An Intertemporal Risk Factor Model

Fousseni Chabi-Yo* Andrei S. Gonçalves† Johnathan Loudis‡

This Version: April 2022§

Link to Most Recent Version

Abstract

We develop a factor model that is tightly linked to intertemporal asset pricing theory. Specifically, we show that a long-term Bayesian investor prices shocks to the market dividend yield and realized variance as they reflect news to long-term expected returns and volatility. Accordingly, we construct intertemporal risk factors as long-short portfolios based on stock exposures to dividend yield and realized variance, and estimate their risk prices, which are consistent with the ICAPM under moderate risk aversion. Our intertemporal factor model performs well relative to previous models in terms of its tangency Sharpe ratio and its pricing of key test assets.

JEL Classification: G10; G11; G12.

Keywords: ICAPM; Intertemporal Risk; Long-term Investors; Factor Models.

*Isenberg School of Management, University of Massachusetts, Amherst, MA 01003. E-mail: fchabiyo@isenberg.umass.edu.

†Kenan-Flagler Business School, University of North Carolina, Chapel Hill, NC 27599. E-mail: Andrei_Goncalves@kenan-flagler.unc.edu.

‡Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556. E-mail: jloudis@nd.edu.

§We thank Andrew Chen for providing us with data on anomaly decile portfolios as well as on stock-level trading costs. We are also grateful for the very helpful comments from Ravi Bansal, John Campbell, Jennifer Conrad, Eric Ghysels, Stefano Giglio, Kewei Hou, Bryan Kelly, Tim Loughran, Giorgio Ottonello (discussant), Sydney Ludvigson, Shrihari Santosh, Ivan Shaliastovich, René Stulz, Lu Zhang, and Irina Zviadadze (discussant) as well as from seminar participants at New York University, The Ohio State University, University of North Carolina at Chapel Hill, University of Notre Dame, the Triangle Macro-Finance Workshop, the 2021 MFA Meeting, and the 2021 LubraFin Conference. All remaining errors are our own. First draft: September 2020.

Introduction

Factor models are ubiquitous in the asset pricing literature. However, the implementation of these models is based on factors that rely on signals that are not directly related to risk. This disconnect is present regardless of whether the factors are motivated empirically (Fama and French (1993) and Carhart (1997)), from a valuation identity (Fama and French (2015)), from firms' optimality conditions (Hou, Xue, and Zhang (2015) and Hou et al. (2020)), or from behavioral arguments (Stambaugh and Yuan (2017) and Daniel, Hirshleifer, and Sun (2020)). Consequently, we do not know whether the tradable factors in these models compensate investors for relevant risks and, if so, through what economic mechanisms.

The issue is that, under the law of one price, a factor model that mimics the tangency portfolio and prices all assets exists regardless of the economic environment (Roll (1977)). As Cochrane (2008) puts it, "*The only content to empirical work in asset pricing is what constraints the author puts on his fishing expedition to avoid rediscovering Roll's theorem.*"

In this paper, we address this issue by developing a tradable factor model that is tightly linked to the Intertemporal CAPM (ICAPM). We then show that this intertemporal risk factor model implies risk prices that are consistent with its underlying structural ICAPM and also that it performs well relative to prominent factor models in several empirical tests.

To start, we build on Binsbergen and Koijen (2010) and Campbell et al. (2018) to show that a long-term Bayesian investor who infers market expected returns and volatility from past observations of prices and dividends dislikes negative dividend yield shocks (positive realized variance shocks) as these imply declines in long-term expected returns (increases in long-term volatility), and thus worse prospects for long-term investing. We then construct tradable intertemporal risk factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) by sorting stocks based on their exposures to changes in the market dividend yield and realized variance, and show that these factors mimick news to long-term expected returns and volatility estimated ex-post from our Bayesian framework. Finally, as implied by the ICAPM, we combine $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ with the market factor (r_m) to form our intertemporal risk factor model, which we explore empirically.

(a) $\mathbb{E}[r]$ from ICAPM Risk Exposures (b) Tangency Portfolio Sharpe Ratios (c) $\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$

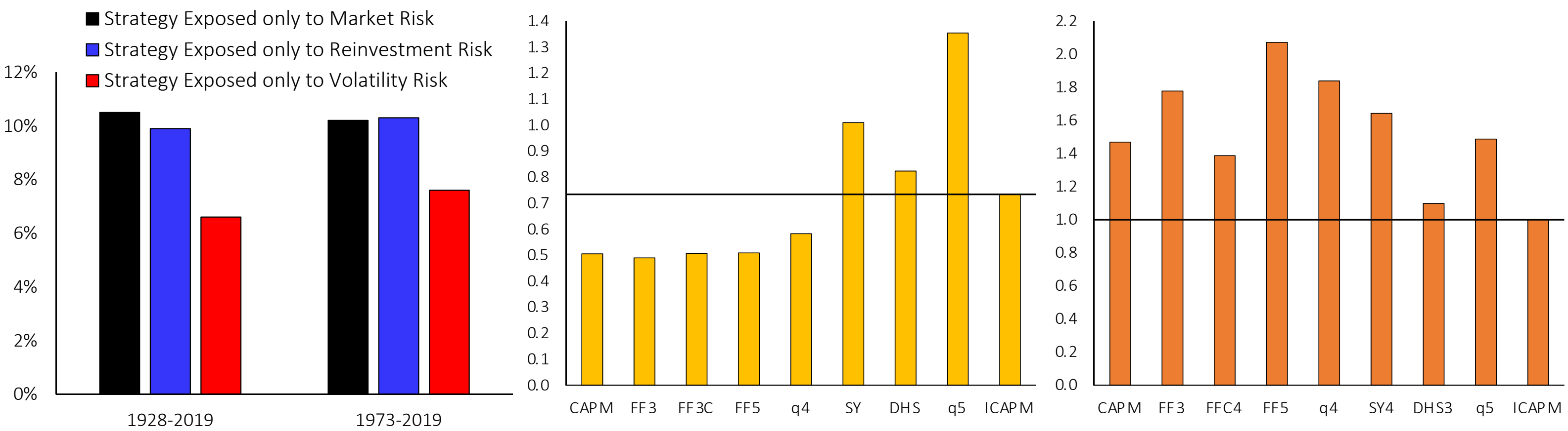

Figure 1

Main Results from our Intertemporal Risk Factor Model

Panel (a) shows average returns on strategies that expose investors to each ICAPM risk factor in isolation (i.e., normalized to be orthogonal to the other ICAPM factors and to match the market volatility). Panel (b) displays tangency Sharpe ratios (out-of-sample and net of trading costs) constructed using the ICAPM factors or using factors from the factor models described at the beginning of Section 3. Panel (c) displays averages of relative pricing errors, $\Sigma\alpha^2/\Sigma\alpha^2_{ICAPM}$, across the different groups of testing assets recommended in Lewellen, Nagel, and Shanken (2010). Further details are provided in Section 3.

In our first empirical exercise, we study our tradable factors and estimate their risk prices. Our risk factors are strongly correlated, which is in line with the ICAPM logic as current market prices move together with investment opportunities. Moreover, we find that strategies that expose investors to marker risk (r_m), reinvestment risk ($r_{\mathbb{E}}$), or volatility risk ($r_{\mathbb{V}}$) in isolation deliver strong and significant risk premia (see Figure 1(a)). While also consistent with the ICAPM, this result is in stark contrast to Herskovic, Moreira, and Muir (2019) as they find that tradable factors that hedge business cycle risk have no risk premia.

To further explore our intertemporal factor model risk prices, we estimate a projection of the Stochastic Discount Factor (SDF) onto r_m, $r_{\mathbb{E}}$, and $r_{\mathbb{V}}$, which is equivalent to a Generalized Method of Moments (GMM) estimation that treats the factors as testing assets. This estimation implies a risk aversion coefficient of 4.8 in our Long Sample (1928-2019) and

5.6 in our Modern Sample (1973-2019), very close to the 5.1 (7.2) estimate from the early period (modern period) structural ICAPM estimation in Campbell et al. (2018). Moreover, consistent with the ICAPM, r_m and $r_{\mathbb{E}}$ have positive risk prices while $r_{\mathbb{V}}$ has a negative risk price. This result holds even after controlling for each of the factor models cited in our first paragraph. The risk price magnitudes are also in line with the structural ICAPM as they remain similar when we reestimate the model imposing the structural restrictions that pin down all risk prices as functions of the risk aversion parameter, which we reestimate to be 6.3 in our Long Sample and 6.9 in our Modern Sample.

In our second empirical exercise, we compare the pricing ability of our intertemporal factor model with that of the prominent factor models cited in our first paragraph. We start by following the argument in Barillas and Shanken (2017) that contrasting the tangency portfolio Sharpe ratios of different factor models is sufficient for comparing them. We find that the ICAPM's in-sample tangency Sharpe ratio is higher than those of the CAPM and the FF3 model (Fama and French (1993)), but lower than those of all other models we explore. However, as Figure 1(b) shows, after partially adjusting for overfitting (through an out-of-sample analysis as in Kan, Wang, and Zheng (2019)) and for trading costs (as in Detzel, Novy-Marx, and Velikov (2020)), the ICAPM's tangency Sharpe ratio is also higher than those of the FFC4, FF5, and q4 models (respectively by Carhart (1997), Fama and French (2015) and Hou, Xue, and Zhang (2015)), remaining lower than only those of the SY4, DHS3, and q5 models (respectively by Stambaugh and Yuan (2017), Daniel, Hirshleifer, and Sun (2020), and Hou et al. (2020)). As such, the ICAPM has a strong tangency Sharpe ratio despite its factors being constrained to reflect risks related to long-term investing, with the use of out-of-sample Sharpe ratios being the key driver of this result.

While focusing on tangency Sharpe ratios is sufficient for comparing factor models in a world without publication biases, directly studying test assets can be important when the publication prospects of proposed factor models correlate with the Sharpe ratios on the proposed factors, which is likely to be the case in the asset pricing literature. As such, we also study the performance of the different factor models in pricing the test assets recommended by Lewellen, Nagel, and Shanken (2010): single stocks, industry portfolios, correlation-clustered portfolios (Ahn, Conrad, and Dittmar (2009)), and bond portfolios. Tests based on these assets are less subject to publication biases (than tests based on Sharpe ratios or anomaly portfolios) as previous factor models were not tested against such assets when originally proposed. As Figure 1(c) shows, we find that the ICAPM pricing errors are lower than those for all alternative factor models we consider.[1]

In summary, we show that a long-term Bayesian investor perceives shocks to the market dividend yield and realized variance as additional risk factors beyond the market portfolio, and use this insight to construct an intertemporal factor model in which the tradable factors capture market risk (r_m), reinvestment risk ($r_{\mathbb{E}}$), and volatility risk ($r_{\mathbb{V}}$). We then explore the model empirically, finding that (i) its risk price signs are consistent with theory and their magnitudes imply a reasonable risk aversion and (ii) it performs well relative to prominent factor models in terms of its tangency Sharpe ratio and its pricing of relevant testing assets.

Our main contribution is to connect the structural ICAPM literature to the reduced-form factor model literature.[2] While structural ICAPM tests provide a direct map from the risk considerations affecting long-term investing to risk prices, they require assumptions outside of the economic environment (e.g., a pre-specified state vector to estimate news) and result in non-tradable factors, with both of these limitations leading to potential specification issues (see Chen and Zhao (2009) and Lewellen, Nagel, and Shanken (2010)). In contrast, reduced-form factor models are fairly robust to mispecification, but do not have a risk-based interpretation, and thus are silent on whether their factors compensate investors for

[1] We also explore anomaly portfolios as test assets and find that the ICAPM pricing errors are only lower than those from the CAPM, FF3, and FF5 models while larger than those from the other five models we consider. Flipping the coin, we show that the ICAPM is able to price decile portfolios sorted on exposures to our $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ factors better than all alternative factor models we consider, which highlights the importance of using test assets that are not based on signals that are connected to the factors when comparing factor models (see Lo and MacKinlay (1990) and Ferson, Sarkissian, and Simin (1999)). In this vein, we also show that a strategy that adds anomaly portfolios to the ICAPM factors in real time has a large chance of ultimately deteriorating the investor's Sharpe ratio.

[2] Some examples of papers in the structural ICAPM literature are Campbell (1993, 1996), Brennan, Wang, and Xia (2004), Campbell and Vuolteenaho (2004), Maio (2013), Campbell et al. (2018), Cederburg (2019), and Gonçalves (2021a,b), with Cederburg (2019) exploring a Bayesian estimation framework.

relevant risks (Kozak, Nagel, and Santosh (2018)). We provide a factor model linked to the risk considerations affecting long-term investing through our Bayesian learning framework while building on important implementation insights from the reduced-form factor model literature. The result is an intertemporal factor model that can be robustly implemented in real time, performs well empirically, and reflects the SDF of a long-term investor with moderate risk aversion, which translate into the positive pricing of market risk (r_m) and reinvestment risk ($r_{\mathbb{E}}$), and the negative pricing of volatility risk ($r_{\mathbb{V}}$).

Some papers explore the pricing of shocks to expected returns or volatility in the context of Merton (1973)'s ICAPM (e.g., Kozak and Santosh (2020) and Ang et al. (2006)). Our contribution to this side of the literature is threefold. First, we develop a fully-specified ICAPM with a Bayesian learning framework, which maps long-term expected returns and volatility to changes in the market dividend yield and realized variance, thereby providing a way to construct tradable ICAPM factors in real time. Second, our framework links each ICAPM risk price to risk aversion in a structural way, allowing us to quantitatively validate our intertemporal factor model. And third, our ICAPM framework implies that the risk price on each ICAPM factor can only be evaluated when controlling for the other two factors. As we detail later, this aspect is important as risk prices are severely distorted if we do not control for all ICAPM factors simultaneously (e.g., the $r_{\mathbb{V}}$ risk price is close to zero).

Several papers have attempted to provide an ex-post ICAPM interpretation for prominent tradable factors by linking them to variation in state variables related to investment opportunities (e.g., Vassalou (2003), Petkova (2006), Maio and Santa-Clara (2012), Boons (2016), Cooper and Maio (2019), and Barroso, Boons, and Karehnke (2020)). Our work is different from (and complementary to) this literature as we construct tradable factors that are consistent with the ICAPM *ex-ante* as oppose to providing an *ex-post* analysis of whether tradable factors are consistent with the ICAPM. Moreover, we focus on a structural ICAPM with state variables determined by a Bayesian learning framework, which imposes stronger restrictions than the consistency tests in this literature as they only rely on the overall logic of Merton (1973) that tradable factors should be related variation in investment opportunities. The result is that our analysis allows us to overcome the ICAPM "fishing license" (Fama (1991)), map each of our tradable factors to a specific source of risk, and infer that the risk price magnitudes are internally consistent with the ICAPM under reasonable risk aversion, tasks that are difficult to accomplish without a fully specified ICAPM framework.

Finally, our work provides a natural response to several papers that reveal important issues in asset pricing tests that raise skepticism about the ability of current models to explain the cross-section of returns. First, we address criticisms related to the measurement of non-tradable risk factors and estimation of their risk prices by relying on tradable factors and requiring the model to price them, which makes our analysis immune to typical issues in the estimation of risk prices and relatively robust to mispecification in the factors' construction.[3] Second, we deal with problems arising from the testing assets used to evaluate models by relying on the testing assets recommended in Lewellen, Nagel, and Shanken (2010), which are reasonably immune to the core issues raised in the literature.[4] And third, we deal with the lack of economic interpretability of factor models (e.g., Ferson, Sarkissian, and Simin (1999) and Kozak, Nagel, and Santosh (2018)) by constructing our risk factors so that they closely reflect the underlying ICAPM risks affecting long-term investing.[5]

The rest of this paper is organized as follows. Section 1 details the structural ICAPM and its factor model implementation, Section 2 explores our intertemporal factor model and its risk prices, Section 3 compares different factor models, Section 4 studies the pricing of anomalies, and Section 5 concludes by summarizing our results and implications. The Internet Appendix contains technical derivations, empirical details, and supplementary results.

[3]Parker and Julliard (2005), Jagannathan and Wang (2007), Savov (2011), and Kroencke (2017) outline important challenges in measuring consumption growth (needed to test consumption CAPMs) and Chen and Zhao (2009) focus on sensitivity issues in measuring news to long-term expected returns (necessary in ICAPM structural tests). Relatedly, Lewellen, Nagel, and Shanken (2010), Kan, Robotti, and Shanken (2013), and Laurinaityte et al. (2020) detail several limitations in the estimation of risk prices of non-tradable factors.

[4]See Lo and MacKinlay (1990), Berk (2000), Grauer and Janmaat (2004), Ahn, Conrad, and Dittmar (2009), Lewellen, Nagel, and Shanken (2010), Cederburg and O'Doherty (2015), Ang, Liu, and Schwarz (2020), and Tian (2021) for the limitations of asset pricing tests based on typical anomaly portfolios.

[5]Models with arbitrary macroeconomic variables face a similar economic interpretability issue in that average returns can be "explained" by macroeconomic variables even when they do not identify risks relevant to investors (Shanken (1992), Reisman (1992), and Nawalkha (1997)).

1 The ICAPM and its Factor Model Implementation

This section introduces the ICAPM framework and details how we implement our intertemporal risk factor model while avoiding the ICAPM "fishing license" (Fama (1991)). The ICAPM structure follows prior work (Campbell et al. (2018) and Gonçalves (2021a)), but our intertemporal risk factor model implementation is new. Subsection 1.1 outlines the structural ICAPM, Subsection 1.2 builds the map from the structural ICAPM to an intertemporal risk factor model, and Subsection 1.3 details how we build our intertemporal risk factor model empirically. Internet Appendix A provides all derivations.

To simplify notation, we use tilde to represent shocks (e.g., $\widetilde{x}_t \equiv x_t - \mathbb{E}_{t-1}[x_t]$), and suppress time subscripts inside conditional moments when convenient (e.g., $\mathbb{E}_t[x] \equiv \mathbb{E}_t[x_{t+1}]$).

1.1 The ICAPM SDF

A long-term (i.e., infinitely lived) investor has Epstein-Zin recursive preferences (Epstein and Zin (1989)) with time discount factor δ, intertemporal elasticity of substitution $\psi = 1$, and relative risk aversion γ.[6] The investor chooses consumption and portfolio allocation to maximize lifetime utility subject to the usual budget constraint. The log Stochastic Discount Factor ($sdf_t = log(SDF_t)$) derived from the investor's optimality conditions with respect to consumption and portfolio allocation is given by[7]

$$sdf_t = \kappa - \gamma \cdot r_{w,t} - (\gamma - 1) \cdot \widetilde{vw}_t, \quad (1)$$

where r_w captures log returns on the investor's wealth portfolio and $vw_t = log(V_t/W_t)$ reflects the investor's log value-wealth ratio.[8]

[6] ψ plays no role in our empirical analysis. We fix $\psi = 1$ in the exposition because in this case the ICAPM implications do not require a log-linear approximation. All ICAPM implications exposed follow (as approximations) if $\psi \neq 1$ as demonstrated in Internet Appendix A.2.

[7] To simplify the exposition, we assume $\mathbb{E}_t[vw] \approx vw_t/\delta$. Otherwise, the intercept has a small time varying component, $\kappa_{t-1} = \kappa + (\gamma - 1) \cdot (vw_{t-1}/\delta - \mathbb{E}_{t-1}[vw])$. Such time varying component has no implications for risk premia (only for interest rate variation), and thus does not play a role in our analysis.

[8] Note that intertemporal risk ($\widetilde{vw}$) is only relevant if long-term prospects are important. Specifically, if the investor had a one period horizon, then the value function would be zero at the end of the period so that $\widetilde{vw}$ would not enter the SDF. Similarly, if investment opportunities did not vary over time, then $\widetilde{vw}$ would

For our factor model implementation, we further decompose $\widetilde{vw}$ into the two components of intertemporal risk: news to long-term expected returns and volatility (i.e., reinvestment risk and volatility risk). Specifically, Internet Appendix A.2 shows that[9]

$$\begin{aligned} vw_t \; &= constant \; + \; \mathbb{E}_t\left[\sum_{h=1}^{\infty} \delta^h \cdot r_{w,t+h}\right] - \frac{(\gamma-1)}{2} \cdot \mathbb{E}_t\left[\sum_{h=1}^{\infty} \delta^h \cdot \mathbb{V}ar_{t+h-1}\left[v_{t+h}\right]\right] \\ &\Downarrow \\ \widetilde{vw}_t \; &= N_{\mathbb{E},t} \; - \; \frac{(\gamma-1)}{2} \cdot N_{\mathbb{V},t} \end{aligned} \tag{2}$$

where

$N_{\mathbb{E},t} = (\mathbb{E}_t - \mathbb{E}_{t-1})\left[\sum_{h=1}^{\infty} \delta^h \cdot r_{w,t+h}\right]$ is expected return news

$N_{\mathbb{V},t} = (\mathbb{E}_t - \mathbb{E}_{t-1})\left[\sum_{h=1}^{\infty} \delta^h \cdot \mathbb{V}ar_{t+h-1}[v_{t+h}]\right]$ is volatility news

Increases in expected returns increase the value function relative to current wealth as wealth is expected to grow at a higher rate, delivering a more valuable future consumption stream. In contrast, higher future volatility decreases the value function relative to current wealth since achieving the given wealth growth requires facing more risk going forward.

Substituting Equation 2 into the log SDF in Equation 1 yields:

$$sdf_t \; = \kappa \; - \; \gamma \cdot r_{w,t} \; - \; (\gamma - 1) \cdot N_{\mathbb{E},t} \; + \; \frac{(\gamma-1)^2}{2} \cdot N_{\mathbb{V},t} \tag{3}$$

which shows that, with $\gamma > 1$, declines in expected returns and increases in expected volatility (holding current wealth fixed) represent bad news to the long-term investor.[10]

In summary, the ICAPM SDF reflects market risk (r_w), reinvestment risk ($N_{\mathbb{E}}$), and volatility risk ($N_{\mathbb{V}}$). Moreover, the last two terms are responsible for the intertemporal risk component in the ICAPM SDF, with the ICAPM reducing to the CAPM in the absence of these intertemporal risk terms, which occurs if long-term investing prospects are not

always be zero. Finally, if $\gamma = 1$, then the investor acts myopically and effectively ignores $\widetilde{vw}$.

[9]The first line in Equation 2 relies on a second order Taylor approximation for $\mathbb{E}_t[SDF_{t+1} \cdot R_{w,t+1}] = 1$ that holds exactly if $\widetilde{v}_t = \widetilde{r}_{w,t} + \widetilde{vw}_t$ is conditionally normal. Otherwise, there are other news terms related to v_t higher order moments. We leave an exploration of such higher order risk factors to future work.

[10]The $\gamma > 1$ condition for the positive $N_{\mathbb{E}}$ risk price is a consequence of two offsetting effects. An asset that comoves positively with expected returns is desirable since it provides more capital to investors at a time when the marginal product of capital is high. However, the asset also exposes investors to reinvestment risk. When $\gamma > 1$, the latter effect dominates so that the $N_{\mathbb{E}}$ risk price is positive.

important (see Footnote 8).

1.2 The Intertemporal Risk Factor Model

We apply a first order Taylor expansion to Equation 3 to obtain

$$SDF_t \;\approx a_o \;-\; \mathbb{E}[SDF] \cdot \left(\gamma \cdot r_{w,t} \;+\; (\gamma - 1) \cdot N_{\mathbb{E}} \;-\; \frac{(\gamma-1)^2}{2} \cdot N_{\mathbb{V}} \right) \tag{4}$$

and let R_t reflect a vector of all asset returns so that projecting the risk factors onto R_t (with the constraint that mimicking factors are long-short portfolios) yields

$$\begin{cases} r_{w,t} \;=\; \zeta_{0,m} \;+\; \zeta_m \cdot r_{m,t} \;+\; \epsilon_{m,t} \\ N_{\mathbb{E},t} \;=\; \zeta_{0,\mathbb{E}} \;+\; \zeta_{\mathbb{E}} \cdot r_{\mathbb{E},t} \;+\; \epsilon_{\mathbb{E},t} \\ N_{\mathbb{V},t} \;=\; \zeta_{0,\mathbb{V}} \;+\; \zeta_{\mathbb{V}} \cdot r_{\mathbb{V},t} \;+\; \epsilon_{\mathbb{V},t} \end{cases} \tag{5}$$

where (for $k = m$, $\mathbb{E}$, $\mathbb{V}$) $r_{k,t} = \pi_k' R_t$ represent tradable mimicking factors, $\Sigma_j \pi_{k,j} = 0$ capture zero-cost portfolio weights, and $\zeta_k \geq 0$ reflect positive normalizing constants.

The projection orthogonality conditions imply $\mathbb{E}[\epsilon_m \cdot r_j] = \mathbb{E}[\epsilon_{\mathbb{E}} \cdot r_j] = \mathbb{E}[\epsilon_{\mathbb{V}} \cdot r_j] = 0$ for any excess return, $r_{j,t} = R_{j,t} - R_{i,t}$. As such, substituting Equation 5 into $\mathbb{E}[SDF \cdot r_j] = 0$ yields the Euler condition $\mathbb{E}[M \cdot r_j] = 0$ for our tradable SDF,

$$M_t \;= a \;-\; b_m \cdot r_{m,t} \;-\; b_{\mathbb{E}} \cdot r_{\mathbb{E},t} \;-\; b_{\mathbb{V}} \cdot r_{\mathbb{V},t}, \tag{6}$$

implying the intertemporal factor model risk premia equation,

$$\mathbb{E}[r_j] \;\;=\;\; \frac{b_m}{\mathbb{E}[M]} \cdot \mathbb{C}ov[r_j, r_m] \;\;+\;\; \frac{b_{\mathbb{E}}}{\mathbb{E}[M]} \cdot \mathbb{C}ov[r_j, r_{\mathbb{E}}] \;\;+\;\; \frac{b_{\mathbb{V}}}{\mathbb{E}[M]} \cdot \mathbb{C}ov[r_j, r_{\mathbb{V}}], \tag{7}$$

with restrictions $b_m \geq 0$, $b_{\mathbb{V}} \leq 0$, and $\text{sign}(b_{\mathbb{E}}) = \text{sign}(\gamma - 1)$ (so $b_{\mathbb{E}} \geq 0$ as long as $\gamma \geq 1$), which we explore in our empirical analysis.[11]

Finally, letting $b = [b_m \;\; b_{\mathbb{E}} \;\; b_{\mathbb{V}}]'$ and $f = [r_m \;\; r_{\mathbb{E}} \;\; r_{\mathbb{V}}]'$, the Euler condition $\mathbb{E}[M \cdot f] = 0$ implies (with Σ_f as the f covariance matrix):

---

[11] Specifically, the risk prices are given by $b_m = \zeta_m \cdot \mathbb{E}[M] \cdot \gamma$, $b_{\mathbb{E}} = \zeta_{\mathbb{E}} \cdot \mathbb{E}[M] \cdot (\gamma - 1)$, and $b_{\mathbb{V}} = -\zeta_{\mathbb{V}} \cdot \mathbb{E}[M] \cdot \frac{(\gamma-1)^2}{2}$. The ICAPM restrictions then follow from $\gamma \geq 0$ and $\zeta_k \geq 0$ for $k = m$, $\mathbb{E}$, $\mathbb{V}$.

$$b \; = \mathbb{E}[M] \cdot \Sigma_f^{-1} \mathbb{E}[f] \tag{8}$$

Since our risk factors are excess returns, the Euler condition $\mathbb{E}[M \cdot f] = 0$ does not identify a. As such, we normalize a by imposing $\mathbb{E}[M] = 1$. This normalization does not affect the model-implied risk premium on any asset, $\mathbb{E}[r_j]$, as can be seeing by substituting Equation 8 into Equation 7 (see Chapter 13.2 in Cochrane (2005) for more details on this normalization).

1.3 The Market and Intertemporal Risk Factors

The key to a valid factor model implementation of the ICAPM is to build r_m, $r_{\mathbb{E}}$, and $r_{\mathbb{V}}$ that represent tradable mimicking factors for market risk (r_w), reinvestment risk ($N_{\mathbb{E}}$), and volatility risk ($N_{\mathbb{V}}$). This subsection details how we approach this crucial task.

1.3.1 The Market Risk Factor

Since $r_{w,t}$ are real log returns on the representative investor's wealth portfolio, if the equity market reflects the wealth portfolio, then we have

$$r_{w,t} \; \approx R_{m,t} - R_{inf,t} \; = \; r_{m,t} \; + \; \epsilon_{m,t} \tag{9}$$

where $r_{m,t} = R_{m,t} - R_{f,t}$ are equity market returns in excess of the risk-free rate and $\epsilon_{m,t} = R_{f,t} - R_{inf,t}$ (roughly) reflects unexpected inflation, which we assume does not price excess returns since inflation affects both legs of any long-short strategy.

The argument above motivates us to use r_m as our market risk factor and implies $\zeta_m = 1$.[12] As such, we have $\gamma = b_m/\mathbb{E}[M] = 1_m' \Sigma_f^{-1} \mathbb{E}[f]$, which allows us to infer the relative risk aversion coefficient from our Σ_f and $\mathbb{E}[f]$ estimates. Using r_m as the ICAPM market factor is also empirically convenient given that effectively all factor models proposed in the literature (including the CAPM) rely on this same r_m as their market factor.

[12] We measure r_m using the market risk factor available in Kenneth French's data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html).

1.3.2 The $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ Proxies

Constructing the intertemporal risk factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) is challenging as it requires real time proxies for news to long-term expected returns ($N_{\mathbb{E}}$) and volatility ($N_{\mathbb{V}}$). To avoid the ICAPM "fishing license" (Fama (1991)) when constructing $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, we do not specify an arbitrary set of state variables for expected returns and volatility, but instead build a simple Bayesian learning framework in which the long-term investor observes only market prices and dividends (similar to Binsbergen and Koijen (2010)). In this framework, the log dividend yield ultimately provides a signal for the (unobservable) mean log return process so that $N_{\mathbb{E}}$ is linked to log dividend yield shocks. Similarly, log realized variance provides a signal for the conditional log variance so that $N_{\mathbb{V}}$ is linked to log realized variance shocks. As a consequence, our Bayesian framework motivates the use of changes in the log dividend yield (Δdp) and log realized variance ($\Delta\sigma^2$) as real time proxies for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, respectively.

Letting $r_{w,t}$ and Δd_t reflect the monthly log wealth return and growth in annual dividends, we assume that $r_{w,t}$, $\mu_t = \mathbb{E}_t[r_{w,t+1}|\mu_t]$, and $g_t = \mathbb{E}_t[\Delta d_{t+1}|g_t]$ have monthly dynamics with

$$r_{w,t+1} = \mu_t + \widetilde{r}^*_{w,t+1} \tag{10}$$

$$\mu_{t+1} = \mu + \phi_\mu \cdot (\mu_t - \mu) + \widetilde{\mu}^*_{t+1} \tag{11}$$

$$g_{t+1} = g + \phi_g \cdot (g_t - g) + \widetilde{g}^*_{t+1} \tag{12}$$

$$\Sigma_t = \Sigma \cdot \mathbb{V}r_t \tag{13}$$

where $[\widetilde{r}^*_{w,t+1}, \widetilde{\mu}^*_{t+1}, \widetilde{g}^*_{t+1}] \sim N(0, \Sigma_t)$ are unobservable shocks and $\mathbb{V}r_t = \mathbb{V}ar_t[\widetilde{r}^*_{w,t+1}]$ captures the wealth portfolio return variance with dynamics detailed later in this subsection.[13]

[13] While the assumption that $\widetilde{r}^*_{w,t}$, $\widetilde{\mu}^*_t$, and $\widetilde{g}^*_t$ have a single volatility factor (i.e., Equation 13) is stylized, it keeps the ICAPM tractable and is analogous to related models that include stochastic volatility in the ICAPM and long-run risks literatures (e.g., Campbell et al. (2018) and Bansal et al. (2014)).

Then, from the log-linear valuation identity in Campbell and Shiller (1989), we have[14]

$$dp_t - dp \;= \Phi_\mu \cdot (\mu_t - \mu) \;-\; \Phi_g \cdot (g_t - g) \tag{14}$$

where $\Phi_\mu = (1-\phi_\mu^{12})/[(1-\phi_\mu)(1-\rho\cdot\phi_\mu^{12})]$, $\Phi_g = (1-\phi_g^{12})/[(1-\phi_g)(1-\rho\cdot\phi_g^{12})]$, and $\rho = e^{-dp}/(1+e^{-dp})$, with $dp_t = log(D_t/P_t)$ reflecting the annual dividend yield and dp capturing the average dp_t.

The long-term investor observes $r_{w,t}$, dp_t, and all model parameters, but not μ_t and g_t.[15] As such, the Bayesian investor forms the expected return process, $\mathbb{E}r_t = \mathbb{E}_t[r_{w,t+1}] = \mathbb{E}_t[\mu_t]$, endogenously from observations of $r_{w,t}$ and dp_t. Specifically, as demonstrated in Internet Appendix A.3, $r_{w,t}$ and dp_t provide signals for μ_t so that

$$\mathbb{E}r_{t+1} \;= \mathbb{E}r \;+\; \phi_{\mathbb{E}} \cdot (\mathbb{E}r_t - \mathbb{E}r) \;+\; \widetilde{\mathbb{E}r}_{t+1} \tag{15}$$

where parameters are given by $\mathbb{E}r = \mu$, $\phi_{\mathbb{E}} = \phi_\mu$, $\nu_o = 1/[\Phi_\mu \cdot (\phi_\mu - \phi_g)]$, and shocks by

$$\widetilde{\mathbb{E}r}_{t+1} \;= \xi_{dp,t} \cdot \widetilde{dp}^{\,o}_{t+1} \;+\; \xi_{r,t} \cdot \widetilde{r}_{w,t+1}, \tag{16}$$

$$\widetilde{dp}^{\,o}_{t+1} \;= \mu \;+\; \nu_o \cdot [(dp_{t+1} - dp) - \phi_g \cdot (dp_t - dp)] \;-\; \mathbb{E}r_t \tag{17}$$

$$\widetilde{r}_{w,t+1} \;= r_{w,t+1} - \mathbb{E}r_t \tag{18}$$

with $\xi_{dp,t}$ and $\xi_{r,t}$ capturing functions of the underlying model parameters and the $\mathbb{V}r_t$ history.

Intuitively, at time $t+1$, the investor forms the Bayesian expectation for μ_{t+1} (i.e., $\mathbb{E}r_{t+1} = \mathbb{E}_{t+1}[\mu_{t+1}]$) by updating the time t expectation ($\mathbb{E}r_t = \mathbb{E}_t[\mu_t]$) using the observed wealth

[14]The Φ_μ and Φ_g parameters in Equation 14 are derived in Internet Appendix A.3 and account for the fact that dp_t is the annual dividend yield while μ_t and g_t have monthly dynamics. The use of the annual dividend yield in a monthly model allows stock exposures to monthly changes in the annual dividend yield to be used in the empirical construction of our reinvestment risk factor (detailed later) while being fully consistent with the model. The focus on monthly betas makes it feasible to rely on a rolling window beta estimation while capturing the possibility that firm-level betas vary over time. At the same time, the use of annual (as opposed to monthly) dividend yield avoids dividend seasonality issues.

[15]The assumption that the investor knows all model parameters is justified by the fact that all parameters can be consistently estimated, for example, by conditional maximum likelihood as we explain in Subsection 1.3.4. In contrast, the investor can never fully recover μ_t. The best that the investor can do is to obtain the expectation of μ_t given the information observed up to time t, which is precisely how we model the belief process of the investor in this subsection.

portfolio return and dividend yield processes from t to $t+1$ as well as the knowledge that μ_t evolves as in Equation 11. As a consequence, the shock to the investor's endogenous expected return process, $\widetilde{\mathbb{E}r}_{t+1}$, is a linear combination of $\widetilde{r}_{w,t+1}$ and $\widetilde{dp}^{o}_{t+1}$.

In terms of the volatility dynamics, we model $\mathbb{V}r_t$ as a Realized log-GARCH process (see Hansen, Huang, and Shek (2012)) so that[16]

$$log(\mathbb{V}r_{t+1}) \; = \omega_{\mathbb{V}r} \; + \; \phi_{\mathbb{V}r} \cdot log(\mathbb{V}r_t) \; + \; \phi_\sigma \cdot \sigma^2_{t+1} \tag{19}$$

$$\sigma^2_{t+1} \; = \omega_\sigma \; + \; log(\mathbb{V}r_t) \; + \; \widetilde{\sigma}^2_{t+1} \tag{20}$$

where $\widetilde{\sigma}^2_t \sim N(0, \sigma^2_\sigma)$ with σ^2_t reflecting the log of the realized variance of r_w over month t.

Internet Appendix A.3 shows that our $\mathbb{E}r_t$ and $\mathbb{V}r_t$ dynamics (in conjunction with the vw_t recursion implied by the ICAPM) result in[17]

$$N_{\mathbb{E},t+1} \; = \theta_{\mathbb{E}} \cdot \widetilde{\mathbb{E}r}_{t+1} \; = \; \theta_{dp,t} \cdot \widetilde{dp}^{o}_{t+1} \; + \; \theta_{r,t} \cdot \widetilde{r}_{w,t+1} \tag{21}$$

and

$$N_{\mathbb{V},t+1} \; \approx \theta_\sigma \cdot \widetilde{\sigma}^2_{t+1} \tag{22}$$

where $\theta_{\mathbb{E}} = \delta/(1 - \delta \cdot \phi_{\mathbb{E}})$, $\theta_{dp,t} = \theta_{\mathbb{E}} \cdot \xi_{dp,t}$, $\theta_{r,t} = \theta_{\mathbb{E}} \cdot \xi_{r,t}$, and $\theta_\sigma > 0$ reflects a function of δ, γ, and the expected return and volatility parameters.

We rely on Equations 21 and 22 to motivate our simple real time proxies for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ (up to a constant of proportionality). Specifically, when constructing $r_{\mathbb{E}}$, we use $\Delta dp_t = dp_t - dp_{t-1}$

[16]Our use of a Realized log-GARCH process effectively treats the conditional variance as observable even though μ_t is not. This approach simplifies exposition and is consistent with prior literature (e.g., Anderson, Ghysels, and Juergens (2009)), with the justification relating back to the work of Merton (1980) and Foster and Nelson (1996). However, our $log(\mathbb{V}r_t)$ process can be derived as the Bayesian posterior of the log conditional variance in a latent stochastic volatility model in which σ^2_t provides a noisy signal for the log conditional variance (see Internet Appendix A.3 for the proof). This Bayesian interpretation of our $\mathbb{V}r_t$ process relies on the normality of $\widetilde{\sigma}^2_t$, which holds approximately in our log-GARCH (but not in a level-GARCH) specification, justifying our modeling approach. Internet Appendix D.4 provides more details on this normality issue and also demonstrates that our log-GARCH model better forecasts long-run variance than an analogous level-GARCH process.

[17]The $N_{\mathbb{E},t} = \theta_{\mathbb{E}} \cdot \widetilde{\mathbb{E}r}_t$ result holds exactly. In contrast, $N_{\mathbb{V},t}$ is generally a non-linear function of $\widetilde{\sigma}^2_t$ so that the $N_{\mathbb{V},t} \approx \theta_{\mathbb{V}} \cdot \widetilde{\sigma}^2_t$ result relies on some linear approximations. In the main text, we use this $N_{\mathbb{V},t} \approx \theta_{\mathbb{V}} \cdot \widetilde{\sigma}^2_t$ approximation for simplicity, but Internet Appendix D.7 shows that our results are similar (but slightly stronger) if we solve for the nonlinear $N_{\mathbb{V},t}$ numerically.

as proxy for $N_{\mathbb{E},t}$. Equation 21 shows that $N_{\mathbb{E},t}$ is a linear combination of $\widetilde{dp}_t^o$ and $\widetilde{r}_{w,t}$ and it is well-known that realized returns provide a very noisy signal for μ_t so that we expect $\theta_{dp,t}$ to be much larger than $\theta_{r,t}$, justifying our use of Δdp_t.[18] Similarly, when constructing $r_{\mathbb{V}}$, we use $\Delta\sigma_t^2 = \sigma_t^2 - \sigma_{t-1}^2$ as a proxy for $N_{\mathbb{V},t}$, which is directly motivated by Equation 22. The next subsection explains our $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ construction.

1.3.3 The Tradable Mimicking Factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$

The typical approach in the literature to construct the zero-cost mimicking factor for a generic nontradable factor x is to[19]

1. Sort stocks based on their $\beta_x = \mathbb{C}ov[x, R]/\mathbb{V}ar[x]$ to form base portfolios

2. Project x onto the base portfolios with a constraint that coefficients add to zero

In our application, real time estimation of the nontradable factors (to estimate $\beta_{N\mathbb{E}}$ and $\beta_{N\mathbb{V}}$ in step 1) would lead to large estimation errors and parameter instability. As such, we replace step 1 with sorts on β_{dp} and β_{σ^2}, so that we are effectively using Δdp and $\Delta\sigma^2$ to proxy for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ (up to a constant of proportionality), as discussed in the previous subsection.[20] Specifically, at each month t (with t from 12/1927 to 11/2019) and for each stock in the CRSP dataset (incorporated in the US and traded on NYSE, AMEX, or NASDAQ), we measure stock-level β_{dp} and β_{σ^2} based on univariate betas of monthly returns on Δdp and $\Delta\sigma^2$, respectively. We use a 5-year rolling window to estimate betas and require stocks to

[18]Our empirical results confirm our expectation. Specifically, in our estimation described in Subsection 1.3.4, the average $\theta_{dp,t}$ is 31 (39) times lager than the average $\theta_{r,t}$ over our Long (Modern) Sample.

[19]The motivation in the literature is as follows. If we project x onto the vector of all asset returns, R_t, we find that the x mimicking factor weights are generally proportional to $\Sigma_R^{-1}\beta_x$, where $\Sigma_R = \mathbb{V}ar[R]$ (see Breeden, Gibbons, and Litzenberger (1989)). Since Σ_R estimates are very noisy when the number of assets in R_t is large, the literature typically uses β_x as a sorting signal to construct a small set of base portfolios (step 1) and then project x directly onto these base portfolios (step 2). We also explore the alternative approach proposed by Giglio and Xiu (2020) to construct the base portfolios in step 1 (see Footnote 25).

[20]Our annual log dividend yield measure, dp_t, follows the construction in Gonçalves (2021a) and our log realized variance is $\sigma_t^2 = log(\frac{21}{N_t} \cdot \Sigma_{i=1}^{N_t} r_{w,t,i}^2)$, with $r_{w,t,i}$ reflecting the log return on the CRSP value-weighted index on day i of month t. Internet Appendix C.1 provides further measurement details for dp_t and σ_t^2.

have the full five years of data available to be included in the construction of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$.[21]

Similarly, the projections in step 2 would likely be unstable if performed in real time (and could lead to large weights on relatively small stocks). As such, we replace step 2 with the construction of value-weighted high minus low returns (following Herskovic, Moreira, and Muir (2019)). Specifically, we form four value-weighted portfolios ($R_{L\mathbb{E}}$, $R_{H\mathbb{E}}$, $R_{L\mathbb{V}}$, $R_{H\mathbb{V}}$) by sorting stocks based on their β_{dp} and β_{σ^2} each month. The $R_{L\mathbb{E}}$ ($R_{L\mathbb{V}}$) portfolio contains stocks that are below the 30% NYSE breakpoint for β_{dp} (β_{σ^2}) while the $R_{H\mathbb{E}}$ ($R_{H\mathbb{V}}$) portfolio contains stocks that are above the 70% NYSE breakpoint for β_{dp} (β_{σ^2}). Finally, our intertemporal risk factors are constructed as $r_{\mathbb{E}} = R_{H\mathbb{E}} - R_{L\mathbb{E}}$ and $r_{\mathbb{V}} = R_{H\mathbb{V}} - R_{L\mathbb{V}}$.[22] Since our betas are measured from 12/1927 to 11/2019, our $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ are available from 01/1928 to 12/2019.

1.3.4 The ex-post Mimicking Factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$

Beyond constructing our tradable mimicking factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) using Δdp and $\Delta\sigma^2$ as proxies for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, we also construct ex-post mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ (which we call $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$) based on ex-post estimated $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ following the model structure in Subsection 1.3.2. While $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ are not tradable (and thus cannot be used for asset pricing tests), they are used in parts of our empirical analysis to validate $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$.

We start by estimating the Realized log-GARCH process in Equations 19 and 20 by non-linear least squares targeting the 10-year subsequent realized variance (Internet Appendix D.6 provides a maximum likelihood estimation). This estimation approach is consistent with our objective of focusing on long-run variance (to measure $N_{\mathbb{V}}$) and is in line with the work of Ederington and Guan (2010), who find that targeting multiperiod realized variance in GARCH estimation tends to result in better long-run variance forecasts. Then, given $\mathbb{V}r_t$, we estimate the $\mathbb{E}r_t$ process parameters (in Equations 15 to 17) by conditional maximum

[21] In the beginning of our sample period, we start with a shorter window (12 months for β_{dp} and 23 months for β_{σ^2}) due to data availability and expand the window for each factor until we have five years of data to estimate the respective beta. Internet Appendix C.2 provides further details and Internet Appendix D.3 shows that our main results are very similar with betas estimated on a 3-year rolling window.

[22] Our approach is analogous to Fama and French (1993), except that we do not control for size or orthogonalize our factors since the ICAPM does not imply factors are orthogonal to size or to each other.

likelihood as in Binsbergen and Koijen (2010). Further estimation details (and parameter estimates) are provided in Internet Appendix B.1.

The above procedure provides us with ex-post estimates for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ from Equations 21 and 22.[23] Using these estimates, we obtain ex-post mimicking portfolios for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ following the formal steps 1 and 2 in the prior subsection. Specifically, we form value-weighted decile portfolios by sorting stocks based on $\beta_{N\mathbb{E}}$ and $\beta_{N\mathbb{V}}$ estimated over the same 5-year rolling window that we use to estimate stock-level β_{dp} and β_{σ^2}. We then project our estimated $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ onto their respective deciles (while requiring weights to add to zero) to obtain the ex-post mimicking factors, $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$, as linear combinations of the returns on the $\beta_{N\mathbb{E}}$ and $\beta_{N\mathbb{V}}$ decile portfolios.

2 The Intertemporal Risk Factor Model: Main Results

This section presents our main empirical results. Subsection 2.1 demonstrates that our intertemporal risk factors reflect tradable mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, Subsection 2.2 explores decile portfolios from sorts on β_{dp} and β_{σ^2}, Subsection 2.3 shows that investors can extract large risk premia in real time from exposures to the ICAPM risk factors, Subsection 2.4 presents the estimated ICAPM risk prices, and Subsection 2.5 outlines the core implications and limitations of our main results.

Our empirical analysis considers two sample periods. The Long Sample (from 1928 to 2019) covers the full period over which we can produce our factors while the Modern Sample (from 1973 to 2019) focuses on (roughly) the second half of our Long Sample, which is the longest period over which we can construct all factors from the other factor models we study.

[23] Strictly speaking, we use $\widetilde{\mathbb{E}r}_t$ and $\widetilde{\sigma}_t^2$ instead of $N_{\mathbb{E}} = \theta_{\mathbb{E}} \cdot \widetilde{\mathbb{E}r}_t$ and $N_{\mathbb{V}} = \theta_{\sigma} \cdot \widetilde{\sigma}_t^2$ so that we avoid estimating $\theta_{\mathbb{E}}$ and θ_{σ} (which depend on δ and γ) when constructing $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$. This approach only affects $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ up to a constant of proportionality, and thus does not impact any of the results we report.

2.1 Validating the Intertemporal Risk Factors

This subsection shows that our tradable intertemporal risk factors, $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$, which are constructed from sorts on β_{dp} and β_{σ^2}, are good real time mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$.

Table 1 reports correlations between different tradable and nontradable risk factors over our Modern and Long samples. Panel A focuses on the correlation matrix for Δdp, $N_{\mathbb{E}}$, $r_{\mathbb{E}}$, and $r_{N\mathbb{E}}$, which reflect reinvestment risk, while Panel B focuses on the correlation matrix for $\Delta\sigma^2$, $N_{\mathbb{V}}$, $r_{\mathbb{V}}$, and $r_{N\mathbb{V}}$, which reflect volatility risk.

The key result from Table 1 is that $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ strongly correlate with our ex-post intertemporal risk factors, $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$. Specifically, the correlations between the ex-post estimated news, $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, and their respective real time proxies, Δdp and $\Delta\sigma^2$, are relatively high with $Cor(\Delta dp, N_{\mathbb{E}}) = 0.91$ and $Cor(\Delta\sigma^2, N_{\mathbb{V}}) = 0.75$ over the Long Sample and $Cor(\Delta dp, N_{\mathbb{E}}) = 0.88$ and $Cor(\Delta\sigma^2, N_{\mathbb{V}}) = 0.92$ over the Modern Sample. As a consequence, we have $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}}) = 0.93$ and $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}}) = 0.81$ over the Long Sample and $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}}) = 0.89$ and $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}}) = 0.84$ over the Modern Sample. Figure 2 provides a visual representation of this result by plotting $r_{\mathbb{E}}$, $r_{N\mathbb{E}}$, $r_{\mathbb{V}}$, and $r_{N\mathbb{V}}$ after a filtering process so that their correlations can be easily visualized.[24]

Table 1 also shows that $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ are almost as good as $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ in mimicking $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ despite the fact that $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ are constructed in real time whereas $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ are designed to reflect ex-post mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$. For instance, over our Long Sample, we have $Cor(r_{\mathbb{E}}, N_{\mathbb{E}}) = Cor(r_{N\mathbb{E}}, N_{\mathbb{E}}) = 0.51$ and $Cor(r_{\mathbb{V}}, N_{\mathbb{V}}) = 0.24 < 0.30 = Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$. Similarly, over our Modern Sample, we have $Cor(r_{\mathbb{E}}, N_{\mathbb{E}}) = 0.36 < 0.39 = Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$ and $Cor(r_{\mathbb{V}}, N_{\mathbb{V}}) = 0.28 < 0.36 = Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$.

The correlations in the previous paragraph are suggestive of market incompleteness (or measurement noise in $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$), which leads the correlations between tradable and nontradable factors to be much lower than one. However, the overall correlation values we find

[24] Following Campbell et al. (2018) and Gonçalves (2021a), the filtering process for a generic risk factor x is based on an exponentially weighted moving average of normalized shocks, $fx_t = (1-\pi)\cdot(x_t - \overline{x})/\sigma_x + \pi \cdot fx_{t-1}$, with a half-life of two years ($\pi = 0.5^{1/24} \approx 0.97$).

are also in line with the correlations between nontradable factors and their mimicking factors reported in the literature (see, e.g., Alekseev et al. (2021)).[25] As such, the level of market incompleteness (or measurement noise) in our setting is comparable to the prior literature.

Finally, note that Table 1 focuses on the ability of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ to act as mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ when the Bayesian framework we introduce in Subsection 1.3.2 is valid. This approach allows us to overcome the ICAPM fishing license (Fama (1991)) as the model dictates how $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ must be constructed (from Equations 21 and 22). However, much of the prior ICAPM literature estimates $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ from vector autoregressions with multiple pre-specified state variables. Internet Appendix D.5 shows that our simple $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ tradable factors are also highly correlated with the ex-post mimicking factors we estimate using vector autoregressions in which the state variables are the ones in Gonçalves (2021a) augmented by realized variance as in Campbell et al. (2018). As such, while our Bayesian framework is useful in disciplining our construction of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$, these tradable factors mimick $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ even if these news are estimated from vector autoregressions.

2.2 Decile Portfolios from Sorts on β_{dp} and β_{σ^2}

This subsection studies decile portfolios from sorts on β_{dp} and β_{σ^2}, which are intrinsically connected to our construction of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$.

Tables 2 and 3 provide betas normalized to market beta units as well as average returns and αs with respect to the CAPM and ICAPM.[26] Table 2 focuses on β_{dp} sorted portfolios

[25]In Internet Appendix D.2, we explore an alternative approach that estimates factor mimicking weights based on the method in Giglio and Xiu (2020) and uses them (instead of betas) to create the base portfolios underlying $r_{\mathbb{E}}$, $r_{\mathbb{V}}$, $r_{N\mathbb{E}}$, and $r_{N\mathbb{V}}$. The mimicking factors obtained from this alternative approach do not yield higher correlations with $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ than our simple method. The crux of the matter is that applying this alternative approach requires out-of-sample estimation of factor mimicking weights, with the benefit of this alternative approach being outweighed by the limitations of out-of-sample estimation in our setting.

[26]For each statistic, we report the values for each decile, the spread between deciles 10 and 1, and the slopes. βs are based on normalized factors (e.g., $x_t = r_{\mathbb{E},t} \cdot \sqrt{\mathbb{V}ar[r_m]}/\sqrt{\mathbb{V}ar[r_{\mathbb{E}}]}$) so that they are in market beta units. The slopes are obtained from panel regressions. In the case of the βs, the regression is specified as $r_{p,t} = a_p + (a + b \cdot \text{Decile}_p) \cdot x_t + \epsilon_{p,t}$. In the case of average returns, the regression is specified as $r_{p,t} = a + b \cdot \text{Decile}_p + \epsilon_{p,t}$. Finally, in the case of αs, the regression is specified as $\widehat{\alpha}_p + \varepsilon_{p,t} = a + b \cdot \text{Decile}_p + \epsilon_{p,t}$, with $\varepsilon_{p,t}$ reflecting time-series residuals from the respective factor model. In all cases, we report $9 \cdot b$ so that the slopes are in the same units as the spread between deciles 10 and 1. The t-statistics are obtained from

while Table 3 focuses on β_{σ^2} sorted portfolios. Panels A and B provide results over our Long and Modern sample periods, respectively.

The first half of Table 2 reports betas of β_{dp} decile-sorted portfolios relative to Δdp, $N_{\mathbb{E}}$, $r_{\mathbb{E}}$, and $r_{N\mathbb{E}}$ (labeled β_{dp}, $\beta_{N\mathbb{E}}$, $\beta_{\mathbb{E}}$, and $\beta_{rN\mathbb{E}}$) while the first half of Table 3 reports betas of β_{σ^2} decile-sorted portfolios relative to $\Delta\sigma^2$, $N_{\mathbb{V}}$, $r_{\mathbb{V}}$, and $r_{N\mathbb{V}}$ (labeled β_{σ^2}, $\beta_{N\mathbb{V}}$, $\beta_{\mathbb{V}}$, and $\beta_{rN\mathbb{V}}$). As it is clear from the tables, higher deciles generally have higher (less negative) betas. Moreover, the beta spreads between deciles 10 and 1 are all strongly significant as are the beta slopes. All these β results (which are visually displayed in Figure 3) hold over both the Long and Modern Samples, and further support the idea that sorting on β_{dp} and β_{σ^2} provides a good way to obtain ex-ante sorts on exposures to $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$.

Interestingly, the second halves of Tables 2 and 3 show that portfolios sorted on β_{dp} provide no statistically detectable spread in average returns and portfolios sorted on β_{σ^2} provide a negative spread in average returns that is statistically insignificant over the Modern Sample. These results differ from what has been reported previously in the factor model literature as previous factors are created from signals that strongly predict returns going forward.

While unusual in the factor model literature, the results in the previous paragraph are not puzzling because the ICAPM does not imply that sorting on β_{dp} or β_{σ^2} should provide a spread in average returns. It implies that each β sort should provide a spread in alphas that control for all ICAPM risk factors except the one related to the β used in the given sort (we label such alphas $\alpha_{m,\mathbb{V}}$ and $\alpha_{m,\mathbb{E}}$ where the subscript identifies the included factors). Moreover, sorting on β_{dp} or β_{σ^2} should not induce a sort on alphas constructed using the full intertemporal factor model (which we label $\alpha_{m,\mathbb{E},\mathbb{V}}$). The results in Tables 2 and 3 confirm these predictions. Sorting on β_{dp} yields a positive $\alpha_{m,\mathbb{V}}$ while sorting on β_{σ^2} yields a negative $\alpha_{m,\mathbb{E}}$, as predicted by the ICAPM. Furthermore, sorting on either β_{dp} or β_{σ^2} induces no spread in $\alpha_{m,\mathbb{E},\mathbb{V}}$ (also as predicted by the ICAPM). For completeness, we also report CAPM alphas (labeled α_m), but the ICAPM does not have a clear prediction about them.

the method in Driscoll and Kraay (1998), which is a generalization of Newey and West (1987, 1994) to panel data and accounts for autocorrelation as well as correlations across portfolios.

2.3 The ICAPM Factor Risk Premia

This subsection shows that strategies that expose investors to each ICAPM risk factor yield substantial risk premia after controlling for other ICAPM factors.

The first part of each panel in Table 4 provides correlations across the ICAPM factors as well as their (annualized) average returns, volatilities, and Sharpe ratios. We find that the ICAPM factors are highly correlated, which is expected since market prices and investment opportunities vary jointly under the ICAPM logic. For instance, $Cor(r_m, r_{\mathbb{E}}) = -0.77$, $Cor(r_m, r_{\mathbb{V}}) = -0.64$, and $Cor(r_{\mathbb{E}}, r_{\mathbb{V}}) = 0.82$ over our Long Sample.[27] In the factor model literature, it is common to design factors to be close to orthogonal. However, as we show in Internet Appendix D.1, this approach distorts risk prices and can even lead to changes their signs. In the context of our analysis, this distortion would be a problem because it would break the structural interpretation of the risk prices we estimate. For instance, we would no longer be able to show that the estimated risk prices are quantitatively reasonable under moderate risk aversion within the structural ICAPM (as we do in the next subsection).

So, instead of forcing our factors to be orthogonal, we study the risk factors jointly as dictated by our underlying ICAPM. For instance, we do not directly interpret the (relatively weak) $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ risk premia presented in the first part of each panel in Table 4. Instead, we recognize that the ICAPM implication is that $r_{\mathbb{E}}$ has a positive risk price ($r_{\mathbb{V}}$ has negative risk price), which must only translate into a positive (negative) risk premium after controlling for r_m and $r_{\mathbb{V}}$ ($r_{\mathbb{E}}$). Consequently, we compute the risk premia for strategies that are only exposed to their respective ICAPM factor, with no exposure to the other two factors. Specifically, we first construct $f_t^o = \Sigma_f^{-1} f_t$ and then apply the normalization $f_{k,t}^{\perp} = \sqrt{\mathbb{V}ar[r_m]/\mathbb{V}ar[f_k^o]} \cdot f_{k,t}^o$ so that each $f_{k,t}^{\perp}$ has the same volatility as the market portfolio, is correlated with $f_{k,t}$, and

[27] The correlations that include nontradable factors are also strong. For instance, we find $Cor(r_m, N_{\mathbb{E}}) = -0.65$, $Cor(r_m, N_{\mathbb{V}}) = -0.38$, and $Cor(N_{\mathbb{E}}, N_{\mathbb{V}}) = 0.28$ over our Long Sample. Similarly, if we use our real time proxies, we have $Cor(r_m, \Delta dp) = -0.87$, $Cor(r_m, \Delta\sigma^2) = -0.34$, and $Cor(\Delta dp, \Delta\sigma^2) = 0.31$. Note that the correlations between tradable factors tend to be stronger than the correlations that include nontradable factors. This result is expected given the level of market incompleteness and/or measurement noise in news estimation discussed in Subsection 2.1.

is uncorrelated with the other factors in f_t.[28] Note that these strategies are still correlated with each other (e.g., $Cor(f^{\perp}_{\mathbb{E},t}, f^{\perp}_{\mathbb{V},t}) \neq 0$ even though $Cor(f^{\perp}_{\mathbb{E},t}, f_{\mathbb{V},t}) = 0$). In Internet Appendix D.1, we show that $\mathbb{E}[f^{\perp}_{k,t}] \propto b_k$, which would not be the case if these strategies were constructed to be uncorrelated with each other as in the factor model literature.

The results for $f^{\perp} = (r^{\perp}_m, r^{\perp}_{\mathbb{E}}, r^{\perp}_{\mathbb{V}})$ are provided in the second part of each panel in Table 4. Each strategy is strongly correlated with its respective risk factor, uncorrelated with the other two factors, and has the same volatility as r_m. Importantly, all three strategies deliver substantial (and statistically significant) risk premia, with $\mathbb{E}[r^{\perp}_m] = 10.5\%$, $\mathbb{E}[r^{\perp}_{\mathbb{E}}] = 9.9\%$, and $\mathbb{E}[r^{\perp}_{\mathbb{V}}] = -6.6\%$ over the Long Sample.

While the analysis above shows that each factor in f_t has a strong risk premium controling for exposure to other factors in f_t, the weights used to construct $r^{\perp}_m$, $r^{\perp}_{\mathbb{E}}$, and $r^{\perp}_{\mathbb{V}}$ are obtained ex-post using the full sample estimate for Σ_f. As such, one could be worried that the results are spurious, reflecting sampling noise in the ex-post Σ_f estimate. In the third part of each panel in Table 4, we address this issue by recreating $f^{\perp} = (r^{\perp}_m, r^{\perp}_{\mathbb{E}}, r^{\perp}_{\mathbb{V}})$ based on weights that use a rolling window of 10 years (other rolling windows deliver similar results). Such real time strategies still result in strong (and statistically significant) risk premia (e.g., $\mathbb{E}[r^{\perp}_m] = 10.2\%$, $\mathbb{E}[r^{\perp}_{\mathbb{E}}] = 10.8\%$, and $\mathbb{E}[r^{\perp}_{\mathbb{V}}] = -7.9\%$ over the Long Sample), implying that Σ_f is stable and the results are not spurious as an investor can extract the ICAPM risk premia using implementable trading strategies.

The results in this subsection get to the core of how our intertemporal factor model differs from prior factor models in the literature. Factors in typical factor models have large risk premia because they inherent this properties from the anomalies they are built to explain. For example, the HML factor has large risk premium because it is created to explain the value premium, which was discovered precisely because of the large average return spread between high and low book-to-market firms. In contrast, our intertemporal risk factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) are built to mimick $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, and thus have no direct connection to anomaly-based risk

[28] To see this result, note that $\mathbb{C}ov[f^o, f] = \mathbb{C}ov[\Sigma_f^{-1} f, f] = \mathrm{I}$ with $\mathbb{V}ar[f^o] = \mathbb{V}ar[\Sigma_f^{-1} f] = \Sigma_f^{-1}$ so that $f^{\perp}_{i,t} = \sqrt{\mathbb{V}ar[r_m]/\mathbb{V}ar[f^o_k]} \cdot f^o_{k,t}$ implies $\mathbb{V}ar[f^{\perp}_k] = \mathbb{V}ar[r_m]$ for $k = m, \mathbb{E}, \mathbb{V}$.

premia that have been previously-established in the literature. The only implication from the ICAPM is that these intertemporal risk factors should have a positive and a negative risk price, respectively, which translate into positive and negative risk premia only on the strategies $r_{\mathbb{E}}^{\perp}$ and $r_{\mathbb{V}}^{\perp}$.

2.4 The ICAPM Risk Prices

The previous subsection shows strategies that expose investors to the ICAPM risk factors yield substantial risk premia. By inspecting how these strategies are constructed, one can see that $\mathbb{E}[f_k^{\perp}] \propto b_k$, where b_k is the respective factor risk price in the ICAPM SDF Equation 6. This subsection explores the bs directly to yield further insights about the ICAPM. In particular, beyond showing that the risk price signs are in line with the ICAPM predictions, we demonstrate that the ICAPM-implied risk aversion, $\gamma = 1_m^{'}\Sigma_f^{-1}\mathbb{E}[f]$, is reasonable (e.g., it is 4.8 over our Long Sample) and that removing either $r_{\mathbb{E}}$ or $r_{\mathbb{V}}$ from the model strongly distorts the risk price estimates for the remaining factors.

Table 5 estimates b by using the sample analogues of Σ_f and $\mathbb{E}[f]$ in Equation 8.[29] Panel A reports the estimated CAPM and ICAPM risk prices, b, and their t-stats. Since the bs are not easily comparable, we report $\sigma_k \cdot b_k$ for each factor f_k so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective f_k (holding other factors fixed). Panel B reports pricing errors for the strategies that expose investors to each risk factor ($r_m^{\perp}$, $r_{\mathbb{E}}^{\perp}$, and $r_{\mathbb{V}}^{\perp}$) to quantify the improvement as we move from the CAPM to the ICAPM (which must price these strategies correctly). Footnote 11 provides the ICAPM structural restrictions that allow us to pin down all risk prices based on γ as long as we have estimates for ζ_m, $\zeta_{\mathbb{E}}$, and $\zeta_{\mathbb{V}}$ (the projection coefficients associated

[29] In Internet Appendix B.2, we show that such a procedure is equivalent to a just-identified GMM estimation in which the risk prices are obtained by requiring the model to price the factors themselves (and show how to obtain standard errors for them). Moreover, we show that such estimator can be motivated from efficiency and/or robustness arguments. In terms of efficiency, adding other testing assets to the GMM estimation leaves the estimator unaffected as long as we rely on the efficient GMM weighting matrix. In terms of robustness, we show that our b estimate converges in probability to the projection of the SDF onto f even if the $M = a + b^{'}f$ model is mispecified, a result that does not hold for other b estimators.

with the ICAPM mimicking portfolios). Column "ICAPM$_\gamma$" in Table 5 provides results based on a GMM estimation of this structural (and single parameter) version of the model.[30]

The ICAPM column in Panel A of Table 5 indicates that market and reinvestment risk (r_m and $r_{\mathbb{E}}$) are always positively priced while volatility risk ($r_{\mathbb{V}}$) is always negatively priced, with these results holding whether we impose the ICAPM structural restrictions or not. That is, holding current wealth fixed, declines in expected returns and increases in expected volatility are associated with high marginal utility because they imply worse prospects for long-term investing. Moreover, the magnitudes are economically large. For instance, over our Long Sample, a one standard deviation movement in $r_{\mathbb{E}}$ induces a 0.32 change in M_t, which is substantial if we consider that $\mathbb{E}[M] = 1$.

Interestingly, the ICAPM$_{\mathbb{E}}$ column shows that $b_{\mathbb{E}}$ is much smaller when we do not control for $r_{\mathbb{V}}$ and the ICAPM$_{\mathbb{V}}$ column demonstrates that $b_{\mathbb{V}}$ is very close to zero (and statistically insignificant) when we do not control for $r_{\mathbb{E}}$. These results highlight the importance of the structural ICAPM we rely on when constructing our intertemporal risk factor model. Without a theoretical framework dictating that both $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ are necessary risk factors to implement the ICAPM, one can mistakenly conclude that the empirical evidence on the $r_{\mathbb{V}}$ risk price is "specification dependent". Internet Appendix D.10 discusses the relation between our risk price results and the prior ICAPM literature estimating the risk prices of shocks to expected returns or volatility separately (e.g., Kozak and Santosh (2020) and Ang et al. (2006)).

Beyond risk prices, the model allows us to infer the implied risk aversion, $\gamma = 1_m' \Sigma_f^{-1} \mathbb{E}[f]$. The risk aversion estimates vary from 4.8 to 6.9 depending on the sample period and whether we impose the ICAPM structure restrictions. These estimates are reasonably close to the $\overline{\gamma} = 5$ benchmark used in the structural ICAPM of Gonçalves (2021a) and the 5.1 (7.2)

[30]We estimate γ using GMM to match the $\mathbb{E}[M \cdot r] = 0$ Euler condition, with $r_m^{\perp}$, $r_{\mathbb{E}}^{\perp}$, and $r_{\mathbb{V}}^{\perp}$ as testing assets and an identity weighting matrix (which is equivalent to using r_m, $r_{\mathbb{E}}$, and $r_{\mathbb{V}}$ as test assets and specifying the weighting matrix to focus on the orthogonal component of each asset). To pin down the projection coefficients, we first note that $\zeta_m = 1$ based on the assumption that the equity market reflects the wealth portfolio (see Subsection 1.3.1). Then, we estimate $\zeta_{\mathbb{E}}$ ($\zeta_{\mathbb{V}}$) as the slope coefficient from a projection of $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) onto $r_{\mathbb{E}}$ ($r_{\mathbb{V}}$). To obtain $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, we follow Croce, Lettau, and Ludvigson (2014) and set $\delta = 0.999$ at monthly horizon ($\delta \approx 0.988$ at annual horizon).

estimate from the early period (modern period) ICAPM structural estimation in Campbell et al. (2018). It is also interesting to note that the CAPM implies even lower risk aversion, with estimates ranging from 2.3 to 2.8. The reason is that accounting for variation in investment opportunities tends to lower the risk of the market portfolio (i.e., equities are safer in the long-run), and thus a higher risk aversion is required in the ICAPM (relative to the CAPM) to justify the equity premium we observe in the data.

The results in Panel B of Table 5 indicate that the CAPM yields high pricing errors for the ICAPM risk factors. Specifically, the CAPM can explain about half of the $r_m^{\perp}$ risk premium and none of the $r_{\mathbb{E}}^{\perp}$ and $r_{\mathbb{V}}^{\perp}$ risk premia (since $r_{\mathbb{E}}^{\perp}$ and $r_{\mathbb{V}}^{\perp}$ are uncorrelated with r_m). Adding either $r_{\mathbb{E}}$ or $r_{\mathbb{V}}$ provides little improvement to the pricing errors while adding both reduces all pricing errors to zero (by construction). We also find that the ICAPM prices $r_m^{\perp}$, $r_{\mathbb{E}}^{\perp}$, and $r_{\mathbb{V}}^{\perp}$ reasonably well even when the model is restricted so that γ pins down all three risk prices (i.e., when the model only has one degree of freedom as in the CAPM). While some αs are statistically significant, they are still much lower in magnitude than under the CAPM. This result indicates that the unconstrained risk prices we focus on are reasonably in line with the implications of a fully structural ICAPM.[31]

2.5 Implications and Limitations of our Main Results

We find that the tradable factors in our intertemporal risk factor model are priced and reflect relevant risks for long-term investing. In our view, these results represent an important step forward to the factor model literature given the lack of risk-related interpretability of

[31]This result may seem surprising given that prior literature (e.g., Chacko and Viceira (2005)) finds that news to long-run variance induces a quantitatively small hedging demand when conditional variance is modeled as a univariate autoregressive process. The crux of the matter is that, as explained in Subsection 1.3.4, we estimate our log-GARCH process by targeting long-run (10-year) realized variance (Internet Appendix D.6 shows that estimating the log-GARCH by maximum likelihood, which effectively targets short-term variance, yields results that are in line with the prior literature). The underlying idea is that targeting long-run expectations provides a robust estimation method if we recognize that any autoregressive model is likely mispecified (see Ederington and Guan (2010) and Jordà and Kozicki (2011)). In Internet Appendix D.5, we further show that our baseline log-GARCH estimation provides better long-run variance forecasts than vector autoregressions that rely on multiple state variables, a result that does not hold if we estimate the log-GARCH by maximum likelihood.

standard factor models (Kozak, Nagel, and Santosh (2018)). Moreover, such a step is crucial to help us move beyond the CAPM when performing risk adjustments in practice.

However, one must be careful when interpreting our results since we model the decision of a single investor (or class of investors) without taking a stand on aggregation or the representative agent in the economy. In a general equilibrium model with heterogeneous agents, the first order conditions of all investors must hold, and thus we construct our intertemporal factor model based on the SDF of a buy-and-hold long-term rational investor. We take this approach because it is sufficient for the purpose of creating a model with tradable factors that reflect the risks associated with long-term investing (i.e., market and intertemporal risk), which are key determinants of risk premia in consumption-based asset pricing models (see Section IV in Gonçalves (2021a)) or even if investors focus on long-run mean-variance analysis (see Cochrane (2014)). A limitation of our approach, however, is that it is partial equilibrium in nature, and thus cannot fully determine the causes of variation in asset prices.

Relatedly, Kozak, Nagel, and Santosh (2018) argue that their result that factor models lack economic interpretability also applies (albeit to a lesser degree) to the ICAPM when investment opportunities vary exogenously because they may be driven by sentiment. This is another limitation of our work as we do not take a stand on what drives variation in expected returns and volatility. While Internet Appendix D.8 shows that our main results are basically identical after controlling for the sentiment index of Baker and Wurgler (2006), it remains possible that other sources of sentiment drive variation in investment opportunities. However, even if sentiment fully drives investment opportunities, our results still shed light on the risks affecting asset prices as we uncover risk factors that dissuade rational long-term investors from fully exploiting the opportunities created by sentiment-driven investors.

3 Comparing the ICAPM with Other Factor Models

The previous section shows that the ICAPM risk factors properly capture market and reinvestment risk and are priced consistently with the ICAPM predictions. In this section, we compare the ICAPM with eight prominent factor models proposed in the literature. Specifically, we consider (ordered by publication year): the Sharpe (1964) CAPM, the Fama and French (1993) 3-Factor model (FF3), the Carhart (1997) 4-Factor model (FFC4), the Fama and French (2015) 5-Factor model (FF5), the Hou, Xue, and Zhang (2015) 4-Factor model (q4), the Stambaugh and Yuan (2017) 4-Factor model (SY4), the Daniel, Hirshleifer, and Sun (2020) 3-Factor Model (DHS3), and the Hou et al. (2020) 5-Factor model (q5).

The rest of this section is organized as follows. Subsection 3.1 estimates the ICAPM risk prices after controlling for other factors, Subsection 3.2 compares the ICAPM with other factor models based on their implied tangency portfolio Sharpe ratios, and Subsection 3.3 extends the comparison to pricing errors on the testing assets recommended by Lewellen, Nagel, and Shanken (2010): single stocks, industry portfolios, correlation-clustered portfolios (Ahn, Conrad, and Dittmar (2009)), and bond portfolios. Internet Appendices B and C provide detailed descriptions of the econometric procedures we rely on and the data sources we use, respectively.

3.1 ICAPM Risk Prices Controlling for Other Factors

This subsection estimates the ICAPM risk prices controlling for other factors. Specifically, we estimate the SDF projection $M_t = a - b' f_t - b_x' x_t$ with x_t reflecting the factors in each of the factor models mentioned above, with the exception that we do not add the market factor to x_t since it is already included in f_t. These SDF projections are analogous to the typical factor spanning tests in the literature, with the added advantage that SDF projections control not only for x_t but also for other f_t factors when testing each f_t factor, which is important in the context of the ICAPM (see Internet Appendix B.3).

For the Modern Sample (1973-2019), we can construct all factors, and thus our estimation is analogous to the one used in Table 5. For the Long Sample (1928-2019), we can only construct the ICAPM risk factors and the factors in FFC4, which are known as SMB, HML, and MOM. However, since the risk price estimates depend only on estimates for Σ_f and $\mathbb{E}[f]$, we are still able to estimate risk prices using the method in Stambaugh (1997), which allows for factors with different time-series lengths in the estimation process.[32]

The ICAPM risk prices, annualized risk premia, and t-statistics after controlling for other factors are reported in Table 6. The ICAPM risk prices remain strong and statistically significant regardless of which factors we control for. The only exception is that, over the Modern Sample, the economically large $b_{\mathbb{E}} = 0.16$ risk price becomes statistically insignificant when controlling for the q5 factors. However, this result seems to be mostly driven by sample size since $b_{\mathbb{E}}$ remains statistically significant in the Long Sample even after controlling for the q5 factors.

In Internet Appendices D.8 and D.9, we estimate SDF projections similar to the ones in Table 6 to control for liquidity (Pástor and Stambaugh (2003)), sentiment (Baker and Wurgler (2006)), and betting against beta (Frazzini and Pedersen (2014)). The risk prices on the ICAPM factors remain strong and statistically significant.

3.2 ICAPM vs Other Factor Models: Maximum Sharpe Ratios

In this subsection, we compare the tangency portfolio Sharpe ratios (i.e., the "maximum Sharpe ratios") of different factor models. Based on Barillas and Shanken (2017), such analysis is sufficient for comparing the pricing ability of different factor models.[33]

Table 7 Panel A shows the (annualized) maximum Sharpe ratios, SR_{max}, of the different

[32]In all cases over the Long Sample, we extend the factors to their first date available and then apply the Stambaugh (1997)'s estimation procedure. For instance, for the q5 model, factor data is available starting in 1967. We then use this factor data and apply the estimation in Stambaugh (1997) to get the Long Sample estimates for b, treating the 1928-2019 period as the "long history" and the 1967-2019 period as the "short history" as per Stambaugh (1997)'s terminology. We always treat the ICAPM and FFC4 factors as available over the "long history" regardless of which SDF projection we are estimating. As such, in the case of the comparison between the ICAPM and the FFC4 model (as well as the FF3 model), the procedure is equivalent to the analysis performed over our Modern Sample. We use bootstrap standard errors to obtain t-statistics for our Long Sample estimates that rely on the method in Stambaugh (1997).

[33]A simplified version of the Barillas and Shanken (2017) argument can be seen directly from the Gibbons, Ross, and Shanken (1989) (GRS) test statistic. Specifically, the GRS statistic is given by $SR^2(f,R) - SR^2(f) = \alpha^{'}\Sigma^{-1}\alpha$, where $SR^2(f)$ is the Sharpe ratio of the tangency portfolio formed with f. If we include all existing assets in the set of testing assets, R, then $SR^2(f,R) = \overline{SR}^2$ is the same for any set of factors we include in f. Therefore, the factor model with the highest $SR^2(f)$ is also the model with the best pricing ability according to the GRS statistic. As such, it is sufficient to compare models based on their maximum Sharpe ratios (i.e., to compare $SR(f)$ for different f) in order to rank their pricing abilities.

factor models, with parentheses reporting the percent of bootstrap simulations over which the ICAPM has higher SR_{max} than the given model. The first row shows in-sample SR_{max} over the entire Modern Sample with models ordered by publication year. Interestingly, SR_{max} increases monotonically with the publication year, with the q5 model displaying the highest in-sample $SR_{max} = 2.10$. In contrast, the ICAPM's $SR_{max} = 0.79$ is only higher than the SR_{max} of the CAPM and the FF3 model.

However, many of these in-sample SR_{max} are too high to be real and effectively represent "near-arbitrage" opportunities (a term from Kozak, Nagel, and Santosh (2018)). For instance, Ross (1976) and Cochrane and Saá-Requejo (2000) argue that any SR_{max} higher than $\sqrt{2} \cdot SR_{CAPM}$ and $2 \cdot SR_{CAPM}$, respectively, is unlikely to be truly available to investors in financial markets (see also Shanken (1992) and MacKinlay (1995)). In contrast, all factor models we explore except the FF3 and the ICAPM have $SR_{max} > 2 \cdot SR_{CAPM}$ in-sample.

As pointed out by Fama and French (2018) and Kan, Wang, and Zheng (2019), one of the issues with in-sample SR_{max} is that it is subject to overfitting since the weights of the tangency portfolio are chosen over the same period over which SR_{max} is calculated. To address this issue, we follow Kan, Wang, and Zheng (2019) and break the Modern Sample into two periods, with the 1st half representing the period from January/1973 to June/1995 and the 2nd half the period from July/1996 to December/2019 (the respective in-sample Sharpe ratios are also provided in the table). We then estimate the SR_{max} weights over the 1st half and obtain our out-of-sample SR_{max} over second half. The results indicate that the ICAPM's SR_{max} is higher than the SR_{max} of all models except SY4, DHS3, and q5.

The out-of-sample analysis above restricts the weights to be estimated over the first half of the Modern Sample. However, some factor models (such as the ICAPM) have data going back further, which allows investors to obtain weights in real time using a much longer sample. As such, we also provide the out-of-sample SR_{max} of each model after allowing the weights to be estimated from all data available up to June/1995. The results get stronger for the ICAPM, with little changes for other models so that the ICAPM's SR_{max} remains higher than the SR_{max} of all models except SY4, DHS3, and q5.

While the analysis above considers the SR_{max} overfitting problem, there is another issue related to differences in implementation costs across different factors. Specifically, each factor is constructed following a different procedure. For instance, the FF3 factors are based on 30% and 70% breakpoints while the DHS3 factors rely on 20% and 80% breakpoints. In the absence of theory, there is no reason to restrict a given factor to a given set of construction rules, but the implementation differences can generate substantial differences in trading costs, which can have a strong effect on Sharpe ratios. To address this issue, we follow Detzel, Novy-Marx, and Velikov (2020) and also report, in Table 7 Panel B, a SR_{max} analysis that relies on net (of trading costs) factor returns.[34] The results are qualitatively similar to Panel A, but quantitatively stronger for the ICAPM, indicating that the trading costs in implementing the ICAPM are smaller than for other models.

The core results from Table 7 are provided visually in Figure 4. In summary, we find that the ICAPM performs well relative to other factor models in terms of its maximum Sharpe ratio despite its factors being constrained to reflect risks that theoretically matter for long-term investing. In fact, only three out of the eight factor models studied have SR_{max} that are higher than the ICAPM SR_{max} once we account for overfitting. Moreover, the ICAPM SR_{max} is stable whether estimated in-sample or out-of-sample and whether we account for trading costs or not, which is not the case for the other multifactor models we study. The ICAPM SR_{max} stability highlights the importance of relying on theory to overcome the natural publication biases that arise in the context of the search for the multifactor model that will replace the CAPM in the years to come.

[34] We follow the same procedure as Detzel, Novy-Marx, and Velikov (2020), except that we use the trading cost measure of Chen and Velikov (2020) for the adjustment (and thank Andrew Chen for sharing the data). As Chen and Velikov (2020) demonstrate, their high frequency trading cost measure more accurately reflects trading costs and implies, on average, lower trading costs than the main alternative measures used in the literature. This choice is conservative from our perspective because, as we show, the ICAPM is less affected by trading costs than the other factor models we consider. This result is a consequence of our factor construction relying on single sorts whereas other models use double or triple sorts, which overweight small stocks.

3.3 ICAPM vs Other Factor Models: Pricing Errors

While focusing on SR_{max} is sufficient for comparing factor models in a world without publication bias, there are important limitations of a SR_{max} analysis when we consider that the publication prospects of proposed factor models likely correlate with the Sharpe ratios of the proposed factors. This issue can be seen directly from Table 7 as SR_{max} increases monotonically with the model's publication year. The out-of-sample analysis in the previous section deals with overfitting but not with publication bias because the factors are still based on a publication process that relied on Sharpe ratios over the period we treat as "out-of-sample." One approach is to wait for several years to perform a truly out-of-sample SR_{max} analysis.

This subsection considers an imperfect, but still useful, alternative solution. Namely, we compare the ICAPM to previous factor models based on the pricing of testing assets that the original studies did not consider. For this task, we focus on the testing assets recommended by Lewellen, Nagel, and Shanken (2010): single stocks, industry portfolios, correlation-clustered portfolios (Ahn, Conrad, and Dittmar (2009)), and bond portfolios. This choice alleviates concerns associated with publication bias in testing assets (Harvey (2017) and Lo and MacKinlay (1990)) and with testing assets that are formed from signals closely related to the factors themselves (e.g., Ferson, Sarkissian, and Simin (1999) and Kogan and Tian (2015)), which would be a problem if we compared models based on, for example, well-known anomalies or the twenty decile portfolios we study in Tables 2 and 3 (but we also provide results for these testing assets in Section 4). When analysing testing assets, we focus on the Modern Sample as it reflects the longest period for which we can construct all factors we explore.

3.3.1 Single Stocks

At each month t, we select all CRSP common stocks of firms incorporated in the United States (shrcd = 10 or 11) that trade on NYSE, AMEX, or NASDAQ (exchcd = 1,2 or 3) and have all returns available over the last five years. We then estimate their pricing errors (αs) over this five-year window based on the usual factor regressions and calculate the squared sum of pricing errors, $\Sigma\alpha^2$.

Table 8 Panel A reports, for each factor model, the time-series averages of these $\Sigma\alpha^2$ normalized by the respective sums of pricing errors computed under risk-neutral pricing ($\alpha_{RN} = \mathbb{E}[r]$) and the ICAPM ($\alpha_{ICAPM}$). We also report the percent of months for which the respective sum is lower under the ICAPM than under the given factor model (e.g., $\%(\Sigma\alpha^2_{ICAPM}<\Sigma\alpha^2)$), which incorporates sampling variability in a manner similar to Fama and MacBeth (1973) regressions. The overall results indicate that the ICAPM produces the smallest pricing errors and the differences are particularly pronounced when comparing the ICAPM with models that were published more recently, such as SY4, DHS3, and q5.

3.3.2 Industry Portfolios

For this analysis, we focus on the Fama and French (1997)'s 30 industry portfolios (following Lewellen, Nagel, and Shanken (2010)). In contrast to single stocks, we have a balanced panel with industry portfolios, and thus calculate one α per portfolio, reporting the same $\Sigma\alpha^2$ (with the same normalizations) as we do with single stocks, but with no time-series average being required. The percent of samples for which the ICAPM has lower pricing errors than each model come from bootstrap simulations in this case.

Table 8 Panel B clearly indicates that the ICAPM produces the smallest pricing errors and the differences are very high, with the ICAPM always outperforming other models in more than 75% of the bootstrap samples.

3.3.3 Correlation-Clustered Portfolios

We follow Ahn, Conrad, and Dittmar (2009) in constructing 10 correlation-clustered portfolios. Specifically, at each time t, we calculate return correlations (on a five year rolling window) between all pairs of stocks used in our analysis of single stocks. We then obtain the distance between stocks i and j as $d_{i,j} = \sqrt{2 \cdot (1 - Cor_{i,j})}$ and use these distance measures in a hierarchical clustering analysis, applying the Ward's minimum variance method to identity groups. Finally, we combine clusters formed at different points in time by assuring that, across adjacent months, each cluster portfolio has the most consistent firm membership possible. As in Ahn, Conrad, and Dittmar (2009), our correlation-clustered portfolios provide a large spread in average returns (as in typical anomaly sorts) and low correlation across portfolios (in contrast to typical anomaly sorts). Further details are provided in Ahn, Conrad, and Dittmar (2009).

Table 8 Panel C reports the results for these correlation-clustered portfolios in the same format as Panel B. The two best performing factor models in this analysis are the FF3 and FFC4. The ICAPM comes next and performs substantially better than all other factor models. For instance, the next best performing model is the CAPM and the ICAPM performs better than it in at least 60% of the bootstrap samples.

3.3.4 Treasury Bond Portfolios

For this analysis, we rely on the Fama bond portfolios available in CRSP, which reflect Treasury bond portfolios with bond maturities up to $h = 1, 2, 3, 4, 5, 10, 30$ years.

Table 8 Panel D reports the results for these Treasury bond portfolios in the same format as Panels B and C. The best performing model is unambiguously the DHS3, with the ICAPM being the next best performing model. Moreover, all other models perform substantially worse than the ICAPM, with the next best performing model being the q5, which performs worse than the ICAPM in at least 58% of the bootstrap samples.

3.3.5 Summarizing α Results

Figure 5 provides a visual summary of the main findings for all four groups of testing assets. For each group, we also create an alternative definition of portfolios to demonstrate that the results are not sensitive to auxiliary empirical decisions. Specifically, we consider a ten-year rolling window for single stocks, the Fama and French (1997)'s 48 industry portfolios, 25 correlation-clustered portfolios, and alternative CRSP US Treasury Indexes that reflect Treasury bond portfolios with bond maturities up to $h = 1, 2, 5, 7, 10, 20, 30$ years.

In summary, we find that the ICAPM is always among the best performing models (in terms of lowest pricing errors) and is the only factor model to consistently do so across all four types of testing assets. Table 9 further extends this statement by ranking models based on normalized $\Sigma\alpha^2$ and also $\Sigma|\alpha|$ for each set of testing assets we study in this section. Regardless of whether we focus on $\Sigma\alpha^2$ or $\Sigma|\alpha|$, the ICAPM is the model with the best average rank among all models we consider.

4 Anomalies

Much of the empirical asset pricing literature has focused on anomalies (i.e., strategies based on signals that were originally proposed as a puzzle to a benchmark asset pricing model, often the CAPM). This fact creates an important publication bias when using anomalies to test a given model or compare across different models (see Lo and MacKinlay (1990)). Such effect is particularly pronounced when some of the factors used to explain anomalies are themselves created from anomaly signals (Ferson, Sarkissian, and Simin (1999) and Kogan and Tian (2015)). Nevertheless, this section studies anomalies for completeness. Subsection 4.1 focuses on a comparison across factor models while Subsection 4.2 asks whether investing in anomalies provides an ex-ante increase in the ICAPM tangency portfolio Sharpe ratio.

4.1 ICAPM vs Other Factor Models: Anomaly Deciles

Table 10, Panel A, provides model comparison results using 158 anomaly decile portfolios (that are value-weighted and based on NYSE breakpoints) from the data made available by Chen and Zimmermann (2020), which gives us a total of 1,580 portfolios.[35] The ICAPM performs better than the CAPM, FF3, and FF5 models in pricing anomalies, but worse than all other models we analyse. The DHS3 and the q5 models are the two best performing models, with the ICAPM performing better than these models in only 8% of the simulations

[35] We begin with the 180 "clear predictors" from Chen and Zimmermann (2020), which reflect anomalies that they classify as being "clearly significant in the original papers". From these 180 significant anomalies, we remove anomalies that do not have return records for all 10 decile portfolios for at least half of our 1973-2019 sample. This procedure yields the 158 anomalies (and the corresponding 1,580 decile portfolios) we explore in Table 10. The data is available at `https://github.com/OpenSourceAP/CrossSection`.

based on $\Sigma\alpha^2$ and in no more than 41% of the simulations based on $\Sigma|\alpha|$.

As previously pointed out, the signals used in the construction of traditional factor models are, by design, strongly connected (and sometimes identical) to the signals used in the construction of anomaly portfolios, which creates important issues when studying factor models (Ferson, Sarkissian, and Simin (1999) and Kogan and Tian (2015)). Table 10, Panel B, gives a rough indication of the problem. It replaces the deciles formed on the 158 anomalies in Panel A with the deciles formed on β_{dp} and β_{σ^2} (studied in Tables 2 and 3). The results indicate the ICAPM is by far the best model in pricing these testing assets. Of course, it would be misleading to conclude that the ICAPM is the best model based on such an analysis because the ICAPM factors are constructed using β_{dp} and β_{σ^2} as signals. The same logic (but in reverse) plagues the analysis in Panel A.

4.2 Can we Increase the ICAPM Tangency Portfolio Sharpe Ratio?

Since the ICAPM performs worse than several other factor models in pricing anomaly deciles, we now ask whether an investor would benefit from adding anomalies to a portfolio that invests in r_m, $r_{\mathbb{E}}$, and $r_{\mathbb{V}}$. We find that this is not necessarily the case. Specifically, we show that it is hard for investors to increase the tangency portfolio Sharpe ratio ex-ante by adding anomalies to the ICAPM factors. Consequently, even if one ignores trading costs and publication biases, it seems investors cannot easily trade on anomalies to improve upon the ICAPM. As such, an investor who ignores anomalies may not be suffering a large utility loss.

To motivate the analysis, note that the result in Gibbons, Ross, and Shanken (1989) implies $\alpha_j^2 = \sigma_{j,\epsilon}^2 \cdot (SR_{max,j}^2 - SR_{max}^2)$, where $\sigma_{j,\epsilon}^2 > 0$ reflects the variance of residuals in a factor regression of portfolio j onto the ICAPM factors. As such, a non-zero α is puzzling because it implies an investor can increase the ICAPM tangency portfolio Sharpe ratio (ex-post) from SR_{max} to $SR_{max,j}$. The black lines in Figure 6 show the distribution of annualized (in-sample estimated) $\Delta SR_j = SR_{max,j} - SR_{max}$ across the 158 anomalies we study. Figure 6(a) considers long-short portfolios based on the three highest and three lowest deciles for each anomaly, which is in line with how the intertemporal factors are created. Figure 6(b) repeats the analysis with long-short portfolios that consider only the highest and lowest deciles for each anomaly. The in-sample results indicate investors can achieve large increases in the ICAPM tangency portfolio Sharpe ratio by trading on anomalies with no risk of decreasing the tangency portfolio Sharpe ratio.

However, as emphasized in Cederburg et al. (2020), the tangency portfolio weights in the black lines (which are implicitly used in αs) are estimated in-sample. To address this issue, we divide our Modern Sample in two periods of equal length, estimating the weights (for both $SR_{max,j}$ and SR_{max}) in the first half and calculating the Sharpe ratios in the second half, which is analogous to how Kan, Wang, and Zheng (2019) (and our Subsection 3.2) compare the tangency Sharpe ratios of different factor models. The blue lines in Figure 6 show the distribution of these out-of-sample annualized ΔSRs. Clearly, the risk of ultimately decreasing the tangency Sharpe ratio by adding an anomaly strategy to the ICAPM tangency portfolio is high (36.7% for both types of anomaly strategies we study). Of course, a real investor would have access not only to the first half of our Modern Sample, but also to the period before 1973. As such, the red lines show the analogous distributions when weights are estimated using all data back to 1928, with the method in Stambaugh (1997) used to estimate weights when the data for a given anomaly starts later than 1928 (as in Subsection 3.1). In this case, the risk of decreasing the tangency Sharpe ratio by adding an anomaly strategy is even higher (49.4% and 42.4% for the two types of anomaly strategies we study). So, investors with access to r_m, $r_{\mathbb{E}}$, and $r_{\mathbb{V}}$ have a high probability of accidentally decreasing their Sharpe ratios as they add anomalies to their portfolios.

5 Conclusion

In this paper, we show that a long-term Bayesian investor perceives shocks to the market dividend yield and realized variance as additional risk factors beyond the market portfolio, and use this insight to construct an intertemporal factor model in which the tradable factors capture market risk (r_m), reinvestment risk ($r_{\mathbb{E}}$), and volatility risk ($r_{\mathbb{V}}$). We then project the SDF onto our tradable factors and find that their risk prices are consistent with an underlying structural ICAPM. Finally, we demonstrate that our intertemporal factor model performs well relative to prominent factor models in several empirical tests.

Our results show that it is fruitful to build models with tradable factors that truly reflect risks (i.e., that are designed to mimick the risks of some underlying theory). Such risk factor models are more informative about investors' motives than traditional factor models, which are not directly linked to investors' preferences and/or beliefs (Kozak, Nagel, and Santosh (2018)). At the same time, risk factor models allow for the identification of risk factors that matter to investors without being subject to the important sensitivity issues associated with estimating the risk prices of non-tradable factors (Lewellen, Nagel, and Shanken (2010)).

We hope the future literature continues to tackle the important task of building risk factor models (or sentiment factor models) closely guided by theory. This can be done by extending the ICAPM to incorporate issues we abstract from such as adding other assets to the wealth portfolio (Cederburg and O'Doherty (2019)) or by implementing factor models that reflect other economic frameworks such as intermediary asset pricing (Adrian, Etula, and Muir (2014)). The ultimate objective of risk factor models must not be to fully explain the cross-section of returns, but rather to understand how much of the cross-section is explained by particular risk-based frameworks. As such, the important unifying theme is that future work should strive to demonstrate that the proposed tradable factors capture the underlying risks for which they intend to proxy.

References

Adrian, T., E. Etula, and T. Muir (2014). "Financial Intermediaries and the Cross-Section of Asset Returns". In: *Journal of Finance* 69.6, pp. 2557–2596.

Ahn, D.-H., J. Conrad, and R. F. Dittmar (2009). "Basis Assets". In: *Review of Financial Studies* 22.12, pp. 5133–5174.

Alekseev, G., S. Giglio, Q. Maingi, J. Selgrad, and J. Stroebel (2021). "A Quantity-based approach to constructing climate risk hedge portfolios". Working Paper.

Anderson, E. W., E. Ghysels, and J. L. Juergens (2009). "The impact of risk and uncertainty on expected returns". In: *Journal of Financial Economics* 94, pp. 233–263.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). "The Cross-Section of Volatility and Expected Returns". In: *Journal of Finance* 61.1, pp. 259–299.

Ang, A., J. Liu, and K. Schwarz (2020). "Using Stocks or Portfolios in Tests of Factor Models". In: *Journal of Financial and Quantitative Analysis* 55.3, pp. 709–750.

Baker, M. and J. Wurgler (2006). "Investor Sentiment and the Cross-Section of Stock Returns". In: *Journal of Finance* 61.4.

Bansal, R., D. Kiku, I. Shaliastovich, and A. Yaron (2014). "Volatility, the Macroeconomy, and Asset Prices". In: *Journal of Finance* 69.6, pp. 2471–2511.

Barillas, F. and J. Shanken (2017). "Which Alpha?" In: *Review of Financial Studies* 30.4, pp. 1316–1338.

Barroso, P., M. Boons, and P. Karehnke (2020). "Time-varying state variable risk premia in the ICAPM". In: *Journal of Financial Economics* Forthcoming.

Berk, J. B. (2000). "Sorting out sorts". In: *Journal of Finance* 55.1, pp. 407–427.

Binsbergen, J. H. v. and R. S. J. Koijen (2010). "Predictive Regressions: A Present-Value Approach". In: *Journal of Finance* 65.4, pp. 1439–1471.

Boons, M. (2016). "State variables, macroeconomic activity, and the cross section of individual stocks". In: *Journal of Financial Economics* 119, pp. 489–511.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989). "Empirical Test of the Consumption-Oriented CAPM". In: *Journal of Finance* 44.2, pp. 231–262.

Brennan, M. J., A. W. Wang, and Y. Xia (2004). "Estimation and Test of a Simple Model of Intertemporal Capital Asset Pricing". In: *Journal of Finance* 59.4, pp. 1743–1775.

Campbell, J. Y. (1993). "Intertemporal Asset Pricing without Comsumption Data". In: *American Economic Review* 83.3, pp. 487–512.

Campbell, J. Y. (1996). “Understanding Risk and Return”. In: *Journal of Political Economy* 104.2, pp. 298–345.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley (2018). “An Intertemporal CAPM with Stochastic Volatility”. In: *Journal of Financial Economics* 128.2, pp. 207–233.

Campbell, J. Y. and R. J. Shiller (1989). “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors”. In: *Review of Financial Studies* 1.3, pp. 195–228.

Campbell, J. Y. and T. Vuolteenaho (2004). “Bad Beta, Good Beta”. In: *American Economic Review* 94.5, pp. 1249–1275.

Carhart, M. M. (1997). “On Persistence in Mutual Fund Performance”. In: *Journal of Finance* 52.1, pp. 57–82.

Cederburg, S. (2019). “Pricing intertemporal risk when investment opportunities are unobservable”. In: *Journal of Financial and Quantitative Analysis* 54.4, pp. 1759–1789.

Cederburg, S. and M. S. O’Doherty (2015). “Asset-pricing anomalies at the firm level”. In: *Journal of Econometrics* 186, pp. 113–128.

Cederburg, S. and M. S. O’Doherty (2019). “Understanding the Risk-Return Relation: The Aggregate Wealth Proxy Actually Matters”. In: *Journal of Business & Economic Statistics* 37.4, pp. 721–735.

Cederburg, S., M. S. O’Doherty, F. Wang, and X. S. Yan (2020). “On the Performance of Volatility-Managed Portfolios”. In: *Journal of Financial Economics* 138.1, pp. 95–117.

Chacko, G. and L. M. Viceira (2005). “Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets”. In: *Review of Financial Studies* 18.4, pp. 1369–1402.

Chen, A. Y. and M. Velikov (2020). “Zeroing in on the Expected Returns of Anomalies”. Working Paper.

Chen, A. Y. and T. Zimmermann (2020). “Open Source Cross-Sectional Asset Pricing”. Working Paper.

Chen, L. and X. Zhao (2009). “Return Decomposition”. In: *Review of Financial Studies* 22.12, pp. 5213–5249.

Cochrane, J. H. (2005). *Asset Pricing*. Revised Edition. Princeton University Press.

Cochrane, J. H. (2008). “Financial Markets and the Real Economy”. In: *Handbook of the Equity Premium*. Ed. by R. Mehra. 1st ed. Elsevier Science. Chap. 7, pp. 237–330.

Cochrane, J. H. (2014). “A Mean-Variance Benchmark for Intertemporal Portfolio Theory”. In: *Journal of Finance* 69.1, pp. 1–49.

Cochrane, J. H. and J. Saá-Requejo (2000). “Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets”. In: *Journal of Political Economy* 108.1, pp. 79–119.

Cooper, I. and P. Maio (2019). “New Evidence on Conditional Factor Models”. In: *Journal of Financial and Quantitative Analysis* 54.5, pp. 1975–2016.

Croce, M. M., M. Lettau, and S. C. Ludvigson (2014). “Investor Information, Long-Run Risk, and the Term Structure of Equity”. In: *Review of Financial Studies* 28.3, pp. 706–742.

Daniel, K., D. Hirshleifer, and L. Sun (2020). “Short- and Long-Horizon Behavioral Factors”. In: *Review of Financial Studies* 4, pp. 1673–1736.

Detzel, A., R. Novy-Marx, and M. Velikov (2020). “Model Selection with Transaction Costs”. Working Paper.

Driscoll, J. C. and A. C. Kraay (1998). “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data”. In: *Review of Economics and Statistics* 80.4, pp. 549–560.

Ederington, L. H. and W. Guan (2010). “Longer-Term Time-Series Volatility Forecasts”. In: *Journal of Financial and Quantitative Analysis* 45.4, pp. 1055–1076.

Epstein, L. G. and S. E. Zin (1989). “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework”. In: *Econometrica* 57.4, pp. 937–969.

Fama, E. F. (1991). “Efficient Capital Markets: II”. In: *Journal of Finance* 46.5, pp. 1575–1617.

Fama, E. F. and K. R. French (1993). “Common Risk Factors in the Returns on Stocks and Bonds”. In: *Journal of Financial Economics* 33, pp. 3–56.

Fama, E. F. and K. R. French (1997). “Industry Costs of Equity”. In: *Journal of Financial Economics* 43.2, pp. 153–193.

Fama, E. F. and K. R. French (2015). “A five-factor asset pricing model”. In: *Journal of Financial Economics* 116, pp. 1–22.

Fama, E. F. and K. R. French (2018). “Choosing Factors”. In: *Journal of Financial Economics* 128, pp. 234–252.

Fama, E. F. and J. D. MacBeth (1973). “Risk, Return and Equilibrium: Empirical Tests”. In: *Journal of Political Economy* 81.3, pp. 607–636.

Ferson, W. E., S. Sarkissian, and T. Simin (1999). “The alpha factor asset pricing model: A parable”. In: *Journal of Financial Markets* 2.1, pp. 49–68.

Foster, D. P. and D. B. Nelson (1996). “Continuous record asymptotics for rollingsample variance estimators”. In: *Econometrica* 64.1, pp. 139–174.

Frazzini, A. and L. H. Pedersen (2014). “Betting Against Beta”. In: *Journal of Financial Economics* 111.1, pp. 1–25.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). “A Test of the Efficiency of a Given Portfolio”. In: *Econometrica* 57.5, pp. 1121–1152.

Giglio, S. and D. Xiu (2020). “Asset pricing with omitted factors”. In: *Journal of Political Economy* forthcoming.

Gonçalves, A. S. (2021a). “Reinvestment Risk and the Equity Term Structure”. In: *Journal of Finance* 76.5, pp. 2153–2197.

Gonçalves, A. S. (2021b). “The Short Duration Premium”. In: *Journal of Financial Economics* 141.3, pp. 919–945.

Grauer, R. R. and J. A. Janmaat (2004). “The unintended consequences of grouping in tests of asset pricing models”. In: *Journal of Banking and Finance* 28, pp. 2889–2914.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). “Realized GARCH: A joint Model for Returns and Realized Measures of Volatility”. In: *Journal of Applied Econometrics* 27, pp. 877–906.

Harvey, C. R. (2017). “Presidential Address: The Scientific Outlook in Financial Economics”. In: *Journal of Finance* 72.4, pp. 1399–1440.

Herskovic, B., A. Moreira, and T. Muir (2019). “Hedging Risk Factors”. Working Paper.

Hou, K., H. Mo, C. Xue, and L. Zhang (2020). “An Augmented q-Factor Model with Expected Growth”. In: *Review of Finance* Forthcoming.

Hou, K., C. Xue, and L. Zhang (2015). “Digesting Anomalies: An Investment Approach”. In: *Review of Financial Studies* 28.3, pp. 650–705.

Jagannathan, R. and Y. Wang (2007). “Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns”. In: *Journal of Finance* 62.4, pp. 1623–1661.

Jordà, Ò. and S. Kozicki (2011). “Estimation and Inference by the Method of Projection Minimum Distance: An Application to the New Keynesian Hybrid Phillips Curve”. In: *International Economic Review* 52.2, pp. 461–487.

Kan, R., C. Robotti, and J. Shanken (2013). “Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology”. In: *Journal of Finance* 68.6, pp. 2617–2649.

Kan, R., X. Wang, and X. Zheng (2019). “In-sample and Out-of-sample Sharpe Ratios of Multi-factor Asset Pricing Models”. Working Paper.

Kogan, L. and M. Tian (2015). “Firm Characteristics and Empirical Factor Models: a Model-Mining Experiment”. Working Paper.

Kozak, S., S. Nagel, and S. Santosh (2018). “Interpreting Factor Models”. In: *Journal of Finance* 73.3, pp. 1183–1223.

Kozak, S. and S. Santosh (2020). “Why do discount rates vary?” In: *Journal of Financial Economics* 137.3, pp. 740–751.

Kroencke, T. A. (2017). “Asset pricing without garbage”. In: *The Journal of Finance* 72.1, pp. 47–98.

Laurinaityte, Meinerding, Schlag, and Thimme (2020). “GMM Weighting Matrices in Cross-Sectional Asset Pricing Tests”. Working Paper.

Lewellen, J., S. Nagel, and J. Shanken (2010). “A Skeptical Appraisal of Asset Pricing Tests”. In: *Journal of Financial Economics* 96.2, pp. 175–194.

Lo, A. W. and A. C. MacKinlay (1990). "Data-Snooping Biases in Tests of Financial Asset Pricing Models". In: *Review of Financial Studies* 3.3, pp. 431–467.

MacKinlay, A. C. (1995). "Multifactor Models do not Explain Deviations from the CAPM". In: *Journal of Financial Economics* 38, pp. 3–28.

Maio, P. (2013). "Intertemporal CAPM with Conditioning Variables". In: *Management Science* 59.1, pp. 122–141.

Maio, P. and P. Santa-Clara (2012). "Multifactor models and their consistency with the ICAPM". In: *Journal of Financial Economics* 106, pp. 586–613.

Merton, R. C. (1973). "An Intertemporal Capital Asset Pricing Model". In: *Econometrica* 41.5, pp. 867–887.

Merton, R. C. (1980). "On Estimating the Expected Return on the Market: An Exploratory Investigation". In: *Journal of Financial Economics* 8, pp. 323–361.

Nawalkha (1997). "A multibeta representation theorem for linear asset pricing theories". In: *Journal of Financial Economics* 46.3, p. 1997.

Newey, W. K. and K. D. West (1987). "A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708.

Newey, W. K. and K. D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". In: *Review of Economic Studies* 61.4, pp. 631–653.

Parker, J. A. and C. Julliard (2005). "Consumption risk and the cross section of expected returns". In: *Journal of Political Economy* 113.1, pp. 185–222.

Pástor, L. and R. F. Stambaugh (2003). "Liquidity Risk and Expected Stock Returns". In: *Journal of Political Economy* 111.3, pp. 642–685.

Petkova, R. (2006). "Do the Fama-French Factors Proxy for Innovations in Predictive Variables?" In: *Journal of Finance* 61.2, pp. 581–612.

Reisman, H. (1992). "Reference Variables, Factor Structure, and the Approximate Multibeta Representation". In: *Journal of Finance* 47.4, pp. 1303–1314.

Roll, R. (1977). "A Critique of the Asset Pricing Theory's Tests". In: *Journal of Financial Economics* 4, pp. 129–176.

Ross, S. A. (1976). “The Arbitrage Theory of Capital Asset Pricing”. In: *Journal of Economic Theory* 13.3, pp. 341–360.

Savov, A. (2011). “Asset Pricing with Garbage”. In: *Journal of Finance* 66.1, pp. 177–201.

Shanken, J. (1992). “The Current State of the Arbitrage Pricing Theory”. In: *Journal of Finance* 47.4, pp. 1569–1574.

Sharpe, W. F. (1964). “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”. In: *Journal of Finance* 19.3, pp. 425–442.

Stambaugh, R. F. (1997). “Analyzing investments whose histories differ in length”. In: *Journal of Financial Economics* 45, pp. 285–331.

Stambaugh, R. F. and Y. Yuan (2017). “Mispricing Factors”. In: *Review of Financial Studies* 4, pp. 1270–1315.

Tian, M. (2021). “Firm Characteristics and Empirical Factor Models: A Model Mining Experiment”. In: *Review of Financial Studies*. Forthcoming.

Vassalou, M. (2003). “News related to fuFlowsGDP growth as a risk factor in equity returns”. In: *Journal of Financial Economics* 68.1, pp. 47–73.

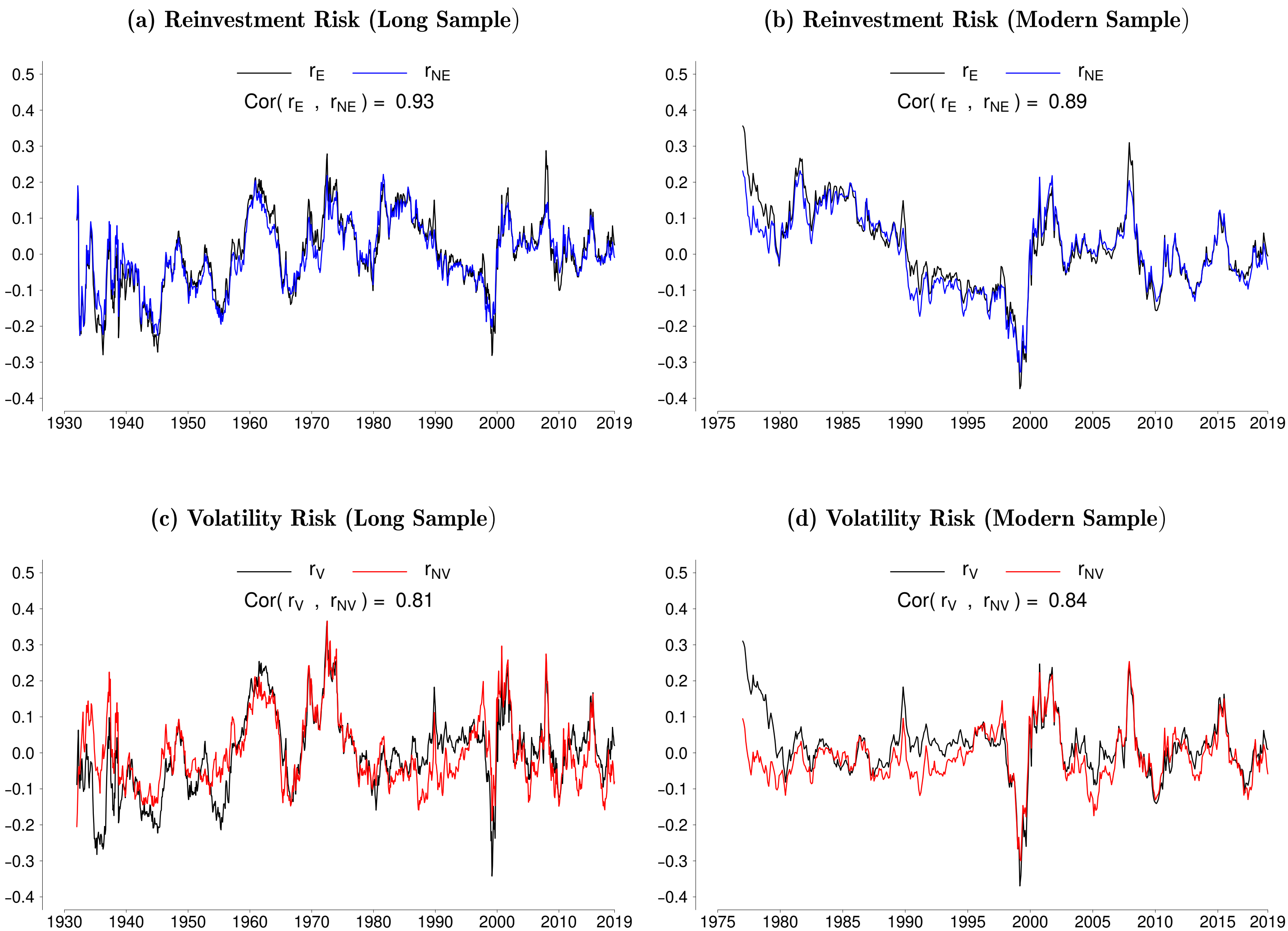

Figure 2
Filtered Intertemporal Risk Factors

These graphs compare filtered versions of ex-ante and ex-post mimicking portfolios for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ (using the filtering procedure in Campbell et al. (2018), which is described in Footnote 24). Panels (a) and (b) compare our tradable reinvestment risk factor ($r_{\mathbb{E}}$) to the ex-post $N_{\mathbb{E}}$ mimicking factor ($r_{N\mathbb{E}}$). Panels (c) and (d) compare our tradable volatility risk factor ($r_{\mathbb{V}}$) to the ex-post $N_{\mathbb{V}}$ mimicking factor ($r_{N\mathbb{V}}$). To construct $r_{\mathbb{E}}$ ($r_{\mathbb{V}}$), we buy a value-weighted portfolio of the stocks with the 30% highest exposures to Δdp ($\Delta\sigma^2$) and sell a value-weighted portfolio of the stocks with the 30% lowest exposures to Δdp ($\Delta\sigma^2$). To construct $r_{N\mathbb{E}}$ ($r_{N\mathbb{V}}$), we project $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) onto returns from decile portfolios constructed by sorting stocks based on their exposure to $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) and imposing that projection coefficients sum to zero (i.e., the factors are zero-net-cost portfolios). The tradable risk factors, $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$, as well as the decile portfolios necessary to obtain $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ are constructed each month using risk exposures estimated on a 5-year rolling window. Expected return news ($N_{\mathbb{E}}$) and volatility news ($N_{\mathbb{V}}$) are based on Equations 21 and 22, and are estimated ex-post using our Long Sample (1928-2019) in Panels (a) and (c) and Modern Sample (1973-2019) in Panels (b) and (d). A brief explanation of our news estimation procedure is provided in Subsection 1.3.4 with a detailed description available in Internet Appendix B.1. Subsection 1.3.3 provides further empirical details on the construction of the tradable risk factors.

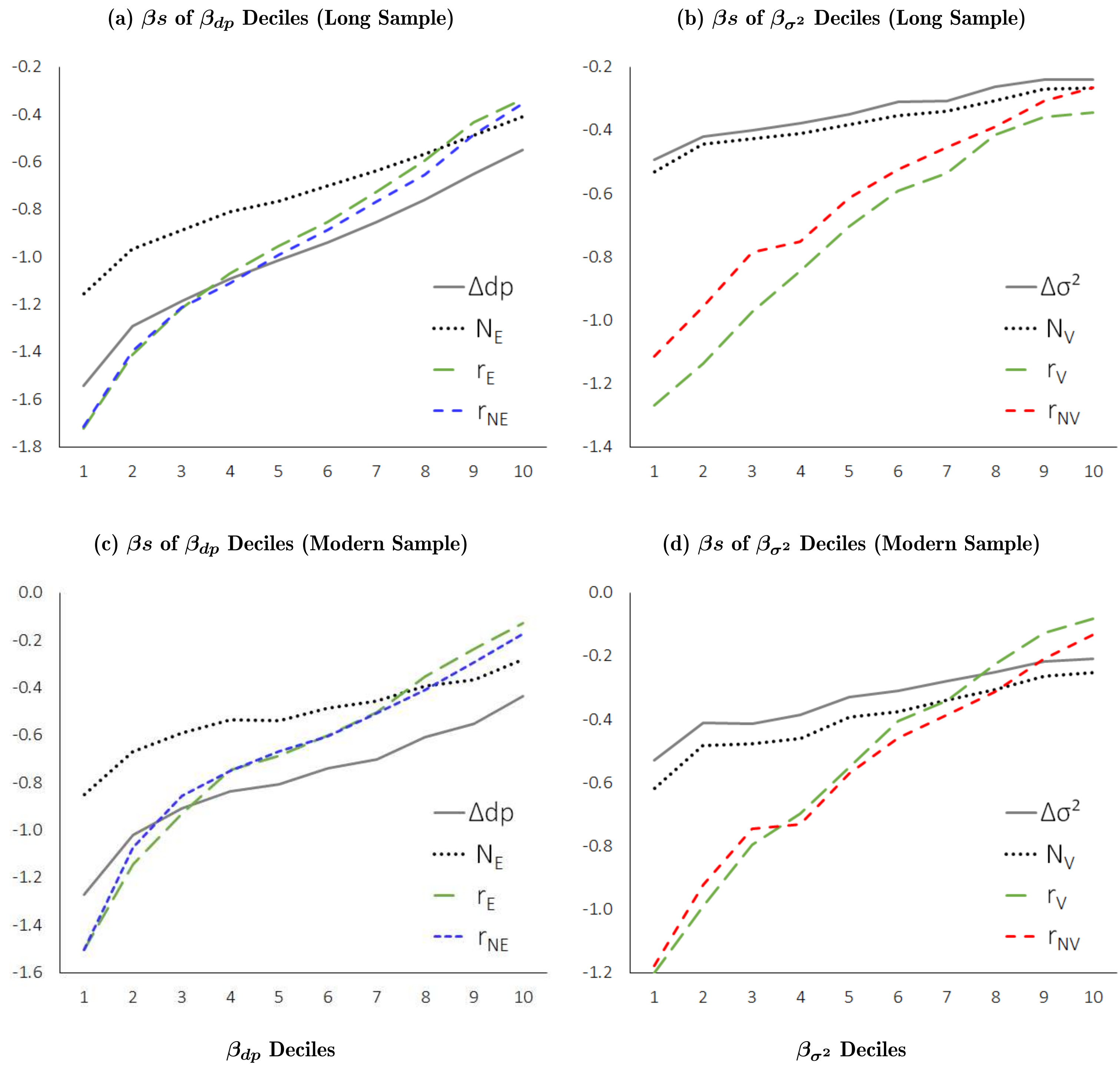

Figure 3

βs of β_{dp} and β_{σ^2} Decile Portfolios

These graphs show ex-post decile portfolio betas in market beta units (see Footnote 26) with various measures of reinvestment risk (Δdp, $N_{\mathbb{E}}$, $r_{\mathbb{E}}$, and $r_{N\mathbb{E}}$) and volatility risk ($\Delta\sigma^2$, $N_{\mathbb{V}}$, $r_{\mathbb{V}}$, and $r_{N\mathbb{V}}$). Panels (a) and (b) use data from our Long Sample (1928-2019). Panels (c) and (d) use data from our Modern Sample (1973-2019). Panels (a) and (c) use portfolios sorted on ex-ante exposure to Δdp. Panels (b) and (d) use portfolios sorted on ex-ante exposure to $\Delta\sigma^2$. See Subsection 1.3 for details related to the measurement of risk factors and decile portfolio construction.

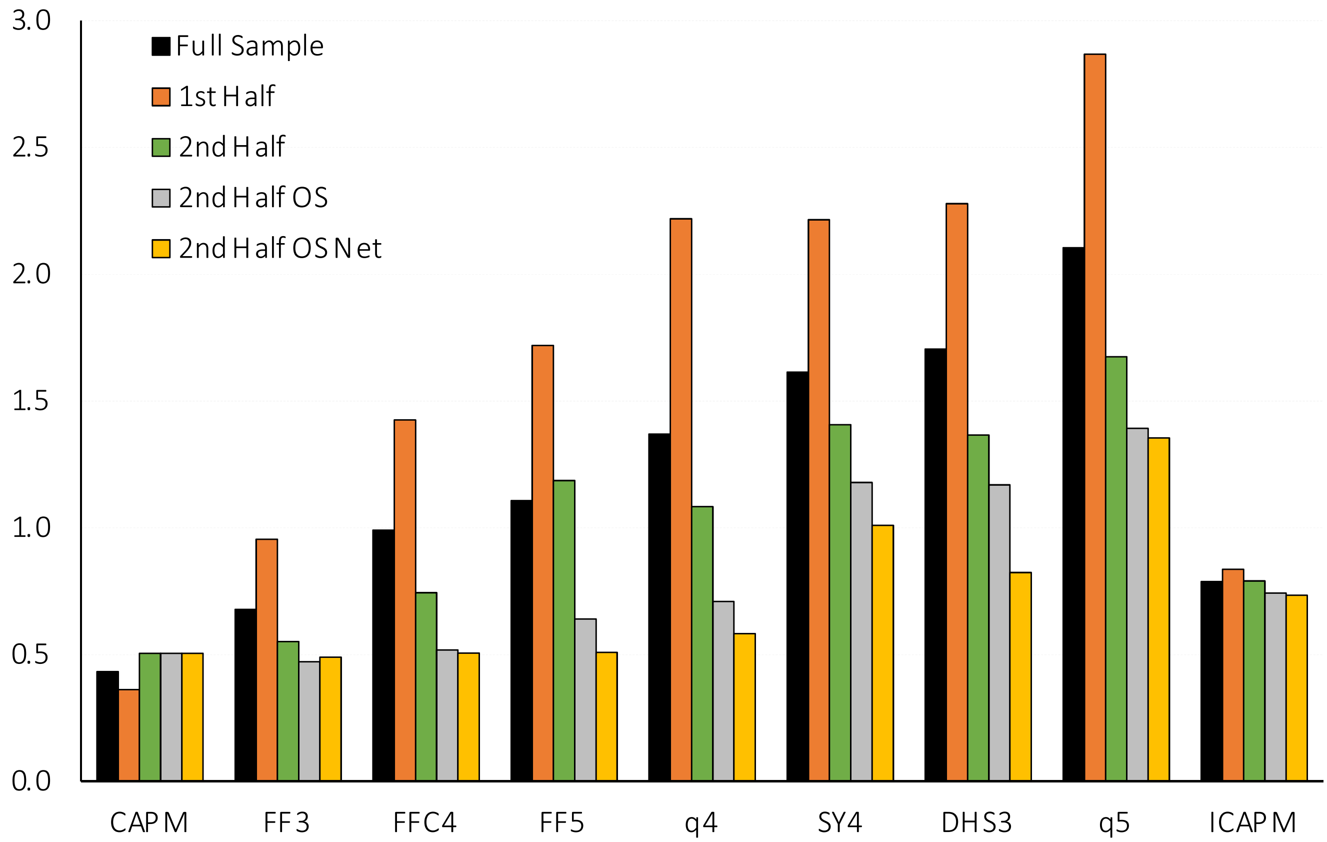

Figure 4

ICAPM vs Other Factor Models: Maximum Sharpe Ratios

This graph shows (in- and out-of-sample) maximum (gross and net-of-trading-cost) Sharpe ratios constructed using the ICAPM factors or using factors from other prominent factor models, which are described at the beginning of Section 3. Results are provided for three different periods: "Full Data" (1973-2019), "1st Half" (1973-1995), and "2nd Half" (1995-2019). For each model, the first three bars display maximum Sharpe ratios constructed using tangency portfolio weights estimated in-sample and applied to factors within each of these three periods. The fourth bar displays the Sharpe ratio that results from applying tangency portfolio weights estimated during the "1st Half" period to construct a portfolio of factors during the "2nd Half" period. The last bar displays the analogous out-of-sample Sharpe ratio when net-of-cost factors are used to both estimate weights and form the factor portfolio (see Footnote 34 for a description of the net-of-cost factor construction).

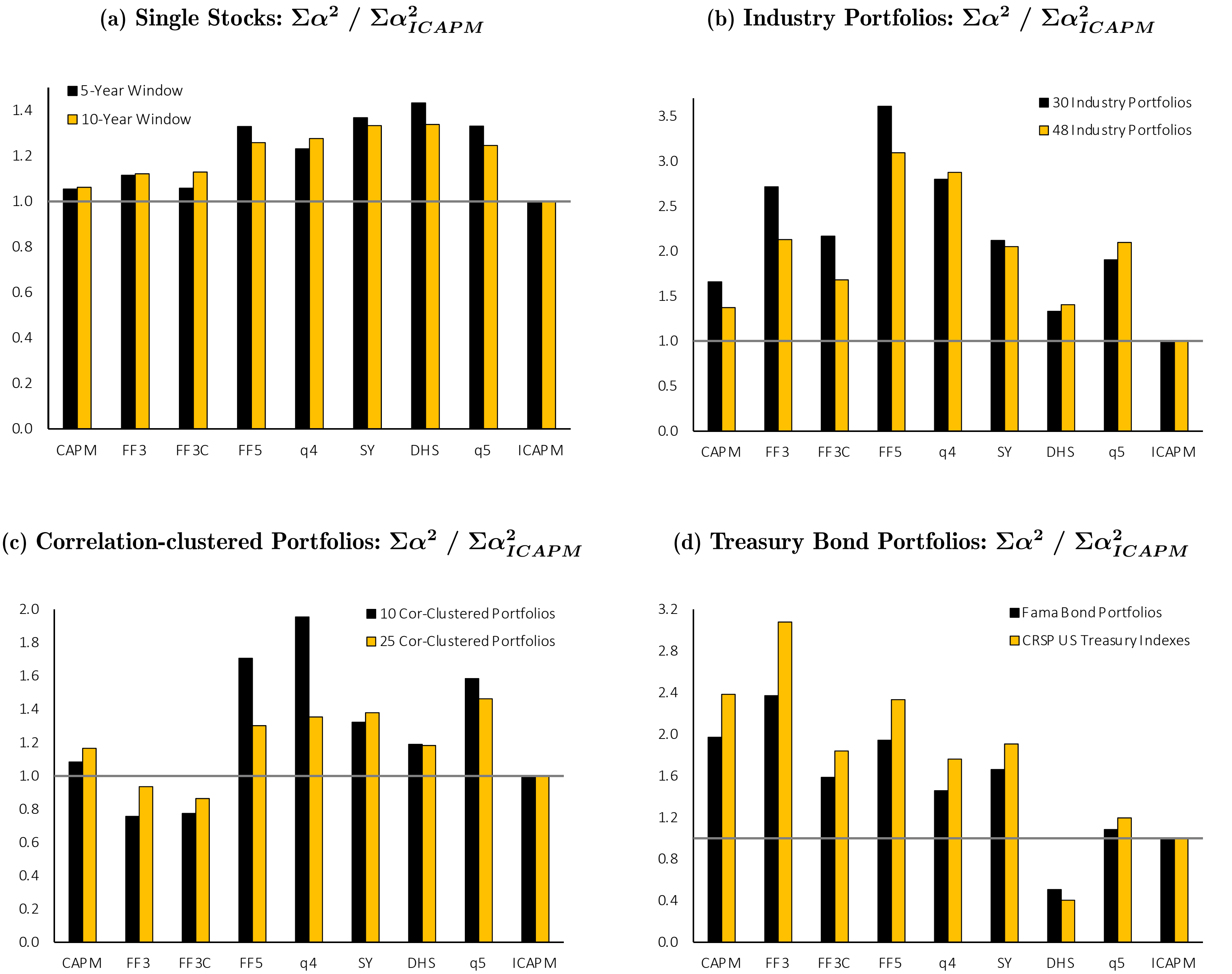

Figure 5

ICAPM vs Other Factor Models: αs for Testing Assets in LNS

These graphs report $\Sigma\alpha^2/\Sigma\alpha^2_{ICAPM}$, which reflect the sum of squared pricing errors from various prominent factor models (described at the beginning of Section 3) relative to that from the ICAPM. Each panel is based on a different group of the test assets recommended in Lewellen, Nagel, and Shanken (2010). Panel A focuses on single stocks with αs for each stock computed on either a 5- or 10-year rolling basis and recorded each month. Panel B focuses on either the 30 or 48 industry portfolios from Fama and French (1997). Panel C focuses on either 10 or 25 correlation-clustered portfolios (constructed using the method in Ahn, Conrad, and Dittmar (2009)). Panel D focuses on either the Fama bond portfolios or the CRSP US Treasury Indexes. In the case of single stocks (Panel A), we compute $\Sigma\alpha^2/\Sigma\alpha^2_{ICAPM}$ each month using only stocks that have full return data in the given rolling period and then plot time-series averages of the monthly $\Sigma\alpha^2/\Sigma\alpha^2_{ICAPM}$. In the case of portfolios (Panels B, C, and D), $\Sigma\alpha^2/\Sigma\alpha^2_{ICAPM}$ is computed using the full panel of portfolio returns. All data are from our Modern Sample (1973-2019).

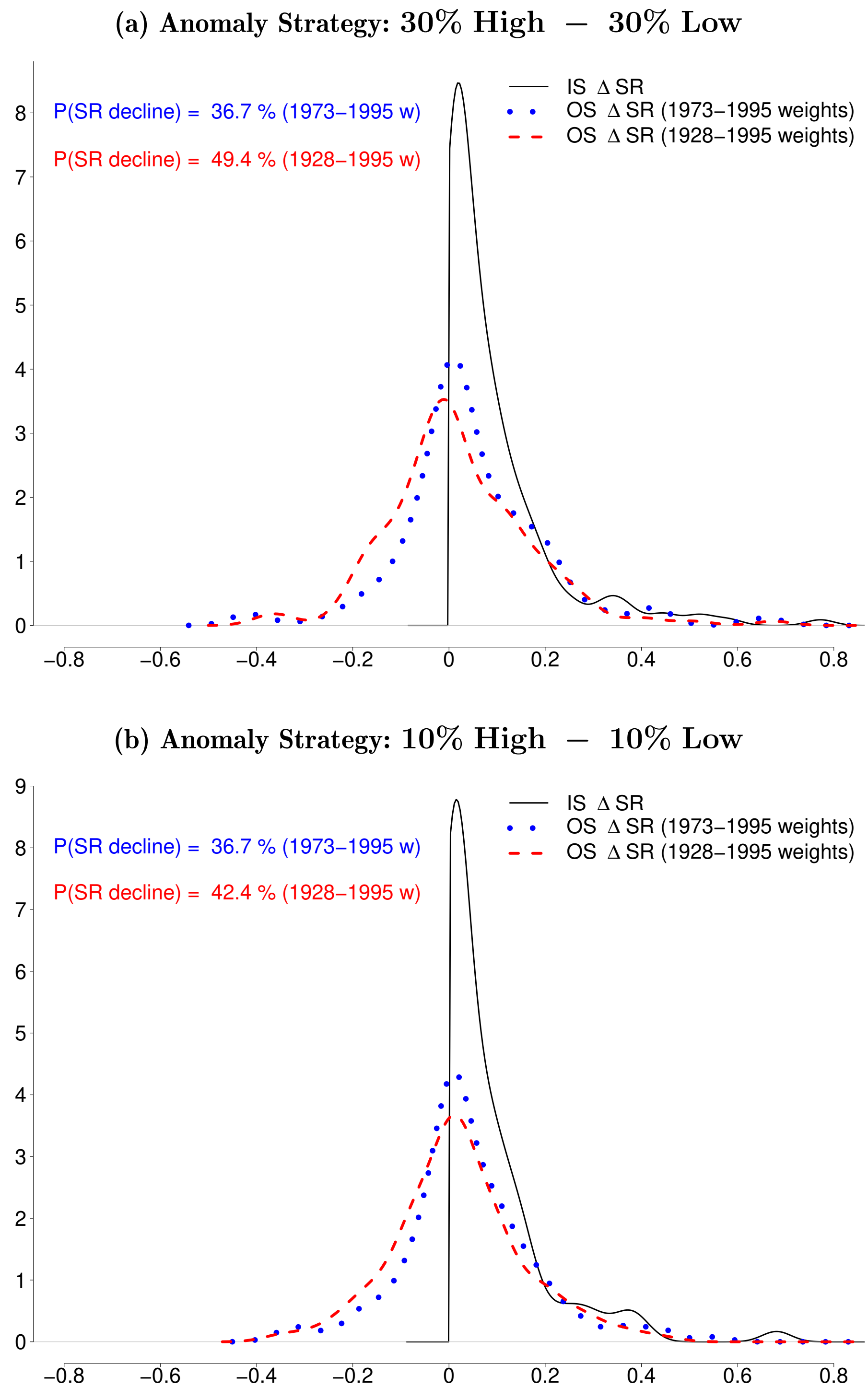

Figure 6

Distribution of Changes to the Tangency Portfolio Sharpe Ratio:

Adding one Anomaly H-L Strategy to ICAPM Tangency Portfolio

These graphs display the empirical density function of $\Delta SR_j = SR_{max,j} - SR_{max}$. SR_{max} reflects the ICAPM tangency portfolio Sharpe ratio and $SR_{max,j}$ reflects the tangency Sharpe ratio that can be achieved by adding each of the High-Low strategies based on the 158 anomalies we study (from Chen and Zimmermann (2020)) to the ICAPM factors. We consider in-sample (IS) Sharpe ratios as well as out-of-sample (OS) Sharpe ratios in which the weights are calculated prior to the Sharpe ratio measurement period. Subsection 4.2 provides further details.

Table 1

Correlations: News Proxies, Tradable Factors, and Ex-post Mimicking Factors

This table reports correlations between our tradable risk factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) and ex-post news ($N_{\mathbb{E}}$ and $N_{\mathbb{V}}$) as well as ex-post news mimicking factors ($r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$). To construct $r_{\mathbb{E}}$ ($r_{\mathbb{V}}$), we buy a value-weighted portfolio of the stocks with the 30% highest exposures to Δdp ($\Delta\sigma^2$) and sell a value-weighted portfolio of the stocks with the 30% lowest exposures to Δdp ($\Delta\sigma^2$). To construct $r_{N\mathbb{E}}$ ($r_{N\mathbb{V}}$), we project $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) onto returns from decile portfolios constructed by sorting stocks based on their exposure to $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) and imposing that projection coefficients sum to zero (i.e., the factors are zero-net-cost portfolios). The news are based on Equations 21 and 22, and are estimated ex-post using our Long Sample (1928-2019) or Modern Sample (1973-2019). The tradable risk factors, $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$, as well as the decile portfolios necessary to obtain the ex-post mimicking factors are constructed each month using risk exposures estimated on a 5-year rolling window. Subsections 1.3.3 and 1.3.4 provide further empirical details on the construction of the tradable and ex-post mimicking risk factors. A detailed description of our news estimation procedure is available in Internet Appendix B.1.

PANEL A: Reinvestment Risk ($N_{\mathbb{E}}$)

	Long Sample (1928-2019)				Modern Sample (1973-2019)			
	Δdp	$N_{\mathbb{E}}$	$r_{\mathbb{E}}$	$r_{N\mathbb{E}}$	Δdp	$N_{\mathbb{E}}$	$r_{\mathbb{E}}$	$r_{N\mathbb{E}}$
Δdp	1	0.91	0.68	0.68	1	0.88	0.54	0.57
$N_{\mathbb{E}}$	0.91	1	0.51	0.51	0.88	1	0.36	0.39
$r_{\mathbb{E}}$	0.68	0.51	1	0.93	0.54	0.36	1	0.89
$r_{N\mathbb{E}}$	0.68	0.51	0.93	1	0.57	0.39	0.889	1

PANEL B: Volatility Risk ($N_{\mathbb{V}}$)

	Long Sample (1928-2019)				Modern Sample (1973-2019)			
	$\Delta\sigma^2$	$N_{\mathbb{V}}$	$r_{\mathbb{V}}$	$r_{N\mathbb{V}}$	$\Delta\sigma^2$	$N_{\mathbb{V}}$	$r_{\mathbb{V}}$	$r_{N\mathbb{V}}$
$\Delta\sigma^2$	1	0.75	0.24	0.27	1	0.92	0.25	0.30
$N_{\mathbb{V}}$	0.75	1	0.24	0.30	0.92	1	0.28	0.36
$r_{\mathbb{V}}$	0.24	0.24	1	0.81	0.25	0.28	1	0.84
$r_{N\mathbb{V}}$	0.27	0.30	0.81	1	0.30	0.36	0.84	1

Table 2
Decile Portfolios Sorted on β_{dp}

This table reports statistics related to monthly returns on 10 β_{dp}-sorted portfolios. Panels A and B provide results from our Long (1928-2019) and Modern (1973-2019) samples, respectively. In the top portion of each panel, we report portfolio return exposures to our expected return news proxy (Δdp), the in-sample expected return news measure ($N_{\mathbb{E}}$), our tradable reinvestment risk factor ($r_{\mathbb{E}}$), and the $N_{\mathbb{E}}$ mimicking portfolio ($r_{N\mathbb{E}}$). Portfolio return exposures to each of these time series are denoted by β_{dp}, $\beta_{N\mathbb{E}}$, $\beta_{\mathbb{E}}$, and $\beta_{N\mathbb{E}}$, respectively, and are normalized to be in market beta units (see Footnote 26). In the bottom portion of each panel, we report portfolio average returns ($\mathbb{E}[r]$) and αs when computed with respect to the CAPM (α_m), the ICAPM excluding $r_{\mathbb{E}}$ ($\alpha_{m,\mathbb{V}}$), and the full ICAPM ($\alpha_{m,\mathbb{E},\mathbb{V}}$). All returns are in percent and annualized (approximately) by multiplying monthly returns by 12. The "Slope" statistic is a measure of the slope of the 10 related portfolio statistics with respect to portfolio decile (see Footnote 26). Portfolios are rebalanced monthly based on individual stock exposures to Δdp with further details provided in Subsection 1.3. The 10-1 portfolio t-statistics are computed according to Newey and West (1987, 1994). The Slope t-statistics are computed according to the method in Driscoll and Kraay (1998) using the procedure in Newey and West (1994) to select the number of lags.

PANEL A: Long Sample (1928-2019)

Dec =	1	2	3	4	5	6	7	8	9	10	10-1	(t_{10-1})	Slope	(t_{Slope})
β_{dp}	-1.54	-1.29	-1.18	-1.09	-1.01	-0.94	-0.85	-0.76	-0.65	-0.55	0.99	(11.5)	0.89	(8.46)
$\beta_{N\mathbb{E}}$	-1.15	-0.97	-0.89	-0.81	-0.77	-0.70	-0.64	-0.57	-0.49	-0.41	0.74	(6.65)	0.67	(5.05)
$\beta_{\mathbb{E}}$	-1.72	-1.41	-1.22	-1.07	-0.95	-0.85	-0.72	-0.59	-0.43	-0.33	1.39	(56.2)	1.29	(55.4)
$\beta_{rN\mathbb{E}}$	-1.71	-1.40	-1.21	-1.11	-0.99	-0.88	-0.77	-0.65	-0.48	-0.35	1.36	(19.2)	1.23	(17.1)
$\mathbb{E}[r]$	9.7%	8.8%	9.2%	9.2%	9.0%	9.5%	8.8%	8.1%	8.3%	6.7%	-3.0%	(-1.11)	-2.0%	(-0.57)
α_m	-3.9%	-2.6%	-1.4%	-0.6%	-0.1%	1.2%	1.2%	1.3%	2.5%	1.9%	5.8%	(2.87)	5.9%	(2.54)
$\alpha_{m,\mathbb{V}}$	-4.0%	-2.7%	-1.4%	-0.6%	-0.1%	1.2%	1.2%	1.3%	2.5%	1.9%	5.8%	(3.77)	5.9%	(3.31)
$\alpha_{m,\mathbb{E},\mathbb{V}}$	0.5%	1.0%	1.0%	0.8%	0.8%	1.9%	1.4%	0.5%	1.0%	0.5%	-0.1%	(-0.08)	0.0%	(-0.02)

PANEL B: Modern Sample (1973-2019)

Dec =	1	2	3	4	5	6	7	8	9	10	10-1	(t_{10-1})	Slope	(t_{Slope})
β_{dp}	-1.27	-1.02	-0.91	-0.84	-0.81	-0.74	-0.70	-0.61	-0.55	-0.44	0.84	(10.4)	0.70	(7.69)
$\beta_{N\mathbb{E}}$	-0.85	-0.67	-0.59	-0.54	-0.54	-0.49	-0.46	-0.39	-0.37	-0.28	0.57	(6.94)	0.47	(4.46)
$\beta_{\mathbb{E}}$	-1.50	-1.14	-0.93	-0.75	-0.69	-0.60	-0.50	-0.35	-0.24	-0.13	1.38	(43.7)	1.22	(56.6)
$\beta_{rN\mathbb{E}}$	-1.50	-1.07	-0.85	-0.75	-0.67	-0.60	-0.51	-0.41	-0.29	-0.17	1.33	(17.5)	1.12	(12.9)
$\mathbb{E}[r]$	7.5%	6.7%	7.0%	7.1%	7.9%	8.0%	7.2%	7.2%	8.2%	6.4%	-1.1%	(-0.31)	0.1%	(0.02)
α_m	-3.3%	-2.2%	-1.0%	-0.3%	0.9%	1.5%	1.0%	1.8%	3.3%	2.6%	5.9%	(2.30)	6.0%	(2.08)
$\alpha_{m,\mathbb{V}}$	-3.5%	-2.3%	-1.0%	-0.2%	1.0%	1.6%	1.1%	1.8%	3.4%	2.7%	6.2%	(3.26)	6.2%	(2.94)
$\alpha_{m,\mathbb{E},\mathbb{V}}$	0.1%	0.9%	1.3%	0.6%	1.7%	2.2%	0.8%	0.4%	1.0%	0.1%	0.0%	(0.01)	-0.1%	(-0.16)

Table 3
Decile Portfolios Sorted on β_{σ^2}

This table reports statistics related to monthly returns on 10 β_{σ^2}-sorted portfolios. Panels A and B provide results from our Long (1928-2019) and Modern (1973-2019) samples, respectively. In the top portion of each panel, we report portfolio return exposures to our volatility news proxy ($\Delta\sigma^2$), the in-sample volatility news measure ($N_{\mathbb{V}}$), our tradable volatility risk factor ($r_{\mathbb{V}}$), and the $N_{\mathbb{V}}$ mimicking portfolio ($r_{N\mathbb{V}}$). Portfolio return exposures to each of these time series are denoted by β_{σ^2}, $\beta_{N\mathbb{V}}$, $\beta_{\mathbb{V}}$, and $\beta_{N\mathbb{V}}$, respectively, and are normalized to be in market beta units (see Footnote 26). In the bottom portion of each panel, we report portfolio average returns ($\mathbb{E}[r]$) and αs when computed with respect to the CAPM (α_m), the ICAPM excluding $r_{\mathbb{V}}$ ($\alpha_{m,\mathbb{E}}$), and the full ICAPM ($\alpha_{m,\mathbb{E},\mathbb{V}}$). All returns are in percent and annualized (approximately) by multiplying monthly returns by 12. The "Slope" statistic is a measure of the slope of the 10 related portfolio statistics with respect to portfolio decile (see Footnote 26). Portfolios are rebalanced monthly based on individual stock exposures to $\Delta\sigma^2$ with further details provided in Subsection 1.3. The 10-1 portfolio t-statistics are computed according to Newey and West (1987, 1994). The Slope t-statistics are computed according to the method in Driscoll and Kraay (1998) using the procedure in Newey and West (1994) to select the number of lags.

PANEL A: Long Sample (1928-2019)

Dec =	1	2	3	4	5	6	7	8	9	10	10-1	(t_{10-1})	Slope	(t_{Slope})
β_{σ^2}	-0.49	-0.42	-0.40	-0.38	-0.35	-0.31	-0.31	-0.26	-0.24	-0.24	0.25	(6.77)	0.24	(5.00)
$\beta_{N\mathbb{V}}$	-0.53	-0.44	-0.43	-0.41	-0.38	-0.35	-0.34	-0.30	-0.27	-0.27	0.27	(6.52)	0.24	(4.53)
$\beta_{\mathbb{V}}$	-1.27	-1.14	-0.97	-0.84	-0.70	-0.59	-0.54	-0.42	-0.36	-0.34	0.93	(34.0)	0.96	(66.0)
$\beta_{rN\mathbb{V}}$	-1.11	-0.96	-0.78	-0.75	-0.61	-0.52	-0.45	-0.39	-0.31	-0.27	0.85	(23.1)	0.83	(20.2)
$\mathbb{E}[r]$	9.8%	10.2%	10.9%	9.3%	9.1%	8.2%	8.0%	7.4%	7.5%	5.8%	-4.0%	(-1.98)	-4.2%	(-1.55)
α_m	-2.0%	-0.3%	1.4%	0.4%	0.8%	0.8%	0.8%	0.8%	0.9%	-1.0%	0.9%	(0.53)	0.8%	(0.38)
$\alpha_{m,\mathbb{E}}$	0.9%	1.8%	2.7%	1.1%	1.0%	0.4%	0.3%	-0.1%	0.0%	-1.8%	-2.6%	(-1.97)	-2.9%	(-1.89)
$\alpha_{m,\mathbb{E},\mathbb{V}}$	-0.4%	0.3%	1.5%	0.3%	0.9%	0.3%	0.5%	0.8%	1.4%	0.1%	0.5%	(0.49)	0.5%	(0.79)

PANEL B: Modern Sample (1973-2019)

Dec =	1	2	3	4	5	6	7	8	9	10	10-1	(t_{10-1})	Slope	(t_{Slope})
β_{σ^2}	-0.53	-0.41	-0.41	-0.39	-0.33	-0.31	-0.28	-0.25	-0.22	-0.21	0.32	(4.94)	0.29	(3.64)
$\beta_{N\mathbb{V}}$	-0.62	-0.48	-0.48	-0.46	-0.39	-0.38	-0.34	-0.30	-0.26	-0.25	0.36	(5.25)	0.33	(3.84)
$\beta_{\mathbb{V}}$	-1.20	-0.99	-0.80	-0.70	-0.55	-0.40	-0.34	-0.23	-0.13	-0.08	1.12	(33.5)	1.10	(70.9)
$\beta_{rN\mathbb{V}}$	-1.18	-0.92	-0.75	-0.73	-0.57	-0.46	-0.38	-0.31	-0.21	-0.13	1.05	(15.2)	0.97	(14.2)
$\mathbb{E}[r]$	8.8%	9.1%	9.5%	7.4%	9.1%	7.8%	7.6%	6.7%	6.6%	4.7%	-4.1%	(-1.47)	-3.8%	(-1.03)
α_m	-0.9%	0.6%	1.8%	-0.1%	2.3%	1.5%	1.6%	0.9%	1.2%	-1.1%	-0.1%	(-0.05)	0.2%	(0.06)
$\alpha_{m,\mathbb{E}}$	2.3%	2.5%	3.0%	0.7%	2.6%	0.9%	0.8%	-0.5%	-0.4%	-2.7%	-5.0%	(-2.91)	-4.6%	(-2.18)
$\alpha_{m,\mathbb{E},\mathbb{V}}$	0.0%	0.3%	1.6%	0.0%	2.3%	1.1%	1.2%	0.7%	1.4%	0.0%	-0.1%	(-0.07)	0.3%	(0.46)

Table 4
The ICAPM Factor Risk Premia

This table reports ICAPM risk factor correlations, expected returns, standard deviations, and Sharpe ratios over our Long (1928-2019) and Modern (1973-2019) Samples (Panels A and B, respectively). Statistics are provided for the raw factors (r) as well as for factors that are orthogonalized in-sample ($IS\ r^{\perp}$) according to the weighting procedure outlined in Subsection 2.3. Out-of-sample versions of the orthogonalized factors ($OS\ r^{\perp}$) are constructed using weights estimated using this same procedure but applied to 10-year rolling windows preceding each month in which the weights are applied. $\mathbb{E}[r]$ and $\sigma[r]$ are the annualized average returns and return standard deviations of each of each factor. The t-statistics are computed according to Newey and West (1987, 1994).

PANEL A: Long Sample (1928-2019)

		$Cor(r, r_m)$	$Cor(r, r_{\mathbb{E}})$	$Cor(r, r_{\mathbb{V}})$	$\mathbb{E}[r]$	$(t_{\mathbb{E}[r]})$	$\sigma[r]$	$\mathbb{E}[r]/\sigma[r]$
r	m	1.00	-0.77	-0.64	7.8%	(3.74)	18.5%	0.42
	$\mathbb{E}$	-0.77	1.00	0.82	-1.3%	(-0.71)	18.0%	-0.07
	$\mathbb{V}$	-0.64	0.82	1.00	-3.7%	(-2.57)	13.7%	-0.27
$IS\ r^{\perp}$	m	0.63	0.00	0.00	10.5%	(4.58)	18.5%	0.57
	$\mathbb{E}$	0.00	0.47	0.00	9.9%	(5.04)	18.5%	0.53
	$\mathbb{V}$	0.00	0.00	0.58	-6.6%	(-3.69)	18.5%	-0.36
$OS\ r^{\perp}$	m	0.56	0.13	0.12	10.2%	(4.88)	18.0%	0.57
	$\mathbb{E}$	-0.02	0.50	0.06	10.8%	(5.88)	17.7%	0.61
	$\mathbb{V}$	-0.04	0.11	0.58	-7.9%	(-3.87)	18.1%	-0.44

PANEL B: Modern Sample (1973-2019)

		$Cor(r, r_m)$	$Cor(r, r_{\mathbb{E}})$	$Cor(r, r_{\mathbb{V}})$	$\mathbb{E}[r]$	$(t_{\mathbb{E}[r]})$	$\sigma[r]$	$\mathbb{E}[r]/\sigma[r]$
r	m	1.00	-0.68	-0.51	6.8%	(2.83)	15.6%	0.43
	$\mathbb{E}$	-0.68	1.00	0.80	0.5%	(0.24)	15.0%	0.04
	$\mathbb{V}$	-0.51	0.80	1.00	-3.2%	(-1.64)	13.1%	-0.24
$IS\ r^{\perp}$	m	0.73	0.00	0.00	10.2%	(4.25)	15.6%	0.65
	$\mathbb{E}$	0.00	0.52	0.00	10.3%	(4.47)	15.6%	0.66
	$\mathbb{V}$	0.00	0.00	0.61	-7.6%	(-3.51)	15.6%	-0.48
$OS\ r^{\perp}$	m	0.74	-0.02	-0.01	9.4%	(3.59)	16.3%	0.58
	$\mathbb{E}$	-0.05	0.49	-0.02	13.2%	(5.04)	17.8%	0.74
	$\mathbb{V}$	0.07	-0.04	0.51	-10.7%	(-3.83)	18.5%	-0.58

Table 5
The ICAPM Risk Prices and Pricing Errors

Panel A reports estimated CAPM and ICAPM risk prices (b) according to Equation 8 while Panel B reports the annualized average returns ($\mathbb{E}[r]$) and associated pricing errors (α) for the three orthogonalized strategies introduced in Subsection 2.3. The ICAPM$_\gamma$ column reports the respective information when imposing the ICAPM structural restrictions in Footnote 11, which imply relative risk aversion, γ, is the only parameter as it determines all three risk prices (see Footnote 30 for estimation details). For the first four columns of each panel, we use $\gamma = b_m = 1_m' \Sigma_f^{-1} \mathbb{E}[f]$ (see Subsection 1.3.1). Since bs are not easily comparable, we report $\sigma_k \cdot b_k$ for each factor $f_{k,t}$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective $f_{k,t}$ (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed according to GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix (see Internet Appendix B.2).

	Long Sample (1928-2019)					**Modern Sample (1973-2019)**				
	CAPM	**ICAPM$_\mathbb{E}$**	**ICAPM$_\mathbb{V}$**	**ICAPM**	**ICAPM$_\gamma$**	**CAPM**	**ICAPM$_\mathbb{E}$**	**ICAPM$_\mathbb{V}$**	**ICAPM**	**ICAPM$_\gamma$**
PANEL A: Risk Prices ($M_t = a + b' f_t$)										
$\boldsymbol{b_m}$	0.12	0.26	0.12	0.26	0.34	0.12	0.24	0.12	0.26	0.29
$\boldsymbol{(t_{stat})}$	(3.05)	(3.57)	(2.35)	(3.59)	(6.00)	(2.44)	(3.23)	(2.13)	(3.45)	(7.19)
$\boldsymbol{b_\mathbb{E}}$		0.18		0.32	0.26		0.18		0.37	0.27
$\boldsymbol{(t_{stat})}$		(3.02)		(4.28)	(5.05)		(2.64)		(4.19)	(6.14)
$\boldsymbol{b_\mathbb{V}}$			-0.00	-0.18	-0.11			-0.01	-0.23	-0.23
$\boldsymbol{(t_{stat})}$			(-0.03)	(-3.34)	(-2.53)			(-0.15)	(-3.52)	(-3.07)
$[\gamma]$	[2.3]	[4.9]	[2.2]	[4.8]	[6.3]	[2.8]	[5.4]	[2.7]	[5.6]	[6.9]
PANEL B: Annualized Pricing Errors (αs)										
$\mathbb{E}[r_m^\perp]$	10.5%	10.5%	10.5%	10.5%	10.5%	10.2%	10.2%	10.2%	10.2%	10.2%
$\boldsymbol{\alpha_m}$	5.6%	-0.1%	5.6%	0.0%	-3.2%	5.2%	0.5%	5.3%	0.0%	-2.2%
$\boldsymbol{(t_{stat})}$	(3.72)	(-3.36)	(5.07)	(0.00)	(-2.18)	(3.32)	(3.26)	(4.47)	(0.00)	(-1.01)
$\mathbb{E}[r_\mathbb{E}^\perp]$	9.9%	9.9%	9.9%	9.9%	9.9%	10.3%	10.3%	10.3%	10.3%	10.3%
$\boldsymbol{\alpha_\mathbb{E}}$	9.9%	4.4%	9.9%	0.0%	2.0%	10.3%	5.4%	10.3%	0.0%	2.6%
$\boldsymbol{(t_{stat})}$	(5.07)	(3.36)	(5.07)	(0.00)	(2.33)	(4.47)	(3.26)	(4.47)	(0.00)	(1.89)
$\mathbb{E}[r_\mathbb{V}^\perp]$	-6.6%	-6.6%	-6.6%	-6.6%	-6.6%	-7.6%	-7.6%	-7.6%	-7.6%	-7.6%
$\boldsymbol{\alpha_\mathbb{V}}$	-6.6%	-6.6%	-6.5%	0.0%	-2.7%	-7.6%	-7.6%	-7.3%	0.0%	-0.2%
$\boldsymbol{(t_{stat})}$	(-3.38)	(-3.36)	(-5.07)	(0.00)	(-1.46)	(-3.29)	(-3.26)	(-4.47)	(0.00)	(-0.12)

Table 6
The ICAPM Risk Prices Controlling for Factors in Other Factor Models

This table reports estimated risk prices (b) for ICAPM risk factors (f_t) according to Equation 8 when controlling for factors from other prominent factor models (x_t), which are described at the beginning of Section 3. Panels A and B cover our Long (1928-2019) and Modern (1973-2019) samples, respectively. In the case of the Long Sample, we include the earliest factor data available for each model (with the starting year listed in parentheses below each model) and use the Stambaugh (1997) procedure to estimate b over the entire Long Sample (see Subsection 3.1 for more details). Since bs are not easily comparable, we report $\sigma_f \cdot b$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective factor (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed using a bootstrap exercise in Panel A (see Internet Appendix B.4) and GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix in Panel B (see Internet Appendix B.2).

PANEL A: Long Sample (1928-2019)

$$M_t = a - b' f_t - b_x' x_t$$

	$x =$	None (1928)	FF3 (1928)	FFC4 (1928)	FF5 (1963)	q4 (1963)	SY4 (1967)	DHS3 (1972)	q5 (1967)
r_m	b	0.26	0.27	0.27	0.31	0.29	0.45	0.34	0.42
	(t_{stat})	(3.59)	(3.60)	(3.93)	(5.03)	(4.57)	(6.44)	(5.23)	(5.47)
$r_{\mathbb{E}}$	b	0.32	0.45	0.38	0.26	0.34	0.34	0.21	0.22
	(t_{stat})	(4.28)	(5.53)	(4.71)	(3.24)	(3.80)	(3.81)	(2.38)	(2.18)
$r_{\mathbb{V}}$	b	-0.18	-0.23	-0.24	-0.31	-0.38	-0.38	-0.35	-0.37
	(t_{stat})	(-3.34)	(-4.72)	(-4.60)	(-4.51)	(-5.33)	(-5.16)	(-4.51)	(-4.49)

PANEL B: Modern Sample (1973-2019)

$$M_t = a - b' f_t - b_x' x_t$$

	$x =$	None	FF3	FFC4	FF5	q4	SY4	DHS3	q5
r_m	b	0.26	0.28	0.27	0.30	0.27	0.41	0.33	0.37
	(t_{stat})	(3.45)	(3.58)	(3.56)	(4.17)	(3.82)	(5.32)	(4.30)	(4.74)
$r_{\mathbb{E}}$	b	0.37	0.37	0.28	0.29	0.27	0.20	0.15	0.16
	(t_{stat})	(4.19)	(4.23)	(3.16)	(3.02)	(2.69)	(2.24)	(1.85)	(1.58)
$r_{\mathbb{V}}$	b	-0.23	-0.23	-0.24	-0.35	-0.38	-0.36	-0.36	-0.35
	(t_{stat})	(-3.52)	(-3.62)	(-3.45)	(-4.58)	(-4.84)	(-4.47)	(-5.09)	(-3.94)

Table 7
ICAPM vs Other Factor Models: Maximum Sharpe Ratios

This table reports the maximum Sharpe ratios constructed (in- and out-of-sample) using the ICAPM factors or using factors from other prominent factor models, which are described at the beginning of Section 3. We also simulate factor data and report the percent of times the ICAPM maximum Sharpe ratio is higher than each alternative model in parenthesis. The simulation results are based on a bootstrap analysis that samples the data 100,000 times with replacement, then recomputes each model's maximum Sharpe ratio within each simulation (see Internet Appendix B.4). Panel A presents results using the gross (of trading cost) factors and Panel B presents results using net-of-trading-cost factors (described in Footnote 34). Results reported for the "Modern Sample", "1st Half", and "2nd Half" are based on portfolio weights (w) estimated in-sample. Results reported for "2nd Half OS (w from 1973-1995)" use factor data from 1973 through the first half of 1995 to estimate w. Results reported in "2nd Half OS (w from 1928-1995)" use factor data from 1928 through the first half of 1995 (or the earliest factor data available for each factor model, which is summarized in Table 6, Panel A) to estimate w.

PANEL A: Gross (of Trading Costs) Factor Returns

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
Modern Sample	0.43	0.68	0.99	1.11	1.37	1.61	1.70	2.10	0.79
(1973-2019)	(100%)	(73%)	(11%)	(3%)	(0%)	(0%)	(0%)	(0%)	-
1st Half	0.36	0.96	1.43	1.72	2.22	2.21	2.28	2.87	0.84
(1973-1995)	(100%)	(27%)	(1%)	(0%)	(0%)	(0%)	(0%)	(0%)	-
2nd Half	0.51	0.55	0.74	1.19	1.08	1.41	1.37	1.67	0.79
(1995-2019)	(100%)	(88%)	(53%)	(2%)	(9%)	(0%)	(1%)	(0%)	-
2nd Half OS	0.51	0.43	0.56	0.66	0.66	1.16	1.17	1.32	0.69
(w from 1973-1995)	(74%)	(77%)	(63%)	(48%)	(48%)	(5%)	(5%)	(2%)	-
2nd Half OS	0.51	0.47	0.52	0.64	0.71	1.18	1.17	1.39	0.74
(w from 1928-1995)	(86%)	(86%)	(75%)	(59%)	(51%)	(7%)	(8%)	(2%)	-

PANEL B: Net (of Trading Costs) Factor Returns

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
Modern Sample	0.43	0.62	0.75	0.87	0.90	1.19	1.23	1.70	0.66
(1973-2019)	(100%)	(61%)	(26%)	(7%)	(7%)	(0%)	(0%)	(0%)	-
1st Half	0.36	0.85	0.99	1.17	1.18	1.36	1.50	2.06	0.60
(1973-1995)	(100%)	(11%)	(3%)	(0%)	(1%)	(0%)	(0%)	(0%)	-
2nd Half	0.51	0.53	0.63	1.06	0.90	1.30	1.20	1.50	0.74
(1995-2019)	(100%)	(87%)	(65%)	(4%)	(20%)	(0%)	(2%)	(0%)	-
2nd Half OS	0.51	0.39	0.48	0.53	0.54	0.95	0.82	1.30	0.68
(w from 1973-1995)	(67%)	(76%)	(69%)	(57%)	(57%)	(14%)	(24%)	(2%)	-
2nd Half OS	0.51	0.49	0.51	0.51	0.58	1.01	0.82	1.35	0.73
(w from 1928-1995)	(84%)	(86%)	(77%)	(70%)	(62%)	(14%)	(31%)	(2%)	-

Table 8
ICAPM vs Other Factor Models: αs for Testing Assets in LNS

This table reports $\Sigma\alpha^2$ for each factor model (described in Section 3) normalized by $\Sigma\mathbb{E}[r]^2$ or $\Sigma\alpha^2_{ICAPM}$. Each panel is based on a different group of the testing assets recommended in Lewellen, Nagel, and Shanken (2010). Panel A focuses on single stocks with αs computed on a 5-year rolling basis and recorded each month. Panel B focuses on the 30 industry portfolios from Fama and French (1997). Panel C focuses on 10 correlation-clustered portfolios (constructed as in Ahn, Conrad, and Dittmar (2009)). Panel D focuses on the Fama bond portfolios. In the case of single stocks (Panel A), we compute $\Sigma\alpha^2$ ratios each month using only stocks that have full return data in the given rolling period and then report time-series averages of the monthly $\Sigma\alpha^2$ ratios. In the case of portfolios (Panels B, C, and D), $\Sigma\alpha^2$ ratios are computed using the full panel of portfolio returns. For each panel and model, we also report $\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$, which reflects the percent of times in which the $\Sigma\alpha^2$ generated by the ICAPM is below that generated by the alternate model. In the case of single stocks (Panel A), $\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$ is based on the time series of $\Sigma\alpha^2$. In the case of portfolios (Panels B to D), $\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$ is based on a bootstrap analysis that samples the data with replacement, then recomputes each model's α ratio metric within each simulation (see Internet Appendix B.4). All data are from our Modern Sample (1973-2019).

PANEL A: Single Stocks

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
$\Sigma\alpha^2$ / $\Sigma\mathbb{E}[r]^2$	0.60	0.62	0.60	0.74	0.69	0.76	0.82	0.75	0.58
$\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$	1.05	1.11	1.06	1.33	1.23	1.37	1.43	1.33	1.00
$\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$	(61%)	(71%)	(58%)	(85%)	(84%)	(87%)	(95%)	(87%)	-

PANEL B: Industry Portfolios

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
$\Sigma\alpha^2$ / $\Sigma\mathbb{E}[r]^2$	0.09	0.15	0.12	0.20	0.16	0.12	0.07	0.11	0.06
$\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$	1.66	2.72	2.17	3.61	2.80	2.12	1.33	1.90	1.00
$\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$	(86%)	(95%)	(92%)	(96%)	(92%)	(85%)	(76%)	(83%)	-

PANEL C: Correlation-clustered Portfolios

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
$\Sigma\alpha^2$ / $\Sigma\mathbb{E}[r]^2$	0.09	0.06	0.06	0.14	0.16	0.11	0.10	0.13	0.08
$\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$	1.08	0.76	0.77	1.71	1.95	1.32	1.19	1.58	1.00
$\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$	(60%)	(17%)	(25%)	(92%)	(96%)	(79%)	(75%)	(86%)	-

PANEL D: Treasury Bond Portfolios

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
$\Sigma\alpha^2$ / $\Sigma\mathbb{E}[r]^2$	0.87	1.04	0.70	0.85	0.64	0.73	0.22	0.48	0.44
$\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$	1.97	2.37	1.59	1.94	1.46	1.66	0.51	1.08	1.00
$\%(\Sigma\alpha^2_{ICAPM} < \Sigma\alpha)$	(98%)	(99%)	(88%)	(96%)	(78%)	(86%)	(23%)	(58%)	-

Table 9

ICAPM vs Other Factor Models: Summarizing α Results

This table reports $\Sigma\alpha^2/\Sigma\mathbb{E}[r]^2$ (Panel A) and $\Sigma|a|/\Sigma|\mathbb{E}[r]|$ (Panel B) ranks across each model within each test asset group recommend in Lewellen, Nagel, and Shanken (2010). Lower ranks correspond to lower ratio values (i.e., better models). The last row in each panel reports the average rank of the ICAPM and each other factor model (described in Section 3) across the eight test asset groups.

PANEL A: Squared Pricing Errors ($\Sigma\alpha^2/\Sigma\mathbb{E}[r]^2$) Ranks

		CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
Single Stocks	5-Year Window	2	4	3	6	5	8	9	7	1
	10-Year Window	2	3	4	6	7	8	9	5	1
Industry Port	30 Portfolios	3	7	6	9	8	5	2	4	1
	48 Portfolios	2	7	4	9	8	5	3	6	1
Cor Clust Port	10 Portfolios	4	1	2	8	9	6	5	7	3
	25 Portfolios	4	2	1	6	7	8	5	9	3
Bond Portfolios	Fama Bond Port	8	9	5	7	4	6	1	3	2
	CRSP Bond Port	8	9	5	7	4	6	1	3	2
	Average Rank =	4.1	5.3	3.8	7.3	6.5	6.5	4.4	5.5	1.8

PANEL B: Absolute Pricing Errors ($\Sigma|\alpha|/\Sigma|\mathbb{E}[r]|$) Ranks

		CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
Single Stocks	5-Year Window	4	1	2	6	5	7	9	8	3
	10-Year Window	3	1	2	5	6	7	9	8	4
Industry Port	30 Portfolios	3	7	6	9	8	5	2	4	1
	48 Portfolios	3	5	4	9	8	7	2	6	1
Cor Clust Port	10 Portfolios	5	2	1	8	9	4	7	6	3
	25 Portfolios	4	2	1	6	7	8	9	5	3
Bond Portfolios	Fama Bond Port	8	9	5	7	4	6	1	3	2
	CRSP Bond Port	8	9	5	7	4	6	1	3	2
	Average Rank =	4.8	4.5	3.3	7.1	6.4	6.3	5.0	5.4	2.4

Table 10
ICAPM vs Other Factor Models: Anomaly Deciles

This table reports $\Sigma\alpha^2$ and $\Sigma|\alpha|$ for each factor model (described at the beginning of Section 3) normalized by (respectively) the sum of squared or absolute average returns ($\Sigma\mathbb{E}[r]^2$ or $\Sigma|\mathbb{E}[r]|$) or ICAPM alphas ($\Sigma\alpha^2_{ICAPM}$ or $\Sigma|\alpha_{ICAPM}|$). Panel A focuses on deciles formed on 158 anomalies from Chen and Zimmermann (2020). Panel B focuses on the 20 deciles formed on β_{dp} and β_{σ^2} sorts. αs for each set of portfolios are estimated once using the full data from our Modern Sample (1973-2019) for each model. We also simulate portfolio return data and report the percent of times in which $\Sigma\alpha^2$ or $\Sigma|a|$ generated by the ICAPM is below that generated by the alternate model in parentheses. The simulation results are based on a bootstrap analysis that samples the data with replacement, then recomputes each model's α ratio metric within each simulation (see Internet Appendix B.4).

PANEL A: Deciles Based on 158 Anomalies

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
$\Sigma\alpha^2$ / $\Sigma\mathbb{E}[r]^2$	0.29	0.32	0.23	0.26	0.22	0.20	0.18	0.19	0.25
$\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$	1.17	1.28	0.93	1.05	0.89	0.83	0.72	0.79	1.00
%($\Sigma\alpha^2_{ICAPM} < \Sigma\alpha^2$)	(89%)	(96%)	(25%)	(59%)	(23%)	(9%)	(8%)	(7%)	-
$\Sigma|\alpha|$ / $\Sigma|\mathbb{E}[r]|$	0.30	0.30	0.25	0.28	0.27	0.25	0.26	0.26	0.28
$\Sigma|\alpha|$ / $\Sigma|\alpha_{ICAPM}|$	1.10	1.08	0.91	1.02	0.97	0.91	0.96	0.93	1.00
%($\Sigma|\alpha_{ICAPM}| < \Sigma|\alpha|$)	(92%)	(84%)	(8%)	(54%)	(42%)	(20%)	(41%)	(29%)	-

PANEL B: Deciles Based on β_{dp} and β_{σ^2} Sorts

	CAPM	FF3	FFC4	FF5	q4	SY4	DHS3	q5	ICAPM
$\Sigma\alpha^2$ / $\Sigma\mathbb{E}[r]^2$	0.05	0.04	0.03	0.06	0.07	0.06	0.09	0.08	0.02
$\Sigma\alpha^2$ / $\Sigma\alpha^2_{ICAPM}$	2.30	1.88	1.14	2.47	2.93	2.54	3.80	3.52	1.00
%($\Sigma\alpha^2_{ICAPM} < \Sigma\alpha^2$)	(98%)	(85%)	(72%)	(92%)	(94%)	(90%)	(97%)	(97%)	-
$\Sigma|\alpha|$ / $\Sigma|\mathbb{E}[r]|$	0.20	0.17	0.14	0.20	0.21	0.21	0.23	0.24	0.12
$\Sigma|\alpha|$ / $\Sigma|\alpha_{ICAPM}|$	1.65	1.41	1.13	1.69	1.77	1.74	1.89	1.97	1.00
%($\Sigma|\alpha_{ICAPM}| < \Sigma|\alpha|$)	(97%)	(83%)	(68%)	(90%)	(92%)	(88%)	(95%)	(95%)	-

Internet Appendix

"An Intertemporal Risk Factor Model"

By Fousseni Chabi-Yo, Andrei S. Gonçalves, and Johnathan Loudis

This Internet Appendix is organized as follows. Section A contains technical derivations required to support the results in the paper. Section B provides some econometric details. Section C describes data sources and measurement for the analysis. Section D describes additional results that supplement the main findings in the paper.

A The ICAPM Derivation

This section derives the ICAPM results we rely on in the main text. Subsection A.1 describes the ICAPM budget constraint, Subsection A.2 derives the general ICAPM SDF, Subsection A.3 specifies investment opportunity dynamics, and Subsection A.4 solves for the ICAPM SDF as a function of our final risk factors.

To simplify notation, we define $r_w \equiv log(R_w)$, use tilde to represent shocks (e.g., $\widetilde{x}_t \equiv x_t - \mathbb{E}_{t-1}[x_t]$), and suppress time subscripts inside first and second moments when convenient (e.g., $\mathbb{E}_t[x] \equiv \mathbb{E}_t[x_{t+1}]$ and $\mathbb{V}ar_t[x] \equiv \mathbb{V}ar_t[x_{t+1}]$).

A long-term (i.e., infinitely lived) investor has Epstein-Zin recursive preferences (Epstein and Zin (1989, 1991) and Weil (1989)) with time discount factor δ, intertemporal elasticity of substitution ψ, and relative risk aversion γ. The investor chooses consumption, C_t, and portfolio allocation, ϖ_t, to maximize lifetime utility subject to the budget constraint $W_{t+1} = (W_t - C_t) \cdot R_{w,t+1}$, with $R_{w,t} = \varpi_t' R_t$ representing the investor's wealth portfolio.

A.1 The ICAPM Budget Constraint

It is instructive to start by rewriting the budget constraint as

$$R_{w,t} \ = \left(\frac{C_{t-1}}{W_{t-1} - C_{t-1}}\right) \cdot \left(\frac{C_t}{C_{t-1}}\right) \cdot \left(\frac{W_t}{C_t}\right), \tag{IA.1}$$

or in logs,

$$r_{w,t} \ = crw_{t-1} + \Delta c_t - cw_t, \tag{IA.2}$$

where $crw_t = -log(e^{-cw_t} - 1)$ is consumption over reinvested wealth.

This alternative way to write the budget constraint demonstrates that shocks to returns on the wealth portfolio can be written as

$$\widetilde{r}_{w,t} \ = \widetilde{\Delta c}_t - \widetilde{cw}_t \ = \ \widetilde{\Delta w}_t. \tag{IA.3}$$

In parts of the derivations, we rely on a log-linear approximation to the consumption-wealth ratio,

$$cw_t \approx k + \overline{\delta} \cdot crw_t \tag{IA.4}$$

which yields the log-linearized budget constraint

$$r_{w,t} \approx -\frac{k}{\overline{\delta}} + \frac{1}{\overline{\delta}} \cdot cw_{t-1} + \Delta c_t - cw_t, \tag{IA.5}$$

where $\overline{\delta} = e^{-\overline{crw}}/(e^{-\overline{crw}} + 1)$ and $k = \overline{\delta} \cdot log(\overline{\delta}) + (1 - \overline{\delta}) \cdot log(1 - \overline{\delta})$ are log-linearization coefficients.[IA.1]

A.2 The ICAPM SDF

This subsection derives the ICAPM SDF

A.2.1 A Quick Derivation for ICAPM SDF Shocks

Start from the well-known form of the Epstein-Zin SDF,

$$SDF_{t+1} = \delta \cdot \left(\frac{C_{t+1}}{C_t}\right)^{-1/\psi} \cdot \left(\frac{V_{t+1}}{\mathbb{E}_t[V_{t+1}^{1-\gamma_t}]^{1/(1-\gamma_t)}}\right)^{-(\gamma-1/\psi)}, \tag{IA.6}$$

and note that Hansen, Heaton, and Li (2008) provide a link between consumption, wealth, and the continuation value function (henceforth "value"),

$$\widetilde{c}_t = \psi \cdot \widetilde{w}_t + (1 - \psi) \cdot \widetilde{v}_t. \tag{IA.7}$$

In this case, shocks to the log SDF are given by:

$$\begin{aligned} \widetilde{sdf}_{t+1} &= -1/\psi \cdot \widetilde{\Delta c}_{t+1} - (\gamma - 1/\psi) \cdot \widetilde{v}_{t+1} \\ &= -\widetilde{w}_{t+1} - (\gamma - 1) \cdot \widetilde{v}_{t+1} \\ &= -\gamma \cdot \widetilde{r}_{w,t+1} - (\gamma - 1) \cdot \widetilde{vw}_{t+1}, \end{aligned} \tag{IA.8}$$

where the second equality uses Equation IA.7 and the third equality uses Equation IA.3.

[IA.1] As we demonstrate below (in Equation IA.15), the optimality conditions yield $cw_t = log(1 - \delta)$ and $crw_t = log((1 - \delta)/\delta)$ if $\psi = 1$, which implies that this log-linear approximation is exact with $\overline{\delta} = \delta$ and $r_{w,t} = \Delta c_t - log(\delta)$ in this case.

In the rest of this subsection we formalize this quick exposition by directly deriving the ICAPM log SDF, $sdf_{t+1} = \frac{\gamma-1}{\bar{\delta}} \cdot \left(vw_t - f(\psi, \delta, \bar{\delta})\right) - \gamma \cdot r_{w,t+1} - (\gamma - 1) \cdot vw_{t+1}$, based on the investor's optimality conditions.

A.2.2 Deriving the ICAPM SDF with $\psi = 1$

With $\psi = 1$, the investor's value function can be written as

$$V(W_t) \; = \underset{\{C_t, \varpi_t\}}{Max} \; C_t^{1-\delta} \cdot \left(\mathbb{E}_t \left[V(W_{t+1})^{1-\gamma}\right]\right)^{\delta/(1-\gamma)}, \tag{IA.9}$$

or in log terms,

$$log(V_t) \; = \underset{\{C_t, \varpi_t\}}{Max} \; (1-\delta) \cdot log(C_t) \; + \; \frac{\delta}{(1-\gamma)} log\left(\mathbb{E}_t\left[V_{t+1}^{1-\gamma}\right]\right), \tag{IA.10}$$

where the second equation simplifies the notation by suppressing the dependence of the value function on wealth.

The consumption first order condition (FOC) then yields:

$$\frac{(1-\delta)}{C_t} \; = \frac{\delta}{\mathbb{E}_t[V_{t+1}^{1-\gamma}]} \mathbb{E}_t[V_{t+1}^{-\gamma} \cdot \partial_W V_{t+1} \cdot R_{w,t+1}], \tag{IA.11}$$

and the Benveniste and Scheinkman (1979) condition relative to wealth implies

$$\partial_W log(V_t) \; = \; \frac{\partial_W V_t}{V_t} \; = \frac{\delta}{\mathbb{E}_t[V_{t+1}^{1-\gamma}]} \mathbb{E}_t[V_{t+1}^{-\gamma} \cdot \partial_W V_{t+1} \cdot R_{w,t+1}], \tag{IA.12}$$

so that combining the two optimality conditions gives

$$\partial_W V_t \; = (1-\delta) \cdot \frac{V_t}{C_t}. \tag{IA.13}$$

Equation IA.13 is the main optimality condition we need to derive the ICAPM SDF. To do so, start by conjecturing that $V(W_t)$ is homogeneous of degree one (i.e., V_t/W_t is not a function of wealth). This conjecture implies that $\partial_W V_t = V_t/W_t$, which, after substituting into Equation IA.13 and using $crw_t = -log(e^{-cw_t} - 1)$, yields

$$C_t/W_t \; = (1-\delta), \tag{IA.14}$$

$$cw_t = log(1-\delta), \qquad \text{and} \qquad crw_t = log((1-\delta)/\delta), \tag{IA.15}$$

so that Equation IA.7 holds when $\psi = 1$.

Now, rewrite the objective function (i.e., Equation IA.10) as

$$\begin{aligned} vw_t &= (1-\delta)\cdot cw_t \;-\; \delta\cdot w_t \;+\; \frac{\delta}{(1-\gamma)} log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot v_{t+1}}\right]\right) \\ &= (1-\delta)\cdot cw_t \;+\; \frac{\delta}{(1-\gamma)} log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot v_{t+1}-(1-\gamma)\cdot w_t}\right]\right) \\ &= (1-\delta)\cdot cw_t \;+\; \frac{\delta}{(1-\gamma)} log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot(log(\delta)+vw_{t+1}+r_{w,t+1})}\right]\right) \\ &= (1-\delta)\cdot log(1-\delta) \;+\; \frac{\delta}{(1-\gamma)} log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot(log(\delta)+vw_{t+1}+r_{w,t+1})}\right]\right), \end{aligned} \tag{IA.16}$$

where the third equality relies on the budget constraint (Equation IA.2) and the last equality uses Equation IA.15.

Equation IA.16 represents a recursion for vw_t that shows that if vw_{t+1} does not depend on wealth, then vw_t also does not. As such, the conjecture that $V(W_t)$ is homogeneous of degree one is valid.

We can further work on Equation IA.16 to get

$$\begin{aligned} log(e^{vw_t}) &= log\left((1-\delta)^{(1-\delta)}\right) \;+\; log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot(log(\delta)+vw_{t+1}+r_{w,t+1})}\right]^{\delta/(1-\gamma)}\right) \\ &= log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot(\frac{1-\delta}{\delta}\cdot log(1-\delta)+log(\delta)+vw_{t+1}+r_{w,t+1})}\right]^{\delta/(1-\gamma)}\right) \\ &\Downarrow \\ 1 &= \mathbb{E}_t\left[e^{(1-\gamma)\cdot(\frac{1-\delta}{\delta}\cdot log(1-\delta)+log(\delta)-\frac{1}{\delta}vw_t+vw_{t+1}+r_{w,t+1})}\right] \\ &= \mathbb{E}_t\left[e^{\frac{(\gamma-1)}{\delta}\cdot[vw_t-f_0(\delta,\delta)]\;-\;\gamma\cdot r_{w,t+1}\;-\;(\gamma-1)\cdot vw_{t+1}}\cdot R_{w,t+1}\right], \end{aligned} \tag{IA.17}$$

where $f_0(z,y) = (1-y)\cdot log(1-z) + y\cdot log(z)$.

Now, rewrite the budget constraint as $W_{t+1} = (W_t - C_t)\cdot(R_{f,t+1} + \varpi_t^{'}(R_{t+1} - R_{f,t+1}))$

and substitute it in $V(W_{t+1})$ so that the FOC with respect to ϖ_t yields

$$
\begin{aligned}
0 &= \mathbb{E}_t\left[\delta \cdot \left(\frac{C_{t+1}}{C_t}\right)^{-1} \cdot \left(\frac{V_{t+1}}{E_t[V_{t+1}^{1-\gamma}]^{1/(1-\gamma)}}\right)^{-(\gamma-1)} \cdot (R_{t+1} - R_{f,t+1})\right] \\
&= \mathbb{E}_t\left[e^{-\widetilde{\Delta c}_{t+1} \,-\, (\gamma-1)\cdot \widetilde{v}_{t+1}} \cdot (R_{t+1} - R_{f,t+1})\right] \\
&= \mathbb{E}_t\left[e^{-\gamma\cdot \widetilde{r}_{w,t+1} \,-\, (\gamma-1)\cdot \widetilde{vw}_{t+1}} \cdot (R_{t+1} - R_{f,t+1})\right] \\
&= \mathbb{E}_t\left[e^{\frac{(\gamma-1)}{\delta}\cdot[vw_t - f_0(\delta,\delta)] \,-\gamma\cdot r_{w,t+1} \,-\, (\gamma-1)\cdot vw_{t+1}} \cdot (R_{t+1} - R_{f,t+1})\right],
\end{aligned} \tag{IA.18}
$$

where the third equality is based on the same derivation as in Equation IA.8 and the second and fourth equalities use the fact that we can multiply any arbitrary variable known as of time t on both sides of this FOC.

Equations IA.17 and IA.18 jointly imply that the ICAPM SDF, given by

$$
sdf_{t+1} \;=\; \frac{(\gamma-1)}{\delta} \cdot [vw_t - f_0(\delta,\delta)] \;-\; \gamma \cdot r_{w,t+1} \;-\; (\gamma-1) \cdot vw_{t+1}, \tag{IA.19}
$$

prices all assets available to the long-term investor.

A.2.3 Deriving the ICAPM SDF with $\psi \neq 1$

With $\psi \neq 1$, the investor's value function can be written as

$$
\underset{\{C_t,\varpi_t\}}{Max} \left\{(1-\delta)\cdot C_t^{1-1/\psi} \;+\; \delta \cdot \mathbb{E}_t\left[V(W_{t+1})^{1-\gamma}\right]^{\frac{1-1/\psi}{1-\gamma}}\right\}^{1/(1-1/\psi)}. \tag{IA.20}
$$

The consumption FOC then yields

$$
(1-\delta)\cdot C_t^{-1/\psi} \;= \delta \cdot \mathbb{E}_t[V_{t+1}^{1-\gamma}]^{\frac{\gamma-1/\psi}{1-\gamma}} \cdot \mathbb{E}_t[V_{t+1}^{-\gamma} \cdot \partial_W V_{t+1} \cdot R_{w,t+1}], \tag{IA.21}
$$

and the Benveniste and Scheinkman (1979) condition relative to wealth implies

$$
\partial_W V_t \;= V_t^{1/\psi} \cdot \delta \cdot \mathbb{E}_t[V_{t+1}^{1-\gamma}]^{\frac{\gamma-1/\psi}{1-\gamma}} \cdot \mathbb{E}_t[V_{t+1}^{-\gamma} \cdot \partial_W V_{t+1} \cdot R_{w,t+1}], \tag{IA.22}
$$

so that combining the two optimality conditions gives

$$
\partial_W V_t \;= (1-\delta) \cdot \left(\frac{V_t}{C_t}\right)^{1/\psi} \tag{IA.23}
$$

Equation IA.23 is the main optimality condition we need to derive the ICAPM SDF. To do so, start by conjecturing that $V(W_t)$ is homogeneous of degree one (i.e., V_t/W_t is not a function of wealth). This conjecture implies that $\partial_W V_t = V_t/W_t$, which, after substituting into Equation IA.23, yields:

$$(V_t/W_t)^{1-1/\psi} = (1-\delta)\cdot(C_t/W_t)^{-1/\psi} \tag{IA.24}$$

$$\Downarrow$$

$$cw_t = \psi\cdot log(1-\delta) + (1-\psi)\cdot vw_t, \tag{IA.25}$$

so that Equation IA.7 also holds when $\psi \neq 1$.

Now, rewrite the objective function (i.e., Equation IA.20) as:

$$(W_t\cdot V_t/W_t)^{1-1/\psi} = (1-\delta)\cdot W_t^{1-1/\psi}\cdot(C_t/W_t)^{1-1/\psi} + \delta\cdot\mathbb{E}_t\left[W_{t+1}^{1-\gamma}\left(\frac{V_{t+1}}{W_{t+1}}\right)^{1-\gamma}\right]^{\frac{1-1/\psi}{1-\gamma}}$$

$$\Downarrow$$

$$\begin{aligned}(V_t/W_t)^{1-1/\psi} &= (1-\delta)\cdot(C_t/W_t)^{1-1/\psi} + \delta\cdot\mathbb{E}_t\left[\left(\frac{W_{t+1}}{W_t}\right)^{1-\gamma}\left(\frac{V_{t+1}}{W_{t+1}}\right)^{1-\gamma}\right]^{\frac{1-1/\psi}{1-\gamma}} \\ &= C_t/W_t\cdot(V_t/W_t)^{1-1/\psi} + \delta\cdot\mathbb{E}_t\left[\left(1-\frac{C_t}{W_t}\right)^{1-\gamma}\left(\frac{V_{t+1}}{W_{t+1}}\right)^{1-\gamma}R_{w,t+1}^{1-\gamma}\right]^{\frac{1-1/\psi}{1-\gamma}} \\ &= \delta\cdot\mathbb{E}_t\left[(1-C_t/W_t)^{1-\frac{1}{\psi}}\left(\frac{V_{t+1}}{W_{t+1}}\right)^{1-\gamma}R_{w,t+1}^{1-\gamma}\right]^{\frac{1-1/\psi}{1-\gamma}},\end{aligned} \tag{IA.26}$$

where the third equality relies on Equation IA.24 and the budget constraint (Equation IA.1).

Equation IA.26 represents a recursion for V_t/W_t that shows that if V_{t+1}/W_{t+1} does not depend on wealth, then V_t/W_t also does not.[IA.2] As such, the conjecture that $V(W_t)$ is homogeneous of degree one is valid.

We can further work on Equation IA.26 to get

[IA.2]The recursion in Equation IA.26 also depends on C_t/W_t. However, Equation IA.24 shows that C_t/W_t is a function of V_t/W_t, and thus the recursion implies that V_t/W_t is a function of the distribution of $[V_{t+1}/W_{t+1}, R_{w,t+1}]$, which does not depend on the wealth level from the perspective of the representative investor.

$$1 = \mathbb{E}_t\left[\left\{\delta^{\frac{1-\gamma}{1-1/\psi}} \cdot (1 - C_t/W_t)^{1-\frac{1}{\psi}} \cdot R_{w,t+1}^{-\gamma} \cdot \left(\frac{V_{t+1}/W_{t+1}}{V_t/W_t}\right)^{-(\gamma-1)}\right\} \cdot R_{w,t+1}\right]$$

$$= \mathbb{E}_t\left[e^{f_{sdf}(\psi,\delta,\gamma,cw_t) - \gamma \cdot r_{w,t+1} - (\gamma-1)\cdot vw_{t+1}} \cdot R_{w,t+1}\right], \tag{IA.27}$$

where $f_{sdf}(\psi, \delta, \gamma, cw_t) = (\gamma - 1) \cdot \left(vw_t + \frac{1}{\psi-1} \cdot [cw_t - crw_t] - \frac{1}{1-1/\psi} \cdot log(\delta)\right)$ is implicitly defined in Equation IA.27.

Now, rewrite the budget constraint as $W_{t+1} = (W_t - C_t) \cdot (R_{f,t+1} + \varpi_t'(R_{t+1} - R_{f,t+1}))$ and substitute it in $V(W_{t+1})$ so that the FOC with respect to ϖ_t yields

$$0 = \mathbb{E}_t\left[\delta \cdot \left(\frac{C_{t+1}}{C_t}\right)^{-1/\psi} \cdot \left(\frac{V_{t+1}}{E_t[V_{t+1}^{1-\gamma}]^{1/(1-\gamma)}}\right)^{-(\gamma-1/\psi)} \cdot (R_{t+1} - R_{f,t+1})\right]$$

$$= \mathbb{E}_t\left[e^{-1/\psi\cdot\widetilde{\Delta c}_{t+1} - (\gamma-1/\psi)\cdot\widetilde{v}_{t+1}} \cdot (R_{t+1} - R_{f,t+1})\right]$$

$$= \mathbb{E}_t\left[e^{-\gamma\cdot\widetilde{r}_{w,t+1} - (\gamma-1)\cdot\widetilde{vw}_{t+1}} \cdot (R_{t+1} - R_{f,t+1})\right]$$

$$= \mathbb{E}_t\left[e^{f_{sdf}(\psi,\delta,\gamma,cw_t) - \gamma\cdot r_{w,t+1} - (\gamma-1)\cdot vw_{t+1}} \cdot (R_{t+1} - R_{f,t+1})\right], \tag{IA.28}$$

where the third equality is based on the same derivation as in Equation IA.8 and the second and fourth equalities use the fact that we can multiply any arbitrary variable known as of time t on both sides of this FOC.

Equations IA.27 and IA.28 jointly imply that the ICAPM SDF, given by

$$sdf_{t+1} = f_{sdf}(\psi, \delta, \gamma, cw_t) - \gamma \cdot r_{w,t+1} - (\gamma - 1) \cdot vw_{t+1}, \tag{IA.29}$$

prices all assets available to the long-term investor.

A.2.4 The General ICAPM SDF

As Equations IA.19 and IA.29 demonstrate, the SDF shocks with Epstein-Zin preferences can be written as

$$\widetilde{sdf}_{t+1} = -\gamma \cdot \widetilde{r}_{w,t+1} - (\gamma - 1) \cdot \widetilde{vw}_{t+1}. \tag{IA.30}$$

The SDF level is more complicated due to the nonlinear $f_{sdf}(\psi, \delta, \gamma, cw_t)$ function. However, we can simplify this function to

$$
\begin{aligned}
f_{sdf}(\psi,\delta,\gamma,cw_t) &= (\gamma-1)\cdot\left(vw_t + \frac{1}{\psi-1}\cdot[cw_t - crw_t] - \frac{1}{1-1/\psi}\cdot log(\delta)\right) \\
&\approx (\gamma-1)\cdot\left(vw_t + \frac{1}{\psi-1}\cdot\left[cw_t - \left(-\frac{k}{\overline{\overline{\delta}}} + \frac{1}{\overline{\overline{\delta}}}\cdot cw_t\right)\right] - \frac{1}{1-1/\psi}\cdot log(\delta)\right) \\
&= \frac{\gamma-1}{\overline{\delta}}\cdot\left(vw_t - \frac{1}{\psi-1}\left[\psi\cdot f_0(\delta,\overline{\delta}) - f_0(\overline{\delta},\overline{\delta})\right]\right) \\
&= \frac{\gamma-1}{\overline{\delta}}\cdot\left(vw_t - f_0(\delta,\overline{\delta}) - \frac{1}{\psi-1}\cdot\left[f_0(\delta,\overline{\delta}) - f_0(\overline{\delta},\overline{\delta})\right]\right), \qquad \text{(IA.31)}
\end{aligned}
$$

where $f_0(z,y) = (1-y)\cdot log(1-z) + y\cdot log(z)$, with the second equality relying on the log-linear approximation to cw_t in Campbell (1993) (Equation IA.4), which is exact if $\psi = 1$, and the third equality using Equation IA.25.

As such, the ICAPM log SDF can be written as

$$
\begin{aligned}
sdf_{t+1} &= \frac{\gamma-1}{\overline{\delta}}\cdot\left(vw_t - f(\psi,\delta,\overline{\delta})\right) \;-\; \gamma\cdot r_{w,t+1} \;-\; (\gamma-1)\cdot vw_{t+1} \\
&= \kappa_t \;-\; \gamma\cdot r_{w,t+1} \;-\; (\gamma-1)\cdot\widetilde{vw}_{t+1} \qquad \text{(IA.32)}
\end{aligned}
$$

where

$$
\kappa_t = (\gamma-1)\cdot\left(vw_t/\overline{\delta} - \mathbb{E}_t[vw] - f(\psi,\delta,\overline{\delta})/\overline{\delta}\right) \qquad \text{(IA.33)}
$$

and

$$
f(\psi,\delta,\overline{\delta}) = \begin{cases} f_0(\delta,\delta) & if\ \psi = 1 \\ f_0(\delta,\overline{\delta}) + \frac{1}{\psi-1}\cdot\left[f_0(\delta,\overline{\delta}) - f_0(\overline{\delta},\overline{\delta})\right] & if\ \psi \neq 1. \end{cases} \qquad \text{(IA.34)}
$$

A.2.5 The ICAPM SDF with $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$

The asset pricing Euler condition, $\mathbb{E}_t[SDF_{t+1} \cdot R_{w,t+1}] = 1$, can be written as

$$\begin{aligned} 0 &= \mathbb{E}_t\left[sdf_{t+1} + r_{w,t+1}\right] \; + \; log\left(\mathbb{E}_t\left[e^{\widetilde{sdf}_{t+1} + \widetilde{r}_{w,t+1}}\right]\right) \\ &= \mathbb{E}_t\left[sdf_{t+1} + r_{w,t+1}\right] \; + \; log\left(\mathbb{E}_t\left[e^{(1-\gamma)\cdot \widetilde{v}_{t+1}}\right]\right) \\ &\approx \mathbb{E}_t\left[sdf_{t+1} + r_{w,t+1}\right] \; + \; \frac{(\gamma-1)^2}{2} \cdot \mathbb{V}ar_t\left[v_{t+1}\right] \end{aligned} \tag{IA.35}$$

where the second line uses the identity $\widetilde{v}_t = \widetilde{r}_{w,t} + \widetilde{vw}_t$ and the third line relies on a second order Taylor approximation that hold exactly if $\widetilde{v}_t$ is conditionally normal.[IA.3]

Then, substituting the sdf_t from Equation IA.32 into Equation IA.35, results in

$$vw_t \; = f(\psi, \delta, \overline{\delta}) \; + \overline{\delta} \cdot \mathbb{E}_t\left[r_{w,t+1}\right] \; + \; \overline{\delta} \cdot \mathbb{E}_t\left[vw_{t+1}\right] \; - \; \overline{\delta} \cdot \frac{(\gamma-1)}{2} \cdot \mathbb{V}ar_t\left[v_{t+1}\right] \tag{IA.36}$$

and a recursive substitution of this equation yields

$$vw_t \; = \frac{f(\psi, \delta, \overline{\delta})}{1-\overline{\delta}} \; + \; \mathbb{E}_t\left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot r_{w,t+h}\right] - \frac{(\gamma-1)}{2} \cdot \mathbb{E}_t\left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{V}ar_{t+h-1}\left[v_{t+h}\right]\right] \tag{IA.37}$$

which implies

$$\widetilde{vw}_{t+1} \; = N_{\mathbb{E},t+1} \; - \; \frac{(\gamma-1)}{2} \cdot N_{\mathbb{V},t+1} \tag{IA.38}$$

where

$N_{\mathbb{E},t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t)\left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot r_{w,t+h}\right]$ is expected return news

[IA.3] To derive the second order Taylor approximation in Equation IA.35, let x_t be a random variable with $\mathbb{E}_t[x_{t+1}] = 0$ (i.e., a shock), then $log(\mathbb{E}_t[e^{x_t}])$ is the conditional entropy of x_t, which is equivalent to $\mathbb{K}(1)$, where $\mathbb{K}(\epsilon) = log(\mathbb{E}_t[e^{\epsilon \cdot x_{t+1}}])$ is the cumulant-generating function of x_t. Then, a second order taylor expansion of $\mathbb{K}(\epsilon)$ around $\epsilon = 0$ yields

$$\begin{aligned} \mathbb{K}(\epsilon) \; &\approx \; \{\partial_\epsilon \mathbb{E}_t[e^{\epsilon \cdot x_{t+1}}]\}_{\epsilon=0} \cdot (\epsilon - 0) \; + \; \frac{1}{2} \cdot \{\partial_{\epsilon\epsilon} \mathbb{E}_t[e^{\epsilon \cdot x_{t+1}}]\}_{\epsilon=0} \cdot (\epsilon - 0)^2 \\ &= \left\{\frac{\mathbb{E}_t[e^{\epsilon \cdot x_{t+1}} \cdot x_{t+1}]}{\mathbb{E}_t[e^{\epsilon \cdot x_{t+1}}]}\right\}_{\epsilon=0} \cdot \epsilon \; + \; \frac{1}{2} \cdot \left\{\frac{\mathbb{E}_t[e^{\epsilon \cdot x_{t+1}} \cdot x_{t+1}^2] \cdot \mathbb{E}_t[e^{\epsilon \cdot x_{t+1}}] - \mathbb{E}_t[e^{\epsilon \cdot x_{t+1}} \cdot x_{t+1}] \cdot \mathbb{E}_t[e^{\epsilon \cdot x_{t+1}} \cdot x_{t+1}]}{\mathbb{E}_t[e^{\epsilon \cdot x_{t+1}}]^2}\right\}_{\epsilon=0} \cdot \epsilon^2 \\ &= \frac{1}{2} \cdot \mathbb{V}ar_t[x_{t+1}] \cdot \epsilon^2 \end{aligned}$$

so that setting $x_t = (1-\gamma) \cdot \widetilde{v}_t$ implies that the asset pricing Euler condition can be written as $\mathbb{K}(1) = 0$, which results in the approximation in Equation IA.35. Note that if $\widetilde{r}_{w,t}$ and $\widetilde{vw}_t$ are conditionally joint normal (so that $\widetilde{v}_t$ is conditionally normal), then this approximation is exact.

$N_{\mathbb{V},t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{V}ar_{t+h}[v_{t+1+h}] \right]$ is volatility news

Then, substituting $\widetilde{vw}_{t+1}$ into the SDF from Equation IA.32 yields

$$\begin{aligned} sdf_{t+1} &= \kappa_t \; - \; \gamma \cdot r_{w,t+1} \; - \; (\gamma - 1) \cdot \widetilde{vw}_{t+1} \\ &= \kappa_t \; - \; \gamma \cdot r_{w,t+1} \; - \; (\gamma - 1) \cdot N_{\mathbb{E},t+1} \; + \; \frac{(\gamma-1)^2}{2} \cdot N_{\mathbb{V},t+1} \end{aligned} \tag{IA.39}$$

In the main text, we simplify the exposition by assuming $\psi = 1$ and $\mathbb{E}_t[vw] \approx vw_t/\delta$ so that $\overline{\delta} = \delta$ and $\kappa_t = \kappa = (\gamma - 1) \cdot ((1 - 1/\delta) \cdot log(1 - \delta) - log(\delta))$. The small time varying component that we effectively ignore (i.e., $\kappa_t - \kappa$) has no implications for risk premia (only for interest rate variation), and thus does not play a role in our analysis.

A.3 The Wealth Return Dynamics

Our ICAPM empirical implementation requires a map from $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ to observable variables. To avoid the ICAPM "fishing license" (Fama (1991)) when constructing $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$, we do not specify an arbitrary set of state variables for the wealth return dynamics, but instead build a simple Bayesian learning framework in which a long-term investor observes only market prices and dividends to derive a belief about the wealth return dynamics (i.e., expected returns and volatility). Our framework can be thought of as an extension of Binsbergen and Koijen (2010) in the sense that we add time-varying volatility to their framework.

A.3.1 The Underlying Environment

Letting $r_{w,t}$ and Δd_t reflect the monthly log wealth return and growth in annual dividends, we assume that $r_{w,t}$, $\mu_t = \mathbb{E}_t[r_{w,t+1}|\mu_t]$, and $g_t = \mathbb{E}_t[\Delta d_{t+1}|g_t]$ have monthly dynamics given by

$$r_{w,t+1} = \mu_t \; + \; \widetilde{r}^*_{w,t+1} \tag{IA.40}$$

$$\mu_{t+1} = \mu \; + \; \phi_\mu \cdot (\mu_t - \mu) \; + \; \widetilde{\mu}^*_{t+1} \tag{IA.41}$$

$$g_{t+1} = g \; + \; \phi_g \cdot (g_t - g) \; + \; \widetilde{g}^*_{t+1} \tag{IA.42}$$

$$\Sigma_t = \Sigma_{\mathbb{V}} \cdot \mathbb{V}r_t \tag{IA.43}$$

where $[\widetilde{r}^*_{w,t+1}, \widetilde{\mu}^*_{t+1}, \widetilde{g}^*_{t+1}] \sim N(0, \Sigma_t)$ are unobservable shocks and $\mathbb{V}r_t = \mathbb{V}ar_t[r^*_{w,t+1}]$ captures the wealth portfolio return variance with dynamics detailed in Subsection A.3.5.

Moreover, we parametrize $\Sigma_{\mathbb{V}}$ such that it is positive definite, which assures a positive semidefinite Σ_t for all t. The $\Sigma_{\mathbb{V}}$ matrices is otherwise fully flexible with the following parameters:[IA.4]

$$\Sigma_{\mathbb{V}} = \begin{bmatrix} 1 & \nu_{r,\mu} & \nu_{r,g} \\ \nu_{r,\mu} & \nu_\mu^2 & \nu_{\mu,g} \\ \nu_{r,g} & \nu_{\mu,g} & \nu_g^2 \end{bmatrix} \tag{IA.44}$$

A.3.2 The μ_t Signals

From the log-linear valuation identity in Campbell and Shiller (1989), we have

$$\begin{aligned} dp_t - dp &= \sum_{h=0}^{\infty} \rho^h \cdot \mathbb{E}_t \left[\Sigma_{j=1}^{12}(r_{w,t+12\cdot h+j} - \mu)|\mu_t\right] - \sum_{h=0}^{\infty} \rho^h \cdot \mathbb{E}_t \left[\Sigma_{j=1}^{12}(\Delta d_{t+12\cdot h+j} - g)|g_t\right] \\ &= \frac{(1-\phi_\mu^{12})}{(1-\phi_\mu)} \cdot \sum_{h=0}^{\infty} \rho^h \cdot \mathbb{E}_t[\mu_{t+12\cdot h} - \mu] - \frac{(1-\phi_g^{12})}{(1-\phi_g)} \cdot \sum_{h=0}^{\infty} \rho^h \cdot \mathbb{E}_t[g_{t+12\cdot h} - g] \\ &= \frac{(1-\phi_\mu^{12})}{(1-\phi_\mu)} \cdot \sum_{h=0}^{\infty} \rho^h \cdot (\phi_\mu^{12})^h \cdot (\mu_t - \mu) - \frac{(1-\phi_g^{12})}{(1-\phi_g)} \cdot \sum_{h=0}^{\infty} \rho^h \cdot (\phi_g^{12})^h \cdot (g_t - g) \\ &= \frac{(1-\phi_\mu^{12})}{(1-\phi_\mu)\cdot(1-\rho\cdot\phi_\mu^{12})} \cdot (\mu_t - \mu) - \frac{(1-\phi_g^{12})}{(1-\phi_g)\cdot(1-\rho\cdot\phi_g^{12})} \cdot (g_t - g) \\ &= \Phi_\mu \cdot (\mu_t - \mu) - \Phi_g \cdot (g_t - g) \end{aligned} \tag{IA.45}$$

where $\Phi_\mu = (1-\phi_\mu^{12})/[(1-\phi_\mu)(1-\rho\cdot\phi_\mu^{12})]$, $\Phi_g = (1-\phi_g^{12})/[(1-\phi_g)(1-\rho\cdot\phi_g^{12})]$, and $\rho = e^{-dp}/(1+e^{-dp})$, with $dp_t = log(D_t/P_t)$ reflecting the annual dividend yield and dp capturing the average dp_t.

Moreover, letting $\widetilde{dp}^*_t = \Phi_\mu \cdot \widetilde{\mu}^*_t - \Phi_g \cdot \widetilde{g}^*_t$ reflect the monthly unobservable shock to dp_t,

[IA.4] In our estimation detailed in Subsection B.1.2, we search only over the space of parameters that yield a positive definite $\Sigma_{\mathbb{V}}$.

we can obtain

$$\mathbb{V}ar_t[\widetilde{dp}^*_{t+1}] = \nu^2_{dp} \cdot \mathbb{V}r_t \tag{IA.46}$$

$$\mathbb{C}ov_t[\widetilde{r}^*_{w,t+1}, \widetilde{dp}^*_{t+1}] = \nu_{r,dp} \cdot \mathbb{V}r_t \tag{IA.47}$$

$$\mathbb{C}ov_t[\widetilde{\mu}^*_{t+1}, \widetilde{dp}^*_{t+1}] = \nu_{\mu,dp} \cdot \mathbb{V}r_t \tag{IA.48}$$

$$\mathbb{C}ov_t[\widetilde{g}^*_{t+1}, \widetilde{dp}^*_{t+1}] = \nu_{g,dp} \cdot \mathbb{V}r_t \tag{IA.49}$$

where

$\nu_{dp} = \sqrt{\Phi^2_\mu \cdot \nu^2_\mu + \Phi^2_g \cdot \nu^2_g - 2 \cdot \Phi_\mu \cdot \Phi_g \cdot \nu_{\mu,g}}$

$\nu_{r,dp} = \Phi_\mu \cdot \nu_{r,\mu} - \Phi_g \cdot \nu_{r,g}$

$\nu_{\mu,dp} = \Phi_\mu \cdot \nu^2_\mu - \Phi_g \cdot \nu_{\mu,g}$

$\nu_{g,dp} = \Phi_\mu \cdot \nu_{\mu,g} - \Phi_g \cdot \nu^2_g$

Repeating Equation IA.40, we have

$$r_{w,t+1} = \mu_t + \widetilde{r}^*_{w,t+1} \tag{IA.50}$$

Moreover, isolating g_t in Equation IA.45 and substituting into Equation IA.42 yields[IA.5]

$$dp^o_{t+1} = \mu_t + \nu_o \cdot \widetilde{dp}^*_{t+1} \tag{IA.51}$$

[IA.5] To derive Equation IA.51, note that Equation IA.45 implies

$$(g_t - g) = (\Phi_\mu/\Phi_g) \cdot (\mu_t - \mu) - (1/\Phi_g) \cdot (dp_t - dp)$$

which we can substitute into Equation IA.42 to get

$$\begin{aligned}
(\Phi_\mu/\Phi_g) \cdot (\mu_{t+1} - \mu) - (1/\Phi_g) \cdot (dp_{t+1} - dp) &= \phi_g \cdot [(\Phi_\mu/\Phi_g) \cdot (\mu_t - \mu) - (1/\Phi_g) \cdot (dp_t - dp)] + \widetilde{g}^*_{t+1} \\
&\Downarrow \\
\Phi_\mu \cdot \left[\phi_\mu \cdot (\mu_t - \mu) + \widetilde{\mu}^*_{t+1}\right] - (dp_{t+1} - dp) &= \phi_g \cdot [\Phi_\mu \cdot (\mu_t - \mu) - (dp_t - dp)] + \Phi_g \cdot \widetilde{g}^*_{t+1} \\
&\Downarrow \\
\Phi_\mu \cdot (\phi_g - \phi_\mu) \cdot (\mu_t - \mu) + (dp_{t+1} - dp) - \phi_g \cdot (dp_t - dp) &= \Phi_\mu \cdot \widetilde{\mu}^*_{t+1} - \Phi_g \cdot \widetilde{g}^*_{t+1} \\
&\Downarrow \\
\mu + \nu_o \cdot [(dp_{t+1} - dp) - \phi_g \cdot (dp_t - dp)] &= \mu_t + \nu_o \cdot \widetilde{dp}^*_{t+1}
\end{aligned}$$

where $\nu_o = 1/[\Phi_\mu \cdot (\phi_\mu - \phi_g)]$ and

$$dp^o_{t+1} \;= \mu \;+\; \nu_o \cdot [(dp_{t+1} - dp) - \phi_g \cdot (dp_t - dp)] \tag{IA.52}$$

Equations IA.50 and IA.51 represent two μ_t signals that the investor observes at time $t+1$, and thus can use when forming the endogenous expected wealth return process, $\mathbb{E}r_t = \mathbb{E}_t[r_{w,t+1}] = \mathbb{E}_t[\mu_t]$.

A.3.3 Some Auxiliary Results

We use the following lemma concerning jointly normal vectors in our derivations (see Section 4.2 of Rencher (2002)):

Lemma 1. *If*

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{bmatrix}, \begin{bmatrix} \mathbb{V}ar[X_1] & \mathbb{C}ov[X_1, X_2]' \\ \mathbb{C}ov[X_1, X_2] & \mathbb{V}ar[X_2] \end{bmatrix}\right) \tag{IA.53}$$

then

$$\mathbb{E}[X_1|X_2 = x_2] \;= \mathbb{E}[X_1] \;+\; \mathbb{C}ov[X_1, X_2]'\mathbb{V}ar[X_2]^{-1} \cdot (x_2 - \mathbb{E}[X_2]) \tag{IA.54}$$

$$\mathbb{V}ar[X_1|X_2 = x_2] \;= \mathbb{V}ar[X_1] \;-\; \mathbb{C}ov[X_1, X_2]'\mathbb{V}ar[X_2]^{-1}\mathbb{C}ov[X_1, X_2] \tag{IA.55}$$

We also rely on a simple result about the expectation operator. Specifically, letting

$$\begin{aligned} \widetilde{dp}^o_{t+1} \;&\equiv dp^o_{t+1} - \mathbb{E}r_t \\ &= \mu_t - \mathbb{E}r_t \;+\; \nu_o \cdot \widetilde{dp}^*_{t+1} \end{aligned} \tag{IA.56}$$

and

$$\begin{aligned} \widetilde{r}_{w,t+1} \;&\equiv r_{w,t+1} - \mathbb{E}r_t \\ &= \mu_t - \mathbb{E}r_t \;+\; \widetilde{r}^*_{w,t+1} \end{aligned} \tag{IA.57}$$

so that $\mathbb{E}_t[\widetilde{dp}^o_{t+1}] = 0$ and $\mathbb{E}_t[\widetilde{r}_{w,t+1}] = 0$, and defining the $\mathcal{F}_t$ subset of the investor's information set such that $\mathcal{F}_{t+1} = \{\mathcal{F}_t \cup (\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1})\}$, we have

$$\begin{aligned}
\mathbb{E}_{t+1}[f(\mu_{t+1})] &= \mathbb{E}_t[f(\mu_{t+1})|\mathcal{F}_{t+1}] \\
&= \mathbb{E}_t[f(\mu_{t+1})|\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}, \mathcal{F}_t] \\
&= \mathbb{E}_t[f(\mu_{t+1})|\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}] + \mathbb{E}_t[f(\mu_{t+1})|\mathcal{F}_t] - \mathbb{E}_t[f(\mu_{t+1})] \\
&= \mathbb{E}_t[f(\mu_{t+1})|\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}]
\end{aligned} \tag{IA.58}$$

for any arbitrary function $f(\mu_{t+1})$. The first line follows from the fact that the investor updates the μ_t belief based only on the signals summarized by the $\mathcal{F}_t$ information set and the third line follows from Theorem 2.4 in Chapter 5 of Anderson and Moore (1979).

Given Lemma 1 and Equation IA.58, so long as we can show that the joint distribution of μ_{t+1}, $\widetilde{dp}^o_{t+1}$, and $\widetilde{r}_{t+1}$ conditioned on time t information is normal, we can obtain $\mathbb{E}r_{t+1} = \mathbb{E}_{t+1}[\mu_{t+1}] = \mathbb{E}_t[\mu_{t+1}|\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}]$ and $\mathbb{U}\mu_{t+1} = \mathbb{V}ar_{t+1}[\mu_{t+1}] = \mathbb{V}ar_t[\mu_{t+1}|\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}]$.

From the dynamics of μ_t, $\widetilde{dp}^o_t$, and $\widetilde{r}_t$ in Equations IA.41, IA.56, and IA.57, we have:

$$\begin{bmatrix} \mu_{t+1} \\ \widetilde{dp}^o_{t+1} \\ \widetilde{r}_{w,t+1} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbb{E}_t[\mu_{t+1}] \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbb{V}ar_t[\mu_{t+1}] & \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] & \mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \\ \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] & \mathbb{V}ar_t[\widetilde{dp}^o_{t+1}] & \mathbb{C}ov_t[\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}] \\ \mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] & \mathbb{C}ov_t[\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}] & \mathbb{V}ar_t[\widetilde{r}_{w,t+1}] \end{bmatrix} \right) \tag{IA.59}$$

where (with $\mathbb{V}r_t = \mathbb{V}ar_t[\widetilde{r}^*_{w,t+1}]$ and $\mathbb{U}\mu_t = \mathbb{V}ar_t[\mu_t]$)

$$\begin{aligned}
\mathbb{E}_t[\mu_{t+1}] &= \mathbb{E}_t[\mu \ + \ \phi_\mu \cdot (\mu_t - \mu) \ + \ \widetilde{\mu}^*_{t+1}] \\
&= \mu \ + \ \phi_\mu \cdot (\mathbb{E}r_t - \mu)
\end{aligned} \tag{IA.60}$$

$$\begin{aligned}
\mathbb{V}ar_t[\widetilde{r}_{w,t+1}] &= \mathbb{V}ar_t[\mu_t - \mathbb{E}r_t \ + \ \widetilde{r}^*_{w,t+1}] \\
&= \mathbb{U}\mu_t \ + \ \mathbb{V}r_t
\end{aligned} \tag{IA.61}$$

$$\begin{aligned}
\mathbb{V}ar_t[\mu_{t+1}] &= \mathbb{V}ar_t[\mu \ + \ \phi_\mu \cdot (\mu_t - \mu) \ + \ \widetilde{\mu}^*_{t+1}] \\
&= \phi^2_\mu \cdot \mathbb{U}\mu_t \ + \ \nu^2_\mu \cdot \mathbb{V}r_t
\end{aligned} \tag{IA.62}$$

$$\begin{aligned}\mathbb{V}ar_t[\widetilde{dp}^{o}_{t+1}] &= \mathbb{V}ar_t[\mu_t - \mathbb{E}r_t \;+\; \nu_o \cdot \widetilde{dp}^{*}_{t+1}] \\ &= \mathbb{U}\mu_t \;+\; \nu_o^2 \cdot \nu_{dp}^2 \cdot \mathbb{V}r_t \end{aligned} \tag{IA.63}$$

$$\begin{aligned}\mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^{o}_{t+1}] &= \mathbb{C}ov_t[\phi_\mu \cdot \mu_t \;+\; \widetilde{\mu}^*_{t+1}, \mu_t - \mathbb{E}r_t \;+\; \nu_o \cdot \widetilde{dp}^{*}_{t+1}] \\ &= \phi_\mu \cdot \mathbb{C}ov_t[\mu_t, \mu_t] \;+\; \nu_o \cdot \mathbb{C}ov_t[\widetilde{\mu}^*_{t+1}, \widetilde{dp}^{*}_{t+1}] \\ &= \phi_\mu \cdot \mathbb{U}\mu_t \;+\; \nu_o \cdot \nu_{\mu,dp} \cdot \mathbb{V}r_t \end{aligned} \tag{IA.64}$$

$$\begin{aligned}\mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] &= \mathbb{C}ov_t[\phi_\mu \cdot \mu_t \;+\; \widetilde{\mu}^*_{t+1}, \mu_t - \mathbb{E}r_t \;+\; \widetilde{r}^*_{w,t+1}] \\ &= \phi_\mu \cdot \mathbb{C}ov_t[\mu_t, \mu_t] \;+\; \mathbb{C}ov_t[\widetilde{\mu}^*_{t+1}, \widetilde{r}^*_{w,t+1}] \\ &= \phi_\mu \cdot \mathbb{U}\mu_t \;+\; \nu_{r,\mu} \cdot \mathbb{V}r_t \end{aligned} \tag{IA.65}$$

$$\begin{aligned}\mathbb{C}ov_t[\widetilde{dp}^{o}_{t+1}, \widetilde{r}_{w,t+1}] &= \mathbb{C}ov_t[\mu_t \;+\; \nu_o \cdot \widetilde{dp}^{*}_{t+1}, \mu_t - \mathbb{E}r_t \;+\; \widetilde{r}^*_{w,t+1}] \\ &= \mathbb{C}ov_t[\mu_t, \mu_t] \;+\; \nu_o \cdot \mathbb{C}ov_t[\widetilde{dp}^{*}_{t+1}, \widetilde{r}^*_{w,t+1}] \\ &= \mathbb{U}\mu_t \;+\; \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t \end{aligned} \tag{IA.66}$$

Moreover, since

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \frac{1}{A \cdot D - B \cdot C} \cdot \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \tag{IA.67}$$

if we define

$$\Omega_t \;=\; \begin{bmatrix} \mathbb{V}ar_t[\widetilde{dp}^{o}_{t+1}] & \mathbb{C}ov_t[\widetilde{dp}^{o}_{t+1}, \widetilde{r}_{w,t+1}] \\ \mathbb{C}ov_t[\widetilde{dp}^{o}_{t+1}, \widetilde{r}_{w,t+1}] & \mathbb{V}ar_t[\widetilde{r}_{w,t+1}] \end{bmatrix} \tag{IA.68}$$

then we also have

$$\Omega_t^{-1} \;=\; \frac{1}{\xi_t} \times \begin{bmatrix} \mathbb{U}\mu_t + \mathbb{V}r_t & -(\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \\ -(\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) & (\mathbb{U}\mu_t + \nu_o^2 \cdot \nu_{dp}^2 \cdot \mathbb{V}r_t) \end{bmatrix} \tag{IA.69}$$

where

$$\xi_t \;=\; \nu_o^2 \cdot (\nu_{dp}^2 - \nu_{r,dp}^2) \cdot \mathbb{V}r_t^2 \;+\; (1 + \nu_o^2 \cdot \nu_{dp}^2 - 2 \cdot \nu_o \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t \cdot \mathbb{U}\mu_t$$

A.3.4 The $\mathbb{E}r_t$ and $\mathbb{U}\mu_t$ Dynamics

Given the results in the previous subsection, $\mathbb{E}r_{t+1}$ is given by

$$\begin{aligned}
\mathbb{E}r_{t+1} &= \mathbb{E}_{t+1}[\mu_{t+1}] \\
&= \mathbb{E}_t[\mu_{t+1} | \widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}] \\
&= \mathbb{E}_t[\mu_{t+1}] \;+\; \begin{bmatrix} \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] \\ \mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \end{bmatrix}' \times \Omega_t^{-1} \times \begin{bmatrix} \widetilde{dp}^o_{t+1} \\ \widetilde{r}_{w,t+1} \end{bmatrix} \\
&= \mu \;+\; \phi_\mu \cdot (\mathbb{E}r_t \;-\; \mu) \;+\; \begin{bmatrix} \xi_{dp,t} \\ \xi_{r,t} \end{bmatrix}' \times \begin{bmatrix} \widetilde{dp}^o_{t+1} \\ \widetilde{r}_{w,t+1} \end{bmatrix} \\
&= \mathbb{E}r \;+\; \phi_{\mathbb{E}} \cdot (\mathbb{E}r_t \;-\mathbb{E}r) \;+\; \widetilde{\mathbb{E}r}_{t+1}
\end{aligned} \tag{IA.70}$$

where $\mathbb{E}r = \mu$, $\phi_{\mathbb{E}} = \phi_\mu$, and

$$\widetilde{\mathbb{E}r}_{t+1} \;= \xi_{dp,t} \cdot \widetilde{dp}^o_{t+1} \;+\; \xi_{r,t} \cdot \widetilde{r}_{w,t+1}, \tag{IA.71}$$

with

$$\begin{aligned}
\xi_{dp,t} &= \frac{1}{\xi_t} \cdot \begin{pmatrix} \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] \cdot (\mathbb{U}\mu_t + \mathbb{V}r_t) \\ -\mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \cdot (\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \end{pmatrix} \\
&= \frac{1}{\xi_t} \cdot \begin{pmatrix} (\phi_\mu \cdot \mathbb{U}\mu_t + \nu_o \cdot \nu_{\mu,dp} \cdot \mathbb{V}r_t) \cdot (\mathbb{U}\mu_t + \mathbb{V}r_t) \\ -(\phi_\mu \cdot \mathbb{U}\mu_t + \nu_{r,\mu} \cdot \mathbb{V}r_t) \cdot (\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \end{pmatrix} \\
&= \frac{1}{\xi_t} \cdot \begin{pmatrix} \nu_o \cdot (\nu_{\mu,dp} - \nu_{r,\mu} \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t^2 \\ +(\phi_\mu \cdot (1 - \nu_o \cdot \nu_{r,dp}) + \nu_o \cdot \nu_{\mu,dp} - \nu_{r,\mu}) \cdot \mathbb{V}r_t \cdot \mathbb{U}\mu_t \end{pmatrix}
\end{aligned} \tag{IA.72}$$

and

$$\begin{aligned}
\xi_{r,t} &= \frac{1}{\xi_t} \cdot \left(\begin{array}{c} -\mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] \cdot (\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \\ +\mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \cdot (\mathbb{U}\mu_t + \nu_o^2 \cdot \nu_{dp}^2 \cdot \mathbb{V}r_t) \end{array} \right) \\
&= \frac{1}{\xi_t} \cdot \left(\begin{array}{c} -(\phi_\mu \cdot \mathbb{U}\mu_t + \nu_o \cdot \nu_{\mu,dp} \cdot \mathbb{V}r_t) \cdot (\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \\ +(\phi_\mu \cdot \mathbb{U}\mu_t + \nu_{r,\mu} \cdot \mathbb{V}r_t) \cdot (\mathbb{U}\mu_t + \nu_o^2 \cdot \nu_{dp}^2 \cdot \mathbb{V}r_t) \end{array} \right) \\
&= \frac{1}{\xi_t} \cdot \left(\begin{array}{c} \nu_o^2 \cdot (\nu_{r,\mu} \cdot \nu_{dp}^2 - \nu_{\mu,dp} \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t^2 \\ +(\nu_{r,\mu} + \phi_\mu \cdot \nu_o^2 \cdot \nu_{dp}^2 - \nu_o \cdot \nu_{\mu,dp} - \phi_\mu \cdot \nu_o \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t \cdot \mathbb{U}\mu_t \end{array} \right) \qquad \text{(IA.73)}
\end{aligned}$$

Similarly, given the results in the previous subsection, $\mathbb{U}\mu_{t+1}$ is given by

$$\begin{aligned}
\mathbb{U}\mu_{t+1} &= \mathbb{V}ar_{t+1}[\mu_{t+1}] \\
&= \mathbb{V}ar_t[\mu_{t+1} | \widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}] \\
&= \mathbb{V}ar_t[\mu_{t+1}] \ - \ \begin{bmatrix} \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] \\ \mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \end{bmatrix}' \times \mathbf{\Omega}_t^{-1} \times \begin{bmatrix} \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] \\ \mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \end{bmatrix} \\
&= \left(\phi_\mu^2 \cdot \mathbb{U}\mu_t \ + \ \nu_\mu^2 \cdot \mathbb{V}r_t\right) \ - \ \begin{bmatrix} \xi_{dp,t} \\ \xi_{r,t} \end{bmatrix}' \times \begin{bmatrix} \mathbb{C}ov_t[\mu_{t+1}, \widetilde{dp}^o_{t+1}] \\ \mathbb{C}ov_t[\mu_{t+1}, \widetilde{r}_{w,t+1}] \end{bmatrix} \\
&= \left(\phi_\mu^2 \cdot \mathbb{U}\mu_t \ + \ \nu_\mu^2 \cdot \mathbb{V}r_t\right) \ - \ \begin{bmatrix} \xi_{dp,t} \\ \xi_{r,t} \end{bmatrix}' \times \begin{bmatrix} \phi_\mu \cdot \mathbb{U}\mu_t + \nu_o \cdot \nu_{\mu,dp} \cdot \mathbb{V}r_t \\ \phi_\mu \cdot \mathbb{U}\mu_t + \nu_{r,\mu} \cdot \mathbb{V}r_t \end{bmatrix} \\
&= \phi_{\mathbb{U},t} \cdot \mathbb{U}\mu_t \ + \ \phi_{\mathbb{U},\mathbb{V},t} \cdot \mathbb{V}r_t \qquad \text{(IA.74)}
\end{aligned}$$

where

$\phi_{\mathbb{U},t} \ = \ \phi_\mu^2 \ - \ \phi_\mu \cdot \xi_{dp,t} \ - \ \phi_\mu \cdot \xi_{r,t}$

$\phi_{\mathbb{U},\mathbb{V},t} \ = \ \nu_\mu^2 \ - \ \nu_o \cdot \nu_{\mu,dp} \cdot \xi_{dp,t} \ - \ \nu_{r,\mu} \cdot \xi_{r,t}$

A.3.5 The $\mathbb{V}r_t$ Dynamics

The framework in the prior subsections is general enough to accommodate any $\mathbb{V}r_t$ process. However, to derive an expression for $N_{\mathbb{V}}$, we need to specify the $\mathbb{V}r_t$ dynamics. We model $\mathbb{V}r_t$ as a Realized log-GARCH process (see Hansen, Huang, and Shek (2012)) so that

$$log(\mathbb{V}r_{t+1}) \ = \omega_{\mathbb{V}r} \ + \ \phi_{\mathbb{V}r} \cdot log(\mathbb{V}r_t) \ + \ \phi_\sigma \cdot \sigma^2_{t+1} \tag{IA.75}$$

$$\sigma^2_{t+1} \ = \omega_\sigma \ + \ log(\mathbb{V}r_t) \ + \ \widetilde{\sigma}^2_{t+1} \tag{IA.76}$$

which can be combined to yield

$$log(\mathbb{V}r_{t+1}) \ = \omega_{\mathbb{V}} \ + \ \phi_{\mathbb{V}} \cdot log(\mathbb{V}r_t) \ + \ \phi_\sigma \cdot \widetilde{\sigma}^2_{t+1} \tag{IA.77}$$

where $\omega_{\mathbb{V}} = \omega_{\mathbb{V}r} + \phi_\sigma \cdot \omega_\sigma$, $\phi_{\mathbb{V}} = \phi_{\mathbb{V}r} + \phi_\sigma$, and $\widetilde{\sigma}^2_t \overset{iid}{\sim} N(0, \sigma^2_{\sigma})$, with σ^2_t reflecting the log of the realized variance of r_w over month t. Moreover, for simplicity, we assume that $\mathbb{C}ov_t[\widetilde{r}_w, \widetilde{\sigma}^2] = \sigma_{r,\sigma}$ and $\mathbb{C}ov_t[\widetilde{dp}^o, \widetilde{\sigma}^2] = \nu_o \cdot \sigma_{dp,\sigma}$.

Our use of a Realized log-GARCH process effectively treats the conditional variance as observable even though μ_t is not. This approach simplifies exposition and is consistent with prior literature (e.g., Anderson, Ghysels, and Juergens (2009)), with the justification being that an econometrician can consistently estimate ex-post variance through sampling returns over arbitrarily short time intervals while the same is not true about average returns (see, e.g., Merton (1980) and Foster and Nelson (1996)). However, our $log(\mathbb{V}r_t)$ process can be derived as the Bayesian posterior of a latent stochastic volatility model in which σ^2_t provides a noisy signal for the log conditional variance, as we demonstrate below.

Consider the following stochastic volatility model in which the log conditional variance, $\overline{\sigma}^2_t$, is unobservable:

$$\overline{\sigma}^2_{t+1} \ = \omega_{\overline{\sigma}} \ + \ \phi_{\overline{\sigma}} \cdot \overline{\sigma}^2_t \ + \ u_{t+1} \tag{IA.78}$$

$$\sigma^2_{t+1} \ = \omega_\sigma \ + \ \overline{\sigma}^2_t \ + \ \epsilon_{t+1} \tag{IA.79}$$

where $u_t \overset{iid}{\sim} N(0, \sigma^2_u)$ and $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2_\epsilon)$ are unobservable shocks with $\mathbb{C}ov_t[\epsilon, u] = \sigma_{\epsilon,u}$.

Now, let $log(\mathbb{V}r_t) \equiv \mathbb{E}_t[\overline{\sigma}_t^2]$ to get

$$\begin{aligned} \widetilde{\sigma}_{t+1}^2 &= \sigma_{t+1}^2 \; - \; \mathbb{E}_t[\sigma_{t+1}^2] \\ &= \sigma_{t+1}^2 \; - \; (\omega_\sigma \; + \; log(\mathbb{V}r_t)) \end{aligned} \tag{IA.80}$$

and define the $\mathcal{F}_t^\sigma$ subset of the investor's information as $\mathcal{F}_{t+1}^\sigma = \{\mathcal{F}_t^\sigma \cup \widetilde{\sigma}_{t+1}^2\}$ so that

$$\begin{aligned} \mathbb{E}_{t+1}[f(\overline{\sigma}_{t+1}^2)] &= \mathbb{E}_t[f(\overline{\sigma}_{t+1}^2)|\mathcal{F}_{t+1}^\sigma] \\ &= \mathbb{E}_t[f(\overline{\sigma}_{t+1}^2)|\widetilde{\sigma}_{t+1}^2, \mathcal{F}_t^\sigma] \\ &= \mathbb{E}_t[f(\overline{\sigma}_{t+1}^2)|\widetilde{\sigma}_{t+1}^2] + \mathbb{E}_t[f(\overline{\sigma}_{t+1}^2)|\mathcal{F}_t^\sigma] - \mathbb{E}_t[f(\overline{\sigma}_{t+1}^2)] \\ &= \mathbb{E}_t[f(\overline{\sigma}_{t+1}^2)|\widetilde{\sigma}_{t+1}^2] \end{aligned} \tag{IA.81}$$

for any arbitrary function $f(\mu_{t+1})$. The first line following from the fact that the investor updates the $\overline{\sigma}_t^2$ belief based only on the signals summarized by the $\mathcal{F}_t^\sigma$ information set and the third line follows from Theorem 2.4 in Chapter 5 of Anderson and Moore (1979).

Given Lemma 1 and Equation IA.81, so long as we can derive the joint distribution of $\overline{\sigma}_{t+1}^2$ and $\widetilde{\sigma}_{t+1}^2$ conditioned on time t information, we can obtain $log(\mathbb{V}r_{t+1}) = \mathbb{E}_{t+1}[\overline{\sigma}_{t+1}^2] = \mathbb{E}_t[\overline{\sigma}_{t+1}^2|\widetilde{\sigma}_{t+1}^2]$.

From the dynamics of $\overline{\sigma}_{t+1}^2$ and $\widetilde{\sigma}_{t+1}^2$ in Equations IA.80 and IA.78, we have:

$$\begin{bmatrix} \overline{\sigma}_{t+1}^2 \\ \widetilde{\sigma}_{t+1}^2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbb{E}_t[\overline{\sigma}_{t+1}^2] \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbb{V}ar_t[\overline{\sigma}_{t+1}^2] & \mathbb{C}ov_t[\overline{\sigma}_{t+1}^2, \widetilde{\sigma}_{t+1}^2] \\ \mathbb{C}ov_t[\overline{\sigma}_{t+1}^2, \widetilde{\sigma}_{t+1}^2] & \mathbb{V}ar_t[\widetilde{\sigma}_{t+1}^2] \end{bmatrix} \right) \tag{IA.82}$$

where (with $\mathbb{U}\overline{\sigma}_t^2 = \mathbb{V}ar_t[\overline{\sigma}_t^2]$)

$$\begin{aligned} \mathbb{E}_t[\overline{\sigma}_{t+1}^2] &= \mathbb{E}_t[\omega_{\overline{\sigma}} \; + \; \phi_{\overline{\sigma}} \cdot \overline{\sigma}_t^2 \; + \; u_{t+1}] \\ &= \omega_{\overline{\sigma}} \; + \; \phi_{\overline{\sigma}} \cdot log(\mathbb{V}r_t) \end{aligned} \tag{IA.83}$$

$$\begin{aligned} \mathbb{V}ar_t[\overline{\sigma}_{t+1}^2] &= \mathbb{V}ar_t[\omega_{\overline{\sigma}} \; + \; \phi_{\overline{\sigma}} \cdot \overline{\sigma}_t^2 \; + \; u_{t+1}] \\ &= \mathbb{U}\overline{\sigma}_t^2 \; + \; \sigma_u^2 \end{aligned} \tag{IA.84}$$

$$
\begin{aligned}
\mathbb{V}ar_t[\widetilde{\sigma}^2_{t+1}] &= \mathbb{V}ar_t[\sigma^2_{t+1} - (\omega_\sigma \ + \ log(\mathbb{V}r_t))] \\
&= \mathbb{V}ar_t[\overline{\sigma}^2_t - log(\mathbb{V}r_t) \ + \ \epsilon_{t+1})] \\
&= \mathbb{U}\overline{\sigma}^2_t \ + \ \sigma^2_\epsilon
\end{aligned} \tag{IA.85}
$$

$$
\begin{aligned}
\mathbb{C}ov_t[\overline{\sigma}^2_{t+1}, \widetilde{\sigma}^2_{t+1}] &= \mathbb{C}ov_t[\omega_{\overline{\sigma}} \ + \ \phi_{\overline{\sigma}} \cdot \overline{\sigma}^2_t \ + \ u_{t+1}, \sigma^2_{t+1} - (\omega_\sigma \ + \ log(\mathbb{V}r_t))] \\
&= \phi_{\overline{\sigma}} \cdot \mathbb{C}ov_t[\overline{\sigma}^2_t, \sigma^2_{t+1}] \ + \ \mathbb{C}ov_t[u_{t+1}\sigma^2_{t+1}] \\
&= \phi_{\overline{\sigma}} \cdot \mathbb{C}ov_t[\overline{\sigma}^2_t, \overline{\sigma}^2_t] \ + \ \mathbb{C}ov_t[u_{t+1}, \epsilon_{t+1}] \\
&= \phi_{\overline{\sigma}} \cdot \mathbb{U}\overline{\sigma}^2_t \ + \ \sigma_{\epsilon,u}
\end{aligned} \tag{IA.86}
$$

From Lemma 1, we then have

$$
\begin{aligned}
log(\mathbb{V}r_{t+1}) &= \mathbb{E}_{t+1}[\overline{\sigma}^2_{t+1}] \\
&= \mathbb{E}_t[\overline{\sigma}^2_{t+1} | \widetilde{\sigma}^2_{t+1}] \\
&= \mathbb{E}_t[\overline{\sigma}^2_{t+1}] \ + \ \frac{\mathbb{C}ov_t[\overline{\sigma}^2_{t+1}, \widetilde{\sigma}^2_{t+1}]}{\mathbb{V}ar_t[\widetilde{\sigma}^2_{t+1}]} \cdot \widetilde{\sigma}^2_{t+1} \\
&= \omega_{\overline{\sigma}} \ + \ \phi_{\overline{\sigma}} \cdot log(\mathbb{V}r_t) \ + \ \frac{\phi_{\overline{\sigma}} \cdot \mathbb{U}\overline{\sigma}^2_t \ + \ \sigma_{\epsilon,u}}{\mathbb{U}\overline{\sigma}^2_t \ + \ \sigma^2_\epsilon} \cdot \widetilde{\sigma}^2_{t+1}
\end{aligned} \tag{IA.87}
$$

and

$$
\begin{aligned}
\mathbb{U}\overline{\sigma}^2_{t+1} &= \mathbb{V}ar_{t+1}[\overline{\sigma}^2_{t+1}] \\
&= \mathbb{V}ar_t[\overline{\sigma}^2_{t+1} | \widetilde{\sigma}^2_{t+1}] \\
&= \mathbb{V}ar_t[\overline{\sigma}^2_{t+1}] \ - \ \frac{\mathbb{C}ov_t[\overline{\sigma}^2_{t+1}, \widetilde{\sigma}^2_{t+1}]^2}{\mathbb{V}ar_t[\widetilde{\sigma}^2_{t+1}]} \\
&= \mathbb{U}\overline{\sigma}^2_t \ + \ \sigma^2_u \ - \ \frac{(\phi_{\overline{\sigma}} \cdot \mathbb{U}\overline{\sigma}^2_t \ + \ \sigma_{\epsilon,u})^2}{\mathbb{U}\overline{\sigma}^2_t \ + \ \sigma^2_\epsilon}
\end{aligned} \tag{IA.88}
$$

which is a deterministic recursion, and thus (under standard regularity conditions for the parameters) converges to $\mathbb{U}\overline{\sigma}^2$.[IA.6]

Without loss of generality, we can then substitute $\mathbb{U}\overline{\sigma}_t^2 = \mathbb{U}\overline{\sigma}^2$ in Equation IA.87 to obtain

$$log(\mathbb{V}r_{t+1}) \ = \omega_{\mathbb{V}} \ + \ \phi_{\mathbb{V}} \cdot log(\mathbb{V}r_t) \ + \ \phi_\sigma \cdot \widetilde{\sigma}_{t+1}^2 \tag{IA.89}$$

where $\omega_{\mathbb{V}} = \omega_{\overline{\sigma}}$, $\phi_{\mathbb{V}} = \phi_{\overline{\sigma}}$, $\phi_\sigma = (\phi_{\overline{\sigma}} \cdot \mathbb{U}\overline{\sigma}^2 \ + \ \sigma_{\epsilon,u})/(\mathbb{U}\overline{\sigma}^2 \ + \ \sigma_\epsilon^2)$.

Consequently, our log-GARCH specification in Equations IA.76 and IA.77 is equivalent to the latent stochastic volatility model in Equations IA.78 and IA.79. This equivalence is often explored in the econometrics literature to provide simple and robust methods to estimate the parameters of latent stochastic volatility models (e.g., Fleming and Kirby (2003)). In our context, we rely on this equivalence to simplify our ICAPM by directly specifying a log-GARCH process for the market conditional variance.

A.4 The $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ Expressions

We now use the $\mathbb{E}r_t$, $\mathbb{V}r_t$, and $\mathbb{U}\mu_t$ dynamics in the previous subsection to derive expressions for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ in terms of measurable variables.

A.4.1 The $N_{\mathbb{E}}$ Expression

From the $\mathbb{E}r_t$ dynamics in Equation IA.70, we have

$$\mathbb{E}_t \left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot r_{w,t+h} \right] \ = \frac{\overline{\delta} \cdot \mathbb{E}r}{1-\overline{\delta}} \ + \ \theta_{\mathbb{E}} \cdot (\mathbb{E}r_t - \mathbb{E}r) \tag{IA.90}$$

[IA.6] While not necessary for our purpose, we can find $\mathbb{U}\overline{\sigma}^2$ as a function of the other model parameters by substituting $\mathbb{U}\overline{\sigma}_{t+1}^2 = \mathbb{U}\overline{\sigma}^2$ and $\mathbb{U}\overline{\sigma}_t^2 = \mathbb{U}\overline{\sigma}^2$ in Equation IA.88 and solving the resulting quadratic equation for $\mathbb{U}\overline{\sigma}^2$.

which implies that $N_{\mathbb{E}}$ in Equation IA.38 is

$$N_{\mathbb{E},t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t)\left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot r_{w,t+h}\right]$$
$$= \theta_{\mathbb{E}} \cdot \widetilde{\mathbb{E}r}_{t+1}$$
$$= \theta_{dp,t} \cdot \widetilde{dp}^o_{t+1} + \theta_{r,t} \cdot \widetilde{r}_{w,t+1} \quad \text{(IA.91)}$$

where $\theta_{\mathbb{E}} = \overline{\delta}/(1 - \overline{\delta} \cdot \phi_{\mathbb{E}})$, $\theta_{dp,t} = \theta_{\mathbb{E}} \cdot \xi_{dp,t}$, and $\theta_{r,t} = \theta_{\mathbb{E}} \cdot \xi_{r,t}$.

A.4.2 The $N_{\mathbb{V}}$ Expression

Now, our objective is to obtain an expression for $N_{\mathbb{V}}$. Unfortunately, given our wealth return dynamics, $N_{\mathbb{V}}$ does not have an analytical solution. As such, this subsection uses several approximations to obtain the analytical approximation $N_{\mathbb{V},t} \approx \theta_\sigma \cdot \widetilde{\sigma}^2_t$, which we use in the main text. In Subsection D.7, we solve for the nonlinear $N_{\mathbb{V}}$ solution numerically (i.e., obviating the need for the approximations to generate the analytical solution) and provide results that are very similar to the baseline results we report in the main text.

The wealth return dynamics in Section A.3 can be fully summarized by

$$r_{w,t+1} = \mathbb{E}r_t + \widetilde{r}_{w,t+1} \quad \text{(IA.92)}$$

$$\mathbb{E}r_{t+1} = \mathbb{E}r + \phi_{\mathbb{E}} \cdot (\mathbb{E}r_t - \mathbb{E}r) + \widetilde{\mathbb{E}r}_{t+1} \quad \text{(IA.93)}$$

$$\mathbb{U}\mu_{t+1} = \phi_{\mathbb{U},t} \cdot \mathbb{U}\mu_t + \phi_{\mathbb{U},\mathbb{V},t} \cdot \mathbb{V}r_t \quad \text{(IA.94)}$$

$$log(\mathbb{V}r_{t+1}) = \omega_{\mathbb{V}r} + \phi_{\mathbb{V}r} \cdot log(\mathbb{V}r_t) + \phi_\sigma \cdot \sigma^2_{t+1} \quad \text{(IA.95)}$$

$$\sigma^2_{t+1} = \omega_\sigma + log(\mathbb{V}r_t) + \widetilde{\sigma}^2_{t+1} \quad \text{(IA.96)}$$

where the shocks are normally distributed with variance and covariance terms given by

$$\mathbb{V}ar_t \begin{bmatrix} \widetilde{r}_{w,t+1} \\ \widetilde{\mathbb{E}r}_{t+1} \\ \widetilde{\sigma}^2_{t+1} \\ \widetilde{\mathbb{U}\mu}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbb{V}ar_t[\widetilde{r}_{w,t+1}] & \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] & \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\sigma}^2_{t+1}] & 0 \\ \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] & \mathbb{V}ar_t[\widetilde{\mathbb{E}r}_{t+1}] & \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\sigma}^2_{t+1}] & 0 \\ \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\sigma}^2_{t+1}] & \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\sigma}^2_{t+1}] & \mathbb{V}ar_t[\widetilde{\sigma}^2_{t+1}] & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{(IA.97)}$$

with[IA.7]

$$\mathbb{V}ar_t[\widetilde{r}_{w,t+1}] = \mathbb{U}\mu_t + \mathbb{V}r_t$$
$$\mathbb{V}ar_t[\widetilde{\mathbb{E}r}_{t+1}] = (\xi^2_{dp,t} + \xi^2_{r,t} + 2 \cdot \xi_{dp,t} \cdot \xi_{r,t}) \cdot \mathbb{U}\mu_t + (\nu^2_o \cdot \nu^2_{dp} \cdot \xi^2_{dp,t} + \xi^2_{r,t} + 2 \cdot \nu_o \cdot \nu_{r,dp} \cdot \xi_{dp,t} \cdot \xi_{r,t}) \cdot \mathbb{V}r_t$$
$$\mathbb{V}ar_t[\widetilde{\sigma}^2_{t+1}] = \sigma^2_\sigma$$
$$\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] = (\xi_{dp,t} + \xi_{r,t}) \cdot \mathbb{U}\mu_t + (\nu_o \cdot \nu_{r,dp} \cdot \xi_{dp,t} + \xi_{r,t}) \cdot \mathbb{V}r_t$$
$$\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\sigma}^2_{t+1}] = \sigma_{r,\sigma}$$

[IA.7] The derivation of the $\widetilde{\mathbb{E}r}_t$ variance and covariance expressions is as follows:

$$\begin{aligned} \mathbb{V}ar_t[\widetilde{\mathbb{E}r}_{t+1}] &= \mathbb{V}ar_t[\xi_{dp,t} \cdot \widetilde{dp}^o_{t+1} \;+\; \xi_{r,t} \cdot \widetilde{r}_{w,t+1}] \\ &= \xi^2_{dp,t} \cdot \mathbb{V}ar_t[\widetilde{dp}^o_{t+1}] \;+\; \xi^2_{r,t} \cdot \mathbb{V}ar_t[\widetilde{r}_{w,t+1}] \;+\; 2 \cdot \xi_{dp,t} \cdot \xi_{r,t} \cdot \mathbb{C}ov_t[\widetilde{dp}^o_{t+1}, \widetilde{r}_{w,t+1}] \\ &= \xi^2_{dp,t} \cdot (\mathbb{U}\mu_t + \nu^2_o \cdot \nu^2_{dp} \cdot \mathbb{V}r_t) \;+\; \xi^2_{r,t} \cdot (\mathbb{U}\mu_t + \mathbb{V}r_t) \;+\; 2 \cdot \xi_{dp,t} \cdot \xi_{r,t} \cdot (\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \\ &= (\xi^2_{dp,t} + \xi^2_{r,t} + 2 \cdot \xi_{dp,t} \cdot \xi_{r,t}) \cdot \mathbb{U}\mu_t \;+\; (\nu^2_o \cdot \nu^2_{dp} \cdot \xi^2_{dp,t} + \xi^2_{r,t} + 2 \cdot \nu_o \cdot \nu_{r,dp} \cdot \xi_{dp,t} \cdot \xi_{r,t}) \cdot \mathbb{V}r_t \end{aligned}$$

$$\begin{aligned} \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] &= \mathbb{C}ov_t[\widetilde{r}_{w,t+1} \;,\; \xi_{dp,t} \cdot \widetilde{dp}^o_{t+1} \;+\; \xi_{r,t} \cdot \widetilde{r}_{w,t+1}] \\ &= \xi_{dp,t} \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{dp}^o_{t+1}] \;+\; \xi_{r,t} \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{r}_{w,t+1}] \\ &= \xi_{dp,t} \cdot (\mathbb{U}\mu_t + \nu_o \cdot \nu_{r,dp} \cdot \mathbb{V}r_t) \;+\; \xi_{r,t} \cdot (\mathbb{V}r_t + \mathbb{U}\mu_t) \\ &= (\xi_{dp,t} + \xi_{r,t}) \cdot \mathbb{U}\mu_t \;+\; (\nu_o \cdot \nu_{r,dp} \cdot \xi_{dp,t} + \xi_{r,t}) \cdot \mathbb{V}r_t \end{aligned}$$

$$\begin{aligned} \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\sigma}^2_{t+1}] &= \mathbb{C}ov_t[\xi_{dp,t} \cdot \widetilde{dp}^o_{t+1} \;+\; \xi_{r,t} \cdot \widetilde{r}_{w,t+1} \;,\; \widetilde{\sigma}^2_{t+1}] \\ &= \xi_{dp,t} \cdot \mathbb{C}ov_t[\widetilde{dp}^o_{t+1}, \widetilde{\sigma}^2_{t+1}] \;+\; \xi_{r,t} \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\sigma}^2_{t+1}] \\ &= \nu_o \cdot \sigma_{dp,\sigma} \cdot \xi_{dp,t} + \sigma_{r,\sigma} \cdot \xi_{r,t} \end{aligned}$$

$\mathbb{C}ov_t[\widetilde{dp}^{o}_{t+1}, \widetilde{\sigma}^2_{t+1}] = \nu_o \cdot \sigma_{dp,\sigma}$

$\mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\sigma}^2_{t+1}] = \nu_o \cdot \sigma_{dp,\sigma} \cdot \xi_{dp,t} + \sigma_{r,\sigma} \cdot \xi_{r,t}$

The first linear approximation we use (derived in Subsection A.4.3) yields

$$\mathbb{V}r_{t+1} \approx \mathbb{V}r + \phi_{\mathbb{V}} \cdot (\mathbb{V}r_t - \mathbb{V}r) + \widetilde{\mathbb{V}r}_{t+1} \tag{IA.98}$$

where

$\mathbb{V}r = e^{\omega_{\mathbb{V}}/(1-\phi_{\mathbb{V}})}$

$\omega_{\mathbb{V}} = \omega_{\mathbb{V}r} + \phi_\sigma \cdot \omega_\sigma$

$\phi_{\mathbb{V}} = \phi_{\mathbb{V}r} + \phi_\sigma$

$\widetilde{\mathbb{V}r}_t = \mathbb{V}r \cdot \phi_\sigma \cdot \widetilde{\sigma}^2_t$

$\mathbb{V}ar_t[\widetilde{\mathbb{V}r}_{t+1}] = \sigma^2_{\mathbb{V}} = \mathbb{V}r^2 \cdot \phi^2_\sigma \cdot \sigma^2_\sigma$

$\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{V}r}_{t+1}] = \sigma_{r,\mathbb{V}} = \mathbb{V}r \cdot \phi_\sigma \cdot \sigma_{r,\sigma}$

$\mathbb{C}ov_t[\widetilde{dp}^{o}_{t+1}, \widetilde{\mathbb{V}r}_{t+1}] = \nu_o \cdot \sigma_{dp,\mathbb{V}} = \nu_o \cdot \mathbb{V}r \cdot \phi_\sigma \cdot \sigma_{dp,\sigma}$

$\mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\mathbb{V}r}_{t+1}] = \nu_o \cdot \sigma_{dp,\mathbb{V}} \cdot \xi_{dp,t} + \sigma_{r,\mathbb{V}} \cdot \xi_{r,t}$

We then conjecture (with verification below) that $N_{\mathbb{V},t+1} = \theta_{\mathbb{V}} \cdot \widetilde{\mathbb{V}r}_{t+1}$, which results in

$$\begin{aligned}
\mathbb{V}ar_t[v_{t+1}] &= \mathbb{V}ar_t[\widetilde{r}_{w,t+1} + \widetilde{vw}_{t+1}] \\
&= \mathbb{V}ar_t[\widetilde{r}_{w,t+1} + \theta_{\mathbb{E}} \cdot \widetilde{\mathbb{E}r}_{t+1} - 0.5 \cdot (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \widetilde{\mathbb{V}r}_{t+1}] \\
&= \mathbb{V}ar_t[\widetilde{r}_{w,t+1} - 0.5 \cdot (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \widetilde{\mathbb{V}r}_{t+1}] + \theta^2_{\mathbb{E}} \cdot \mathbb{V}ar_t[\widetilde{\mathbb{E}r}_{t+1}] \\
&\quad + 2 \cdot \theta_{\mathbb{E}} \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] - (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \theta_{\mathbb{E}} \cdot \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\mathbb{V}r}_{t+1}] \\
&= \mathbb{V}ar_t[\widetilde{r}_{w,t+1}] + (0.5 \cdot (\gamma - 1) \cdot \theta_{\mathbb{V}})^2 \cdot \mathbb{V}ar_t[\widetilde{\mathbb{V}r}_{t+1}] - (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{V}r}_{t+1}] \\
&\quad + \theta^2_{\mathbb{E}} \cdot \mathbb{V}ar_t[\widetilde{\mathbb{E}r}_{t+1}] + 2 \cdot \theta_{\mathbb{E}} \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] - (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \theta_{\mathbb{E}} \cdot \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\mathbb{V}r}_{t+1}] \\
&= (0.5 \cdot (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \sigma_{\mathbb{V}})^2 - (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \sigma_{r,\mathbb{V}} + \mathbb{U}\mu_t + \mathbb{V}r_t + \theta^2_{\mathbb{E}} \cdot \mathbb{V}ar_t\left[\widetilde{\mathbb{E}r}_{t+1}\right] \\
&\quad + 2 \cdot \theta_{\mathbb{E}} \cdot \mathbb{C}ov_t\left[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}\right] - (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \theta_{\mathbb{E}} \cdot \mathbb{C}ov_t\left[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\mathbb{V}r}_{t+1}\right] \\
&= \varphi_{0,t} + \varphi_{\mathbb{V},t} \cdot \mathbb{V}r_t + \varphi_{\mathbb{U},t} \cdot \mathbb{U}\mu_t
\end{aligned} \tag{IA.99}$$

where

$$\begin{aligned}
\varphi_{0,t} \ &= (0.5 \cdot (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \sigma_{\mathbb{V}})^2 \ - \ (\gamma - 1) \cdot \theta_{\mathbb{V}} \cdot \sigma_{r,\mathbb{V}} \\
&\quad -(\gamma - 1) \cdot \nu_o \cdot \theta_{\mathbb{E}} \cdot \theta_{\mathbb{V}} \cdot \sigma_{dp,\mathbb{V}} \cdot \xi_{dp,t} \ - \ (\gamma - 1) \cdot \theta_{\mathbb{E}} \cdot \theta_{\mathbb{V}} \cdot \sigma_{r,\mathbb{V}} \cdot \xi_{r,t} \\
\varphi_{\mathbb{V},t} &= 1 \ + \ 2 \cdot \theta_{\mathbb{E}} \cdot \nu_o \cdot \nu_{r,dp} \cdot \xi_{dp,t} \ + \ 2 \cdot \theta_{\mathbb{E}} \cdot \xi_{r,t} \\
&\quad + \theta_{\mathbb{E}}^2 \cdot \nu_o^2 \cdot \nu_{dp}^2 \cdot \xi_{dp,t}^2 \ + \ \theta_{\mathbb{E}}^2 \cdot \xi_{r,t}^2 \ + \ 2 \cdot \theta_{\mathbb{E}}^2 \cdot \nu_o \cdot \nu_{r,dp} \cdot \xi_{dp,t} \cdot \xi_{r,t} \\
\varphi_{\mathbb{U},t} &= 1 \ + \ 2 \cdot \theta_{\mathbb{E}} \cdot \xi_{dp,t} \ + \ 2 \cdot \theta_{\mathbb{E}} \cdot \xi_{r,t} \ + \ \theta_{\mathbb{E}}^2 \cdot \xi_{dp,t}^2 \ + \ \theta_{\mathbb{E}}^2 \cdot \xi_{r,t}^2 \ + \ 2 \cdot \theta_{\mathbb{E}}^2 \cdot \xi_{dp,t} \cdot \xi_{r,t}
\end{aligned}$$

Equations IA.99 and IA.74 imply that $\mathbb{V}ar_t[v_{t+1}]$ and $\mathbb{U}\mu_{t+1}$ are non-linear functions of $\mathbb{V}r_t$ and $\mathbb{U}\mu_t$. In Subsection A.4.3, we provide first order Taylor approximations that allow us to write

$$\mathbb{V}ar_t \left[v_{t+1}\right] \ \approx \varphi_0 \ + \ \varphi_{\mathbb{V}} \cdot \mathbb{V}r_t \ + \ \varphi_{\mathbb{U}} \cdot \mathbb{U}\mu_t \tag{IA.100}$$

and

$$\mathbb{U}\mu_{t+1} \ \approx \mathbb{U}\mu \ + \ \phi_{\mathbb{U}} \cdot (\mathbb{U}\mu_t - \mathbb{U}\mu) \ + \ \phi_{\mathbb{U},\mathbb{V}} \cdot (\mathbb{V}r_t - \mathbb{V}r) \tag{IA.101}$$

where the expressions for φ_0, $\varphi_{\mathbb{V}}$, $\varphi_{\mathbb{U}}$, $\mathbb{U}\mu$, $\phi_{\mathbb{U}}$, and $\phi_{\mathbb{U},\mathbb{V}}$ are provided in Subsection A.4.3.

Using the approximation in Equation IA.100 and then the approximation in Equation IA.101, we have

$$\begin{aligned}
\mathbb{E}_t \left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{V}ar_{t+h-1} \left[v_{t+h}\right] \right] \ &\approx \frac{\overline{\delta} \cdot \varphi_0}{1 - \overline{\delta}} + \varphi_{\mathbb{V}} \cdot \sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{E}_t \left[\mathbb{V}r_{t+h-1}\right] + \varphi_{\mathbb{U}} \cdot \sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{E}_t \left[\mathbb{U}\mu_{t+h-1}\right] \\
&= \frac{\overline{\delta} \cdot (\varphi_0 + \varphi_{\mathbb{V}} \cdot \mathbb{V}r + \varphi_{\mathbb{U}} \cdot \mathbb{U}\mu)}{1 - \overline{\delta}} + \theta_{\mathbb{V}} \cdot (\mathbb{V}r_t - \mathbb{V}r) + \theta_{\mathbb{U}} \cdot (\mathbb{U}\mu_t - \mathbb{U}\mu)
\end{aligned} \tag{IA.102}$$

so that $N_{\mathbb{V}}$ in Equation IA.38 is

$$\begin{aligned}
N_{\mathbb{V},t} \ &= (\mathbb{E}_{t+1} - \mathbb{E}_t) \left[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{V}ar_{t+h}[v_{t+1+h}] \right] \\
&\approx \theta_{\mathbb{V}} \cdot \widetilde{\mathbb{V}r}_t \\
&= \theta_{\sigma} \cdot \widetilde{\sigma}_t^2
\end{aligned} \tag{IA.103}$$

which approximately confirms our $N_{\mathbb{V},t+1} = \theta_{\mathbb{V}} \cdot \widetilde{\mathbb{V}r}_{t+1}$ conjecture and further implies (when added to Equations IA.37 and IA.90)

$$vw_t \;=\; \frac{f(\psi,\delta,\overline{\delta}) + \overline{\delta}\cdot\mathbb{E}r}{1-\overline{\delta}} \;+\; \theta_{\mathbb{E}}\cdot(\mathbb{E}r_t - \mathbb{E}r) \;-\; \frac{(\gamma-1)}{2}\cdot\mathbb{E}_t\left[\sum_{h=1}^{\infty}\overline{\delta}^h\cdot\mathbb{V}ar_{t+h-1}\left[v_{t+h}\right]\right]$$

$$\approx \theta_0 \;+\; \theta_{\mathbb{E}}\cdot\mathbb{E}r_t \;-\; \frac{(\gamma-1)}{2}\cdot\theta_{\mathbb{V}}\cdot\mathbb{V}r_t \;-\; \frac{(\gamma-1)}{2}\cdot\theta_{\mathbb{U}}\cdot\mathbb{U}\mu_t \qquad \text{(IA.104)}$$

where[IA.8]

$\theta_0 = \frac{f(\psi,\delta,\overline{\delta})-0.5\cdot(\gamma-1)\cdot\overline{\delta}\cdot\varphi_0}{1-\overline{\delta}} + \left(\frac{\overline{\delta}}{1-\overline{\delta}} - \theta_{\mathbb{E}}\right)\cdot\mathbb{E}r - \frac{(\gamma-1)}{2}\cdot\left(\frac{\overline{\delta}\cdot\varphi_{\mathbb{V}}}{1-\overline{\delta}} - \theta_{\mathbb{V}}\right)\cdot\mathbb{V}r - \frac{(\gamma-1)}{2}\cdot\left(\frac{\overline{\delta}\cdot\varphi_{\mathbb{U}}}{1-\overline{\delta}} - \theta_{\mathbb{U}}\right)\cdot\mathbb{U}\mu$

$\theta_{\mathbb{E}} = \overline{\delta}/(1-\overline{\delta}\cdot\phi_{\mathbb{E}})$

$\theta_{\mathbb{V}} = \varphi_{\mathbb{V}}\cdot[\overline{\delta}/(1-\overline{\delta}\cdot\phi_{\mathbb{V}})] + \varphi_{\mathbb{U}}\cdot(\overline{\delta}^2\cdot\phi_{\mathbb{U},\mathbb{V}})/[(1-\overline{\delta}\cdot\phi_{\mathbb{V}})\cdot(1-\overline{\delta}\cdot\phi_{\mathbb{U}})]$

$\theta_{\mathbb{U}} = \varphi_{\mathbb{U}}\cdot\overline{\delta}/(1-\overline{\delta}\cdot\phi_{\mathbb{U}})$

$\theta_{\sigma} = \mathbb{V}r\cdot\phi_{\sigma}\cdot\theta_{\mathbb{V}}$

with the $\theta_{\mathbb{V}}$ expression representing a recursive equation since it expresses $\theta_{\mathbb{V}}$ as an implicit function of $\theta_{\mathbb{V}}$ and other parameters.

A.4.3 The Approximations Used to Derive the $N_{\mathbb{V}}$ Expression

We now derive all approximations used in the previous subsection to obtain our $N_{\mathbb{V},t} \approx \theta_{\mathbb{V}}\cdot\widetilde{\mathbb{V}r_t}$ approximation.

[IA.8] To derive the $\theta_{\mathbb{V}}$ and $\theta_{\mathbb{U}}$ expressions, note that

$$\mathbb{E}_t\left[\begin{pmatrix}\mathbb{V}r_{t+h-1}-\mathbb{V}r\\ \mathbb{U}\mu_{t+h-1}-\mathbb{U}\mu\end{pmatrix}\right] = \begin{bmatrix}\phi_{\mathbb{V}} & 0\\ \phi_{\mathbb{U},\mathbb{V}} & \phi_{\mathbb{U}}\end{bmatrix}^{h-1}\times\begin{pmatrix}\mathbb{V}r_t-\mathbb{V}r\\ \mathbb{U}\mu_t-\mathbb{U}\mu\end{pmatrix}$$

so that

$$\sum_{h=1}^{\infty}\overline{\delta}^h\cdot\mathbb{E}_t\left[\begin{pmatrix}\mathbb{V}r_{t+h-1}-\mathbb{V}r\\ \mathbb{U}\mu_{t+h-1}-\mathbb{U}\mu\end{pmatrix}\right] = \overline{\delta}\cdot\begin{bmatrix}1-\overline{\delta}\cdot\phi_{\mathbb{V}} & 0\\ -\overline{\delta}\cdot\phi_{\mathbb{U},\mathbb{V}} & 1-\overline{\delta}\cdot\phi_{\mathbb{U}}\end{bmatrix}^{-1}\times\begin{pmatrix}\mathbb{V}r_t-\mathbb{V}r\\ \mathbb{U}\mu_t-\mathbb{U}\mu\end{pmatrix}$$

$$= \overline{\delta}\cdot\frac{1}{(1-\overline{\delta}\cdot\phi_{\mathbb{V}})\cdot(1-\overline{\delta}\cdot\phi_{\mathbb{U}})}\cdot\begin{bmatrix}1-\overline{\delta}\cdot\phi_{\mathbb{U}} & 0\\ \overline{\delta}\cdot\phi_{\mathbb{U},\mathbb{V}} & 1-\overline{\delta}\cdot\phi_{\mathbb{V}}\end{bmatrix}\times\begin{pmatrix}\mathbb{V}r_t-\mathbb{V}r\\ \mathbb{U}\mu_t-\mathbb{U}\mu\end{pmatrix}$$

$$= \begin{pmatrix}\frac{\overline{\delta}}{1-\overline{\delta}\cdot\phi_{\mathbb{V}}}\cdot(\mathbb{V}r_t-\mathbb{V}r)\\ \frac{\overline{\delta}^2\cdot\phi_{\mathbb{U},\mathbb{V}}}{(1-\overline{\delta}\cdot\phi_{\mathbb{V}})\cdot(1-\overline{\delta}\cdot\phi_{\mathbb{U}})}\cdot(\mathbb{V}r_t-\mathbb{V}r) \;+\; \frac{\overline{\delta}}{1-\overline{\delta}\cdot\phi_{\mathbb{U}}}\cdot(\mathbb{U}\mu_t-\mathbb{U}\mu)\end{pmatrix}$$

which yields the $\theta_{\mathbb{V}}$ and $\theta_{\mathbb{U}}$ expressions.

(a) Approximation for $\mathbb{V}r_t$ Dynamics

Our log-GARCH dynamics for $\mathbb{V}r_t$ are given by

$$log(\mathbb{V}r_{t+1}) \;= \omega_{\mathbb{V}r} \;+\; \phi_{\mathbb{V}r} \cdot log(\mathbb{V}r_t) \;+\; \phi_\sigma \cdot \sigma^2_{t+1} \tag{IA.105}$$

$$\sigma^2_{t+1} \;= \omega_\sigma \;+\; log(\mathbb{V}r_t) \;+\; \widetilde{\sigma}^2_{t+1} \tag{IA.106}$$

which can be combine to yield

$$log(\mathbb{V}r_{t+1}) \;= \omega_{\mathbb{V}} \;+\; \phi_{\mathbb{V}} \cdot log(\mathbb{V}r_t) \;+\; \phi_\sigma \cdot \widetilde{\sigma}^2_{t+1} \tag{IA.107}$$

where $\omega_{\mathbb{V}} = \omega_{\mathbb{V}r} + \phi_\sigma \cdot \omega_\sigma$, $\phi_{\mathbb{V}} = \phi_{\mathbb{V}r} + \phi_\sigma$, and $\widetilde{\sigma}^2_t \overset{iid}{\sim} N(0, \sigma^2_\sigma)$.

As such, we have:

$$\begin{aligned} \mathbb{V}r_{t+1} &= \mathbb{V}r_t^{\phi_{\mathbb{V}}} \cdot e^{\omega_{\mathbb{V}} \,+\, \phi_\sigma \cdot \widetilde{\sigma}^2_{t+1}} \\ &\approx \mathbb{V}r^{\phi_{\mathbb{V}}} \cdot e^{\omega_{\mathbb{V}}} \;+\; \phi_{\mathbb{V}} \cdot \mathbb{V}r^{\phi_{\mathbb{V}}-1} \cdot e^{\omega_{\mathbb{V}}} \cdot (\mathbb{V}r_t - \mathbb{V}r) \;+\; \mathbb{V}r^{\phi_{\mathbb{V}}} \cdot e^{\omega_{\mathbb{V}}} \cdot \phi_\sigma \cdot \widetilde{\sigma}^2_{t+1} \end{aligned} \tag{IA.108}$$

where the second line follows from a first order Taylor expansion around $\mathbb{V}r_t = \mathbb{V}r$ and $\widetilde{\sigma}^2_{t+1} = 0$.

Now, taking unconditional expectation on both sides of Equation IA.108 implies $\mathbb{V}r = e^{\omega_{\mathbb{V}}/(1-\phi_{\mathbb{V}})}$, which allows us to write Equation IA.108 as

$$\mathbb{V}r_{t+1} \;\approx \mathbb{V}r \;+\; \phi_{\mathbb{V}} \cdot (\mathbb{V}r_t - \mathbb{V}r) \;+\; \widetilde{\mathbb{V}r}_{t+1} \tag{IA.109}$$

where

$\widetilde{\mathbb{V}r}_t = \mathbb{V}r \cdot \phi_\sigma \cdot \widetilde{\sigma}^2_t$

$\mathbb{V}ar_t[\widetilde{\mathbb{V}r}_{t+1}] = \sigma^2_{\mathbb{V}} = \mathbb{V}r^2 \cdot \phi^2_\sigma \cdot \sigma^2_\sigma$

$\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{V}r}_{t+1}] = \sigma_{r,\mathbb{V}} = \mathbb{V}r \cdot \phi_\sigma \cdot \sigma_{r,\sigma}$

$\mathbb{C}ov_t[\widetilde{dp}^o_{t+1}, \widetilde{\mathbb{V}r}_{t+1}] = \nu_o \cdot \sigma_{dp,\mathbb{V}} = \nu_o \cdot \mathbb{V}r \cdot \phi_\sigma \cdot \sigma_{dp,\sigma}$

$\mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\mathbb{V}r}_{t+1}] = \nu_o \cdot \sigma_{dp,\mathbb{V}} \cdot \xi_{dp,t} \;+\; \sigma_{r,\mathbb{V}} \cdot \xi_{r,t}$

(b) Approximation for $\mathbb{V}ar_t\left[v_{t+1}\right]$ Dynamics

For notation convenience, if $x_t = f(\mathbb{V}r_t, \mathbb{U}\mu_t)$, then we define the expressions $\overline{x} \equiv f(\mathbb{V}r, \mathbb{U}\mu)$, $\partial_{\mathbb{V}r}\overline{x} \equiv \partial_{\mathbb{V}r}\left\{x_t\right\}_{(\mathbb{V}r_t,\mathbb{U}\mu_t)=(\mathbb{V}r,\mathbb{U}\mu)}$, and $\partial_{\mathbb{U}\mu}\overline{x} \equiv \partial_{\mathbb{U}\mu}\left\{x_t\right\}_{(\mathbb{V}r_t,\mathbb{U}\mu_t)=(\mathbb{V}r,\mathbb{U}\mu)}$.

Applying a first order Taylor expansion around $\mathbb{V}r_t \approx \mathbb{V}r$ and $\mathbb{U}\mu_t \approx \mathbb{U}\mu$ on Equation IA.99, we have

$$
\begin{aligned}
\mathbb{V}ar_t\left[v_{t+1}\right] &\approx \overline{\mathbb{V}ar\left[v\right]} \;+\; \partial_{\mathbb{V}r}\overline{\mathbb{V}ar\left[v\right]}\cdot(\mathbb{V}r_t - \mathbb{V}r) \;+\; \partial_{\mathbb{U}\mu}\overline{\mathbb{V}ar\left[v\right]}\cdot(\mathbb{U}\mu_t - \mathbb{U}\mu) \\
&= \left(\overline{\mathbb{V}ar\left[v\right]} - \partial_{\mathbb{V}r}\overline{\mathbb{V}ar\left[v\right]}\cdot\mathbb{V}r - \partial_{\mathbb{U}\mu}\overline{\mathbb{V}ar\left[v\right]}\cdot\mathbb{U}\mu\right) \;+\; \partial_{\mathbb{V}r}\overline{\mathbb{V}ar\left[v\right]}\cdot\mathbb{V}r_t \;+\; \partial_{\mathbb{U}\mu}\overline{\mathbb{V}ar\left[v\right]}\cdot\mathbb{U}\mu_t \\
&= \varphi_0 \;+\; \varphi_{\mathbb{V}}\cdot\mathbb{V}r_t \;+\; \varphi_{\mathbb{U}}\cdot\mathbb{U}\mu_t
\end{aligned}
\tag{IA.110}
$$

where

$$
\begin{aligned}
\varphi_0 \;&=\; \overline{\varphi}_0 \;+\; (\overline{\varphi}_{\mathbb{V}} - \varphi_{\mathbb{V}})\cdot\mathbb{V}r \;+\; (\overline{\varphi}_{\mathbb{U}} - \varphi_{\mathbb{U}})\cdot\mathbb{U}\mu \\
\varphi_{\mathbb{V}} \;&=\; \partial_{\mathbb{V}r}\overline{\varphi}_0 \;+\; \partial_{\mathbb{V}r}\overline{\varphi}_{\mathbb{V}}\cdot\mathbb{V}r \;+\; \overline{\varphi}_{\mathbb{V}} \;+\; \partial_{\mathbb{V}r}\overline{\varphi}_{\mathbb{U}}\cdot\mathbb{U}\mu \\
\varphi_{\mathbb{U}} \;&=\; \partial_{\mathbb{U}\mu}\overline{\varphi}_0 \;+\; \partial_{\mathbb{U}\mu}\overline{\varphi}_{\mathbb{V}}\cdot\mathbb{V}r \;+\; \partial_{\mathbb{U}\mu}\overline{\varphi}_{\mathbb{U}}\cdot\mathbb{U}\mu \;+\; \overline{\varphi}_{\mathbb{U}} \\
\partial_{\mathbb{V}r}\overline{\varphi}_0 \;&=\; -\nu_o\cdot\theta_{\mathbb{E}}\cdot\theta_{\mathbb{V}}\cdot(\gamma-1)\cdot\sigma_{dp,\mathbb{V}}\cdot\partial_{\mathbb{V}r}\overline{\xi}_{dp} \;-\; (\gamma-1)\cdot\theta_{\mathbb{V}}\cdot\theta_{\mathbb{E}}\cdot\sigma_{r,\mathbb{V}}\cdot\partial_{\mathbb{V}r}\overline{\xi}_r \\
\partial_{\mathbb{U}\mu}\overline{\varphi}_0 \;&=\; -\nu_o\cdot\theta_{\mathbb{E}}\cdot\theta_{\mathbb{V}}\cdot(\gamma-1)\cdot\sigma_{dp,\mathbb{V}}\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_{dp} \;-\; (\gamma-1)\cdot\theta_{\mathbb{V}}\cdot\theta_{\mathbb{E}}\cdot\sigma_{r,\mathbb{V}}\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_r \\
\partial_{\mathbb{V}r}\overline{\varphi}_{\mathbb{V}} \;&=\; 2\cdot\theta_{\mathbb{E}}\cdot\nu_o\cdot\left(\nu_{r,dp} + \theta_{\mathbb{E}}\cdot\nu_o\cdot\nu_{dp}^2\cdot\overline{\xi}_{dp}\right)\cdot\partial_{\mathbb{V}r}\overline{\xi}_{dp} + \;2\cdot\theta_{\mathbb{E}}\cdot\left(1+\theta_{\mathbb{E}}\cdot\overline{\xi}_r\right)\cdot\partial_{\mathbb{V}r}\overline{\xi}_r \\
&\quad + 2\cdot\theta_{\mathbb{E}}^2\cdot\nu_o\cdot\nu_{r,dp}\cdot\partial_{\mathbb{V}r}\overline{\xi_r\cdot\xi_{dp}} \\
\partial_{\mathbb{U}\mu}\overline{\varphi}_{\mathbb{V}} \;&=\; 2\cdot\theta_{\mathbb{E}}\cdot\nu_o\cdot\left(\nu_{r,dp} + \theta_{\mathbb{E}}\cdot\nu_o\cdot\nu_{dp}^2\cdot\overline{\xi}_{dp}\right)\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_{dp} + \;2\cdot\theta_{\mathbb{E}}\cdot\left(1+\theta_{\mathbb{E}}\cdot\overline{\xi}_r\right)\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_r \\
&\quad + 2\cdot\theta_{\mathbb{E}}^2\cdot\nu_o\cdot\nu_{r,dp}\cdot\partial_{\mathbb{U}\mu}\overline{\xi_r\cdot\xi_{dp}} \\
\partial_{\mathbb{V}r}\overline{\varphi}_{\mathbb{U}} \;&=\; 2\cdot\theta_{\mathbb{E}}\cdot(1+\theta_{\mathbb{E}}\cdot\overline{\xi}_{dp})\cdot\partial_{\mathbb{V}r}\overline{\xi}_{dp} \;+\; 2\cdot\theta_{\mathbb{E}}\cdot(1+\theta_{\mathbb{E}}\cdot\overline{\xi}_r)\cdot\partial_{\mathbb{V}r}\overline{\xi}_r \;+\; 2\cdot\theta_{\mathbb{E}}^2\cdot\partial_{\mathbb{V}r}\overline{\xi_r\cdot\xi_{dp}} \\
\partial_{\mathbb{U}\mu}\overline{\varphi}_{\mathbb{U}} \;&=\; 2\cdot\theta_{\mathbb{E}}\cdot(1+\theta_{\mathbb{E}}\cdot\overline{\xi}_{dp})\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_{dp} \;+\; 2\cdot\theta_{\mathbb{E}}\cdot(1+\theta_{\mathbb{E}}\cdot\overline{\xi}_r)\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_r \;+\; 2\cdot\theta_{\mathbb{E}}^2\cdot\partial_{\mathbb{U}\mu}\overline{\xi_r\cdot\xi_{dp}} \\
\partial_{\mathbb{V}r}\overline{\xi_r\cdot\xi_{dp}} \;&=\; \partial_{\mathbb{V}r}\overline{\xi_r}\cdot\overline{\xi}_{dp} \;+\; \partial_{\mathbb{V}r}\overline{\xi_{dp}}\cdot\overline{\xi}_r \\
\partial_{\mathbb{U}\mu}\overline{\xi_r\cdot\xi_{dp}} \;&=\; \partial_{\mathbb{U}\mu}\overline{\xi_r}\cdot\overline{\xi}_{dp} \;+\; \partial_{\mathbb{U}\mu}\overline{\xi_{dp}}\cdot\overline{\xi}_r \\
\partial_{\mathbb{V}r}\overline{\xi_{dp}} \;&=\; (\partial_{\mathbb{V}r}\overline{\xi\cdot\xi_{dp}}\cdot\overline{\xi} - \partial_{\mathbb{V}r}\overline{\xi}\cdot\overline{\xi\cdot\xi_{dp}})/\overline{\xi^2} \\
\partial_{\mathbb{U}\mu}\overline{\xi_{dp}} \;&=\; (\partial_{\mathbb{U}\mu}\overline{\xi\cdot\xi_{dp}}\cdot\overline{\xi} - \partial_{\mathbb{U}\mu}\overline{\xi}\cdot\overline{\xi\cdot\xi_{dp}})/\overline{\xi^2} \\
\partial_{\mathbb{V}r}\overline{\xi_r} \;&=\; (\partial_{\mathbb{V}r}\overline{\xi\cdot\xi_r}\cdot\overline{\xi} - \partial_{\mathbb{V}r}\overline{\xi}\cdot\overline{\xi\cdot\xi_r})/\overline{\xi^2} \\
\partial_{\mathbb{U}\mu}\overline{\xi_r} \;&=\; (\partial_{\mathbb{U}\mu}\overline{\xi\cdot\xi_r}\cdot\overline{\xi} - \partial_{\mathbb{U}\mu}\overline{\xi}\cdot\overline{\xi\cdot\xi_r})/\overline{\xi^2}
\end{aligned}
$$

$$
\begin{aligned}
\partial_{\mathbb{V}r}\overline{\xi} \;&=\; 2\cdot\nu_o^2\cdot(\nu_{dp}^2-\nu_{r,dp}^2)\cdot\mathbb{V}r \;+\; (1+\nu_o^2\cdot\nu_{dp}^2-2\cdot\nu_o\cdot\nu_{r,dp})\cdot\mathbb{U}\mu\\
\partial_{\mathbb{U}\mu}\overline{\xi} \;&=\; (1+\nu_o^2\cdot\nu_{dp}^2-2\cdot\nu_o\cdot\nu_{r,dp})\cdot\mathbb{V}r\\
\partial_{\mathbb{V}r}\overline{\xi\cdot\xi_{dp}} \;&=\; 2\cdot\nu_o\cdot(\nu_{\mu,dp}-\nu_{r,\mu}\cdot\nu_{r,dp})\cdot\mathbb{V}r\\
&\quad+(\phi_\mu\cdot(1-\nu_o\cdot\nu_{r,dp})+\nu_o\cdot\nu_{\mu,dp}-\nu_{r,\mu})\cdot\mathbb{U}\mu\\
\partial_{\mathbb{U}\mu}\overline{\xi\cdot\xi_{dp}} \;&=\; (\phi_\mu\cdot(1-\nu_o\cdot\nu_{r,dp})+\nu_o\cdot\nu_{\mu,dp}-\nu_{r,\mu})\cdot\mathbb{V}r\\
\partial_{\mathbb{V}r}\overline{\xi\cdot\xi_r} \;&=\; 2\cdot\nu_o^2\cdot(\nu_{r,\mu}\cdot\nu_{dp}^2-\nu_{\mu,dp}\cdot\nu_{r,dp})\cdot\mathbb{V}r\\
&\quad+(\nu_{r,\mu}+\phi_\mu\cdot\nu_o^2\cdot\nu_{dp}^2-\nu_o\cdot\nu_{\mu,dp}-\phi_\mu\cdot\nu_o\cdot\nu_{r,dp})\cdot\mathbb{U}\mu\\
\partial_{\mathbb{U}\mu}\overline{\xi\cdot\xi_r} \;&=\; (\nu_{r,\mu}+\phi_\mu\cdot\nu_o^2\cdot\nu_{dp}^2-\nu_o\cdot\nu_{\mu,dp}-\phi_\mu\cdot\nu_o\cdot\nu_{r,dp})\cdot\mathbb{V}r
\end{aligned}
$$

(c) Approximation for $\mathbb{U}\mu_t$ Dynamics

Applying a first order Taylor expansion around $\mathbb{V}r_t \approx \mathbb{V}r$ and $\mathbb{U}\mu_t \approx \mathbb{U}\mu$ on Equation IA.74, we have

$$
\begin{aligned}
\mathbb{U}\mu_{t+1} \;&\approx \overline{\mathbb{U}\mu} \;+\; \partial_{\mathbb{U}\mu}\overline{\mathbb{U}\mu}\cdot(\mathbb{U}\mu_t-\mathbb{U}\mu) \;+\; \partial_{\mathbb{V}r}\overline{\mathbb{U}\mu}\cdot(\mathbb{V}r_t-\mathbb{V}r)\\
&= \mathbb{U}\mu \;+\; \phi_{\mathbb{U}}\cdot(\mathbb{U}\mu_t-\mathbb{U}\mu) \;+\; \phi_{\mathbb{U},\mathbb{V}}\cdot(\mathbb{V}r_t-\mathbb{V}r)
\end{aligned}
\tag{IA.111}
$$

where

$$
\begin{aligned}
\mathbb{U}\mu \;&=\; \overline{\mathbb{U}\mu} \;=\; \overline{\phi_0} \;+\; \overline{\phi_{\mathbb{U}}}\cdot\mathbb{U}\mu \;+\; \overline{\phi_{\mathbb{U},\mathbb{V}}}\cdot\mathbb{V}r^{\text{IA.9}}\\
\phi_{\mathbb{U}} \;&=\; \partial_{\mathbb{U}\mu}\overline{\mathbb{U}\mu} \;=\; \partial_{\mathbb{U}\mu}\overline{\phi_{\mathbb{U}}}\cdot\mathbb{U}\mu \;+\; \overline{\phi_{\mathbb{U}}} \;+\; \partial_{\mathbb{U}\mu}\overline{\phi_{\mathbb{U},\mathbb{V}}}\cdot\mathbb{V}r\\
\phi_{\mathbb{U},\mathbb{V}} \;&=\; \partial_{\mathbb{V}r}\overline{\mathbb{U}\mu} \;=\; \partial_{\mathbb{V}r}\overline{\phi_{\mathbb{U}}}\cdot\mathbb{U}\mu \;+\; \partial_{\mathbb{V}r}\overline{\phi_{\mathbb{U},\mathbb{V}}}\cdot\mathbb{V}r_t \;+\; \overline{\phi_{\mathbb{U},\mathbb{V}}}\\
\partial_{\mathbb{V}r}\overline{\phi_{\mathbb{U}}} \;&=\; -\,\phi_\mu\cdot\partial_{\mathbb{V}r}\overline{\xi}_{dp} \;-\; \phi_\mu\cdot\partial_{\mathbb{V}r}\overline{\xi}_r\\
\partial_{\mathbb{U}\mu}\overline{\phi_{\mathbb{U}}} \;&=\; -\,\phi_\mu\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_{dp} \;-\; \phi_\mu\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_r\\
\partial_{\mathbb{V}r}\overline{\phi_{\mathbb{U},\mathbb{V}}} \;&=\; -\,\nu_o\cdot\nu_{\mu,dp}\cdot\partial_{\mathbb{V}r}\overline{\xi}_{dp} \;-\; \nu_{r,\mu}\cdot\partial_{\mathbb{V}r}\overline{\xi}_r\\
\partial_{\mathbb{U}\mu}\overline{\phi_{\mathbb{U},\mathbb{V}}} \;&=\; -\,\nu_o\cdot\nu_{\mu,dp}\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_{dp} \;-\; \nu_{r,\mu}\cdot\partial_{\mathbb{U}\mu}\overline{\xi}_r
\end{aligned}
$$

with expressions for $\partial_{\mathbb{V}r}\overline{\xi}_{dp}$, $\partial_{\mathbb{V}r}\overline{\xi}_r$, $\partial_{\mathbb{U}\mu}\overline{\xi}_{dp}$, and $\partial_{\mathbb{U}\mu}\overline{\xi}_r$ provided in the prior subsection.

[IA.9] Note that this is a non-linear equation for $\mathbb{U}\mu$ since both sides of the equation (including $\overline{\phi_0}$, $\overline{\phi_{\mathbb{U}}}$, and $\overline{\phi_{\mathbb{U},\mathbb{V}}}$) depend on $\mathbb{U}\mu$.

(d) Accuracy of Approximations Used to Derive $N_{\mathbb{V}}$

Figures IA.1(a) to IA.1(c) show that the approximations for the dynamics of $\mathbb{V}r_t$, $\mathbb{U}\mu_t$, and $\mathbb{V}ar_t[v]$ (i.e., Equations IA.109 to IA.111) are fairly accurate given our estimated parameters under the structural ICAPM (with $\gamma = 6.3$ as per our long sample GMM estimation). However, the accuracy that ultimately matters for our purpose is of $N_{\mathbb{V}} = (\mathbb{E}_{t+1} - \mathbb{E}_t)(\sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{V}ar_{t+h}[v_{t+1+h}])$. Figure IA.1(d) plots the $N_{\mathbb{V}}$ expression that results from our approximations (i.e., Equation IA.103) against the same quantity solved numerically following the procedure detailed in Subsection D.7 (but fixing $\gamma = 6.3$). As it is clear from the figure, our solution is very accurate. Hence, it is not surprising that we find (in Subsection D.7) results that are very similar to the ones reported in the main text after solving for the nonlinear $N_{\mathbb{V}}$ numerically. If anything, the nonlinear $N_{\mathbb{V}}$ is slightly more volatile than the approximate $N_{\mathbb{V}}$, which results in slightly stronger $b_{\mathbb{V}}$ in our GMM estimation.

B Econometric Details

This section provides econometric details related to some of the results described in the main text. Subsection B.1 explains the estimation of the wealth return dynamics, Subsection B.2 focuses on our estimation of risk prices, Subsection B.3 shows the equivalence of factor spanning tests and our SDF tests, and Subsection B.4 describes our bootstrap simulations.

B.1 Estimating the Wealth Return Dynamics

This subsection explains how we estimate the wealth return dynamics summarized by the $\mathbb{V}r_t$ and $\mathbb{E}r_t$ processes.

B.1.1 Estimating $\mathbb{V}\boldsymbol{r_t}$

The Realized log-GARCH dynamics for $\mathbb{V}r_t$ are given by

$$log(\mathbb{V}r_{t+1}) \;= \omega_{\mathbb{V}r} \;+\; \phi_{\mathbb{V}r} \cdot log(\mathbb{V}r_t) \;+\; \phi_\sigma \cdot \sigma^2_{t+1} \tag{IA.112}$$

$$\sigma^2_{t+1} \;= \omega_\sigma \;+\; log(\mathbb{V}r_t) \;+\; \widetilde{\sigma}^2_{t+1} \tag{IA.113}$$

where $\widetilde{\sigma}^2_t \overset{iid}{\sim} N(0, \sigma^2_\sigma)$, $\mathbb{C}ov_t[\widetilde{r}_w, \widetilde{\sigma}^2] = \sigma_{r,\sigma}$, and $\mathbb{C}ov_t[\widetilde{dp}^o, \widetilde{\sigma}^2] = \nu_o \cdot \sigma_{dp,\sigma}$.

The $\mathbb{V}r_t$ parameters are $\Theta_\mathbb{V} = (\omega_{\mathbb{V}r}, \omega_\sigma, \phi_{\mathbb{V}r}, \phi_\sigma, \sigma^2_\sigma, \sigma_{r,\sigma}, \sigma_{dp,\sigma})$. We first estimate the parameters $(\omega_{\mathbb{V}r}, \omega_\sigma, \phi_{\mathbb{V}r}, \phi_\sigma)$, which allows us to obtain the $\mathbb{V}r_t$ process and the $\widetilde{\sigma}^2_t$ shocks. We then estimate $(\sigma^2_\sigma, \sigma_{r,\sigma}, \sigma_{dp,\sigma})$ from the sample covariance matrix of our $\widetilde{\sigma}^2$, $\widetilde{r}_w$, and $\widetilde{dp}^o$ shocks (with the construction of $\widetilde{r}_w$ and $\widetilde{dp}^o$ detailed in the next subsection). Note that while we describe the estimation of $(\omega_{\mathbb{V}r}, \omega_\sigma, \phi_{\mathbb{V}r}, \phi_\sigma)$ and $(\sigma^2_\sigma, \sigma_{r,\sigma}, \sigma_{dp,\sigma})$ in two steps, all seven parameters are jointly estimated (the same way that the slope and the residual variance are jointly estimated in a regression) with the $(\sigma^2_\sigma, \sigma_{r,\sigma}, \sigma_{dp,\sigma})$ parameters matching the sample variances and covariances given the $(\omega_{\mathbb{V}r}, \omega_\sigma, \phi_{\mathbb{V}r}, \phi_\sigma)$ parameters.

We estimate $\phi_{\mathbb{V}r}$ and ϕ_σ by Nonlinear Least Squares (NLS) with the objective function

$$\underset{\{\phi_{\mathbb{V}r}\,,\,\phi_\sigma\}}{Min} \sum_{t=1}^{T-H} \left(\sum_{h=1}^{H} RV_{t+h} - \mathbb{E}_t \left[\sum_{h=1}^{H} RV_{t+h} \right] \right)^2 \tag{IA.114}$$

where $RV_t = e^{\sigma_t^2}$ is the realized variance of market returns in month t and $H = 120$ months in our estimation. As such, we are effectively estimating the $\mathbb{V}r_t$ parameters by targeting long-run (10-year) variance dynamics.

The parameter $\omega_{\mathbb{V}r}$ is obtained inside the objective function to impose that the average $\mathbb{V}r_t$ matches the unconditional variance of r_w. Similarly, the parameter ω_σ is obtained inside the objective function to impose $\mathbb{E}[\widetilde{\sigma}^2] = 0$ over our sample period (i.e., $\omega_\sigma = \mathbb{E}[\sigma^2] - \mathbb{E}[log(\mathbb{V}r_t)]$). This approach is equivalent to including $\omega_{\mathbb{V}r}$ and ω_σ as optimizing parameters in the objective function IA.114 while adding two constraints to the optimization (that $\mathbb{E}[\widetilde{\sigma}^2] = 0$ and $\mathbb{E}[\mathbb{V}r_t] = \mathbb{V}ar[r_w]$ must hold).

We now explain how we obtain the $\mathbb{E}_t[\sum_{h=1}^{H} RV_{t+h}]$ term inside the objective function given all relevant parameter values. To start, note that Equation IA.113 implies

$$\mathbb{E}_t\left[RV_{t+1}\right] = \mathbb{E}_t\left[e^{\sigma_{t+1}^2}\right] = e^{\omega_\sigma + 0.5\cdot\sigma_\sigma^2}\cdot\mathbb{V}r_t \tag{IA.115}$$

so that

$$\mathbb{E}_t\left[\sum_{h=1}^{H} RV_{t+h}\right] = e^{\omega_\sigma + 0.5\cdot\sigma_\sigma^2}\cdot\sum_{h=1}^{H}\mathbb{E}_t\left[\mathbb{V}r_{t+h-1}\right] \tag{IA.116}$$

Now, we need an expression for $\mathbb{E}_t[\mathbb{V}r_{t+h}]$ given an arbitrary h. To obtain such an expression, note that combining Equations IA.112 and IA.113 yields

$$log(\mathbb{V}r_{t+1}) = \omega_{\mathbb{V}} + \phi_{\mathbb{V}}\cdot log(\mathbb{V}r_t) + \phi_\sigma\cdot\widetilde{\sigma}_{t+1}^2 \tag{IA.117}$$

where $\omega_{\mathbb{V}} = \omega_{\mathbb{V}r} + \phi_\sigma\cdot\omega_\sigma$ and $\phi_{\mathbb{V}} = \phi_{\mathbb{V}r} + \phi_\sigma$. As such, we have

$$\mathbb{E}_t\left[\mathbb{V}r_{t+1}^{\tau}\right] = \mathbb{E}_t\left[e^{\tau\cdot(\omega_{\mathbb{V}} + \phi_\sigma\cdot\widetilde{\sigma}_{t+1}^2)}\cdot\mathbb{V}r_t^{\tau\cdot\phi_{\mathbb{V}}}\right] = A(\tau)\cdot\mathbb{V}r_t^{\tau\cdot\phi_{\mathbb{V}}} \tag{IA.118}$$

where $A(\tau) = e^{\tau\cdot\omega_{\mathbb{V}} + 0.5\cdot\tau^2\cdot\phi_\sigma^2\cdot\sigma_\sigma^2}$. Finally, we can use the generic Equation IA.118 to obtain

$$\mathbb{E}_t\left[\mathbb{V}r_{t+1}\right] = A(1)\cdot\mathbb{V}r_t^{\phi_{\mathbb{V}}}$$

$$\mathbb{E}_t\left[\mathbb{V}r_{t+2}\right] = \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\mathbb{V}r_{t+2}\right]\right] = \mathbb{E}_t\left[A(1)\cdot\mathbb{V}r_{t+1}^{\phi_{\mathbb{V}}}\right] = A(1)\cdot A(\phi_{\mathbb{V}})\cdot\mathbb{V}r_t^{\phi_{\mathbb{V}}^2}$$

$$\mathbb{E}_t\left[\mathbb{V}r_{t+3}\right] = \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\mathbb{V}r_{t+3}\right]\right] = \mathbb{E}_t\left[A(1)\cdot A(\phi_{\mathbb{V}})\cdot\mathbb{V}r_{t+1}^{\phi_{\mathbb{V}}^2}\right] = A(1)\cdot A(\phi_{\mathbb{V}})\cdot A(\phi_{\mathbb{V}}^2)\cdot\mathbb{V}r_t^{\phi_{\mathbb{V}}^3}$$

and so on, which generalizes to

$$\mathbb{E}_t\left[\mathbb{V}r_{t+h}\right] \;=\; \left(\prod_{j=0}^{h-1} A(\phi_{\mathbb{V}}^{j})\right)\cdot \mathbb{V}r_t^{\phi_{\mathbb{V}}^{h}} \tag{IA.119}$$

and, when combined with Equation IA.116, yields

$$\mathbb{E}_t\left[\sum_{h=1}^{H} RV_{t+h}\right] \;= e^{\omega_\sigma+0.5\cdot\sigma_\sigma^2}\cdot \sum_{h=1}^{H}\left(\prod_{j=0}^{h-2} A(\phi_{\mathbb{V}}^{j})\right)\cdot \mathbb{V}r_t^{\phi_{\mathbb{V}}^{h-1}} \tag{IA.120}$$

Our $\Theta_{\mathbb{V}}$ estimates are reported in Panel B of Table IA.1.

B.1.2 Estimating $\mathbb{E}r_t$

The $\mathbb{E}r_t$ parameters are given by $\Theta_{\mathbb{E}} = (\rho, \mu, \phi_\mu, \phi_g, \nu_\mu, \nu_g, \nu_{r,\mu}, \nu_{r,g}, \nu_{\mu,g})$ and they can be combined with observations of dp_t, $r_{w,t}$, and $\mathbb{V}r_t$ (with the latter obtained in the prior subsection), to recover

$$\mathbb{E}r_{t+1} \;= \mathbb{E}r \;+\; \phi_{\mathbb{E}}\cdot(\mathbb{E}r_t \;-\mathbb{E}r) \;+\; \xi_{dp,t}\cdot \widetilde{dp}^{o}_{t+1} \;+\; \xi_{r,t}\cdot \widetilde{r}^{o}_{w,t+1} \tag{IA.121}$$

and

$$\mathbb{U}\mu_{t+1} \;= \phi_{\mathbb{U},t}\cdot \mathbb{U}\mu_t \;+\; \phi_{\mathbb{U},\mathbb{V},t}\cdot \mathbb{V}r_t \tag{IA.122}$$

where

$\mathbb{E}r = \mu$

$\phi_{\mathbb{E}} = \phi_\mu$

$\Phi_\mu = (1-\phi_\mu^{12})/[(1-\phi_\mu)(1-\rho\cdot\phi_\mu^{12})]$

$\Phi_g = (1-\phi_g^{12})/[(1-\phi_g)(1-\rho\cdot\phi_g^{12})]$

$\nu_o = 1/[\Phi_\mu\cdot(\phi_\mu-\phi_g)]$

$\nu_{dp} = \sqrt{\Phi_\mu^2\cdot\nu_\mu^2+\Phi_g^2\cdot\nu_g^2-2\cdot\Phi_\mu\cdot\Phi_g\cdot\nu_{\mu,g}}$

$\nu_{r,dp} = \Phi_\mu\cdot\nu_{r,\mu}-\Phi_g\cdot\nu_{r,g}$

$\nu_{\mu,dp} = \Phi_\mu\cdot\nu_\mu^2-\Phi_g\cdot\nu_{\mu,g}$

$\nu_{g,dp} = \Phi_\mu\cdot\nu_{\mu,g}-\Phi_g\cdot\nu_g^2$

$\widetilde{dp}^{o}_{t+1} = \mu \;+\; \nu_o\cdot[(dp_{t+1}-dp)-\phi_g\cdot(dp_t-dp)]-\mathbb{E}r_t$

$$\widetilde{r}^{o}_{w,t+1} = r_{w,t+1} \; - \; \mathbb{E}r_t$$

$$\xi_t \;=\; \nu_o^2 \cdot (\nu_{dp}^2 - \nu_{r,dp}^2) \cdot \mathbb{V}r_t^2 \;+\; (1 + \nu_o^2 \cdot \nu_{dp}^2 - 2 \cdot \nu_o \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t \cdot \mathbb{U}\mu_t$$

$$\xi_{dp,t} = (1/\xi_t) \cdot [\nu_o \cdot (\nu_{\mu,dp} - \nu_{r,\mu} \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t^2$$
$$+(\phi_\mu \cdot (1 - \nu_o \cdot \nu_{r,dp}) + \nu_o \cdot \nu_{\mu,dp} - \nu_{r,\mu}) \cdot \mathbb{V}r_t \cdot \mathbb{U}\mu_t]$$

$$\xi_{r,t} = (1/\xi_t) \cdot [\nu_o^2 \cdot (\nu_{r,\mu} \cdot \nu_{dp}^2 - \nu_{\mu,dp} \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t^2$$
$$+(\nu_{r,\mu} + \phi_\mu \cdot \nu_o^2 \cdot \nu_{dp}^2 - \nu_o \cdot \nu_{\mu,dp} - \phi_\mu \cdot \nu_o \cdot \nu_{r,dp}) \cdot \mathbb{V}r_t \cdot \mathbb{U}\mu_t]$$

$$\phi_{\mathbb{U},t} \;=\; \phi_\mu^2 \;-\; \phi_\mu \cdot \xi_{dp,t} \;-\; \phi_\mu \cdot \xi_{r,t}$$

$$\phi_{\mathbb{U},\mathbb{V},t} \;=\; \nu_\mu^2 \;-\; \nu_o \cdot \nu_{\mu,dp} \cdot \xi_{dp,t} \;-\; \nu_{r,\mu} \cdot \xi_{r,t}$$

with initial conditions given by $\mathbb{E}r_0 = \mathbb{E}r$ and $\mathbb{U}\mu_0 = \mathbb{U}\mu$.[IA.10]

We estimate μ and dp (needed for $\rho = e^{-dp}/(1 + e^{-dp})$) from the average values of $r_{w,t}$ and dp_t. We then obtain the other parameters in $\Theta_{\mathbb{E}}$ by Maximum Likelihood Estimation (MLE). Specifically, we have

$$\begin{bmatrix} r_{w,t+1} \\ dp_{t+1} \end{bmatrix} | \, \mathcal{F}_t \; \sim N \left(\begin{bmatrix} \mathbb{E}_t[r_{w,t+1}] \\ \mathbb{E}_t[dp_{t+1}] \end{bmatrix} , \begin{bmatrix} \mathbb{V}ar_t[r_{w,t+1}] & \mathbb{C}ov_t[r_{w,t+1}, dp_{t+1}] \\ \mathbb{C}ov_t[r_{w,t+1}, dp_{t+1}] & \mathbb{V}ar_t[dp_{t+1}] \end{bmatrix} \right) \tag{IA.123}$$

where[IA.11]

$$\mathbb{E}_t[r_{w,t+1}] = \mathbb{E}r_t$$

$$\mathbb{E}_t[dp_{t+1}] = dp \;+\; \phi_g \cdot (dp_t - dp) \;+\; (1/\nu_o) \cdot (\mathbb{E}r_t - \mathbb{E}r)$$

$$\mathbb{V}ar_t[r_{w,t+1}] = \mathbb{V}ar_t[\mu_t + \widetilde{r}_{w,t+1}] = \mathbb{U}\mu_t + \mathbb{V}r_t$$

$$\mathbb{V}ar_t[dp_{t+1}] = \mathbb{V}ar_t[(1/\nu_o) \cdot (\mu_t - \mu) + \widetilde{dp}^{*}_{t+1}] = (1/\nu_o^2) \cdot \mathbb{U}\mu_t + \nu_{dp}^2 \cdot \mathbb{V}r_t$$

$$\mathbb{C}ov_t[r_{w,t+1}, dp_{t+1}] = \mathbb{C}ov_t[\mu_t + \widetilde{r}^{*}_{w,t+1}, (1/\nu_o) \cdot (\mu_t - \mu) + \widetilde{dp}^{*}_{t+1}] = (1/\nu_o) \cdot \mathbb{U}\mu_t + \nu_{r,dp} \cdot \mathbb{V}r_t$$

Hence, letting $u_{t+1} = [r_{w,t+1} - \mathbb{E}_t[r_{w,t+1}] \;,\; dp_{t+1} - \mathbb{E}_t[dp_{t+1}]]' = [\widetilde{r}_{w,t+1} \;,\; \widetilde{dp}_{t+1}]$ and $\Sigma_{u,t}$

[IA.10] To obtain the unconditional μ_t uncertainty ($\mathbb{U}\mu_0 = \mathbb{U}\mu$), we solve Equation IA.122 at steady state. Specifically, we solve for $\mathbb{U}\mu$ based on the non-linear equation $\mathbb{U}\mu = \overline{\phi}_{\mathbb{U}} \cdot \mathbb{U}\mu + \overline{\phi}_{\mathbb{U},\mathbb{V}} \cdot \mathbb{V}r$, with $\mathbb{V}r$ obtained from the prior subsection and $\overline{\phi}_{\mathbb{U}}$ and $\overline{\phi}_{\mathbb{U},\mathbb{V}}$ obtained by substituting $\mathbb{U}\mu_t = \mathbb{U}\mu$ and $\mathbb{V}r_t = \mathbb{V}r$ into $\phi_{\mathbb{U},t}$ and $\phi_{\mathbb{U},\mathbb{V},t}$.

[IA.11] Note that Equation IA.51 implies

$$dp_{t+1} \;= dp \;+\; \phi_g \cdot (dp_t - dp) \;+\; (1/\nu_o) \cdot (\mu_t - \mu) \;+\; \widetilde{dp}^{*}_{t+1}$$

reflect its covariance matrix in Equation IA.123, the likelihood function of the data is

$$det(2 \cdot \pi \cdot \Sigma_{u,t})^{-1/2} \cdot exp\left\{-\frac{1}{2} \cdot u'_{t+1}\Sigma_{u,t}^{-1}u_{t+1}\right\} \tag{IA.124}$$

which implies that the MLE we rely on is given by

$$\Theta_{\mathbb{E}}^{(MLE)} = \underset{\Theta_{\mathbb{E}}}{argmin} \sum_{t=1}^{T}\left[log\left(det(\Sigma_{u,t})\right) + u'_{t+1}\Sigma_{u,t}^{-1}u_{t+1}\right] \tag{IA.125}$$

Our $\Theta_{\mathbb{E}}$ estimates are reported in Panel A of Table IA.1.

B.2 Estimating Risk Prices

Equation 8 shows that $b = \mathbb{E}[M] \cdot \Sigma_f^{-1}\mathbb{E}[f]$ and we estimate b from the sample analogue $\widehat{b} = 1 \cdot \widehat{\Sigma}_f^{-1}\widehat{\mathbb{E}}[f]$ (with a $\mathbb{E}[M] = 1$ normalization). Despite its simplicity, our estimation approach has a clear economic justification (i.e., it represents the projection of the true SDF onto the factors) and can be motivated by efficiency and/or robustness arguments. This subsection elaborates on these aspects and shows how to compute $\widehat{b}$ standard errors.

We begin by motivating our risk price estimation using a linear regression framework. Next, we show how this maps directly into a GMM estimate of the same risk prices and use GMM theory to develop asymptotic standard errors. Finally, we show that this methodology can be motivated by efficiency and robustness arguments. In terms of efficiency, if we were to add other testing assets to the analogous GMM framework, this would leave our estimator unaffected as long as we rely on the efficient GMM weighting matrix. In terms of robustness, we show that our b estimate using the linear regression framework converges in probability to the projection of the SDF onto f even if the $M = a + b'f$ model is mispecified, a result that does not hold for other b estimators.

We consider an environment that is more general than our intertemporal factor model so that we can understand what our estimation procedure uncovers if the intertemporal factor model is false. As such, for the rest of this section, M_t represents the true SDF and $\dddot{M}_t$ reflects a given model for M_t. Obviously, any model's implication is that $M_t = \dddot{M}_t$, but the implication might not hold in reality (i.e., the model might be misspecified), in which case it is important to consider the distinction between M_t and $\dddot{M}_t$. We consider models of the form $\dddot{M}_t = a - b' f_t$, where f_t are excess returns on traded factors and a risk-free asset exists, since this is by far the most common case in the empirical asset pricing literature (and nests our ICAPM risk factor model).

B.2.1 Linear Regression Framework

Since any model of interest is likely misspecified, we want a procedure that is consistent for a and b if $M_t = \dddot{M}_t = a - b' f_t$ holds, but that is "reasonable" (i.e., delivers $\dddot{M}_t = a - b' f_t$ with some useful properties) if $M_t \neq \dddot{M}_t$ holds instead.

To find such a procedure, start by projecting M_t onto f_t, which can always be done whether $\dddot{M}_t$ is well specified or not. Then, we have $M_t = a - b' f_t - \epsilon_t$ where $\mathbb{E}\left[\epsilon \cdot f\right] = 0$ and $\mathbb{E}\left[\epsilon\right] = 0$. Thus, M_t can always be understood as a linear factor model with $k+1$ factors in which the last factor, ϵ_t, is orthogonal to the first k. This means that for any given set of factors, the linearity assumption can always be justified as long as we recognize that we might have a missing factor. In this case, the original misspecification is incorporated into this unobservable factor, making the linearity of M_t on f_t well specified. If $\dddot{M}_t$ is well specified, then we have $M_t = \dddot{M}_t = a - b' f_t$ and $\epsilon_t = 0 \, \forall \, t$. However, if it is misspecified, we have $M_t \neq \dddot{M}_t = a - b' f_t$ and $\dddot{M}_t$ recovers the best linear predictor of M_t given f_t or $\dddot{M}_t = Proj\left(M_t | f_t\right)$. Moreover, ϵ_t does not help pricing f_t.

If we demean M_t, we have $M_t - \mathbb{E}\left[M\right] = -b'\left(f_t - \mathbb{E}\left[f\right]\right)$. Then, a projection of $M_t - \mathbb{E}\left[M\right]$ onto $f_t - \mathbb{E}\left[f\right]$ yields $b = -\Sigma_f^{-1}\mathbb{E}\left[\left(f_t - \mathbb{E}\left[f\right]\right)\left(M_t - \mathbb{E}\left[M\right]\right)\right]$ which, after some algebra, becomes:

$$b = \mathbb{E}\left[M\right] \cdot \Sigma_f^{-1}\mathbb{E}\left[f\right] \tag{IA.126}$$

where Σ_f is the covariance matrix of the factors.

Note that Equation IA.126 is equivalent to the expression for b in Equation 8, and that a can be easily recovered from $a = \mathbb{E}\left[M\right] + b'\mathbb{E}\left[f\right]$. Simply plugging in sample analogues to the moments in these expressions gives a consistent estimator:

$$\begin{cases} \widehat{a} = 1 + \widehat{b}'\widehat{\mathbb{E}}\left[f\right] \\ \widehat{b} = 1 \cdot \widehat{\Sigma}_f^{-1}\widehat{\mathbb{E}}\left[f\right] \end{cases} \tag{IA.127}$$

where $\widehat{\Sigma}_f = \widehat{\mathbb{E}}[(f - \widehat{\mathbb{E}}[f])(f - \widehat{\mathbb{E}}[f])']$ represents the sample covariance matrix, $\widehat{\mathbb{E}}\left[\cdot\right]$ represents the sample average, and we use the normalization $\mathbb{E}\left[M\right] = 1$.

In summary, Equation IA.127 delivers consistent estimates for a and b under the validity of $\dddot{M}_t$ and recovers the projection of M_t onto f_t when $\dddot{M}_t$ is misspecified, transforming the misspecification into an orthogonal missing factor, ϵ_t, such that $\mathbb{E}\left[\epsilon \cdot f\right] = 0$ and $\mathbb{E}\left[\epsilon\right] = 0$.

B.2.2 GMM Framework

We formalize the GMM procedure that leads to the $\widehat{a}$ and $\widehat{b}$ estimates in the previous subsection and provide standard errors for it. For the rest of this subsection, we stack the risk-free payoff (a constant 1) together with the factors. We do the same for prices and parameters (with the normalization that the price of a payoff of 1 equals $\mathbb{E}\left[M\right] = 1$). Thus, we define the new terms $F_t \equiv (1\ ,\ f_t')'$, $P_{Ft-1} \equiv (1\ ,\ 0')'$, and $\theta \equiv (a\ ,\ -b')'$.

We have $k+1$ Euler conditions summarized by $\mathbb{E}[M_t F_t] = \mathbb{E}[P_{Ft-1}]$. Then, substituting $M_t = \theta' F_t - \epsilon_t$ and $\mathbb{E}[\epsilon \cdot F] = 0$ into the Euler conditions yields $\mathbb{E}[\theta' F_t F_t] = \mathbb{E}[P_{Ft-1}]$. Given this just identified system, $\widehat{\theta}$ is the solution to $\widehat{\mathbb{E}}[F_t F_t' \widehat{\theta}] = \widehat{\mathbb{E}}[P_{Ft-1}]$, which is given by:[IA.12]

$$\widehat{\theta} = \widehat{\mathbb{E}}\left[F_t F_t'\right]^{-1} \widehat{\mathbb{E}}\left[P_{Ft-1}\right]. \tag{IA.128}$$

We can use GMM to get the asymptotic covariance matrix for $\widehat{\theta}$ as follows. Let $u_t(\theta) = \theta' F_t F_t - P_{Ft-1}$ and $g(\theta) = \mathbb{E}\left[u_t(\theta)\right]$. Then, from general GMM theory (see Hansen

[IA.12] Note that $\widehat{\mathbb{E}}[F_t F_t' \widehat{\theta}] = \widehat{\mathbb{E}}[P_{Ft-1}]$ is equivalent to

$$\begin{cases} \widehat{\mathbb{E}}\left[\widehat{a} - \widehat{b}' f\right] = 1 \\ \widehat{\mathbb{E}}\left[\left(\widehat{a} - \widehat{b}' f\right) f\right] = 0. \end{cases}$$

which yields the same estimates as in Equation IA.127 once we solve for $\widehat{a}$ and $\widehat{b}$.

(1982)), we have that $aVar(\widehat{\theta}) = \left(\frac{\partial g(\theta)}{\partial \theta}' S^{-1} \frac{\partial g(\theta)}{\partial \theta}\right)^{-1}$, where $S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left[u_t(\theta) u_{t-j}(\theta)'\right]$. Thus, simply substituting terms in the expression for the asymptotic variance-covariance matrix gives:

$$aVar(\widehat{\theta}) = \frac{1}{T}\left\{\mathbb{E}\left[F_t F_t'\right]' \left(\sum_{j=-\infty}^{\infty} \mathbb{E}\left[u_t(\theta) u_{t-j}(\theta)'\right]\right)^{-1} \mathbb{E}\left[F_t F_t'\right]\right\}^{-1}. \tag{IA.129}$$

To estimate the asymptotic variance-covariance matrix, we plug in $\widehat{\mathbb{E}}[F_t F_t']$ as an estimator for $\mathbb{E}[F_t F_t']$, $\widehat{\theta}$ to obtain $u_t(\widehat{\theta})$, and use Newey and West (1987, 1994) to estimate the spectral density matrix (i.e., the infinity summation in Equation IA.129). We then use this $aVar(\widehat{\theta})$ estimate to compute standard errors for our risk price estimates.

B.2.3 Efficiency and Robustness

a) Efficiency

In general, $b = \mathbb{E}[M] \cdot \Sigma_f^{-1}\mu_f$, where μ_f is the vector of risk premia for the factors (see Ludvigson (2013)). Thus, we effectively estimate b by the respective sample moments, $\widehat{b} = 1 \cdot \widehat{\Sigma}_f^{-1}\widehat{\mathbb{E}}[f]$. The first two moments are obviously fine, and thus we should question how efficient it is to estimate μ_f using $\widehat{\mathbb{E}}[f]$.

Suppose we try to estimate μ_f using a cross-sectional regression of betas on average returns of testing assets including not only the factors, but also all other assets in the economy. In this case, the *efficient* Generalized Least Squares (GLS) gives $\widehat{\mu}_f = \mathbb{E}_T[f_t]$ and ignores the cross-section information on all other assets (see Cochrane (2005)). A similar result holds for the maximum likelihood estimator of μ_f in factor regressions (Gibbons, Ross, and Shanken (1989)) and on two-pass regressions (Shanken (1992)). Hence, $\widehat{\mathbb{E}}[f]$ is the efficient choice of estimator for μ_f.

This well-known result extends to our $\widehat{b}$. Suppose we try to estimate b directly using GMM, but relying on a set of testing assets that includes f_t as well as other assets. Then, the *efficient* GMM ignores (asymptotically) all testing assets other than f_t (see Section 3 in Nagel (2013)), which makes our $\widehat{b}$ asymptotically efficient.

b) Robustness

If we include other testing assets in the GMM, then, in finite samples, the estimates for a and b vary with the assets included. This is quite problematic as the estimates will depend not only on the model under analysis, but also on the testing assets used. The problem gets worse if the model is misspecified. In this case, not only are the estimates dependent on the cross-section of assets used, but so is their probability limit (as T grows). This means that, if the model is mispecified, *what* we are estimating varies with the testing assets under GMM.

To see this point, consider a well specified model such that $M_t = \dddot{M}_t = a - b' f_t$. In this case, although $\widehat{a}$ and $\widehat{b}$ depend on the cross-section of asset used, GMM with any set of testing assets provides estimates that converge to the same (and correct) a and b. That is not the case if the model is misspecified. Under misspecification, the $\widehat{a}$ and $\widehat{b}$ converge to the a and b in the projection $M_t = a - b' f_t - \epsilon_t$ if and only if $\widehat{\theta}$ converge to $\mathbb{E}[F_t F_t']^{-1}\mathbb{E}[M_t F_t]$. This is the case with our estimator because $\mathbb{E}[P_{Ft-1}] = \mathbb{E}[M_t F_t]$, and thus $\widehat{\theta} = \widehat{\mathbb{E}}[F_t F_t']^{-1}\widehat{\mathbb{E}}[P_{Ft-1}] \xrightarrow{p} \mathbb{E}[F_t F_t']^{-1}\mathbb{E}[P_{Ft-1}] = \mathbb{E}[F_t F_t']^{-1}\mathbb{E}[M_t F_t]$.

Therefore, our $\widehat{a} - \widehat{b}' f_t$ has the robust interpretation that it always reflects the projection of M_t onto f_t regardless of whether $M_t = \dddot{M}_t$ or not. The same does not hold true for GMM estimators that rely on other testing assets and a non-optimal weighting matrix.

B.3 SDF Projections as Factor Spanning Tests

We would like to understand whether adding f_t to x_t in a factor model improves the pricing of testing assets (i.e., increases the Sharpe ratio of the tangency portfolio). It is well-known in the literature that $f_{k,t}$ adds to x_t if and only if $\alpha_k \neq 0$ and that the sign of α_k tells us whether we would like to long or short f_k. Tests of α_i are called factor spanning tests. Below, we show that the b in the SDF $M_t = a - b' f_t - b_x' x_t$ provides the same economic information as factor spanning tests, with the added advantage that the SDF version of the test controls not only for x_t but also for other f_t factors when testing each f_t factor, which is important in the context of the ICAPM as discussed in Subsection 2.3.

To start, generalize the framework to allow for mispecification as in Subsection B.2. That is, consider $M_t = \dddot{M}_t - \epsilon_t$ in which M_t is the true SDF and $\dddot{M}_t$ is a model under consideration. Moreover, note that the α of any asset r_j relative to any arbitrary model $\dddot{M}_t$ is given by $\alpha_j = -\mathbb{E}[\epsilon \cdot r_j]/\mathbb{E}[\dddot{M}]$.[IA.13] Then, the proposition below assures the SDF projection $M_t = a - b' f_t - b_x' x_t - \epsilon_t$ contains the same economic information as a factor spanning test.

Proposition 2. *The SDF projection* $M_t = a - b' f_t - b_x' x_t - \epsilon_t$ *results in* $b_k = 0$ *if and only if* $\alpha_k = 0$*, where* α_k *represents the pricing error of* $f_{k,t}$ *relative to the model* $\dddot{M}_t = \dddot{a} - \dddot{b}' f_{-k,t} - \dddot{b}_x' x_t$*, with* $f_{-k,t}$ *reflecting all factors in* f_t *except* $f_{k,t}$*.*

Proof. Suppose $b_k = 0$. In this case, estimating $M_t = \dddot{M}_t - \epsilon_t$ results in the same ϵ_t obtained from the SDF projection $M_t = a - b' f_t - b_x' x_t - \epsilon_t$. Then, we have $\alpha_k = -\mathbb{E}[\epsilon \cdot f_k]/\mathbb{E}[\dddot{M}] = 0$ since, by construction, the SDF projection implies $\mathbb{E}[\epsilon \cdot f_k] = 0$.

Alternatively, suppose $\alpha_k = 0$. Then, we must have $\mathbb{E}[\epsilon \cdot f_k] = 0$ since $\alpha_k = -\mathbb{E}[\epsilon \cdot f_k]/\mathbb{E}[\dddot{M}]$. As such, estimating the SDF projection $M_t = a - b' f_t - b_x' x_t - \epsilon_t$ yields $b_k = 0$ because the ϵ_t in $M_t = \dddot{a} - \dddot{b}' f_{-k,t} - \dddot{b}_x' x_t - \epsilon_t$ is already orthogonal to $f_{k,t}$. □

B.4 Details Related to our Bootstrap Analyses

We use bootstrap simulations for two purposes. The first is to estimate t-statistics associated with the Long-Sample risk prices (and risk premia) estimated in Table 6 when applying the Stambaugh (1997) procedure. The second is to estimate sampling statistics related to Sharpe ratios (Table 7) and sum-squared-alpha metrics (Tables 8 and 10) across different factor models. We describe each of the related bootstrapping methodologies below.

B.4.1 Long-Sample Risk Price t-Statistics

The Stambaugh (1997) procedure allows us to estimate both Σ_f and $\mathbb{E}[f]$ needed to compute risk prices according to Equation 8 for a set of factors over the Long Sample (1928-2019)

[IA.13] To see this result, note that $\dddot{M}$ implies $\dddot{\mathbb{E}}[r_j] = -Cov(\dddot{M}, r_j)/\mathbb{E}[\dddot{M}]$, and thus

$$\alpha_j = \mathbb{E}[r_j] - \dddot{\mathbb{E}}[r_j] = \mathbb{E}[r_j] - \frac{Cov(\dddot{M}, r_j)}{\mathbb{E}[\dddot{M}]} = \mathbb{E}[r_j] - \frac{Cov(M, r_j)}{\mathbb{E}[M]} - \frac{Cov(\epsilon, r_j)}{\mathbb{E}[\dddot{M}]} = -\frac{\mathbb{E}[\epsilon \cdot r_j]}{\mathbb{E}[\dddot{M}]}$$

when some of these factors are not available during the early period ("short factors" from the "short sample") where other factors are available ("long factors" from the "Long Sample"). For example, we have ICAPM risk factor data available from 1928-2019, but only have FF5 factor data available from 1963-2019. We use this procedure to estimate ICAPM risk prices over the Long Sample when controlling for factors from the following models: FF5, q4, SY4, DHS3, and q5.

We would like to estimate Σ_f and $\mathbb{E}[f]$ for all factors in both models over the Long Sample so that we can estimate the Long Sample risk prices in Table 6. The Stambaugh (1997) procedure allows us to compute consistent estimates of these variables by projecting short factors onto long factors that are available during both the long and short samples,[IA.14] then using this projection to extrapolate over the Long Sample. The fact that both Σ_f and $\mathbb{E}[f]$ are estimated with error results in complications for the asymptotic theory needed to estimate standard errors on the resulting b estimates. As opposed to using something like the Delta Method to estimate asymptotic standard errors, we instead choose to estimate standard errors using simulation.

We run bootstrap simulations as follows. Let T be the total number of months in the Long Sample (i.e., the number of months from 1928-2019). For a given alternative factor model (such as the FF5 model), let T_S be the total number of months in the short sample over which data for the alternative model is available (e.g., in the case of the FF5 model, the number of months from 1963-2019). We randomly select T_S months of factor data (with replacement) from the the short sample to estimate the projection of the short factors onto the long factors. Next, we randomly select $T - T_S$ months of data from the early period (1928-1962 in the example) with replacement and merge the data with the short-period simulated data to estimate Σ_f and $\mathbb{E}[f]$ following Stambaugh (1997). Finally, we calculate the resulting Long Sample b according to Equation 8. We repeat this procedure 100,000 times and record the estimated b in each simulation. We use the standard deviation from the resulting distribution

[IA.14]We use the ICAPM and FFC4 factors, which are available during the entire Long Sample (1928-2019) as long factors in the Stambaugh (1997) procedure.

of bs as an estimate of the standard error on our b estimated from the full data using the Stambaugh (1997) procedure.

B.4.2 Sharpe Ratio Statistics

We use bootstrap simulations to estimate the percent of times the ICAPM Sharpe ratio is *higher* than that from alternative models under various IS/OS constructions given bootstrapped distributions. These values are reported in parentheses in Table 7

The simulation procedure is slightly different depending on whether we are concerned with IS Sharpe ratios (i.e., those reported for the "Modern Sample", "1st Half", and "2nd Half"), OS Sharpe ratios using weights estimated from 1973-1995 (i.e., those reported for "2nd Half OS (w from 1973-1995)"), or OS Sharpe ratios using weight estimates from 1928-1995 (i.e., those reported for "2nd Half OS (w from 1928-1995)"). We will describe each of the three simulation procedures in turn below. Regardless of the procedure, we repeat it 100,000 and note the resulting Sharpe ratios for each model in each simulation. We then compute the reported metric as the percent of simulations in which the ICAPM Sharpe ratio was higher than the Sharpe ratio for a particular alternative model. These simulations are motivated by a similar sampling procedure in Fama and French (2018) and similar data-splitting procedure in Kan, Wang, and Zheng (2019).

We simulate IS Sharpe ratios for three different periods: (i) The "Modern Sample" from 1973-2019, (ii) The "1st Half" of the Modern Sample from 1973-1995, and (iii) The "2nd Half" of the Modern Sample from 1995-2019. Let the particular period of interest contain T months of factor data. Within each simulation, we randomly sample T months of data from the original sample period (with replacement) and then use these simulated data to compute the maximum Sharpe ratios for the ICAPM and each alternative model. We record these Sharpe ratios and repeat the process 100,000 times. Reported values are the percent of times the ICAPM Sharpe ratio is greater than that for the alternative model across all simulations.

We simulate the "2nd Half OS (w from 1973-1995)" Sharpe ratios as follows.[IA.15] Let there be T months in the Modern Sample from 1973-2019. Within each simulation, we randomly select $T/2$ months of data from the first half of the Modern Sample (i.e., from 1973-1995) with replacement. For each of these randomly-selected months, t, we pair it with month $t + T/2$ from the second half of the Modern Sample (1995-2019). We use the simulated data from the first half of the sample to construct factor weights, w, that produce the maximum in-sample Sharpe ratio for each model, then apply these weights to the paired simulated factor data from the second half of the sample and compute the resulting OS Sharpe ratios for each model. We record these Sharpe ratios and repeat the process 100,000 times. Reported values are the percent of times the ICAPM Sharpe ratio is greater than that for the alternative model across all simulations.

The simulation methodology we use for "2nd Half OS (w from 1928-1995)" is similar to that used for "2nd Half OS (w from 1973-1995)", but allows us to use factor data from before 1973 (where available) to estimate in-sample weights, w. Within each simulation, we select IS and OS data from 1973-1995 and 1995-2019, respectively, in the same manner as for the "2nd Half OS (w from 1973-1995)" simulations. For a given model, let there be N months of data available before 1973. We also randomly sample N months from this data (with replacement) and combine it with the data sampled from 1973-1995. We the use this combined data to estimate factor weights that produce the maximum in-sample Sharpe ratio, then apply these weights to the randomly-sampled data from 1995-2019 and compute the OS Sharpe ratio from this data. In this way, we make use of data available before 1973 (when available), which helps improve the stability of the estimated weights, w. We record these Sharpe ratios and repeat the process 100,000 times. Reported values are the percent of times the ICAPM Sharpe ratio is greater than that for the alternative model across all simulations.

[IA.15] Recall that we use data from 1973-1995 to estimate weights, w, applied to the factors to construct maximum Sharpe ratios in the 1973-1995 sample, and then apply these weights to factor data from 1995-2019 to construct the OS Sharpe ratios.

B.4.3 Comparing $\Sigma\alpha^2/\Sigma\mathbb{E}[r]^2$ Across Factor Models

We use bootstrap simulations to estimate the percent of times the ICAPM $\Sigma\alpha^2/\Sigma\mathbb{E}[r]^2$ ratio is *lower* than those from alternative models under the bootstrapped distribution when the models are applied to various testing assets (Table 8) and 158 anomalies from Chen and Zimmermann (2020) (Table 10).

The simulation procedure is as follows. Let there be T months of data from 1973-2019. In each simulation, we randomly select T months of data (with replacement) from the 1973-2019 period and record the associated factor values and test asset returns. We then compute the $\Sigma\alpha^2/\Sigma\mathbb{E}[r]^2$ metric for each model given the sampled data and record these values. We repeat the process 100,000 times for all results except for those reported in Table 10 related to anomalies. In that case, we repeat the process 10,000 times due to data processing limitations associated with simulating returns from 1,580 anomaly portfolios. In all cases, reported values are the percent of times the ICAPM $\Sigma\alpha^2/\Sigma\mathbb{E}[r]^2$ metric is less than that for the alternative model across all simulations.

C Data Sources and Measurement

This section contains details on the data sources and measurement beyond the information provided in the main text. Subsection C.1 details the aggregate variables necessary to obtain $r_{\mathbb{E}}$, $r_{\mathbb{V}}$, $N_{\mathbb{E}}$, and $N_{\mathbb{V}}$, Subsection C.2 describes our estimation of betas to construct $r_{\mathbb{E}}$, $r_{\mathbb{V}}$, $r_{N\mathbb{E}}$, and $r_{N\mathbb{V}}$, Subsection C.3 focuses on the factor models we study in Sections 3 and 4 (beyond the intertemporal factor model), and Subsection C.4 focuses on the anomaly deciles we explore in Section 4.

C.1 Aggregate Variables used to Measure $r_{\mathbb{E}}$, $r_{\mathbb{V}}$, $N_{\mathbb{E}}$, and $N_{\mathbb{V}}$

Constructing our tradable intertemporal risk factors requires the measurement of dp_t and σ_t^2. Similarly, the estimation of $\mathbb{E}r_t$ and $\mathbb{V}r_t$ (described in Subsection B.1) requires the measurement of $r_{w,t}$, dp_t, and σ_t^2. As such, measuring $r_{\mathbb{E}}$, $r_{\mathbb{V}}$, $N_{\mathbb{E}}$, and $N_{\mathbb{V}}$ for our baseline analysis requires only the measurement of $r_{w,t}$, dp_t, and σ_t^2.

Our monthly log realized variance measure is given by $\sigma_t^2 = log(\frac{21}{N_t} \cdot \Sigma_{i=1}^{N_t} r_{w,t,i}^2)$ with t indexing the month and i the day. $r_{w,t,i}$ is measured from daily log returns on the CRSP value-weighted index and N_t represents the number of trading days on month t.

Our wealth log return and dividend yield, $r_{w,t}$ and dp_t, are based on a value-weighted portfolio containing all common stocks available in the CRSP dataset and their measurement accounts for delistings and mergers and acquisitions (M&A) paid in cash. We do not use the CRSP value-weighted index because accounting for delistings and M&A activity requires a "bottom-up" approach.

We start by adjusting returns for delistings. For each firm for which we can identify a delisting (delisting code available and different from 100), we adjust the (ex- and cum-dividend) return for the month in which the distribution of proceeds took place by assigning the delisting return to that month. If no delisting return is available, we base the delisting return on the findings in Shumway (1997) and assign to the delisting month a return of -30% if the delisting was for cause (delisting code between 400 and 599) and of 0% otherwise. We assign a 0% return to all months between delisting and distribution when there is a temporal gap between the two events.

With ex- and cum-dividend returns accounting for delistings, we construct returns based on a value-weighted equity portfolio. We start by selecting all common shares (share codes 10 and 11) listed on NYSE, NASDAQ, or AMEX (exchange code 1, 2, and 3) and then calculate value-weighted cum- and ex-dividend monthly returns ($R^{cum}_{mon,t}$ and $R^{ex}_{mon,t}$).

Since our dividend measurement accounts for M&A paid in cash (as suggested in Allen and Michaely (2003) and Sabbatucci (2015)), we also construct a monthly "M&A yield" ($M\&Ay$) at the aggregate level. Specifically, each month we sum all proceeds from distributions that can be classified as originating from an M&A paid in cash (distribution code between 3000 and 3400) across all firms that have lagged market equity available, and we divide this value by the sum of the lagged market equity for these firms.

To get dividends that incorporate M&A activity, we calculate a normalized aggregate price series, P_t, by cumulating $R^{ex}_{mon,t} - M\&Ay$. We then calculate dividends from cum- and ex-dividend returns as is standard in the literature (see Koijen and Nieuwerburgh (2011)), but relying on the adjusted ex-dividend return so that $D_{mon,t} = \left(R^{cum}_{mon,t} - R^{ex}_{mon,t} + M\&Ay\right) \cdot P_{t-1}$. Finally, we follow Binsbergen and Koijen (2010) to get the monthly series of annual dividends (D_t) as the sum of the monthly dividends ($D_{mon,t}$) over the respective period.

Our annual dividend yield is given by $dp_t = log(D_t/P_t)$ and our monthly real log return by $r_{w,t} = log((P_t + D_{mon,t})/P_{t-1}) - log(CPI_t/CPI_{t-1})$, where CPI is the consumer price index obtained from the Federal Reserve of St. Louis webpage.

C.2 Betas Estimated on a Rolling Window

As we explain in the main text, we use a 5-year rolling window to estimate β_{dp} and β_{σ^2} and require stocks to have the full five years of data available to be included in the construction of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$. The only exception to the 5-year rolling window rule is in the beginning of our sample period, before we have five years of data on Δdp and $\Delta\sigma^2$ available to estimate betas. Our first observation for Δdp_t is on January/1927 and for $\Delta\sigma^2_t$ is on February/1926. As such, at the end of December/1927, we use stock returns going back to January/1927 for β_{dp} (i.e., 12 months of data) and February/1926 for β_{σ^2} (i.e., 23 months of data). For the subsequent months, we expand the window for each factor (keeping the January/1927 and February/1926 starting points) until we have five years of data to estimate the respective beta. Once we have five years of data to estimate beta (December/1931 for β_{dp} and January/1931 for β_{σ^2}) we start applying the five year rolling window procedure. To assure the universe of stocks used to construct the factors are always the same, we impose that stocks need to have returns available for the entire β_{σ^2} measurement window (which starts earlier than the β_{dp} measurement window before December/1931).

We face a related challenge when estimating $\beta_{N\mathbb{E}}$ and $\beta_{N\mathbb{V}}$ to build the decile portfolios necessary to obtain $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$. Specifically, since $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ are estimated in-sample (using either our Long Sample or our Modern Sample), we cannot estimate $\beta_{N\mathbb{E}}$ and $\beta_{N\mathbb{V}}$ on a 5-year rolling window during the first five years of the relevant sample period. As such, for the first five years in each sample period, we use the $\beta_{N\mathbb{E}}$ and $\beta_{N\mathbb{V}}$ estimated over the first five years to create the deciles instead of using a rolling window. This approach allows us to obtain $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ over the same period we have our tradable risk factors, $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$.

C.3 Replicating Factors from Prominent Factor Models

We obtain the factor data for all factor models from the original authors.[IA.16] We also replicate factors from all the factor models we investigate for two reasons. First, replicating the factors allows us to extend factors from the SY4 and DHS3 models beyond their publicly-available end dates (2016 and 2018, respectively) to the end of our sample (2019). In these cases, we use our replicated versions of the factors from 2017-2019 and in 2019, respectively, in our main results. Second, we need to reconstruct the factors each month to create the net-

[IA.16] 1928-2019 data for the factors in FF3 and FFC4 as well as 1963-2019 data for the factors in FF5 are obtained from Kenneth French's data library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). 1967-2019 data for the factors in q4 and q5 are obtained from the global-q data library (http://global-q.org/index.html). 1963-2016 data for the factors in SY4 are obtained from Robert Stambaugh's webpage (http://finance.wharton.upenn.edu/~stambaug/). 1972-2018 data for the factors in DHS3 are obtained from Lin Sun's webpage (https://sites.google.com/view/linsunhome).

trading-cost factors used to produce results in Figure 4 and in Table 7 (Panel B). We follow Detzel, Novy-Marx, and Velikov (2020) in the cost adjustment, with the necessary details provided in Footnote 34.

We describe how we obtain or construct signals used to create factors from each model we investigate below. Once the signals are constructed, we follow the sorting procedures described in the original papers to create the respective factors. For instance, to construct the FF3 SMB and HML factors, we use the 2x3 sorting procedure described in Fama and French (1993). Note that we need not account for trading costs associated with the market factor since it is a passive strategy that just holds the market portfolio.

a) FF3 factors (SMB and HML)

SMB and HML are constructed annually at the end of June in each year t using the same procedure as in Fama and French (1993). We use market equity, computed by multiplying CRSP shares outstanding (shrout) with absolute price (prc) at the end of June in year t, as the sorting signal for size. The sorting signal for HML is book-to-market equity, where book equity is based on accounting data that is available at the end of December in year $t-1$ and market equity is from the same month. We construct book equity using COMPUSTAT data as described in Fama and French (1993) with two slight modifications. First, we exclude deferred taxes from the calculation for fiscal years starting in 1993 due to changes in tax treatments.[IA.17] Second, we augment the COMPUSTAT book equity data with hand-collected book equity data from Moody's Manuals.[IA.18] Finally, we use NYSE breakpoints for the size and book-to-market signals for sorting, which we obtain from Kenneth French's website. We reconstruct SMB and HML over our Long Sample (1928-2019) and achieve correlations of 99.8% and 99.6% with the original factors, respectively, over that period.

[IA.17] Note that this modification is also used to construct the publicly available SMB and HML factors. Further details can be found on Kenneth French's webpage.

[IA.18] This data is available on Kenneth French's website and is based on Davis, Fama, and French (2000).

b) FFC4 factors (MOM)

MOM is constructed monthly using a stock's cumulative returns over the past 12 months and skipping the last month (i.e., the "12-2 return") as in Carhart (1997). The factor is then constructed using an independent double-sorting procedure that sorts on the momentum measure and market equity, computed by multiplying CRSP shares outstanding (shrout) with absolute price (prc), at the end of each month as in Carhart (1997). We reconstruct MOM over our Long Sample (1928-2019) and achieve a correlation of 99.9% with the original factor over that period.

c) FF5 factors (SMB, CMA, and RMW)

The FF5 model uses the same HML factor as the FF3 model. However, its construction of SMB is slightly different than that in the FF3 model. The FF5 model also adds CMA and RMW factors. Therefore, we reconstruct the FF5 SMB, CMA, and RMW factors annually at the end of June in each year t using the same procedure as in Fama and French (2015). To construct the CMA and RMW, we use COMPUSTAT data to construct asset growth ("investment") and profitability ("operating profitability") signals as described in Fama and French (2015). We use NYSE breakpoints for the investment and operating profitability signals for sorting, which we obtain from Kenneth French's website. We reconstruct the FF5 SMB, CMA, and RMW factors from 1963-2019 and achieve correlations of 99.7%, 98.2%, and 98.9% with the original factors, respectively, during the period in which they are available (1963-2019).

d) q4 and q5 models (ME, IA, ROE, and EG)

We reconstruct the ME, IA, ROE, and EG factors using the same procedure described in Hou, Xue, and Zhang (2015) and Hou et al. (2020) using signals that were provided directly by the authors.[IA.19] Our reconstructed ME, IA, ROE, and EG factors achieve correlations of

[IA.19]We thank the authors of these papers for sharing these data.

99.9%, 99.6%, 99.8%, and 99.7% with the original factors, respectively, during the period in which they are available (1967-2019).

e) SY4 model (SMB, MGMT, and PERF)

We reconstruct the SY4 SMB, MGMT, and PERF factors using the same procedure described in Stambaugh and Yuan (2017). For the SY4 SMB factor, the signal is the same as in the SMB factor of the FF3 model even though the factor construction differs from the approach in FF3. Therefore we rely on the same size signal we use in the FF3 SMB construction to obtain the SY4 size signal. The MGMT factor requires six different signals: net stock issues, composite equity issues, accruals, net operating assets, asset growth, and the investment-to-assets ratio. We obtain four of these signals (accruals, net operating assets, asset growth, and investment-to-asset ratio) from the signal data Chen and Zimmermann (2020) make publicly available (the respective labels are "Accruals", "NOA", "AG", and "InvestPPEInv").[IA.20] We construct the other two signals (net stock issues and composite equity issues) directly from CRSP and COMPUSTAT data following Stambaugh and Yuan (2017) because Chen and Zimmermann (2020)'s construction of the two analogous signals differs slightly from the construction in Stambaugh and Yuan (2017).[IA.21] The PERF factor requires five different signals: financial distress, O-score, 12-2 momentum, gross profitability, and quarterly return on assets. We use the same momentum signal as in our MOM factor. The remaining four signals (financial distress, O-score, gross profitability, and quarterly return on assets) are obtained from the signal data in Chen and Zimmermann (2020) (the respective labels are "FailureProbability", "OScore", "GP", and "roaq"). Given these signals, we construct SMB, MGMT, and PERF as in Stambaugh and Yuan (2017). Our reconstructed SMB, MGMT, and PERF factors achieve correlations of 95.7%, 96.7%, and 93.0% with the original factors, respectively, during the period in which they are available (1963-2016). Note that we reconstruct these factors over the 1963-2019 period and augment the original factors with our reconstructed factors only

[IA.20] See `https://github.com/OpenSourceAP/CrossSection`

[IA.21] Note that the construction of net stock issues and composite equity issues in Stambaugh and Yuan (2017) differs from the construction used in the DHS3 FIN factor of Daniel, Hirshleifer, and Sun (2020).

from 2017 to 2019 in our main results.

f) DHS3 model (FIN and PEAD)

We construct FIN and PEAD factors using the same procedure described in Daniel, Hirshleifer, and Sun (2020). The FIN factor requires two signals: net stock issues and composite equity issues. We construct net stock issues as in Daniel, Hirshleifer, and Sun (2020) directly using COMPUSTAT data.[IA.22] We use the "CompEqIss" signal from Chen and Zimmermann (2020) for composite equity issues since it is constructed in the same way as the corresponding signal in Daniel, Hirshleifer, and Sun (2020). The PEAD factor uses the four-day cumulative returns around earnings announcements as a signal. We use the "AnnouncementReturn" signal from Chen and Zimmermann (2020) since it is constructed in the same way as the corresponding signal in Daniel, Hirshleifer, and Sun (2020). Our reconstructed FIN and PEAD factors achieve correlations of 97.3% and 97.2% with the original factors, respectively, during the period in which they are available (1972-2018). Note that we reconstruct these factors over the 1972-2019 period and augment the original factors with our reconstructed factors only for 2019 in our main results.

C.4 Anomaly Portfolios

We obtain 158 anomaly decile portfolios (that are value-weighted and based on NYSE breakpoints) from Chen and Zimmermann (2020) (see Footnote IA.20), which gives us a total of 1,580 portfolios. In particular, we begin with the 180 "clear predictors" from Chen and Zimmermann (2020), which reflect anomalies that they classify as being "clearly significant in the original papers". From these 180 significant anomalies, we remove anomalies that do not have return records for all 10 decile portfolios for at least half of our 1973-2019 sample. This procedure yields the 158 anomalies (and the corresponding 1,580 decile portfolios) we use as test assets in Section 4.

[IA.22] Note that the construction of net stock issues and composite equity issues in Daniel, Hirshleifer, and Sun (2020) differs from the construction used in the SY4 MGMT factor of Stambaugh and Yuan (2017).

D Supplementary Empirical Results

This section provides empirical results that supplement our analysis in the main text. Subsection D.1 discusses the risk price distortion that arises if one relies on orthogonalized factors in the SDF estimation, Subsection D.2 explores principal components to estimate the mimicking factor weights used as sorting signals in our analysis, Subsection D.3 considers 3-year rolling window betas instead our our baseline 5-year rolling window betas, Subsection D.4 compares our log-GARCH with a level-GARCH specification, Subsection D.5 studies $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ constructed from vector autoregressions, Subsection D.6 considers a maximum likelihood estimation of our log-GARCH process, Subsection D.7 provides results with a nonlinear $N_{\mathbb{V}}$ solved numerically, Subsection D.8 reports the ICAPM risk prices controlling for sentiment and liquidity, Subsection D.9 provides further risk price results controlling for the betting against beta factor, and Subsection D.10 relates our risk price results to some of the prior ICAPM literature.

D.1 Risk Price Distortion with Orthogonalized Factors

In Subsection 2.3, we discuss the fact that our tradable risk factors (r_m, $r_{\mathbb{E}}$, and $r_{\mathbb{V}}$) are strongly correlated. While this result is in line with the ICAPM logic that market prices and investment opportunities vary jointly, it is unusual in the factor model literature as typical factor models are constructed such that their tradable factors are (at least close to) orthogonal. This aspect raises the question of whether we should design our ICAPM factors to be orthogonal. This subsection shows that doing so would distort the risk prices we estimate, and thus break their link to risk aversion and our structural ICAPM more broadly.

We assume that the true SDF is generally given by

$$M_t = a - b' f_t \tag{IA.130}$$

with f_t reflecting a vector of K risk factors with covariance matrix Σ_f.

Now, note that if Λ is the positive definite square-root matrix of Σ_f (i.e., $\Sigma_f = \Lambda\Lambda'$), then we have $\mathbb{V}ar[\Lambda^{-1}f_t] = \Lambda^{-1}\Sigma_f\Lambda^{-1'} = \Lambda^{-1}\Lambda\Lambda'\Lambda^{-1'} = \mathrm{I}$.[IA.23] As such, the orthogonal factors

$$f_t^d = \Gamma\Lambda^{-1}f_t \tag{IA.131}$$

satisfy

$$\mathbb{V}ar[f_t^d] = \begin{bmatrix} \sigma_{1,d}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{2,d}^2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{K,d}^2 \end{bmatrix} \tag{IA.132}$$

so long as Γ is a diagonal matrix with elements $\sigma_{k,d}$

As such, if we create f_t^d, and estimate the SDF $M_t = a - b_d' f_t^d$, then we recover the true SDF in Equation IA.130, but with distorted risk prices

$$b_d = \Gamma^{-1}\Lambda b \tag{IA.133}$$

so that $b_d = b$ if and only if $\Lambda = \Gamma$, which requires the square-root matrix Λ to be diagonal. That is, unless f_t is already orthogonal, relying on orthogonal factors for the SDF estimation will distort risk prices. In the context of our analysis, f_t is unlikely to be orthogonal because market prices and investment opportunities vary jointly, and thus relying on f_t^d would complicate the interpretation of risk prices.

In the main text, we use f_t in our SDF but also consider the risk premia on $f_t^{\perp} = \Psi f_t^o = \Psi\Sigma_f^{-1}f_t$, where Ψ is a diagonal matrix with elements $\sqrt{\mathbb{V}ar[r_m]/\mathbb{V}ar[f_k^o]}$. We refer to $f_t^{\perp}$ as orthogonal factors, but the meaning of "orthogonal" in the context of $f_t^{\perp}$ is different from the meaning we use above when discussing f_t^d. Specifically, $f_t^{\perp}$ are orthogonal to f_t since $\mathbb{C}ov[f^{\perp}, f] = \mathbb{C}ov[\Psi\Sigma_f^{-1}f, f] = \Psi$, but are not orthogonal to each other since $\mathbb{V}ar[f^{\perp}] = \mathbb{V}ar[\Psi\Sigma_f^{-1}f] = \Psi\Sigma_f^{-1}\Psi$. These $f_t^{\perp}$ factors are useful because their variances match the market variance, $\mathbb{V}ar[f_{k,t}^{\perp}] = \mathbb{V}ar[r_m]$, whereas their risk premia are proportional to the f_t risk prices, $\mathbb{E}[f_t^{\perp}] = \Psi\Sigma_f^{-1}\mathbb{E}[f_t] \propto b$. However, using $f_t^{\perp}$ to estimate the SDF $M_t = a - b_{\perp}' f_t^{\perp}$

[IA.23] Note that there exists an infinite number of matrices Γ such that $\Sigma_f = \Gamma\Gamma'$, but there is a unique positive definite Γ such that $\Sigma_f = \Gamma\Gamma'$.

would also lead to risk price distortion since we have $b_\perp = \Psi^{-1}\Sigma_f b$ so that, similar to above, $b_\perp = b$ if and only if $\Sigma_f = \Psi$, which requires Σ_f to be diagonal.

D.2 Mimicking Portfolio Weights from PCA

As explained in Subsections 1.3.3 and 1.3.4, to construct $r_{\mathbb{E}}$, $r_{N\mathbb{E}}$, $r_{\mathbb{V}}$, and $r_{N\mathbb{V}}$, we build basis portfolios by sorting stocks based on their univariate betas relative to the respective non-tradable factors (similar to Herskovic, Moreira, and Muir (2019)). A stock's univariate beta relative to a generic risk factor, x_t, is a natural sorting signal given its direct connection to the mimicking factor weight (see Footnote 19).

To understand whether we can improve upon our univariate beta signals, we now explore a more sophisticated method to estimate the mimicking factor weights (from Giglio and Xiu (2020)) that relies on Principal Component Analysis (PCA). Specifically, over each 5-year rolling window, we obtain the first ten principle components of stock returns and project Δdp and $\Delta\sigma^2$ onto them with the restriction that the final stock weights ($\pi_{\mathbb{E}}$ and $\pi_{\mathbb{V}}$) sum to zero so that we have long-short portfolios. This method provides direct estimates of $\pi_{\mathbb{E}}$ and $\pi_{\mathbb{V}}$, which we use as a replacement for the $\beta_{\Delta dp}$ and $\beta_{\Delta\sigma^2}$ signals we rely on in our baseline analysis. The decile portfolios needed to construct $r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$ follow a similar procedure except that the $\pi_{\mathbb{E}}$ and $\pi_{\mathbb{V}}$ weights are based on projections of $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ onto the principle components.

Table IA.2 adds the resulting factors, $r_{\mathbb{E}}^{PCA}$, $r_{N\mathbb{E}}^{PCA}$, $r_{\mathbb{V}}^{PCA}$, and $r_{N\mathbb{V}}^{PCA}$, to the correlations in Table 1. In general, we have $Cor(r_{\mathbb{E}}^{PCA}, N_{\mathbb{E}}) < Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$ and $Cor(r_{\mathbb{V}}^{PCA}, N_{\mathbb{V}}) < Cor(r_{\mathbb{V}}, N_{\mathbb{V}})$ as well as $Cor(r_{N\mathbb{E}}^{PCA}, N_{\mathbb{E}}) < Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$ and $Cor(r_{N\mathbb{V}}^{PCA}, N_{\mathbb{V}}) < Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$. Moreover, $Cor(r_{\mathbb{E}}^{PCA}, r_{N\mathbb{E}}^{PCA}) < Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$ and $Cor(r_{\mathbb{V}}^{PCA}, r_{N\mathbb{V}}^{PCA}) < Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$. So, all correlations become weaker when we use a more sophisticated method to obtain mimicking factor weights as sorting signals. Moreover, using five or fifteen principle components yields similar results (not tabulated).

It is important to point out that the results described in the prior paragraph are likely related to the fact that using single stocks to construct mimicking factors yields a large cross-section with very few time periods, which can be detrimental to the performance of principle component methods. However, we are reluctant to change from single stocks as base assets to typical anomaly portfolios used in the literature for two reasons. First, most anomaly portfolios can only be constructed after COMPUSTAT variables are available, which (after accounting for some initial training period to estimate the principle components) would only allow us to construct tradable risk factors starting (at the earliest) in the 1980s. Second, the choice of characteristics used to obtain anomalies is subject to serious statistical/empirical problems (see, for example, Berk (2000), Cederburg and O'Doherty (2015), Ang, Liu, and Schwarz (2020), and Tian (2021)) as well as publication biases (see Lo and MacKinlay (1990) and Harvey (2017)). These issues can be avoided by relying on single stocks to construct our risk factors, with Table 1 and Figures 2 and 3 demonstrating that our tradable factors are already good ex-ante proxies for the $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ mimicking portfolios.

D.3 Exposure to Δdp and $\Delta\sigma^2$ using Three-Year Rolling Windows

In this subsection, we construct portfolios and factors based on stock exposures to Δdp and $\Delta\sigma^2$ (β_{dp} and β_{σ^2}, respectively) using 3-year rolling estimation windows rather than 5-year windows as in our baseline analysis.

Table IA.3 reports results analogous to those in Tables 2 and 3 when stocks are sorted into decile portfolios based on either β_{dp} (Panel A) or β_{σ^2} (Panel B) estimated using the 3-year rolling windows. As in our main results, these portfolios sort well on ex-post exposure to the original expected return and variance news proxies (Δdp and $\Delta\sigma^2$) as well as the news components themselves ($N_{\mathbb{E}}$ and $N_{\mathbb{V}}$). For brevity, we only provide results for our Long Sample (1928-2019), but the results for our Modern Sample (1973-2019) are also similar to our main results for that period.

Table IA.4 reports estimated ICAPM risk prices and risk premia when the ICAPM factors are constructed from stocks based on β_{dp} and β_{σ^2} estimated using the 3-year rolling windows. Results are quantitatively similar to our main results (with 5-year rolling windows), implying that the consistency of ICAPM factor model risk prices with ICAPM theory is robust to modifying the β estimation window.

D.4 $\mathbb{V}r_t$ from Log-GARCH vs Level-GARCH

As described in Subsection A.3.5, we model the market variance, $\mathbb{V}r_t$, as a log-GARCH and demonstrate that our $log(\mathbb{V}r_t)$ process can be derived as the Bayesian posterior for the log conditional variance in a stochastic volatility model in which log realized variance, σ_t^2, provides a noisy signal for the unobservable log conditional variance. However, we could have alternatively modeled $\mathbb{V}r_t$ as a level-GARCH and used the same latent stochastic volatility framework to provide a Bayesian interpretation for the model so long as we interpreted $\overline{\sigma}_t$ and σ_t as the level (and not the log) of the realized and conditional variances respectively. This subsection explains why we use a log-GARCH instead of a level-GARCH in our empirical analysis.

First, while level-GARCH models are often used in the literature for their linearity, there is little tractability gains in our framework given that the closed-form solution $N_{\mathbb{V}} \approx \theta_{\mathbb{V}} \cdot \widetilde{\sigma}_t$ requires linear approximations whether we use a log-GARCH or a level-GARCH. As such, the key benefit of level-GARCH models does not apply in our context while the main benefit of the log-GARCH framework (i.e., better ability to forecast variance) is present in our setting as can be seen in Figures IA.2(a) and Figures IA.2(b). Specifically, these figures demonstrate that the sum of squared errors (when forecasting subsequent realized variance at different horizons) of a level-GARCH are higher than those of a log-GARCH.

And second, our Bayesian framework requires normality of the unobservable shocks (otherwise the Bayesian updating is non-linear), which implies normality of $\widetilde{\sigma}_t^2$ in our GARCH model. While asymptotically (as the number of observations used to estimate σ_t^2 grows) $\widetilde{\sigma}_t^2$ converges to a normal distribution whether σ_t^2 is the realized variance or its log (Barndorff-Nielsen and Shephard (2002)), in small samples nether model yields normal σ_t^2 shocks, but $\widetilde{\sigma}_t^2$ is approximately normal in the log-GARCH model whereas it is far from normal in the level-GARCH model (Andersen et al. (2001a,b) and Barndorff-Nielsen and Shephard (2005)). Figures IA.2(c) and IA.2(d) provide a visualization of this well-established result in the context of our empirical analysis (we estimate the level-GARCH using a procedure analogous to our log-GARCH estimation in Subsection B.1.1). While neither model yields normal shocks, the log-GARCH $\widetilde{\sigma}_t^2$ can be approximated by a normal distribution so that our Bayesian interpretation is reasonable whereas the level-GARCH $\widetilde{\sigma}_t^2$ is far from normally distributed, which would render the Bayesian interpretation much less reasonable.

D.5 $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ from Vector Autoregressions

In the main text, we avoid the ICAPM "fishing license" (Fama (1991)) when constructing $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ as we do not specify an arbitrary set of state variables for expected returns and volatility, but instead build a simple Bayesian learning framework in which a long-term investor observes only market prices and dividends (similar to Binsbergen and Koijen (2010)). In our framework, $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ are ultimately linked to log dividend yield shocks and log realized variance shocks, respectively. While our approach is parsimonious and theoretically founded, much of the prior ICAPM literature relies on vector autoregressions (VARs) with multiple state variables that have some logical link to expected returns and volatility (e.g., Campbell et al. (2018) and Gonçalves (2021a)).

In this subsection, we show that our tradable intertemporal risk factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) are good ex-ante mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ even if these news terms are constructed from a VAR. To estimate the VAR, we assume that $z_t = (r_{w,t},\, s_t)$ evolves as (with s_t reflecting a state vector that includes dp_t and $RV_t = e^{\sigma_t^2}$)

$$z_{t+1} - \overline{z} \ = B(z_t - \overline{z}) \ + \ \widetilde{z}_{t+1} \tag{IA.134}$$

We then estimate B in-sample (using OLS equation-by-equation) and define $B_\infty = \delta \cdot B\,(I - \delta \cdot B)^{-1}$, $1_r' z_t = r_{w,t}$, and $1_{RV}' z_t = RV_t$, so that news to long-term expected returns and volatility are given by $N_{\mathbb{E},t} = 1_r' B_\infty \widetilde{z}_t$ and $N_{\mathbb{V},t} \propto N_{RV} = 1_{RV}' B_\infty \widetilde{z}_t$. Following Gonçalves (2021a), we fix the first B column to be zero before the estimation because r_w is not a good predictor of future returns or variance, but the $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ constructed without this step are very similar. In our most extensive specification, s_t contains dp_t, RV_t, and four extra state variables (all measured in natural log units): the credit spread (CS), Treasury yield (TY), term spread (TS), and value spread (VS).[IA.24]

Table IA.5 reports $Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$, $Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$, and $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$ (as well as the analogous $N_{\mathbb{V}}$, $r_{N\mathbb{V}}$, and $r_{\mathbb{V}}$ correlations) over our long and modern sample periods for different VAR specifications (as well as for our Bayesian framework for comparability). Moreover, Figure IA.3 provides ex-post $\beta_{rN\mathbb{E}}$ and $\beta_{rN\mathbb{V}}$ of decile portfolios sorted on Δdp and $\Delta\sigma^2$.

Table IA.5, Panel B sets $s_t = (dp_t,\ \sigma_t^2)$ so that the dividend yield and realized variance are the only state variables used to construct $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$. The idea behind this specification is to check whether cross-predictability (i.e., dp_t forecasting realized variance and σ_t^2 forecasting returns) makes a VAR specification sufficiently different from our Bayesian framework as to render $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ bad at mimicking $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$. We find that this is not the case. Specifically, $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$ and $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$ are quite high in this specification (0.95 and 0.83 over the Long Sample and 0.90 and 0.76 over the Modern Sample) and Figure IA.3 shows large spreads in betas across the decile portfolios, suggesting that cross-predictability is not a problem for our analysis.

Table IA.5, Panel C sets $s_t = (dp_t,\ \sigma_t^2,\ CS_t)$ to explore a specification that contains the three state variables that, according to Campbell et al. (2018), jointly capture long-term volatility.[IA.25] We continue to observe high values for $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$ and $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$ (0.93

[IA.24] Our measurement of these variables largely follows Gonçalves (2021a). The credit spread (CS_t) is the difference between Moody's corporate BAA and AAA log yields with both coming from the Federal Reserve of St. Louis website. The Treasury yield (TY) is the annualized 3-month log Treasury bill rate and comes from Global Financial Data until December 1933 and from the Federal Reserve of St. Louis webpage after that. The term spread (TS) is the difference between the 10-year log Treasury yield and TY, where the former comes from Global Financial Data until March 1953 and from the Federal Reserve of St. Louis website thereafter. The value spread (VS) is the difference between the log book-to-market ratios of the value and growth portfolios formed based on small stocks and adjusting for within-year movements in market equity. The data come from Kenneth French's data library and measurement follows Campbell and Vuolteenaho (2004).

[IA.25] Campbell et al. (2018) argue that long-term volatility news is jointly captured by three state variables: the realized volatility, the smooth earnings yield, and the credit spread. Our specification uses the same three variables, except that, following Gonçalves (2021a), we use dp_t as a valuation ratio instead of the smooth earnings yield. The reason is that, as our Bayesian framework makes it clear, dp_t provides a signal for expected returns. Moreover, Campbell (1991) log-linear approximation for stock returns implies that state vectors that do not span dp_t are mispecified (see Engsted, Pedersen, and Tanggaard (2012) for details).

and 0.77 over the Long Sample and 0.89 and 0.75 over the Modern Sample) and large beta spreads across decile portfolios (see Figure IA.3), results that indicate that our tradable intertemporal factors continue to capture well ex-post mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ even when the credit spread is included in the state vector.

Table IA.5, Panel D sets $s_t = (dp_t,\ \sigma_t^2,\ CS_t,\ TY_t,\ TS_t)$, which includes the three state variables that capture long-term volatility (according to Campbell et al. (2018)) as well as the Treasury yield, TY, and the term spread, TS, which are classical interest rate and equity premium predictors (Fama and Schwert (1977), Fama (1981), Campbell (1987), and Fama and French (1989)). Remarkably, the $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$ and $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$ values remain high and very similar to what we observe in Panels B and C (0.92 and 0.77 over the Long Sample and 0.90 and 0.77 over the Modern Sample). Moreover, Figure IA.3 continues to display beta spreads across decile portfolios that are similar to what we observe for other specifications. As such, our tradable intertemporal factors capture well ex-post mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ even when ex-post news are constructed from a multivariate system that includes several traditional state variables in the predictability literature.

Table IA.5, Panel E sets $s_t = (dp_t,\ \sigma_t^2,\ CS_t,\ TY_t,\ TS_t,\ VS_t)$, which includes all state variables in our analysis. The beta spreads in Figure IA.3 remain large and the $Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$ level remains very high (0.91 over the Long Sample and 0.86 over the Modern Sample). However, $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$ sharply declines (0.46 over the Long Sample and 0.63 over the Modern Sample). This result indicates that there is a component of long-run variance captured by the value spread that is not fully reflected in our $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ factors. Interestingly, this limitation of $r_{\mathbb{V}}$ as a proxy for $r_{N\mathbb{V}}$ under this particular VAR specification concentrates in the late 1990s and early 2000s, which is the tech boom and bust period, suggesting that it is an isolated incident. In fact, $Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$ is 0.63 over our Long Sample and 0.70 over our Modern Sample if we exclude the years from 1999 to 2002. So, overall, we conclude that $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ are generally good ex-ante proxies for the $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ mimicking factors even in a specification that includes all state variables we explore.

The results in Table IA.5 and Figure IA.3 indicate that our interpretation of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ as mimicking factors for $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ is robust to the use of a VAR framework to estimate $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$. However, the strength of the $r_{\mathbb{V}}$ interpretation does weaken a bit as we include the value spread in the VAR. To address this issue, we perform an extra exercise that asks whether VAR models with several state variables are better at forecasting long-term variance than our simple log-GARCH framework. Specifically, Figures IA.4(a) and IA.4(b) display the sum of squared forecasting errors for different VAR specifications (relative to the same quantity for our log-GARCH model) at different forecasting horizons from 1 month to 120 months (10 years). As it is clear from the figures, our simple log-GARCH model is better at forecasting long-run variance than all VAR specifications we explore. Moreover, the VAR that includes the value spread is not the best performing VAR in terms of predicting long-run variance. As such, our long-run variance predictability results limit any remaining concerns with the interpretation of our $r_{\mathbb{V}}$ factor.

D.6 ICAPM Risk Prices with log-GARCH Estimated by MLE

Using our log-GARCH, we find that the $r_{\mathbb{V}}$ risk price that is restricted by the ICAPM (with γ as the only parameter) is large and comparable to the unrestricted $r_{\mathbb{V}}$ risk price. This result may seem surprising given that prior literature (e.g., Chacko and Viceira (2005)) finds that news to long-term variance induces a quantitatively small hedging demand when conditional variance is modeled as a univariate autoregressive process.

The crux of the matter is that, as explained in Subsection B.1.1, we estimate our log-GARCH process by targeting long-run (10-year) realized variance. The underlying idea is that targeting long-run expectations provides a robust estimation method if we recognize that any autoregressive model is likely mispecified (see Ederington and Guan (2010) and Jordà and Kozicki (2011)). To demonstrate that the targeting of long-run variance is the driver of the strong $r_{\mathbb{V}}$ risk price, we reestimate the log-GARCH by maximum likelihood, which effectively focuses on short-term variance dynamics as it assumes that the intertemporal variance dynamics in the log-GARCH specification are exactly correct.[IA.26] We then reestimate the

[IA.26] The log-GARCH implies $\sigma^2_{t+1} \sim N(\omega_\sigma + log(\mathbb{V}r_t), \sigma^2_\sigma)$ so that the maximum likelihood estimation that

$N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ dynamics using this alternative $\mathbb{V}r_t$ process and obtain new ICAPM risk prices, with the final results reported in Table IA.6 under the column "MLE $N_{\mathbb{V}}$". As it is clear from the table, the $r_{\mathbb{V}}$ risk price is quantitatively small (albeit significant) when the log-GARCH process is estimated by maximum likelihood, a result that is in line with the weak volatility hedging channel in the prior literature that models conditional variance as a univariate autoregressive process (e.g., Chacko and Viceira (2005)).

Figures IA.4(c) and IA.4(d) show why estimating our log-GARCH by maximum likelihood is not appropriate if the main objective of the estimation is to capture long-term variance dynamics. In particular, VAR models tend to perform better than this log-GARCH estimated by maximum likelihood at forecasting long-term variance (for most horizons), in stark contrast with the results we find in Figures IA.4(a) and IA.4(b) when we estimate the log-GARCH by a NLS that targets long-run variance.

D.7 ICAPM Risk Prices with Nonlinear $N_{\mathbb{V}}$

Our baseline estimate of ex-post variance news, $N_{\mathbb{V}} \approx \theta_\sigma \cdot \widetilde{\sigma}^2_\sigma$, relies on three linear approximations (see Subsection A.4.3 for details on the approximations). In this subsection, we solve for vw_t numerically without relying on these linearizations and use the resulting solution to obtain a nonlinear $N_{\mathbb{V}}$. We then use this nonlinear $N_{\mathbb{V}}$ in our ICAPM estimation, with the results reported in Table IA.6 under the column "Nonlinear $N_{\mathbb{V}}$". As can be seen by comparing the "Baseline $N_{\mathbb{V}}$" and "Nonlinear $N_{\mathbb{V}}$" columns, this nonlinear $N_{\mathbb{V}}$ yields results that are quantitatively similar to (and even a bit stronger than) our baseline results. The rest of this subsection explains the numerical procedure we rely on to obtain our nonlinear $N_{\mathbb{V}}$.

Let $r_{w,t}$ reflect wealth portfolio returns and s_t be a state vector so that $z_t = [r_{w,t}, s_t]$ is

---

we rely on is based on

$$\underset{\{\omega_{\mathbb{V}r},\omega_\sigma,\phi_{\mathbb{V}r},\phi_\sigma,\sigma^2_\sigma\}}{Min} \sum_{t=1}^{T} \left[log\left(\sigma^2_\sigma\right) + \frac{\left(\sigma^2_{t+1} - \omega_\sigma - log(\mathbb{V}r_t)\right)^2}{\sigma^2_\sigma} \right]$$

with $log(\mathbb{V}r_{t+1}) = \omega_{\mathbb{V}r} + \phi_{\mathbb{V}r} \cdot log(\mathbb{V}r_t) + \phi_\sigma \cdot \sigma^2_{t+1}$.

modeled as

$$r_{w,t+1} \; = \mathbb{E}r(s_t) \; + \widetilde{r}_{w,t+1} \tag{IA.135}$$

$$s_{t+1} \; = \mathbb{E}s(s_t) \; + \widetilde{s}_{t+1}. \tag{IA.136}$$

$$\mathbb{V}ar_t[\widetilde{r}_{w,t+1}] \; = \mathbb{V}r(s_t) \tag{IA.137}$$

$$\mathbb{V}ar_t[\widetilde{s}_{t+1}] \; = \mathbb{V}s(s_t) \tag{IA.138}$$

$$\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{s}_{t+1}] \; = \mathbb{C}rs(s_t) \tag{IA.139}$$

where the functions $\mathbb{E}r(s_t)$, $\mathbb{E}s(s_t)$, $\mathbb{V}r(s_t)$, $\mathbb{V}s(s_t)$, and $\mathbb{C}rs(s_t)$ are known and can be measured in the data.

We know that the log value-wealth ratio is a function of s_t, $vw_t = vw(s_t)$, and we would like to solve for the $vw(s_t)$ function using the recursive Equation IA.36, which is reproduced here for convenience:

$$vw_t \; = f(\psi, \delta, \overline{\delta}) \; + \overline{\delta} \cdot \mathbb{E}_t[r_{w,t+1}] \; + \overline{\delta} \cdot \mathbb{E}_t[vw_{t+1}] \; - \overline{\delta} \cdot \frac{(\gamma - 1)}{2} \cdot \mathbb{V}ar_t[v_{t+1}] \tag{IA.140}$$

First, let's rewrite the recursive equation in a way that makes the dependence on s_t explicit:

$$\begin{aligned} vw(s_t) \; &= f(\psi, \delta, \overline{\delta}) \; + \overline{\delta} \cdot \; \mathbb{E}r(s_t) \; + \; \overline{\delta} \cdot \mathbb{E}_t[vw(s_{t+1})] \\ &\quad - \; \overline{\delta} \cdot \frac{(\gamma - 1)}{2} \cdot (\mathbb{V}r(s_t) \; + \; \mathbb{V}ar_t[vw(s_{t+1})] \; + \; 2 \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, vw(s_{t+1})]) \end{aligned} \tag{IA.141}$$

with this result relying on $\mathbb{V}ar_t[v_{t+1}] = \mathbb{V}ar_t[r_{w,t+1} + vw_{t+1}]$. For notational convenience, the rest of the derivations suppress the dependence of some of the functions on s_t by defining $\mathbb{E}s_t = \mathbb{E}s(s_t)$ and $\mathbb{V}s_t = \mathbb{V}s(s_t)$.

In principle, Equation IA.141 can be solved numerically as a function iteration on the data grid with interpolation for $vw(s_t)$ and numerical integration to compute the vw moments, $\mathbb{E}_t[vw(s_{t+1})]$, $\mathbb{V}ar_t[vw(s_{t+1})]$, and $\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, vw(s_{t+1})]$. However, depending on the application, this approach is too costly from a computational standpoint. We now rely on a conditional second order Taylor approximation for $vw(s_{t+1})$ to solve for the vw moments so that the recursion can be efficiently evaluated.

Given a current guess for $vw(s_t)$ (we later detail the algorithm to initiate and update the $vw(s_t)$ function), we have

$$vw_0(s_{t+1}, \mathbb{E}s_t) \approx vw(\mathbb{E}s_t) + \mathcal{G}(\mathbb{E}s_t)'(s_{t+1} - \mathbb{E}s_t) + \frac{1}{2} \cdot (s_{t+1} - \mathbb{E}s_t)'\mathcal{H}(\mathbb{E}s_t)(s_{t+1} - \mathbb{E}s_t) \quad \text{(IA.142)}$$

where $\mathcal{G}(x)$ is the vw gradient vector evaluated at x and $\mathcal{H}(x)$ is the vw hessian matrix evaluated at x.

Using the approximation in Equation IA.142, we then have:[IA.27]

$$\mathbb{E}_t[vw(s_{t+1})] \;= vw(\mathbb{E}s_t) \;+\; \frac{1}{2} \cdot tr\left(\mathcal{H}(\mathbb{E}s_t)\mathbb{V}s_t\right) \quad \text{(IA.143)}$$

$$\mathbb{V}ar_t[vw(s_{t+1})] \;= \mathcal{G}(\mathbb{E}s_t)\mathbb{V}s_t\mathcal{G}(\mathbb{E}s_t)' \;+\; \frac{1}{2} \cdot tr\left(\mathbb{V}s_t\mathcal{H}(\mathbb{E}s_t)\mathbb{V}s_t\mathcal{H}(\mathbb{E}s_t)\right) \quad \text{(IA.144)}$$

$$\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, vw(s_{t+1})] \;= \mathcal{G}(\mathbb{E}s_t)'\mathbb{C}rs(s_t) \quad \text{(IA.145)}$$

where $tr(X)$ is the trace of matrix X.

Substituting Equations IA.143, IA.144, and IA.145 into Equation IA.141, we have

$$\begin{aligned} vw(s_t) \;&= f(\psi, \delta, \overline{\delta}) \;+\; \overline{\delta} \cdot \mathbb{E}r(s_t) \;+\; \overline{\delta} \cdot \left(vw(\mathbb{E}s_t) + \frac{1}{2} \cdot tr\left(\mathcal{H}(\mathbb{E}s_t)\mathbb{V}s_t\right)\right) \quad \text{(IA.146)} \\ &- \;\overline{\delta} \cdot \frac{(\gamma - 1)}{2} \cdot \mathbb{V}r(s_t) \\ &- \;\overline{\delta} \cdot \frac{(\gamma - 1)}{2} \cdot \left(\mathcal{G}(\mathbb{E}s_t)\mathbb{V}s_t\mathcal{G}(\mathbb{E}s_t)' + \frac{1}{2} \cdot tr\left(\mathbb{V}s_t\mathcal{H}(\mathbb{E}s_t)\mathbb{V}s_t\mathcal{H}(\mathbb{E}s_t)\right)\right) \\ &- \;\overline{\delta} \cdot (\gamma - 1) \cdot \mathcal{G}(\mathbb{E}s_t)'\mathbb{C}rs(s_t) \end{aligned}$$

We now specialize the state vector to include the state variables in our model. In particular,

[IA.27] Note that our second order Taylor approximation is conditional and is only being used to approximate the vw moments: $\mathbb{E}_t[vw(s_{t+1})]$, $\mathbb{V}ar_t[vw(s_{t+1})]$, and $\mathbb{C}ov_t[\widetilde{r}_{w,t+1}, vw(s_{t+1})]$. As such, our numerical solution method is still global and our $vw(s_t)$ function is still allowed to be arbitrarily nonlinear. This approach is commonly use in the nonlinear filtering literature (see Chapter 8 of Anderson and Moore (1979)).

we have $s_t = [\mathbb{E}r_t,\ log(\mathbb{V}r_t),\ \mathbb{U}\mu_t]$ with

$$\mathbb{E}r(s_t) \ = \mathbb{E}r_t, \tag{IA.147}$$

$$\mathbb{E}s(s_t) \ = \begin{bmatrix} \mathbb{E}r \ + \ \phi_{\mathbb{E}} \cdot (\mathbb{E}r_t - \mathbb{E}r) \\ \omega_{\mathbb{V}} \ + \ \phi_{\mathbb{V}} \cdot log(\mathbb{V}r_t) \\ \phi_{\mathbb{U},t} \cdot \mathbb{U}\mu_t \ + \ \phi_{\mathbb{U},\mathbb{V},t} \cdot \mathbb{V}r_t \end{bmatrix}, \tag{IA.148}$$

$$\mathbb{V}r(s_t) \ = \mathbb{V}ar_t[\widetilde{r}_{w,t+1}], \tag{IA.149}$$

$$\mathbb{V}s(s_t) \ = \begin{bmatrix} \mathbb{V}ar_t[\widetilde{\mathbb{E}r}_{t+1}] & \phi_\sigma \cdot \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\sigma}^2_{t+1}] & 0 \\ \phi_\sigma \cdot \mathbb{C}ov_t[\widetilde{\mathbb{E}r}_{t+1}, \widetilde{\sigma}^2_{t+1}] & \phi_\sigma^2 \cdot \mathbb{V}ar_t[\widetilde{\sigma}^2_{t+1}] & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and} \tag{IA.150}$$

$$\mathbb{C}rs(s_t) \ = \begin{bmatrix} \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\mathbb{E}r}_{t+1}] \\ \phi_\sigma \cdot \mathbb{C}ov_t[\widetilde{r}_{w,t+1}, \widetilde{\sigma}^2_{t+1}] \\ 0 \end{bmatrix}. \tag{IA.151}$$

The expressions for all variance and covariance terms above are provided below Equation IA.97 and are functions of $\mathbb{V}r_t$ and $\mathbb{U}\mu_t$ (and thus of s_t). Note that $\phi_{\mathbb{U},t}$, $\phi_{\mathbb{U},\mathbb{V},t}$, $\xi_{dp,t}$, and $\xi_{r,t}$ are also functions of $\mathbb{V}r_t$ and $\mathbb{U}\mu_t$ (and thus of s_t).

Equation IA.146 provides an algorithm for solving for $vw(s_t)$ as a function of γ:

1. Create a three-dimensional grid that covers reasonable state space variable values[IA.28]

[IA.28] To define the state space upper and lower bounds, we simulate the state vector given parameter estimates for 100,000 months. We take the resulting maximum and minimum simulated state variable values as the boundaries in the state space on which we solve for the value function. Furthermore, we create a $11 \times 11 \times 11$ three-dimensional state space grid using equal-spacings across each state variable dimension. We choose to use 11 grid points in each state variable dimension to balance granularity of our solutions with computation time. In practice, we find that the converged vw_t is smooth with low curvature, so this grid spacing is unlikely to significantly impact our results related to the nonlinear $N_{\mathbb{V}}$.

2. Start from an arbitrary function $vw^{n-1}(s_t)$ at points on the grid (e.g., $vw^{n-1}(s_t) = 0$, or a previous solution possibly from a different γ_0 close to the current γ)

3. Calculate (numerically) the $\mathcal{G}(\mathbb{E}s_t)$ vector and the $\mathcal{H}(\mathbb{E}s_t)$ matrix at the $\mathbb{E}s_t$ value associated with each grid point (i.e., the gradient and hessian evaluated at the expected state vector value next period given the state vector value on the grid)[IA.29]

4. Evaluate the right hand side of Equation IA.146 at each dataset grid point and call the resulting values $vw^n(s_t)$

5. Repeat steps 2 to 4 until $vw^n(s_t)$ converges for all points in the dataset

6. Given a converged solution, $vw(s_t)$, compute $N_{\mathbb{V}}$ at each date in our sample according to Equation IA.38[IA.30]

The algorithm described above solves for the value function given a pre-specified γ. In order to identify the optimal γ through our GMM procedure given the nonlinear $N_{\mathbb{V}}$, we have an outter algorithm:

1. Solve for vw_t over a grid of γ values (we begin with an initial grid that ranges from 1 to 10)

2. Evaluate the GMM objective given $N_{\mathbb{V}}$ associated with each of these γ grid points and identify the grid point with the lowest objective

[IA.29] To numerically estimate the $\mathcal{G}(\mathbb{E}s_t)$ vector and the $\mathcal{H}(\mathbb{E}s_t)$ matrix at each iteration given $vw^{n-1}(s_t)$, we use linear interpolation to estimate $vw^{n-1}(s_t)$ off the grid points at points necessary for the numerical derivative approximations. In all cases, we use standard expressions for the numerical derivatives that are second-order accurate in the step size used to estimate the derivatives. We use step sizes equal to 1/4 of the grid spacings in each state variable dimension. Since we estimate $vw^{n-1}(s_t)$ off the grid points using linear interpolation for tractability and stability, we cannot use very small step sizes to estimate numerical derivatives since this would lead to second derivative estimates equal to zero. We find this choice to be innocuous, again because the value function solution has relatively low curvature on the state space grid.

[IA.30] To compute $\widetilde{vw}_t$ at each date in our sample, we use Equation IA.143 to compute $\mathbb{E}_{t-1}[vw(s_t)]$. We then approximate the realized vw_t via linear interpolation given our converged solution on the grid, $vw(s_t)$, and obtain $\widetilde{vw}_t = vw_t - \mathbb{E}_{t-1}[vw(s_t)]$. Based on Equation IA.38, we then have $N_{\mathbb{V}} = 2 \cdot (N_{\mathbb{E},t} - \widetilde{vw}_t)/(\gamma - 1)$, with $N_{\mathbb{E}}$ based on Equation IA.91, which is an exact solution that is independent of the $N_{\mathbb{V}}$ solution method.

3. Create a new γ grid that spans the next lowest and next highest γ values on the current grid relative to the value that produced the lowest objective

4. Repeat steps 1-3 until the γ grid converges[IA.31]

Our final $N_{\mathbb{V}}$ time series used in Table IA.6 is that associated with this optimal γ. Note that we implement this procedure separately in our Long and Modern Samples.

D.8 ICAPM Risk Prices Controlling for Sentiment and Liquidity

In this subsection, we explore whether the ICAPM risk prices are robust to controlling for two other factors: sentiment and liquidity.

These factors are outside the models we investigate in the main paper, but have been shown to explain some aspects of the cross section of average returns. The sentiment factor we use is a tradable version of the sentiment index in Baker and Wurgler (2006), which we construct by creating a mimicking portfolio for changes to sentiment using the same methodology as we do to construct our intertemporal risk factors, $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$. We obtain the sentiment data from Jeffrey Wurgler's website. Note that these data are only available until December 2018, so our analysis using sentiment ends one year before our standard ending period of December 2019. For liquidity, we use the tradable liquidity factor in Pástor and Stambaugh (2003), and we obtain the relevant data from CRSP.

Table IA.7 provides results analogous to those in Table 6 when controlling for the sentiment and liquidity factors. Panel A provides results over our Long Sample (1928-2019) and Panel B provides results over our Modern Sample (1973-2019). In all cases, our main results are qualitatively unchanged. The market and reinvestment factors have positive and statistically significant risk prices. The volatility factor has a negative and statistically significant price of risk. The risk prices remain quantitatively similar to their original values as well, except for the case of the volatility risk price when controlling for the liquidity factor. In this case, the volatility risk price magnitude decreases slightly, which is reasonable since volatility and

[IA.31] In practice, we iterate until the final γ grid range is lower than 0.0001.

liquidity tend to be correlated. However, the volatility risk price remains large in magnitude and statistically significant.

D.9 ICAPM Risk Prices Controlling for the Betting Against Beta Factor

In this subsection, we further explore whether the ICAPM risk prices are robust to controlling for the betting against beta (BaB) factor of Frazzini and Pedersen (2014).

We start by replicating their BaB factor following their exact methodology.[IA.32] The risk prices related to this replication are provided in Table IA.8 under the header "Original BaB". As the table shows, the ICAPM risk prices are strong and statistically significant after controlling for the BaB factor. Interestingly, the only exception to this result is that the market price of risk becomes much weaker and only marginally significant over our Long Sample after all four factors are included in the SDF. However, as we detail below, this result is an artifact of the fact that the BaB factor has an embedded long position on the market portfolio to hedge the negative market beta of the rest of the portfolio.

As Novy-Marx and Velikov (2022) demonstrate, the returns on the BaB factor can be written as:[IA.33]

$$\begin{aligned} r_{BaB,t+1}^{Original} &= \frac{R_{L,t+1} - R_{f,t+1}}{\beta_{m,L}} - \frac{R_{H,t+1} - R_{f,t+1}}{\beta_{m,H}} \\ &= r_{BaB,t+1}^{noHedge} + r_{BaB,t+1}^{Hedge} \end{aligned} \tag{IA.152}$$

where

$r_{BaB,t+1}^{noHedge} = R_{L,t+1} - R_{H,t+1}$

$r_{BaB,t+1}^{Hedge} = \left((\beta_{m,L}^{-1} - 1) \cdot R_{L,t+1} + (1 - \beta_{m,H}^{-1}) \cdot R_{H,t+1}\right) - (\beta_{m,L}^{-1} - \beta_{m,H}^{-1}) \cdot R_{f,t+1}$

The factor $r_{BaB}^{noHedge}$ is a BaB factor with no hedge. That is, this factor is negatively exposed to market risk since it longs low beta stocks and shorts high beta stocks. The factor r_{BaB}^{Hedge}

[IA.32] Over the period we have data from the original paper (04/1929 to 03/2012), our replicating BaB factor has a correlation of 0.99 with the original BaB factor. Moreover, over the same period, our replicating BaB factor has an average return of 0.73% with a volatility of 3.34%, which are close to the average return (0.70%) and volatility (3.10%) of the original BaB factor.

[IA.33] We set $\beta_{m,L,t} = \beta_{m,L}$ and $\beta_{m,H,t} = \beta_{m,H}$ to simplify the exposition, but we account for the time-varying betas when constructing the BaB factor (and the arguments we make remain valid with time-varying betas).

is a hedging portfolio that buys equities financed by borrowing at the risk-free rate. The r_{BaB}^{Hedge} market beta is positive with the same magnitude as the (negative) market beta of the $r_{BaB}^{noHedge}$ factor. As such, $r_{BaB}^{Original} = r_{BaB}^{noHedge} + r_{BaB}^{Hedge}$ has zero market beta (i.e., it is hedged against market risk).

As Novy-Marx and Velikov (2022) point out, implementing r_{BaB}^{Hedge} is non-standard and very costly. In fact, it would be much more natural to hedge against the negative market risk of $r_{BaB}^{noHedge}$ by buying the market portfolio itself. In this case, the BaB factor would be

$$r_{BaB,t+1}^{MktHedged} = r_{BaB,t+1}^{noHedge} + (\beta_{m,H} - \beta_{m,L}) \cdot r_{m,t+1} \tag{IA.153}$$

with $r_{BaB}^{MktHedged}$ also having zero market beta. The risk prices that use $r_{BaB}^{MktHedged}$ as the BaB factor are provided in Table IA.8 under the header "Mkt Hedged BaB". The ICAPM risk prices remain strong and statistically significant. Perhaps more importantly, once the hedge is based on the market factor, the BaB factor is the weakest from a statistical standpoint (but it remains statistically significant).

Equation IA.153 also clarifies why the market risk price seems relatively weak when the BaB factor is included in the SDF. Specifically, we have

$$\begin{aligned} M_{t+1} &= a + b_m \cdot r_{m,t+1} + b_{BaB} \cdot r_{BaB,t+1}^{MktHedged} \\ &= a + b_m \cdot r_{m,t+1} + b_{BaB} \cdot r_{BaB,t+1}^{noHedge} + b_{BaB} \cdot (\beta_{m,H} - \beta_{m,L}) \cdot r_{m,t+1} \\ &= a + [b_m + b_{BaB} \cdot (\beta_{m,H} - \beta_{m,L})] \cdot r_{m,t+1} + b_{BaB} \cdot r_{BaB,t+1}^{noHedge} \end{aligned} \tag{IA.154}$$

and analogously in the SDF specification that controls for the intertemporal risk factors. As such, the market hedge embedded in $r_{BaB}^{MktHedged}$ incorporates a portion of the market risk price, and thus artificially decreases the estimated b_m. To address this issue, we further provide risk prices for a specification that uses $r_{BaB}^{noHedge}$ as the BaB factor in Table IA.8 under the header "Not Hedged BaB". As it is clear from the table, the BaB factor remains the weakest from a statistical standpoint (but it also remains statistically significant over our Modern Sample). However, the market risk price controlling for the BaB factor is now at least as strong as (and typically stronger than) the market risk price in the specifications that do not control for the BaB factor.

In summary, the ICAPM risk prices remain strong and statistically significant after controlling for the BaB factor. This result is true whether we use the original BaB factor or alternative versions that either rely on the market portfolio for hedging purpose or that do not hedge at all. In contrast, the risk price of the original BaB factor is extremely strong, but its statistical power weakens substantially if we use the market portfolio for hedging purpose or do not hedge at all when constructing the BaB factor.

D.10 Relation Between our Risk Prices and the Prior ICAPM Literature

Some prior papers estimate the risk prices of shocks to market expected returns and volatility in the context of an ICAPM without controlling for each other. This subsection relates our findings to this literature.

(a) $r_{\mathbb{E}}$ Risk Price

In terms of the $r_{\mathbb{E}}$ risk price, the ICAPM$_{\mathbb{E}}$ column in Table 5 shows that $b_{\mathbb{E}}$ is much smaller when we do not control for $r_{\mathbb{V}}$, but it is still strong and statistically significant. This result is consistent with several papers in the ICAPM literature that estimate the risk price of news to long-term expected returns through auxiliary econometric models (e.g., Campbell and Vuolteenaho (2004), Brennan, Wang, and Xia (2004), Maio (2013), Cederburg (2019), and Gonçalves (2021b)). However, one paper in the ICAPM literature (Kozak and Santosh (2020)) finds that the risk price of shocks to expected returns is actually negative. This paper is important because it does not require an auxiliary econometric model to estimate expected returns. In particular, Kozak and Santosh (2020) show that the covariance of an asset, $r_{j,t}$, relative to market expected return shocks can be measured using the risk factor $\Sigma_{h=1}^{H}\kappa^{h-1} \cdot r_{m,t+h}$ since future returns are composed of expected returns plus an error term that is orthogonal to $r_{j,t}$.

To understand this apparent contradiction to our results, we follow our $r_{\mathbb{E}}$ construction to create a tradable factor based on the Kozak and Santosh (2020) measure of expected returns.

We call this tradable factor $r_{\mathbb{E}}^{ST}$ and report SDF regressions that include this factor in Table IA.9. Consistent with Kozak and Santosh (2020), the $r_{\mathbb{E}}^{ST}$ factor has a negative risk price (although the noise makes it statistically weak in the Modern Sample), but controlling for it has basically no effect on the risk prices of our ICAPM factors (and vice versa). Table IA.10 shows that the reason is that the factor in Kozak and Santosh (2020) captures relatively short-term expected returns. Specifically, market returns display long-term reversal, which implies that $\Sigma_{h=1}^{H} \kappa^{h-1} \cdot r_{m,t+h}$ is negatively correlated with subsequent long-term returns. As such, $\Sigma_{h=1}^{H} \kappa^{h-1} \cdot r_{m,t+h}$ is a good measure for short-term expected returns, but not for long-term expected returns. From the perspective of Kozak and Santosh (2020), this aspect is not a problem since their goal is to estimate the risk price on expected return shocks. However, our goal is to construct a tradable factor that captures news to long-term expected returns as implied by our ICAPM, and thus the Kozak and Santosh (2020) measure is not appropriate for our purpose. In any case, the results in Tables IA.9 and IA.10 demonstrate that there is no contradiction between our results and the ones in Kozak and Santosh (2020).

(b) $r_{\mathbb{V}}$ Risk Price

In terms of the $r_{\mathbb{V}}$ risk price, the ICAPM$_{\mathbb{V}}$ column in Table 5 shows that $b_{\mathbb{V}}$ is close to zero (and statistically insignificant) when we control for r_m but not for $r_{\mathbb{E}}$. This result seems to contradict the findings in Ang et al. (2006) and Adrian and Rosenberg (2008) but, as we explain below, it does not.

First, Ang et al. (2006) find a negative volatility risk premium that is statistically significant (although not that large economically) in a model that includes the market portfolios. To understand their result, it is important to note that their volatility exposure is based on multivariate betas that control for the market factor, not univariate betas or covariances. This aspect implies that the risk price they report is proportional to the univariate projection coefficient of the SDF onto the volatility factor, not the multivariate projection coefficient of the SDF onto the volatility factor controlling for other factors (see Chapter 13.4 in Cochrane (2005)). In other words, their result is comparable to our $\mathbb{E}[r_{\mathbb{V}}]$ and not our $b_{\mathbb{V}}$. As such, our results are in line with Ang et al. (2006) as our $\mathbb{E}[r_{\mathbb{V}}]$ is negative (although statistically weak in the Modern Sample).

Second, Adrian and Rosenberg (2008) estimate a volatility model that has short- and long-run (nontradable) volatility components and find that volatility risk is strongly priced controlling for market returns. While their risk prices are estimated using covariances as risk exposures, and thus reflect bs in our terminology, they effectively control for expected returns indirectly. Specifically, when estimating their volatility model, they impose a general equilibrium condition that the market expected return is linear in their two volatility components. As such, our results are consistent with theirs since $b_{\mathbb{V}}$ is strong and statistically significant after we control for $r_{\mathbb{E}}$ in our SDF projections. As explained in Subsection 2.5, we treat our ICAPM as a partial equilibrium model so that we do not impose the equilibrium condition that the market risk premium is linear in our volatility factor because we would like to accommodate the possibility that variation in expected returns are also driven by other channels such as sentiment (as in Kozak, Nagel, and Santosh (2018)) and risk aversion (as in Gonçalves (2021a)) without explicitly modeling such effects.

References for Internet Appendix

Adrian, T. and J. Rosenberg (2008). "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk". In: *Journal of Finance* 63.6, pp. 2997–3030.

Allen, F. and R. Michaely (2003). "Payout Policy". In: *Handbook of the Economics of Finance.* Ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. Vol. 1. A. Elsevier Science. Chap. 7, pp. 337–429.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001a). "The distribution of realized stock return volatility". In: *Journal of Financial Economics* 61, pp. 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b). "The Distribution of Realized Exchange Rate Volatility". In: *American Statistical Association* 96.453, pp. 42–55.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Ed. by T. Kailath. Prentice-Hall, Inc.

Anderson, E. W., E. Ghysels, and J. L. Juergens (2009). "The impact of risk and uncertainty on expected returns". In: *Journal of Financial Economics* 94, pp. 233–263.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). "The Cross-Section of Volatility and Expected Returns". In: *Journal of Finance* 61.1, pp. 259–299.

Ang, A., J. Liu, and K. Schwarz (2020). "Using Stocks or Portfolios in Tests of Factor Models". In: *Journal of Financial and Quantitative Analysis* 55.3, pp. 709–750.

Baker, M. and J. Wurgler (2006). "Investor Sentiment and the Cross-Section of Stock Returns". In: *Journal of Finance* 61.4.

Barndorff-Nielsen, O. E. and N. Shephard (2002). "Econometric Analysis of Realized Volatility and Its Use in Estimating StochasticVolatility Models". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64.2, pp. 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2005). "How accurate is the asymptotic approximation to the distribution of realized variance?" In: *Identification and inference for econo-*

metric models. A Festschrift in honour of TJ Rothenberg. Ed. by D. W. K. Andrews and J. H. Stock. Cambridge University Press.

Benveniste, L. and J. Scheinkman (1979). "On the Differentiability of the Value Function in Dynamic Models of Economics". In: *Econometrica* 47.3, pp. 727–732.

Berk, J. B. (2000). "Sorting out sorts". In: *Journal of Finance* 55.1, pp. 407–427.

Binsbergen, J. H. v. and R. S. J. Koijen (2010). "Predictive Regressions: A Present-Value Approach". In: *Journal of Finance* 65.4, pp. 1439–1471.

Brennan, M. J., A. W. Wang, and Y. Xia (2004). "Estimation and Test of a Simple Model of Intertemporal Capital Asset Pricing". In: *Journal of Finance* 59.4, pp. 1743–1775.

Campbell, J. Y. (1987). "Stock Returns and the Term Structure". In: *Journal of Financial Economics* 18, pp. 373–399.

Campbell, J. Y. (1991). "A Variance Decomposition for Stock Returns". In: *The Economic Journal* 101.405, pp. 157–179.

Campbell, J. Y. (1993). "Intertemporal Asset Pricing without Comsumption Data". In: *American Economic Review* 83.3, pp. 487–512.

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley (2018). "An Intertemporal CAPM with Stochastic Volatility". In: *Journal of Financial Economics* 128.2, pp. 207–233.

Campbell, J. Y. and R. J. Shiller (1989). "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors". In: *Review of Financial Studies* 1.3, pp. 195–228.

Campbell, J. Y. and T. Vuolteenaho (2004). "Bad Beta, Good Beta". In: *American Economic Review* 94.5, pp. 1249–1275.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance". In: *Journal of Finance* 52.1, pp. 57–82.

Cederburg, S. (2019). "Pricing intertemporal risk when investment opportunities are unobservable". In: *Journal of Financial and Quantitative Analysis* 54.4, pp. 1759–1789.

Cederburg, S. and M. S. O'Doherty (2015). "Asset-pricing anomalies at the firm level". In: *Journal of Econometrics* 186, pp. 113–128.

Chacko, G. and L. M. Viceira (2005). "Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets". In: *Review of Financial Studies* 18.4, pp. 1369–1402.

Chen, A. Y. and T. Zimmermann (2020). "Open Source Cross-Sectional Asset Pricing". Working Paper.

Cochrane, J. H. (2005). *Asset Pricing*. Revised Edition. Princeton University Press.

Dalla, V., L. Giraitis, and P. C. B. Phillips (2020). "Robust Test for White Noise and Cross-Correlation". In: *Econometric Theory*, pp. 1–29.

Daniel, K., D. Hirshleifer, and L. Sun (2020). "Short- and Long-Horizon Behavioral Factors". In: *Review of Financial Studies* 4, pp. 1673–1736.

Davis, J. L., E. F. Fama, and K. R. French (2000). "Characteristics, Covariances, and Average Returns: 1929-1997". In: *Journal of Finance* 55.1, pp. 389–406.

Detzel, A., R. Novy-Marx, and M. Velikov (2020). "Model Selection with Transaction Costs". Working Paper.

Driscoll, J. C. and A. C. Kraay (1998). "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data". In: *Review of Economics and Statistics* 80.4, pp. 549–560.

Ederington, L. H. and W. Guan (2010). "Longer-Term Time-Series Volatility Forecasts". In: *Journal of Financial and Quantitative Analysis* 45.4, pp. 1055–1076.

Engsted, T., T. Q. Pedersen, and C. Tanggaard (2012). "Pitfalls in VAR based return decomposition: A clarification". In: *Journal of Banking and Finance* 36.5, pp. 1255–1265.

Epstein, L. G. and S. E. Zin (1989). "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework". In: *Econometrica* 57.4, pp. 937–969.

Epstein, L. G. and S. E. Zin (1991). "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis". In: *Journal of Political Economy* 99.2, pp. 263–286.

Fama, E. F. (1981). "Stock Returns, Real Activity, Inflation, and Money". In: *American Economic Review* 71.4, pp. 545–565.

Fama, E. F. (1991). "Efficient Capital Markets: II". In: *Journal of Finance* 46.5, pp. 1575–1617.

Fama, E. F. and K. R. French (1989). "Business conditions and expected returns on stocks and bonds". In: *Journal of Financial Economics* 25.1, pp. 23–49.

Fama, E. F. and K. R. French (1993). "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics* 33, pp. 3–56.

Fama, E. F. and K. R. French (2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116, pp. 1–22.

Fama, E. F. and K. R. French (2018). "Choosing Factors". In: *Journal of Financial Economics* 128, pp. 234–252.

Fama, E. F. and G. W. Schwert (1977). "Asset Returns and Inflation". In: *Journal of Financial Economics* 5, pp. 115–146.

Fleming, J. and C. Kirby (2003). "A Closer Look at the Relation Between GARCH and Stochastic Autoregressive Volatility". In: *Journal of Financial Econometrics* 1.3, pp. 365–419.

Foster, D. P. and D. B. Nelson (1996). "Continuous record asymptotics for rollingsample variance estimators". In: *Econometrica* 64.1, pp. 139–174.

Frazzini, A. and L. H. Pedersen (2014). "Betting Against Beta". In: *Journal of Financial Economics* 111.1, pp. 1–25.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). "A Test of the Efficiency of a Given Portfolio". In: *Econometrica* 57.5, pp. 1121–1152.

Giglio, S. and D. Xiu (2020). "Asset pricing with omitted factors". In: *Journal of Political Economy* forthcoming.

Gonçalves, A. S. (2021a). "Reinvestment Risk and the Equity Term Structure". In: *Journal of Finance* 76.5, pp. 2153–2197.

Gonçalves, A. S. (2021b). "The Short Duration Premium". In: *Journal of Financial Economics* 141.3, pp. 919–945.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". In: *Econometrica* 50.4, pp. 1029–1054.

Hansen, L. P., J. C. Heaton, and N. Li (2008). "Consumption Strikes Back? Measuring Long-Run Risk". In: *Journal of Political Economy* 116.2, pp. 260–302.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). "Realized GARCH: A joint Model for Returns and Realized Measures of Volatility". In: *Journal of Applied Econometrics* 27, pp. 877–906.

Harvey, C. R. (2017). "Presidential Address: The Scientific Outlook in Financial Economics". In: *Journal of Finance* 72.4, pp. 1399–1440.

Herskovic, B., A. Moreira, and T. Muir (2019). "Hedging Risk Factors". Working Paper.

Hou, K., H. Mo, C. Xue, and L. Zhang (2020). "An Augmented q-Factor Model with Expected Growth". In: *Review of Finance* Forthcoming.

Hou, K., C. Xue, and L. Zhang (2015). "Digesting Anomalies: An Investment Approach". In: *Review of Financial Studies* 28.3, pp. 650–705.

Jordà, Ò. and S. Kozicki (2011). "Estimation and Inference by the Method of Projection Minimum Distance: An Application to the New Keynesian Hybrid Phillips Curve". In: *International Economic Review* 52.2, pp. 461–487.

Kan, R., X. Wang, and X. Zheng (2019). "In-sample and Out-of-sample Sharpe Ratios of Multi-factor Asset Pricing Models". Working Paper.

Koijen, R. S. J. and S. V. Nieuwerburgh (2011). "Predictability of Returns and Cash Flows". In: *Annual Review of Financial Economics* 3, pp. 467–491.

Kozak, S., S. Nagel, and S. Santosh (2018). "Interpreting Factor Models". In: *Journal of Finance* 73.3, pp. 1183–1223.

Kozak, S. and S. Santosh (2020). "Why do discount rates vary?" In: *Journal of Financial Economics* 137.3, pp. 740–751.

Lo, A. W. and A. C. MacKinlay (1990). “Data-Snooping Biases in Tests of Financial Asset Pricing Models”. In: *Review of Financial Studies* 3.3, pp. 431–467.

Ludvigson, S. C. (2013). “Advances in Consumption-Based Asset Pricing: Empirical Tests”. In: *Handbook of the Economics of Finance*. Ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. 1st ed. Vol. 2. B. Elsevier Science. Chap. 12, pp. 799–906.

Maio, P. (2013). “Intertemporal CAPM with Conditioning Variables”. In: *Management Science* 59.1, pp. 122–141.

Merton, R. C. (1980). “On Estimating the Expected Return on the Market: An Exploratory Investigation”. In: *Journal of Financial Economics* 8, pp. 323–361.

Nagel, S. (2013). “Empirical cross-sectional asset pricing”. In: *Annual Review of Financial Economics* 5, pp. 167–199.

Newey, W. K. and K. D. West (1987). “A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”. In: *Econometrica* 55.3, pp. 703–708.

Newey, W. K. and K. D. West (1994). “Automatic Lag Selection in Covariance Matrix Estimation”. In: *Review of Economic Studies* 61.4, pp. 631–653.

Novy-Marx, R. and M. Velikov (2022). “Betting against betting against beta”. In: *Journal of Financial Economics* 143.1, pp. 80–106.

Pástor, L. and R. F. Stambaugh (2003). “Liquidity Risk and Expected Stock Returns”. In: *Journal of Political Economy* 111.3, pp. 642–685.

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. Second Edition. Wiley-Interscience.

Sabbatucci, R. (2015). “Are Dividends and Stock Returns Predictable? New Evidence Using M&A Cash Flows”. Working Paper.

Shanken, J. (1992). “On the Estimation of Beta-Pricing Models”. In: *Review of Financial Studies* 5.1, pp. 1–33.

Shumway, T. (1997). “The Delisting Bias in CRSP Data”. In: *Journal of Finance* 52.1, pp. 327–340.

Stambaugh, R. F. (1997). “Analyzing investments whose histories differ in length”. In: *Journal of Financial Economics* 45, pp. 285–331.

Stambaugh, R. F. and Y. Yuan (2017). “Mispricing Factors”. In: *Review of Financial Studies* 4, pp. 1270–1315.

Tian, M. (2021). “Firm Characteristics and Empirical Factor Models: A Model Mining Experiment”. In: *Review of Financial Studies*. Forthcoming.

Weil, P. (1989). “The Equity Premium Puzzle and the Risk-Free Rate Puzzle”. In: *Journal of Monetary Economics* 24.3, pp. 401–421.

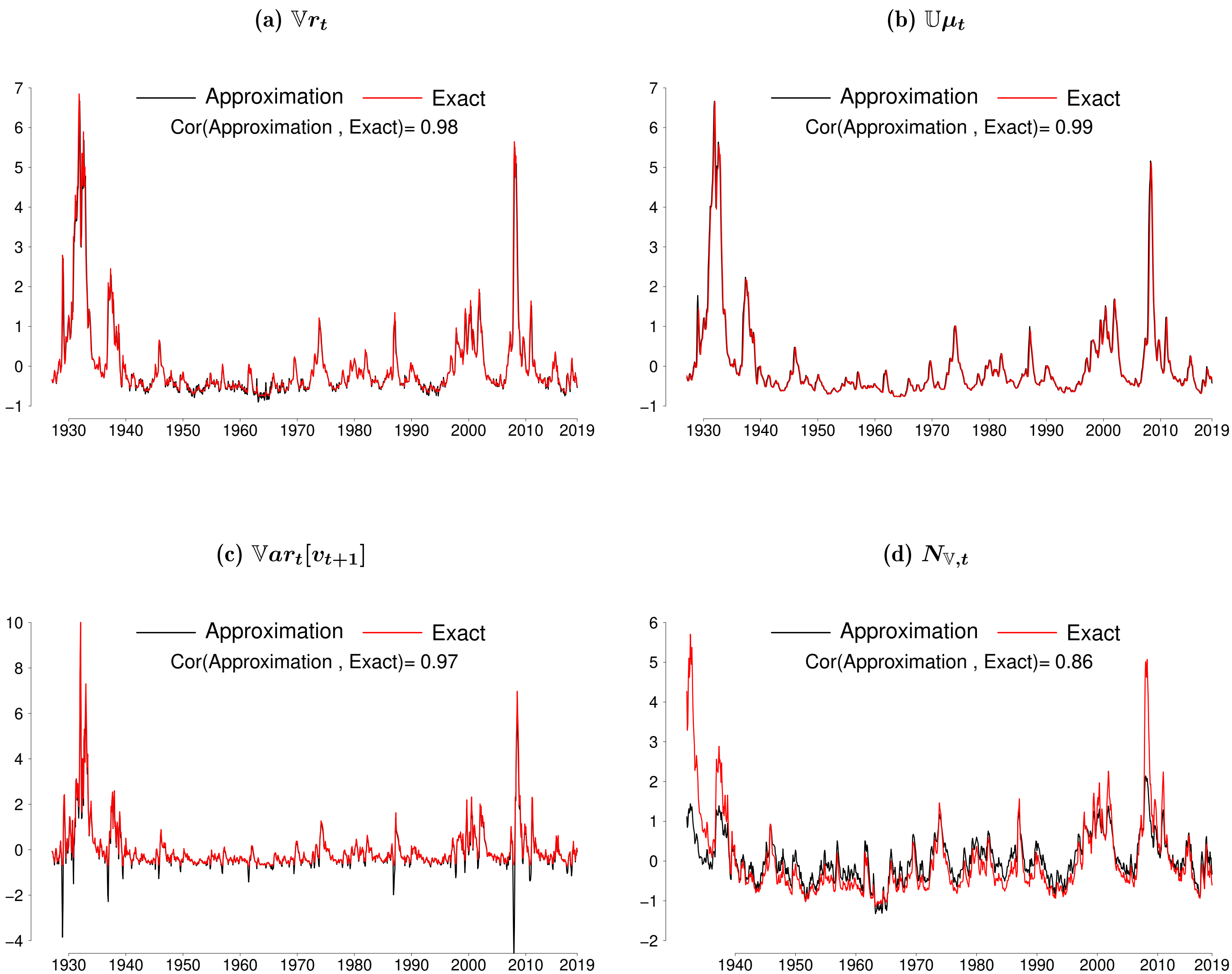

Figure IA.1

Accuracy of Approximations Used to Derive $N_{\mathbb{V},t} \approx \theta_{\mathbb{V}} \cdot \widetilde{\sigma}_t^2$

Panels (a) to (c) show the approximations for (and exact values of) $\mathbb{V}r_t$ (Equation IA.109), $\mathbb{U}\mu_t$ (Equation IA.111), and $\mathbb{V}ar_t[v]$ (Equation IA.110). These approximations are used to derive an approximate expression for $N_{\mathbb{V},t} = (\mathbb{E}_{t+1} - \mathbb{E}_t)[\sum_{h=1}^{\infty} \overline{\delta}^h \cdot \mathbb{V}ar_{t+h}[v]]$ (Equation IA.103), which is displayed in panel (d) together with its numerical solution (solved globally as described in Subsection D.7), which we still label “exact”. To help visualize the correlation in panel (d), we use a filter analogous to the one in Figure 2, except that we do not divide the shocks by their respective standard deviations so that differences in standard deviation between the exact and approximate series are displayed (note that the correlation number displayed is still based on the original news). To normalize units, all panels report the z-scores for each variable but with means and standard deviations calculated always based on the exact series so that the figures display any difference in means and standard deviations between the exact and approximate solutions. All results are reported using our benchmark estimation for $\mathbb{E}r_t$ and $\mathbb{V}r_t$ in TableIA.1 over our Long Sample (1928-2019) and, for panels (c) and (d), the γ value obtained in our structural estimation in Table 5 over our Long Sample.

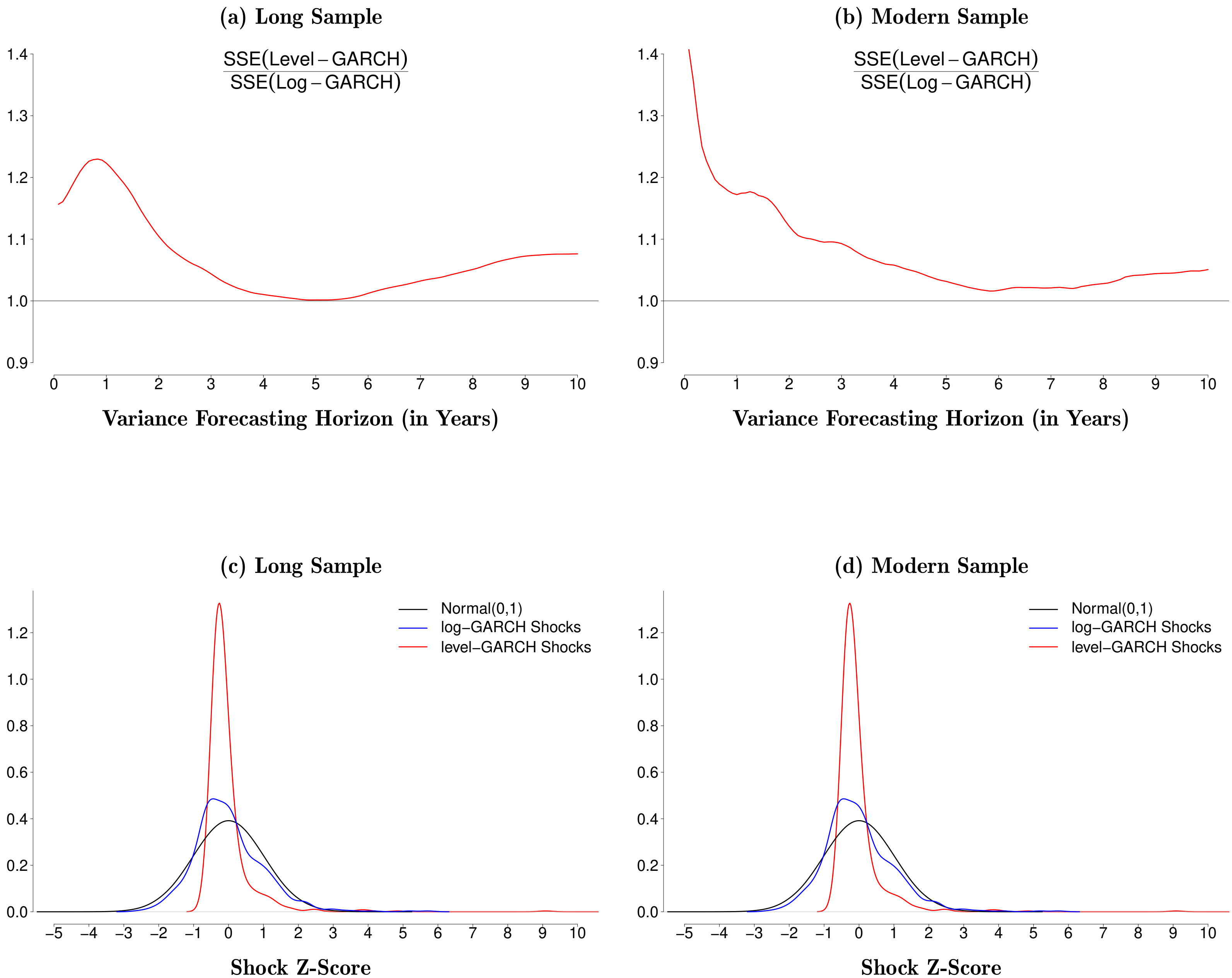

Figure IA.2
Log-GARCH vs Level-GARCH

Panels (a) and (b) provide the forecasting sum of squared errors (SSE) of a level-GARCH relative to our baseline log-GARCH model. We provide the relative SSE values for different forecasting horizons going from one month to one hundred and twenty months (ten years). Panels (c) and (d) provide empirical density estimates for the shocks of the same log-GARCH and level-GARCH models. Subsection D.4 provides further details.

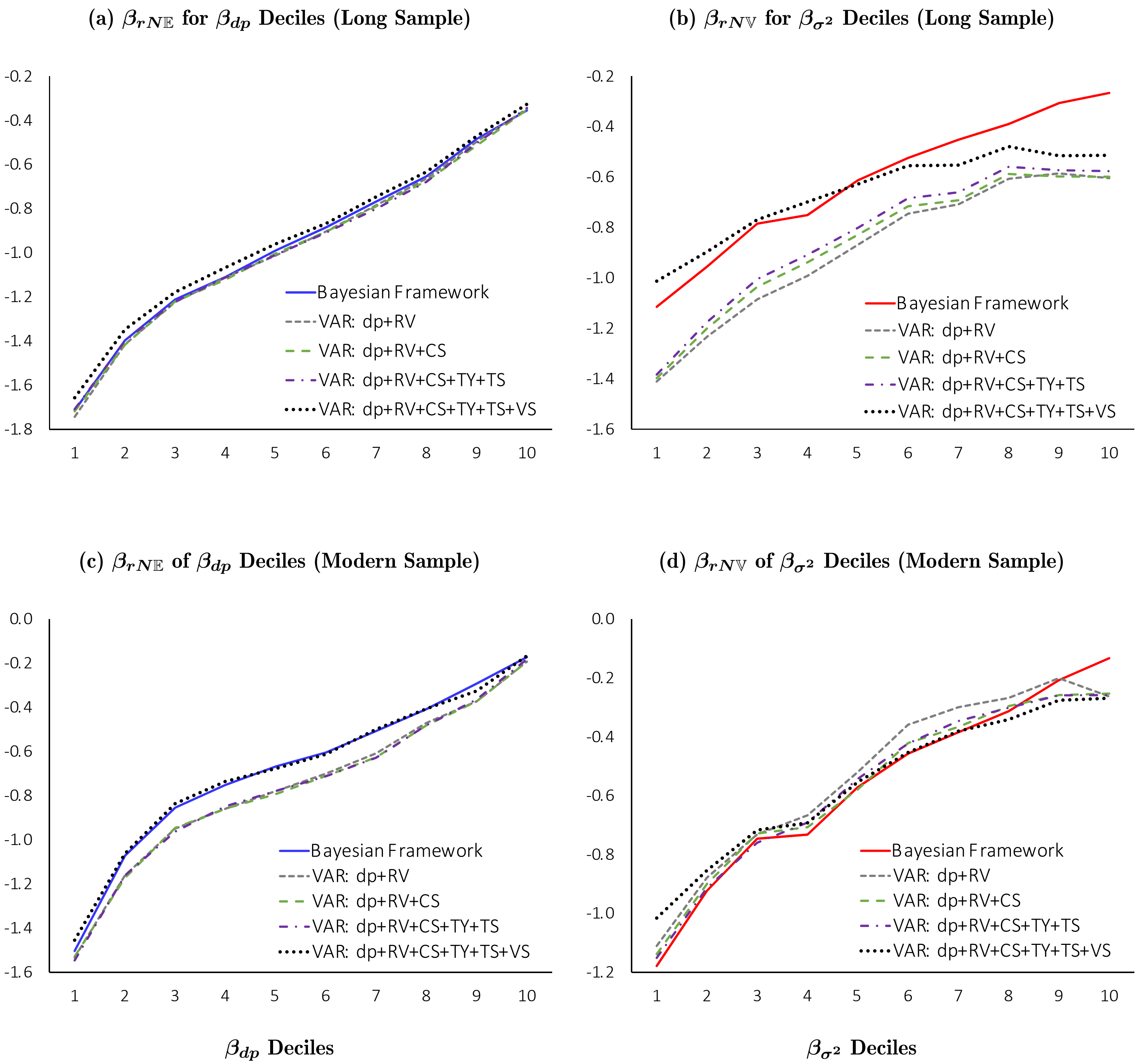

Figure IA.3

$\beta_{rN\mathbb{E}}$ and $\beta_{rN\mathbb{V}}$ of β_{dp} and β_{σ^2} Decile Portfolios: $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ from a VAR System

These graphs show ex-post decile portfolio betas (in market beta units as explained in Footnote 26) with respect to the $N_{\mathbb{E}}$ and $N_{\mathbb{V}}$ mimicking factors ($r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$) when news are computed under our baseline Bayesian framework in Subsection 1.3.2 or under different vector autoregressive (VAR) specifications. In particular, we estimate news using four alternative VAR specifications that include different combinations of the following state variables: dividend yield (dp), realized variance (RV), credit spread (CS), Treasury yield (TY), term spread (TS), and value spread (VS). Panels (a) and (b) use data from our Long Sample (1928-2019) whereas Panels (c) and (d) use data from our Modern Sample (1973-2019). Panels (a) and (c) use portfolios sorted on ex-ante exposures to Δdp (β_{dp}) whereas Panels (b) and (d) use portfolios sorted on ex-ante exposures to $\Delta\sigma^2$ (β_{σ^2}). See Subsection D.5 for more details.

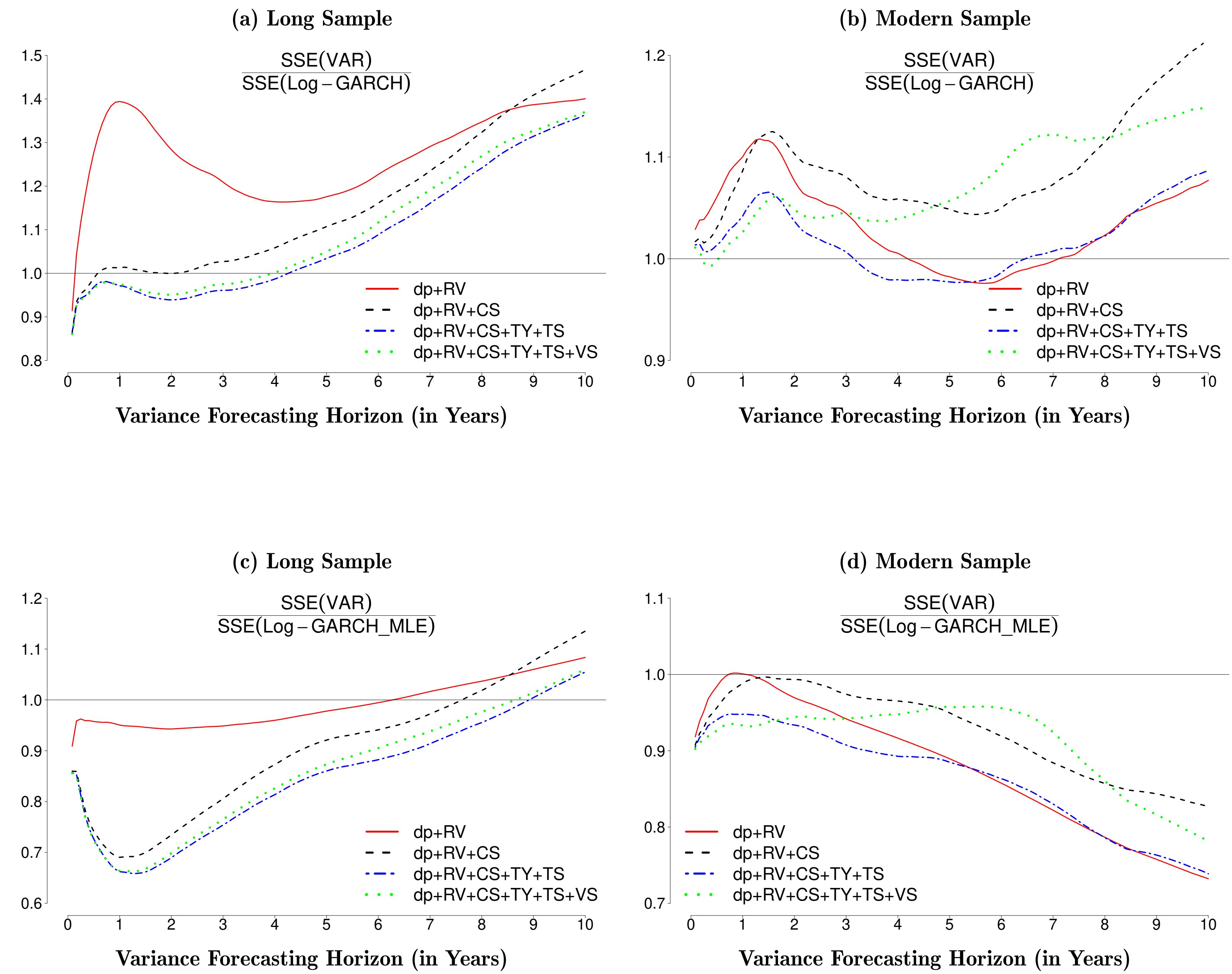

Figure IA.4

VAR Forecasting Sum Squared Errors

Panels (a) and (b) provide the forecasting sum of squared errors (SSE) of different VAR specifications relative to our baseline log-GARCH model, which is estimated by targeting long-run variance (see Subsection B.1.1), while Panels (c) and (d) provide analogous results when the log-GARCH model is estimated by maximum likelihood. We consider four alternative VAR specifications that include different combinations of the following state variables: dividend yield (dp), realized variance (RV), credit spread (CS), Treasury yield (TY), term spread (TS), and value spread (VS). Subsections D.5 and D.6 provide further details.

Table IA.1
$\mathbb{E}r_t$ and $\mathbb{V}r_t$ Parameters

Panel A reports the parameter estimates related to the $\mathbb{E}r_t$ process while Panel B reports the parameter estimates related to the $\mathbb{V}r_t$ process. The estimation procedure is described in Subsection B.1. Parameter estimates are reported in their original units with two exceptions. First, some parameters are annualized for interpretability, with their label providing the explicit annualization (e.g., $12 \cdot \mu$ is the annualized mean r_w). Second, the units of parameters related to covariances ($\nu_{r,\mu}$, $\nu_{r,g}$, $\nu_{\mu,g}$, $\sigma_{r,\sigma}$, $\sigma_{dp,\sigma}$) is hard to interpret, and thus we instead report correlation values. Specifically, we report $Cor_{r,\mu} = \nu_{r,\mu}/\nu_\mu$, $Cor_{r,g} = \nu_{r,g}/\nu_g$, $Cor_{\mu,g} = \nu_{\mu,g}/(\nu_\mu \cdot \nu_g)$, $Cor_{r,\sigma} = \sigma_{r,\sigma}/(\sqrt{\mathbb{V}r} \cdot \sigma_\sigma)$, and $Cor_{dp,\sigma} = \sigma_{dp,\sigma}/(\sqrt{\mathbb{V}r} \cdot \nu_{dp} \cdot \sigma_\sigma)$, where $\mathbb{V}r = e^{\omega_{\mathbb{V}}/(1-\phi_{\mathbb{V}})}$.

PANEL A: $\mathbb{E}r_t$ Process

	Long Sample (1928-2019)	Modern Sample (1973-2019)
ρ	0.964	0.968
$12 \cdot \mu$	0.064	0.062
ϕ_μ^{12}	0.9473	0.967
ϕ_g^{12}	0.182	0.237
ν_μ	0.011	0.012
ν_g	0.209	0.272
$Cor_{r,\mu}$	-0.635	-0.518
$Cor_{r,g}$	0.050	-0.054
$Cor_{\mu,g}$	0.736	0.859

PANEL B: $\mathbb{V}r_t$ Process

	Long Sample (1928-2019)	Modern Sample (1973-2019)
$\sqrt{\mathbb{V}r}$	0.045	0.040
$\omega_{\mathbb{V}}$	-0.160	-0.597
$\omega_{\mathbb{V}r}$	-0.026	-0.502
ω_σ	-0.520	-0.169
$\phi_{\mathbb{V}}$	0.974	0.908
$\phi_{\mathbb{V}r}$	0.716	0.343
ϕ_σ	0.258	0.564
σ_σ	0.692	0.646
$Cor_{r,\sigma}$	-0.455	-0.456
$Cor_{dp,\sigma}$	0.430	0.396

Table IA.2
Correlations: News Proxies, Tradable Factors, and Ex-post Mimicking Factors (Adding Factors based on PCA)

This table reports correlations between our tradable risk factors ($r_{\mathbb{E}}$, $r_{\mathbb{V}}$, $r_{\mathbb{E}}^{PCA}$, and $r_{\mathbb{V}}^{PCA}$) and ex-post news ($N_{\mathbb{E}}$ and $N_{\mathbb{V}}$) as well as ex-post news mimicking factors ($r_{N\mathbb{E}}$, $r_{N\mathbb{V}}$, $r_{N\mathbb{E}}^{PCA}$, $r_{N\mathbb{V}}^{PCA}$). To construct $r_{\mathbb{E}}$ ($r_{\mathbb{V}}$), we buy value-weighted portfolios of the stocks with the 30% highest exposures to Δdp ($\Delta\sigma^2$) and sell a value-weighted portfolio of the stocks with the 30% lowest exposures to Δdp ($\Delta\sigma^2$). Similarly, to construct $r_{\mathbb{E}}^{PCA}$ ($r_{\mathbb{V}}^{PCA}$), we buy value-weighted portfolios of the stocks with the 30% highest weights on the Δdp ($\Delta\sigma^2$) mimicking portfolio and sell a value-weighted portfolio of the stocks with the 30% lowest weights on the Δdp ($\Delta\sigma^2$) mimicking portfolio, with mimicking weights estimated by Principal Component Analysis (PCA) using the method in Giglio and Xiu (2020). To construct $r_{N\mathbb{E}}$ and $r_{N\mathbb{E}}^{PCA}$ ($r_{N\mathbb{V}}$ and $r_{N\mathbb{V}}^{PCA}$), we project $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) onto returns from decile portfolios constructed by sorting stocks based on the same signals used for $r_{\mathbb{E}}$ and $r_{\mathbb{E}}^{PCA}$ ($r_{\mathbb{V}}$ and $r_{\mathbb{V}}^{PCA}$) while imposing that projection coefficients sum to zero (i.e., the factors are zero-net-cost portfolios). The news are based on Equations 21 and 22, and are estimated ex-post using our Long Sample (1928-2019) or Modern Sample (1973-2019). The tradable risk factors as well as the decile portfolios necessary to obtain the ex-post mimicking factors are constructed each month using their respective signals estimated on a 5-year rolling window. Subsections 1.3.3, 1.3.4, and D.2 provide further empirical details on the construction of the tradable and ex-post mimicking risk factors. A detailed description of our news estimation procedure is available in Subsection B.1.

PANEL A: Reinvestment Risk ($N_{\mathbb{E}}$)

	Long Sample (1928-2019)						Modern Sample (1973-2019)					
	Δdp	$N_{\mathbb{E}}$	$r_{\mathbb{E}}$	$r_{N\mathbb{E}}$	$r_{\mathbb{E}}^{PCA}$	$r_{N\mathbb{E}}^{PCA}$	Δdp	$N_{\mathbb{E}}$	$r_{\mathbb{E}}$	$r_{N\mathbb{E}}$	$r_{\mathbb{E}}^{PCA}$	$r_{N\mathbb{E}}^{PCA}$
Δdp	1	0.91	0.68	0.68	0.39	0.57	1	0.88	0.54	0.57	0.41	0.47
$N_{\mathbb{E}}$	0.91	1	0.51	0.51	0.31	0.44	0.88	1	0.36	0.39	0.28	0.36
$r_{\mathbb{E}}$	0.68	0.51	1	0.93	0.53	0.81	0.54	0.36	1	0.89	0.74	0.68
$r_{N\mathbb{E}}$	0.68	0.51	0.93	1	0.44	0.76	0.57	0.39	0.89	1	0.66	0.69
$r_{\mathbb{E}}^{PCA}$	0.39	0.31	0.53	0.44	1	0.63	0.41	0.28	0.74	0.66	1	0.71
$r_{N\mathbb{E}}^{PCA}$	0.57	0.44	0.81	0.76	0.63	1	0.47	0.36	0.68	0.69	0.71	1

PANEL B: Volatility Risk ($N_{\mathbb{V}}$)

	Long Sample (1928-2019)						Modern Sample (1973-2019)					
	$\Delta\sigma^2$	$N_{\mathbb{V}}$	$r_{\mathbb{V}}$	$r_{N\mathbb{V}}$	$r_{\mathbb{V}}^{PCA}$	$r_{N\mathbb{V}}^{PCA}$	$\Delta\sigma^2$	$N_{\mathbb{V}}$	$r_{\mathbb{V}}$	$r_{N\mathbb{V}}$	$r_{\mathbb{V}}^{PCA}$	$r_{N\mathbb{V}}^{PCA}$
$\Delta\sigma^2$	1	0.75	0.24	0.27	0.09	0.24	1	0.92	0.25	0.30	0.18	0.28
$N_{\mathbb{V}}$	0.75	1	0.24	0.30	0.12	0.26	0.92	1	0.28	0.36	0.22	0.32
$r_{\mathbb{V}}$	0.24	0.24	1	0.81	0.47	0.70	0.25	0.28	1	0.84	0.69	0.79
$r_{N\mathbb{V}}$	0.27	0.30	0.81	1	0.42	0.72	0.30	0.36	0.84	1	0.57	0.74
$r_{\mathbb{V}}^{PCA}$	0.09	0.12	0.47	0.42	1	0.53	0.18	0.22	0.69	0.57	1	0.68
$r_{N\mathbb{V}}^{PCA}$	0.24	0.26	0.70	0.72	0.53	1	0.28	0.32	0.79	0.74	0.68	1

Table IA.3
Decile Portfolios Sorted on β_{dp} and β_{σ^2} (Three-Year βs)

This table reports statistics related to monthly returns on 10 β_{dp}-sorted portfolios (Panel A) and 10 β_{σ^2}-sorted portfolios (Panel B) when sorting on exposure to either Δdp or $\Delta\sigma^2$, respectively, estimated over three-year rolling windows. All data is from our Long Sample (1928-2019). The portion of Panel A reports portfolio return exposures to our expected return news proxy (Δdp), the in-sample expected return news measure ($N_{\mathbb{E}}$), our tradable reinvestment risk factor ($r_{\mathbb{E}}$), and the $N_{\mathbb{E}}$ mimicking portfolio ($r_{N\mathbb{E}}$). Portfolio return exposures to each of these time series are denoted by β_{dp}, $\beta_{N\mathbb{E}}$, $\beta_{\mathbb{E}}$, and $\beta_{rN\mathbb{E}}$, respectively, and are normalized to be in market beta units. The portion of Panel B reports portfolio return exposures to our volatility news proxy ($\Delta\sigma^2$), the in-sample volatility news measure ($N_{\mathbb{V}}$), our tradable volatility risk factor ($r_{\mathbb{V}}$), and the $N_{\mathbb{V}}$ mimicking portfolio ($r_{N\mathbb{V}}$). Portfolio return exposures to each of these time series are denoted by β_{σ^2}, $\beta_{N\mathbb{V}}$, $\beta_{\mathbb{V}}$, and $\beta_{rN\mathbb{V}}$, respectively, and are normalized to be in market beta units (see Footnote 26). In the bottom portion of each panel, we report portfolio average returns ($\mathbb{E}[r]$) and αs when computed with respect to the CAPM (α_m), the ICAPM excluding $r_{\mathbb{E}}$ ($\alpha_{m,\mathbb{V}}$) or $r_{\mathbb{V}}$ ($\alpha_{m,\mathbb{E}}$), and the full ICAPM ($\alpha_{m,\mathbb{E},\mathbb{V}}$). All returns are in percent and annualized (approximately) by multiplying monthly returns by 12. The "Slope" statistic is a measure of the slope of the 10 related portfolio statistics with respect to portfolio decile (see Footnote 26). Portfolios are rebalanced monthly based on individual stock exposures to Δdp or σ^2 with further details provided in Subsection 1.3.1. The 10-1 portfolio t-statistics are computed according to Newey and West (1987, 1994). The Slope t-statistics are computed according to the method in Driscoll and Kraay (1998) using the procedure in Newey and West (1994) to select the number of lags.

PANEL A: β_{dp}-Sorted Portfolios

Dec	1	2	3	4	5	6	7	8	9	10	10-1	(t_{10-1})	Slope	(t_{Slope})
β_{dp}	-1.52	-1.30	-1.18	-1.07	-1.00	-0.94	-0.83	-0.74	-0.67	-0.57	0.96	(9.93)	0.88	(7.98)
$\beta_{N\mathbb{E}}$	-1.14	-0.98	-0.88	-0.80	-0.75	-0.71	-0.62	-0.55	-0.50	-0.42	0.72	(6.16)	0.66	(4.88)
$\beta_{\mathbb{E}}$	-1.70	-1.41	-1.21	-1.04	-0.94	-0.85	-0.71	-0.55	-0.45	-0.34	1.36	(55.1)	1.27	(56.0)
$\beta_{rN\mathbb{E}}$	-1.68	-1.41	-1.18	-1.06	-0.93	-0.84	-0.71	-0.59	-0.47	-0.34	1.35	(32.5)	1.24	(35.5)
$\mathbb{E}[r]$	9.4%	8.7%	9.1%	9.3%	8.8%	9.3%	9.5%	8.2%	7.6%	6.8%	-2.7%	(-0.98)	-1.9%	(-0.54)
α_m	-4.1%	-2.9%	-1.4%	-0.3%	-0.2%	0.9%	2.1%	1.6%	1.6%	1.8%	5.9%	(2.98)	5.9%	(2.50)
$\alpha_{m,\mathbb{V}}$	-4.3%	-3.1%	-1.5%	-0.3%	-0.2%	1.0%	2.2%	1.7%	1.7%	1.9%	6.2%	(4.00)	6.2%	(3.27)
$\alpha_{m,\mathbb{E},\mathbb{V}}$	0.0%	0.1%	0.8%	1.2%	0.6%	1.4%	2.3%	0.6%	0.3%	0.1%	0.1%	(0.10)	0.3%	(0.45)

PANEL B: β_{σ^2}-Sorted Portfolios

Dec	1	2	3	4	5	6	7	8	9	10	10-1	(t_{10-1})	Slope	(t_{Slope})
β_{σ^2}	-0.50	-0.41	-0.38	-0.37	-0.33	-0.30	-0.29	-0.29	-0.26	-0.26	0.24	(6.55)	0.21	(5.02)
$\beta_{N\mathbb{V}}$	-0.54	-0.44	-0.42	-0.41	-0.37	-0.34	-0.33	-0.31	-0.29	-0.30	0.25	(6.35)	0.22	(4.62)
$\beta_{\mathbb{V}}$	-1.23	-1.09	-0.96	-0.80	-0.65	-0.59	-0.50	-0.39	-0.36	-0.38	0.85	(15.8)	0.90	(43.9)
$\beta_{rN\mathbb{V}}$	-1.16	-1.00	-0.87	-0.75	-0.61	-0.55	-0.45	-0.40	-0.34	-0.38	0.78	(14.5)	0.81	(17.9)
$\mathbb{E}[r]$	10.8%	10.1%	10.5%	9.6%	9.0%	8.7%	7.8%	7.6%	6.5%	6.2%	-4.5%	(-2.27)	-4.7%	(-1.82)
α_m	-0.6%	-0.4%	1.0%	0.8%	1.0%	1.3%	0.7%	0.8%	-0.3%	-1.1%	-0.5%	(-0.27)	-0.2%	(-0.10)
$\alpha_{m,\mathbb{E}}$	1.7%	1.4%	2.1%	1.4%	1.1%	1.1%	0.0%	0.1%	-0.9%	-1.6%	-3.3%	(-2.38)	-3.3%	(-2.03)
$\alpha_{m,\mathbb{E},\mathbb{V}}$	0.1%	0.0%	0.8%	0.8%	0.8%	0.9%	0.1%	1.0%	0.3%	0.0%	-0.1%	(-0.07)	0.0%	(0.01)

Table IA.4
The ICAPM Risk Prices and Pricing Errors (Three-Year βs)

Panel A reports estimated CAPM and ICAPM risk prices (b) according to Equation 8 while Panel B reports the annualized average returns ($\mathbb{E}[r]$) and associated pricing errors (α) for the three orthogonalized strategies introduced in Subsection 2.3. The ICAPM$_\gamma$ column reports the respective information when imposing the ICAPM structural restrictions in Footnote 11, which imply relative risk aversion, γ, is the only parameter as it determines all three risk prices (see Footnote 30 for estimation details). For the CAPM and ICAPM columns of each panel, we use $\gamma = b_m = 1_m' \Sigma_f^{-1} \mathbb{E}[f]$ (see Subsection 1.3.1). Since bs are not easily comparable, we report $\sigma_k \cdot b_k$ for each factor $f_{k,t}$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective $f_{k,t}$ (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed according to GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix (see Subsection B.2).

	Long Sample (1928-2019)					**Modern Sample (1973-2019)**				
		(5-Year Betas)		**(3-Year Betas)**			**(5-Year Betas)**		**(3-Year Betas)**	
	CAPM	**ICAPM**	**ICAPM$_\gamma$**	**ICAPM**	**ICAPM$_\gamma$**	**CAPM**	**ICAPM**	**ICAPM$_\gamma$**	**ICAPM**	**ICAPM$_\gamma$**
PANEL A: Risk Prices ($M_t = a + b' f_t$)										
b_m	0.12	0.26	0.34	0.25	0.31	0.12	0.26	0.29	0.24	0.268
(t_{stat})	(3.05)	(3.59)	(6.00)	(3.31)	(5.49)	(2.44)	(3.45)	(7.19)	(3.20)	(6.41)
$b_{\mathbb{E}}$		0.32	0.26	0.28	0.24		0.37	0.27	0.30	0.23
(t_{stat})		(4.28)	(5.05)	(3.70)	(4.55)		(4.19)	(6.14)	(3.82)	(5.37)
$b_{\mathbb{V}}$		-0.18	-0.11	-0.15	-0.08		-0.23	-0.23	-0.17	-0.17
(t_{stat})		(-3.34)	(-2.53)	(-3.06)	(-2.28)		(-3.52)	(-3.07)	(-3.01)	(-2.69)
$[\gamma]$	[2.3]	[4.8]	[6.3]	[4.6]	[5.9]	[2.8]	[5.6]	[6.9]	[5.3]	[6.2]
PANEL B: Annualized Pricing Errors (αs)										
$\mathbb{E}[r_m^\perp]$	10.5%	10.5%	10.5%	10.5%	10.1%	10.2%	10.2%	10.2%	10.2%	9.7%
α_m	5.6%	0.0%	-3.2%	0.0%	-2.7%	5.2%	0.0%	-2.2%	0.0%	-1.07%
(t_{stat})	(3.72)	(0.00)	(-2.18)	(0.00)	(-2.05)	(3.32)	(0.00)	(-1.01)	(0.00)	(-0.80)
$\mathbb{E}[r_{\mathbb{E}}^\perp]$	9.9%	9.9%	9.9%	9.9%	9.3%	10.3%	10.3%	10.3%	10.3%	9.2%
$\alpha_{\mathbb{E}}$	9.9%	0.0%	2.0%	0.0%	1.4%	10.3%	0.0%	2.6%	0.0%	2.2%
(t_{stat})	(5.07)	(0.00)	(2.33)	(0.00)	(1.77)	(4.47)	(0.00)	(1.89)	(0.00)	(1.61)
$\mathbb{E}[r_{\mathbb{V}}^\perp]$	-6.6%	-6.6%	-6.6%	-6.6%	-6.0%	-7.6%	-7.6%	-7.6%	-7.6%	-5.9%
$\alpha_{\mathbb{V}}$	-6.6%	0.0%	-2.7%	0.0%	-2.7%	-7.6%	0.0%	-0.2%	0.0%	-0.1%
(t_{stat})	(-3.38)	(0.00)	(-1.46)	(0.00)	(-1.45)	(-3.29)	(0.00)	(-0.12)	(0.00)	(0.04)

Table IA.5
Correlations: News Proxies, Tradable Factors, and Ex-post Mimicking Factors (News from Bayesian Framework and VAR Specifications)

This table reports correlations between our tradable risk factors ($r_{\mathbb{E}}$ and $r_{\mathbb{V}}$) and ex-post news ($N_{\mathbb{E}}$ and $N_{\mathbb{V}}$) as well as ex-post news mimicking factors ($r_{N\mathbb{E}}$ and $r_{N\mathbb{V}}$). To construct $r_{\mathbb{E}}$ ($r_{\mathbb{V}}$), we buy a value-weighted portfolio of the stocks with the 30% highest exposures to Δdp ($\Delta\sigma^2$) and sell a value-weighted portfolio of the stocks with the 30% lowest exposures to Δdp ($\Delta\sigma^2$). To construct $r_{N\mathbb{E}}$ ($r_{N\mathbb{V}}$), we project $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) onto returns from decile portfolios constructed by sorting stocks based on their exposure to $N_{\mathbb{E}}$ ($N_{\mathbb{V}}$) and imposing that projection coefficients sum to zero (i.e., the factors are zero-net-cost portfolios). The news are estimated ex-post over our Long (1928-2019) or Modern (1973-2019) Sample based on the Bayesian framework (Equations 21 and 22) of the VAR in Equation IA.134, with different panels using different state vectors. The tradable risk factors, $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$, as well as the decile portfolios necessary to obtain the ex-post mimicking factors are constructed each month using risk exposures estimated on a 5-year rolling window. Subsections 1.3.3, 1.3.4, and D.5 provide further details on the state variables and construction of risk factors.

PANEL A: News from Bayesian Framework

Sample	$Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$	$Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$
Long	0.51	0.51	0.93	0.30	0.24	0.81
Modern	0.39	0.36	0.89	0.36	0.28	0.84

PANEL B: News from VAR with $s = (dp,\ RV)$

Sample	$Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$	$Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$
Long	0.69	0.66	0.95	0.60	0.48	0.83
Modern	0.58	0.50	0.90	0.22	0.17	0.76

PANEL C: News from VAR with $s = (dp,\ RV,\ CS)$

Sample	$Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$	$Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$
Long	0.64	0.62	0.93	0.62	0.46	0.77
Modern	0.58	0.50	0.89	0.34	0.22	0.75

PANEL D: News from VAR with $s = (dp,\ RV,\ CS,\ TY,\ TS)$

Sample	$Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$	$Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$
Long	0.67	0.64	0.92	0.57	0.43	0.77
Modern	0.59	0.51	0.90	0.37	0.25	0.77

PANEL E: News from VAR with $s = (dp,\ RV,\ CS,\ TY,\ TS,\ VS)$

Sample	$Cor(r_{N\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, N_{\mathbb{E}})$	$Cor(r_{\mathbb{E}}, r_{N\mathbb{E}})$	$Cor(r_{N\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, N_{\mathbb{V}})$	$Cor(r_{\mathbb{V}}, r_{N\mathbb{V}})$
Long	0.61	0.57	0.91	0.56	0.25	0.46
Modern	0.60	0.51	0.86	0.32	0.19	0.63

Table IA.6
The ICAPM Risk Prices and Pricing Errors (Alternative $N_{\mathbb{V}}$)

Panel A reports estimated ICAPM risk prices (b) while Panel B reports the annualized average returns ($\mathbb{E}[r]$) and associated pricing errors (α) for the three orthogonalized strategies introduced in Subsection 2.3. All columns impose the ICAPM structural restrictions in Footnote 11, which imply relative risk aversion, γ, is the only parameter as it determines all three risk prices (see Footnote 30 for estimation details). Since bs are not easily comparable, we report $\sigma_k \cdot b_k$ for each factor $f_{k,t}$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective $f_{k,t}$ (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed according to GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix (see Internet Appendix B.2). Columns differ in the $N_{\mathbb{V}}$ specification, with the first column of each panel reporting our baseline results. Column "MLE $N_{\mathbb{V}}$" estimates our log-GARCH process used to obtain $N_{\mathbb{V}}$ by maximum likelihood (see Subsection D.6), and column "Nonlinear $N_{\mathbb{V}}$" obtains the nonlinear $N_{\mathbb{V}}$ numerically (see Subsection D.7).

	Long Sample (1928-2019)			**Modern Sample (1973-2019)**		
	(Baseline $N_{\mathbb{V}}$)	**(MLE $N_{\mathbb{V}}$)**	**(Nonlinear $N_{\mathbb{V}}$)**	**(Baseline $N_{\mathbb{V}}$)**	**(MLE $N_{\mathbb{V}}$)**	**(Nonlinear $N_{\mathbb{V}}$)**
	ICAPM$_{\gamma}$	**ICAPM$_{\gamma}$**	**ICAPM$_{\gamma}$**	**ICAPM$_{\gamma}$**	**ICAPM$_{\gamma}$**	**ICAPM$_{\gamma}$**
PANEL A: Risk Prices ($M_t = a + b' f_t$)						
b_m	0.34	0.32	0.33	0.29	0.34	0.27
(t_{stat})	(6.00)	(5.18)	(6.80)	(7.19)	(5.23)	(7.97)
$b_{\mathbb{E}}$	0.26	0.26	0.25	0.27	0.30	0.23
(t_{stat})	(5.05)	(4.30)	(5.68)	(6.14)	(4.54)	(6.65)
$b_{\mathbb{V}}$	-0.11	-0.02	-0.14	-0.23	-0.08	-0.26
(t_{stat})	(-2.53)	(-2.15)	(-2.84)	(-3.07)	(-2.27)	(-3.33)
$[\gamma]$	[6.3]	[5.9]	[6.1]	[6.9]	[7.6]	[6.0]
PANEL B: Annualized Pricing Errors (αs)						
$\mathbb{E}[r_m^{\perp}]$	10.5%	10.5%	10.5%	10.2%	10.2%	10.2%
α_m	-3.2%	-2.3%	-2.7%	-2.2%	-3.5%	-0.7%
(t_{stat})	(-2.18)	(-2.33)	(-1.58)	(-1.01)	(-2.17)	(-0.30)
$\mathbb{E}[r_{\mathbb{E}}^{\perp}]$	9.9%	9.9%	9.9%	10.3%	10.3%	10.3%
$\alpha_{\mathbb{E}}$	2.0%	1.9%	2.3%	2.6%	1.8%	3.7%
(t_{stat})	(2.33)	(1.88)	(2.38)	(1.89)	(1.50)	(2.30)
$\mathbb{E}[r_{\mathbb{V}}^{\perp}]$	-6.6%	-6.6%	-6.6%	-7.6%	-7.6%	-7.6%
$\alpha_{\mathbb{V}}$	-2.7%	-5.7%	-1.2%	-0.2%	-4.8%	1.0%
(t_{stat})	(-1.46)	(-2.83)	(-0.71)	(-0.12)	(-2.28)	(0.74)

Table IA.7
The ICAPM Risk Prices Controlling for Sentiment and Liquidity Factors

This table reports estimated risk prices (b) for ICAPM risk factors (f_t) according to Equation 8 when controlling for sentiment and liquidity factors (x_t) in the SDF $M_t = a - b' f_t - b_x' x_t$. The sentiment factor is a tradable version of the index in Baker and Wurgler (2006), with construction analogous to $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$, and data (until 2018) obtained from Jeffrey Wurgler's website. For liquidity, we use the tradable liquidity factor in Pástor and Stambaugh (2003) available on CRSP. Panels A and B cover our Long (1928-2019) and Modern (1973-2019) samples, respectively. In the case of the Long Sample, we include the earliest factor data available for each model (1970 for sentiment and 1968 for liquidity) and use the Stambaugh (1997) procedure to estimate b over the entire Long Sample (see Subsection 3.1 for more details). Since bs are not easily comparable, we report $\sigma_f \cdot b$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective factor (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed using a bootstrap exercise in Panel A (see Internet Appendix B.4) and GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix in Panel B (see Subsection B.2).

PANEL A: Long Sample (1928-2019)

	$x =$	CAPM	ICAPM	CAPM+SENT	ICAPM+SENT	CAPM+LIQ	ICAPM+LIQ
r_m	b	0.12	0.26	0.14	0.25	0.12	0.26
	(t_{stat})	(3.05)	(3.59)	(2.86)	(4.50)	(2.45)	(4.70)
$r_{\mathbb{E}}$	b		0.32		0.32		0.29
	(t_{stat})		(4.28)		(4.57)		(4.03)
$r_{\mathbb{V}}$	b		-0.18		-0.19		-0.14
	(t_{stat})		(-3.34)		(-3.38)		(-2.30)
r_{SENT}	b			0.08	0.05	0.11	0.09
or r_{LIQ}	(t_{stat})			(1.79)	(1.28)	(2.73)	(2.05)

PANEL B: Modern Sample (1973-2019)

	$x =$	CAPM	ICAPM	CAPM+SENT	ICAPM+SENT	CAPM+LIQ	ICAPM+LIQ
r_m	b	0.12	0.26	0.13	0.24	0.15	0.25
	(t_{stat})	(2.44)	(3.45)	(2.39)	(3.31)	(2.34)	(3.44)
$r_{\mathbb{E}}$	b		0.37		0.34		0.34
	(t_{stat})		(4.19)		(3.97)		(3.78)
$r_{\mathbb{V}}$	b		-0.23		-0.23		-0.19
	(t_{stat})		(-3.52)		(-3.69)		(-2.80)
r_{SENT}	b			0.11	0.05	0.30	0.09
or r_{LIQ}	(t_{stat})			(2.64)	(1.15)	(3.62)	(1.87)

Table IA.8
The ICAPM Risk Prices Controlling for the Betting Against Beta Factor

This table reports estimated risk prices (b) for ICAPM risk factors (f_t) according to Equation 8 when controlling for the betting against beta (BaB) factor of Frazzini and Pedersen (2014) in the SDF $M_t = a - b' f_t - b_{BaB} \cdot r_{BaB,t}$. The columns under "Original BaB" rely on our replication of the BaB factor in Frazzini and Pedersen (2014). The columns under "Mkt Hedged BaB" adjust the BaB factor so that its implicit market hedge is based on a position on the market portfolio. The columns under "Not Hedged BaB" fully remove the market hedging position from the BaB factor. Further details associated with the different column headers are provided in Subsection D.9. Panels A and B cover our Long (1928-2019) and Modern (1973-2019) samples, respectively. Since bs are not easily comparable, we report $\sigma_f \cdot b$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective factor (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed using GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix (see Subsection B.2).

PANEL A: Long Sample (1928-2019)

	$x =$	CAPM	ICAPM	Original BaB		Mkt Hedged BaB		Not Hedged BaB	
r_m	b	0.12	0.26	0.14	0.12	0.14	0.19	0.26	0.29
	(t_{stat})	(3.05)	(3.59)	(3.09)	(1.86)	(3.18)	(3.12)	(3.26)	(3.46)
$r_{\mathbb{E}}$	b		0.32		0.14		0.22		0.23
	(t_{stat})		(4.28)		(2.03)		(3.15)		(3.29)
$r_{\mathbb{V}}$	b		-0.18		-0.21		-0.19		-0.19
	(t_{stat})		(-3.34)		(-3.45)		(-3.36)		(-3.33)
r_{BaB}	b			0.24	0.26	0.13	0.11	0.19	0.14
	(t_{stat})			(4.34)	(4.27)	(2.73)	(2.04)	(2.62)	(1.68)

PANEL B: Modern Sample (1973-2019)

	$x =$	CAPM	ICAPM	Original BaB		Mkt Hedged BaB		Not Hedged BaB	
r_m	b	0.12	0.26	0.15	0.16	0.15	0.16	0.32	0.36
	(t_{stat})	(2.44)	(3.45)	(2.34)	(2.29)	(2.63)	(2.50)	(3.05)	(3.40)
$r_{\mathbb{E}}$	b		0.37		0.21		0.22		0.22
	(t_{stat})		(4.19)		(2.34)		(2.47)		(2.38)
$r_{\mathbb{V}}$	b		-0.23		-0.28		-0.29		-0.30
	(t_{stat})		(-3.52)		(-3.26)		(-3.43)		(-3.38)
r_{BaB}	b			0.30	0.32	0.19	0.22	0.27	0.33
	(t_{stat})			(3.62)	(3.38)	(2.76)	(2.37)	(2.74)	(2.29)

Table IA.9
The ICAPM Risk Prices Controlling for a Short-term Expected Return Factor

This table reports estimated risk prices (b) for ICAPM risk factors (f_t) according to Equation 8 when controlling for a tradable version of the short-term expected return factor in Kozak and Santosh (2020) through the SDF $M_t = a - b' f_t - b_{\mathbb{E}}^{ST} \cdot r_{\mathbb{E},t}^{ST}$. The construction of $r_{\mathbb{E}}^{ST}$ is analogous to the construction of $r_{\mathbb{E}}$ and $r_{\mathbb{V}}$ except that the betas used for sorting are based on stock exposures to the factor $\Sigma_{h=1}^{H} \kappa^{h-1} \cdot r_{m,t+h}$ and the beta estimation window ends H months before the sorting date to avoid a look-ahead bias. We rely on $\kappa = 0.90$ and the columns differ in the H used. The baseline analysis in Kozak and Santosh (2020) is based on $\kappa = 0.90$ and H=12 months, with H=36 months being the longest horizon they explore. Further details on $r_{\mathbb{E}}^{ST}$ are provided in Subsection D.10. Panels A and B cover our Long (1928-2019) and Modern (1973-2019) samples, respectively. Since bs are not easily comparable, we report $\sigma_f \cdot b$ so that the reported values can be interpreted as the change in M_t induced by a one standard deviation change in the respective factor (holding other factors fixed). b is estimated by Generalized Method of Moments (GMM) and the t-statistics are computed using GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix (see Subsection B.2).

PANEL A: Long Sample (1928-2019)

	$x =$	CAPM	ICAPM	H=12 months		H=24 months		H=36 months	
r_m	b	0.12	0.26	0.13	0.28	0.12	0.27	0.12	0.27
	(t_{stat})	(3.05)	(3.59)	(3.48)	(3.88)	(3.20)	(3.82)	(3.08)	(3.75)
$r_{\mathbb{E}}$	b		0.32		0.31		0.32		0.32
	(t_{stat})		(4.28)		(3.83)		(3.92)		(3.97)
$r_{\mathbb{V}}$	b		-0.18		-0.15		-0.15		-0.16
	(t_{stat})		(-3.34)		(-2.66)		(-2.69)		(-2.74)
$r_{\mathbb{E}}^{ST}$	b			-0.08	-0.07	-0.07	-0.07	-0.07	-0.07
	(t_{stat})			(-2.44)	(-2.08)	(-2.05)	(-1.94)	(-1.89)	(-1.92)

PANEL B: Modern Sample (1973-2019)

	$x =$	CAPM	ICAPM	H=12 months		H=24 months		H=36 months	
r_m	b	0.12	0.26	0.12	0.25	0.11	0.25	0.11	0.24
	(t_{stat})	(2.44)	(3.45)	(2.36)	(3.42)	(2.23)	(3.32)	(2.14)	(3.25)
$r_{\mathbb{E}}$	b		0.37		0.36		0.37		0.36
	(t_{stat})		(4.19)		(4.03)		(4.11)		(4.10)
$r_{\mathbb{V}}$	b		-0.23		-0.22		-0.22		-0.22
	(t_{stat})		(-3.52)		(-3.19)		(-3.25)		(-3.30)
$r_{\mathbb{E}}^{ST}$	b			-0.07	-0.07	-0.06	-0.07	-0.05	-0.06
	(t_{stat})			(-1.61)	(-1.59)	(-1.33)	(-1.47)	(-1.09)	(-1.23)

Table IA.10
Return Reversals over Long Horizons

This table reports correlations between relatively short term market returns, $\Sigma_{h=0}^{H-1}\kappa^h \cdot r_{m,t-h}$ and subsequent long-term market returns $\Sigma_{h=1}^{120}\delta^h \cdot r_{m,t+h}$. We also report 95% confidence intervals for these correlations, which are robust to autocorrelation and heteroskedasticity (see Dalla, Giraitis, and Phillips (2020)). The rows differ in the κ value used ($\kappa = 0.90$ and $\kappa = 0.97$) and the columns differ in the H value used (12 months, 24 months, and 36 months). The baseline analysis in Kozak and Santosh (2020) is based on $\kappa = 0.90$ and H=12 months, with $\kappa = 0.97$ being the highest discounting they explore and H=36 months being the longest horizon they explore.

PANEL A: Long Sample (1928-2019)

		H=12 months	**H=24 months**	**H=36 months**
$\boldsymbol{\kappa = 0.90}$	$Cor\left(\sum_{h=0}^{H-1} \kappa^h \cdot r_{m,t-h}\ ,\ \sum_{h=1}^{120} \delta^h \cdot r_{m,t+h}\right)$	-0.06	-0.06	-0.07
	[Robust 95% Confidence Interval]	[-0.12 ; 0.00]	[-0.12 ; 0.00]	[-0.13 ; 0.00]
$\boldsymbol{\kappa = 0.97}$	$Cor\left(\sum_{h=0}^{H-1} \kappa^h \cdot r_{m,t-h}\ ,\ \sum_{h=1}^{120} \delta^h \cdot r_{m,t+h}\right)$	-0.07	-0.08	-0.07
	[Robust 95% Confidence Interval]	[-0.13 ; -0.01]	[-0.14 ; -0.01]	[-0.14 ; 0.00]

PANEL B: Modern Sample (1973-2019)

		H=12 months	**H=24 months**	**H=36 months**
$\boldsymbol{\kappa = 0.90}$	$Cor\left(\sum_{h=0}^{H-1} \kappa^h \cdot r_{m,t-h}\ ,\ \sum_{h=1}^{120} \delta^h \cdot r_{m,t+h}\right)$	-0.24	-0.34	-0.35
	[Robust 95% Confidence Interval]	[-0.36 ; -0.12]	[-0.47 ; -0.20]	[-0.49 ; -0.21]
$\boldsymbol{\kappa = 0.97}$	$Cor\left(\sum_{h=0}^{H-1} \kappa^h \cdot r_{m,t-h}\ ,\ \sum_{h=1}^{120} \delta^h \cdot r_{m,t+h}\right)$	-0.26	-0.38	-0.45
	[Robust 95% Confidence Interval]	[-0.37 ; -0.14]	[-0.51 ; -0.25]	[-0.59 ; -0.31]