1}^e . Thus, the errors have similar variance and are highly correlated with one another across models.

tribution under the null. The p -values are one-tailed for the R^2 s and for coefficients whose signs the model predicts, and two-tailed otherwise, as noted in the tables. Tests of dividend growth predictability impose the analogous null of no dividend growth predictability. Confidence intervals for the news moments come from a separate bootstrap that simulates the estimated VAR in equation (12), re-estimates the VAR on each simulated sample, and recomputes the moments. The intervals are equal-tailed percentile intervals. Additional details are in Internet Appendix C.2.

3 Data and descriptive statistics

Our sample is monthly from January 1927 to December 2021 and consists of the excess market return and 25 return predictors, each available over the full period. The market return and risk-free rate are from Kenneth French’s website.⁵ We work with log excess returns in percent. The validation tests in Section 4.3 also use aggregate dividend growth. Following Kojien and Van Nieuwerburgh (2011), we construct monthly dividends on the CRSP value-weighted market portfolio without reinvestment, sum them over each 12-month window to remove seasonality, and measure annual log dividend growth each month as the log ratio of the trailing 12-month sum to the sum over the preceding 12 months. Internet Appendix B describes the construction of each predictor.

Our predictor set begins with 14 widely studied variables from Goyal and Welch (2008): valuation ratios (dividend-price ratio, dividend yield, earnings-price ratio, and book-to-market ratio), corporate decisions (dividend-earnings ratio and net stock issues), stock variance, and macroeconomic indicators (Treasury bill yield, long-term yield, long-term return, term spread, default spread, default return, and inflation).⁶

We augment these predictors with three variables that have no close analogues in Goyal and Welch (2008), all from John Campbell’s work on return decompositions: the relative

⁵We thank Kenneth French for making these data available (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

⁶We thank Amit Goyal for making these data available (<https://sites.google.com/view/agoyal145>).

bill rate from Campbell (1991), the smoothed earnings-price ratio from Shiller (2000) and Campbell and Vuolteenaho (2004), and the small-stock value spread from Campbell and Vuolteenaho (2004).

We also include eight variables from recent studies in the top three finance journals. These predictors measure corporate decisions [net payout yield from Boudoukh, Michaely, Richardson, and Roberts (2007)], statistical features of returns [average correlation from Pollet and Wilson (2010); tail risk from Kelly and Jiang (2014); and average skewness from Jondeau, Zhang, and Zhu (2019)], market valuations [nearness to Dow historical high from Li and Yu (2012) and partial least squares (PLS) aggregated book-to-market ratio from Kelly and Pruitt (2013)], and market liquidity [Amihud (2002) illiquidity measure from Chen, Eaton, and Paye (2018) and zero return frequency from Lesmond, Ogden, and Trzcinka (1999) and Chen, Eaton, and Paye (2018)].

We standardize each predictor to mean zero and standard deviation one. For each predictor, Table I reports the coefficient, bootstrap p -value under the null of no return predictability, and R^2 from the univariate predictive regression in equation (10), the AR(1) coefficient from equation (5), and the persistence group.

Table I reveals considerable heterogeneity in predictive power across variables. Of the 25 predictors, only six demonstrate statistically significant predictability: Treasury bill yield and PLS aggregated book-to-market at the 5% level, and long-term yield, dividend yield, net payout yield, and net stock issues at the 10% level. The economic magnitude of predictability is modest. The strongest predictor, PLS aggregated book-to-market, achieves an R^2 of 2.23%, while the others explain 0.70% or less of return variation.⁷

The persistence of the predictors also varies substantially. The valuation ratios show high persistence (AR(1) coefficients above 0.98), while some macroeconomic indicators (long-term return, default return, and inflation) and average skewness exhibit lower persistence (AR(1)

⁷We extensively validate our constructed predictors: (i) for variables with partial time series available from the original authors, we ensure high correlations in the overlapping periods; (ii) for several variables without available time series, we obtain return predictability results similar to the original studies if we match the sample period; and (iii) for variables for which replication is challenging, we closely match the independent replication attempts from Kothari and O’Doherty (2023) and Goyal, Welch, and Zafirov (2024).

coefficients below 0.50). The persistence sort places $K_1 = 7$ predictors in the low-persistence group, $K_2 = 9$ in the medium group, and $K_3 = 9$ in the high group. The high-persistence group spans AR(1) coefficients from 0.988 to 0.997, a width of about 0.01, while the low-persistence group spans -0.103 to 0.777 , a width of about 0.88. Dispersion in persistence is costly mainly where the amplification weights are large [Proposition 1(e)], and the sort leaves almost none there.

4 Results

4.1 Forecast combinations

Figure 2 shows the forecast combinations for different predictor groupings. Each series is a fitted expected excess return in percent per month. Panel A shows the case in which all 25 predictors form a single group ($k = 1$). The fitted expected return rises sharply during NBER recessions (shaded areas) and shows a secular decline over the sample period, consistent with estimates by Fama and French (2002) and Avdis and Wachter (2017). The combination thus captures economically meaningful variation in expected returns, in support of Rapach, Strauss, and Zhou’s (2010) arguments for forecast combination.

Panels B and C of Figure 2 group the predictors by persistence into two and three groups, respectively.⁸ The combinations remain broadly countercyclical, but their distinct persistence levels generate notably different time-series dynamics. In the base case ($k = 3$), the low-persistence combination varies monthly around a stable level, the medium-persistence combination swings over several years, and the high-persistence combination moves over decades and carries the secular decline.

These persistence differences determine how each combination enters discount-rate news. The AR(1) coefficients are 0.390 (half-life below one month), 0.965 (half-life of roughly 19 months), and 0.993 (half-life of roughly 94 months), and the amplification weights in equation (8) evaluated at these persistence levels are 1.63, 26.37, and 98.61. A shock to

⁸For $k = 2$, the single cutoff is at a half-life of four years.

the high-persistence combination therefore enters discount-rate news about 60 times more heavily than an equal-sized shock to the low-persistence combination.

Figure 3 shows how the predictive relation between each forecast combination and future returns accumulates with the horizon. For each combination, we regress the cumulative log excess market return over months $t + 1$ through $t + h$ on the standardized period- t combination and plot the coefficient $b(h)$ for horizons from one to 240 months. Under an AR(1) process with persistence ϕ_l , the cumulative coefficient converges to $b(1)/(1 - \phi_l)$ as the horizon grows, and the figure marks these limits with diamonds.

We find that the cumulative coefficient for the low-persistence combination peaks within the first year and remains near zero at longer horizons, while the coefficient for the medium-persistence combination increases over several years and fluctuates around its limit thereafter. The high-persistence combination has the smallest coefficient at the one-month horizon but the largest cumulative coefficient after 20 years, and its coefficient is still rising at the end of the window. This pattern is consistent with its persistence, since an AR(1) with $\phi_3 = 0.993$ implies a cumulative coefficient that reaches only about 80% of its limit after 240 months. Thus, the figure shows that the low-persistence combination forecasts returns at short horizons only, whereas the high-persistence combination is a relevant predictor for decades into the future. Because discount-rate news sums across all horizons, a combination's importance for discount-rate news depends on both its one-month predictive power and its persistence [equation (8)].

4.2 VAR estimates and news components

Panel A of Table II shows the matrix of VAR coefficients A , the R^2 from each regression, and the residual covariance matrix Σ . In the return equation, the medium-persistence forecast combination is significant at the 5% level, whereas the lagged excess market return and the low- and high-persistence combinations are insignificant. The insignificant one-month coefficient on the high-persistence combination is consistent with Figure 3, where that combination has the smallest coefficient at one month but the largest cumulative coefficient after

20 years. The return equation tests only the former; the latter enters the VAR through its dynamics, as the small one-month effect of this highly persistent combination accumulates over decades. In untabulated multiple regressions of cumulative returns at longer horizons on the four state variables, the high-persistence combination has the largest point estimate.

The predictive regression R^2 is 4.93%, which is high for market return predictability with a monthly horizon. Inference is based on the bootstrap (Section 2.4), which repeats every stage of the estimation under the null, so the hurdle for significance is the fit that the procedure generates from unpredictable returns. The observed R^2 clears it with a p -value of 0.000. A conventional VAR that uses three of the 25 predictors as state variables produces much less predictability: of the 2,298 such models, the average return-equation R^2 is 1.76%, 86% have R^2 s below 2.00%, and just six (0.26%) exceed 4.93%.⁹ Our approach uses information from a larger set of predictors, so it achieves better spanning of expected returns relative to the conventional approach [Proposition 1(a)].

The VAR coefficient estimates in Panel A of Table II also show that each forecast combination is the dominant predictor of itself with the expected persistence level. The covariance matrix estimate implies short-term continuation (due to a positive covariance between the shocks to r_{t+1}^e and the low-persistence combination $\tilde{r}_{1,t+2}^e$) and mean reversion in the medium and long terms (due to negative covariances between shocks to r_{t+1}^e and both $\tilde{r}_{2,t+2}^e$ and $\tilde{r}_{3,t+2}^e$).

Panel B of Table II reports summary statistics for the time series of discount-rate news and cash-flow news along with 90% bootstrap confidence intervals. Similar to Campbell (1991), we find that discount-rate news has a higher standard deviation (5.90% per month) than cash-flow news (2.84% per month). The 90% confidence intervals for the two news estimates are tight, ranging from 5.08% to 7.13% and 2.60% to 3.64% for discount-rate news and cash-flow news, respectively. Different from Campbell (1991), we find a positive correlation of 0.47 between the news components, with a confidence interval of 0.25 to 0.70. This positive correlation estimate is comparable to more recent estimates by Vuolteenaho

⁹Two of the $\binom{25}{3} = 2,300$ predictor sets are exactly collinear, so we omit those models. The dividend-earnings ratio equals the log dividend-price ratio minus the log earnings-price ratio, and the term spread equals the long-term yield minus the Treasury bill yield.

(2002), Campbell, Giglio, and Polk (2013), and Gonçalves (2021).

Discount-rate news is a weighted sum of revisions to expected returns at every horizon, and Figure 4 examines horizon-specific contributions to discount-rate news for different numbers of persistence groups (i.e., $k = 1, 2$, or 3). Panel A shows the contribution to discount-rate news variance from variation in expected returns at each horizon from one to 120 months. Specifically, given equation (3), discount-rate news variance is

$$\text{Var}(N_{DR,t+1}) = e_1' \rho A (I - \rho A)^{-1} \Sigma (I - \rho A)^{-1'} A' \rho e_1. \quad (13)$$

We decompose this expression by horizon as

$$\text{Var}(N_{DR,t+1}) = \sum_{j=1}^{\infty} e_1' \rho^j A^j \Sigma (I - \rho A)^{-1'} A' \rho e_1, \quad (14)$$

and we plot the j th term of the sum scaled by $\text{Var}(N_{DR,t+1})$. Panel B shows the cumulative sum of equation (14) for horizons of one to 120 months. The diamonds in Panel B denote the total discount-rate news variance from the infinite sum for each model.

The total discount-rate news variance is similar across the three models, between 30 and 35 (in %²), but its distribution across horizons is not. With the exception of some short-term dynamics related to the lagged market return predictor, Figure 4 shows that the single forecast combination model (i.e., $k = 1$) displays approximately exponential decay for discount-rate effects. These simple dynamics, combined with the relatively short half-life of the single combination (about 16 months), imply that the discount-rate news component from this VAR places heavy emphasis on short-horizon effects: 45% of its variance is from horizons of one year or less, 85% is within three years, and almost 100% is within ten years. With $k = 2$, Panel A shows a fast-decaying short-horizon component and a flat long-horizon tail, and with $k = 3$ the medium-persistence combination fills in the intermediate horizons. For the base case with $k = 3$, 26% of the variance is from horizons of one year or less, 61% is within three years, and 93% is within ten years. The single-combination model is restricted to one rate of decay, whereas the base case has the flexibility of three.

4.3 Validation of the news components

Equation (1) contains three unobservable terms that the VAR estimates: the expected excess return, discount-rate news, and cash-flow news. The testable implication for the first term is that the fitted expected return forecasts returns, and the return equation in Table II, with an R^2 of 4.93% (p -value of 0.000), satisfies it. Table II cannot validate the two news components, because equations (3) and (4) produce components that sum to the unexpected return for any set of state variables. The definitions of the news components do, however, imply testable restrictions. Discount-rate news should forecast future excess returns but not future cash flows beyond what cash-flow news forecasts, and cash-flow news should forecast future cash flows but not future excess returns beyond what discount-rate news forecasts.

Because news is the revision to expectations over all future horizons, the expectation at any date reflects accumulated past news. To aggregate past news, we follow Campbell, Giglio, and Polk (2013) and use a trailing exponentially weighted moving average.¹⁰ They place a weight of 0.08 on the current quarterly estimate and a weight of 0.92 on the past moving average. We use a monthly analogue with weights of $1 - 0.92^{1/3}$ and $0.92^{1/3}$, so the moving average for cash-flow news, for example, is

$$\text{MA}_t(N_{CF}) = 0.0274 \times N_{CF,t} + 0.9726 \times \text{MA}_{t-1}(N_{CF}). \quad (15)$$

Figure 5 plots $\text{MA}_t(N_{CF})$ and $-\text{MA}_t(N_{DR})$, where the negative sign reflects the negative relation between discount-rate news and unexpected returns in equation (1), so that a positive value of either series indicates a positive contribution to recent returns. Our estimates indicate that the price run-ups before the market crashes of 1929 and 2000 were accompanied by sharp declines in expected returns, and that cash-flow news contributed strongly to the price declines at the onset of the Great Depression, upon the bursting of the Internet bubble, and in late 2008 as the Great Recession deepened.

¹⁰When expected returns follow a single AR(1) process, the moving average of discount-rate news with decay equal to the persistence of expected returns is proportional to the expected return. With several persistence groups, no single decay rate recovers expected returns, and we use the Campbell, Giglio, and Polk (2013) weights as a convention.

Table III reports forecasts of annual returns and dividend growth from monthly regressions with overlapping annual observations. Columns (1) and (2) forecast the annual log excess market return,

$$r_{t \rightarrow t+12}^e = a_r + b_r \text{MA}_t(N_{DR}) + c_r \text{MA}_t(N_{CF}) + e_{r,t}, \quad (16)$$

with c_r restricted to zero or left unrestricted. Columns (3) and (4) forecast the annual log dividend growth in excess of the log risk-free rate,

$$\Delta d_{t \rightarrow t+12} - r_{f,t \rightarrow t+12} = a_d + b_d \text{MA}_t(N_{DR}) + c_d \text{MA}_t(N_{CF}) + e_{d,t}, \quad (17)$$

with b_d restricted to zero or left unrestricted. Inference uses the bootstrap of Section 2.4, which repeats every stage of the estimation, including the moving averages, and reproduces the overlapping structure of the annual regressions under the null.

Table III shows that each news component forecasts its own target and does not forecast the other. Discount-rate news forecasts annual excess returns in the restricted regression with an R^2 of 20.22%, significant at the 1% level. Cash-flow news (coefficient of 0.29, p -value of 0.753) adds nothing to the forecast of annual excess returns, as the R^2 is unchanged in the unrestricted regression.¹¹ For dividend growth, cash-flow news is a successful predictor with a coefficient of 13.78 and an R^2 of 18.13%, both significant at the 1% level. Including discount-rate news moves the R^2 only to 18.17%, and its coefficient of -0.31 is statistically insignificant with a p -value of 0.807.

Table IV repeats the forecasting regressions with real-time news estimates. Table III estimates the news components on the full sample to test the predictability relations, whereas the real-time analysis repeats at each date t every stage of the method, including the persistence sort, to estimate the news components using only data through t . The real-time test

¹¹We do not report p -values for the coefficients on discount-rate news in columns (1) and (2). In sample, $\text{MA}_t(N_{DR})$ is approximately a constant multiple of the VAR's fitted expected return and a regression of returns on their own fitted values has a slope of one, so the slope on $\text{MA}_t(N_{DR})$ is close to the reciprocal of the constant regardless of predictability. The R^2 is the statistic that tests predictability in these two regressions (Internet Appendix C.2).

is more stringent because the news at date t uses no data from the period it forecasts and because every stage is estimated anew each month, so any drift in the persistence sort, the combinations, or the VAR coefficients as data accumulate makes the news noisy and weakens its forecasting power.

The news validation pattern survives in real time. Discount-rate news forecasts returns in column (1) with a coefficient of 3.70 (p -value of 0.048) and an R^2 of 2.92% (p -value of 0.094), and cash-flow news forecasts dividend growth in column (3) with a coefficient of 7.66 (p -value of 0.000) and an R^2 of 19.89% (p -value of 0.000). Cash-flow news does not forecast returns [coefficient of -2.81 and p -value of 0.687 in column (2)], and discount-rate news does not forecast dividend growth [coefficient of -1.54 and p -value of 0.340 in column (4)].

Cash-flow news is estimated as a residual, the part of the unexpected return that discount-rate news does not account for, so it is not directly fitted to dividend growth. Nonetheless, cash-flow news forecasts dividend growth with an R^2 near 20% using in-sample or real-time news, while adding nothing to the return forecast. Taken as a whole, the validation results show our approach cleanly separates information about future discount rates from information about future cash flows.

To benchmark these results, Figure 6 compares our approach with the set of 2,298 VAR models that directly use three of the 25 predictor variables as state variables. Panels A and B summarize in-sample results for return and dividend growth forecasts, and Panels C and D show the corresponding real-time results. Our forecast combination approach outperforms the conventional approach. In sample, our predictive return regression achieves a higher R^2 than any of the conventional VARs (20.22% versus a maximum of 16.24%), and the dividend growth R^2 of 18.13% exceeds the R^2 s for 81% of the VARs. In real time, our return R^2 of 2.92% is higher than the R^2 s for 73% of VARs, and our dividend growth R^2 of 19.89% is higher than for every VAR (maximum of 13.71%). The 19% of VARs that beat our approach on in-sample dividend growth and the 27% that beat it on real-time returns are largely different models. Fewer than 4% beat it on both, and none beats it on three of the four comparisons.

4.4 Stability of news across design choices

Figure 1 shows that discount-rate news estimates from a conventional three-predictor VAR vary widely with which three predictors are chosen. We examine whether our estimates share that sensitivity, varying each design input in turn: how many total predictors are used, whether predictors are grouped by persistence or at random, how many persistence groups there are, and where the cutoffs between groups fall. Each comparison is the time-series correlation between the discount-rate news of an alternative model and that of the base case of Section 4.2. Proposition 1 predicts the patterns these comparisons should show. Using more predictors should improve the estimates through both spanning and within-group averaging, models that group predictors by persistence should be more successful than models that group them at random, adding groups should help until the sampling error from the added VAR coefficients outweighs the incremental reduction in bias, and the cutoffs should matter mainly through whether they produce a high-persistence group that is narrow in persistence but well populated.

Figure 7 varies the predictors in the three forecast combinations. Each panel holds the number of predictors per combination fixed, from one in Panel A to five in Panel E, so the total number of predictors ranges from three to 15. The blue histogram in each panel includes every model that draws that number of predictors from each of the three persistence groups, so its combinations are persistence sorted as in the base case. The orange histogram includes 100,000 models that draw the same number of predictors per combination at random from the full set of 25, so its combinations are not persistence sorted. In Panel A, the orange case is equivalent to a conventional three-predictor VAR (i.e., the set of models in Figure 1), and the blue case is the subset of those models with one predictor from each persistence group. In every panel, the only difference between the blue and orange plots is the persistence sort.

Figure 7 demonstrates that both persistence sorting and forecast combination are crucial. In Panel A, the two histograms are wide and similar, with many models having low or negative correlations with the base case. These specifications permit no within-group averaging to reduce estimation noise [Proposition 1(b)]. As predictors are added, the histograms

diverge. The blue distribution shifts toward one and tightens from Panel B onward, the joint effect of spanning more of expected returns and averaging more predictors within each group [Proposition 1(a) and (b)]. The orange distribution improves much less, because the random models mix predictors with different persistence and retain the bias of a common amplification weight [Proposition 1(c)]. In Panel E, with five predictors per combination and 15 in total, the median correlation is 0.92 for persistence-sorted models and the tenth percentile is 0.81, versus a median of 0.54 and a tenth percentile of 0.35 for random models. Successful models require both (i) a reasonably large set of predictors, to span expected returns and to gain from forecast combination, and (ii) persistence sorting, to reduce bias.

Table V varies the number of persistence groups from one to 12. The base case is $k = 3$, and for other values of k the predictors are sorted by persistence into groups of near-equal size. The table reports correlations of discount-rate news across the 12 models and the news moments for each model. With one group, the discount-rate news estimate has a correlation of just 0.59 with the base case. The correlation rises to 0.88 with two groups and 0.95 with four, and stays between 0.97 and 0.99 from five through seven groups. Each added group narrows the range of persistence and reduces the bias of a common amplification weight [Proposition 1(d)], and that gain is largely realized by the third group. Beyond seven groups the results are less stable from one value of k to the next, with moments swinging between adjacent models and the correlation with the base case falling to between 0.86 and 0.88 from $k = 9$ onward. These models have at least 81 VAR coefficients versus 16 in the base case, and the instability across adjacent values of k is consistent with growing sampling error in the coefficients [Proposition 1(d)]. The estimate is stable for any choice from three to seven groups, the range over which the bias-variance tradeoff appears flat.

Figure 8 varies the persistence cutoffs. Panel A holds the high-persistence half-life cutoff at its base value of 48 months and moves the low-persistence cutoff from one month to 36 months, and Panel B holds the low-persistence half-life cutoff at three months and moves the high-persistence cutoff from 12 months to 120 months. The groups change only when a cutoff crosses a predictor’s half-life, so each distinct grouping is included in the histogram

once. The estimates are insensitive to the low-persistence cutoff, as every correlation in Panel A exceeds 0.97. The high-persistence cutoff matters more, although every correlation in Panel B exceeds 0.83. The lowest values sit at the two extremes. A half-life cutoff between 109 and 116 months leaves only three predictors in the high-persistence group, and a half-life cutoff below 22 months pushes 14 or more highly heterogeneous predictors into it. These results highlight the importance of creating a high-persistence group that is narrow in persistence and dense in predictors [Proposition 1(e)].

4.5 Alternative VAR specifications

Our base specification follows two common conventions. First, we model discount-rate news using excess returns, because the predictors proposed in the literature are shown to forecast excess returns. A natural consequence is that the cash-flow news term in equation (1) includes the risk-free rate. Second, we estimate cash-flow news as the residual in equation (4). This subsection relaxes each convention in turn, with the risk-free rate as a standalone state variable in the first case and the log dividend-price ratio in the second, and asks whether the news estimates change. Table VI reports the return-equation R^2 , the news moments, and the correlation of each alternative's news with the base case. Panel A repeats the base case statistics for convenience. Internet Appendix D repeats the validation regressions of Section 4.3 for each alternative.

The first alternative separates dividend growth and the risk-free rate, given a potential concern with equation (1) that expectations about risk-free rates enter cash-flow news,

$$\begin{aligned}
r_{t+1}^e - E_t r_{t+1}^e &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{f,t+1+j} \\
&\quad - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}^e \\
&= N_{CF,t+1} - N_{RF,t+1} - N_{RP,t+1},
\end{aligned} \tag{18}$$

where $N_{RF,t+1}$ is risk-free rate news and $N_{RP,t+1}$ is risk-premium news. We then define

discount-rate news as $N_{DR,t+1} = N_{RF,t+1} + N_{RP,t+1}$. The VAR adds the log risk-free rate to the state vector and drops the Treasury bill yield from the high-persistence combination (see Internet Appendix D for the state vector and news formulas). Panel B of Table VI shows that the news moments are close to the base case. The two discount-rate news series correlate at 0.98, and cash-flow news, which no longer contains the risk-free rate, correlates at 0.89 with the base case. The validation regressions in Internet Appendix D show the same pattern as Tables III and IV, with discount-rate news forecasting returns, cash-flow news forecasting dividend growth, and neither forecasting the other’s target.

The second alternative estimates cash-flow news directly. As Section 2.1 notes, the direct and residual methods coincide when the log dividend-price ratio is a standalone state variable [Chen (2010); Engsted, Pedersen, and Tanggaard (2012)], so we add $d_t - p_t$ to the state vector and drop it from the high-persistence combination (see Internet Appendix D for the state vector). Panel C of Table VI shows that both news series are more volatile than in the base case and more highly correlated with each other. The estimates nonetheless agree with the base case, with correlations of 0.94 for discount-rate news and 0.96 for cash-flow news, and the validation regressions in Internet Appendix D show the same pattern as Tables III and IV. For our approach, the choice between the residual and direct methods makes little difference.

5 Conclusion

We propose a simple way to tame the estimation of discount-rate news and cash-flow news within Campbell’s (1991) VAR approach. A bias-variance framework underscores the importance of (i) retaining a large set of predictors to span expected returns, (ii) using forecast combination to diversify estimation error and to keep the VAR dimension manageable, and (iii) persistence sorting to minimize bias. Our approach implements each of these. The resulting discount-rate and cash-flow news each forecast their own target and not the other, in sample and in real time, so the method cleanly separates information about future discount

rates from information about future cash flows. The estimates are stable across reasonable modeling choices, and the sensitivity that remains appears where the bias-variance framework predicts it. Our approach represents a step toward restoring confidence in VAR-based estimates of discount-rate news and cash-flow news.

References

- Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.
- Anarkulova, Aizhan, Scott Cederburg, and Yi Zhou, 2026, The stock market’s two truths: Subjective beliefs and objective reality, Working paper, Emory University.
- Atanasov, Victoria, and Thomas Nitschka, 2015, Foreign currency returns and systematic risks, *Journal of Financial and Quantitative Analysis* 50, 231–250.
- Avdis, Efstathios, and Jessica A. Wachter, 2017, Maximum likelihood estimation of the equity premium, *Journal of Financial Economics* 125, 589–609.
- Bates, John M., and Clive W. J. Granger, 1969, The combination of forecasts, *Operational Research Quarterly* 20, 451–468.
- Bekaert, Geert, and Robert J. Hodrick, 1992, Characterizing predictable components in excess returns on equity and foreign exchange markets, *Journal of Finance* 47, 467–509.
- Bernanke, Ben S., and Kenneth N. Kuttner, 2005, What explains the stock market’s reaction to Federal Reserve policy?, *Journal of Finance* 60, 1221–1257.
- Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.
- Campbell, John Y., 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.
- , 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.
- Campbell, John Y., Stefano Giglio, and Christopher Polk, 2013, Hard times, *Review of Asset Pricing Studies* 3, 95–132.
- Campbell, John Y., and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.
- , 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.
- Campbell, John Y., and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.
- Chen, Long, 2010, On return and beta decompositions: Governing principles and applications, Working paper, Washington University in St. Louis.
- Chen, Long, and Xinlei Zhao, 2009, Return decomposition, *Review of Financial Studies* 22, 5213–5249.

- Chen, Yong, Gregory W. Eaton, and Bradley S. Paye, 2018, Micro(structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity, *Journal of Financial Economics* 130, 48–73.
- Chiang, I-Hsuan Ethan, and Xi Nancy Mo, 2026, A “bad beta, good beta” anatomy of currency risk premiums and trading strategies, *Review of Finance* 30, 681–715.
- Clemen, Robert T., 1989, Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting* 5, 559–583.
- Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.
- Engsted, Tom, Thomas Q. Pedersen, and Carsten Tanggaard, 2012, Pitfalls in VAR based return decompositions: A clarification, *Journal of Banking and Finance* 36, 1255–1265.
- Fama, Eugene F., and Kenneth R. French, 2002, The equity premium, *Journal of Finance* 57, 637–659.
- Fisher, Walter D., 1958, On grouping for maximum homogeneity, *Journal of the American Statistical Association* 53, 789–798.
- Gonçalves, Andrei S., 2021, Reinvestment risk and the equity term structure, *Journal of Finance* 76, 2153–2197.
- Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.
- Goyal, Amit, Ivo Welch, and Athanasse Zafirov, 2024, A comprehensive 2022 look at the empirical performance of equity premium prediction, *Review of Financial Studies* 37, 3490–3557.
- Greene, William H., 2018, *Econometric Analysis*. Pearson, New York, 8th edn.
- Hill, Bruce M., 1975, A simple general approach to inference about the tail of a distribution, *Annals of Statistics* 3, 1163–1174.
- Jondeau, Eric, Qunzi Zhang, and Xiaoneng Zhu, 2019, Average skewness matters, *Journal of Financial Economics* 134, 29–47.
- Kelly, Bryan, and Hao Jiang, 2014, Tail risk and asset prices, *Review of Financial Studies* 27, 2841–2871.
- Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.
- Kendall, Maurice G., 1954, Note on bias in the estimation of autocorrelation, *Biometrika* 41, 403–404.

- Kilian, Lutz, 1998, Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* 80, 218–230.
- Koijen, Ralph S. J., and Stijn Van Nieuwerburgh, 2011, Predictability of returns and cash flows, *Annual Review of Financial Economics* 3, 467–491.
- Kothari, Pratik, and Michael S. O’Doherty, 2023, Job postings and aggregate stock returns, *Journal of Financial Markets* 64, 100804.
- Lamont, Owen A., 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.
- Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka, 1999, A new estimate of transaction costs, *Review of Financial Studies* 12, 1113–1141.
- Li, Jun, and Jianfeng Yu, 2012, Investor attention, psychological anchors, and stock return predictability, *Journal of Financial Economics* 104, 401–419.
- Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Palm, Franz C., and Arnold Zellner, 1992, To combine or not to combine? Issues of combining forecasts, *Journal of Forecasting* 11, 687–701.
- Patelis, Alex D., 1997, Stock return predictability and the role of monetary policy, *Journal of Finance* 52, 1951–1972.
- Pollet, Joshua M., and Mungo Wilson, 2010, Average correlation and stock market returns, *Journal of Financial Economics* 96, 364–380.
- Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.
- Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.
- Shiller, Robert J., 2000, *Irrational Exuberance*. Princeton University Press, Princeton, NJ.
- Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Timmermann, Allan, 2006, Forecast combinations, in Graham Elliott, Clive W. J. Granger, and Allan Timmermann, eds.: *Handbook of Economic Forecasting* (North-Holland, Amsterdam).
- Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns?, *Journal of Finance* 57, 233–264.
- White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Table I
Return predictability and persistence of the predictor variables.

The table reports estimates from univariate predictive regressions and autoregressions for 25 market return predictors. The predictor variables are standardized. For each variable, we regress the log excess market return on the lagged predictor and report the coefficient, the two-tailed bootstrap p -value under the null of no return predictability, and the R^2 (in percent). The table also reports the AR(1) coefficient from a regression of the predictor on its own lag, along with the predictor's persistence group: H (half-life of more than four years), M (three months to four years), or L (under three months). The sample period is January 1927 to December 2021. An * (**) [***] on the predictive coefficient estimate indicates statistical significance at the 10% (5%) [1%] level based on the bootstrap p -value.

Variable	Coefficient	p -value	R^2 (%)	AR(1)	Group
Long-term yield	-0.24*	[0.087]	0.21	0.997	H
Dividend-price ratio	0.18	[0.759]	0.11	0.995	H
Dividend yield	0.29*	[0.063]	0.30	0.994	H
Treasury bill yield	-0.32**	[0.043]	0.35	0.994	H
Dividend-earnings ratio	-0.07	[0.718]	0.02	0.992	H
Smoothed earnings-price ratio	0.35	[0.502]	0.42	0.992	H
Nearness to Dow historical high	-0.08	[0.807]	0.02	0.990	H
Earnings-price ratio	0.25	[0.279]	0.22	0.988	H
Book-to-market ratio	0.32	[0.495]	0.35	0.988	H
Small-stock value spread	-0.07	[0.770]	0.02	0.984	M
Net payout yield	0.45*	[0.050]	0.70	0.983	M
Net stock issues	-0.41*	[0.064]	0.58	0.981	M
Default spread	0.11	[0.852]	0.04	0.975	M
PLS aggregated book-to-market ratio	0.80**	[0.034]	2.23	0.968	M
Term spread	0.22	[0.171]	0.17	0.961	M
Amihud illiquidity measure	0.16	[0.474]	0.09	0.946	M
Average correlation	0.24	[0.302]	0.20	0.882	M
Zero return frequency	0.27	[0.231]	0.25	0.879	M
Tail risk	0.12	[0.473]	0.05	0.777	L
Relative bill rate	-0.15	[0.347]	0.08	0.741	L
Stock variance	-0.06	[0.842]	0.01	0.577	L
Inflation	-0.21	[0.327]	0.15	0.482	L
Average skewness	0.16	[0.541]	0.09	0.106	L
Long-term return	0.26	[0.131]	0.24	0.049	L
Default return	0.25	[0.337]	0.22	-0.103	L

Table II
Vector autoregression estimates.

The table reports estimates from the baseline vector autoregression (VAR). The market return predictors are sorted into three groups by their AR(1) coefficients, and the VAR state variables are the log excess market return and the standardized forecast combinations from the low-persistence ($\tilde{r}_{1,t+1}^e$, half-lives less than three months), medium-persistence ($\tilde{r}_{2,t+1}^e$, three months to four years), and high-persistence ($\tilde{r}_{3,t+1}^e$, more than four years) predictor groups. Panel A reports the VAR coefficient matrix A , the R^2 (in percent) from each regression, and the covariance matrix Σ , with p -values (in brackets) below the coefficients and R^2 s. In the return equation, the one-tailed p -values for the coefficients and the R^2 are from a bootstrap under the null of no return predictability. All other coefficient p -values are two-tailed, based on White (1980) heteroskedasticity-robust standard errors, and the remaining R^2 p -values are from Wald tests that all slopes in the equation are zero. Panel B reports the standard deviations (in percent per month) and correlation of the discount-rate news (N_{DR}) and cash-flow news (N_{CF}), with 90% confidence intervals (in brackets) from a parametric bootstrap centered at the estimated VAR. The sample period is January 1927 to December 2021. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

Panel A: VAR estimates									
	A				R^2	Σ			
	r_t^e	$\tilde{r}_{1,t+1}^e$	$\tilde{r}_{2,t+1}^e$	$\tilde{r}_{3,t+1}^e$		r_{t+1}^e	$\tilde{r}_{1,t+2}^e$	$\tilde{r}_{2,t+2}^e$	$\tilde{r}_{3,t+2}^e$
r_{t+1}^e	0.04 [0.166]	0.38 [0.837]	0.99** [0.037]	0.08 [0.905]	4.93*** [0.000]	27.27	1.86	-0.24	-0.55
$\tilde{r}_{1,t+2}^e$	-0.03*** [0.000]	0.45*** [0.000]	0.07* [0.054]	-0.02 [0.523]	17.63*** [0.000]	1.86	0.82	-0.02	-0.03
$\tilde{r}_{2,t+2}^e$	-0.01*** [0.000]	-0.01 [0.400]	0.97*** [0.000]	0.02 [0.109]	93.76*** [0.000]	-0.24	-0.02	0.06	0.01
$\tilde{r}_{3,t+2}^e$	-0.00 [0.496]	-0.00 [0.599]	-0.02*** [0.005]	1.00*** [0.000]	98.57*** [0.000]	-0.55	-0.03	0.01	0.01
Panel B: News component summary statistics									
Statistic	$\sigma(N_{DR})$	$\sigma(N_{CF})$	$\rho(N_{DR}, N_{CF})$						
Estimate	5.90	2.84	0.47						
90% CI	[5.08, 7.13]	[2.60, 3.64]	[0.25, 0.70]						

Table III

Return and dividend growth forecasts from discount-rate and cash-flow news.

The table reports estimates from regressions that forecast annual log excess market return in columns (1) and (2) and annual log dividend growth in excess of the log risk-free rate in columns (3) and (4). The predictors are smoothed versions of the discount-rate news ($MA_t(N_{DR})$) and cash-flow news ($MA_t(N_{CF})$) estimates from the baseline VAR, constructed as trailing exponentially weighted moving averages following Campbell, Giglio, and Polk (2013). The sample period is January 1927 to December 2021. The p -values (in brackets) below the coefficients and R^2 s are from a bootstrap under the null of no return predictability in columns (1) and (2) and the null of no dividend growth predictability in columns (3) and (4). The p -values are one-tailed for the R^2 s and the $MA_t(N_{CF})$ coefficients in the dividend growth forecasts, and two-tailed otherwise. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

	Return		Div. growth	
	(1)	(2)	(3)	(4)
$MA_t(N_{DR})$	10.64	10.58		−0.31 [0.807]
$MA_t(N_{CF})$		0.29 [0.753]	13.78*** [0.001]	14.18*** [0.001]
R^2 (%)	20.22*** [0.003]	20.22*** [0.006]	18.13*** [0.000]	18.17*** [0.003]

Table IV
Return and dividend growth forecasts from real-time discount-rate and cash-flow news.

The table reports estimates from regressions that forecast annual log excess market return in columns (1) and (2) and annual log dividend growth in excess of the log risk-free rate in columns (3) and (4). The predictors are real-time estimates of the smoothed discount-rate news ($MA_t(N_{DR})$) and cash-flow news ($MA_t(N_{CF})$). Each month, on an expanding window of data available through that month, we sort predictors into three persistence groups using bias-corrected AR(1) coefficients, estimate the baseline VAR, and form the smoothed news estimates from that vintage only. The sample period is January 1927 to December 2021. The p -values (in brackets) below the coefficients and R^2 s are from a bootstrap under the null of no return predictability in columns (1) and (2) and the null of no dividend growth predictability in columns (3) and (4). The p -values are one-tailed for the R^2 s, the $MA_t(N_{DR})$ coefficients in the return forecasts, and the $MA_t(N_{CF})$ coefficients in the dividend growth forecasts, and two-tailed otherwise. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

	Return (real-time)		Div. growth (real-time)	
	(1)	(2)	(3)	(4)
$MA_t(N_{DR})$	3.70** [0.048]	4.77* [0.055]		-1.54 [0.340]
$MA_t(N_{CF})$		-2.81 [0.687]	7.66*** [0.000]	9.02*** [0.000]
R^2 (%)	2.92* [0.094]	3.40 [0.204]	19.89*** [0.000]	21.15*** [0.000]

Table V
Correlation of discount-rate news across VARs with different numbers of persistence groups.

The table reports time-series correlations of the discount-rate news estimates from VARs whose state variables are the log excess market return and k forecast combinations, for k ranging from one to 12. For $k = 3$, the predictors are sorted into the low-, medium-, and high-persistence groups of the baseline model using half-life cutoffs of three months and four years. For other values of k , the predictors are sorted by their AR(1) coefficients and split into k groups of near-equal size. The table also reports the average correlation of each model's discount-rate news with that of the other 11 models. The bottom rows report the standard deviations (in percent per month) and correlation of the discount-rate news (N_{DR}) and cash-flow news (N_{CF}) from each VAR. The sample period is January 1927 to December 2021.

k	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00											
2	0.59	1.00										
3	0.59	0.88	1.00									
4	0.61	0.97	0.95	1.00								
5	0.61	0.88	0.99	0.95	1.00							
6	0.59	0.94	0.98	0.98	0.98	1.00						
7	0.55	0.89	0.97	0.95	0.96	0.97	1.00					
8	0.53	0.87	0.96	0.94	0.94	0.95	0.98	1.00				
9	0.51	0.75	0.86	0.83	0.85	0.85	0.87	0.86	1.00			
10	0.51	0.76	0.87	0.83	0.86	0.85	0.88	0.87	0.99	1.00		
11	0.51	0.78	0.88	0.85	0.87	0.87	0.89	0.89	0.98	0.99	1.00	
12	0.51	0.78	0.87	0.85	0.86	0.86	0.89	0.90	0.97	0.98	0.99	1.00
Average	0.56	0.83	0.89	0.88	0.89	0.89	0.89	0.88	0.85	0.85	0.86	0.86
$\sigma(N_{DR})$	5.64	6.08	5.90	6.42	5.92	5.85	5.75	5.33	5.62	5.59	5.42	4.90
$\sigma(N_{CF})$	5.72	3.85	2.84	3.66	3.11	3.02	2.84	1.94	3.55	3.46	3.24	2.98
$\rho(N_{DR}, N_{CF})$	0.57	0.53	0.47	0.58	0.48	0.46	0.43	0.25	0.43	0.42	0.37	0.21

Table VI
Return predictability and news component moments across VAR specifications.

The table reports the R^2 (in percent) from the log excess market return equation of the VAR and the standard deviations (in percent per month), correlation of the discount-rate news (N_{DR}) and cash-flow news (N_{CF}), and correlations of alternative VAR news estimates with baseline VAR estimates. Panel A reports results for the baseline VAR, Panel B for the VAR that directly includes the log risk-free rate, and Panel C for the VAR that directly includes the log dividend-price ratio. The R^2 p -values (in brackets) are one-tailed, from a bootstrap under the null of no return predictability. The 90% confidence intervals (in brackets) are from a parametric bootstrap centered at the estimated VAR. The sample period is January 1927 to December 2021. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the bootstrap p -value.

R^2 (%)	$\sigma(N_{DR})$	$\sigma(N_{CF})$	$\rho(N_{DR}, N_{CF})$	$\rho(N_{DR}, N_{DR}^{base})$	$\rho(N_{CF}, N_{CF}^{base})$
Panel A: Baseline VAR					
4.93*** [0.000]	5.90 [5.08, 7.13]	2.84 [2.60, 3.64]	0.47 [0.25, 0.70]		
Panel B: Risk-free rate VAR					
4.99*** [0.000]	5.37 [4.87, 6.55]	2.41 [2.16, 3.06]	0.29 [0.14, 0.62]	0.98	0.89
Panel C: Dividend-price ratio VAR					
5.29*** [0.000]	7.10 [6.18, 10.28]	5.08 [3.07, 7.44]	0.68 [0.57, 0.89]	0.94	0.96

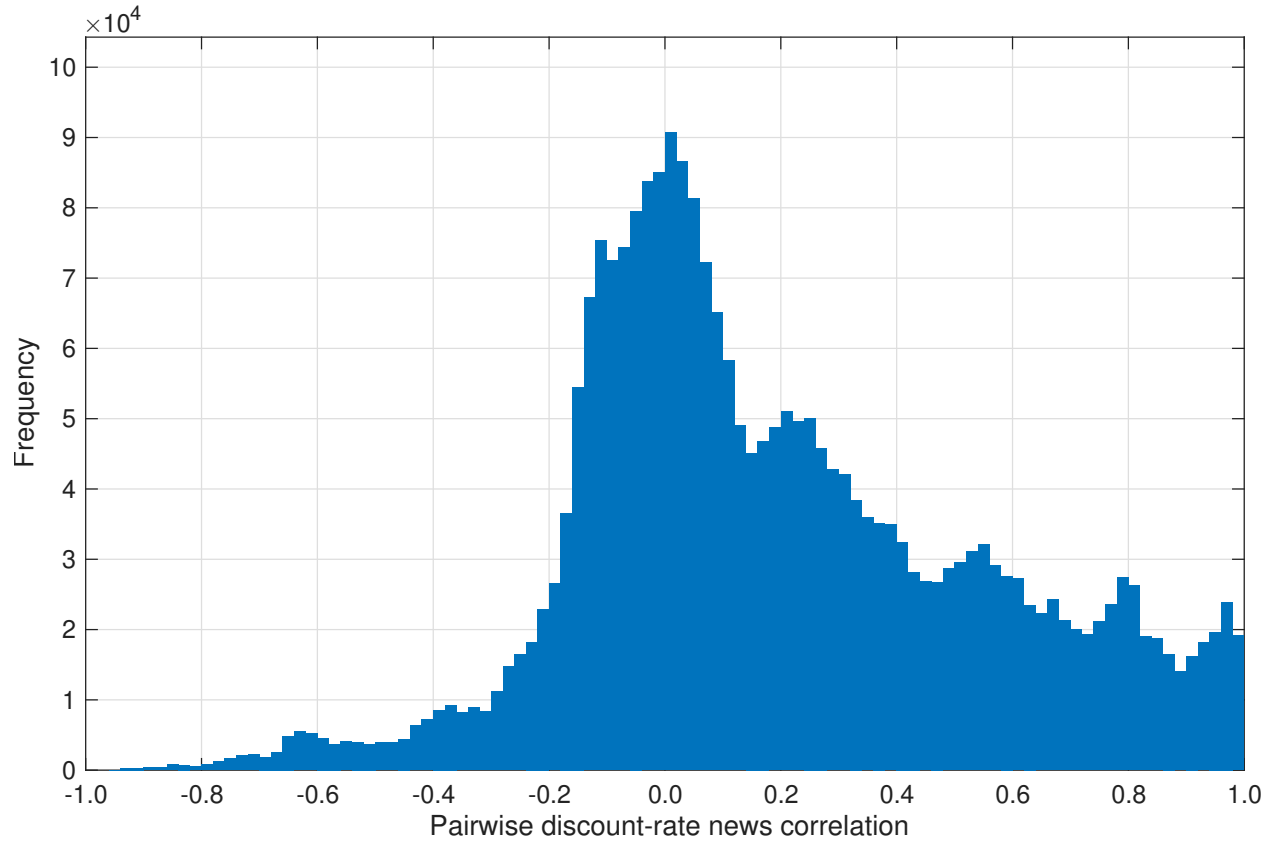


Figure 1. Correlation of discount-rate news across model pairs. The figure shows a histogram of the pairwise time-series correlations of the discount-rate news estimates from all 2,298 VARs that each include the log excess market return and three of the 25 predictor variables. The sample period is January 1927 to December 2021.

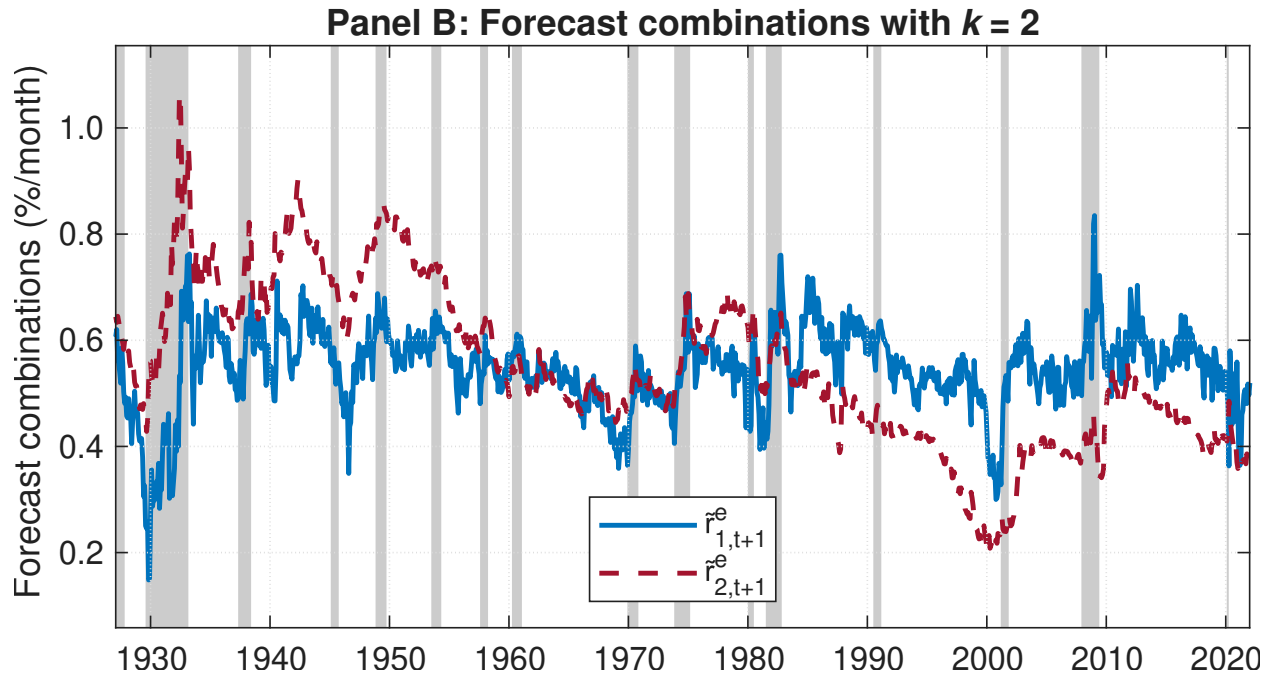
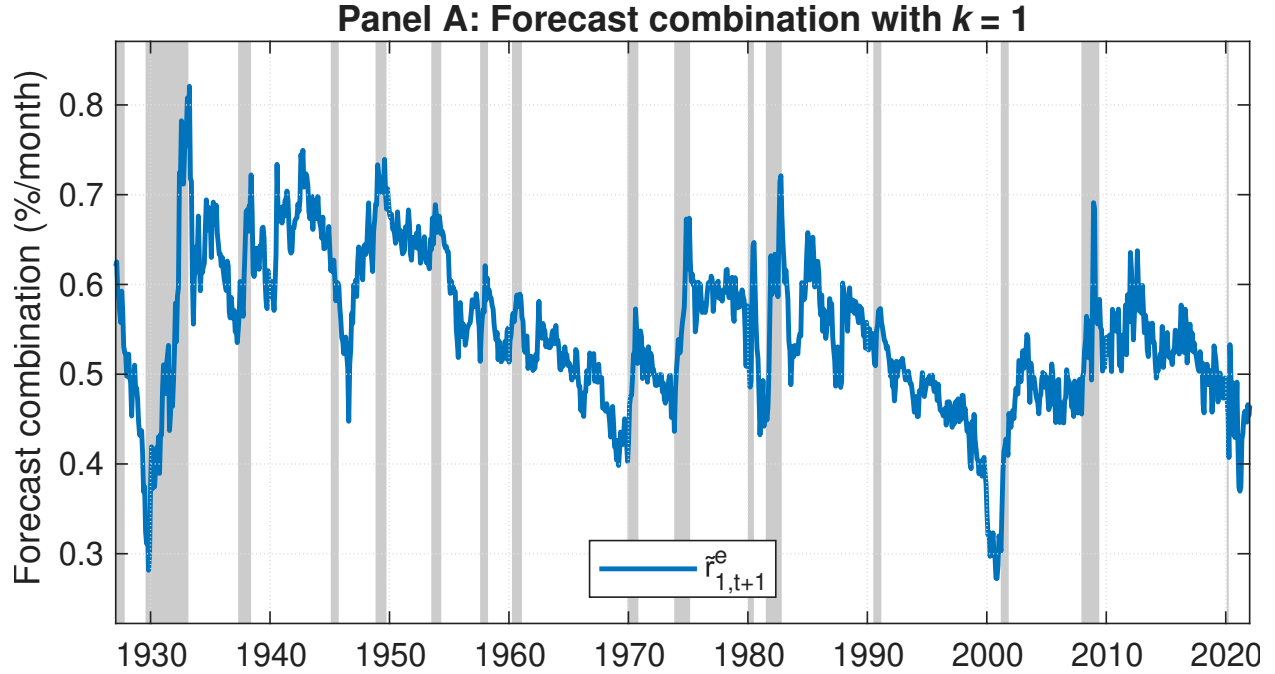


Figure 2. Forecast combinations. The figure shows the time series of forecast combinations for one (Panel A), two (Panel B), and three (Panel C) persistence groups. Each combination is the equal-weighted average of the univariate log excess market return forecasts (in percent per month) from the predictors in its group. NBER recessions are shaded. The sample period is January 1927 to December 2021.

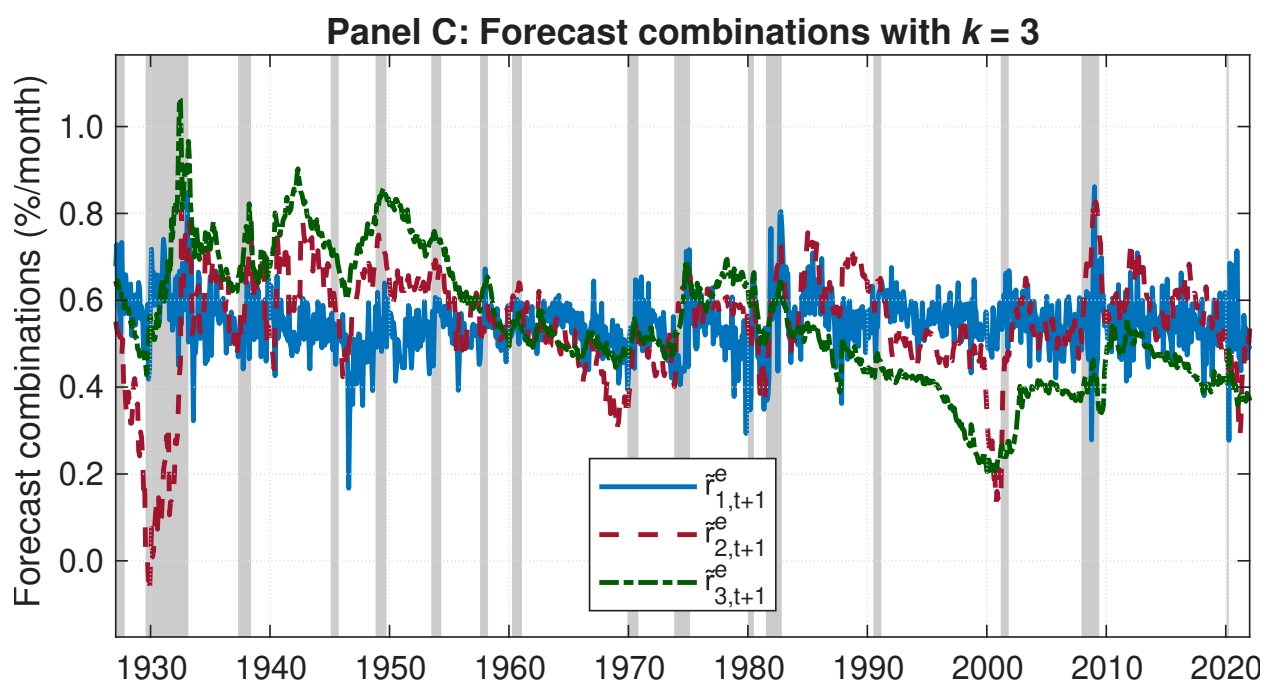


Figure 2. Forecast combinations (continued).

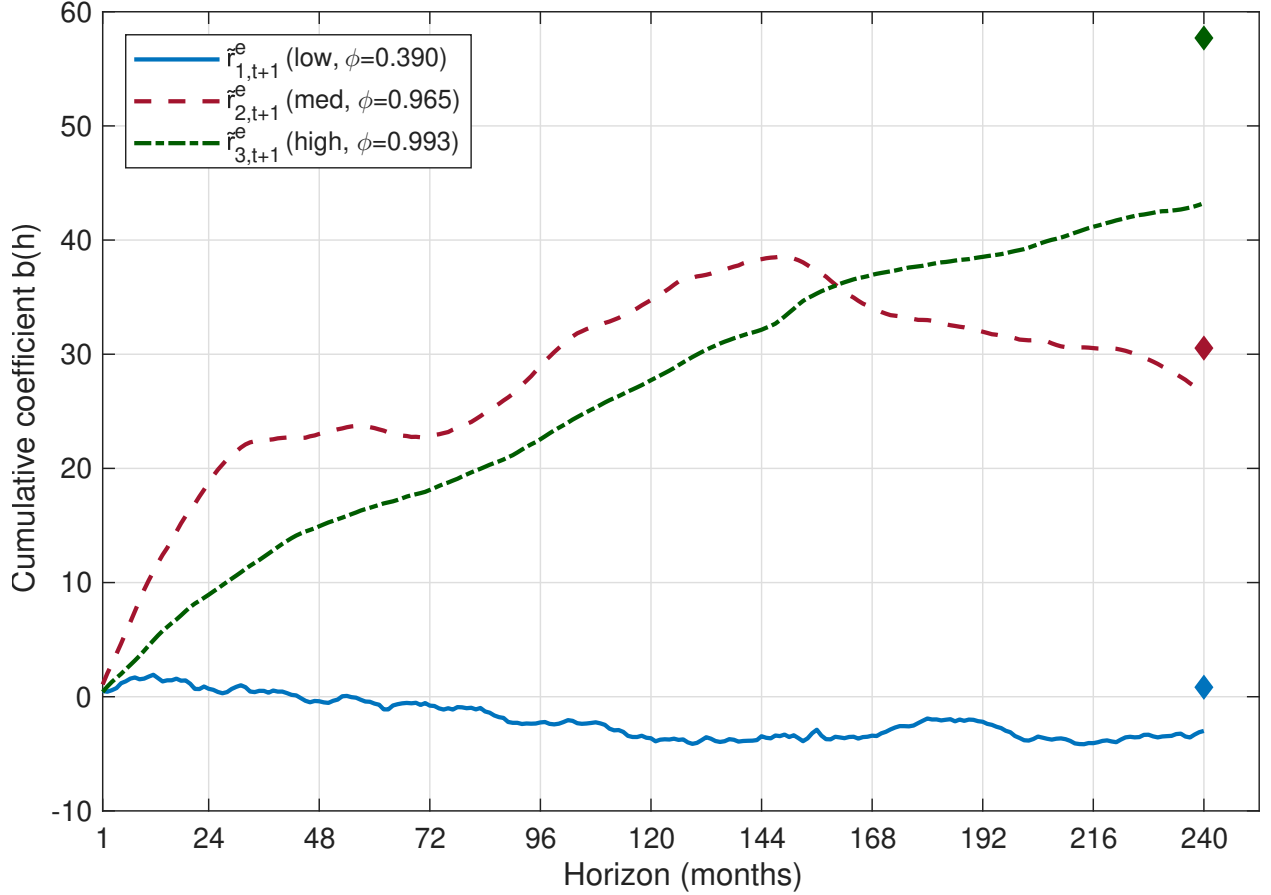


Figure 3. Long-horizon cumulative return coefficients by persistence group. The figure shows the coefficient $b(h)$ from a regression of the cumulative log excess market return (in percent) over months $t + 1$ through $t + h$ on the period- t forecast combination, for horizons h from one to 240 months. The forecast combinations are standardized. The three lines correspond to the low-persistence ($\tilde{r}_{1,t+1}^e$, AR(1) coefficient $\phi = 0.390$, half-life below one month), medium-persistence ($\tilde{r}_{2,t+1}^e$, $\phi = 0.965$, half-life of roughly 19 months), and high-persistence ($\tilde{r}_{3,t+1}^e$, $\phi = 0.993$, half-life of roughly 94 months) forecast combinations. The diamonds at the 240-month horizon show the theoretical limits $b(1)/(1 - \phi)$. The sample period is January 1927 to December 2021.

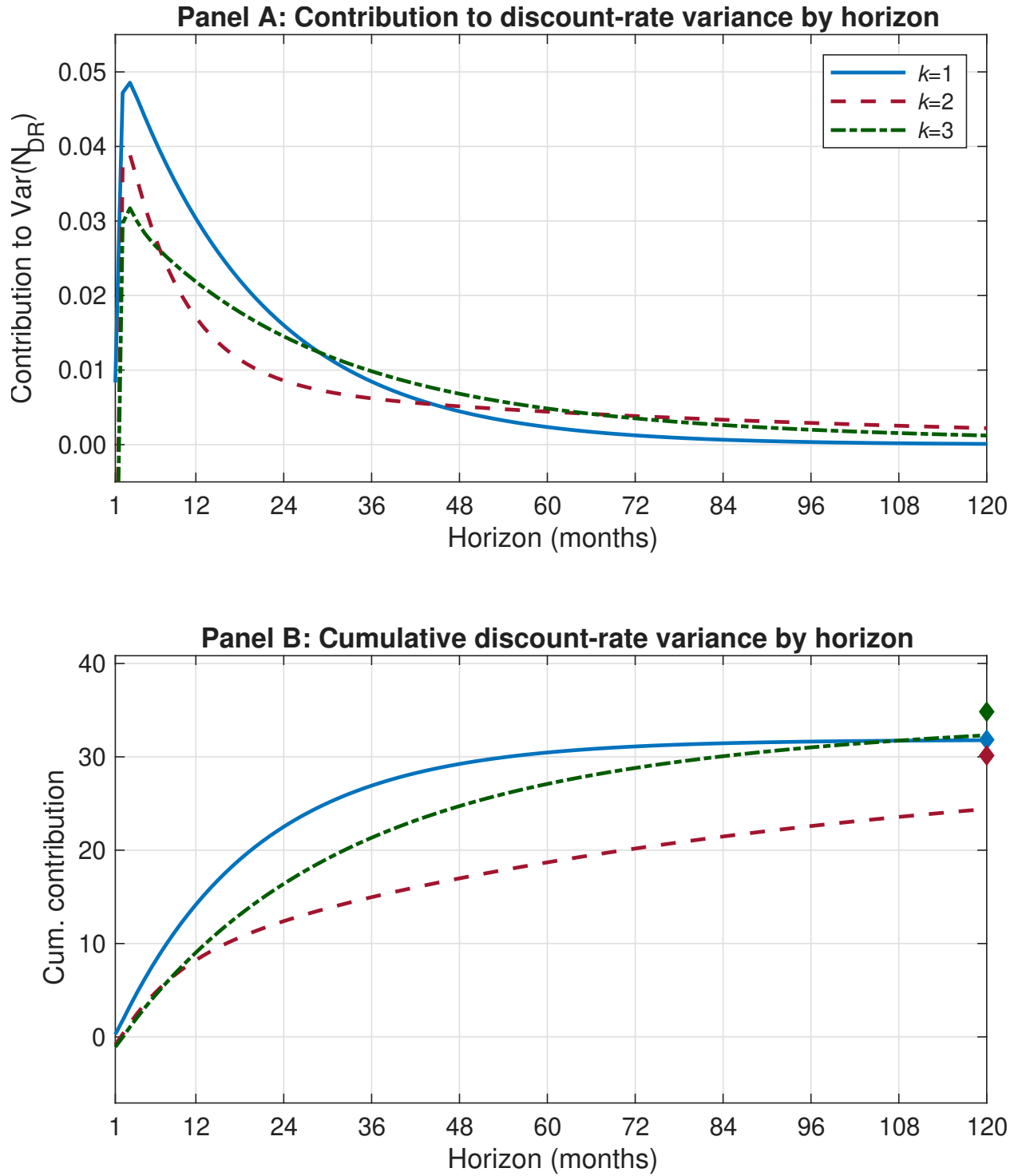


Figure 4. Discount-rate news variance by horizon for one to three persistence groups. The figure decomposes discount-rate news variance by horizon for VARs with one ($k = 1$), two ($k = 2$), and three ($k = 3$) persistence groups. Panel A shows the contribution from each horizon as a share of total variance. Panel B shows the cumulative contribution (in $\%^2$) across horizons, and the diamonds at the right edge show the total discount-rate news variance for each VAR. The sample period is January 1927 to December 2021.

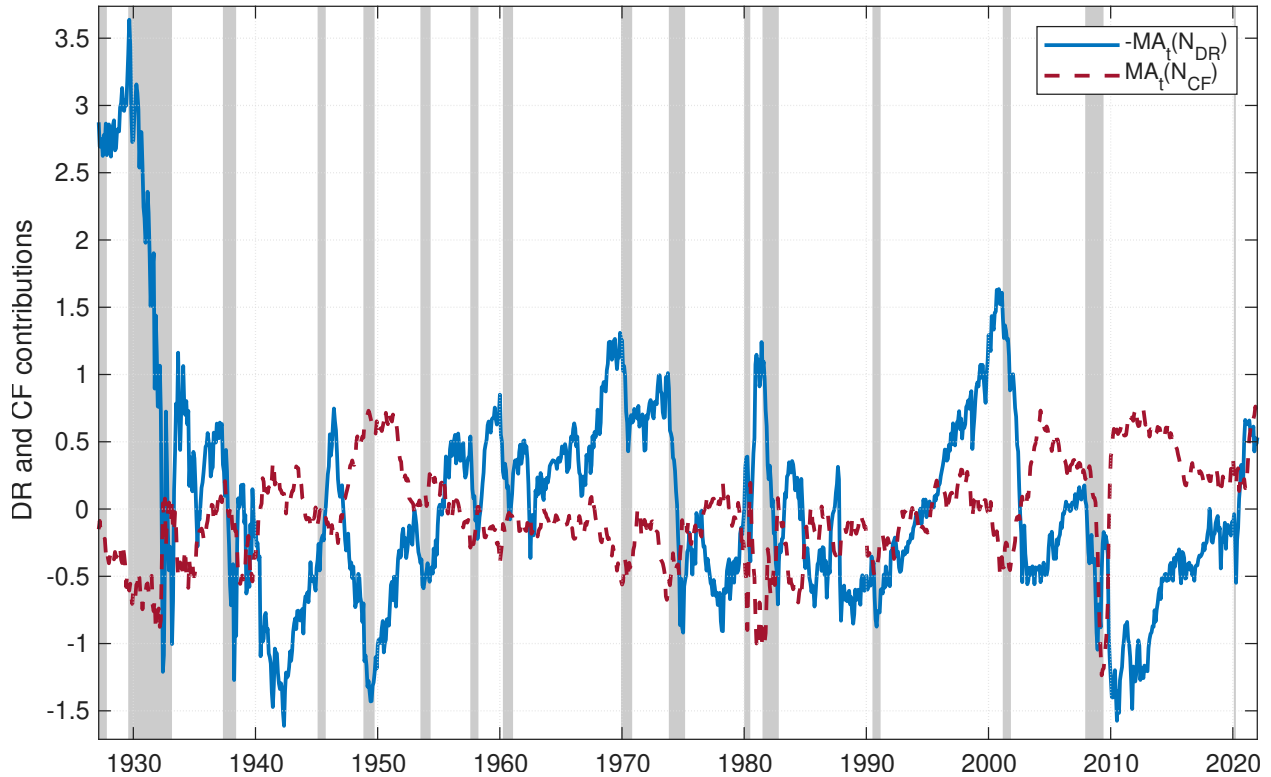


Figure 5. Smoothed discount-rate news and cash-flow news. The figure plots $-MA_t(N_{DR})$ (solid blue line) and $MA_t(N_{CF})$ (dashed red line), the trailing exponentially weighted moving averages of the discount-rate news and cash-flow news estimates, following Campbell, Giglio, and Polk (2013). A positive value of either series indicates a positive contribution to recent excess market returns. NBER recessions are shaded. The sample period is January 1927 to December 2021.

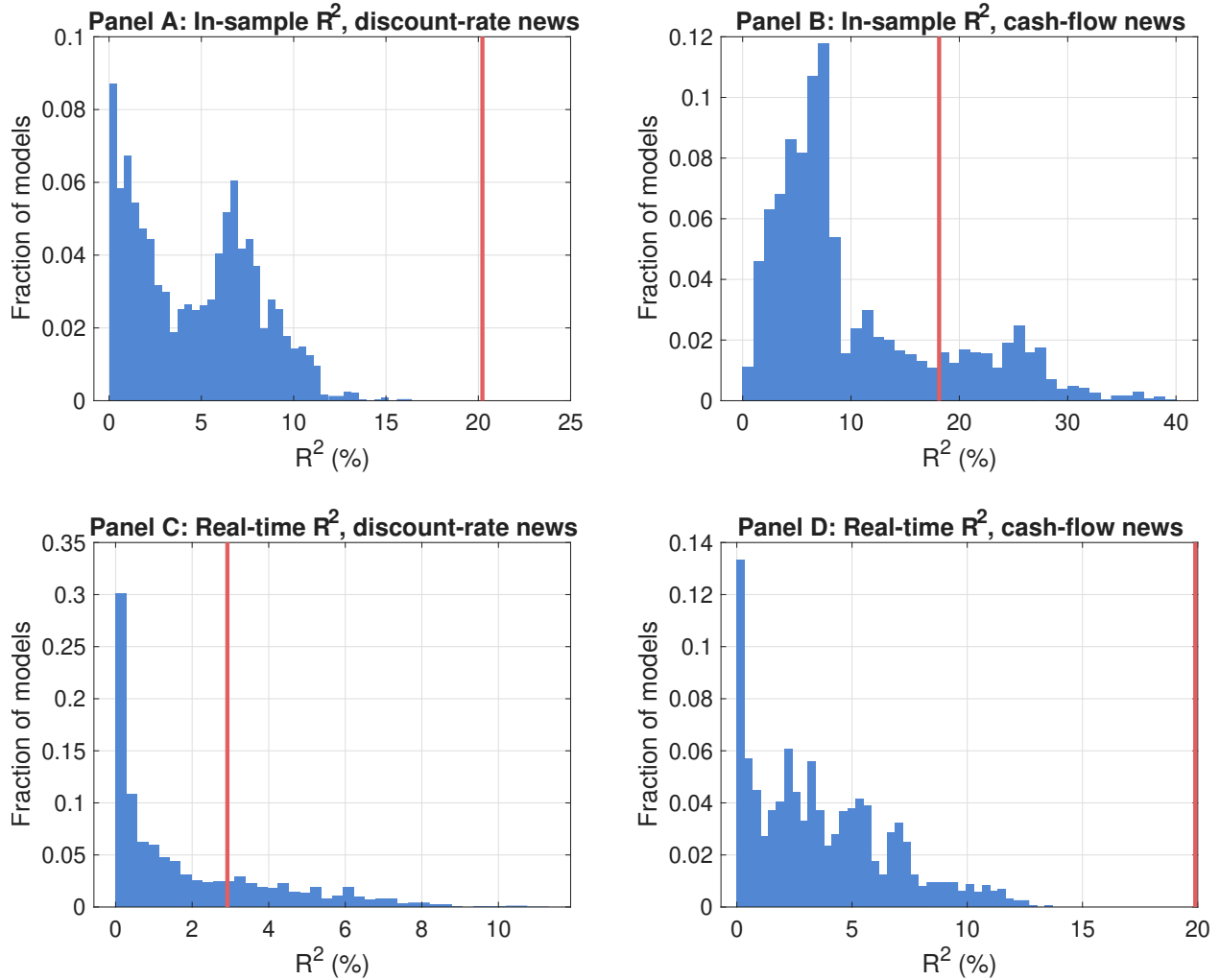


Figure 6. Return and dividend growth forecast R^2 s across models, in sample and in real time. The figure shows R^2 s (in percent) from predictive regressions of annual log excess market returns on smoothed discount-rate news (Panels A and C) and of annual log dividend growth in excess of the log risk-free rate on smoothed cash-flow news (Panels B and D). Panels A and B use in-sample news estimates from the full-sample VAR. Panels C and D use real-time news estimates in which the VAR is re-estimated each month on an expanding window of data available through that month. The blue histograms show the distribution across the 2,298 VARs that use the log excess market return and three of the 25 predictor variables as state variables, and the red line marks the baseline VAR. The sample period is January 1927 to December 2021.

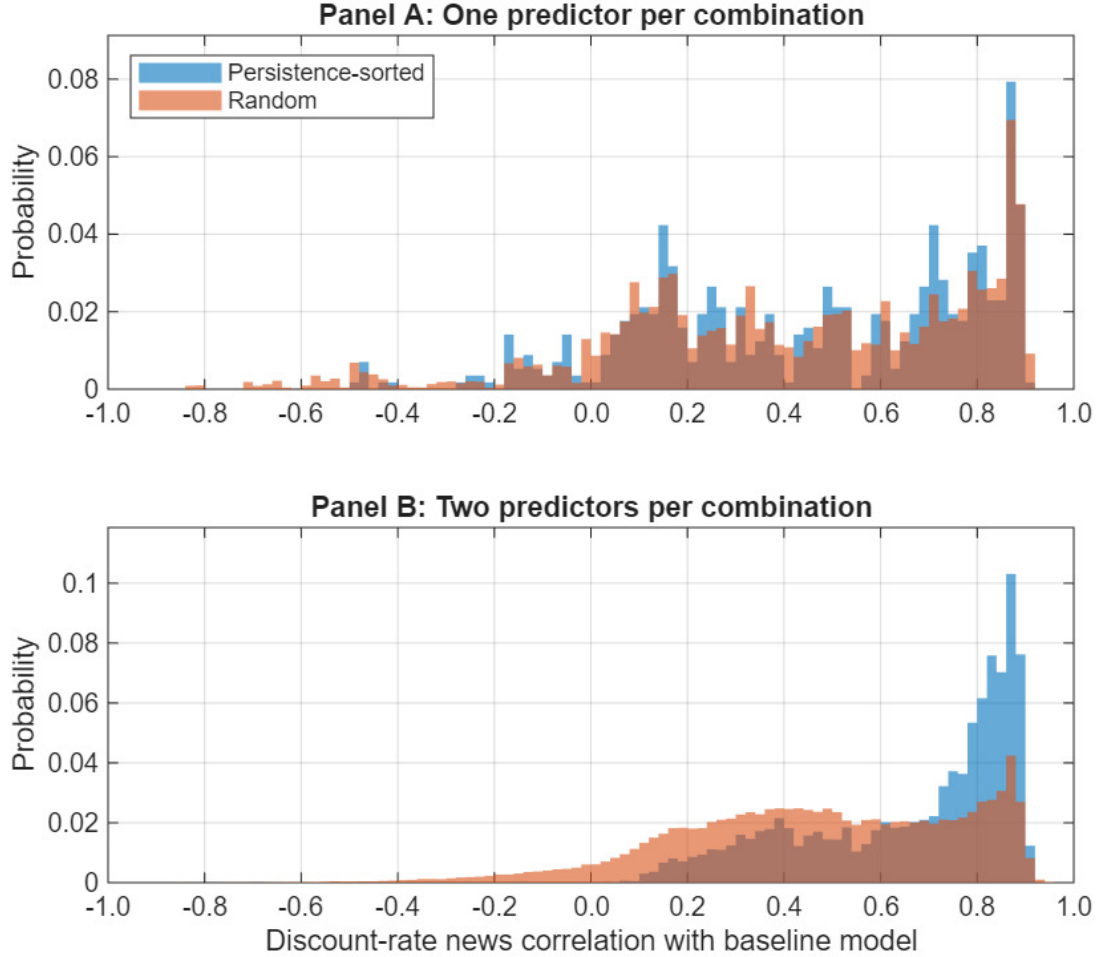


Figure 7. Correlation of discount-rate news with the baseline model by number of predictors per combination. The figure shows histograms of the time-series correlation between the discount-rate news estimates of the baseline model and those of alternative models. All models contain the log excess market return and three forecast combinations as state variables, with the number of predictors per combination ranging from one (Panel A) to five (Panel E). The blue histogram (persistence-sorted) in each panel includes all models that draw the stated number of predictors from each of the three persistence groups. The orange histogram (random) includes 100,000 models that draw the same number of predictors for each combination at random from all 25 predictors, ignoring the persistence groups. The sample period is January 1927 to December 2021.

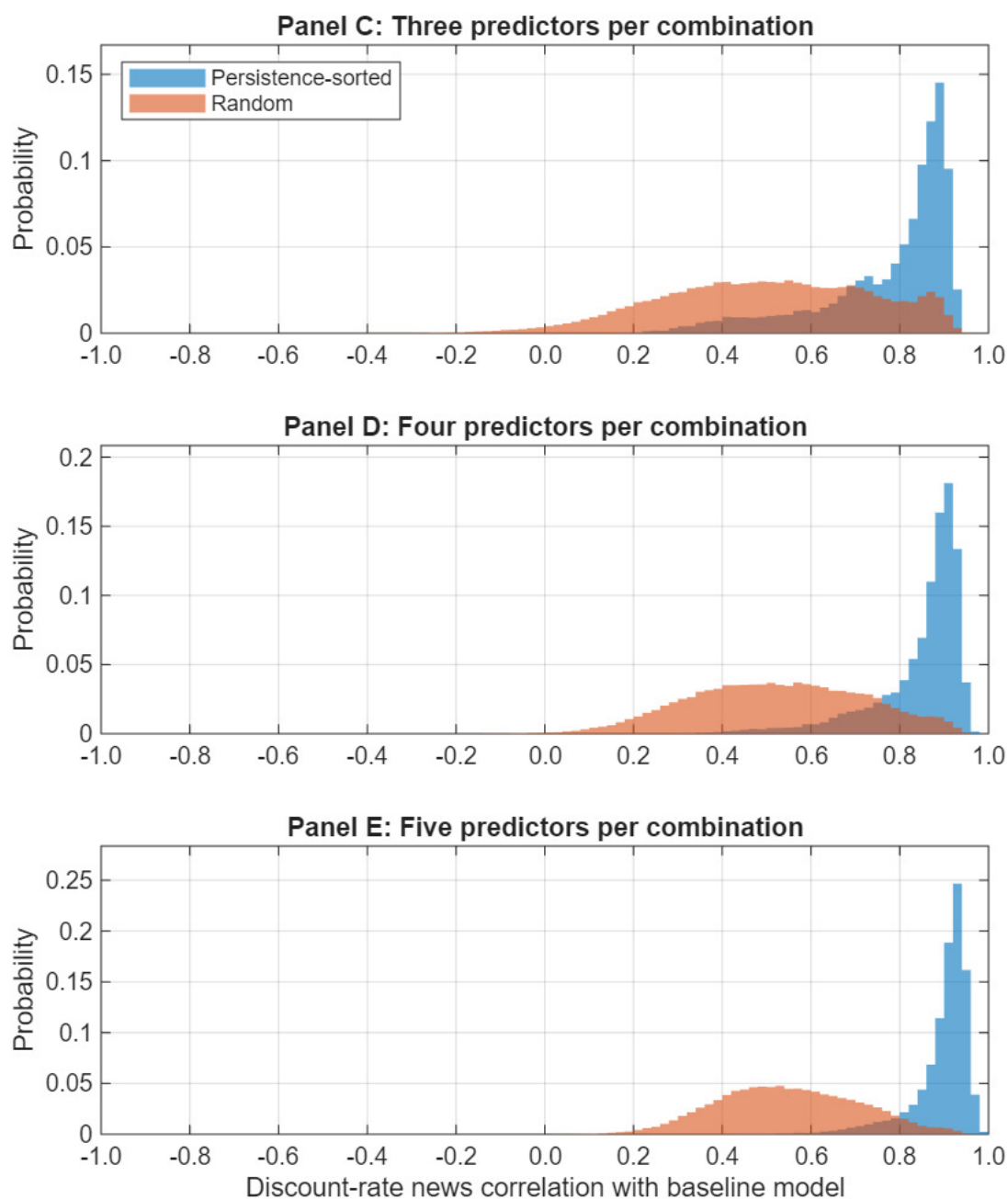


Figure 7. Correlation of discount-rate news with the baseline model by number of predictors per combination (continued).

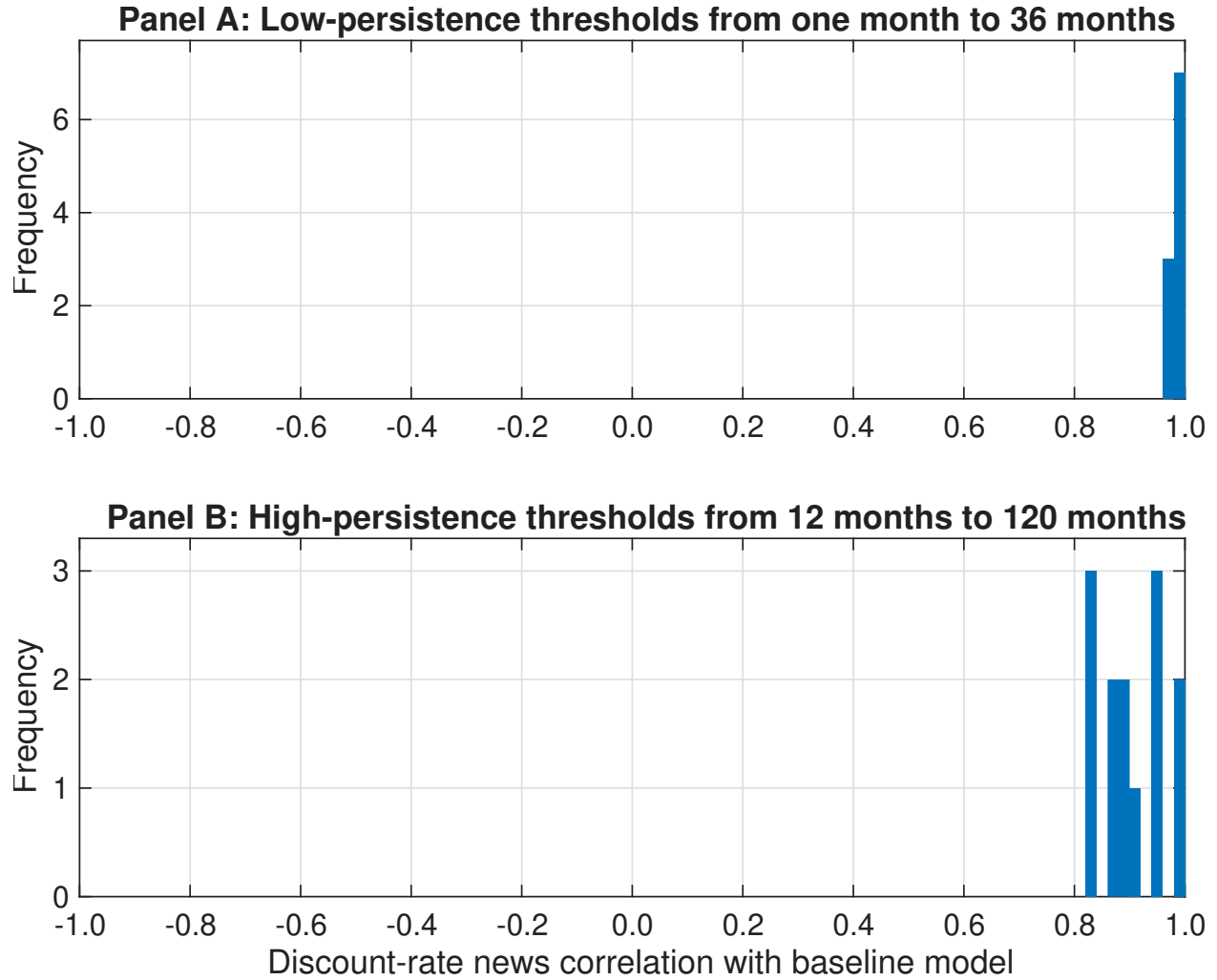


Figure 8. Correlation of discount-rate news with the baseline model across persistence cutoffs. The figure shows histograms of the time-series correlation between the discount-rate news estimates of the baseline model and those of models with alternative half-life cutoffs for the persistence groups. All models contain the log excess market return and three forecast combinations as state variables. Panel A keeps the high-persistence cutoff at its baseline value of 48 months and varies the low-persistence cutoff from one month to 36 months. Panel B keeps the low-persistence cutoff at its baseline value of three months and varies the high-persistence cutoff from 12 months to 120 months. The sample period is January 1927 to December 2021.

Internet Appendix

“Taming the VAR”

A Bias-variance tradeoff

This appendix derives the MSE decomposition in equation (9), gives conditions under which the VAR recovers spanned discount-rate news exactly, characterizes the return-channel remainder, and proves the five claims of Proposition 1.

A.1 Setup

The first-order VAR stacks the excess return and the forecast combinations in the state vector $z_t = (r_t^e, \tilde{r}_{1,t+1}^e, \dots, \tilde{r}_{k,t+1}^e)'$ and estimates the transition matrix $\hat{A}$ with residual $\hat{\varepsilon}_{t+1} = z_{t+1} - \hat{A}z_t$, where variables are demeaned and intercepts are suppressed throughout the appendix. The news estimate reads news off the return equation,

$$\hat{N}_{DR,t+1} = e_1' \rho \hat{A} (I - \rho \hat{A})^{-1} \hat{\varepsilon}_{t+1}, \quad (\text{A1})$$

where e_1 selects the first (return) component. Throughout, all variables are covariance stationary with finite second moments, and the population VAR satisfies $\rho \max |\text{eig}(A)| < 1$, so the discounted sum in the news operator converges. Let μ_{t+1} denote the conditional expectation of $\hat{N}_{DR,t+1}$ given the evaluation period, meaning all date- t levels and date- $(t+1)$ innovations, spanned and unspanned (the date- t predictor levels x_t , return r_t^e , and unspanned component q_t , and the date- $(t+1)$ predictor innovations v_{t+1} , return innovation $u_{r,t+1} = r_{t+1}^e - E_t r_{t+1}^e$, and the unspanned news $n_{q,t+1}$). Write the target from equation (8) as $N_{DR,t+1} = \tilde{N}_{DR,t+1} + n_{q,t+1}$ with $\tilde{N}_{DR,t+1} = \sum_{i=1}^K \omega_i \beta_i v_{i,t+1}$ the part of news spanned by the K predictor innovations. Then

$$\hat{N}_{DR,t+1} - N_{DR,t+1} = (\hat{N}_{DR,t+1} - \mu_{t+1}) + (\mu_{t+1} - \tilde{N}_{DR,t+1}) - n_{q,t+1}. \quad (\text{A2})$$

Because μ_{t+1} is a conditional expectation given the evaluation period, the first term has conditional mean zero by construction, while the second term and $n_{q,t+1}$ are functions of the evaluation period alone. By iterated expectations, both cross terms involving $\hat{N}_{DR,t+1} - \mu_{t+1}$ are exactly zero; this uses only the definition of the conditional mean and holds under full-sample estimation. The cross term between $\mu_{t+1} - \tilde{N}_{DR,t+1}$ and $n_{q,t+1}$ depends on the dynamics of the predictors and of the unspanned component, which we now restrict.

Assumption 1. (i) Each predictor follows equation (5) with $E_t[v_{i,t+1}] = 0$ for all i , where E_t is the conditional expectation operator of equation (1); equivalently, $E_t[x_{i,t+1}] = \phi_i x_{i,t}$, so the conditional-mean dynamics of the predictor vector are diagonal. (ii) The predictors do not forecast the unspanned component: $\text{Cov}(q_{t+j}, x_{i,t}) = 0$ for all i and $j \geq 1$, for each

*predictor set under consideration.*¹

Assumption 1 defines the benchmark environment in which we characterize the MSE components, but the estimation procedure of Section 2.3 does not impose it. Under Assumption 1, the predictor innovations are orthogonal to $n_{q,t+1}$: for each horizon $j \geq 1$,

$$\text{Cov}((E_{t+1} - E_t)q_{t+j}, v_{i,t+1}) = \text{Cov}(q_{t+j}, v_{i,t+1}) = \text{Cov}(q_{t+j}, x_{i,t+1}) - \phi_i \text{Cov}(q_{t+j}, x_{i,t}) = 0, \quad (\text{A3})$$

where the first equality holds because $E_t q_{t+j}$ is a date- t quantity uncorrelated with the innovation $v_{i,t+1}$ by Assumption 1(i), and the final equality applies the projection defining q_t to the first covariance at $j = 1$ and Assumption 1(ii) to all remaining covariances. Summing over horizons with weights ρ^j gives $\text{Cov}(n_{q,t+1}, v_{i,t+1}) = 0$. The return-equation innovation, however, contains the unexpected return and hence $n_{q,t+1}$, so the cross term between $\mu_{t+1} - \tilde{N}_{DR,t+1}$ and $n_{q,t+1}$ is generally non-zero. The news operator loads on it, and the decomposition carries the remainder $\Delta = -2E[(\mu_{t+1} - \tilde{N}_{DR,t+1})n_{q,t+1}]$. The expected square of equation (A2) therefore gives

$$\begin{aligned} \underbrace{E[(\hat{N}_{DR,t+1} - N_{DR,t+1})^2]}_{\text{MSE}} &= \underbrace{E[(\mu_{t+1} - \tilde{N}_{DR,t+1})^2]}_{B(k, \{G_l\}, T)} + \underbrace{E[(\hat{N}_{DR,t+1} - \mu_{t+1})^2]}_{V(k, \{G_l\}, T)} + \underbrace{E[n_{q,t+1}^2]}_{S(K)} \\ &\quad + \underbrace{[-2E[(\mu_{t+1} - \tilde{N}_{DR,t+1})n_{q,t+1}]]}_{\Delta(K, k, \{G_l\}, T)}. \end{aligned}$$

Here, B is the squared bias (with limit $E[(\bar{N}_{DR,t+1} - \tilde{N}_{DR,t+1})^2]$ as $T \rightarrow \infty$, where $\bar{N}_{DR,t+1} = e'_1 \rho A (I - \rho A)^{-1} \varepsilon_{t+1}$ uses the population VAR coefficients $A = \text{plim } \hat{A}$), V is the sampling variation of the news estimate (and goes to zero as $T \rightarrow \infty$), S is irreducible given a predictor set, and Δ is the return-channel remainder characterized in Appendix A.4.²

The terms of the MSE measure distinct errors for a common realization of the evaluation-period shocks. The target $\tilde{N}_{DR,t+1} = \sum_{i=1}^K \omega_i \beta_i v_{i,t+1}$ weights each predictor's shock by its own persistence amplification ω_i and its multiple regression coefficient β_i . The estimate $\hat{N}_{DR,t+1}$ applies the news operator $e'_1 \rho \hat{A} (I - \rho \hat{A})^{-1}$ to fitted innovations of one sample, and μ_{t+1} is the average of this estimate over estimation samples at the same evaluation-period shocks. The variance is the dispersion of this estimate across samples, and the bias is the systematic gap between the weight the procedure places on each predictor's shock, through its group's persistence and loading, and the target weight $\omega_i \beta_i$. Because the VAR carries k forecast states rather than K predictors, it cannot reproduce the per-predictor weights: it imposes one persistence per group, and it fits a single loading on each group's mean forecast rather than separate coefficients on the individual predictors. The irreducible term is the news no predictor is informative about. The remainder is the return channel: the news operator also loads on the return-equation innovation, which is the unexpected return and contains unspanned news.

¹Condition (ii) weakens as the predictor set grows. Better spanning shrinks q_t , and the condition holds trivially when spanning is complete.

²At finite T , B differs from its population limit by an $O(1/T)$ term of undetermined sign, arising from $\hat{A}$'s finite-sample bias and sampling variance. The proofs of Proposition 1 use the population limit of B and approximate V to first order in the sampling error of $\hat{b}_i$ and $\hat{A}$.

A.2 Construction

The methodology splits the K predictors into groups $G_1, \dots, G_k$ and forms one forecast combination per group as the mean univariate forecast,

$$\tilde{r}_{l,t+1}^e = \sum_{i \in G_l} c_i x_{i,t}, \quad c_i = b_i / K_l, \quad (\text{A4})$$

a time- t state variable that forecasts the time- $t+1$ return, where b_i is the univariate predictive coefficient of equation (10). Let ϕ_l be the population first-order regression coefficient for the group- l state variable from a regression of $\tilde{r}_{l,t+2}^e$ on $\tilde{r}_{l,t+1}^e$, let $\omega_l = \rho / (1 - \rho\phi_l)$ be the corresponding amplification weight, and let θ_l denote the population predictive coefficient for group l from the VAR return equation. Substituting the AR(1) from equation (5) into the group forecast, its one-period residual is

$$\tilde{r}_{l,t+2}^e - \phi_l \tilde{r}_{l,t+1}^e = \sum_{i \in G_l} c_i v_{i,t+1} + \sum_{i \in G_l} c_i (\phi_i - \phi_l) x_{i,t}. \quad (\text{A5})$$

A.3 Exact recovery in the benchmark

When within-group persistence is equal ($\phi_i = \phi_l$ for all $i \in G_l$ in every group), the one-period residual of each group state variable reduces to $\sum_{i \in G_l} c_i v_{i,t+1}$ [equation (A5)], which is orthogonal to all date- t variables by Assumption 1(i). Each combination row of the population VAR matrix A therefore has ϕ_l on its own lag and zeros elsewhere, the corresponding population innovation is $\sum_{i \in G_l} c_i v_{i,t+1}$, and the return row is $(\alpha, \theta_1, \dots, \theta_k)$, where α is the coefficient on the lagged return. For this A , $(A^j)_{1,1+l} = \theta_l (\phi_l^j - \alpha^j) / (\phi_l - \alpha)$ (for $\phi_l \neq \alpha$; the expression at $\phi_l = \alpha$ follows by continuity), and summing with weights ρ^j gives the news operator weight on group l 's innovation,

$$e'_1 \rho A (I - \rho A)^{-1} e_{1+l} = \frac{\rho \theta_l}{(1 - \rho\phi_l)(1 - \rho\alpha)} = \frac{\omega_l \theta_l}{1 - \rho\alpha}, \quad (\text{A6})$$

while the weight on the return-equation innovation is $\rho\alpha / (1 - \rho\alpha)$.

The following lemma states the conditions under which the design achieves its target: the news estimate from the small VAR coincides exactly with spanned discount-rate news.

Lemma 1. *Impose Assumption 1 and suppose (i) within-group persistence is equal in every group; (ii) the multiple regression coefficients are proportional to the univariate coefficients within every group, $\beta_i = \kappa_l b_i$ for all $i \in G_l$; and (iii) the lagged return carries no marginal forecasting power in the VAR return equation. Then $\bar{N}_{DR,t+1} = \tilde{N}_{DR,t+1}$: the $(k+1)$ -state VAR recovers spanned discount-rate news exactly in population, and the MSE of $\hat{N}_{DR,t+1}$ converges to the spanning floor S as T grows.*

Proof. Under (ii), the spanned part of expected returns is exactly linear in the combinations, $\sum_{i=1}^K \beta_i x_{i,t} = \sum_{l=1}^k \kappa_l K_l \tilde{r}_{l,t+1}^e$. The remaining components of the return, q_t and the return innovation, are orthogonal to the combinations, so under (iii) the population return row is $(0, \theta_1, \dots, \theta_k)$ with $\theta_l = \kappa_l K_l$, and hence $\theta_l c_i = \kappa_l K_l (b_i / K_l) = \beta_i$ for every i . With $\alpha = 0$ and

the combination rows diagonal under (i), the first column of A is zero, so the news operator places no weight on the return-equation innovation and, by equation (A6), weight $\omega_l \theta_l$ on group l 's innovation. Therefore

$$\bar{N}_{DR,t+1} = \sum_{l=1}^k \omega_l \theta_l \sum_{i \in G_l} c_i v_{i,t+1} = \sum_{i=1}^K \omega_{l(i)} \theta_{l(i)} c_i v_{i,t+1} = \sum_{i=1}^K \omega_i \beta_i v_{i,t+1} = \tilde{N}_{DR,t+1},$$

using (i) for $\omega_{l(i)} = \omega_i$ and the proportionality above for $\theta_{l(i)} c_i = \beta_i$. As $T \rightarrow \infty$, μ_{t+1} converges to $\bar{N}_{DR,t+1}$, and $\bar{N}_{DR,t+1} = \tilde{N}_{DR,t+1}$ from above. The limit of B is $E[(\bar{N}_{DR,t+1} - \tilde{N}_{DR,t+1})^2] = 0$, and the limit of Δ is $-2E[(\bar{N}_{DR,t+1} - \tilde{N}_{DR,t+1})n_{q,t+1}] = 0$. The variance term goes to zero as the sample grows. The MSE therefore converges to $S(K)$. $\square$

A.4 Return-channel remainder

As $T \rightarrow \infty$, μ_{t+1} converges to $\bar{N}_{DR,t+1} = e_1' \rho A (I - \rho A)^{-1} \varepsilon_{t+1}$. At finite T , the two differ by an $O(1/T)$ term arising from $\hat{A}$'s finite-sample bias and sampling variance (footnote 2). Therefore, using $\text{Cov}(\bar{N}_{DR,t+1}, n_{q,t+1}) = 0$ from Appendix A.1, $\Delta = -2e_1' \rho A (I - \rho A)^{-1} \text{Cov}(\varepsilon_{t+1}, n_{q,t+1}) + O(1/T)$. Each combination-equation innovation in ε_{t+1} (the innovation of the equation for a forecast combination) is a sum of predictor innovations plus date- t quantities [equation (A5)]. The predictor innovations are orthogonal to $n_{q,t+1}$ by Assumption 1 [equation (A3)], and the date- t quantities are orthogonal to $n_{q,t+1}$ because it is a forecast revision with $E_t n_{q,t+1} = 0$. Only the return equation contributes. Its innovation is the unexpected return plus a date- t term, the gap between the conditional expectation of the return and the VAR's forecast of it. The date- t term is uncorrelated with $n_{q,t+1}$ because $n_{q,t+1}$ is a forecast revision, and by equation (1) and the orthogonality of $\tilde{N}_{DR,t+1}$ to $n_{q,t+1}$,

$$\begin{aligned} \text{Cov}(\varepsilon_{1,t+1}, n_{q,t+1}) &= \text{Cov}(N_{CF,t+1} - N_{DR,t+1}, n_{q,t+1}) \\ &= \text{Cov}(N_{CF,t+1}, n_{q,t+1}) - \text{Cov}(\tilde{N}_{DR,t+1} + n_{q,t+1}, n_{q,t+1}) \\ &= \text{Cov}(N_{CF,t+1}, n_{q,t+1}) - \text{Var}(n_{q,t+1}). \end{aligned} \quad (\text{A7})$$

Therefore,

$$\Delta = 2g_0 [\text{Var}(n_{q,t+1}) - \text{Cov}(N_{CF,t+1}, n_{q,t+1})] + O(1/T), \quad g_0 = e_1' \rho A (I - \rho A)^{-1} e_1, \quad (\text{A8})$$

where g_0 is the weight the news operator places on the return-equation innovation. Under the structure of Appendix A.3, $g_0 = \rho \alpha / (1 - \rho \alpha)$, which is zero under condition (iii) of Lemma 1. Three properties follow. First, the term of Δ that does not shrink with the sample size is zero when $g_0 = 0$, which holds when the lagged return contributes no forecasting power to the VAR (the first column of A is zero). Second, the leading term of Δ is bounded by $2|g_0| \sqrt{S} (\sqrt{S} + \sqrt{E[N_{CF,t+1}^2]}) = 2|g_0| (S + \sqrt{S} \sqrt{E[N_{CF,t+1}^2]})$ by the Cauchy-Schwarz inequality, so Δ shrinks as spanning improves and goes to zero with S : the large predictor set that motivates the methodology also limits the remainder. Third, the sign of Δ is indeterminate, depending on the sign of g_0 and on the covariance between cash-flow news and unspanned discount-rate news. The design choices affect the term of Δ that does not shrink with the

sample size only through g_0 , and Δ is outside the claims of Proposition 1, which concerns B , V , and S .

A.5 Proof of Proposition 1

Proof of part (a)

By equation (6), $S(K) = E[n_{q,t+1}^2]$ is the variance of news in the residual q_t from the population regression of expected returns on the K predictors. Under Assumption 1, $\text{Cov}(n_{q,t+1}, v_{i,t+1}) = 0$ for every predictor in the set [equation (A3)], so equation (8) decomposes $N_{DR,t+1}$ into a term that is spanned by $\{v_{1,t+1}, \dots, v_{K,t+1}\}$ plus a residual that is orthogonal to these innovations. Thus, $\sum_{i=1}^K \omega_i \beta_i v_{i,t+1}$ is the linear projection of $N_{DR,t+1}$ onto the space of the K predictor innovations, and $S(K)$ is the variance of the residual from this projection. Adding a predictor adds its innovation to this set and leaves the others unchanged, because each $v_{i,t+1}$ is defined by predictor i 's own autoregression [equation (5)]. The target $N_{DR,t+1}$ is fixed by equation (1), so its projection is onto a weakly larger space as K grows and the residual variance weakly falls, $S(K) \geq S(K+1)$. The sorting and aggregation rules combine the same predictors into state variables without changing the predictor set, so they leave q_t and S unchanged. $\square$

Proof of part (b)

Sampling error in $\hat{N}_{DR,t+1}$ comes from the estimated coefficients at two levels: the predictive coefficients $\hat{b}_i$ that determine the state variables through forecast combination, and the VAR matrix $\hat{A}$ that maps the innovations into discount-rate news. Proposition 1(b) concerns the first. With fixed k , adding a predictor to a group changes that group's state variable but not the number of state variables, so $\hat{A}$ has $(k+1)^2$ coefficients regardless of K_l . The dimension channel is treated separately in the proof of part (d).

Group l 's combination forecast carries the estimation error $\frac{1}{K_l} \sum_{i \in G_l} e_i$ relative to its population combination, where $e_i = (\hat{b}_i - b_i)x_{i,t}$ is the sampling error in predictor i 's fitted forecast: the univariate slope error from equation (10) times the evaluation-date predictor level, with the evaluation period held fixed. This error passes into the group's state variable, and the news operator introduces its innovation into $\hat{N}_{DR,t+1}$ with the group's amplification weight, so it contributes to V . Let $\bar{V}_l$ denote the average of $\text{Var}(e_i)$ over $i \in G_l$ and $\bar{C}_l$ denote the average of $\text{Cov}(e_i, e_j)$ over pairs $i \neq j$ in G_l . Expanding the variance of the average gives the identity

$$\text{Var}\left(\frac{1}{K_l} \sum_{i \in G_l} e_i\right) = \bar{C}_l + \frac{\bar{V}_l - \bar{C}_l}{K_l},$$

with $\bar{C}_l \leq \bar{V}_l$ [since $\text{Cov}(e_i, e_j) \leq \frac{1}{2}\text{Var}(e_i) + \frac{1}{2}\text{Var}(e_j)$] and equality only when the e_i are perfectly positively correlated with a common variance. The combination's noise is therefore the common comovement of the errors plus an idiosyncratic contribution weighted by $1/K_l$. Averaging removes the idiosyncratic differences among the estimation errors and retains their common component, the standard variance reduction from combining forecasts [Timmermann (2006)]. Adding a predictor to the group increases K_l and shrinks the weight

on the idiosyncratic contribution, but it also changes $\bar{V}_l$ and $\bar{C}_l$, so the net change in the combination's noise is not signed in general. The reduction in the idiosyncratic component is the sense in which averaging tends to reduce V . $\square$

Proof of part (c)

Lemma 2. *With non-collinear predictors, the residual of the group state variable $\tilde{r}_{l,t+2}^e - \phi_l \tilde{r}_{l,t+1}^e$ is orthogonal to every time- t predictor level if and only if ϕ_i is equal across $i \in G_l$ with $c_i \neq 0$. In particular, the group state variable is exactly AR(1) when within-group persistence is equal.*

Proof. The residual is the right-hand side of equation (A5). Its first sum is orthogonal to all time- t variables by Assumption 1(i). The second sum, $\sum_{i \in G_l} c_i(\phi_i - \phi_l)x_{i,t}$, is measurable at time t and orthogonal to every $x_{j,t}$ only if it is zero, which, with non-collinear predictors, requires $\phi_i = \phi_l$ for all $i \in G_l$ with $c_i \neq 0$. $\square$

When within-group persistence is equal, Lemma 2 gives an exact AR(1) for the group state variable with coefficient $\phi_l = \phi_i$, so the group inherits each predictor's persistence and the methodology incurs no persistence bias. If within-group persistence is not equal, imposing a single ϕ_l affects the news estimate through two channels, both functions of the within-group gaps $\phi_i - \phi_l$.

First, the second sum in equation (A5) is a component of the group state variable's innovation that is measurable at time t . If the remaining state variables in the VAR predictive regression do not completely remove this component, the remainder enters the VAR innovation and contributes to B a predictable quantity that the methodology misclassifies as news. This term is zero when $\phi_i = \phi_l$ for all $i \in G_l$ and continuous in the gaps $\phi_i - \phi_l$.

Second, the group persistence provides a common amplification to the predictor innovations. The target weights predictor i 's shock by its own amplification $\omega_i = \rho/(1 - \rho\phi_i)$, whereas the group applies the common $\omega_l = \rho/(1 - \rho\phi_l)$, a gap $\omega_l - \omega_i$ for each $t + 1$ shock. Since ω is continuous in ϕ , this gap goes to zero as $\phi_l \rightarrow \phi_i$ and is continuous in the gaps $\phi_i - \phi_l$.

Both channels are governed by the within-group dispersion in ϕ_i and go to zero as it shrinks, so the bias from imposing a common group-level persistence goes to zero as within-group dispersion in ϕ_i shrinks.

It remains to show that the smallest within-group dispersion in ϕ_i for a given k is attained by a persistence-sorted partition. The best grouping puts predictors with similar persistence together, since dispersion penalizes spread around each group mean. In one dimension, the best groups are consecutive ranges of persistence with no overlap, so they can be formed by sorting predictors on ϕ_i and dividing the sorted list into k groups [Fisher (1958)]. Sorting on any other characteristic gives groups that are consecutive in that characteristic, not in persistence, and so reduces persistence dispersion only when the characteristic happens to correlate with persistence. Random assignment ignores persistence altogether.

At zero dispersion the persistence-related bias components are zero but B may still be positive. By equation (A6), the news operator weights group l 's innovation by $\omega_l\theta_l/(1 - \rho\alpha)$. The factor $1/(1 - \rho\alpha)$ is common to every group and equals one when the lagged return carries no forecasting power, so we absorb it into the loadings. Therefore, $\bar{N}_{DR,t+1}$ weights

predictor i 's shock by $\omega_{l(i)}\theta_{l(i)}c_i$, where $\theta_{l(i)}$ is the coefficient on the forecast combination containing predictor i , whereas $\tilde{N}_{DR,t+1}$ weights it by $\omega_i\beta_i$ with $\omega_i = \omega_{l(i)}$ at zero dispersion. The remaining gap is between $\theta_{l(i)}c_i$ and β_i . This coefficient channel reflects the aggregation rule rather than the sorting rule, and is a component of B that does not go to zero as persistence dispersion shrinks. The gap $\theta_{l(i)}c_i - \beta_i$ is zero for every predictor in a group only if the univariate coefficients are proportional to the multiple regression coefficients within that group: with $c_i = b_i/K_l$, the requirement $\theta_l c_i = \beta_i$ for all $i \in G_l$ forces $\beta_i = (\theta_l/K_l)b_i$. The single loading θ_l can rescale the combination but cannot reweight predictors within it. In addition, $\tilde{N}_{DR,t+1}$ loads on the return-equation innovation, which $\tilde{N}_{DR,t+1}$ excludes. The systematic part of this channel contributes to B , and its covariance with $n_{q,t+1}$ is the remainder Δ of equation (A8). Proposition 1(c) concerns the persistence channel. $\square$

Proof of part (d)

Fix the predictor set. Throughout, we work within the exact-recovery benchmark of Lemma 1, imposing its conditions where stated. We compare a k -group partition with the $(k+1)$ -group partition that divides one of its groups into two and leaves the others unchanged. The state variables are standardized in the estimation (Section 2.3). Standardization rescales each state variable by a constant, which leaves the news estimate unchanged. Here, we work in unstandardized units for convenience.

The k -group VAR is the $(k+1)$ -group VAR estimated under linear restrictions, and the rest of the proof compares the two on that basis. Let z_t and z_t^+ denote the state vectors of the k -group and $(k+1)$ -group VARs, respectively, and let G_l be the group of the k -group partition that the $(k+1)$ -group partition separates into G_a and G_b , with $K_a + K_b = K_l$. Because each combination is an equal-weighted mean, $\tilde{r}_{l,t+1}^e = (K_a/K_l)\tilde{r}_{a,t+1}^e + (K_b/K_l)\tilde{r}_{b,t+1}^e$. Without loss of generality, order the groups so that G_l is the last group of the k -group partition and G_a and G_b are the last two groups of the $(k+1)$ -group partition. Let M be the $(k+1) \times (k+2)$ matrix that satisfies $z_t = Mz_t^+$,

$$M = \begin{pmatrix} I_k & 0 & 0 \\ 0 & K_a/K_l & K_b/K_l \end{pmatrix}.$$

The k -group VAR regresses z_{t+1} on z_t , which is equivalent to regressing Mz_{t+1}^+ on Mz_t^+ ,

$$Mz_{t+1}^+ = AMz_t^+ + \varepsilon_{t+1}, \tag{A9}$$

so the matrix of coefficients on z_t^+ is AM . Writing $A = \begin{bmatrix} A_{(1:k+1,1:k)} & A_{(1:k+1,k+1)} \end{bmatrix}$, where $A_{(1:k+1,k+1)}$ is the last column of A (i.e., the coefficients on $\tilde{r}_{l,t+1}^e$ in the $k+1$ equations),

$$AM = \begin{bmatrix} A_{(1:k+1,1:k)} & \frac{K_a}{K_l}A_{(1:k+1,k+1)} & \frac{K_b}{K_l}A_{(1:k+1,k+1)} \end{bmatrix}.$$

The last two columns of AM are proportional in every row, with the ratio K_a/K_b , which is the restriction under which the k -group VAR estimates the $k+1$ equations. The $(k+1)$ -group VAR regresses z_{t+1}^+ on z_t^+ with coefficient matrix A^+ . To compare the two systems on

the same equations, premultiply both sides of the $(k+1)$ -group VAR by M ,

$$Mz_{t+1}^+ = MA^+z_t^+ + M\varepsilon_{t+1}^+. \quad (\text{A10})$$

Reconciling equations (A9) and (A10), the k -group and $(k+1)$ -group systems agree in population if and only if $MA^+ = AM$, in which case $\varepsilon_{t+1} = M\varepsilon_{t+1}^+$.

Bias. Suppose the $(k+1)$ -group partition satisfies the three conditions of Lemma 1. The k -group partition merges G_a and G_b into G_l , so it satisfies the same conditions if and only if $\phi_a = \phi_b$ [condition (i) for G_l] and $\kappa_a = \kappa_b$ [condition (ii) for G_l], since condition (iii) carries over to the k -group VAR when condition (ii) holds for G_l . When $\phi_a = \phi_b$ and $\kappa_a = \kappa_b$, both VARs recover spanned discount-rate news exactly, so $B = 0$ under both partitions and there is no bias to reduce. When either equality fails, the $(k+1)$ -group VAR still recovers spanned news exactly, whereas the k -group VAR carries the bias of a common persistence in G_l when $\phi_a \neq \phi_b$ and the bias of a common loading when $\kappa_a \neq \kappa_b$ [proof of part (c)], so $B > 0$ for the k -group VAR and the additional group removes it. The reduction is small when G_a and G_b differ little in ϕ , by the continuity established in the proof of part (c), or in κ , because the coefficient-channel gap is then proportional to $\kappa_a - \kappa_b$.

Variance. Suppose both partitions satisfy the conditions of Lemma 1, so that $\phi_a = \phi_b = \phi_l$ and $\kappa_a = \kappa_b = \kappa_l$. Under these conditions, Appendix A.3 gives the population innovation of each group state variable as $\sum_{i \in G_l} c_i v_{i,t+1}$ with $c_i = b_i/K_l$, so the innovation of the equation for $\tilde{r}_{l,t+1}^e$ is the weighted average of the innovations of the equations for $\tilde{r}_{a,t+1}^e$ and $\tilde{r}_{b,t+1}^e$:

$$\begin{aligned} \sum_{i \in G_l} c_i v_{i,t+1} &= \frac{1}{K_l} \left(\sum_{i \in G_a} b_i v_{i,t+1} + \sum_{i \in G_b} b_i v_{i,t+1} \right) \\ &= \frac{K_a}{K_l} \sum_{i \in G_a} \frac{b_i}{K_a} v_{i,t+1} + \frac{K_b}{K_l} \sum_{i \in G_b} \frac{b_i}{K_b} v_{i,t+1}, \end{aligned}$$

where the two sums on the right are the innovations for $\tilde{r}_{a,t+1}^e$ and $\tilde{r}_{b,t+1}^e$. The proof of Lemma 1 further gives the return equation's residual as q_t plus the return innovation under either partition. Hence, $\varepsilon_{t+1} = M\varepsilon_{t+1}^+$ and equations (A9) and (A10) imply $MA^+ = AM$ because the state variables are not collinear. From $MA^+ = AM$, induction gives $A^j M = MA^{+j}$ for every j , so $(I - \rho A)^{-1} M = M(I - \rho A^+)^{-1}$, and because $e_1' M = e_1'$, premultiplying by $e_1' \rho A$ gives $e_1' \rho A (I - \rho A)^{-1} M = e_1' \rho A^+ (I - \rho A^+)^{-1}$.

Conditional on the first-stage estimates $\hat{b}_i$, which the two VARs share, expand each news estimate to first order in its coefficient sampling error (footnote 2). For the $(k+1)$ -group VAR, let $\delta^+ = \hat{A}^+ - A^+$ be the coefficient sampling error, so that $\hat{\varepsilon}_{t+1}^+ = \varepsilon_{t+1}^+ - \delta^+ z_t^+$. To first order, $(I - \rho \hat{A}^+)^{-1} = (I - \rho A^+)^{-1} + (I - \rho A^+)^{-1} \rho \delta^+ (I - \rho A^+)^{-1}$. Because $\rho \hat{A}^+ (I - \rho \hat{A}^+)^{-1} = (I - \rho \hat{A}^+)^{-1} - I$, the estimated news operator differs from the population operator by $(I - \rho A^+)^{-1} \rho \delta^+ (I - \rho A^+)^{-1}$ to first order. Writing the estimated operator as the population operator plus this difference, and the fitted innovation as $\varepsilon_{t+1}^+ - \delta^+ z_t^+$, to first order the product has three terms: the population news, the difference in the operator

applied to ε_{t+1}^+ , and the population operator applied to $-\delta^+ z_t^+$. Hence

$$\hat{N}_{DR,t+1}^+ - \bar{N}_{DR,t+1}^+ = \rho e_1' (I - \rho A^+)^{-1} \delta^+ (I - \rho A^+)^{-1} \varepsilon_{t+1}^+ - e_1' \rho A^+ (I - \rho A^+)^{-1} \delta^+ z_t^+$$

to first order. From $(I - \rho A)^{-1} M = M(I - \rho A^+)^{-1}$ and $e_1' M = e_1'$, we have $e_1' (I - \rho A^+)^{-1} = e_1' (I - \rho A)^{-1} M$ and $e_1' \rho A^+ (I - \rho A^+)^{-1} = e_1' \rho A (I - \rho A)^{-1} M$, so

$$\hat{N}_{DR,t+1}^+ - \bar{N}_{DR,t+1}^+ = \rho e_1' (I - \rho A)^{-1} M \delta^+ (I - \rho A^+)^{-1} \varepsilon_{t+1}^+ - e_1' \rho A (I - \rho A)^{-1} M \delta^+ z_t^+, \quad (\text{A11})$$

which depends on δ^+ only through $M\delta^+$, the sampling error in the coefficients of the $k+1$ equations in equation (A10). For the k -group VAR, let $\delta = \hat{A} - A$. The same expansion, with $\varepsilon_{t+1} = M\varepsilon_{t+1}^+$, $z_t = Mz_t^+$, and $(I - \rho A)^{-1} M = M(I - \rho A^+)^{-1}$, gives equation (A11) with δM in place of $M\delta^+$. The two first-order errors are therefore the same linear function of the error in estimating the coefficient matrix $AM = MA^+$ of the $k+1$ shared equations, which the $(k+1)$ -group VAR estimates without restriction, as $M\hat{A}^+$ with error $M\delta^+$, and the k -group VAR estimates under the restriction, as $\hat{A}M$ with error δM .

With identical regressors in every equation and conditionally homoskedastic innovations, the restricted least-squares estimator has weakly smaller covariance than the unrestricted one [Greene (2018)]. The difference between the two covariance matrices is positive semidefinite, so relaxing the restriction weakly raises the sampling variance of every coefficient already in the system and adds the sampling variance of the new ones. Because the two news errors are the same linear function of those coefficients, the leading term of V is weakly larger with $k+1$ groups than with k . $\square$

Proof of part (e)

The amplification weight satisfies $d\omega_l/d\phi_l = \rho^2/(1 - \rho\phi_l)^2 = \omega_l^2 > 0$, so ω_l is increasing in ϕ_l , and increasingly steeply: an error in the persistence of a high-persistence group state variable moves its amplification weight with sensitivity ω_l^2 , orders of magnitude larger than for a low-persistence group. The news operator amplifies a group's innovation through the persistence of the group state variable, by a factor that is increasing in ϕ_l . The bias and variance terms of the MSE both inherit this weighting: from part (c) the misclassified predictable component enters B amplified by the group's persistence, and from part (b) the group's estimation noise enters V amplified the same way. Both contributions are therefore largest in the high-persistence groups, so B and V concentrate there.

For a fixed predictor set and number of groups, the persistence bias contribution scales with within-group ϕ_i dispersion amplified by the group's persistence. Therefore, dispersion is most costly at high ϕ_l , favoring narrow bins for high-persistence groups. The variance contribution scales with estimation noise, amplified by the group's persistence. By equation (A6), an error in the estimated persistence of group l enters the news operator weight with sensitivity $\theta_l \omega_l^2/(1 - \rho\alpha)$, and an error in the estimated loading with sensitivity $\omega_l/(1 - \rho\alpha)$. Both grow with ω_l , and the loading channel does not depend on the size of θ_l , so the concentration of V at high persistence does not require a large loading. Forecast combination removes the idiosyncratic part of this noise at rate $1/K_l$ [part (b)]. Therefore, averaging over many predictors (i.e., large K_l) is most valuable at high ϕ_l , favoring dense bins for high-persistence

groups. At low persistence, the amplification is small, so bin width and predictor count matter less. Minimizing the MSE therefore favors partitions narrow in ϕ_i and dense in predictors at high persistence. $\square$

B Variable descriptions

This appendix describes the return predictor variables used in the vector autoregressions. We require predictors to have data back to January 1927. We estimate VARs using monthly data, so we also require that we can calculate each variable on a monthly basis.

We consider a total of 25 predictors. The first 14 predictors are from Goyal and Welch (2008).³

1. **Dividend-price ratio:** The log of the ratio of dividends paid on the S&P 500 over the past 12 months and the index level of the S&P 500.
2. **Dividend yield:** The log of the ratio of dividends paid on the S&P 500 over the past 12 months and the 12-month lag of the index level of the S&P 500.
3. **Earnings-price ratio:** The log of the ratio of earnings on the S&P 500 over the past 12 months and the index level of the S&P 500.
4. **Book-to-market ratio:** The ratio of book value and market value of the Dow Jones Industrial Average.
5. **Dividend-earnings ratio:** The log of the ratio of dividends on the S&P 500 over the past 12 months and earnings on the S&P 500 over the past 12 months.
6. **Stock variance:** The sum of squared daily returns on the S&P 500 over the past month.
7. **Net stock issues:** The ratio of the sum of net issues by NYSE stocks over the past 12 months and the total market capitalization of NYSE stocks.
8. **Treasury bill yield:** The yield on three-month Treasury bills.
9. **Long-term yield:** The yield on long-term Treasury bonds.
10. **Long-term return:** The return on long-term Treasury bonds over the past month.
11. **Term spread:** The difference between the yield on long-term Treasury bonds and the yield on three-month Treasury bills.
12. **Default spread:** The difference between the yield on BAA-rated corporate bonds and the yield on AAA-rated corporate bonds.
13. **Default return:** The difference between the return on long-term corporate bonds and the return on long-term Treasury bonds over the past month.

³We thank Amit Goyal for making these data available (<https://sites.google.com/view/agoyal145>).

14. **Inflation:** The one-month percentage change in the Consumer Price Index (All Urban Consumers) lagged by one month to account for reporting delays.

The next three predictors are variables used in VARs by John Campbell that are not analogous to any of the variables listed above.

15. **Relative bill rate:** Following Campbell (1991), we compute the difference between the yield on one-month Treasury bills and the average yield on one-month Treasury bills over the prior 12 months. Data for one-month Treasury bill yields are from Kenneth French’s website.⁴
16. **Smoothed earnings-price ratio:** Following Campbell and Vuolteenaho (2004), the smoothed earnings-price ratio is the log of the ratio of the average annual earnings on the S&P 500 over the past ten years and the index level of the S&P 500. We thank Amit Goyal for providing these data.
17. **Small-stock value spread:** Following Campbell and Vuolteenaho (2004), the small-stock value spread is the log of the ratio of the book-to-market ratio of the small value portfolio and the book-to-market ratio of the small growth portfolio, where the small value and small growth portfolios are from a two-by-three sort on size and book-to-market. The June book-to-market ratio for each portfolio is the value-weighted average of the book value from fiscal year $t - 1$ divided by the market value from December of year $t - 1$. The log difference in the book-to-market ratios in subsequent months is adjusted by adding the log of the cumulative value-weighted return on the small growth portfolio since the past June and subtracting the log of the cumulative value-weighted return on the small value portfolio since the past June. All data are from Kenneth French’s website.

The remaining eight predictors are from recent studies published in the *Journal of Finance*, *Journal of Financial Economics*, or *Review of Financial Studies*.

18. **Net payout yield:** Following Boudoukh, Michaely, Richardson, and Roberts (2007), we compute the net payout yield as the log of 0.1 plus the ratio of the aggregate net payout by nonfinancial firms over the past 12 months and the aggregate market capitalization of nonfinancial firms. We use data from CRSP for NYSE-, AMEX-, and Nasdaq-listed firms with share codes of 10 and 11, and we exclude firms with Standard Industrial Classification (SIC) codes between 6000 and 6999. The aggregate dividend for each month is calculated from firm-level dividends that are imputed from the returns and ex-dividend returns on nonfinancial firms. The aggregate net issuance for each month is calculated from firm-level net issuance, which is the product of the split-adjusted price per share and the monthly change in the split-adjusted number of shares outstanding, in which the split-adjusted price per share is the month-beginning and month-ending average. Aggregate net payout is aggregate dividends minus aggregate net issuances. We calculate $PY_t = \log(0.1 + NP_t)$, where NP_t is the ratio of the

⁴We thank Kenneth French for making these data available (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

aggregate net payout by nonfinancial firms over the past 12 months and the aggregate market capitalization of nonfinancial firms.

19. **Average correlation:** Following Pollet and Wilson (2010), we calculate the value-weighted average cross-sectional correlation of daily returns for the 500 largest stocks by market capitalization at the end of each month. We use daily returns from CRSP for NYSE-, AMEX-, and Nasdaq-listed firms with share codes of 10 and 11. For each month t , we compute the standard deviation for each stock i , $\hat{\sigma}_{i,t}$, and the covariance between each pair of stocks i and j , $\hat{\sigma}_{ij,t}$, over the period from month $t - 2$ to month t using non-missing daily returns. We require that each firm exists for all three months by ensuring it has a price observation at the end of month $t - 3$. We then compute the correlation between each pair of stocks i and j as

$$\hat{\rho}_{ij,t} = \frac{\hat{\sigma}_{ij,t}}{\hat{\sigma}_{i,t}\hat{\sigma}_{j,t}}. \quad (\text{B1})$$

Finally, we calculate the average correlation measure as

$$CORR_t = \sum_{i=1}^{500} \sum_{j \neq i} w_{i,t} w_{j,t} \hat{\rho}_{ij,t}, \quad (\text{B2})$$

where $w_{i,t}$ and $w_{j,t}$ are the value weights of stocks i and j at the end of month t in a portfolio of the largest 500 stocks at the end of month t .

20. **Nearness to Dow historical high:** Following Li and Yu (2012), we calculate the ratio of the Dow Jones Industrial Average at the end of each month t and the historical high of daily closing index values of the Dow Jones Industrial Average. Data are from Bloomberg.
21. **Partial least squares aggregated book-to-market ratio:** Following Kelly and Pruitt (2013), we aggregate information from the book-to-market ratios of portfolios sorted on size and book-to-market using a partial least squares (PLS) approach. We use data on returns and book-to-market ratios of 100 portfolios sorted on size and book-to-market and data on returns of the CRSP value-weighted stock market index from Kenneth French's website. The June book-to-market ratio for each portfolio is the value-weighted average of the book value from fiscal year $t - 1$ divided by the market value from June of year t . The log book-to-market ratios in subsequent months for each portfolio are adjusted by subtracting the log of the cumulative value-weighted return on the portfolio since the past June. In the first step of the PLS approach, we estimate the time-series regression

$$\log(BM_{i,t}) = \phi_{i,0} + \phi_{i,1} \log R_{t+1} + e_{i,t}, \quad (\text{B3})$$

for each portfolio i , where $BM_{i,t}$ is the book-to-market of portfolio i in month t and R_{t+1} is the stock market return in month $t + 1$. In the second step, we estimate the

cross-sectional regression

$$\log(BM_{i,t}) = c_t + ABM_t \hat{\phi}_{i,1} + w_{i,t}, \quad (\text{B4})$$

for each month t . The ABM_t measure is the estimated latent factor in the cross-sectional regressions.⁵ We estimate the model using data from January 1927 to January 2022 to obtain a time series of the predictor that spans January 1927 to December 2021.

22. **Tail risk:** Following Kelly and Jiang (2014), we estimate tail risk using Hill’s (1975) power law estimator. We use daily returns from CRSP for NYSE-, AMEX-, and Nasdaq-listed firms with share codes of 10 and 11. The tail risk measure in month t is calculated as

$$TAIL_t = \frac{1}{K_t} \sum_{k=1}^{K_t} \log \left(\frac{R_{k,t}}{u_t} \right), \quad (\text{B5})$$

where $R_{k,t}$ is the k th daily return in the pooled cross section in month t that falls below the extreme value threshold (u_t) and K_t is the total number of returns below the threshold in month t . The threshold parameter is the fifth percentile of the pooled cross-sectional distribution of daily returns in month t . We standardize the measure to have a mean of zero and a standard deviation of one.

23. **Amihud illiquidity measure:** Following Chen, Eaton, and Paye (2018), we measure the log of the equal-weighted average of the Amihud (2002) illiquidity measure with adjustments for breaks and market volatility. We use daily returns from CRSP for NYSE-listed stocks with share codes of 10, 11, and 12. For each firm i in each month t , the Amihud illiquidity measure is calculated as

$$AMI_{i,t} = \left(\frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \frac{|R_{i,d,t}|}{DVOL_{i,d,t}} \right) CPI_t, \quad (\text{B6})$$

where $D_{i,t}$ is the number of trading days with positive volume for stock i in month t , $|R_{i,d,t}|$ is the absolute value of the return on stock i on day d of month t , $DVOL_{i,d,t}$ is the dollar trading volume in millions for stock i on day d , and CPI_t is the Consumer Price Index (CPI). CPI data are from the Federal Reserve Bank of St. Louis.⁶ We use the real-time break adjustment procedure from Chen, Eaton, and Paye (2018) to adjust the time series for the effects of changes in tick sizes in 1997 and 2001. Finally, we subtract the market volatility component as the cross-sectional average of the log of estimated firm-level volatility, $vol_t = \overline{\log(\hat{\sigma}_{i,t})}$.

24. **Zero return frequency:** Following Chen, Eaton, and Paye (2018), we measure the log of the equal-weighted average of the proportion of zero-return days with an adjustment for breaks. We use daily returns from CRSP for NYSE-listed stocks with share codes of 10, 11, and 12. For each firm i in each month t , we calculate the proportion of

⁵We thank Seth Pruitt for making code available (<https://sethpruitt.net/research/>).

⁶See <https://fred.stlouisfed.org/series/CPIAUCNS>.

trading days with a return of zero. We use the real-time break adjustment procedure from Chen, Eaton, and Paye (2018) to adjust the time series for the effects of changes in tick sizes in 1997 and 2001.

25. **Average skewness:** Following Jondeau, Zhang, and Zhu (2019), we measure the value-weighted average of the skewness of excess daily returns within each month. We use daily returns from CRSP for NYSE-, AMEX-, and Nasdaq-listed firms with share codes of 10 and 11. We exclude stocks with an Amihud (2002) illiquidity measure in the top 0.1% or with prices below \$1. For a given month, we use all stocks that have at least ten valid return observations.

C Additional methodology

C.1 Persistence cutoff selection procedure

We group K predictors into k groups according to estimated predictor persistence. To perform the grouping, we need to define persistence cutoff points for $\hat{\phi}_i$ from equation (5). We have two goals in determining the cutoff points. First, the resulting forecast combinations should be successful in forecasting excess market returns across both short and long horizons. This goal is important in our context because equation (1) shows that discount-rate news depends on excess return forecasts in all future periods, so properly modeling expected returns across horizons is necessary for estimating discount-rate news. Second, each group should contain at least five predictors. Averaging within a group reduces estimation error [Proposition 1(b)], and in-sample overfitting is likelier when a group contains only one or two predictors.

The basis of our procedure to select persistence cutoffs is a set of excess return forecasts at horizons from one to 60 months. The return forecasts are designed to be analogous to those from the VAR, so the predictors are the lagged excess market return and the k forecast combinations. Given a candidate set of cutoff points for the k groups, we form the forecast combinations $\tilde{r}_{1,t+1}^e, \dots, \tilde{r}_{k,t+1}^e$ and run 60 forecasting regressions with horizon h ranging from one to $H = 60$ months,

$$r_{t+h}^e = a + b_0 r_t^e + b_1 \tilde{r}_{1,t+1}^e + \dots + b_k \tilde{r}_{k,t+1}^e + \eta_{t+h}. \quad (\text{C1})$$

Assuming $\eta_{t+h} \sim N(0, \sigma_{\eta,h}^2)$, we compute the log-likelihood, $\log \mathcal{L}_h$, from this regression for each horizon h . Finally, we sum the Akaike information criterion across the H horizons,

$$AIC = 2(k+3)H - 2 \sum_{h=1}^H \log \mathcal{L}_h, \quad (\text{C2})$$

as a measure of overall fit.

We select persistence cutoffs that achieve strong return forecasting performance across horizons, subject to the constraint on minimum group size. For the $k = 2$ case, we need to choose a single persistence cutoff. The lowest value of AIC occurs for a cutoff that groups 16 predictors into the low-persistence group and nine predictors into the high-persistence

group. Based on the estimated persistence levels, this grouping occurs for any cutoff on $\hat{\phi}_i$ between 0.9844 and 0.9876, or, equivalently, half-lives of about 44 to 55 months. Accordingly, we set a persistence cutoff of 48 months in this $k = 2$ case. The next four highest-ranked models correspond to persistence cutoffs with half-lives of 42, 36, 24, and 60 months, so the procedure is naturally generating a divide between predictors that are quite persistent and those that are less so.

For our base case of $k = 3$, we choose two predictor cutoffs. The lowest value of AIC is achieved by a model with cutoffs of 30- and 42-month half-lives. This model, however, produces group sizes of 13, two, and ten predictors, so its medium group falls below the floor. Among models that have at least five predictors in each group, there are two models that achieve a virtual tie for the lowest AIC . The model with the lower AIC has cutoff points corresponding to three- and 42-month half-lives with groupings of seven, eight, and ten predictors, and the next model has three- and 48-month half-lives with groupings of seven, nine, and nine predictors. These two models generate virtually identical results for all analyses in the paper. We choose the model with the 48-month half-life (in addition to the three-month half-life) in the $k = 3$ case for symmetry with the 48-month half-life in the $k = 2$ case.

C.2 Bootstrap procedures

This appendix describes the three simulation procedures behind the inference in the paper: (i) a bootstrap that imposes a null of no return predictability and produces the p -values for the return-predictability statistics, (ii) a bootstrap that imposes a null of no dividend growth predictability and produces the p -values for the dividend-growth-predictability statistics, and (iii) a bootstrap at the estimated system that produces the confidence intervals for the news moments. The three procedures share the same simulation approach and differ in the process generating the simulated data.

C.2.1 Null of no return predictability

The predictor set for each specification is as in the data. The base specification includes all $K = 25$ predictors, whereas the risk-free rate specification excludes the Treasury bill yield and the dividend-price ratio specification excludes the dividend-price ratio, as the log risk-free rate and the log dividend-price ratio enter the respective VARs directly as state variables. For each draw $m = 1, \dots, M$ with $M = 9,999$:

1. We resample dates with replacement and collect the innovation vector of each resampled date, which contains the demeaned excess return, the estimated innovations of the predictor autoregressions, and the estimated innovation of the risk-free rate. Resampling entire dates preserves the contemporaneous correlations among the innovations and removes any time-series dependence.
2. We simulate excess returns as the sample mean plus the resampled return innovations, so returns are unpredictable by construction.

3. We simulate each predictor from its autoregression in equation (5) using the resampled predictor innovations and the bias-corrected persistence coefficient $\phi_i^{bc} = \min\{\hat{\phi}_i + (1 + 3\hat{\phi}_i)/(T - 1), 0.999\}$ [Kendall (1954)], as simulating at the uncorrected estimates would understate the persistence of the predictors under the null. We simulate the log risk-free rate in the same manner from its own autoregression. The resampled innovations are the unadjusted estimated residuals.
4. We simulate $T + 2,000$ months and discard the first 2,000 as a burn-in.
5. We apply the complete estimation procedure to the simulated sample, in which we (i) estimate the predictor autoregressions, (ii) estimate the univariate predictive regressions, (iii) sort the predictors into persistence groups by applying the half-life cutoffs to the estimated persistence coefficients as in the data, (iv) construct the forecast combinations, (v) estimate the VAR, (vi) compute the news series and their moving averages, and (vii) estimate the forecasting regressions. The annual dependent variables are cumulated from the simulated monthly observations to reproduce the overlapping structure of the data regressions, with the simulated risk-free rate used to construct the targets defined relative to it.

Draws for which the procedure fails to produce finite statistics are discarded and replaced, so that all p -values are computed from M valid draws. The real-time statistics are tested against the same null with the real-time procedure repeated inside each draw. Beginning from an initial window of 120 months, we re-estimate the predictor autoregressions, the persistence sorting, the forecast combinations, and the VAR on each expanding vintage of the simulated data, construct the news series from each vintage, and estimate the forecasting regressions on the resulting real-time regressors. As in the data, the real-time sorting applies the half-life cutoffs to bias-corrected persistence coefficients.

This null generates the p -values for every return-predictability statistic in the paper, namely the univariate predictive regressions in Table I, the VAR return equations in Tables II, VI, D.I, and D.III, and the return forecasting regressions in Tables III, IV, D.II, and D.IV.

In the in-sample regressions of annual returns on the moving average of discount-rate news, the regressor is built from the fitted values of the VAR return equation, so the null distribution of the slope is centered near a model-implied value rather than zero. In the real-time regressions the news series are constructed in real time and the null distribution of the slope is centered near zero.

C.2.2 Null of no dividend growth predictability

For the dividend-growth-predictability statistics, we impose a null of no dividend growth predictability. The draw construction differs from Appendix C.2.1 only in the process generating the simulated returns and dividend growth. The resampled innovation vector replaces the return innovation with the estimated innovation of monthly dividend growth. We simulate dividend growth as the sample mean plus the resampled innovations, so dividend growth is unpredictable by construction, and we derive the simulated returns from the Campbell–Shiller identity,

$$r_t^* = (d_{t-1} - p_{t-1}) - \rho(d_t - p_t) + \Delta d_t^*,$$

up to a constant, so that the simulated system is internally consistent [Cochrane (2008)]. Under this null, the population coefficient in a regression of dividend growth on cash-flow news is zero, and returns inherit the predictability implied by the identity. The predictor simulation, burn-in, per-draw re-estimation, and validity screening are identical to Appendix C.2.1. This null generates the p -values for the dividend growth forecasting regressions in Tables III, IV, D.II, and D.IV.

C.2.3 Confidence intervals for the news moments

The confidence intervals for the standard deviations of discount-rate news and cash-flow news and for their correlation in Tables II and VI are from a bootstrap at the estimated system rather than under a null. For each specification, we estimate the VAR and bias-correct the estimated system by applying the correction of Appendix C.2.1 to the own-persistence coefficients, applying the stationarity adjustment of Kilian (1998) if the correction produces an explosive system. We simulate from the bias-corrected system by resampling the fitted innovations jointly by date with a 2,000-month burn-in, re-estimate the VAR on each simulated sample, discard any draw in which the re-estimated system is explosive, and recompute the news moments. The intervals are equal-tailed percentile intervals from $M = 4,999$ draws, and the reported point estimates are the estimates from the data. Neither the adjustment nor the discard is needed in any specification, as the bias-corrected systems are stationary and every draw produces stationary estimates.

C.2.4 Test statistics and p -values

The test statistic for each coefficient is its t -statistic, and the test statistic for each regression fit is the R^2 , computed identically for the data and for each draw. The t -statistics use heteroskedasticity-robust White (1980) standard errors in the monthly regressions and Newey and West (1987) standard errors with 12 lags in the annual overlapping regressions. One-tailed p -values are $p = (r + 1)/(M + 1)$, where r is the number of draws in which the statistic is at least as large as its value in the data. Two-tailed p -values double the smaller of the two one-tailed p -values and are capped at one. A displayed value of [0.000] indicates $p < 0.0005$. One-tailed tests apply to the R^2 s and to the coefficients whose signs the model predicts, namely (i) the return-equation coefficients of the VAR, including the forecast combinations, (ii) the discount-rate news coefficients in the return forecasts, and (iii) the cash-flow news coefficients in the dividend growth forecasts. Two-tailed tests apply otherwise, including the univariate predictive regressions and the cross-news coefficients. The brackets on the state-variable dynamics equations of the VAR are analytic rather than bootstrap-based, with two-tailed p -values based on White (1980) standard errors for the coefficients and Wald tests that all slopes in the equation are zero for the R^2 s.

D Alternative VAR specifications

This appendix gives the VAR specifications for the two alternatives in Section 4.5 and reports their estimates and validation regressions.

D.1 Risk-free rate as a state variable

The first alternative separates risk-free rate news from cash-flow news as in equation (18). The VAR is

$$\begin{bmatrix} r_{t+1}^e \\ r_{f,t+2} \\ \tilde{r}_{1,t+2}^e \\ \tilde{r}_{2,t+2}^e \\ \tilde{r}_{3,t+2}^e \end{bmatrix} = a + A \begin{bmatrix} r_t^e \\ r_{f,t+1} \\ \tilde{r}_{1,t+1}^e \\ \tilde{r}_{2,t+1}^e \\ \tilde{r}_{3,t+1}^e \end{bmatrix} + \varepsilon_{t+1}, \quad (\text{D1})$$

where the log risk-free rate enters as a state variable and the Treasury bill yield is excluded from the high-persistence combination $\tilde{r}_{3,t+1}^e$. The news components are

$$N_{RP,t+1} = e'_1 \rho A (I - \rho A)^{-1} \varepsilon_{t+1}, \quad (\text{D2})$$

$$N_{RF,t+1} = e'_2 \rho (I - \rho A)^{-1} \varepsilon_{t+1}, \quad (\text{D3})$$

$$N_{CF,t+1} = (e'_1 + e'_1 \rho A (I - \rho A)^{-1} + e'_2 \rho (I - \rho A)^{-1}) \varepsilon_{t+1}. \quad (\text{D4})$$

Table D.I reports the estimates. The risk-free rate is highly persistent, with a coefficient of 0.98 on its own lag and an R^2 of 95.40%, and the remaining estimates are close to those of the base case in Table II.

Using this VAR specification, discount-rate news should forecast returns (rather than excess returns as in Tables III and IV) and cash-flow news should forecast dividend growth (rather than dividend growth in excess of the risk-free rate). We estimate

$$r_{t \rightarrow t+12} = a_r + b_r \text{MA}_t(N_{DR}) + c_r \text{MA}_t(N_{CF}) + e_{r,t} \quad (\text{D5})$$

and

$$\Delta d_{t \rightarrow t+12} = a_d + b_d \text{MA}_t(N_{DR}) + c_d \text{MA}_t(N_{CF}) + e_{d,t} \quad (\text{D6})$$

in place of equations (16) and (17). Table D.II reports the results. In sample, discount-rate news forecasts returns with an R^2 of 18.97%, significant at the 5% level, and cash-flow news adds nothing to that regression (coefficient of 0.57, p -value of 0.741). Cash-flow news forecasts dividend growth with an R^2 of 19.47%, significant at the 1% level, and discount-rate news adds nothing (coefficient of -0.03 , p -value of 0.846). In real time, cash-flow news forecasts dividend growth with a coefficient of 8.48 and an R^2 of 16.68%, both significant at the 1% level, and discount-rate news again adds nothing. The real-time return forecast is somewhat weaker than in the base case, with a coefficient of 3.79 that just misses statistical significance at the 10% level (p -value of 0.106) and an R^2 of 2.65%.

D.2 Dividend-price ratio as a state variable

The second alternative adds the log dividend-price ratio to the state vector,

$$\begin{bmatrix} r_{t+1}^e \\ d_{t+1} - p_{t+1} \\ \tilde{r}_{1,t+2}^e \\ \tilde{r}_{2,t+2}^e \\ \tilde{r}_{3,t+2}^e \end{bmatrix} = a + A \begin{bmatrix} r_t^e \\ d_t - p_t \\ \tilde{r}_{1,t+1}^e \\ \tilde{r}_{2,t+1}^e \\ \tilde{r}_{3,t+1}^e \end{bmatrix} + \varepsilon_{t+1}, \quad (\text{D7})$$

and excludes it from the high-persistence combination. With $d_t - p_t$ as a standalone state variable, the residual and direct methods for cash-flow news coincide [Chen (2010); Engsted, Pedersen, and Tanggaard (2012)], and equations (3) and (4) apply unchanged. Table D.III reports the estimates. The return equation R^2 is 5.29% against 4.93% in the base case, and the coefficient on the dividend-price ratio in the return equation is negative and insignificant once the combinations are included. Table D.IV reports the validation regressions. In sample, discount-rate news forecasts returns with an R^2 of 21.41% and cash-flow news forecasts dividend growth with an R^2 of 16.41%, both significant at the 1% level. Cash-flow news adds nothing to the return regression (coefficient of -1.69 , p -value of 0.761), and discount-rate news adds nothing to the dividend growth regression (coefficient of 1.51, p -value of 0.523). In real time, discount-rate news forecasts returns with a coefficient of 3.78 and an R^2 of 3.05%, both significant at the 10% level, and cash-flow news forecasts dividend growth with a coefficient of 4.98, significant at the 5% level, and an R^2 of 13.37%, significant at the 1% level. Adding cash-flow news to the real-time return regression raises the R^2 to 6.21%, but the coefficient of -7.15 is not significant (p -value of 0.201), and adding discount-rate news to the dividend growth regression changes little.

Table D.I
Vector autoregression for the risk-free rate model.

The table reports estimates from the vector autoregression (VAR) that directly includes the log risk-free rate as a state variable. The market return predictors are sorted into three groups by their AR(1) coefficients, and the VAR state variables are the log excess market return, the log risk-free rate, and the standardized forecast combinations from the low-persistence ($\tilde{r}_{1,t+1}^e$, half-lives less than three months), medium-persistence ($\tilde{r}_{2,t+1}^e$, three months to four years), and high-persistence ($\tilde{r}_{3,t+1}^e$, more than four years) predictor groups. The table reports the VAR coefficient matrix A , the R^2 (in percent) from each regression, and the covariance matrix Σ , with p -values (in brackets) below the coefficients and R^2 's. In the return equation, the one-tailed p -values for the coefficients and the R^2 are from a bootstrap under the null of no return predictability. All other coefficient p -values are two-tailed, based on White (1980) heteroskedasticity-robust standard errors, and the remaining R^2 p -values are from Wald tests that all slopes in the equation are zero. The sample period is January 1927 to December 2021. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

	A					R ²	Σ				
	r _t ^e	r _{f,t+1}	$\tilde{r}_{1,t+1}^e$	$\tilde{r}_{2,t+1}^e$	$\tilde{r}_{3,t+1}^e$		r _{t+1} ^e	r _{f,t+2}	$\tilde{r}_{1,t+2}^e$	$\tilde{r}_{2,t+2}^e$	$\tilde{r}_{3,t+2}^e$
r _{t+1} ^e	0.04 [0.172]	-0.57 [0.171]	0.37 [0.830]	0.98* [0.051]	0.05 [0.904]	4.99*** [0.000]	27.25	-0.01	1.85	-0.24	-0.58
r _{f,t+2}	0.00*** [0.002]	0.98*** [0.000]	-0.00 [0.101]	-0.00 [0.520]	0.00 [0.939]	95.40*** [0.000]	-0.01	0.00	-0.00	-0.00	0.00
$\tilde{r}_{1,t+2}^e$	-0.03*** [0.000]	-0.11 [0.406]	0.45*** [0.000]	0.07* [0.066]	-0.02 [0.469]	17.72*** [0.000]	1.85	-0.00	0.82	-0.02	-0.04
$\tilde{r}_{2,t+2}^e$	-0.01*** [0.000]	-0.03 [0.332]	-0.01 [0.390]	0.97*** [0.000]	0.02 [0.106]	93.77*** [0.000]	-0.24	-0.00	-0.02	0.06	0.00
$\tilde{r}_{3,t+2}^e$	-0.00 [0.769]	-0.01 [0.509]	-0.01 [0.164]	-0.02*** [0.003]	1.00*** [0.000]	98.52*** [0.000]	-0.58	0.00	-0.04	0.00	0.01

Table D.II
Return and dividend growth forecasts: risk-free rate VAR.

The table reports estimates from regressions that forecast annual log market return in columns (1) and (2) and annual log dividend growth in columns (3) and (4). The predictors are smoothed versions of the discount-rate news ($MA_t(N_{DR})$) and cash-flow news ($MA_t(N_{CF})$) estimates from the VAR that directly includes the log risk-free rate, constructed as trailing exponentially weighted moving averages following Campbell, Giglio, and Polk (2013). Panel A uses in-sample news estimates. Panel B uses real-time news estimates: each month, on an expanding window of data available through that month, we sort predictors into three persistence groups using bias-corrected AR(1) coefficients, estimate the VAR, and form the smoothed news estimates from that vintage only. The sample period is January 1927 to December 2021. The p -values (in brackets) below the coefficients and R^2 s are from a bootstrap under the null of no return predictability in columns (1) and (2) and the null of no dividend growth predictability in columns (3) and (4). The p -values are one-tailed for the R^2 s, the $MA_t(N_{DR})$ coefficients in the return forecasts, and the $MA_t(N_{CF})$ coefficients in the dividend growth forecasts, and two-tailed otherwise. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

	Return		Div. growth	
	(1)	(2)	(3)	(4)
Panel A: In-sample				
$MA_t(N_{DR})$	10.71	10.61		−0.03 [0.846]
$MA_t(N_{CF})$		0.57 [0.741]	16.13*** [0.006]	16.17*** [0.005]
R^2 (%)	18.97** [0.018]	18.97** [0.034]	19.47*** [0.000]	19.47*** [0.000]
Panel B: Real-time				
$MA_t(N_{DR})$	3.79 [0.106]	5.50* [0.060]		−0.99 [0.522]
$MA_t(N_{CF})$		−5.43 [0.240]	8.48*** [0.002]	9.42*** [0.001]
R^2 (%)	2.65 [0.163]	3.90 [0.202]	16.68*** [0.000]	17.16*** [0.000]

Table D.III
Vector autoregression for the dividend-price ratio model.

The table reports estimates from the vector autoregression (VAR) that directly includes the log dividend-price ratio as a state variable. The market return predictors are sorted into three groups by their AR(1) coefficients, and the VAR state variables are the log excess market return, the log dividend-price ratio, and the standardized forecast combinations from the low-persistence ($\tilde{r}_{1,t+1}^e$, half-lives less than three months), medium-persistence ($\tilde{r}_{2,t+1}^e$, three months to four years), and high-persistence ($\tilde{r}_{3,t+1}^e$, more than four years) predictor groups. The table reports the VAR coefficient matrix A , the R^2 (in percent) from each regression, and the covariance matrix Σ , with p -values (in brackets) below the coefficients and R^2 s. In the return equation, the one-tailed p -values for the coefficients and the R^2 are from a bootstrap under the null of no return predictability. All other coefficient p -values are two-tailed, based on White (1980) heteroskedasticity-robust standard errors, and the remaining R^2 p -values are from Wald tests that all slopes in the equation are zero. The sample period is January 1927 to December 2021. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

	A					Σ					
	r_t^e	$d_t - p_t$	$\tilde{r}_{1,t+1}^e$	$\tilde{r}_{2,t+1}^e$	$\tilde{r}_{3,t+1}^e$	R^2	r_{t+1}^e	$d_{t+1} - p_{t+1}$	$\tilde{r}_{1,t+2}^e$	$\tilde{r}_{2,t+2}^e$	$\tilde{r}_{3,t+2}^e$
r_{t+1}^e	0.03 [0.229]	-1.43 [0.993]	0.44 [0.763]	0.93* [0.074]	0.72 [0.393]	5.29*** [0.000]	27.17	-0.27	1.88	-0.24	-0.54
$d_{t+1} - p_{t+1}$	-0.00 [0.818]	1.00*** [0.000]	-0.01** [0.040]	-0.01*** [0.002]	-0.00 [0.820]	98.69*** [0.000]	-0.27	0.00	-0.02	0.00	0.01
$\tilde{r}_{1,t+2}^e$	-0.03*** [0.001]	0.37*** [0.008]	0.43*** [0.000]	0.09** [0.018]	-0.19*** [0.008]	18.30*** [0.000]	1.88	-0.02	0.82	-0.01	-0.03
$\tilde{r}_{2,t+2}^e$	-0.01*** [0.000]	-0.04 [0.203]	-0.01 [0.525]	0.97*** [0.000]	0.04** [0.036]	93.77*** [0.000]	-0.24	0.00	-0.01	0.06	0.01
$\tilde{r}_{3,t+2}^e$	-0.00 [0.514]	0.01 [0.445]	-0.00 [0.711]	-0.02*** [0.009]	0.99*** [0.000]	98.55*** [0.000]	-0.54	0.01	-0.03	0.01	0.01

Table D.IV
Return and dividend growth forecasts: dividend-price ratio VAR.

The table reports estimates from regressions that forecast annual log excess market return in columns (1) and (2) and annual log dividend growth in excess of the log risk-free rate in columns (3) and (4). The predictors are smoothed versions of the discount-rate news ($MA_t(N_{DR})$) and cash-flow news ($MA_t(N_{CF})$) estimates from the VAR that directly includes the log dividend-price ratio, constructed as trailing exponentially weighted moving averages following Campbell, Giglio, and Polk (2013). Panel A uses in-sample news estimates. Panel B uses real-time news estimates: each month, on an expanding window of data available through that month, we sort predictors into three persistence groups using bias-corrected AR(1) coefficients, estimate the VAR, and form the smoothed news estimates from that vintage only. The sample period is January 1927 to December 2021. The p -values (in brackets) below the coefficients and R^2 s are from a bootstrap under the null of no return predictability in columns (1) and (2) and the null of no dividend growth predictability in columns (3) and (4). The p -values are one-tailed for the R^2 s, the $MA_t(N_{DR})$ coefficients in the return forecasts, and the $MA_t(N_{CF})$ coefficients in the dividend growth forecasts, and two-tailed otherwise. An * (**) [***] indicates statistical significance at the 10% (5%) [1%] level based on the corresponding p -value.

	Return		Div. growth	
	(1)	(2)	(3)	(4)
Panel A: In-sample				
$MA_t(N_{DR})$	9.57	10.17		1.51 [0.523]
$MA_t(N_{CF})$		-1.69 [0.761]	8.52*** [0.001]	6.78** [0.011]
R^2 (%)	21.41*** [0.003]	21.53*** [0.005]	16.41*** [0.000]	17.42*** [0.004]
Panel B: Real-time				
$MA_t(N_{DR})$	3.78* [0.082]	8.24** [0.015]		-1.05 [0.629]
$MA_t(N_{CF})$		-7.15 [0.201]	4.98** [0.020]	5.95** [0.011]
R^2 (%)	3.05* [0.082]	6.21** [0.049]	13.37*** [0.000]	13.74*** [0.001]