Dynamic Multivariate Learning with Generalized Information: Asset Pricing Implications

Praveen Kumar and James Yae*

This version: March 15, 2021

Abstract

Agents are generally uncertain about multiple, and possibly time-varying, structural parameters that drive consumption and financial payoffs but learn through noisy correlated signals, such as aggregate or macroeconomic news. We find that dynamic learning of multivariate time-varying parameters with correlated signals generates endogenous long-run risks resulting in large and never-decaying equity risk premium. In general, the risk premium is driven by *intertemporal co-uncertainty*, that is, the dynamic covariance of posterior means, rather than uncertainty (i.e., variance of beliefs) that is highlighted in the literature. Signal correlation structure plays a crucial role in the dynamics of beliefs and asset prices and hence the determination of the equity premium. Apart from its quantitative implications, signal correlation generates non-monotone effects of information quality on the equity premium. We also present empirical evidence of the prevalence of highly correlated signals. Our general learning framework highlights the economic effects of correlated signals on Bayesian learning.

Keywords: Learning, Parameter Uncertainty, Signal Correlation, Equity Premium
JEL Classification: D83, G12

*We thank Hendrik Bessembinder, John Campbell, Timothy Cogley, Evgenia Gvozdeva, Peter Hammond, Satish Iyengar, Ravi Jagannathan, Michael Johannes, Jay Kadane, Andrew Lo, Maureen O'Hara, Lubos Pastor, and Pradeep Yadav for helpful comments or discussions. Praveen Kumar and James Yae are at the C. T. Bauer College of Business, University of Houston. Emails: pkumar@uh.edu for Praveen Kumar and syae.uh@gmail.com for James Yae. This paper was earlier circulated under the title "Joint-Learning, Belief-Covariation, and Asset Prices".

1 Introduction

Investors' beliefs about risks and returns continually evolve as they process new information about the market and the economy. While the rational expectations assumption is analytically convenient, in practice, agents expend enormous resources on developing information, or generating signals, to reduce uncertainty regarding unknown structural parameters that drive consumption and dividends. Therefore, a challenge for researchers is to model the dynamic evolution of investors' beliefs based on realistic signals (or information structures) with quantitative asset pricing implications that are consistent with the data (see, e.g., Hansen 2007). However, for analytic tractability, the existing literature examines the role of parameter learning only with specialized cases of uncertainty and information environments. In this paper, we examine general equilibrium asset pricing implications of learning multivariate and possibly time-varying valuation-related structural parameters through noisy correlated signals (such as, aggregate or macroeconomic news).[1]

We find that multivariate learning on both consumption and dividend mean-growth-rates, under empirically plausible correlation structures of aggregate signals, endogenously generates *intertemporal co-uncertainty*, that is, the dynamic covariance of agents' (Bayesian) parameter estimates. In turn, intertemporal co-uncertainty creates a large equilibrium equity premium with recursive utility (Epstein and Zin 1989) even if dividends have no fundamental link to consumption in an endowment economy. Indeed, we show that this learning-induced risk premium never decays when the unknown parameters are time-varying, i.e., there are latent states.

We develop a generalized framework for *joint learning* when agents confront multivariate and possibly time-varying valuation-related structural parameters but dynamically learn through noisy *joint signals*—that is, signals that are interconnected with (or linked to) multiple unknown parameters. To fix ideas, consider the case where agents face unknown parameters that drive stock returns and output growth. Then, aggregate signals such as changes in corporate tax rates or economic assessments by monetary authorities affect beliefs

[1]The empirical importance of noisy signals, such as macroeconomic news shocks, in dynamic settings is highlighted by a growing literature (see Kurmann and Otrok 2013, Ghosh and Constantinides 2017, and Ai and Bansal 2018 among many others). Also, see Kumar et al. (2008) and the references therein for the examples in static settings.

on both future returns and output growth. Such signals, which are empirically prevalent, are irreducible to independent signals on individual (unknown) parameters; rather, they are examples of signals with correlated noisy components. Thus, our joint learning setting includes as special cases many asset pricing models with learning in the literature.[2]

But what are the novel economic implications of considering asset pricing in this generalized learning setting? To motivate our main results, it is helpful to consider a concrete joint learning setting. Suppose an agent learns about the mean consumption and dividend growth rates, μ_c and μ_d, via possibly correlated aggregate signals $s_c = \mu_c + e_{sc}$ and $s_d = \mu_d + \rho e_{sc} + e_{sd}$, which include the information from observed growth rates. Then the agent's Bayesian parameter uncertainty comprises of four distinct components: static uncertainty, static co-uncertainty, intertemporal uncertainty, and intertemporal co-uncertainty. Here, *static uncertainty* refers to the conditional (or posterior) variance of μ_c (at each information set). This is the usual parameter uncertainty or estimation risk in a one-period model with a single unknown parameter. With two unknowns, μ_c and μ_d, *static co-uncertainty* refers to the contemporaneous conditional covariance of the unknown parameters. However, because estimates of μ_c and μ_d (or the posterior means) dynamically evolve due to the Bayesian learning process, *intertemporal uncertainty* and *intertemporal co-uncertainty* refer to the dynamic variance and covariance, respectively, of the agent's Bayesian estimates.

Our paper presents two main results. First, we find that intertemporal co-uncertainty, as a by-product of learning through joint signals, translates into two key elements of the equity premium: (1) an endogenous long-run risk (Bansal and Yaron 2004) created by time-varying estimates on μ_c when the agent prefers early resolution of uncertainty (Collin-Dufresne et al. 2016), and (2) the endogenous exposure of dividends to this long-run risk. Intuitively, positive intertemporal co-uncertainty aggravates future consumption-risk diversification and investment opportunities, implying a high negative covariance between future marginal utility and stock returns. The demand for risky assets is, therefore, ceteris paribus declining in

[2] Johannes et al. (2016) also describe their learning problem as "joint learning" in the sense that agents face multiple unknown parameters. The key difference is that our joint learning framework allows correlated signals (as aggregate joint signals). Meanwhile, the signals in Kumar et al. (2008) are indeed joint signals consistent with our definition although they study the cross-section of equity returns based on the unknown fixed parameters in partial equilibrium. Bansal and Shaliastovich (2010), and Harbaugh et al. (2016) use multiple *independent* signals that, therefore, reduce to a single signal with the sample mean and variance as sufficient statistics for the reduced signal.

intertemporal co-uncertainty.

Second, the correlation structure of signals is a key determinant of intertemporal co-uncertainty and its effect on equity risk premium. We show that with dynamic learning on multivariate parameters (positive) intertemporal co-uncertainty equals the expected reduction in static co-uncertainty by future signals. To facilitate intuition, consider the case of univariate parameter uncertainty, where highly accurate signals have dramatic effects on Bayesian parameter estimation and substantially reduce static uncertainty. Therefore, intertemporal variance of parameter estimates depends on the decline in static (parameter) uncertainty. Our result is the multivariate cross-moment extension of this univariate case. Hence, reduction in static co-uncertainty through processing of correlated signals is accompanied by dynamically covarying parameter estimates (i.e., intertemporal co-uncertainty). This "trade-off" (dictated by the Bayesian learning process) has a substantial impact on asset demand because intertemporal co-uncertainty is associated with long-run risk, while static co-uncertainty is associated with short-run risk.[3] Therefore, learning with correlated signals can amplify (or boost) long-run risk.

Considering correlated signals has other implications as well. For example, the relation of information quality and equity premium becomes non-monotonic when the signals are negatively correlated, in contrast to some of the literature.[4] In addition, the equity risk premium increases with the volatility of idiosyncratic persistent components in dividends.

In our empirical analysis, we find evidence that the signals on the GDP and corporate earnings are highly correlated. From the Survey of Professional Forecasters (SPF) conducted by Federal Reserve Bank of Philadelphia, we estimate the implied signal correlations by applying Bayes' theorem to the observed forecast errors from 1968:Q4 to 2019:Q3. The estimation results are consistent with our long-run risk model calibrations that generate the historical equity premium.

To summarize, our study makes three contributions. First, we separate intertemporal co-uncertainty from conventional parameter uncertainty and show its significant effects on general equilibrium asset pricing. Second, our analysis highlights the importance of cor-

[3]We relate this trade-off to Frankel and Kamenica (2019)'s argument: the value of information and cost of uncertainty should be measured specific to the agent's decision problem.

[4]See Veronesi (2000), Ai (2010), and Brevik and d'Addona (2010).

related signals in asset pricing with learning. Third, we suggest how to estimate signal correlation from survey data.

Our study presents a novel channel for the effects of learning on the equity risk premium in general equilibrium. The existing literature presents two main channels for parameter uncertainty to affect the equity premium. The first channel analyzed by Weitzman (2007) exploits the fact that an agent' posterior belief on the consumption growth follows a *t-distribution* when the consumption growth follows a normal distribution with unknown mean and volatility.[5] As Geweke (2001) shows, expected utility diverges with heavy-tailed *t*-distributed consumption growth and, therefore, any level of equity premium can be attained by manipulating prior beliefs or truncation points in tails. The learning models in Gvozdeva et al. (2015) and Johannes et al. (2016) partially benefit from this channel when they generate the large risk premium.

The second channel in the literature emphasizes endogenous long run risks that arise with Bayesian updating of fixed unknown parameters (Gvozdeva et al. 2015 and Collin-Dufresne et al. 2016). This channel fundamentally differs from our joint learning framework in two aspects. First, our learning-induced risk premium never decays when unknown parameters are time-varying (i.e., in a stationary latent state model).[6] In contrast, endogenous long-run risk and risk premium with learning on fixed unknowns gradually dissipate, exhibiting a non-stationary process as learning reduces (static) uncertainty regarding the fixed parameters. Hence, these papers neither recognize nor fully utilize intertemporal co-uncertainty due to absence of learning through joint signals. Instead, they rely on the first channel (heavy-tail effects) or amplified static uncertainty that slows down the learning process to allow matching of observed historical equity premium before further decay.

Second, intertemporal co-uncertainty can *create*, not just amplify, the risk premium. In contrast, learning in Gvozdeva et al. (2015) and Collin-Dufresne et al. (2016) only amplifies existing priced macroeconomic risk, implying that their learning models cannot increase the risk premium if an asset generates cash flows independent of aggregate consumption.

[5] As Bakshi and Skoulakis (2010) point out, the equity premium is still difficult to justify with a power utility if volatility (as a normal scale mixture) is truncated at a reasonable level.

[6] Time-varying multiple unknowns are motivated by the large theoretical and empirical asset pricing literatures on time-varying risks (e.g., Merton 1973, Engle et al. 1987, Bollerslev et al. 1988, Bansal and Yaron 2004). These stationary dynamics are of interest from an econometric viewpoint.

In our model, the agent knows with certainty that asset payoffs and the economy are not intrinsically linked. Yet, she also expects that her estimates on the asset's profitability and the macroeconomic fundamentals will covary due to correlated aggregate signals, creating endogenous long-run risk and dividend exposure.[7] Our stationary latent state model can generate the historical equity risk premium (6% per year) without an exogenously imposed link between consumption and dividends or with low elasticity of intertemporal substitution (EIS) (for example, 0.5).[8]

The paper proceeds as follows. Section 2 motivates the analysis of joint learning. Section 3 studies general relationships among estimation uncertainties in joint learning. Section 4 further develops asset pricing implications of joint learning. Section 5 addresses empirical aspects of joint learning. Section 6 discusses future research directions and concludes.

2 Motivation

In this section, we motivate our analysis by demonstrating significant novel implications of learning on multiple unknowns through joint signals (that is, *joint learning*). We do so by extending a well known model in the literature where there is learning on a fixed unknown parameter. In particular, we show that joint learning can create a risk premium for an asset that is fundamentally idiosyncratic, that is, when there is no correlation between payout and consumption risk.

2.1 Joint Learning in a Bivariate-Normal Model

To demonstrate the possibilities introduced by joint learning, we generalize the i.i.d. learning model in Collin-Dufresne et al. (2016) as follows. Consider a representative-agent endowment economy. The log of growth rates $\mathbf{g}_{t+1} = [g_{c,t+1} \;\; g_{d,t+1}]^\top$ of consumption C_t and dividends D_t of a risky portfolio are drawn from an independent and identically distributed (i.i.d.)

[7] Dynamic multivariate learning connects news on the asset's future cash flows with news on future macroeconomic shocks, rather than scaling up the risk prices of macroeconomic shocks.

[8] Note this is not due to behavioral biases or mistakes in valuations, but a by-product of Bayesian learning.

bivariate normal distribution:

$$\mathbf{g}_{t+1} = \boldsymbol{\mu} + \mathbf{e}_{g,t+1}, \text{ where } \mathbf{e}_{g,t+1} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_g) \text{ and } \boldsymbol{\Sigma}_g = \begin{bmatrix} \sigma_c^2 & \rho_g \sigma_c \sigma_d \\ \rho_g \sigma_c \sigma_d & \sigma_d^2 \end{bmatrix}, \quad (1)$$

where $g_{c,t+1} = \log \frac{C_{t+1}}{C_t}$, $g_{d,t+1} = \log \frac{D_{t+1}}{D_t}$, and $\boldsymbol{\mu} = [\mu_c \;\; \mu_d]^\top$. The representative agent knows the variance-covariance matrix $\boldsymbol{\Sigma}_g$ exactly, including the correlation ρ_g between consumption and dividend growth rates, but does not know the mean-growth-rate vector $\boldsymbol{\mu}$.

The agent observes an aggregate signal $\mathbf{s}_{t+1} = [s_{c,t+1} \;\; s_{d,t+1}]^\top$ at $t+1$, which represents all the information available to the agent including growth realizations $\mathbf{g}_{t+1}$:

$$\mathbf{s}_{t+1} = \boldsymbol{\mu} + \mathbf{e}_{s,t+1}, \text{ where } \mathbf{e}_{s,t+1} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_s) \text{ and } \boldsymbol{\Sigma}_s = \begin{bmatrix} \sigma_{sc}^2 & \rho_s \sigma_{sc} \sigma_{sd} \\ \rho_s \sigma_{sc} \sigma_{sd} & \sigma_{sd}^2 \end{bmatrix}. \quad (2)$$

We denote the observed history of aggregate signals at t by the profile $\mathbf{s}^t = \{\mathbf{s}_1, ..., \mathbf{s}_t\}$. The agent then makes inference on $\boldsymbol{\mu}$ at time t, conditional only on $\mathbf{s}^t$, because $\mathbf{s}^t$ includes all available information by construction. Therefore, the growth history $\mathbf{g}^t$, conditional on $\mathbf{s}^t$, contains no additional information; that is, $p(\boldsymbol{\mu}|\mathbf{s}^t, \mathbf{g}^t) = p(\boldsymbol{\mu}|\mathbf{s}^t)$. In general, the growth shocks $\mathbf{e}_{g,t+1} = [e_{gc,t+1} \;\; e_{gd,t+1}]^\top$ and the signal noises $\mathbf{e}_{s,t+1} = [e_{sc,t+1} \;\; e_{sd,t+1}]^\top$ can be correlated, and their cross-covariance matrix is as follows.

$$\mathbf{J}_{gs} = \begin{bmatrix} cov(e_{gc,t+1}, e_{sc,t+1}) & cov(e_{gc,t+1}, e_{sd,t+1}) \\ cov(e_{gd,t+1}, e_{sc,t+1}) & cov(e_{gd,t+1}, e_{sd,t+1}) \end{bmatrix} \text{ and } \mathbf{J}_{sg} = \mathbf{J}_{gs}^\top. \quad (3)$$

The agent is ex-ante unbiased and her prior beliefs are equivalent to observing n periods of aggregate signals. Hence, at $t = 0$, the agent has prior beliefs given by

$$\boldsymbol{\mu} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_0, \boldsymbol{\Sigma}_{\mu,0}), \text{ where } \hat{\boldsymbol{\mu}}_0 = \begin{bmatrix} \mu_c \\ \mu_d \end{bmatrix} \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{\mu},0}^{-1} = n\boldsymbol{\Sigma}_s^{-1}. \quad (4)$$

Using Bayes' theorem, we express the agent's posterior beliefs on $\boldsymbol{\mu}$ at any $t \geq 1$, conditional

on her information set $\mathcal{F}_t = \{\mathbf{s}^t\}$, as follows.

$$\boldsymbol{\mu}|\mathcal{F}_t \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_t, \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}\right), \quad \text{where} \quad \hat{\boldsymbol{\mu}}_t = \begin{bmatrix} \hat{\mu}_{c,t} \\ \hat{\mu}_{d,t} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{\mu},t} = \begin{bmatrix} \sigma^2_{c,t} & \rho_{\boldsymbol{\mu},t}\sigma_{c,t}\sigma_{d,t} \\ \rho_{\boldsymbol{\mu},t}\sigma_{c,t}\sigma_{d,t} & \sigma^2_{d,t} \end{bmatrix}, \quad (5)$$

where the posterior mean and covariance matrix $(\hat{\boldsymbol{\mu}}_t, \boldsymbol{\Sigma}_{\boldsymbol{\mu},t})$ can be recursively computed as

$$\begin{aligned} \hat{\boldsymbol{\mu}}_t &= \hat{\boldsymbol{\mu}}_{t-1} + \mathbf{K}_{s,t}(\mathbf{s}_t - \hat{\boldsymbol{\mu}}_{t-1}), \\ \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu},t} &= \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu},t-1} + \boldsymbol{\Sigma}^{-1}_{s}. \end{aligned} \quad (6)$$

Here, $\mathbf{K}_{s,t} = \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}\boldsymbol{\Sigma}_s^{-1} = \frac{1}{n+t}\mathbf{I}$ is the (2×2) Kalman gain matrix associated with aggregate signals and $\mathbf{I}$ is an identity matrix.[9] Equation (6) indicates how joint signals can create a risk premium for an idiosyncratic asset ($\rho_g = 0$). The agent's Bayes-rational estimates on μ_c and μ_d can covary due to learning from positively correlated aggregate signals $\mathbf{s}_t$. As we will establish in the next section, the endogenous intertemporal covariation in these estimates can generate a risk premium in equilibrium even though the agent knows that the consumption and dividend dynamics are independent. This intuition is only a precursor to the deeper implications of joint learning that we develop in the analysis below and Section 3.

2.2 Joint-Learning and the Equity Premium

We now examine the effect of joint learning on the agent's beliefs and risk premium in the model setup in the previous section. Following Epstein and Zin (1989), the stochastic discount factor (SDF) for recursive preferences has the form

$$M_{t+1} = \delta^\theta G_{c,t+1}^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1}, \quad (7)$$

where $G_{c,t+1} = \frac{C_{t+1}}{C_t}$ is the aggregate gross growth rate of per-capita consumption and $R_{c,t+1} = \frac{W_{t+1}+C_{t+1}}{W_t}$ is the gross return on the total wealth portfolio W_t that pays out the aggregate consumption stream $\{C_t\}$. Among the preference parameters, $0 < \delta < 1$ is a time

[9] We can simplify Equation (6) to $\hat{\boldsymbol{\mu}}_t = \frac{n}{n+t}\boldsymbol{\mu}_0 + \frac{t}{n+t}\bar{\mathbf{s}}^t$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t} = \frac{1}{n+t}\boldsymbol{\Sigma}_s$, where $\bar{\mathbf{s}}^t$ is the sample average of the signals $\mathbf{s}_\tau$ for $\tau = 1, 2, ..., t$. However, we keep the recursive form in (6) because it can easily extend to consider situations with time-varying parameters.

discount factor, $\gamma \geq 0$ is the risk aversion parameter, $\psi \geq 0$ is the elasticity of intertemporal substitution (EIS), and $\theta = (1-\gamma)/(1-\frac{1}{\psi})$. The stochastic discount factor in (7) implies the following Euler equation conditional on the agent's information set $\mathcal{F}_t$:

$$E[\delta^{\theta} G_{c,t+1}^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1} R_{i,t+1} | \mathcal{F}_t] = 1, \tag{8}$$

where $R_{i,t+1}$ is the gross return of any asset i in the economy. Using this Euler equation, we solve for the wealth-consumption and the price-dividend ratios by log-linearizing asset returns recursively from the limiting case with perfect information at sufficiently large t.[10]

Following Campbell and Shiller (1988), the log wealth-consumption ratio $z_{c,t}$ and the log price-dividend ratio $z_{d,t}$ are linear functions of the state variable $\hat{\boldsymbol{\mu}}_t$ given in Equation (5):

$$z_{c,t} = \log \frac{W_t}{C_t} = A_{0c,t} + \mathbf{A}_{1c} \hat{\boldsymbol{\mu}}_t = A_{0c,t} + \left(\frac{1-\psi^{-1}}{1-\kappa_{1c}} \right) \hat{\mu}_{c,t}, \tag{9}$$

$$z_{d,t} = \log \frac{P_t}{D_t} = A_{0d,t} + \mathbf{A}_{1d} \hat{\boldsymbol{\mu}}_t = A_{0d,t} + \left(\frac{-\psi^{-1}}{1-\kappa_{1d}} \right) \hat{\mu}_{c,t} + \left(\frac{1}{1-\kappa_{1d}} \right) \hat{\mu}_{d,t}, \tag{10}$$

where both κ_{1c} and κ_{1d} are the constants between zero and one, determined by log-linearization. In (9) and (10), uncertainty regarding unknown parameters is a deterministic state variable (that is, it is non-random), consistent with a bivariate normal learning setting with unknown means and a known variance-covariance matrix. Therefore, $A_{0c,t}$ and $A_{0d,t}$ contain information about $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$. However, the impact of EIS (ψ) on $\mathbf{A}_{1c}$ and $\mathbf{A}_{1d}$ through intertemporal substitution and wealth effects is similar to that elucidated in Bansal and Yaron (2004).

The innovation in the log of the SDF, m_{t+1}, conditional on the agent's information set is

$$m_{t+1} - E[m_{t+1}|\mathcal{F}_t] = -\boldsymbol{\lambda}_g (\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) - \boldsymbol{\lambda}_{s,t} (\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \tag{11}$$

$$= -\boldsymbol{\lambda}_g \mathbf{e}_{g,t+1} - \boldsymbol{\lambda}_{s,t} \mathbf{e}_{s,t+1} - (\boldsymbol{\lambda}_g + \boldsymbol{\lambda}_{s,t}) \mathbf{u}_{\boldsymbol{\mu},t+1}, \tag{12}$$

where $\boldsymbol{\lambda}_g = \gamma \mathbf{I}_c = \gamma [1 \;\; 0]$, $\boldsymbol{\lambda}_{s,t} = (1-\theta)\kappa_{1c} \mathbf{A}_{1c} \mathbf{K}_{s,t+1}$, and $\hat{\boldsymbol{\mu}}_t = E[\mathbf{g}_{t+1}|\mathcal{F}_t] = E[\mathbf{s}_{t+1}|\mathcal{F}_t]$. The last shock $\mathbf{u}_{\boldsymbol{\mu},t+1} = \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\mu},t})$ is a random variable that represents the parameter uncertainty (or estimation risk) regarding $\boldsymbol{\mu}$ at time t in the agent's posterior beliefs, where

[10] The approach is similar to Bansal and Yaron (2004), extending to the case of unknown parameters. The details are provided in the online Appendix A.1.

$\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ is specified in (5). On the other hand, the innovation in the log of risky portfolio return $(r_{m,t+1})$ is

$$r_{m,t+1} - E[r_{m,t+1}|\mathcal{F}_t] = \boldsymbol{\beta}_g(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \boldsymbol{\beta}_{s,t}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \tag{13}$$

$$= \boldsymbol{\beta}_g \mathbf{e}_{g,t+1} + \boldsymbol{\beta}_{s,t}\mathbf{e}_{s,t+1} + (\boldsymbol{\beta}_g + \boldsymbol{\beta}_{s,t})\mathbf{u}_{\boldsymbol{\mu},t+1}, \tag{14}$$

where $\boldsymbol{\beta}_g = \mathbf{I}_d = [0 \;\; 1]$ and $\boldsymbol{\beta}_{s,t} = \kappa_{1d}\mathbf{A}_{1d}\mathbf{K}_{s,t+1}$. Using (12) and (14), we finally express the log-return counterpart of the equity risk premium $RP_t = -Cov(r_{m,t+1}, m_{t+1}|\mathcal{F}_t)$ as

$$E[r_{m,t+1}|\mathcal{F}_t] - r_{f,t} + \frac{1}{2}Var[r_{m,t+1}|\mathcal{F}_t] = \underbrace{\boldsymbol{\beta}_g\boldsymbol{\Sigma}_g\boldsymbol{\lambda}_g^\top}_{\text{from growth}} + \underbrace{\boldsymbol{\beta}_{s,t}\boldsymbol{\Sigma}_s\boldsymbol{\lambda}_{s,t}^\top}_{\text{from signal}} + \underbrace{\boldsymbol{\beta}_{s,t}\mathbf{J}_{sg}\boldsymbol{\lambda}_g^\top}_{\text{signal}\times\text{growth}} + \underbrace{\boldsymbol{\beta}_g\mathbf{J}_{gs}\boldsymbol{\lambda}_{s,t}^\top}_{\text{growth}\times\text{signal}}$$
$$+ \underbrace{(\boldsymbol{\beta}_g + \boldsymbol{\beta}_{s,t})\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}(\boldsymbol{\lambda}_g + \boldsymbol{\lambda}_{s,t})^\top}_{\text{from static (co)-uncertainty in } \boldsymbol{\mu}}. \tag{15}$$

The first term $\boldsymbol{\beta}_g\boldsymbol{\Sigma}_g\boldsymbol{\lambda}_{g,t}^\top$ on the right hand side of (15) is the conventional risk premium $\gamma Cov(g_{c,t}, g_{d,t})$ when the mean growth rates are known. The last term is related to $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, which represents current static parameter uncertainty at t. Quantitatively, the last term is negligible because $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ is only a fraction $1/n$ of $\boldsymbol{\Sigma}_s$ initially at $t = 0$ (see Equation (4)) , and it further decreases deterministically over time at the rate $O(t^{-1})$.

Most importantly, the three terms in the middle of the right hand side of (15) represent the risk premium associated with the agent's learning process. While integrating these three terms with the last term in (15), we rewrite them in detail using (6), (11) and (13) to explain how the effects of joint learning differ from its *special* cases that are studied in the literature.

$$RP_{ss} \triangleq \boldsymbol{\beta}_{s,t}\boldsymbol{\Sigma}_s^*\boldsymbol{\lambda}_{s,t}^\top = \kappa_{1c}^*\kappa_{1d}^*\left(\gamma - \frac{1}{\psi}\right)Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)\left[\frac{Cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t)}{Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)} - \frac{1}{\psi}\right], \tag{16}$$

$$RP_{sg} \triangleq \boldsymbol{\beta}_{s,t}\mathbf{J}_{sg}^*\boldsymbol{\lambda}_g^\top = \kappa_{1d}^*\gamma\left[Cov(\hat{\mu}_{d,t+1}, e_{gc,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{\mu}_{c,t+1}, e_{gc,t+1}|\mathcal{F}_t)\right], \tag{17}$$

$$RP_{gs} \triangleq \boldsymbol{\beta}_g\mathbf{J}_{gs}^*\boldsymbol{\lambda}_{s,t}^\top = \kappa_{1c}^*\left(\gamma - \frac{1}{\psi}\right)Cov(\hat{\mu}_{c,t+1}, e_{gd,t+1}|\mathcal{F}_t), \tag{18}$$

where $\boldsymbol{\Sigma}_s^* = \boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, $\mathbf{J}_{sg}^* = \mathbf{J}_{sg}^* + \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, $\mathbf{J}_{gs}^* = \mathbf{J}_{gs}^* + \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, $\kappa_{1c}^* = \kappa_{1c}/(1-\kappa_{1c}) > 0$, and $\kappa_{1d}^* = \kappa_{1d}/(1-\kappa_{1d}) > 0$.

Note RP_{ss} in (16) captures the joint-learning-induced long-run risk. If an agent has preferences for early resolution of uncertainty ($\gamma > \frac{1}{\psi}$), then RP_{ss} is positively related to

$$(\text{Belief Beta})_t \triangleq \frac{(\text{Intemporal Co-Uncertainty})_t}{(\text{Intemporal Uncertainty})_t} = \frac{Cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t)}{Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)}. \tag{19}$$

Here, we define *belief beta*, analogous to the CAPM beta, as the ratio of *intertemporal co-uncertainty*, $Cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t)$, to *intertemporal uncertainty*, $Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)$, when the agent's estimates $\hat{\mu}_{c,t+1}$ and $\hat{\mu}_{d,t+1}$ covary over time. If intertemporal co-uncertainty is large relative to intertemporal uncertainty, the belief beta can exceed a threshold ψ^{-1} and joint learning raises the risk premium. The belief beta, given intertemporal uncertainty, is therefore positively related to the risk premium when there is a preference for early resolution of uncertainty. The other terms RP_{sg} and RP_{gs} in the risk premium represent interaction effects between the growth shocks and aggregate signals, which we explore in Section 4. In our i.i.d. bivariate normal model, however, we assume $\mathbf{J}_{gs} = \mathbf{J}_{sg} = 0$ to 'shut down' these secondary channels; therefore, both terms RP_{sg} and RP_{gs} are zero here.

In general, intertemporal co-uncertainty (that is, the numerator of belief beta) has a more unambiguous effect on the risk premium, given ($\gamma > \frac{1}{\psi}$), than intertemporal uncertainty (that is, the denominator). For example, when we increase intertemporal uncertainty by more accurate signals, the risk premium in (16) can either increase or decrease depending on the sign of the last term, which consists of belief beta and EIS. In contrast, when we change signal correlation, intertemporal uncertainty remains the same but the risk premium increases with intertemporal co-uncertainty, regardless of the value of EIS.

Comparisons to the Learning Models in the Literature

We now return to the benchmark i.i.d.-learning model of Collin-Dufresne et al. (2016). The model presumes that (1) consumption and dividend growth rates are perfectly correlated, $g_{d,t} = \phi_\mu g_{c,t}$, and (2) realized growth rates coincide with aggregate signals, $\mathbf{s}^t = \mathbf{g}^t$. These assumptions imply restrictions in parameters, such as $\mu_d = \phi_\mu \mu_c$, $\rho_g = \rho_s = 1$, and $\mathbf{\Sigma}_s = \mathbf{\Sigma}_g = \mathbf{J}_{sg} = \mathbf{J}_{gs}$, in our general setup. In particular, they use $\phi_\mu = 1$ and $\psi = 1$; therefore,

the model implies $RP_{ss} = RP_{sg} = 0$, and RP_{gs} reduces to, by applying (6) and $\mathbf{\Sigma}_{\boldsymbol{\mu},t} = \frac{1}{n+t}\mathbf{\Sigma}_s$,

$$RP_{gs} = \kappa_{1c}^* \left[\gamma - \frac{1}{\psi}\right] Var(\mu_c|\mathcal{F}_{t+1}). \tag{20}$$

We caution that Equation (20) can give a somewhat misleading impression that the risk premium generated by learning is, in general, driven by the parameter uncertainty in the mean consumption growth rate μ_c, say, $Var(\mu_c|\mathcal{F}_t)$, which we call *static uncertainty* to distinguish it from static co-uncertainty $Cov(\mu_c, \mu_d|\mathcal{F}_t)$ and intertemporal (co)-uncertainty $Cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t)$. As we mentioned above, in our general setup this static uncertainty has only a minimal quantitative effect through the last term of (15). However, this point is obscured in the literature because seemingly innocuous assumptions make intertemporal (co)-uncertainty—which is generated by the dynamics of beliefs in the learning process—indistinguishable from static (co)-uncertainty—which is based only on a 'snapshot' of beliefs.

Furthermore, the extant literature on the effect of learning on asset prices in general equilibrium suggests that learning from signals—for example, from scheduled macroeconomic announcements (Ai and Bansal 2018)—*always* increases the risk premium when an agent prefers early resolution of uncertainty ($\gamma > \frac{1}{\psi}$). Our general learning setup shows this is not necessarily true. Note the last term in (16) can be negative if EIS is low or belief beta is small. Joint learning, therefore, can increase the risk premium even with preferences for late resolution of uncertainty ($\gamma < \frac{1}{\psi}$).

In summary, our general learning environment highlights the presence of learning-induced risk premium channels that appear to have been ignored in the literature. The analysis indicates that with joint signals on multiple unknown parameters, an asset's risk premium mirrors how agents' posterior beliefs on economic and financial fundamentals covary over time, rather than how the fundamentals themselves are truly correlated or how uncertain they are. The next section quantitatively confirms this argument.

2.3 Numerical Examples

Correlated signals on multiple unknowns affect the risk premium through intertemporal co-uncertainty. To demonstrate this, we compare three different correlation structures for

growth shocks and aggregate signals in the i.i.d. bivariate model at hand:

(i) $Corr(g_{c,t}, g_{d,t}) = \rho_g = 0.55$, $\mathbf{s}^t = \mathbf{g}^t$, and thus, $Corr(s_{c,t}, s_{d,t}) = \rho_s = 0.55$,

(ii) $Corr(g_{c,t}, g_{d,t}) = \rho_g = 0$ and $Corr(s_{c,t}, s_{d,t}) = \rho_s = 0.55$ where $\mathbf{s}^t$ is independent of $\mathbf{g}^t$,

(iii) $Corr(g_{c,t}, g_{d,t}) = \rho_g = 0$ and $Corr(s_{c,t}, s_{d,t}) = \rho_s = 0.80$ where $\mathbf{s}^t$ is independent of $\mathbf{g}^t$,

where the correlation coefficient 0.55 matches the estimate from Bansal and Yaron (2004). The first case (i) emulates the benchmark model in Collin-Dufresne et al. (2016) because the annunciated assumptions imply $\mathbf{\Sigma}_s = \mathbf{\Sigma}_g = \mathbf{J}_{sg} = \mathbf{J}_{gs}$. The only difference is that Collin-Dufresne et al. (2016) additionally assume $g_{d,t} = g_{c,t}$, resulting in $\rho_g = \rho_s = 1$ and $\hat{\mu}_{d,t} = \hat{\mu}_{c,t}$. We find, however, that case (i) inherits all asset pricing implications from their i.i.d. model; hence, case (i) is an appropriate benchmark specification for cases (ii) and (iii). In cases (ii) and (iii), we isolate the main joint learning channel (16) by assuming $\mathbf{s}^t$ and $\mathbf{g}^t$ are independent; that is, $\mathbf{J}_{gs} = 0$. Aggregate signals $\mathbf{s}^t$, however, include information from $\mathbf{g}^t$; therefore, they are possibly correlated. Section 4 finds that such cases can further boost the risk premium with joint learning.

We solve our model at quarterly frequency while calibrating the parameters following Collin-Dufresne et al. (2016) and Bansal and Yaron (2004). The calibration details and computation results are reported in Table 1. For all three cases, static uncertainty in consumption mean growth rate $Var(\mu_c|\mathcal{F}_t)$ and intertemporal co-uncertainty $Cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t)$ gradually dissipate as the agent learns about the true fixed mean-growth-rates $\boldsymbol{\mu}$. Therefore, the annualized conditional (cumulative) equity premium declines over time, consistent with the benchmark model in Collin-Dufresne et al. (2016).

Two new patterns emerge regarding cases (ii) and (iii), however. First, joint learning with correlated aggregate signals can create an equity premium even if the underlying growth rates of consumption and dividends are uncorrelated. For the first 100 years, the annualized average equity premium is 5.88%, 4.83%, and 8.33% for (i)-(iii), respectively. Although this i.i.d. model is not fully realistic, the equity premium generated by joint learning is still comparable to the benchmark case (i). Second, increasing the signal correlation ρ_s leads to a higher equity premium. We note that these patterns are not a coincidence. To see this,

note that the last two terms in Equation (16), by applying (28), can be written

$$Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)\left[\frac{Cov(\hat{\mu}_{c,t+1},\hat{\mu}_{d,t+1}|\mathcal{F}_t)}{Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)}-\frac{1}{\psi}\right]=\frac{\sigma_{sc}\sigma_{sd}}{(n+t)(n+t+1)}\left[\rho_s-\frac{1}{\psi}\frac{\sigma_{sc}}{\sigma_{sd}}\right]. \tag{21}$$

Other things being equal, the risk premium increases with signal correlation ρ_s and can be positive if ρ_s is sufficiently large. Intuitively, highly correlated signals amplify the covariance in agents' point estimates $\hat{\mu}_{c,t+1}$ and $\hat{\mu}_{d,t+1}$. As a result, the risky portfolio becomes more exposed to consumption risk (based on the the representative agents' posterior beliefs), and the asset ceteris paribus requires a higher risk premium.

A comparison of (ii) and (iii) also reveals that the risk premium is mainly driven by intertemporal co-uncertainty and not by static uncertainty. For both cases, static uncertainty is identical in any given quarter t; therefore, static uncertainty cannot explain the equity premium difference between these two cases. On the other hand, intertemporal co-uncertainty increases with signal correlation similar to the increase in the equity premium observed in (ii) and (iii). Thus, by varying the signal correlation, we can clarify the differential impact of static uncertainty versus intertemporal co-uncertainty on the equilibrium risk premium because the signal correlation affects only off-diagonal elements in the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\mu},t}=\frac{1}{n+t}\mathbf{\Sigma}_s$ from (6).

The simplicity of this i.i.d. model, however, is accompanied by many issues. First, learning on fixed parameters results in a gradually declining equity premium with non-stationary belief dynamics. This can generate the misconception that learning-induced equity premium is always decaying and asymptotically unsustainable. Second, it may appear that a simple i.i.d. model with the assumptions $\mu_c=\mu_d$ and $\rho_g=0$ generates the same learning-induced risk premium for an idiosyncratic asset, regardless of the signal correlation ρ_s. Therefore, one might want to conclude that considering multiple unknowns is unnecessary. However, when the parameters are time-varying, the asset is no longer idiosyncratic under such assumptions. Finally, learning on fixed parameters unwittingly distorts the relationship between signal correlation and intertemporal co-uncertainty. We address these issues in Section 3 and 4 through an analysis of joint learning with multivariate unknown time-varying parameters.

3 Co-Uncertainty, Learning, and Correlated Signals

The foregoing analysis illustrates the significant economic implications of joint learning on equilibrium asset prices. We now examine in more generality some fundamental concepts for joint learning—that is, Bayesian learning in generalized information environments—that we developed above. This analysis will be useful for understanding the effects of signal correlation on Bayesian learning on multiple unknowns.

3.1 Static and Intertemporal Covariance of Beliefs

For concise notation, we first define static and intertemporal co-uncertainty matrices. The following definition generally applies to unknown parameters and distributions, and is not restricted to location parameters or an i.i.d. bivariate normal distribution considered in the previous section.

Definition 1. (Static and Intertemporal Beliefs) *If an agent has posterior beliefs* $\hat{\boldsymbol{\mu}}_t = E(\boldsymbol{\mu}|\mathcal{F}_t)$ *on a vector of unknowns* $\boldsymbol{\mu}$ *at time* t *conditional on her information set* $\mathcal{F}_t$*, then static and intertemporal co-uncertainty matrices are defined, respectively, as*

$$\text{Static co-uncertainty matrix: } \boldsymbol{\Sigma}_{\boldsymbol{\mu},t} \triangleq Cov(\boldsymbol{\mu}|\mathcal{F}_t), \tag{22}$$

$$\text{Intertemporal co-uncertainty matrix: } \boldsymbol{\Sigma}_t^{\hat{\boldsymbol{\mu}}_{t+1}} \triangleq Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t), \tag{23}$$

In words, the static co-uncertainty matrix is the contemporaneous conditional variance-covariance matrix of the unknowns and provides a 'snapshot' of posterior beliefs. In contrast, the intertemporal co-uncertainty matrix is the variance-covariance matrix of innovations in the posterior means (or Bayesian point estimates) of the unknowns. Note that $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ includes static uncertainty and co-uncertainty as its diagonal and off-diagonal elements, respectively, whereas $\boldsymbol{\Sigma}_t^{\hat{\boldsymbol{\mu}}_{t+1}}$ does the same with respect to intertemporal uncertainty and co-uncertainty. The following result dynamically links static and intertemporal co-uncertainty matrices.[11]

Proposition 1. (Trade-off Between Static and Intertemporal Beliefs) *In Bayesian learning on a vector of unknowns* $\boldsymbol{\mu}$*, the static and intertemporal co-uncertainty matrices, if*

[11] Equation (24) is a multivariate version of Equation (9) in Fogli and Veldkamp (2011).

they exist, jointly satisfy the following condition for each t:

$$\boldsymbol{\Sigma}_t^{\hat{\boldsymbol{\mu}}_{t+1}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu},t} - E(\boldsymbol{\Sigma}_{\boldsymbol{\mu},t+1}|\mathcal{F}_t), \tag{24}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t} = Cov(\boldsymbol{\mu}|\mathcal{F}_t)$ and $\boldsymbol{\Sigma}_t^{\hat{\boldsymbol{\mu}}_{t+1}} = Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t)$ are the static and intertemporal co-uncertainty matrices at time t, respectively, and $\hat{\boldsymbol{\mu}}_{t+1} = E(\boldsymbol{\mu}|\mathcal{F}_{t+1})$ is the posterior mean of $\boldsymbol{\mu}$ at time $t+1$.

The proof of this proposition follows from noting that $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t+1}$ denote prior and posterior static co-uncertainty matrices, respectively, and hence (24) is equivalent to the law of total covariance:

$$Cov(\boldsymbol{\mu}|\mathcal{F}_t) = Cov(E(\boldsymbol{\mu}|\mathcal{F}_{t+1})|\mathcal{F}_t) + E(Cov(\boldsymbol{\mu}|\mathcal{F}_{t+1})|\mathcal{F}_t) = Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t) + E(\boldsymbol{\Sigma}_{\boldsymbol{\mu},t+1}|\mathcal{F}_t).$$

We note that Equation (24) is not an implication of a stylized asset pricing model; rather, it is a representation of a general mathematical fact: the intertemporal co-uncertainty matrix at time t equals the expected reduction (or negative innovation) in the static co-uncertainty matrix for the next period.[12]

Proposition 1 implies a fundamental trade-off in learning in generalized environments considered in this paper. Suppose an agent faces static (co)-uncertainty regarding multiple unknowns but has some information available about them. Incorporating this information in a Bayesian learning process can reduce the posterior static (co)-uncertainty. However, Equation (24) implies that now the agent faces a new risk of (co)-varying parameter estimates in exchange for reducing the posterior static (co)-uncertainty. In asset pricing contexts, this trade-off can be either beneficial or costly. In some cases, the agent can be worse-off with learning and requires a higher risk premium. In Section 4, we generalize this intuition by considering cases of multivariate parameter uncertainty where learning can either increase or decrease static co-uncertainty depending on signal correlation.

Furthermore, Proposition 1 motivates a useful heuristic on the relative economic impor-

[12] Frankel and Kamenica (2019) show that information and uncertainty are coupled in that the expected reduction in uncertainty always equals the expected amount of information generated if measures of information and uncertainty arise from the same decision problem.

tance of static and intertemporal (co)-uncertainty. Consider a static or one-period version of (24) in which the true values of parameters are revealed at $t+1$. Then the static and intertemporal co-uncertainty covariance matrices become identical, that is, $\mathbf{\Sigma}_t^{\hat{\boldsymbol{\mu}}_{t+1}} = \mathbf{\Sigma}_{\boldsymbol{\mu},t}$ and the model solution presents static (co)-uncertainty as the key variable of interest. But in a dynamic multivariate normal setting—such as the i.i.d bivariate normal model considered in the previous section—the agent's Bellman equation can be written

$$V(W_t, \hat{\boldsymbol{\mu}}_t, \mathbf{\Sigma}_{\boldsymbol{\mu},t}) = \max \left((1-\beta) c_t^{\frac{\psi-1}{\psi}} + \beta \mathcal{R} \left[V(W_{t+1}, \hat{\boldsymbol{\mu}}_{t+1}, \mathbf{\Sigma}_{\boldsymbol{\mu},t+1}) | \mathcal{F}_t \right]^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}}, \tag{25}$$

where the state variables at t are wealth W_t and the first two moments $(\hat{\boldsymbol{\mu}}_t, \mathbf{\Sigma}_{\boldsymbol{\mu},t})$ that represent the posterior distribution of parameters. Intertemporal co-uncertainty resides inside $\mathcal{R}[\cdot|\mathcal{F}_t]$ of Equation (25) and therefore will generally will be an integral component in agents' dynamic optimization and hence equilibrium risk premium. In particular, holding fixed $(W_t, \hat{\boldsymbol{\mu}}_t, \mathbf{\Sigma}_{\boldsymbol{\mu},t})$, a different signal correlation structure can dramatically change the value function through the intertemporal co-uncertainty channel.

3.2 Correlated Signals for Fixed vs. Time-Varying Parameters

A possible intuition that may be advanced to explain the results of the i.i.d. model in Section 2 proceeds along the following lines: There are two parameters, each of which has its own signal; and if these two signals are positively correlated, the Bayes' estimates on the parameters tend to change in the same direction. Somewhat surprisingly, this straightforward intuition is incorrect, at least conditionally. As a counter example, consider the simple bivariate normal model in Equation (2). If static co-uncertainty is zero, i.e., the two parameters are uncorrelated in the agent's posterior beliefs, then from (24) a positive signal correlation implies negative intertemporal co-uncertainty. But if, instead, the posterior static co-uncertainty is positive, then uncorrelated signals generate a positive intertemporal co-uncertainty. These facts are puzzling with respect to the candidate intuition given above, but they are consistent with the more general result below, which is a direct application of Proposition 1.[13]

[13]The proof is on the online Appendix A.2.

Proposition 2. (Sign of Intertemporal Co-Uncertainty) *If the signal and the posterior beliefs on unknown location parameters follow Equations (2) and (5), respectively, then intertemporal co-uncertainty at t is positive, that is, $cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t) > 0$, if and only if*

$$\rho_s < \rho_{\boldsymbol{\mu},t} \left((1 - \rho^2_{\boldsymbol{\mu},t}) + \frac{\sigma^2_{sc}}{\sigma^2_{c,t}} + \frac{\sigma^2_{sd}}{\sigma^2_{d,t}} \right) \frac{\sigma_{c,t}\sigma_{d,t}}{\sigma_{sc}\sigma_{sd}} \frac{1}{1 + \rho^2_{\boldsymbol{\mu},t}}. \tag{26}$$

Note that all the terms in the right-hand side of (26) are positive except the first term $\rho_{\boldsymbol{\mu},t}$. Therefore, roughly speaking, intertemporal co-uncertainty is positive when the signal correlation is lower than a (scaled) prior static co-uncertainty *in terms of* correlation $\rho_{\boldsymbol{\mu},t}$.[14]

To better understand Proposition 2, it is useful to rewrite Proposition 1 for our i.i.d. model by relocating terms.

$$\underbrace{Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t)}_{\text{Intertemporal}} + \underbrace{\boldsymbol{\Sigma}_{\boldsymbol{\mu},t+1}}_{\substack{\text{Posterior} \\ \text{static}}} = \underbrace{\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}}_{\substack{\text{Prior} \\ \text{static}}} = \text{ (Fixed at time } t\text{)}, \tag{27}$$

where $E(\boldsymbol{\Sigma}_{\boldsymbol{\mu},t+1}|\mathcal{F}_t) = \boldsymbol{\Sigma}_{\boldsymbol{\mu},t+1}$ is known at t because of the deterministic evolution of $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$. The precision matrix additivity property, $\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu},t+1} = \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu},t} + \boldsymbol{\Sigma}^{-1}_{s}$ by Bayes' theorem, implies that a higher signal correlation results in higher posterior static co-uncertainty, which lowers intertemporal co-uncertainty since the right-hand-side of (27) is fixed. This trade-off effect is shown graphically in Figure 1. The horizontal axes show the signal correlation; the vertical axes show posterior static co-uncertainty in Panel (a) and intertemporal co-uncertainty in Panel (b), respectively. The three thin black lines correspond to different levels of prior static co-uncertainty (in terms of the correlation): $\rho_{\boldsymbol{\mu},t} = (-0.7, 0, +0.7)$. These graphs show two natural determinants for static and intertemporal co-uncertainties. First, higher signal correlation conditionally increases posterior static co-uncertainty but decreases intertemporal co-uncertainty. Second, holding fixed the signal correlation, higher prior static co-uncertainty conditionally increases both posterior static and intertemporal co-uncertainties.[15]

Propositions 1 and 2, therefore, raise a seeming contradiction to the results in Section 2 which show positive correlations create positive intertemporal co-uncertainty. The resolution,

[14] Signal noisiness also affects intertemporal co-uncertainty. See Figure A.1 in the online Appendix A.3.
[15] The two counter examples mentioned previously combine these two facts.

however, is in the prior specification and non-stationary dynamics of posterior beliefs in the benchmark models in the literature. In particular, the initial prior covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\mu},0}$ is assumed to be proportional to the signal covariance matrix, i.e., $\mathbf{\Sigma}_{\boldsymbol{\mu},0} = n^{-1}\mathbf{\Sigma}_s$ where n is a scalar. This prior specification is common in the literature and seemingly reasonable because prior information is equivalent to observing n signals at $t = 0$ (Collin-Dufresne et al. 2016). However, this seemingly innocuous assumption forces both the posterior static co-uncertainty matrix ($\mathbf{\Sigma}_{\boldsymbol{\mu},t+1}$) and intertemporal co-uncertainty matrix, $Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t)$, to be proportional to the signal covariance matrix $\mathbf{\Sigma}_s$ as follows.

$$Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t) = \mathbf{\Sigma}_{\mu,t} - \mathbf{\Sigma}_{\mu,t+1} = \frac{1}{n+t}\mathbf{\Sigma}_s - \frac{1}{n+t+1}\mathbf{\Sigma}_s = \frac{1}{(n+t)(n+t+1)}\,\mathbf{\Sigma}_s, \quad (28)$$

where $\mathbf{\Sigma}^{-1}_{\boldsymbol{\mu},t} = \mathbf{\Sigma}^{-1}_{\boldsymbol{\mu},t-1} + \mathbf{\Sigma}^{-1}_s = (n+t)\mathbf{\Sigma}^{-1}_s$. This result falsely suggests that positively correlated signals generally produce (both) large static co-uncertainty and large intertemporal co-uncertainty in the i.i.d. model.[16] Note that when we increase signal correlation under this prior specification, we automatically also increase prior static co-uncertainty. Consequently, we observe the patterns in Table 1, consistent with the upward-sloping thick gray lines in Figure 1 instead of thin black lines.

The special aspects of the conventional prior specification are, however, not the main issue. A deeper issue is the non-stationary dynamics of beliefs in i.i.d. models. Even if we properly specify the prior distribution when we change signal correlation, the impact of prior beliefs attenuates over time. The static and intertemporal co-uncertainty matrices, therefore, converge to the one that is proportional to the signal covariance matrix, that is, $T\mathbf{\Sigma}_{\boldsymbol{\mu},T} \to \mathbf{\Sigma}_s$ and $T^2 Cov(\hat{\boldsymbol{\mu}}_{T+1}|\mathcal{F}_T) \to \mathbf{\Sigma}_s$ as $T \to \infty$. Consequently, higher signal correlation will eventually lead to larger intertemporal co-uncertainty although it is small or even negative at the outset.

Instead, if the unknown parameters are time-varying, then innovations in parameter-dynamics allow the posterior beliefs to follow a stationary process where (co)-uncertainty never dissipates through learning and signal correlation only partially affects belief dynamics. As a result, we recover the original conditional relationships implied by Proposition 2 and

[16] Intertemporal co-uncertainty is positive because static co-uncertainty *in terms of covariance* falls over time although static co-uncertainty *in terms of correlation* remains the same.

Figure 1 Panel (b); in particular, lower signal correlation generates larger intertemporal co-uncertainty, which is confirmed in the next section.

4 Joint-Learning on Time-Varying Parameters

4.1 Model Setup

This section analyzes joint learning on time-varying parameters, i.e., latent states, in the well known long-run risk framework of Bansal and Yaron (2004). The dynamics of hidden long-run components $\mathbf{x}_t = [x_{c,t} \quad x_{d,t}]^\top$, aggregate signals $\mathbf{s}_t = [s_{c,t} \quad s_{d,t}]^\top$, and realized consumption and dividend growth shocks $\mathbf{g}_t = [g_{c,t} \quad g_{d,t}]^\top$ are as follows.[17]

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{e}_{x,t+1}, \\ \mathbf{s}_{t+1} &= \mathbf{x}_t + \mathbf{e}_{s,t+1}, \\ \mathbf{g}_{t+1} &= \boldsymbol{\mu} + \mathbf{x}_t + \mathbf{e}_{g,t+1}, \end{aligned} \tag{29}$$

$$\text{where } \mathbf{F} = \begin{bmatrix} F_x & 0 \\ 0 & F_x \end{bmatrix}, \ \mathbf{e}_{x,t+1} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_x), \text{ and } \boldsymbol{\Sigma}_x = \begin{bmatrix} \sigma_{xc}^2 & \rho_x \sigma_{xc} \sigma_{xd} \\ \rho_x \sigma_{xc} \sigma_{xd} & \sigma_{xd}^2 \end{bmatrix}. \tag{30}$$

To isolate the effects of joint learning, we assume that the shocks $\mathbf{e}_{x,t+1}$ are independent of both $\mathbf{e}_{g,t+1}$ and $\mathbf{e}_{s,t+1}$. The conditional mean-growth-rate vector $\mathbf{x}_t$ in (29) is unknown while the unconditional mean-growth-rate vector $\boldsymbol{\mu} = [\mu_c \quad \mu_d]^\top$ is known. The agent's information set is $\mathcal{F}_t = \{\mathbf{g}^t, \mathbf{s}^t\}$ but this is summarized by $\mathcal{F}_t = \{\mathbf{s}^t\}$ because the aggregate signal history $\mathbf{s}^t$ contains all available information, including the realized growth rates history $\mathbf{g}^t$, by definition. Therefore, the short-run growth shocks $\mathbf{e}_{g,t+1} = [e_{gc,t+1} \quad e_{gd,t+1}]^\top$ and the signal noises $\mathbf{e}_{s,t+1} = [e_{sc,t+1} \quad e_{sd,t+1}]^\top$ in aggregate signals are possibly positively correlated

[17] Alternatively, we can model the aggregate signal $\mathbf{s}_{t+1}$ as a forward-looking signal: $\mathbf{s}_{t+1} = \mathbf{x}_{t+1} + \mathbf{e}_{s,t+1}$. This alternative specification does not alter the overall results, however.

as follows:

$$\begin{bmatrix} \mathbf{e}_{g,t+1} \\ \mathbf{e}_{s,t+1} \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left[\mathbf{0}, \mathbf{\Sigma}_{gs}\right], \quad \mathbf{\Sigma}_{gs} = \begin{bmatrix} \mathbf{\Sigma}_g & \mathbf{J}_{gs} \\ \mathbf{J}_{gs}^{\top} & \mathbf{\Sigma}_s \end{bmatrix}, \; \mathbf{\Sigma}_g = \begin{bmatrix} \sigma_c^2 & \rho_g \sigma_c \sigma_d \\ \rho_g \sigma_c \sigma_d & \sigma_d^2 \end{bmatrix},$$
$$\mathbf{\Sigma}_s = \begin{bmatrix} \sigma_{sc}^2 & \rho_s \sigma_{sc} \sigma_{sd} \\ \rho_s \sigma_{sc} \sigma_{sd} & \sigma_{sd}^2 \end{bmatrix}, \text{ and } \mathbf{J}_{gs} = \begin{bmatrix} \rho_{gs} \sigma_{gc} \sigma_{sc} & 0 \\ 0 & \rho_{gs} \sigma_{gd} \sigma_{sd} \end{bmatrix}, \tag{31}$$

where $\rho_g \geq 0$ and $\rho_{gs} = Corr(e_{gc,t}, e_{sc,t}) = Corr(e_{gd,t}, e_{sd,t}) \geq 0$. In the cross-covariance matrix $\mathbf{J}_{gs}$ from (3), we set the cross-correlations $Corr(e_{gd,t}, e_{sc,t})$ and $Corr(e_{gc,t}, e_{sd,t})$ to zero because their signs are ambiguous.

In sum, at each decision point, the representative agent knows all the parameters except $\mathbf{x}^t$ and makes Bayesian inference on it conditional on the history of signals. By solving a Kalman filter, the agent's posterior belief on $\mathbf{x}_t$ at t converge to the stationary process:

$$\mathbf{x}_t | \mathcal{F}_t \sim \mathcal{N}\left(\hat{\mathbf{x}}_t, \mathbf{\Pi}\right) \quad \text{and} \quad \hat{\mathbf{x}}_t = \mathbf{F}\hat{\mathbf{x}}_{t-1} + \mathbf{K}_s(\mathbf{s}_t - \hat{\mathbf{x}}_{t-1}), \tag{32}$$

where $\mathbf{K}_s = \mathbf{F}\mathbf{\Pi}(\mathbf{\Pi} + \mathbf{\Sigma}_s)^{-1}$ is the Kalman gain matrix and $\mathbf{\Pi}$ is the conditional covariance matrix of $\mathbf{x}_t$ under the agent's posterior beliefs after the Kalman filter converges to the steady state. We consider the state variable $\hat{\mathbf{x}}_t$ as *predictive* in that $E[\mathbf{g}_{t+1}|\mathcal{F}_t] = \boldsymbol{\mu} + \hat{\mathbf{x}}_t$ and $E[\mathbf{s}_{t+1}|\mathcal{F}_t] = \hat{\mathbf{x}}_t$.

To examine the effects of joint learning, we generalize consumption and dividend dynamics of the benchmark constant-volatility model in Bansal and Yaron (2004). In our model, the consumption and dividend growth rates have their own separate long-run components $x_{c,t}$ and $x_{d,t}$, respectively, so that the agent learns on multiple unknowns. Nonetheless, our specification is linked to Bansal and Yaron (2004) through the following implied parameters:

$$\phi_{xd} = \frac{Cov(x_{c,t}, x_{d,t}|\mathbf{x}_{t-1})}{Var(x_{c,t}|\mathbf{x}_{t-1})} \quad \text{and} \quad \chi_{xd} = \frac{\sigma_{xd}}{\sigma_{xc}} = \frac{SD(x_{d,t}|\mathbf{x}_{t-1})}{SD(x_{c,t}|\mathbf{x}_{t-1})}, \tag{33}$$

where ϕ_{xd} measures the exposure of the long-run component in dividend growth to the corresponding long-run consumption growth, while χ_{xd} measures the scale of $x_{d,t}$ as a volatility ratio of the innovations in $x_{d,t}$ and $x_{c,t}$. We note that Bansal and Yaron (2004) examine the

special case $\chi_{xd} = \phi_{xd}$ since $x_{d,t} = \phi_{xd} x_{c,t}$ in their specification. In contrast, we explore the implications of multiple unknowns, i.e., $x_{d,t} \neq \phi_{xd} x_{c,t}$, under the condition $\chi_{xd} \geq \phi_{xd}$ when the dividend process has an idiosyncratic persistent component that is independent of consumption. If ρ_x—that is, the correlation between innovations in $x_{c,t}$ and $x_{d,t}$—is zero, then the dividend long-run component is completely idiosyncratic because there is no underlying link to consumption.

Parameter Calibration and Detection Error Probability

We set the baseline parameters at monthly frequency following Bansal and Yaron (2004), as shown in Table 2. Calibrating the parameters in long-run risk models, however, is controversial; the main criticism in the literature targets F_x, the persistence of long-run consumption growth. On top of that, we find new calibration challenges due to unobservable long-run components. Consider, in particular, the parameters ϕ_{xd} and χ_{xd}. Estimating these parameters with high precision can be very difficult because the long-run components in consumption and dividends are unobservable to the agent, as well as econometricians. Rather than focusing on issues relating to estimation methods or data, we present the underlying estimation uncertainty using detection error probability metrics in the context of parameter calibration.

The detection error probability (henceforth, DEP) measures the probability that an econometrician errs when attempting to detect the correct parameter values. Suppose that the econometrician considers two differently calibrated models: an original calibration (O) and an alternative calibration (A). With equal prior weights on (O) and (A), the DEP is the average probability of the two possible calibration error likelihoods:

$$\text{Detection Error Probability} = \frac{\Pr(\text{choose A} \mid \text{O is true}) + \Pr(\text{choose O} \mid \text{A is true})}{2}. \quad (34)$$

In short, DEP is the average of Type-I errors; two parameter sets are indistinguishable by a statistical test if DEP is larger than 5%, or even 10% depending on the significance level.

We report the DEPs associated with ϕ_{xd} and χ_{xd} in the online Appendix A.4. The DEP between $\phi_{xd} = 0$ and $\phi_{xd} = 3$ is higher than 10% with 70 years of monthly data, so is the

DEP between $\chi_{xd} = 3$ and $\chi_{xd} = 6$. The results suggest that the long-run component in dividends may be purely idiosyncratic and larger than is typically assumed. Furthermore, as shown in Jagannathan and Liu (2019), the parameters in the dividends process are not very stable; therefore, the agent faces large uncertainty in dividend dynamics. This high level of uncertainty may translate to high volatility (σ_{xd}) of the idiosyncratic long-run component in dividends under the agent's posterior beliefs. We, therefore, explore the long-run risk model with the values of ϕ_{xd} and χ_{xd} in these ranges.

4.2 Risk Premium in Joint Learning

Model Solutions We solve the model using log-linearization. The solution expressions are, therefore, similar to those of the i.i.d. model in Section 2. In particular, the intertemporal co-uncertainty component in the equity premium is almost identical to Equation (16) because

$$\boldsymbol{\beta}_s \boldsymbol{\Sigma}_s \boldsymbol{\lambda}_s^\top = \kappa_{1c}^{**} \kappa_{1d}^{**} \left(\gamma - \frac{1}{\psi}\right) Var(\hat{x}_{c,t+1}|\mathcal{F}_t) \left[\frac{Cov(\hat{x}_{c,t+1}, \hat{x}_{d,t+1}|\mathcal{F}_t)}{Var(\hat{x}_{c,t+1}|\mathcal{F}_t)} - \frac{1}{\psi}\right], \tag{35}$$

where $\kappa_{1c}^{**} = \kappa_{1c}/(1 - \kappa_{1c}F_x)$ and $\kappa_{1d}^{**} = \kappa_{1d}/(1 - \kappa_{1d}F_x)$. The only notable difference is that the unknown state is now $\mathbf{x}_t$ rather than $\boldsymbol{\mu}$. As suggested by the similarity of the expressions, our long-run risk model inherits the asset pricing implications of intertemporal co-uncertainty from the i.i.d. model in Section 2. In particular, intertemporal co-uncertainty drives the risk premium in a recursive utility framework. The main difference, however, is that now negatively correlated signals—rather than positively correlated signals—generate large intertemporal co-uncertainty and a high risk premium, as explained in Section 3.2. We summarize the main results in Table 2 and include the solution details in the online Appendix A.5

Learning Solely from Realized Growth Rates Panel A of Table 2 shows the benchmark cases in which an agent learns only from realized growth rates. This specification has multiple unknowns $x_{c,t}$ and $x_{d,t}$, yet it is still a special case of learning with the restrictions. Most importantly, the agent has no access to signals other than realized growth rates, that is, $\mathbf{s}_t = \mathbf{g}_t$; therefore, aggregate signal correlation ρ_s always equals the correlation ρ_g between

short-run consumption and dividend growth shocks. With this strong, and arguably unrealistic, restriction on the information environment, the effect of learning on the risk premium is negligible, although it exists. Rows #2 to #4 show that the equity premium is directly linked to ϕ_{xd}, which is the intrinsic exposure of dividends to the consumption long-run component. The case of no exposure $\phi_{xd} = 0$ in row #4 results in absence of intertemporal co-uncertainty and risk premium.[18] Also, from the rows #1 and #2, we confirm that idiosyncratic dividend long-run risk has an insignificant effect in this conventional learning environment.

Signal Correlation Table 2 Panel B shows cases where the agent learns from aggregate signals, as in Equation (29). Given the other parameter values, decreasing signal correlation ρ_s increases intertemporal co-uncertainty and risk premium, as shown in rows #5 to #7 and in Figure 2 Panel (a). When $\rho_s = +0.8$ in row #5, intertemporal co-uncertainty is negative and the equity premium is -3.40% per year even though the agent knows that dividends are intrinsically exposed to long-run consumption risk (note that $\phi_{xd} = 1$). In essence, positively correlated signals simply dominate the effects of the intrinsic exposure to consumption and distort the risk premium. In contrast, when $\rho_s = -0.9$ in row #7, intertemporal co-uncertainty is positive, and the equity premium is $+6.58\%$ per year. Furthermore, given ρ_{gs} in Panel B, the equity premium follows the magnitude of intertemporal co-uncertainty (in the last column in Table 2) rather than follows static uncertainty in the consumption long-run component. These results confirm the analysis in Section 3.2.

Systematic and Idiosyncratic Long-Run Dividend Risks Proposition 1 implies that not only the signal correlation but also prior static co-uncertainty can affect intertemporal co-uncertainty and hence the risk premium. Considering Figure 1 Panel (b) along with Figure 2 Panel (a) confirms this implication. In particular, the three lines (1)–(3) in Figure 2 Panel (a) show that risk premium decreases with signal correlation but increases with ϕ_{xd}, namely, the parameter that governs prior static co-uncertainty of $\mathbf{x}_t$ in Kalman filter recursions. Figure 1 Panel (b) shows the same patterns with intertemporal co-uncertainty instead. Rows #8 and #9 in Table 2 also show consistent results.[19] On the other hand, comparing rows #7 and

[18] The risk premium is even negative due to the difference between net and log returns.

[19] See the online Appendix A.5.6 for more extensive comparisons.

#8 in Table 2 indicates that the idiosyncratic long-run component in dividends significantly contributes to the equity premium. That is, equity premium increases with χ_{xd} while ϕ_{xd} is fixed. The intuition here is straightforward: Higher idiosyncratic volatility in the long-run component $x_{d,t}$ of dividends implies higher total volatility and, hence, greater intertemporal co-uncertainty (in terms of covariance) and a higher risk premium, other things held fixed.

Growth-to-Signal Correlation Unlike the i.i.d. model in Section 2.3, our long-run risk model allows a positive correlation ρ_{gs} between short-run growth shocks and noisy components of aggregate signals. Since the aggregate signals are the source of endogenous long-run risk, a financial asset requires a high risk premium for correlations to the aggregate signals. Rows #9 and #10 in Table 2 show the substantial effects of ρ_{gs} on the risk premium components in (17) and (18). However, the role of ρ_{gs} is only to multiplicatively scale up the equity premium rather than add to it. Roughly speaking, signal correlation ρ_s sets the sign of the equity premium whereas the growth-to-signal correlation ρ_{gs} amplifies the equity premium (either positive or negative). To see this, compare rows #5 and #12 in Table 2 or the two lines (4) and (5) in Figure 2 Panel (a). Nevertheless, ρ_{gs} does not affect intertemporal co-uncertainty since the agent still learns only from the aggregate signals (as seen by comparing rows #9 and #10). The effects of ρ_{gs}, therefore, differ from the main joint learning channel in Equation (16). Finally, even without a risk premium boost (from $\rho_{gs} = 0$) or a fundamental link ($\phi_{xd} = 0$), we find that negatively correlated signals can create a risk premium as large as 4% per year (row #11).

Preference Parameters In our joint learning model, the equity premium increases with the risk aversion γ and EIS ψ, as in Bansal and Yaron (2004), when the last terms in Equations (16)–(18) are positive. In Table 3, we investigate if our joint learning model can generate the observed equity premium even with lower risk aversion and EIS than those commonly used in the literature. We calibrate other parameters realistically (see Table 3 for details). With low risk aversion $\gamma = 5$, our model generates annual equity premium of 5.33% and the risk-free rate of 1.12% at $\psi = 1.5$. Also, unlike other models in the literature, EIS larger than unity is not a necessary condition to generate a high equity premium in

our model. Note that the last term in Equation (35) can be still positive and large even with low EIS only if intertemporal co-uncertainty is large enough to compensate low EIS. With low EIS $\psi = 0.9$, the model can generate the annual equity premium of 6.43% and the risk-free rate of 1.68% at $\gamma = 7.5$. These results are useful because there is no consensus in the literature on whether the EIS is less or more than 1. For comparison, we report three benchmark cases: 1) zero signal correlation, 2) learning only from growth rates, and 3) known $\mathbf{x}_t$ through signals without noises. In all three cases, the equity premium is around 3.50% even with high risk aversion $\gamma = 10$ and high EIS $\psi = 1.5$, while our joint learning model produces 9.09% for the same parameter values.[20]

4.3 Information Quality and Risk Premium

The relationship between information quality and equity premium depends on the definition of signals in the model. First, we can define signals, following our joint learning model, as *aggregate* signals that include information from growth rates. In this case, an agent learns only from aggregate signals. Alternatively, following the conventional definition in the literature, we can introduce *additional* signals on top of growth rates. In this case, the agent learns from both the additional signals and growth rates although the agent acquires the same information as in the first case. To simplify comparison of model solutions with different definitions of signals, we fix the preference parameters at $\gamma = 10$ and $\psi = 1.5$ following Bansal and Yaron (2004).

Benchmark Cases The first column of Table 4 Panel A shows a benchmark case with *aggregate* signals when the consumption and dividend long-run components are tightly linked and covary together ($x_{d,t}/x_{c,t} = \phi_{xd} = \chi_{xd}$). The equity premium increases with information quality of *aggregate* signals because the agent's beliefs fluctuate more if signals are more accurate. This is consistent with the conditional equity premium results in Figure 3 of Brevik and d'Addona (2010). By contrast, the first three columns of Table 4 Panel B show that the equity premium decreases with information quality of *additional* signals because accurate

[20] When risk aversion is high, the average price-dividend ratio is low, and so is κ_{1d}. Therefore, higher risk aversion generates lower equity volatility in contrast to Bansal and Yaron (2004). We note that this is not an artifact of the Taylor approximation.

additional signals suppress the effects of *learning from growth rates*; this is consistent with the unconditional equity premium results in Brevik and d'Addona (2010).

Quality of Aggregate Signals The next four columns of Table 4 Panel A and Figure 2 Panel (b) show that the equity premium is non-monotonically associated with information quality of *aggregate* signals when the consumption and dividend long-run components are loosely linked ($\phi_{xd} < \chi_{xd}$).[21] In this situation, the joint learning channel (Equation 16) generates a sizable risk premium, which shrinks, however, if aggregate signals are too noisy or too accurate. It is because extremely noisy signals do not generate significant intertemporal co-uncertainty whereas there is essentially no learning in the presence of highly accurate signals. Given $(\phi_{xd}, \rho_{gs}, \rho_s)$, the last five columns of Table 4 Panel A confirm that intertemporal co-uncertainty, and not necessarily information quality, determines the equity premium.

Quality of additional Signals The three columns in the middle of Table 4 Panel B show that the equity premium can also be non-monotonically associated with information quality of negatively correlated *additional* signals if $\phi_{xd} < \chi_{xd}$. This non-monotonicity arises due to a confounding effect: changes in precision of correlated *additional* signals also change the correlation of implied *aggregate* signals—which by definition combine the additional signals and growth rates—as shown in the last three columns of Table 4 Panel B. This confounding effect, thus, distorts the risk premium generated from the joint learning channel (Equation 16), which is governed by the correlation of aggregate signals. As a result, the total equity premium has a non-monotonic relationship with information quality of *additional* signals.[22]

5 Empirical Aspects of Joint Learning

The foregoing analysis has highlighted the novel implications of joint learning in asset pricing contexts. The theoretical aspects of joint learning can be intriguing, but are they empirically

[21] In other words, when idiosyncratic long-run dividend risk exists.

[22] Power utility (Veronesi 2000) in our setup implies $RP_{ss} = RP_{gs} = 0$. Also, the term RP_{sg}, with an additional assumption of $C_t = D_t$, further reduces to $RP_{sg} = \boldsymbol{\beta}_s \mathbf{J}^*_{sg} \boldsymbol{\lambda}_g^\top = \kappa^{**}_{1d} \gamma (1-\gamma) Cov(\hat{x}_{c,t+1}, e_{gc,t+1} | \mathcal{F}_t)$. This expression explains two implications of Veronesi (2000): 1) the equity premium increases with information quality of additional signals if $\gamma > 1$ and 2) the equity premium is non-monotonically related to risk aversion.

relevant? In this Section, we address two possible concerns regarding joint learning. First, is joint learning a general characteristic of learning by economic agents? Second, economic and financial signals are typically presumed to be positively correlated and hence joint learning appears to exacerbate, rather than explains, the equity premium puzzle in the long-run risk framework. We, therefore, discuss and provide some empirical support for the prevalence of negatively correlated signals.

5.1 Joint Signals and Joint Learning: A General Perspective

We argue that learning through correlated signals is not a special case of learning in economic environments; rather, its absence, which we call *independent* learning should be a specialized or rare occurrence. We support this argument by providing a general characterization of joint signals and joint learning.

Definition 2. (Joint Signals and Joint Learning) *Suppose* $\mathbf{s} = [s_1, s_2, ..., s_q]^\top$ *is a* $(q \times 1)$ *signal vector on a* $(m \times 1)$ *vector of parameters* $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_m]^\top$ *where* $q \geq 1$ *and* $m \geq 2$, $p(\cdot)$ *is a subjective probability measure and* $\mathbf{c} = \{c_1, c_2, ..., c_m\}$ *is an exact cover of* $\mathbf{s}$. *That is, the subsets* $c_1, c_2, ..., c_m$ *are pairwise disjoint and their union equals* $\mathbf{s}$. *Then the signal* $\mathbf{s}$ *is a joint signal with respect to* $\boldsymbol{\mu}$ *under the beliefs* $p(\cdot)$ *if there does not exist any* $\mathbf{c}$ *that satisfies*

$$p(\boldsymbol{\mu}|\mathbf{s}) = \prod_{i=1}^{m} p(\mu_i|c_i). \tag{36}$$

Joint learning is then defined as any learning associated with joint signals. However, learning is defined as independent if the condition (36) is satisfied.

Thus, in independent learning, an agent's learning on each unknown parameter is a separate process (from independent signals) such that there are no interaction effects between learning on different parameters. An example of independent learning on multiple parameters is when signals are mutually independent and so are the parameters in prior beliefs. But in practice such cases of completely independent learning would appear to occur only in specialized or rare circumstances. To fix ideas, consider learning on the economic effects of exogenous events such as the COVID-19 pandemic that affects almost every part of the

global economy and financial markets. Therefore, received signals are unlikely to be mutually independent. But even if the signals are independent, the unknown parameters governing effects on employment, inflation, corporate earnings, etc. would generally be interconnected in plausible representations of agents' prior beliefs. Therefore, independent learning, rather than joint learning, appears contrived and not generally empirically plausible. A useful econometric heuristic is that having two independent learning processes in a model is similar to running a separate regression for each of two related dependent variables with disjoint sets of explanatory variables that are independent of each other.

We also point out that transforming joint signals to independent signals through re-parametrization does not eliminate joint learning. For example, define $\mathbf{s}^*_{t+1} = \Sigma_s^{-\frac{1}{2}}\mathbf{s}_{t+1}$ and $\boldsymbol{\mu}^* = \Sigma_s^{-\frac{1}{2}}\boldsymbol{\mu}$ in Equation (2) and then specify a prior distribution of $\boldsymbol{\mu}^*$ as an uncorrelated bivariate normal distribution. Now, there is no joint learning with respect to $\boldsymbol{\mu}^*$, but there is still joint learning with respect to $\boldsymbol{\mu}$ in that $E[g_{c,t+1}|\mathcal{F}_t]$ and $E[g_{d,t+1}|\mathcal{F}_t]$ covary generating the learning-induced risk premium. Re-parametrization, therefore, has no effect on joint learning other than the use of different notation in the model solutions.

5.2 Evidence from Survey Data

Empirical evidence on correlated signals can be indirectly measured from survey data. For example, we collect the time-series of median analyst forecasts for GDP growth $g_{c,t}$ and corporate profit growth $g_{d,t}$ for 204 quarterly periods from 1968:Q4 to 2019:Q3 in the Survey of Professional Forecasters (SPF) from Federal Reserve Bank of Philadelphia.[23] We focus on consensus forecasts, because the asset pricing models considered in this paper are all

[23]GDP growth is computed from a SPF variable NGDP, which changes from quarterly levels of nominal GNP to nominal GDP (seasonally adjusted, annualized, billions of dollars) during 1992. Thus, we drop observations in 1992, which are potentially contaminated by forecasters transition between the two data regimes. Corporate profit growth is computed from CPROF, which consists of forecasts for quarterly nominal corporate profits after tax, excluding IVA and CCAdj (seasonally adjusted, annualized, in billions of dollars). Beginning with the survey of 2006:Q1, this variable includes IVA and CCAdj. Thus, we drop observations in 2006.

representative-agent models. For estimation, we use the following identity:

$$\begin{aligned}\boldsymbol{\Sigma}^{(g)}_{t|t-h} &\triangleq Cov(g_{c,t}, g_{d,t}|\mathcal{F}_{t-h}) = Cov(g_{c,t} - E[g_{c,t}|\mathcal{F}_{t-h}], g_{d,t} - E[g_{d,t}|\mathcal{F}_{t-h}]) \\ &= Cov(FE^{(gc)}_{t|t-h}, FE^{(gd)}_{t|t-h}),\end{aligned} \tag{37}$$

where $E[g_{c,t}|\mathcal{F}_{t-h}]$ and $E[g_{d,t}|\mathcal{F}_{t-h}]$ are the h-period ahead SPF median forecasts for $g_{c,t}$ and $g_{d,t}$, respectively. Therefore, $FE^{(gc)}_{t|t-h}$ and $FE^{(gd)}_{t|t-h}$ are the resulting forecast errors for GDP and corporate profits in SPF. We then use the ex-post sample covariance matrix of the forecast errors as an estimate for $Cov(FE^{(gc)}_{t|t-h}, FE^{(gd)}_{t|t-h})$ and the static co-uncertainty matrix $\boldsymbol{\Sigma}^{(g)}_{t|t-h}$ from $h = 1$ to $h = 4$ quarters. Next, we infer the implied signal precision matrix $\{\boldsymbol{\Sigma}^{(s)}_{(t-h):(t-h')}\}^{-1}$ by inserting estimates for $\boldsymbol{\Sigma}^{(g)}_{t|t-h'}$ and $\boldsymbol{\Sigma}^{(g)}_{t|t-h}$ in the Bayesian updating equation for bivariate normal distributions:

$$\{\boldsymbol{\Sigma}^{(g)}_{t|t-h'}\}^{-1} = \{\boldsymbol{\Sigma}^{(g)}_{t|t-h}\}^{-1} + \{\boldsymbol{\Sigma}^{(s)}_{(t-h):(t-h')}\}^{-1} \quad \text{for} \quad h > h'. \tag{38}$$

Our indirect estimation approach requires bivariate normality assumptions on signals and prior/posterior beliefs. Therefore, the resulting signal covariance matrix estimate is not guaranteed to be positive semidefinite when this assumption is violated. The following generalizes the definition of signal correlation to handle such cases.

Definition 3. (Generalized Signal Correlation Coefficient) *Suppose* $\{\boldsymbol{\Sigma}^{(s)}_{(t-h):(t-h')}\}^{-1}$ *denotes the precision matrix for the signal between time* $t-h$ *and* $t-h'$ *where* $h > h'$*. Then generalized signal correlation coefficient* ρ^*_s *is defined as follows.*

$$\rho^*_s = \frac{\pi_{cd}}{\sqrt{\pi_{cc}\pi_{dd}}} \quad \textit{where} \quad \{\boldsymbol{\Sigma}^{(s)}_{(t-h):(t-h')}\}^{-1} = \begin{bmatrix} \pi_{cc} & -\pi_{cd} \\ -\pi_{cd} & \pi_{dd} \end{bmatrix}. \tag{39}$$

Note that this generalized signal correlation definition ρ^*_s coincides with ρ_s in Equation (2), (31), or (40) if the signal covariance matrix is positive semi-definite. If not, the generalized signal correlation can be lower than -1 or higher than $+1$. The intuition behind this generalization is as follows. Signal correlation changes static co-uncertainty (off-diagonal elements) in the posterior static co-uncertainty matrix; and the lower is the signal correlation,

the lower is the posterior static co-uncertainty as seen in Figure 1 Panel (a). Therefore, generalized signal correlation is interpreted to be lower than -1 if the decrease in posterior static co-uncertainty is greater than the maximum decrease under the bivariate normality assumption.[24]

We find two patterns in the survey data. First, uncertainty regarding GDP and corporate profit growth gradually decrease as the forecasting horizons diminish, as seen in the third and forth columns of Table 5 Panel A. This can be evidence that forecasters learn from informative signals. Second, static co-uncertainty, in terms of both covariance and correlation, falls over time, as shown in the first and second columns of Panel A. This reduction is statistically significant (see the first and second columns in Panel B). These results imply that signal correlation is possibly negative according to Proposition 1. Indeed, our estimation shows that generalized signal correlation is highly negative, as shown in the third column in Panel B. The point estimates on generalized signal correlation are all less than -1 although the bootstrap-based confidence intervals are rather wide.[25]

5.3 Implications of Negatively Correlated Signals

Although the notion of negatively correlated signals may appear somewhat counter-intuitive, we argue that such signals are ubiquitous and have a natural interpretation. Consider the following examples of joint signals. An agent learns about an unknown parameter vector $\boldsymbol{\mu} = [\mu_A \;\; \mu_B]^\top$ through a multivariate normal signal $\mathbf{s}_t \sim \mathcal{N}(\mathbf{H}\boldsymbol{\mu}, \boldsymbol{\Sigma}_s)$. If a left inverse $\mathbf{H}^{-1} = (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top$ exists, the signal can be transformed by re-parametrization to a bivariate signal $\mathbf{s}_t^* \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_s^*)$ where $\mathbf{s}_t^* = \mathbf{H}^{-1}\mathbf{s}_t$ and $\boldsymbol{\Sigma}_s^* = \mathbf{H}^{-1}\boldsymbol{\Sigma}_s(\mathbf{H}^{-1})^\top$ as follows.

$$\text{Correlated signal:} \quad \mathbf{s}_t^* \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}_s^*\right) \quad \text{where} \quad \boldsymbol{\Sigma}_s^* = \sigma_s^2 \begin{bmatrix} \nu & \rho_s \\ \rho_s & \nu^{-1} \end{bmatrix}. \tag{40}$$

[24] Likewise, the generalized signal correlation is higher than $+1$ if the increase in posterior static co-uncertainty is greater than the maximum increase under the bivariate normality assumption.

[25] We follow Efron (1987) to compute bias-corrected and accelerated (BCa) bootstrap confidence intervals. It corrects a bias and skewness in bootstrap distribution.

The only case when $\mathbf{H}^{-1}$ does not exist is when $\mathbf{H}$ has a unit rank. Similarly, this case can be reduced by re-parameterization to a univariate signal as follows.

$$\text{Combined signal:} \quad s_t \sim \mathcal{N}\left(\mathbf{H}\boldsymbol{\mu}, \sigma_s^2\right) \quad \text{where} \quad \mathbf{H} = \begin{bmatrix} h_A & h_B \end{bmatrix}, \tag{41}$$

which we call a *combined* signal in that s_t has information on the sum $h_A\mu_A + h_B\mu_B$.

Combined Signals We generally expect signals received by multiple agents to be positively correlated *across the agents* because of common information sources such as mass media, government agencies, and social networks. However, is it reasonable to expect that an individual agent who learns on multiple unknowns will generally receive positively correlated signals *across the unknowns*? Take, for example, news about high future earnings growth for a cloud computing firm 'A'. Investors often interpret it as good news for other similar firms since their earnings are positively correlated (historically) or the news can also reveal the common mean-growth-rate in the cloud computing industry. But note that this type of earnings news is a combined signal. Revealing the common industry mean growth rate implies a signal structure such as $s_t = w\mu_A + (1-w)\mu_B + e_t$; where $0 < w < 1$, μ_A is the mean earnings growth rate for the firm 'A,' μ_B is the mean growth rate for other firms in the cloud computing industry, and e_t follows a univariate normal distribution with mean zero. Thus, s_t is a combined signal with $\mathbf{H} = [w \quad (1-w)]$ in (41). This kind of combined signal structure is widely prevalent: from signals on local market growth for specific products to signals on the change in average temperatures across the globe. The prevalence of combined signals is in fact not surprising since they can be more cost-effective compared with collecting signals on individual components and aggregating them (Simon 1955).

Negative Signal Correlation and Combined Signals The earnings news in the example above is, indeed, a joint signal that produces positive intertemporal co-uncertainty. First, historical positive correlation in earnings generates positive prior static co-uncertainty. Second, the combined signal corresponds to negatively correlated signals. Therefore, intertemporal co-uncertainty is positive by Proposition 2. A simple calculation shows why the combined signal corresponds to negatively correlated signals. Imagine an agent who

wants to learn about two parameters μ_A and μ_B. She conducts some research and receives a combined signal $s_{1,t} = \mu_A + \mu_B + e_{1,t}$. Then, she repeats a similar, but independent, research exercise to receive a second signal $s_{2,t} = \mu_A + .99\mu_B + e_{2,t}$ where $e_{1,t}$ and $e_{2,t}$ are independently drawn from a normal distribution with mean zero. These two signals together correspond to a bivariate joint signal $\mathbf{s}_t \sim \mathcal{N}(\mathbf{H}\boldsymbol{\mu}, \boldsymbol{\Sigma}_s)$ and transform into $\mathbf{s}_t^*$ in (40) where the signal correlation is highly negative $\rho_s = -.999987$. On the other hand, a negatively correlated bivariate signal can be transformed into two independent combined signals. For example, consider the following example where $\mathbf{s}_t^*$ is a negatively correlated bivariate signal ($\rho_s = -.994475$) in (40).

$$\mathbf{s}_t^* = \boldsymbol{\mu} + \begin{bmatrix} 1 & -.9 \\ -.9 & 1 \end{bmatrix} \mathbf{e}_{s,t}, \Longleftrightarrow \mathbf{s}_t = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix} \mathbf{s}_t^* = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix} \boldsymbol{\mu} + 0.19\ \mathbf{e}_{s,t}, \tag{42}$$

where $\mathbf{e}_{s,t}$ is a 2×1 vector of a random variables from a standard bivariate normal distribution with zero correlation. Again, the signals in the right hand side of (42) are two independent combined signals on $(\mu_A + .9\mu_B)$ and $(.9\mu_A + \mu_B)$, respectively.[26] Note signal noisiness σ_s is seven times larger in $\mathbf{s}_t^*$ than in $\mathbf{s}_t$.[27] Roughly speaking, two combined signals of this kind are equivalent to two highly negatively correlated signals of low precision, consistent with our calibration for noisiness and correlation in aggregate signals.

Combined Signals on Consumption and Dividend Growth We can provide additional arguments for the prevalence of negatively correlated signals. Take, for example, two independent combined signals on $(\mu_A + 0.0001\mu_B)$ and $(\mu_A + 0.000099\mu_B)$. They look like two independent signals on μ_A with a minute 'contamination' by μ_B, yet they imply two highly negatively correlated signals on μ_A and μ_B with exactly the same correlation

[26] We can transform negatively correlated signals in many different ways. For example, Cholesky decomposition produces one highly accurate combined signal and one that is a noisy individual signal on one unknown. Alternatively, by spectral decomposition, we have two independent combined signals: one highly accurate signal on $(\mu_c + \mu_d)$ and the other—quite noisy—signal on $(\mu_c - \mu_d)$. The latter signal becomes completely uninformative as the original signal correlation goes to -1. Therefore, our conclusion remains unchanged, regardless of how we transform signals.

[27] $\sqrt{1 + (-0.9)^2}/0.19 \approx 7.08$

coefficient -0.999987 as in the previous example.[28] Similarly, consider the mean consumption growth rate that can be decomposed to mean labor and dividend growth rates in an endowment economy: $x_{c,t} = 0.9x_{l,t} + 0.1x_{d,t}$. An investor can look for a signal on consumption growth but, instead, directs attention towards dividends because she can directly trade financial assets but not human capital, say, $s_t = 0.9x_{l,t} + 1.1x_{d,t} + e_t$. Then, this signal can be seen as a combined signal on consumption and dividend growth rates since $s_t = 0.9x_{l,t} + 0.1x_{d,t} + e_t = x_{l,t} + x_{d,t} + e_t$, which implies negative signal correlations in (40).

6 Conclusion

Economic agents generally confront uncertainty on multiple and possibly time-varying structural parameters that determine consumption and financial payoffs. Furthermore, in practice, agents develop information or generate signals to infer or learn these parameters. We theoretically and empirically find that dynamic learning of multivariate time-varying structural parameters with noisy signals generates endogenous long-run risks resulting in large and never-decaying equity risk premium. Moreover, the correlation structure of signals plays a crucial role in the economic effects of learning. Apart from understanding risk premium dynamics, the interplay of multivariate unknowns and aggregate signals, such as macroeconomic news, can also potentially explain apparent mispricing and high risk premium of new assets without intrinsic links to aggregate consumption (such as crypto-currencies like Bitcoin).

In the context at hand, we study only the first moment of the unknowns, but these definitions can be generalized to higher moments. For example, the posterior variance, skewness, or kurtosis of an individual unknown can covary with those of other unknowns over time resulting in higher equity premium due to joint learning. We note that the joint learning in Definition 2 pertains to the whole posterior distribution and not just posterior mean.

While we focus on the log-linear approximation for tractability and comparisons with

[28] This is not a coincidence. The signal correlation ρ_s in (40) does not change as long as each column vector in $\mathbf{H}$ remains as its scalar multiple.

the literature, we may also consider higher posterior moments of Bayes' estimates $\hat{\boldsymbol{\mu}}_t$. Intertemporal co-uncertainty, however, will still play a major role with joint learning since the risk premium, by definition, comes from covariance. Considering higher moments of $\hat{\boldsymbol{\mu}}_t$ can potentially explain higher moments and other properties of asset returns.

References

Ai, Hengjie, 2010, Information quality and long-run risk: Asset pricing implications, *Journal of Finance* 65, 1333–1367.

Ai, Hengjie, and Ravi Bansal, 2018, Risk preferences and the macroeconomic announcement premium, *Econometrica* 86, 1383–1430.

Bakshi, Gurdip, and Georgios Skoulakis, 2010, Do subjective expectations explain asset pricing puzzles?, *Journal of Financial Economics* 98, 462–477.

Bansal, Ravi, and Ivan Shaliastovich, 2010, Confidence risk and asset prices, *American Economic Review* 100, 537–41.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Bollerslev, Tim, Robert F Engle, and Jeffrey M Wooldridge, 1988, A capital asset pricing model with time-varying covariances, *Journal of political Economy* 96, 116–131.

Brevik, Frode, and Stefano d'Addona, 2010, Information quality and stock returns revisited, *Journal of Financial and Quantitative Analysis* 1419–1446.

Campbell, John Y, and Robert J Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* 1, 195–228.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A Lochstoer, 2016, Parameter learning in general equilibrium: The asset pricing implications, *American Economic Review* 106, 664–98.

Efron, Bradley, 1987, Better bootstrap confidence intervals, *Journal of the American statistical Association* 82, 171–185.

Engle, Robert F, David M Lilien, and Russell P Robins, 1987, Estimating time varying risk premia in the term structure: The arch-m model, *Econometrica* 391–407.

Epstein, Larry G, and Stanley E Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 937–969.

Fogli, Alessandra, and Laura Veldkamp, 2011, Nature or nurture? learning and the geography of female labor force participation, *Econometrica* 79, 1103–1138.

Frankel, Alexander, and Emir Kamenica, 2019, Quantifying information and uncertainty, *American Economic Review* 109, 3650–80.

Geweke, John, 2001, A note on some limitations of CRRA utility, *Economics letters* 71, 341–345.

Ghosh, Anisha, and George M Constantinides, 2017, What information drives asset prices? Working Paper.

Gvozdeva, Evgenia, Praveen Kumar, and James Yae, 2015, Unknown consumption and financial risks and asset pricing puzzles, *University of Houston,* Working Paper.

Hansen, Lars Peter, 2007, Beliefs, doubts and learning: Valuing economic risk, *American Economic Review* 97.

Harbaugh, Rick, John Maxwell, and Kelly Shue, 2016, Consistent good news and inconsistent bad news, *Wokring Paper* .

Jagannathan, Ravi, and Binying Liu, 2019, Dividend dynamics, learning, and expected stock index returns, *The Journal of Finance* 74, 401–448.

Johannes, Michael, Lars A Lochstoer, and Yiqun Mou, 2016, Learning about consumption dynamics, *The Journal of finance* 71, 551–600.

Kumar, Praveen, Sorin M Sorescu, Rodney D Boehme, and Bartley R Danielsen, 2008, Estimation risk, information, and the conditional capm: Theory and evidence, *The Review of Financial Studies* 21, 1037–1075.

Kurmann, André, and Christopher Otrok, 2013, News shocks and the slope of the term structure of interest rates, *American Economic Review* 103, 2612–32.

Li, Jun, and Harold H Zhang, 2017, Short-run and long-run consumption risks, dividend processes, and asset returns, *The Review of Financial Studies* 30, 588–630.

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica* 867–887.

Simon, Herbert A, 1955, A behavioral model of rational choice, *The quarterly journal of economics* 69, 99–118.

Veronesi, Pietro, 2000, How does information quality affect stock returns?, *Journal of Finance* 55, 807–837.

Weitzman, Martin L, 2007, Subjective expectations and asset-return puzzles, *The American Economic Review* 1102–1130.

Figure 1: Signal Correlations vs. Static and Intertemporal Co-Uncertainties

This figure shows how signal correlation ρ_s affects static and intertemporal co-uncertainties at a given time in a bivariate normal setting in Equation (2) and (5). The other parameters for priors are fixed at $\sigma_{sc} = \sigma_{sd} = 2$ and $\sigma_{c,t} = \sigma_{d,t} = 1$. Panel (a) plots posterior static co-uncertainty against signal correlations at three different levels of prior static co-uncertainty, *in terms of correlation* (-0.7, 0, +0.7). Panel (b) repeats Panel (a) but plots intertemporal co-uncertainty instead. The gray points and lines refer to the cases where signal correlation and prior static co-uncertainty—in terms of correlation—are identical by construction.

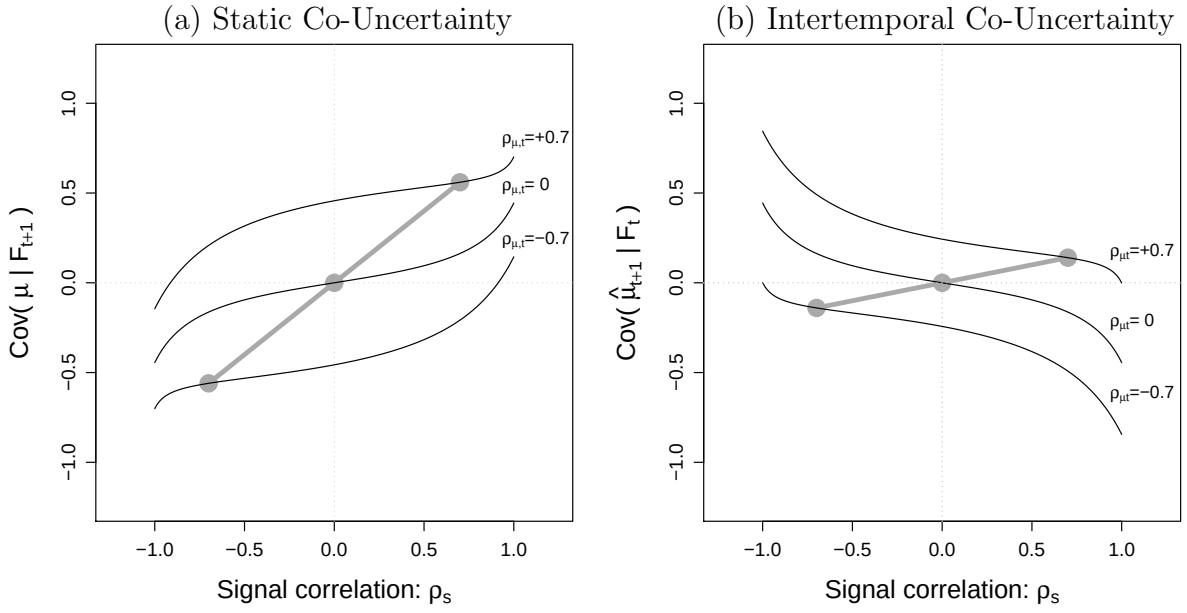


Figure 2: Signal Correlations, Information Quality, and Equity Premium

This figure shows the determinants of the equity premium in our long-run risk models in which the baseline parameters are calibrated following Table 2. Panel (a) shows the equity premium versus signal correlation at different levels of $(\rho_{gs}, \chi_{xd}, \phi_{xd})$ where the correlation between consumption and dividend growth shocks is set as $\rho_g = 0.3$. Panel (b) shows the equity premium versus signal noisiness $\phi_s^{(g)}$, which is a scaled version of the overall signal noisiness σ_s (from Equation 40) of the *aggregate* signals $\mathbf{s}_t$. At $\phi_s^{(g)} = 1$, static co-uncertainty $Var(x_{c,t}|\mathcal{F}_t)$ equals that of the benchmark case in which the growth rates are the aggregate signal $\mathbf{s}_t = \mathbf{g}_t$. Other parameters in Panel (b) are set as follows: the long-run exposure $\phi_{xd} = 1$, dividend long-run component volatility ratio $\chi_{xd} = 4.5$, and growth-to-signal correlation $\rho_{gs} = 0$ if not stated otherwise in the plot.

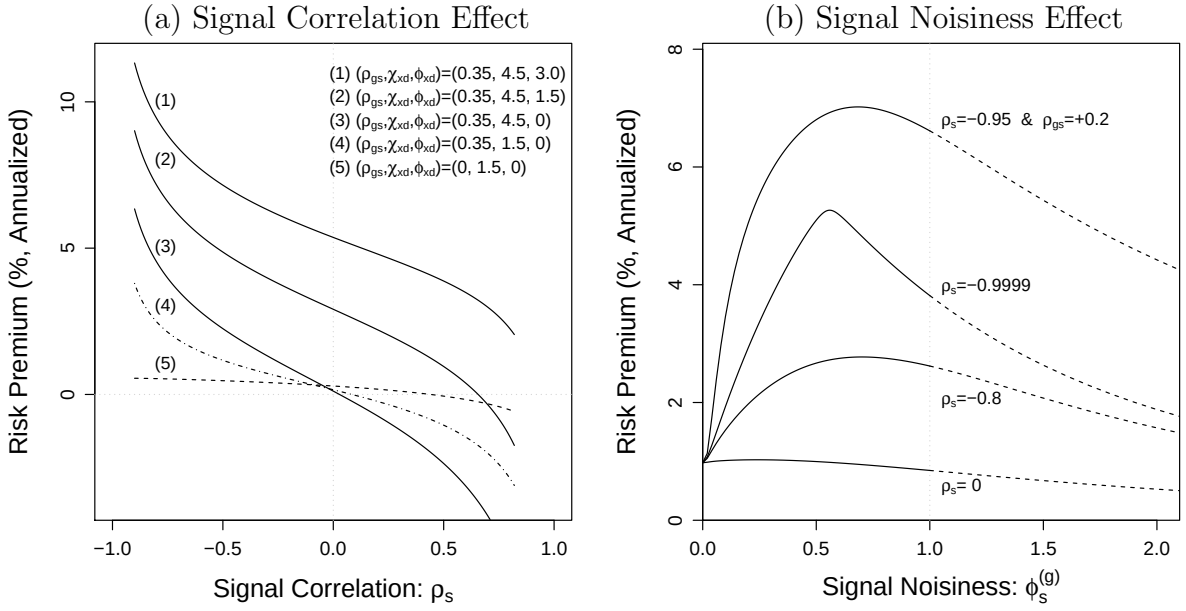


Table 1: Equity Premium and (Co)-Uncertainty in I.I.D. Growth Models

The table shows the results for three cases (i)–(iii) in Section 2.3. The parameters are calibrated at the quarterly frequency following Collin-Dufresne et al. (2016): $\beta = 0.994$, $\gamma = 10$, $\psi = 1.5$, $\mu_c = \mu_d = 0.0045$, and $\sigma_c = 0.0135$. We adopt parameters for dividend growth process from Bansal and Yaron (2004): $\sigma_d = 4.5\sigma_c$ and $\rho_g = 0.55$. The numbers in the first row correspond to the results after $t = 100$ quarters since the agent's learning starts. The first three columns report the annualized cumulative equity premium from the beginning of learning. The next three columns show static uncertainty regarding the annualized mean consumption growth rate. The last three columns show intertemporal co-uncertainty in the annualized mean consumption and dividend growth rates.

Correlation in	Annualized equity premium (cumulative, %) $E(R_m - R_f)$			Static uncertainty in mean consumption growth ($\times 10^4$) $Var(\mu_c \mid \mathcal{F}_t)$			Intertemporal co-uncertainty in mean growth ($\times 10^6$) $Cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1} \mid \mathcal{F}_t)$		
growth rates: ρ_g	0.550	0.000	0.000	0.550	0.000	0.000	0.550	0.000	0.000
aggreg. signals: ρ_s	0.550	0.550	0.800	0.550	0.550	0.800	0.550	0.550	0.800
beliefs: $\rho_{\mu,t}$	0.550	0.550	0.800	0.550	0.550	0.800	0.550	0.550	0.800
t (quarters)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
100	13.24	15.48	27.05	0.139	0.139	0.139	0.244	0.244	0.356
200	8.90	8.92	15.45	0.081	0.081	0.081	0.083	0.083	0.121
400	5.88	4.83	8.33	0.044	0.044	0.044	0.025	0.025	0.036
800	3.95	2.52	4.34	0.023	0.023	0.023	0.007	0.007	0.010
∞	1.21	0.00	0.00	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: Asset Moments and (Co)-Uncertainty in Long-Run Risk Models

The table shows the effects of parameters (ϕ_{xd}, χ_{xd}, ρ_{gs}, ρ_s) on equilibrium asset pricing, static uncertainty, and intertemporal co-uncertainty in long-run risk models calibrated at the monthly frequency. The baseline parameter values are from Bansal and Yaron (2004): $\gamma = 10$, $\psi = 1.5$, $\mu_c = \mu_d = 0.0015$, $\sigma_c = 0.0078$, $\sigma_d = 4.5\sigma_c$, $\sigma_{xc} = 0.044\sigma_c$, $F_x = 0.979$, and $\rho_g = 0$ except for $\beta = 0.999$ to set the risk-free rate at 1%/year in the row #9. The benchmark parameter calibration of Bansal and Yaron (2004) corresponds to $\phi_{xd} = \chi_{xd} = 3$, but we explore larger values of χ_{xd} as well, motivated by the detection error probability results in the online Appendix A.4. The noisiness of aggregate signals is set as $(\sigma_{sc}, \sigma_{sd}) = (\sigma_c, \sigma_d)$. As implied by the patterns in static uncertainty, aggregate signals become more accurate as their signal correlations get further apart from zero even if $(\sigma_{sc}, \sigma_{sd})$ are fixed. This is because Shannon's entropy of a multivariate normal distribution depends on the determinant of its covariance matrix. Therefore, correlated aggregate signals in this table are more informative than the realized growth rates (as signals) even if we set $(\sigma_{sc}, \sigma_{sd}) = (\sigma_c, \sigma_d)$.

	Systematic exposure to $x_{c,t}$ and σ_{xd}/σ_{xc} $\{\phi_{xd}, \chi_{xd}\}$	Growth-to-signal correlation ρ_{gs}	Aggregate signal correlation ρ_s	Annualized equity premium (%) $E(R_m - R_f)$	Annualized risk-free rates (%) $E(R_f)$	Static uncertainty in $x_{c,t}$ ($\times 10^6$) $Var(x_{c,t}\|\mathcal{F}_t)$	Intertemporal co-uncertainty ($\times 10^7$) $Cov(\hat{\mathbf{x}}_{\mathbf{t+1}}\|\mathcal{F}_t)$
Panel A: Learning $\mathbf{x}_t$ from consumption and dividend growth rates $\mathbf{s}_t = \mathbf{g}_t$							
#1	{3, 6.0}	1.0	0.0	5.83	0.79	1.64	2.12
#2	{3, 4.5}	1.0	0.0	5.81	0.80	1.61	1.91
#3	{1, 4.5}	1.0	0.0	1.32	0.77	1.72	0.65
#4	{0, 4.5}	1.0	0.0	-1.09	0.76	1.73	0.00
Panel B: Learning $\mathbf{x}_t$ from aggregate signals $\mathbf{s}_t$ including $\mathbf{g}_t$ information							
#5	{1, 4.5}	0.3	0.8	-3.40	1.18	1.60	-0.20
#6	{1, 4.5}	0.3	-0.5	2.85	1.27	1.59	1.16
#7	{1, 4.5}	0.3	-0.9	6.58	0.88	1.27	1.91
#8	{1, 6.0}	0.3	-0.9	7.98	0.95	1.39	2.26
#9	{0, 6.0}	0.3	-0.9	6.24	1.00	1.49	1.46
#10	{0, 6.0}	0.0	-0.9	2.58	1.57	1.49	1.46
#11	{0, 6.0}	0.0	-1.0	4.00	1.53	1.28	2.18
#12	{1, 4.5}	0.0	0.8	-1.32	1.59	1.60	-0.20
#13	{3, 6.0}	0.3	-0.9	11.03	0.83	1.12	3.78

Table 3: Preference Parameters and Asset Moments in Long-Run Risk Models

The table shows asset moments with different preference parameter values. We set the baseline parameter values following Table 2 except for $\sigma_{sc} = 0.9\sigma_{gc}$, $\sigma_{sd} = 0.9\sigma_{gd}$, $\rho_g = 0.3$, $\rho_{gs} = 0.35$, $\chi_{xd} = 4.5$, $\rho_s = -0.9$, and $\beta = 0.9994$ consistent with Li and Zhang (2017). Then we explore different values of risk aversion γ and EIS ψ. We also present three benchmark cases with $\gamma = 10$.

	Equity premium (%) $E(R_m - R_f)$, annualized				Equity volatility (%) $SD(r_{m,t+1})$, annualized				Mean risk-free rate (%) $E(R_f)$, annualized			
ψ	0.5	0.9	1.1	1.5	0.5	0.9	1.1	1.5	0.5	0.9	1.1	1.5
γ												
4.0	2.09	3.48	3.86	4.34	23.78	26.50	27.20	28.07	4.02	2.25	1.82	1.30
5.0	2.85	4.41	4.81	5.33	23.20	25.59	26.20	26.95	3.91	2.09	1.65	1.12
7.5	4.60	6.43	6.88	7.43	21.60	23.41	23.85	24.41	3.63	1.68	1.22	0.66
10.0	6.18	8.09	8.54	9.09	20.02	21.47	21.83	22.26	3.36	1.28	0.80	0.22
Benchmark #1: signal correlations $\rho_s = 0$												
10.0	1.15	2.69	3.05	3.50	22.83	23.09	23.19	23.33	3.27	1.34	0.89	0.35
Benchmark #2: growth = signals ($\mathbf{s}_t = \mathbf{g}_t$)												
10.0	1.47	2.91	3.25	3.66	22.79	23.04	23.12	23.25	3.03	1.20	0.77	0.26
Benchmark #3: known $\mathbf{x}_t$ by signals												
10.0	-0.05	2.10	2.65	3.35	25.12	25.24	25.38	25.61	4.49	2.10	1.52	0.83

Table 4: Information Quality and Equity Premium in Long-Run Risk Models

The table shows how the equity premium changes with different levels of signal noisiness (or an inverse of information quality). In Panel A, signal noisiness is a scaled version of the overall signal noisiness σ_s from (40) of the *aggregate* signals which also include information in growth rates. At 'Noisiness = 1', $Var(x_{c,t}|\mathcal{F}_t)$ equals that of the benchmark case in which the growth rates are the aggregate signal $\mathbf{s}_t = \mathbf{g}_t$. In panel A, we set the baseline parameter values following Table 2, except for $\chi_{xd} = 4.5$. Each column explores different values of ρ_s, ρ_{gs}, and ϕ_{xd}. In Panel B, signal noisiness is defined as a scaled version of the overall signal noisiness of the *additional* signals on top of the growth rates so that an agent learns from both the additional signal and the growth rates. We set $(\sigma_{sc}, \sigma_{sd})$ as a multiple of (σ_c, σ_d) so that $\sigma_{sc} = \sigma_c$ and $\sigma_{sd} = \sigma_d$ at 'Noisiness = 1'. The first six columns in Panel B show the equity premium with different levels of (ρ_s, ϕ_{xd}, χ_{xd}). The last three columns in Panel B show the signal correlations of the implied aggregate signals (i.e., additional signals plus growth rates as a whole) for both cases $\phi_{xd} = \chi_{xd} = 3$ and $(\phi_{xd}, \chi_{xd}) = (3, 4.5)$. In Panel B, we set the baseline parameter values following Table 2, except for $\rho_{gs} = 0$.

Panel A: Noisiness of aggregate signals

	Equity premium (%)					Intertemporal co-uncertainty ($\times 10^7$)				
ϕ_{xd}	4.5	1.0	1.0	1.0	4.5	4.5	1.0	1.0	1.0	4.5
ρ_{gs}	0	0	0	0.2	0.2	0	0	0	0.2	0.2
ρ_s	0	0	-1.0	-0.95	-0.9	0	0	-1.0	-0.95	-0.9
Noisiness										
0	8.50	0.98	0.98	0.98	5.60	5.08	1.13	1.13	1.13	3.39
0.25	7.50	1.03	3.37	5.53	8.44	4.38	1.01	2.12	1.79	3.55
1	4.90	0.85	3.82	6.61	6.68	2.67	0.65	2.09	1.84	2.25
2	2.88	0.53	1.88	4.42	4.34	1.47	0.36	0.97	0.93	1.16
10	0.37	0.07	0.15	0.89	0.88	0.11	0.02	0.06	0.06	0.08

Panel B: Noisiness of additional signals

additional	$\phi_{xd} = \chi_{xd} = 3$			$\phi_{xd} = 3$, $\chi_{xd} = 4.5$			Aggregate signal's ρ_s		
signal's ρ_s	0.5	-0.5	-0.9	0.5	-0.5	-0.9	0.5	-0.5	-0.9
Noisiness									
0	5.42	5.42	5.42	5.60	5.60	5.60	0.50	-0.50	-0.90
0.3	5.56	5.50	5.43	5.65	5.45	5.34	0.46	-0.46	-0.88
1	5.79	5.69	5.45	5.77	5.56	5.34	0.28	-0.28	-0.75
∞	5.91	5.91	5.91	5.81	5.81	5.81	0.00	-0.00	-0.00

Table 5: Signal Correlation and Co-Uncertainty in Survey Data (SPF)

The table shows estimates on static (co)-uncertainty and implied signal correlations for a median forecaster in the Survey of Professional Forecasters (SPF). The estimates are from the forecast errors associated with GDP and corporate profits growth rates, up to the four-quarter horizon. The numbers in parenthesis are 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals from Efron (1987). Changes in covariance, $\Delta Cov(g_{c,t}, g_{d,t}|\mathcal{F}_{t-h})^*$ in Panel B is scaled by a product of prior standard deviations in the last two columns of Panel A. The last column in Panel B show the generalized signal correlation coefficients in Definition 3, implied by Equation (38).

Panel A: Static co-uncertainty / Static uncertainty				
Horizon (h: quarter)	$Corr(g_{c,t}, g_{d,t}\|\mathcal{F}_{t-h})$	$Cov(g_{c,t}, g_{d,t}\|\mathcal{F}_{t-h})$ $\times 10^4$	$SD(g_{c,t}\|\mathcal{F}_{t-h})$ $\times 10^2$	$SD(g_{d,t}\|\mathcal{F}_{t-h})$ $\times 10^2$
1	.40 (.25, .54)	1.90 (1.02, 2.87)	.78 (.65, .91)	6.10 (5.00,7.15)
2	.51 (.38, .63)	3.08 (1.94, 4.33)	.92 (.78, 1.06)	6.54 (5.46,7.57)
3	.59 (.47, .69)	4.09 (2.62, 5.74)	.99 (.85, 1.13)	6.98 (5.80, 8.08)
4	0.58 (.46, .69)	4.06 (2.63, 5.62)	1.00 (.86, 1.14)	6.96 (5.77, 8.10)

Panel B: Changes in static uncertainty / Signal correlations			
$h \to h'$	$\Delta Corr(g_{c,t}, g_{d,t}\|\mathcal{F}_{t-h})$	$\Delta Cov(g_{c,t}, g_{d,t}\|\mathcal{F}_{t-h})^*$ scaled by prior standard deviations	Generalized signal correlation $Corr(s_{c,t}, s_{d,t})$
$2 \to 1$	$-.11$ $(-.17, -.06)$	$-.20$ $(-.27, -.13)$	-1.63 $(-53.48 - 0.71)$
$3 \to 1$	$-.19$ $(-.27, -.12)$	$-.32$ $(-.40, -.23)$	-2.08 $(-41.69 - 0.86)$
$3 \to 2$	$-.08$ $(-.12, -.03)$	$-.15$ $(-.20, -.09)$	-3.96 $(-134.38 - 0.60)$

Online Appendix

(Not Intended for Publication)

for

Dynamic Multivariate Learning with Generalized Information: Asset Pricing Implications

Praveen Kumar and James Yae

March 15, 2021

A.1 Solutions for the I.I.D. Model

Before presenting solutions for the i.i.d. model from Section 2, we fist discuss two important points: 1) a representative agent's information set and 2) truncation of a prior distribution.

First, remember we introduce an aggregate bivariate signal $\mathbf{s}_t = [s_{c,t} \;\; s_{d,t}]^\top$ observed at t, which represents the whole available information to the agent including growth realizations $\mathbf{g}_t$. Therefore, the agent's information set $\mathcal{F}_t$ at t equals the history of aggregate signals up to t, that is, $\mathbf{s}^t = \{\mathbf{s}_1, ..., \mathbf{s}_t\}$, and we have $p(\boldsymbol{\mu}|\mathcal{F}_t) = p(\boldsymbol{\mu}|\mathbf{s}^t) = p(\boldsymbol{\mu}|\mathbf{s}^t, \mathbf{g}^t)$. By Bayes' theorem, the agent's belief at t follows $\boldsymbol{\mu}|\mathcal{F}_t \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_t, \boldsymbol{\Sigma}_{\boldsymbol{\mu},t})$ where the posterior mean and covariance matrix $(\hat{\boldsymbol{\mu}}_t, \boldsymbol{\Sigma}_{\boldsymbol{\mu},t})$ can be recursively computed as

$$\begin{aligned} \hat{\boldsymbol{\mu}}_t &= \hat{\boldsymbol{\mu}}_{t-1} + \mathbf{K}_{s,t}(\mathbf{s}_t - \hat{\boldsymbol{\mu}}_{t-1}), \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}^{-1} &= \boldsymbol{\Sigma}_{\boldsymbol{\mu},t-1}^{-1} + \boldsymbol{\Sigma}_s^{-1}. \end{aligned} \tag{A.1}$$

Here, $\mathbf{K}_{s,t} = \mathbf{\Sigma}_{\boldsymbol{\mu},t}\mathbf{\Sigma}_s^{-1} = \frac{1}{n+t}\mathbf{I}$ is the (2×2) Kalman gain matrix associated with aggregate signals and $\mathbf{I}$ is an identity matrix. Note the expression for Kalman gain is still simple even with non-zero cross-covariance matrix, $J_{gs} = Cov(\mathbf{e}_{g,t}, \mathbf{e}_{s,t}) \neq 0$, because the agent learns only from the aggregate signals. Alternatively, the expression (A.1) can be written as

$$\hat{\boldsymbol{\mu}}_t = \frac{n}{n+t}\boldsymbol{\mu}_0 + \frac{t}{n+t}\bar{\mathbf{s}}^t \quad \text{and} \quad \mathbf{\Sigma}_{\boldsymbol{\mu},t} = \frac{1}{n+t}\mathbf{\Sigma}_s, \tag{A.2}$$

where $\bar{\mathbf{s}}^t$ is the sample average of the signals $\mathbf{s}_\tau$ for $\tau = 1, 2, ..., t$. However, we choose to use the recursive form in (A.1) to make it easy to compare the i.i.d. case with time-varying parameter cases.

Second, as shown in Collin-Dufresne et al. (2016), truncation of a prior distribution has minimal effects on final results.[29] The intuition is simple. In their benchmark i.i.d model, the high risk premium arises with learning because the agent's parameter estimate $\hat{\mu}_{c,t}$ (or posterior mean) varies over time by learning, not because the consumption mean growth rate μ_c can have arbitrarily high or low values. In other words, the risk premium arises because learning creates intertemporal uncertainty of the consumption mean-growth-rate parameter μ_c rather than its static uncertainty. Likewise, in our i.i.d. model, the risk premium arises because learning makes the agent's parameter estimates $(\hat{\mu}_{c,t}, \hat{\mu}_{d,t})$ on the mean growth rates (μ_c, μ_d) of consumption and dividends covary over time (intertemporal co-uncertainty). Extreme values of (μ_c, μ_d) hardly affect the covariance of $(\hat{\mu}_{c,t}, \hat{\mu}_{d,t})$; therefore, truncation of (μ_c, μ_d) has no major effects on final results.

A.1.1 Wealth-Consumption Ratio

Throughout the section A.1, we use boldface symbols for non-scalar vectors and matrices. Suppose $R_{c,t+1} = \frac{W_{t+1}+C_{t+1}}{W_t}$ denotes the gross return on the total wealth portfolio that is valued at W_t and pays out the aggregate consumption stream $\{C_t\}$. Following Campbell and Shiller (1988), the log wealth-consumption ratio $z_{c,t}$ is a linear function of the state variable

[29] Conceptually, truncation of a prior distribution for $\boldsymbol{\mu}$ makes sense in that *unconditional* mean growth rates are unlikely to be extreme values. From a technical perspective, truncation of the prior distribution is considered so that a transversality condition is met even with EIS (ψ) larger than unity.

$\hat{\boldsymbol{\mu}}_t$ given in Equation (5).

$$z_{c,t} = \log \frac{W_t}{C_t} = A_{0c,t} + \mathbf{A}_{1c,t}\hat{\boldsymbol{\mu}}_t. \tag{A.3}$$

Note the intercept term $A_{0c,t}$ is time-dependent because parameter uncertainty gradually decreases by learning. Therefore, $A_{0c,t}$ contains the term $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ that represents such (bivariate) parameter uncertainty. However, we do not express $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ as an explicit state variable because it has no randomness as shown in (A.2). Next, we approximate the log return of the total wealth, $r_{c,t+1} = \log R_{c,t+1}$, as

$$\begin{aligned} r_{c,t+1} &= \log\left(\exp(z_{c,t+1}) + 1\right) - z_{c,t} + g_{c,t+1}, \\ &\approx \kappa_{0c,t} + \kappa_{1c,t} z_{c,t+1} - z_{c,t} + g_{c,t+1}, \end{aligned} \tag{A.4}$$

where we define Taylor approximation coefficients as

$$\begin{aligned} \kappa_{1c,t} &= \frac{\exp(\bar{z}_{c,t+1})}{\exp(\bar{z}_{c,t+1}) + 1}, \\ \kappa_{0c,t} &= \log \frac{1}{1 - \kappa_{1c,t}} - \kappa_{1c,t}\bar{z}_{c,t+1}, \\ \bar{z}_{c,t+1} &= A_{0c,t+1} + \mathbf{A}_{1c,t+1}\boldsymbol{\mu}, \end{aligned}$$

where $\kappa_{1c,t}$ is a constant between zero and one. Note the average log wealth-consumption ratio $\bar{z}_{c,t+1}$ depends on time due to the time-dependence of $A_{0c,t}$. We then re-calculate Taylor approximation coefficients $(\kappa_{0c,t}, \kappa_{1c,t})$ with respect to $\bar{z}_{c,t+1}$ at each time t for better approximation accuracy. As a result, a 1×2 row vector $\mathbf{A}_{1c,t}$ becomes also time-dependent. However, we omit the time subscript of $\mathbf{A}_{1c,t}$ in the main text for simple notation.

Following Epstein and Zin (1989), the stochastic discount factor (SDF) for recursive preferences has the form:

$$M_{t+1} = \delta^{\theta} G_{c,t+1}^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1}, \tag{A.5}$$

where $G_{c,t+1} = \frac{C_{t+1}}{C_t}$ is the aggregate gross growth rate of per-capita consumption. Among the preference parameters, $0 < \delta < 1$ is a time discount factor, $\gamma \geq 0$ is the risk aversion parameter, $\psi \geq 0$ is the elasticity of intertemporal substitution (EIS), and $\theta = (1-\gamma)/(1-\frac{1}{\psi})$. The stochastic discount factor in (A.5) implies the following Euler equation conditional on

the agent's information set $\mathcal{F}_t$:

$$E[\delta^{\theta} G_{c,t+1}^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1} R_{i,t+1} | \mathcal{F}_t] = 1, \tag{A.6}$$

where $R_{i,t+1}$ is the gross return of any asset i in the economy. Using this Euler equation with $R_{i,t+1} = R_{c,t+1}$, we solve for the wealth-consumption ratio, using a backward recursion from the limiting case of known-parameters at sufficiently large t. For concise notation, we first define the following terms

$$\boldsymbol{\lambda}_g = \gamma \mathbf{I}_c = \gamma [1 \;\; 0] \quad \text{and} \quad \boldsymbol{\lambda}_{s,t} = (1-\theta)\kappa_{1c,t} \mathbf{A}_{1c,t} \mathbf{K}_{s,t+1}, \tag{A.7}$$

and their scaled versions

$$\boldsymbol{\lambda}_g^* = \frac{1-\gamma}{\gamma} \boldsymbol{\lambda}_g \quad \text{and} \quad \boldsymbol{\lambda}_{s,t}^* = \frac{\theta}{1-\theta} \boldsymbol{\lambda}_{s,t}. \tag{A.8}$$

Then, we have the following identity from (A.3) and (A.4).

$$\begin{aligned}
\log\left(G_{c,t+1}^{-\frac{1}{\psi}} R_{c,t+1}\right) &= \left(1-\psi^{-1}\right) g_{c,t+1} + \kappa_{0c,t} + \kappa_{1c,t}(A_{0c,t+1} + \mathbf{A}_{1c,t+1}\hat{\boldsymbol{\mu}}_{t+1}) - (A_{0c,t} + \mathbf{A}_{1c,t}\hat{\boldsymbol{\mu}}_t), \\
&= (\kappa_{0c,t} + \kappa_{1c,t} A_{0c,t+1} - A_{0c,t}) \\
&\quad + \left(\kappa_{1c,t}\mathbf{A}_{1c,t+1} - \mathbf{A}_{1c,t} + (1-\psi^{-1})\mathbf{I}_c\right)\hat{\boldsymbol{\mu}}_t \\
&\quad + (1-\psi^{-1})\mathbf{I}_c(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \kappa_{1c,t}\mathbf{A}_{1c,t+1}(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t),
\end{aligned} \tag{A.9}$$

where the last two terms multiplied by θ are expressed as

$$\theta(1-\psi^{-1})\mathbf{I}_c(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) = \frac{1-\gamma}{\gamma}\boldsymbol{\lambda}_g(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) = \boldsymbol{\lambda}_g^*(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t), \tag{A.10}$$

$$\begin{aligned}
\theta\kappa_{1c,t}\mathbf{A}_{1c,t+1}(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t) &= \theta\kappa_{1c,t}\mathbf{A}_{1c,t+1}\mathbf{K}_{s,t+1}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\
&= \frac{\theta}{1-\theta}\boldsymbol{\lambda}_{s,t}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) = \boldsymbol{\lambda}_{s,t}^*(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t).
\end{aligned} \tag{A.11}$$

Note a vector of random variables $(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t)$ and $(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t)$ follows a multivariate normal

distribution under the agent's posterior belief:

$$\begin{bmatrix} \mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t \\ \mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t \end{bmatrix} | \mathcal{F}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*_{gs}\right) \quad \text{where} \quad \boldsymbol{\Sigma}^*_{gs} = \begin{bmatrix} \boldsymbol{\Sigma}^*_g & \mathbf{J}^*_{gs} \\ \mathbf{J}^*_{sg} & \boldsymbol{\Sigma}^*_s \end{bmatrix}, \tag{A.12}$$

where $\boldsymbol{\Sigma}^*_g = \boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, $\boldsymbol{\Sigma}^*_s = \boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, $\mathbf{J}^*_{sg} = \mathbf{J}_{sg} + \boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$, and $\mathbf{J}^*_{gs} = (\mathbf{J}^*_{sg})^\top$, all of which include static (co)-uncertainty matrix $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ because of the following decompositions:

$$\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t = (\mathbf{g}_{t+1} - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t) \quad \text{and} \quad \mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t = (\mathbf{s}_{t+1} - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t), \tag{A.13}$$

where $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t) = \mathbf{u}_{\boldsymbol{\mu},t+1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\mu},t})$ is a random variable that represents static (co)-uncertainty of $\boldsymbol{\mu}$ under the agent's posterior belief.

Finally, we plug (A.10)–(A.12) to the Euler equation (A.6) and solve the expectation using a log-normal distribution. The resulting equation should be satisfied regardless of $\hat{\boldsymbol{\mu}}_t$, and so the coefficient on $\hat{\boldsymbol{\mu}}_t$ should be zero while the remaining term should be also zero; therefore we have the following recursive equations at each t.

$$\mathbf{A}_{1c,t} = (1 - \psi^{-1})\mathbf{I}_c + \kappa_{1c,t}\mathbf{A}_{1c,t+1}, \tag{A.14}$$

$$A_{0c,t} = \log\delta + \kappa_{0c,t} + \kappa_{1c,t}A_{0c,t+1} + \frac{1}{2\theta}\boldsymbol{\lambda}^*_t\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*_t\right)^\top, \tag{A.15}$$

where $\boldsymbol{\lambda}^*_t = [\boldsymbol{\lambda}^*_g \;\; \boldsymbol{\lambda}^*_{s,t}]$.

On the other hand, as $t \to \infty$, $\boldsymbol{\mu}$ is revealed and there is no further learning. Therefore, Euler equation, as $t \to \infty$, approaches

$$z_c = \log\delta + (1 - \psi^{-1})\mathbf{I}_c\boldsymbol{\mu} + \log\left(\exp(z_c) + 1\right) + \frac{1}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top, \tag{A.16}$$

where $\boldsymbol{\lambda}^* = [\boldsymbol{\lambda}^*_g \;\; 0 \;\; 0]$. If $\boldsymbol{\mu}$ is revealed, the wealth-consumption ratio is constant and so Taylor-approximation is exact. By plugging $z_c = A_{0c} + \mathbf{A}_{1c}\boldsymbol{\mu}$ and $\log\left(\exp(z_c) + 1\right) = \kappa_{0c} + \kappa_{1c}z_c$ into (A.16), we have a boundary condition:

$$A_{0c} = \log\delta + \kappa_{0c} + \kappa_{1c}A_{0c} + \frac{1}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top + \left((\kappa_{1c} - 1)\mathbf{A}_{1c} + (1 - \psi^{-1})\mathbf{I}_c\right)\boldsymbol{\mu}, \tag{A.17}$$

which implies the following.

$$\mathbf{A}_{1c} = \frac{1-\psi^{-1}}{1-\kappa_{1c}}\mathbf{I}_c, \tag{A.18}$$

$$A_{0c} = \frac{1}{1-\kappa_{1c}}\left(\log\delta + \kappa_{0c} + \frac{1}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top\right). \tag{A.19}$$

Note these equations are the limiting cases (i.e., $t \to \infty$) of (A.14) and (A.15).

The role of the EIS regarding the wealth and substitution effects follows Bansal and Yaron (2004). The sign of the coefficient on $\hat{\mu}_{c,t}$ in the log wealth-consumption ratio depends on EIS ψ. When $\psi > 1$, as calibrated in Bansal and Yaron (2004) for example, the first element of $\mathbf{A}_{1c}$ is positive, and the intertemporal substitution effect dominates the wealth effect. Thus, the agent is ceteris paribus willing to pay more for the asset that generates the aggregate consumption stream $\{C_t\}$ when they are more optimistic about future expected growth, say, in good times.[30]

A.1.2 Price-Dividend Ratio

We compute the price-dividend ratio following the wealth-consumption ratio. Let $R_{m,t+1} = \frac{P_{t+1}+D_{t+1}}{P_t}$ denote the gross return on a risky asset that is priced at P_t and pays out the aggregate dividend stream $\{D_t\}$. Then, the log price-dividend ratio is written as

$$z_{d,t} = \log\frac{P_t}{D_t} = A_{0d,t} + \mathbf{A}_{1d,t}\hat{\boldsymbol{\mu}}_t. \tag{A.20}$$

Like the case of wealth-consumption ratio, the intercept term $A_{0d,t}$ is time-dependent because parameter uncertainty gradually decreases by learning. Therefore, $A_{0d,t}$ contains the term $\boldsymbol{\Sigma}_{\boldsymbol{\mu},t}$ that represents such (bivariate) parameter uncertainty. Next, the log return of the risky asset, $r_{m,t+1} = \log R_{m,t+1}$, is approximated as

$$r_{m,t+1} \approx \kappa_{0d,t} + \kappa_{1d,t}z_{d,t+1} - z_{d,t} + g_{d,t+1}, \tag{A.21}$$

[30] We can apply a similar interpretation to the equity that generates the dividend stream $\{D_t\}$ when considering the special case where the agent believes $\mu_c = \mu_d$, and hence $\hat{\mu}_{c,t} = \hat{\mu}_{d,t}$.

where we recalculate Taylor approximation coefficients $\kappa_{0d,t}$, $\kappa_{1d,t}$, and $\bar{z}_{d,t+1}$ following the method used for the wealth-consumption ratio. Therefore, we omit the time subscript of $\mathbf{A}_{1d,t}$ in the main text for simple notation. We first define the following terms for concise notations.

$$\boldsymbol{\beta}_g = \mathbf{I}_d = [0 \;\; 1] \quad \text{and} \quad \boldsymbol{\beta}_{s,t} = \kappa_{1d}\mathbf{A}_{1d}\mathbf{K}_{s,t+1}. \tag{A.22}$$

Then, by plugging $R_{i,t+1} = R_{m,t+1}$ into the Euler equation (A.6), we first compute the following:

$$\begin{aligned}
\log\left(G_{c,t+1}^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1} R_{m,t+1}\right) = &\left(-\frac{\theta}{\psi}\mathbf{I}_c + (\theta-1)\mathbf{I}_c + \mathbf{I}_d\right)\mathbf{g}_{t+1} \\
&+ (\theta-1)\left(\kappa_{0c,t} + \kappa_{1c,t}(A_{0c,t+1} + \mathbf{A}_{1c,t+1}\hat{\boldsymbol{\mu}}_{t+1}) - A_{0c,t} - \mathbf{A}_{1c,t}\hat{\boldsymbol{\mu}}_t\right) \\
&+ \left(\kappa_{0d,t} + \kappa_{1d,t}(A_{0d,t+1} + \mathbf{A}_{1d,t+1}\hat{\boldsymbol{\mu}}_{t+1}) - A_{0d,t} - \mathbf{A}_{1d,t}\hat{\boldsymbol{\mu}}_t\right) \\
=&(\theta-1)(\kappa_{0c,t} + \kappa_{1c,t}A_{0c,t+1} - A_{0c,t}) + (\kappa_{0d,t} + \kappa_{1d,t}A_{0d,t+1} - A_{0d,t}) \\
&+ (\mathbf{I}_d - \boldsymbol{\lambda}_g + (\theta-1)(\kappa_{1c,t}\mathbf{A}_{1c,t+1} - \mathbf{A}_{1c,t}) + \kappa_{1d,t}\mathbf{A}_{1d,t+1} - \mathbf{A}_{1d,t})\,\hat{\boldsymbol{\mu}}_t \\
&+ (\mathbf{I}_d - \boldsymbol{\lambda}_g)(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\
&+ [(\theta-1)\kappa_{1c,t}\mathbf{A}_{1c,t+1} + \kappa_{1d,t}\mathbf{A}_{1d,t+1}]\,(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t).
\end{aligned} \tag{A.23}$$

Note the last term can be simplified, by (A.1), to

$$[(\theta-1)\kappa_{1c,t}\mathbf{A}_{1c,t+1} + \kappa_{1d,t}\mathbf{A}_{1d,t+1}]\,(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t) = (\boldsymbol{\beta}_{s,t} - \boldsymbol{\lambda}_{s,t})(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t). \tag{A.24}$$

Finally, plug (A.23) and (A.24) to the Euler equation (A.6) and apply $\kappa_{1c,t}\mathbf{A}_{1c,t+1} - \mathbf{A}_{1c,t} = (\psi^{-1}-1)\mathbf{I}_c$ from (A.14). Then we have the following condition since (A.23) should be satisfied regardless of $\hat{\boldsymbol{\mu}}_t$.

$$\begin{aligned}
0 &= \mathbf{I}_d - \boldsymbol{\lambda}_g + (\theta-1)(\kappa_{1c,t}\mathbf{A}_{1c,t+1} - \mathbf{A}_{1c,t}) + \kappa_{1d,t}\mathbf{A}_{1d,t+1} - \mathbf{A}_{1d,t} \\
&= \mathbf{I}_d - \psi^{-1}\mathbf{I}_c + \kappa_{1d,t}\mathbf{A}_{1d,t+1} - \mathbf{A}_{1d,t}.
\end{aligned}$$

That is, we can recursively calculate $\mathbf{A}_{1d,t}$ from

$$\mathbf{A}_{1d,t} = -\frac{1}{\psi}\mathbf{I}_c + \mathbf{I}_d + \kappa_{1d,t}\mathbf{A}_{1d,t+1}. \tag{A.25}$$

The remaining terms in the Euler equation should be zero as well, and so we have

$$\begin{aligned} A_{0d,t} &= \theta \log\delta + (\theta-1)(\kappa_{0c,t} + \kappa_{1c,t}A_{0c,t+1} - A_{0c,t}) + \kappa_{0d,t} + \kappa_{1d,t}A_{0d,t+1} + \frac{1}{2}\boldsymbol{\beta}_t^*\boldsymbol{\Sigma}_{gs}^*\left(\boldsymbol{\beta}_t^*\right)^\top \\ &= \log\delta + \kappa_{0d,t} + \kappa_{1d,t}A_{0d,t+1} + \frac{1-\theta}{2\theta}\boldsymbol{\lambda}_t^*\boldsymbol{\Sigma}_{gs}^*\left(\boldsymbol{\lambda}_t^*\right)^\top + \frac{1}{2}\boldsymbol{\beta}_t^*\boldsymbol{\Sigma}_{gs}^*\left(\boldsymbol{\beta}_t^*\right)^\top, \end{aligned} \tag{A.26}$$

where the last equation comes from (A.15) and $\boldsymbol{\beta}_t^* = \begin{bmatrix}(\boldsymbol{\beta}_g - \boldsymbol{\lambda}_g) & (\boldsymbol{\beta}_{s,t} - \boldsymbol{\lambda}_{s,t})\end{bmatrix}$. On the other hand, as $t \to \infty$, $\boldsymbol{\mu}$ is revealed, and we have the following boundary conditions:

$$\mathbf{A}_{1d} = \frac{-\psi^{-1}}{1-\kappa_{1d}}\mathbf{I}_c + \frac{1}{1-\kappa_{1d}}\mathbf{I}_d, \tag{A.27}$$

$$A_{0d} = \frac{1}{1-\kappa_{1d}}\left(\log\delta + \kappa_{0d} + \frac{1-\theta}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}_{gs}^*\left(\boldsymbol{\lambda}^*\right)^\top + \frac{1}{2}\boldsymbol{\beta}^*\boldsymbol{\Sigma}_{gs}^*\left(\boldsymbol{\beta}^*\right)^\top\right), \tag{A.28}$$

where $\boldsymbol{\beta}^* = \begin{bmatrix}(\boldsymbol{\beta}_g - \boldsymbol{\lambda}_g) & 0 & 0\end{bmatrix}$.

A.1.3 Stochastic Discount Factor and Equity Returns

Using the results from Section (A.1.1), we express the log of the stochastic discount factor as follows.

$$\begin{aligned} m_{t+1} =& \theta\log\delta - \frac{\theta}{\psi}g_{c,t+1} + (\theta-1)r_{c,t+1} \\ \approx& \theta\log\delta - \frac{\theta}{\psi}g_{c,t+1} + (\theta-1)\left(\kappa_{0c,t} + \kappa_{1c,t}(A_{0c,t+1} + \mathbf{A}_{1c,t+1}\hat{\boldsymbol{\mu}}_{t+1}) - A_{0c,t} - \mathbf{A}_{1c,t}\hat{\boldsymbol{\mu}}_t + g_{c,t+1}\right) \\ =& \theta\log\delta + (\theta-1)\left(\kappa_{0c,t} + \kappa_{1c,t}A_{0c,t+1} - A_{0c,t}\right) + (\theta-1)(\kappa_{1c,t}\mathbf{A}_{1c,t+1} - \mathbf{A}_{1c,t})\hat{\boldsymbol{\mu}}_t \\ & - \left(\frac{\theta}{\psi} + 1 - \theta\right)g_{c,t+1} + (\theta-1)\kappa_{1c,t}\mathbf{A}_{1c,t+1}(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\ =& \theta\log\delta + (\theta-1)\left(\kappa_{0c,t} + \kappa_{1c,t}A_{0c,t+1} - A_{0c,t}\right) + \left(-\gamma + (\theta-1)(\psi^{-1}-1)\right)\mathbf{I}_c\hat{\boldsymbol{\mu}}_t \\ & - \gamma\mathbf{I}_c(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) - (1-\theta)\kappa_{1c,t}\mathbf{A}_{1c,t+1}\mathbf{K}_{s,t+1}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t). \end{aligned} \tag{A.29}$$

Therefore, the innovation in m_{t+1} conditional on the agent's information set is

$$\begin{aligned} m_{t+1} - E[m_{t+1}|\mathcal{F}_t] &\approx -\gamma \mathbf{I}_c(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) - (1-\theta)\kappa_{1c,t}\mathbf{A}_{1c,t+1}\mathbf{K}_{s,t+1}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\ &= -\boldsymbol{\lambda}_g(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) - \boldsymbol{\lambda}_{s,t}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \qquad \text{(A.30)} \\ &= -\boldsymbol{\lambda}_g \mathbf{e}_{g,t+1} - \boldsymbol{\lambda}_{s,t}\mathbf{e}_{s,t+1} - (\boldsymbol{\lambda}_g + \boldsymbol{\lambda}_{s,t})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t), \qquad \text{(A.31)} \end{aligned}$$

where $\mathbf{I}_c = [1 \quad 0]$ and the last expression comes from (A.13). On the other hand, the log return, $r_{m,t+1}$, of the risky asset can be written as

$$\begin{aligned} r_{m,t+1} &\approx \kappa_{0d,t} + \kappa_{1d,t} z_{d,t+1} - z_{d,t} + g_{d,t+1} \\ &= \kappa_{0d,t} + \kappa_{1d,t} A_{0d,t+1} - A_{0d,t} + \kappa_{1d,t}\mathbf{A}_{1d,t+1}\hat{\boldsymbol{\mu}}_{t+1} - \mathbf{A}_{1d,t}\hat{\boldsymbol{\mu}}_t + g_{d,t+1} \\ &= \kappa_{0d,t} + \kappa_{1d,t} A_{0d,t+1} - A_{0d,t} + \kappa_{1d,t}\mathbf{A}_{1d,t+1}(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t) + (\kappa_{1d,t}\mathbf{A}_{1d,t+1} - \mathbf{A}_{1d,t})\hat{\boldsymbol{\mu}}_t + g_{d,t+1} \\ &= (\kappa_{0d,t} + \kappa_{1d,t} A_{0d,t+1} - A_{0d,t}) + (\psi^{-1}\mathbf{I}_c)\hat{\boldsymbol{\mu}}_t + \mathbf{I}_d(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \kappa_{1d,t}\mathbf{A}_{1d,t+1}(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t). \end{aligned}$$

Therefore, we have the innovation in $r_{m,t+1}$ as follows.

$$\begin{aligned} r_{m,t+1} - E[r_{m,t+1}|\mathcal{F}_t] &\approx \mathbf{I}_d(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \kappa_{1d,t}\mathbf{A}_{1d,t+1}\mathbf{K}_{s,t+1}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\ &= \boldsymbol{\beta}_g(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \boldsymbol{\beta}_{s,t}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \qquad \text{(A.32)} \\ &= \boldsymbol{\beta}_g \mathbf{e}_{g,t+1} + \boldsymbol{\beta}_{s,t}\mathbf{e}_{s,t+1} + (\boldsymbol{\beta}_g + \boldsymbol{\beta}_{s,t})(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t). \qquad \text{(A.33)} \end{aligned}$$

A.1.4 Decomposition of the Equity Premium

The innovation in the log of the SDF m_{t+1} conditional on the agent's information set is

$$\begin{aligned} m_{t+1} - E[m_{t+1}|\mathcal{F}_t] &= -\boldsymbol{\lambda}_g(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) - \boldsymbol{\lambda}_{s,t}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\ &= -\gamma \mathbf{I}_c(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) - \kappa^*_{1c,t}\left(\gamma - \frac{1}{\psi}\right)\mathbf{I}_c(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t), \qquad \text{(A.34)} \end{aligned}$$

where $\mathbf{I}_c = [1 \quad 0]$, $\kappa^*_{1c,t} = \kappa_{1c,t}/(1-\kappa_{1c,t}) > 0$, and the last expression comes from (A.1). Note the sign of the coefficient on the belief shock $(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t)$ depends on whether $\gamma > \frac{1}{\psi}$

or not. On the other hand, the innovation in the log of risky portfolio return ($r_{m,t+1}$) is

$$\begin{aligned} r_{m,t+1} - E[r_{m,t+1}|\mathcal{F}_t] &= \boldsymbol{\beta}_g(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \boldsymbol{\beta}_{s,t}(\mathbf{s}_{t+1} - \hat{\boldsymbol{\mu}}_t) \\ &= \mathbf{I}_d(\mathbf{g}_{t+1} - \hat{\boldsymbol{\mu}}_t) + \kappa^*_{1d,t}\begin{bmatrix} -\frac{1}{\psi} & 1 \end{bmatrix}(\hat{\boldsymbol{\mu}}_{t+1} - \hat{\boldsymbol{\mu}}_t), \end{aligned} \tag{A.35}$$

where $\boldsymbol{\beta}_g = \mathbf{I}_d = [0 \quad 1]$ and $\kappa^*_{1d,t} = \kappa_{1d,t}/(1-\kappa_{1d,t}) > 0$.[31] Since $(m_{t+1}, r_{m,t+1})$ follows a bivariate normal distribution, Euler equation for the risky asset $1 = E[M_{t+1}R_{m,t+1}|\mathcal{F}_t]$ becomes

$$0 = E[m_{t+1}|\mathcal{F}_t] + E[r_{m,t+1}|\mathcal{F}_t] + \frac{1}{2}Var(m_{t+1}|\mathcal{F}_t) + \frac{1}{2}Var(r_{m,t+1}|\mathcal{F}_t) + Cov(m_{t+1}, r_{m,t+1}|\mathcal{F}_t),$$

where the log risk-free rate is $r_{f,t} = -E[m_{t+1}|\mathcal{F}_t] - \frac{1}{2}Var(m_{t+1}|\mathcal{F}_t)$. Therefore, we can express the log-return counterpart of the equity risk premium $RP_t = -Cov(r_{m,t+1}, m_{t+1}|\mathcal{F}_t) = \boldsymbol{\beta}_t\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\lambda}_t$ where $\boldsymbol{\beta}_t = [\boldsymbol{\beta}_g \;\; \boldsymbol{\beta}_{s,t}]$ and $\boldsymbol{\lambda}_t = [\boldsymbol{\lambda}_g \;\; \boldsymbol{\lambda}_{s,t}]$, which is

$$\begin{aligned} E[r_{m,t+1}|\mathcal{F}_t] - r_{f,t} + \frac{1}{2}Var[r_{m,t+1}|\mathcal{F}_t] =& \underbrace{\gamma Cov(g_{c,t+1}, g_{d,t+1}|\mathcal{F}_t)}_{\text{from growth}} \\ &+ \underbrace{\kappa^*_{1c,t}\kappa^*_{1d,t}\left(\gamma - \frac{1}{\psi}\right)\begin{bmatrix} -\frac{1}{\psi} & 1 \end{bmatrix} Cov(\hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t)\mathbf{I}_c^\top}_{\text{from signal}} \\ &+ \underbrace{\kappa^*_{1d,t}\gamma\begin{bmatrix} -\frac{1}{\psi} & 1 \end{bmatrix} Cov(\hat{\boldsymbol{\mu}}_{t+1}, \mathbf{g}_{t+1}|\mathcal{F}_t)\mathbf{I}_c^\top}_{\text{from signal}\times\text{growth}} \\ &+ \underbrace{\kappa^*_{1c,t}\left(\gamma - \frac{1}{\psi}\right)\mathbf{I}_d Cov(\mathbf{g}_{t+1}, \hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t)\mathbf{I}_c^\top}_{\text{from growth}\times\text{signal}}. \end{aligned} \tag{A.36}$$

Note each of four terms here include static (co)-uncertainties from the last term of (15) in the main text. Also, all covariance expressions here refer to a 2×2 covariance matrix except for the first scalar term. The covariances in the third and fourth terms are, in particular, cross-covariance matrices; therefore, $Cov(\mathbf{g}_{t+1}, \hat{\boldsymbol{\mu}}_{t+1}|\mathcal{F}_t) = Cov(\hat{\boldsymbol{\mu}}_{t+1}, \mathbf{g}_{t+1}|\mathcal{F}_t)^\top$. Finally, we

[31] $\kappa^*_{1c,t}$ and $\kappa^*_{1d,t}$ are positive because $\kappa_{1c,t}$ and $\kappa_{1d,t}$ are the constants from log-linearization and lie between zero and one.

can rewrite the last three terms from (A.36) in scalar terms as follows.

$$RP_{ss} \triangleq \boldsymbol{\beta}_{s,t}\boldsymbol{\Sigma}^*_s\boldsymbol{\lambda}^\top_{s,t} = \kappa^*_{1c,t}\kappa^*_{1d,t}\left[\gamma - \frac{1}{\psi}\right] Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)\left[\frac{Cov(\hat{\mu}_{c,t+1},\hat{\mu}_{d,t+1}|\mathcal{F}_t)}{Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)} - \frac{1}{\psi}\right],$$
$$RP_{sg} \triangleq \boldsymbol{\beta}_{s,t}\mathbf{J}^*_{sg}\boldsymbol{\lambda}^\top_{g} = \kappa^*_{1d,t}\gamma\left[Cov(\hat{\mu}_{d,t+1},g_{c,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{\mu}_{c,t+1},g_{c,t+1}|\mathcal{F}_t)\right],$$
$$RP_{gs} \triangleq \boldsymbol{\beta}_{g}\mathbf{J}^*_{gs}\boldsymbol{\lambda}^\top_{s,t} = \kappa^*_{1c,t}\left[\gamma - \frac{1}{\psi}\right] Cov(\hat{\mu}_{c,t+1},g_{d,t+1}|\mathcal{F}_t).$$

A.1.5 Equity Volatility and Risk-free Rate

From (A.12) and (A.35), we can decompose the conditional variance of the risky asset's returns, $Var[r_{m,t+1}|\mathcal{F}_t] = \boldsymbol{\beta}_t\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\beta}^\top_t$, as follows.

$$\begin{aligned} Var[r_{m,t+1}|\mathcal{F}_t] =& \underbrace{Var(g_{d,t+1}|\mathcal{F}_t)}_{\text{from growth}} \\ &+ \underbrace{(\kappa^*_{1d,t})^2 Var\left(\hat{\mu}_{d,t+1} - \frac{1}{\psi}\hat{\mu}_{c,t+1}|\mathcal{F}_t\right)}_{\text{from signal}} \\ &+ \underbrace{\kappa^*_{1d,t}\left(Cov(\hat{\mu}_{d,t+1},g_{d,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{\mu}_{d,t+1},g_{c,t+1}|\mathcal{F}_t)\right)}_{\text{from signal}\times\text{growth}} \\ &+ \underbrace{\kappa^*_{1d,t}\left(Cov(\hat{\mu}_{d,t+1},g_{d,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{\mu}_{c,t+1},g_{d,t+1}|\mathcal{F}_t)\right)}_{\text{from growth}\times\text{signal}}. \end{aligned} \tag{A.37}$$

On the other hand, the log risk-free rate is

$$\begin{aligned} r_{f,t} &= -E[m_{t+1}|\mathcal{F}_t] - \frac{1}{2}Var(m_{t+1}|\mathcal{F}_t) \\ &= -\theta\log\delta - (\theta-1)\left(\kappa_{0c,t} + \kappa_{1c,t}A_{0c,t+1} - A_{0c,t}\right) + \frac{1}{\psi}\mathbf{I}_c\hat{\boldsymbol{\mu}}_t - \frac{1}{2}\boldsymbol{\lambda}_t\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\lambda}^\top_t \\ &= -\log\delta + \frac{1}{\psi}\hat{\mu}_{c,t} - \frac{1-\theta}{2\theta}\boldsymbol{\lambda}^*_t\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*_t\right)^\top - \frac{1}{2}\boldsymbol{\lambda}_t\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\lambda}^\top_t. \end{aligned} \tag{A.38}$$

Extending Table 1, we report the annualized conditional cumulative market return volatility, the risk-free rate, and the market Sharpe ratio in Table A.1. Similar to the conditional equity premium in Table 1, the annualized market return volatility and Sharpe ratios remain high, and the risk-free rate remains relatively low, even after 50 years of learning. The risk-free rate is the same in cases (ii) and (iii) since static uncertainty $Var(\mu_c|\mathcal{F}_t)$ are identical for (ii) and (iii).

Table A.1: Asset Moments and (Co)-Uncertainty in I.I.D. Growth Models

Correlation in	Annualized volatility (cumulative, %) $SD(R_m)$			Annualized risk-free rate (cumulative, %) $E(R_f)$			Annualized Sharpe ratio (cumulative, %) $E(R_m - R_f)/SD(R_m)$		
growth rates: ρ_g	0.550	0.000	0.000	0.550	0.000	0.000	0.550	0.000	0.000
aggregate signals: ρ_s	0.550	0.550	0.800	0.550	0.550	0.800	0.550	0.550	0.800
beliefs (implied): $\rho_{\mu,t}$	0.550	0.550	0.800	0.550	0.550	0.800	0.550	0.550	0.800
t (quarters)									
100	32.51	54.01	51.72	1.40	2.62	2.62	0.41	0.29	0.52
200	26.71	41.74	40.02	2.01	2.89	2.89	0.33	0.21	0.39
400	21.71	31.55	30.30	2.47	3.06	3.06	0.27	0.15	0.28
800	17.78	23.83	22.97	2.78	3.16	3.16	0.22	0.11	0.19
∞	9.90	9.90	9.90	3.27	3.27	3.27	0.12	0.00	0.00

A.2 Proof of Proposition 2

We first define $\sigma_{scd} = \rho_s \sigma_{sc}\sigma_{sd}$ and $\sigma_{cd,t} = \rho_{\mu,t}\sigma_{c,t}\sigma_{d,t}$, and then show the following inequality.

$$\begin{aligned}
(\sigma_{scd} + \sigma_{cd,t})^2 &= (\rho_s \sigma_{sc}\sigma_{sd} + \rho_{\mu,t}\sigma_{c,t}\sigma_{d,t})^2 \\
&< (\sigma_{sc}\sigma_{sd} + \sigma_{c,t}\sigma_{d,t})^2 = \sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{sc}^2\sigma_{sd}^2 + 2\sigma_{sc}\sigma_{d,t}\sigma_{sd}\sigma_{c,t} \\
&< \sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{sc}^2\sigma_{sd}^2 + 2\sigma_{sc}\sigma_{d,t}\sigma_{sd}\sigma_{c,t} + (\sigma_{sc}\sigma_{d,t} - \sigma_{sd}\sigma_{c,t})^2 \\
&= \sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{sc}^2\sigma_{sd}^2 + (\sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2). \qquad \text{(A.39)}
\end{aligned}$$

Next, we prove Proposition 2 as follows.

$$\begin{aligned}
0 &< cov(\hat{\mu}_{c,t+1}, \hat{\mu}_{d,t+1}|\mathcal{F}_t) \\
&= cov(\mu_c, \mu_d|\mathcal{F}_t) - cov(\mu_c, \mu_d|\mathcal{F}_{t+1}) \qquad (\because \text{Proposition 1}) \\
&= \sigma_{cd,t} - \frac{\sigma_{cd,t}^2\sigma_{scd} + \sigma_{cd,t}\sigma_{scd}^2 - \sigma_{sc}^2\sigma_{sd}^2\sigma_{cd,t} - \sigma_{c,t}^2\sigma_{d,t}^2\sigma_{scd}}{-\sigma_{scd}^2 - \sigma_{cd,t}^2 - 2\sigma_{scd}\sigma_{cd,t} + \sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2 + \sigma_{sc}^2\sigma_{sd}^2}.
\end{aligned}$$

(The last step is done by brute force.)

$$\begin{aligned}
\iff 0 &< \frac{-(\sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{cd,t}^2)\sigma_{scd} + (\sigma_{c,t}^2\sigma_{d,t}^2 - \sigma_{cd,t}^2 + \sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2)\sigma_{cd,t}}{-(\sigma_{scd} + \sigma_{cd,t})^2 + \sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2 + \sigma_{sc}^2\sigma_{sd}^2} \\
\iff 0 &< -(\sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{cd,t}^2)\sigma_{scd} + (\sigma_{c,t}^2\sigma_{d,t}^2 - \sigma_{cd,t}^2 + \sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2)\sigma_{cd,t}, \qquad (\because \text{Equation A.39}) \\
\iff \sigma_{scd} &< \frac{\sigma_{c,t}^2\sigma_{d,t}^2 - \sigma_{cd,t}^2 + \sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2}{\sigma_{c,t}^2\sigma_{d,t}^2 + \sigma_{cd,t}^2}\sigma_{cd,t}, \\
&= \frac{\sigma_{c,t}^2\sigma_{d,t}^2(1 - \rho_{\boldsymbol{\mu},t}^2) + \sigma_{sc}^2\sigma_{d,t}^2 + \sigma_{sd}^2\sigma_{c,t}^2}{\sigma_{c,t}^2\sigma_{d,t}^2(1 + \rho_{\boldsymbol{\mu},t}^2)}\sigma_{cd,t}, \\
&= \left((1 - \rho_{\boldsymbol{\mu},t}^2) + \frac{\sigma_{sc}^2}{\sigma_{c,t}^2} + \frac{\sigma_{sd}^2}{\sigma_{d,t}^2}\right)\frac{\sigma_{cd,t}}{1 + \rho_{\boldsymbol{\mu},t}^2}, \\
\iff \rho_s &< \rho_{\boldsymbol{\mu},t}\left((1 - \rho_{\boldsymbol{\mu},t}^2) + \frac{\sigma_{sc}^2}{\sigma_{c,t}^2} + \frac{\sigma_{sd}^2}{\sigma_{d,t}^2}\right)\frac{\sigma_{c,t}\sigma_{d,t}}{\sigma_{sc}\sigma_{sd}}\frac{1}{1 + \rho_{\boldsymbol{\mu},t}^2}.
\end{aligned}$$

A.3 Signal noisiness and intertemporal co-uncertainty

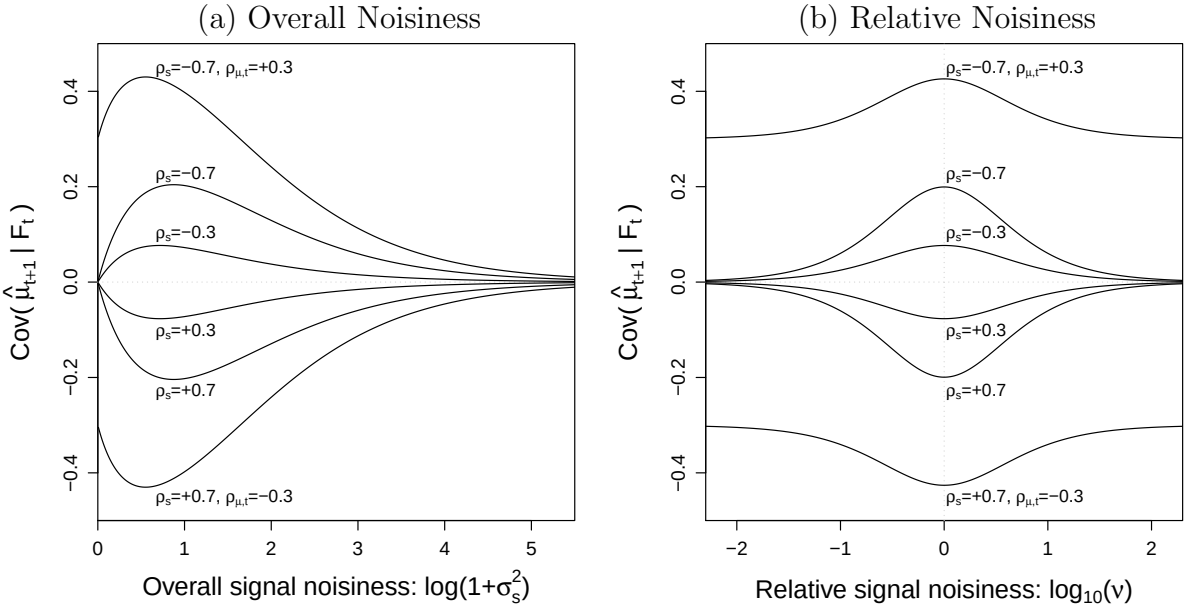
In addition to signal correlations, the noisiness of signals also affects intertemporal co-uncertainty. Consider a bivariate normal signal $\mathbf{s}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_s)$ on an unknown bivariate vector $\boldsymbol{\mu}$, where $\boldsymbol{\Sigma}_s$ is given by Equation (40). Then the signal is characterized by

1. (Signal correlation) $= \rho_s$
2. (Overall signal noisiness) $= \sigma_s$
3. (Relative signal noisiness) $= \nu$

Figure A.1 shows how intertemporal co-uncertainty is related to overall noisiness in Panel (a) and relative noisiness in Panel (b). The non-monotonic pattern in Panel (a) is consistent with Figure 2 Panel (b).

Figure A.1: Intertemporal Co-Uncertainty vs. Signal Noisiness

Prior static uncertainty, $var(\mu_c)$ and $var(\mu_d)$, are unity, and prior static co-uncertainty $\rho_{\mu,t}$, *in terms of correlation*, is zero if not stated otherwise. Noisiness is expressed as log-scales.



A.4 Detection Error Probability

Table A.2 reports detection error probabilities (DEP) in the calibration of parameters in a long-run risk model where DEP $= \frac{1}{2}\left[\Pr(\text{choose A} \mid \text{O is true}) + \Pr(\text{choose O} \mid \text{A is true})\right]$. In short, the DEP in this table measures intrinsic estimation uncertainty in parameters in terms of average Type-I errors from the original (O) and alternative (A) model calibrations. First, we set all parameters consistent with Bansal and Yaron (2004). Panel A shows the DEP associated with $\phi_{xd} = \frac{Cov(x_{c,t}, x_{d,t}|x_{t-1})}{Var(x_{c,t}|x_{t-1})}$ while we fix $\chi_{xd} = \frac{\sigma_{xd}}{\sigma_{xc}} = \frac{SD(x_{d,t}|x_{t-1})}{SD(x_{c,t}|x_{t-1})}$ at 3. The numbers in Panel A are the DEP of the long-run risk model with $\phi_{xd} = 3$ against other values of ϕ_{xd} ranging from -1 to 2.5. Panel B shows the DEP associated with χ_{xd} while ϕ_{xd} is fixed at 3. The numbers in Panel B are the DEP of the long-run risk model with $\chi_{xd} = 3$ against other values of χ_{xd} ranging from 3.5 to 7.

Table A.2: Detection Error Probabilities in Long-run Risk Models

Panel A: Detection error probabilities associated with ϕ_{xd}									
Parameter to compare	Baseline	Parameter values in alternative calibrations							
ϕ_{xd}	3.0	2.5	2.0	1.5	1.0	0.5	0.0	-0.5	-1.0
Sample size (months)		Detection error probability (%)							
600	50	40.2	32.8	26.9	22.0	17.9	14.4	11.7	9.2
840	50	38.3	29.7	23.0	17.7	13.4	10.2	7.7	5.6
1200	50	35.9	26.0	18.6	13.1	9.1	6.3	4.2	2.8
Panel B: Detection error probabilities associated with χ_{xd}									
Parameter to compare	Baseline	Parameter values in alternative calibrations							
χ_{xd}	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0
Sample size (months)		Detection error probability (%)							
600	50	43.2	36.0	29.7	24.2	19.4	15.4	12.0	9.3
840	50	41.7	33.6	26.4	20.3	15.3	11.3	8.2	5.9
1200	50	40.1	30.6	22.4	15.9	11.1	7.4	4.7	3.0

A.5 Solutions for Long-Run Risk Models

In this section, we solve a more general specification for the long-run risk model than the one we analyze in Section 4. The resulting solutions, thus, can be applied to the case of a combined signal in (41). We do not report here, yet the asset pricing implications of combined signals are very similar to those of negatively correlated signals, as expected from Section 5.3. The solution approach here is identical to Section A.1; therefore, we recycle notations and often omit repetitive steps. Throughout the section, we use boldface symbols for non-scalar vectors and matrices. The dynamics of hidden long-run components $\mathbf{x}_t = [x_{c,t} \;\; x_{d,t}]^\top$, aggregate signals $\mathbf{s}_t = [s_{c,t} \;\; s_{d,t}]^\top$, and realized growth shocks $\mathbf{g}_t = [g_{c,t} \;\; g_{d,t}]^\top$ are as follows.

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{F}\mathbf{x}_t + \mathbf{e}_{x,t+1}, \\ \mathbf{s}_{t+1} &= \mathbf{H}\mathbf{x}_t + \mathbf{e}_{s,t+1}, \\ \mathbf{g}_{t+1} &= \boldsymbol{\mu} + \mathbf{x}_t + \mathbf{e}_{g,t+1}, \end{aligned} \tag{A.40}$$

$$\text{where} \quad \begin{bmatrix} \mathbf{e}_{x,t+1} \\ \mathbf{e}_{g,t+1} \\ \mathbf{e}_{s,t+1} \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{xgs}\right), \quad \boldsymbol{\Sigma}_{xgs} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \mathbf{J}_{xg} & \mathbf{J}_{xs} \\ \mathbf{J}_{xg}^\top & \boldsymbol{\Sigma}_g & \mathbf{J}_{gs} \\ \mathbf{J}_{xs}^\top & \mathbf{J}_{gs}^\top & \boldsymbol{\Sigma}_s \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma}_{gs} = \begin{bmatrix} \boldsymbol{\Sigma}_g & \mathbf{J}_{gs} \\ \mathbf{J}_{gs}^\top & \boldsymbol{\Sigma}_s \end{bmatrix}. \tag{A.41}$$

Note our solutions here are valid even with general specifications of $\mathbf{F}$ and $\mathbf{H}$. At each decision point, the representative agent knows all the parameters except $\mathbf{x}^t$ and makes Bayesian inference on it conditional on the history of signals. By solving a Kalman filter, the agent's posterior belief on $\mathbf{x}_t$ at t converge to the stationary process:

$$\mathbf{x}_t|\mathcal{F}_t \sim \mathcal{N}\left(\hat{\mathbf{x}}_t, \boldsymbol{\Pi}\right) \quad \text{and} \quad \hat{\mathbf{x}}_t = \mathbf{F}\hat{\mathbf{x}}_{t-1} + \mathbf{K}_s(\mathbf{s}_t - \mathbf{H}\hat{\mathbf{x}}_{t-1}), \tag{A.42}$$

where $\mathbf{K}_s = (\mathbf{F}\boldsymbol{\Pi}\mathbf{H}^\top + \mathbf{J}_{xs})(\mathbf{H}\boldsymbol{\Pi}\mathbf{H}^\top + \boldsymbol{\Sigma}_s)^{-1}$ is the Kalman gain matrix and $\boldsymbol{\Pi}$ is the conditional covariance matrix of $\mathbf{x}_t$ under the agent's posterior beliefs after the Kalman filter converges to the steady state. Note this distribution equals a Bayesian posterior distribution $p(\mathbf{x}_t|\mathbf{s}^t)$ since $\mathcal{F}_t = \mathbf{s}^t$. To find the conditional covariance matrix $\boldsymbol{\Pi}$ at the steady state, we first define $\boldsymbol{\Pi}_t = Cov(\mathbf{x}_t|\mathbf{s}^t)$ and $\mathbf{K}_{s,t} = (\mathbf{F}\boldsymbol{\Pi}_t\mathbf{H}^\top + \mathbf{J}_{xs})(\mathbf{H}\boldsymbol{\Pi}_t\mathbf{H}^\top + \boldsymbol{\Sigma}_s)^{-1}$. Then apply

Proposition 1 to the unknown $\mathbf{x}_{t+1}$ as follows.

$$Cov(\hat{\mathbf{x}}_{t+1}|\mathbf{s}^t) = Cov(\mathbf{x}_{t+1}|\mathbf{s}^t) - E[Cov(\mathbf{x}_{t+1}|\mathbf{s}^{t+1})|\mathbf{s}^t],$$

where the left-hand side equals $\mathbf{K}_{s,t}(\mathbf{H}\boldsymbol{\Pi}_t\mathbf{H}^\top + \boldsymbol{\Sigma}_s)\mathbf{K}_{s,t}^\top$ due to (A.42) and the right-hand side equals $(\mathbf{F}\boldsymbol{\Pi}_t\mathbf{F}^\top + \boldsymbol{\Sigma}_x) - \boldsymbol{\Pi}_{t+1}$. Therefore, we find $\boldsymbol{\Pi}$ by computing the following equation recursively until $\boldsymbol{\Pi}_t$ converges to $\boldsymbol{\Pi}$.

$$\boldsymbol{\Pi}_{t+1} = \mathbf{F}\boldsymbol{\Pi}_t\mathbf{F}^\top + \boldsymbol{\Sigma}_x - \mathbf{K}_{s,t}(\mathbf{H}\boldsymbol{\Pi}_t\mathbf{H}^\top + \boldsymbol{\Sigma}_s)\mathbf{K}_{s,t}^\top. \tag{A.43}$$

Note this equation coincides with a standard Kalman filter recursion equation, and we consider the state variable $\hat{\mathbf{x}}_t$ as *predictive* in that $E[\mathbf{g}_{t+1}|\mathcal{F}_t] = \boldsymbol{\mu} + \hat{\mathbf{x}}_t$ and $E[\mathbf{s}_{t+1}|\mathcal{F}_t] = \mathbf{H}\hat{\mathbf{x}}_t$.

A.5.1 Wealth-Consumption Ratio

To solve our long-run risk model, we apply the same approach used for the i.i.d. model in Section A.1. The log wealth-consumption ratio $z_{c,t}$ is a linear function of the state variable $\hat{\mathbf{x}}_t$ as follows.

$$z_{c,t} = \log\frac{W_t}{C_t} = A_{0c} + \mathbf{A}_{1c}\hat{\mathbf{x}}_t. \tag{A.44}$$

Next, we approximate the log return of the total wealth, $r_{c,t+1} = \log R_{c,t+1}$, as

$$\begin{aligned} r_{c,t+1} &= \log\left(\exp(z_{c,t+1}) + 1\right) - z_{c,t} + g_{c,t+1}, \\ &\approx \kappa_{0c} + \kappa_{1c}z_{c,t+1} - z_{c,t} + g_{c,t+1}, \end{aligned} \tag{A.45}$$

where κ_{0c} and κ_{1c} are Taylor approximation coefficients at $\bar{z}_c = A_{0c}$. For concise notation, we first define the following terms

$$\boldsymbol{\lambda}_g = \gamma\mathbf{I}_c = \gamma[1 \;\; 0] \quad \text{and} \quad \boldsymbol{\lambda}_s = (1-\theta)\kappa_{1c}\mathbf{A}_{1c}\mathbf{K}_s, \tag{A.46}$$

and their scaled versions

$$\boldsymbol{\lambda}_g^* = \frac{1-\gamma}{\gamma}\boldsymbol{\lambda}_g \quad \text{and} \quad \boldsymbol{\lambda}_s^* = \frac{\theta}{1-\theta}\boldsymbol{\lambda}_s. \tag{A.47}$$

Then, the time-varying-parameter version of (A.9) is

$$\begin{aligned}\log\left(G_{c,t+1}^{-\frac{1}{\psi}}R_{c,t+1}\right) &= (\kappa_{0c}+\kappa_{1c}A_{0c}-A_{0c}) + (1-\psi^{-1})\mathbf{I}_c\boldsymbol{\mu} \\ &\quad + \left(\kappa_{1c}\mathbf{A}_{1c}\mathbf{F} - \mathbf{A}_{1c} + (1-\psi^{-1})\mathbf{I}_c\right)\hat{\mathbf{x}}_t \\ &\quad + (1-\psi^{-1})\mathbf{I}_c(\mathbf{g}_{t+1}-\boldsymbol{\mu}-\hat{\mathbf{x}}_t) + \kappa_{1c}\mathbf{A}_{1c}(\hat{\mathbf{x}}_{t+1}-\mathbf{F}\hat{\mathbf{x}}_t),\end{aligned} \tag{A.48}$$

where the last two terms multiplied by θ are expressed as

$$\theta(1-\psi^{-1})\mathbf{I}_c(\mathbf{g}_{t+1}-\boldsymbol{\mu}-\hat{\mathbf{x}}_t) = \boldsymbol{\lambda}_g^*(\mathbf{g}_{t+1}-E[\mathbf{g}_{t+1}|\mathcal{F}_t]), \tag{A.49}$$

$$\theta\kappa_{1c}\mathbf{A}_{1c}(\hat{\mathbf{x}}_{t+1}-\mathbf{F}\hat{\mathbf{x}}_t) = \boldsymbol{\lambda}_s^*(\mathbf{s}_{t+1}-\mathbf{H}\hat{\mathbf{x}}_t) = \boldsymbol{\lambda}_s^*(\mathbf{s}_{t+1}-E[\mathbf{s}_{t+1}|\mathcal{F}_t]). \tag{A.50}$$

Note a vector of random variables $(\mathbf{g}_{t+1}-E[\mathbf{g}_{t+1}|\mathcal{F}_t])$ and $(\mathbf{s}_{t+1}-E[\mathbf{s}_{t+1}|\mathcal{F}_t])$ follows a multivariate normal distribution under the agent's posterior belief:

$$\begin{bmatrix}\mathbf{g}_{t+1}-E[\mathbf{g}_{t+1}|\mathcal{F}_t]\\ \mathbf{s}_{t+1}-E[\mathbf{s}_{t+1}|\mathcal{F}_t]\end{bmatrix}|\mathcal{F}_t \sim \mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}_{gs}^*\right) \quad \text{where} \quad \boldsymbol{\Sigma}_{gs}^* = \begin{bmatrix}\boldsymbol{\Sigma}_g^* & \mathbf{J}_{gs}^*\\ \mathbf{J}_{sg}^* & \boldsymbol{\Sigma}_s^*\end{bmatrix}. \tag{A.51}$$

Here, we define new matrices such as $\boldsymbol{\Sigma}_g^* = \boldsymbol{\Sigma}_g + \boldsymbol{\Pi}$, $\boldsymbol{\Sigma}_s^* = \boldsymbol{\Sigma}_s + \mathbf{H}\boldsymbol{\Pi}\mathbf{H}^\top$, $\mathbf{J}_{gs}^* = \mathbf{J}_{gs} + \boldsymbol{\Pi}\mathbf{H}^\top$, and $\mathbf{J}_{sg}^* = (\mathbf{J}_{gs}^*)^\top$, all of which include static (co)-uncertainty matrix $\boldsymbol{\Pi}$ because of the following decompositions:

$$\mathbf{g}_{t+1}-E[\mathbf{g}_{t+1}|\mathcal{F}_t] = (\mathbf{g}_{t+1}-\boldsymbol{\mu}-\mathbf{x}_t) + (\mathbf{x}_t-\hat{\mathbf{x}}_t), \tag{A.52}$$

$$\mathbf{s}_{t+1}-E[\mathbf{s}_{t+1}|\mathcal{F}_t] = (\mathbf{s}_{t+1}-\mathbf{H}\mathbf{x}_t) + \mathbf{H}(\mathbf{x}_t-\hat{\mathbf{x}}_t), \tag{A.53}$$

where $(\mathbf{x}_t-\hat{\mathbf{x}}_t) = \mathbf{u}_{\mathbf{x},t} \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Pi})$ is a random variable that represents static (co)-uncertainties of $\mathbf{x}_t$ under the agent's posterior belief. Finally, we plug (A.49)–(A.51) to the Euler equation (A.6) and solve the expectation using a log-normal distribution. With $\boldsymbol{\lambda}^* = [\boldsymbol{\lambda}_g^* \;\; \boldsymbol{\lambda}_s^*]$ and an

identity matrix $\mathbf{I}$, we obtain the following equations:

$$\mathbf{A}_{1c} = \left(1 - \frac{1}{\psi}\right) \mathbf{I}_c (\mathbf{I} - \kappa_{1c}\mathbf{F})^{-1}, \tag{A.54}$$

$$A_{0c} = \frac{1}{1-\kappa_{1c}} \left(\log\delta + \kappa_{0c} + (1-\psi^{-1})\mu_c + \frac{1}{2\theta} \boldsymbol{\lambda}^* \boldsymbol{\Sigma}^*_{gs} \left(\boldsymbol{\lambda}^*\right)^\top \right). \tag{A.55}$$

A.5.2 Price-Dividend Ratio

Let $R_{m,t+1} = \frac{P_{t+1}+D_{t+1}}{P_t}$ denote the gross return on the risky asset that is priced at P_t and pays out the aggregate dividend stream $\{D_t\}$. Then, the log price-dividend ratio is

$$z_{d,t} = \log \frac{P_t}{D_t} = A_{0d} + \mathbf{A}_{1d}\hat{\mathbf{x}}_t, \tag{A.56}$$

where A_{0d} contains the static (co)-uncertainty term $\boldsymbol{\Pi}$. Next, the log return of the risky asset, $r_{m,t+1} = \log R_{m,t+1}$, is approximated as

$$r_{m,t+1} \approx \kappa_{0d} + \kappa_{1d} z_{d,t+1} - z_{d,t} + g_{d,t+1}, \tag{A.57}$$

where κ_{0d} and κ_{1d} are Taylor approximation coefficients at $\bar{z}_d = A_{0d}$. For concise notation, we first define the following terms

$$\boldsymbol{\beta}_g = \mathbf{I}_d = [0 \;\; 1] \quad \text{and} \quad \boldsymbol{\beta}_s = \kappa_{1d}\mathbf{A}_{1d}\mathbf{K}_s. \tag{A.58}$$

Then, by plugging $R_{i,t+1} = R_{m,t+1}$ into the Euler equation (A.6), we first compute the time-varying-parameter version of (A.23) as follows.

$$\begin{aligned}
\log\left(G_{c,t+1}^{-\frac{\theta}{\psi}} R_{c,t+1}^{\theta-1} R_{m,t+1}\right) = {} & (\theta-1)(\kappa_{0c} + \kappa_{1c}A_{0c} - A_{0c}) + (\kappa_{0d} + \kappa_{1d}A_{0d} - A_{0d}) \\
& + (\mathbf{I}_d - \boldsymbol{\lambda}_g)\boldsymbol{\mu} \\
& + (\mathbf{I}_d - \boldsymbol{\lambda}_g + (\theta-1)(\kappa_{1c}\mathbf{A}_{1c}\mathbf{F} - \mathbf{A}_{1c}) + \kappa_{1d}\mathbf{A}_{1d}\mathbf{F} - \mathbf{A}_{1d})\,\hat{\mathbf{x}}_t \\
& + (\mathbf{I}_d - \boldsymbol{\lambda}_g)(\mathbf{g}_{t+1} - \boldsymbol{\mu} - \hat{\mathbf{x}}_t) \\
& + [(\theta-1)\kappa_{1c}\mathbf{A}_{1c} + \kappa_{1d}\mathbf{A}_{1d}]\,(\hat{\mathbf{x}}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_t).
\end{aligned} \tag{A.59}$$

Note the last term can be simplified, by (A.42), to

$$[(\theta-1)\kappa_{1c}\mathbf{A}_{1c}+\kappa_{1d}\mathbf{A}_{1d}]\,(\hat{\mathbf{x}}_{t+1}-\mathbf{F}\hat{\mathbf{x}}_t)=(\boldsymbol{\beta}_s-\boldsymbol{\lambda}_s)(\mathbf{s}_{t+1}-\mathbf{H}\hat{\mathbf{x}}_t). \tag{A.60}$$

Finally, plug (A.59) and (A.60) to the Euler equation (A.6) and apply $\kappa_{1c}\mathbf{A}_{1c}\mathbf{F}-\mathbf{A}_{1c}=(\psi^{-1}-1)\mathbf{I}_c$ from (A.54). Then we have the following condition since (A.59) should be satisfied regardless of $\hat{\mathbf{x}}_t$.

$$\begin{aligned}0&=\mathbf{I}_d-\boldsymbol{\lambda}_g+(\theta-1)(\kappa_{1c}\mathbf{A}_{1c}\mathbf{F}-\mathbf{A}_{1c})+\kappa_{1d}\mathbf{A}_{1d}\mathbf{F}-\mathbf{A}_{1d} &\text{(A.61)}\\ &=\mathbf{I}_d-\psi^{-1}\mathbf{I}_c+\kappa_{1d}\mathbf{A}_{1d}\mathbf{F}-\mathbf{A}_{1d}. &\text{(A.62)}\end{aligned}$$

The remaining terms in the Euler equation should be zero as well, and so we have

$$\begin{aligned}A_{0d}&=\theta\log\delta+(\theta-1)(\kappa_{0c}+\kappa_{1c}A_{0c}-A_{0c})+\kappa_{0d}+\kappa_{1d}A_{0d}+(\mathbf{I}_d-\boldsymbol{\lambda}_g)\boldsymbol{\mu}+\frac{1}{2}\boldsymbol{\beta}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\beta}^*\right)^\top\\ &=\log\delta+\kappa_{0d}+\kappa_{1d}A_{0d}+(\mathbf{I}_d-\psi^{-1}\mathbf{I}_c)\boldsymbol{\mu}+\frac{1-\theta}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top+\frac{1}{2}\boldsymbol{\beta}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\beta}^*\right)^\top, \quad \text{(A.63)}\end{aligned}$$

where the last equation comes from (A.55) and $\boldsymbol{\beta}^*=\begin{bmatrix}(\boldsymbol{\beta}_g-\boldsymbol{\lambda}_g) & (\boldsymbol{\beta}_s-\boldsymbol{\lambda}_s)\end{bmatrix}$. Therefore, we have

$$\mathbf{A}_{1d}=(\mathbf{I}_d-\psi^{-1}\mathbf{I}_c)(\mathbf{I}-\kappa_{1d}\mathbf{F})^{-1} \tag{A.64}$$

$$A_{0d}=\frac{1}{1-\kappa_{1d}}\left(\log\delta+\kappa_{0d}+(\mathbf{I}_d-\psi^{-1}\mathbf{I}_c)\boldsymbol{\mu}+\frac{1-\theta}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top+\frac{1}{2}\boldsymbol{\beta}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\beta}^*\right)^\top\right). \tag{A.65}$$

A.5.3 Stochastic Discount Factor and Equity Returns

Using the results from Section (A.5.1), we express the log of the stochastic discount factor as follows, similarly to (A.29).

$$\begin{aligned}
m_{t+1} \approx & \theta \log \delta - \frac{\theta}{\psi} g_{c,t+1} + (\theta - 1)\left(\kappa_{0c} + \kappa_{1c}(A_{0c} + \mathbf{A}_{1c}\hat{\mathbf{x}}_{t+1}) - A_{0c} - \mathbf{A}_{1c}\hat{\mathbf{x}}_t + g_{c,t+1}\right) \\
= & \theta \log \delta + (\theta - 1)\left(\kappa_{0c} + \kappa_{1c}A_{0c} - A_{0c}\right) + (\theta - 1)(\kappa_{1c}\mathbf{A}_{1c}\mathbf{F} - \mathbf{A}_{1c})\hat{\mathbf{x}}_t \\
& - \left(\frac{\theta}{\psi} + 1 - \theta\right) g_{c,t+1} + (\theta - 1)\kappa_{1c}\mathbf{A}_{1c}(\hat{\mathbf{x}}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_t) \\
= & \theta \log \delta + (\theta - 1)\left(\kappa_{0c} + \kappa_{1c}A_{0c} - A_{0c}\right) - \gamma \mathbf{I}_c \boldsymbol{\mu} \\
& + \left(-\gamma + (\theta - 1)(\psi^{-1} - 1)\right)\mathbf{I}_c\hat{\mathbf{x}}_t \\
& - \gamma \mathbf{I}_c(\mathbf{g}_{t+1} - \boldsymbol{\mu} - \hat{\mathbf{x}}_t) - (1 - \theta)\kappa_{1c}\mathbf{A}_{1c}\mathbf{K}_s(\mathbf{s}_{t+1} - \mathbf{H}\hat{\mathbf{x}}_t).
\end{aligned}$$

Therefore, the innovation in m_{t+1} conditional on the agent's information set is

$$\begin{aligned}
m_{t+1} - E[m_{t+1}|\mathcal{F}_t] &\approx -\boldsymbol{\lambda}_g(\mathbf{g}_{t+1} - E[\mathbf{g}_{t+1}|\mathcal{F}_t]) - \boldsymbol{\lambda}_s(\mathbf{s}_{t+1} - E[\mathbf{s}_{t+1}|\mathcal{F}_t])) \\
&= -\boldsymbol{\lambda}_g \mathbf{e}_{g,t+1} - \boldsymbol{\lambda}_s \mathbf{e}_{s,t+1} - (\boldsymbol{\lambda}_g + \boldsymbol{\lambda}_s \mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_t), \qquad \text{(A.66)}
\end{aligned}$$

where $\mathbf{I}_c = [1 \quad 0]$ and the last expression comes from (A.53). On the other hand, the log return, $r_{m,t+1}$, of the risky asset can be written as

$$\begin{aligned}
r_{m,t+1} &\approx \kappa_{0d} + \kappa_{1d} z_{d,t+1} - z_{d,t} + g_{d,t+1} \\
&= \kappa_{0d} + \kappa_{1d}A_{0d} - A_{0d} + \kappa_{1d}\mathbf{A}_{1d}\hat{\boldsymbol{\mu}}_{t+1} - \mathbf{A}_{1d}\hat{\boldsymbol{\mu}}_t + g_{d,t+1} \\
&= \kappa_{0d} + \kappa_{1d}A_{0d} - A_{0d} + \kappa_{1d}\mathbf{A}_{1d}(\hat{\mathbf{x}}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_t) + (\kappa_{1d}\mathbf{A}_{1d}\mathbf{F} - \mathbf{A}_{1d})\hat{\mathbf{x}}_t + g_{d,t+1} \\
&= (\kappa_{0d} + \kappa_{1d}A_{0d} - A_{0d}) + \mathbf{I}_d\boldsymbol{\mu} + (\psi^{-1}\mathbf{I}_c)\hat{\mathbf{x}}_t + \mathbf{I}_d(\mathbf{g}_{t+1} - \boldsymbol{\mu} - \hat{\mathbf{x}}_t) + \kappa_{1d}\mathbf{A}_{1d}(\hat{\mathbf{x}}_{t+1} - \mathbf{F}\hat{\mathbf{x}}_t).
\end{aligned}$$

Therefore, we have the innovation in $r_{m,t+1}$ as follows.

$$\begin{aligned}
r_{m,t+1} - E[r_{m,t+1}|\mathcal{F}_t] &\approx \boldsymbol{\beta}_g(\mathbf{g}_{t+1} - E[\mathbf{g}_{t+1}|\mathcal{F}_t]) + \boldsymbol{\beta}_s(\mathbf{s}_{t+1} - E[\mathbf{s}_{t+1}|\mathcal{F}_t]) \qquad \text{(A.67)} \\
&= \boldsymbol{\beta}_g \mathbf{e}_{g,t+1} + \boldsymbol{\beta}_s \mathbf{e}_{s,t+1} + (\boldsymbol{\beta}_g + \boldsymbol{\beta}_s \mathbf{H})(\mathbf{x}_t - \hat{\mathbf{x}}_t). \qquad \text{(A.68)}
\end{aligned}$$

A.5.4 Decomposition of the Equity Premium

To compare with the i.i.d. model, we decompose the equity premium for the specification in Section 4. That is, we have

$$\mathbf{F} = \begin{bmatrix} F_x & 0 \\ 0 & F_x \end{bmatrix} \quad \text{and} \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{A.69}$$

Then we can simply $\mathbf{A}_{1c}$ and $\mathbf{A}_{1d}$ as follows.

$$\mathbf{A}_{1c} = \frac{1}{1-\kappa_{1c}F_x}\begin{bmatrix} \left(1-\psi^{-1}\right) & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{A}_{1d} = \frac{1}{1-\kappa_{1d}F_x}\begin{bmatrix} -\psi^{-1} & 1 \end{bmatrix}. \tag{A.70}$$

The innovation in the log of the SDF m_{t+1} conditional on the agent's information set is

$$\begin{aligned} m_{t+1} - E[m_{t+1}|\mathcal{F}_t] &\approx -\boldsymbol{\lambda}_g(\mathbf{g}_{t+1} - E[\mathbf{g}_{t+1}|\mathcal{F}_t]) - \boldsymbol{\lambda}_s(\mathbf{s}_{t+1} - E[\mathbf{s}_{t+1}|\mathcal{F}_t])) \\ &= -\gamma\mathbf{I}_c(\mathbf{g}_{t+1} - E[\mathbf{g}_{t+1}|\mathcal{F}_t]) - \kappa_{1c}^{**}\left(\gamma - \frac{1}{\psi}\right)\mathbf{I}_c(\hat{\mathbf{x}}_{t+1} - E[\hat{\mathbf{x}}_{t+1}|\mathcal{F}_t]), \end{aligned}$$

where $\mathbf{I}_c = [1 \;\; 0]$, $\kappa_{1c}^{**} = \kappa_{1c}/(1-\kappa_{1c}F_x)$, and the last expression comes from (A.42). On the other hand, the innovation in the log of market portfolio return ($r_{m,t+1}$) is

$$\begin{aligned} r_{m,t+1} - E[r_{m,t+1}|\mathcal{F}_t] &\approx \boldsymbol{\beta}_g(\mathbf{g}_{t+1} - E[\mathbf{g}_{t+1}|\mathcal{F}_t]) + \boldsymbol{\beta}_s(\mathbf{s}_{t+1} - E[\mathbf{s}_{t+1}|\mathcal{F}_t]) \\ &= \mathbf{I}_d(\mathbf{g}_{t+1} - E[\mathbf{g}_{t+1}|\mathcal{F}_t]) + \kappa_{1d}^{**}\begin{bmatrix} -\frac{1}{\psi} & 1 \end{bmatrix}(\hat{\mathbf{x}}_{t+1} - E[\hat{\mathbf{x}}_{t+1}|\mathcal{F}_t]), \end{aligned} \tag{A.71}$$

where $\boldsymbol{\beta}_g = \mathbf{I}_d = [0 \;\; 1]$ and $\kappa_{1d}^{**} = \kappa_{1d}/(1-\kappa_{1d}F_x)$.[32] Therefore, we can express the log-return counterpart of the equity risk premium $RP_t = -Cov(r_{m,t+1}, m_{t+1}|\mathcal{F}_t) = \boldsymbol{\beta}\boldsymbol{\Sigma}_{gs}^*\boldsymbol{\lambda}$ where

[32] κ_{1c}^{**} and κ_{1d}^{**} are positive because κ_{1c}, κ_{1d}, and F_x lie between zero and one.

$\boldsymbol{\beta} = [\boldsymbol{\beta}_g \;\; \boldsymbol{\beta}_s]$ and $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_g \;\; \boldsymbol{\lambda}_s]$, which is

$$\begin{aligned} E[r_{m,t+1}|\mathcal{F}_t] - r_{f,t} + \frac{1}{2}Var[r_{m,t+1}|\mathcal{F}_t] =& \underbrace{\gamma Cov(g_{c,t+1}, g_{d,t+1}|\mathcal{F}_t)}_{\text{from growth}} \\ &+ \underbrace{\kappa_{1c}^{**}\kappa_{1d}^{**}\left(\gamma - \frac{1}{\psi}\right)\begin{bmatrix} -\frac{1}{\psi} & 1 \end{bmatrix} Cov(\hat{\mathbf{x}}_{t+1}|\mathcal{F}_t)\mathbf{I}_c^\top}_{\text{from signal}} \\ &+ \underbrace{\kappa_{1d}^{**}\gamma\begin{bmatrix} -\frac{1}{\psi} & 1 \end{bmatrix} Cov(\hat{\mathbf{x}}_{t+1}, \mathbf{g}_{t+1}|\mathcal{F}_t)\mathbf{I}_c^\top}_{\text{from signal}\times\text{growth}} \\ &+ \underbrace{\kappa_{1c}^{**}\left(\gamma - \frac{1}{\psi}\right)\mathbf{I}_d Cov(\mathbf{g}_{t+1}, \hat{\mathbf{x}}_{t+1}|\mathcal{F}_t)\mathbf{I}_c^\top}_{\text{from growth}\times\text{signal}}. \end{aligned} \tag{A.72}$$

Note each of four terms here include static (co)-uncertainties like the i.i.d. model case. Also, all covariance expressions here refer to a 2×2 covariance matrix except for the first scalar term. The covariance matrices in the third and fourth terms are, in particular, cross-covariance matrices; therefore, $Cov(\mathbf{g}_{t+1}, \hat{\mathbf{x}}_{t+1}|\mathcal{F}_t) = Cov(\hat{\mathbf{x}}_{t+1}, \mathbf{g}_{t+1}|\mathcal{F}_t)^\top$. Finally, we can rewrite the last three terms from (A.72) in scalar terms as follows.

$$RP_{ss} \triangleq \boldsymbol{\beta}_s \boldsymbol{\Sigma}_s^* \boldsymbol{\lambda}_s^\top = \kappa_{1c}^{**}\kappa_{1d}^{**}\left[\gamma - \frac{1}{\psi}\right] Var(\hat{x}_{c,t+1}|\mathcal{F}_t)\left[\frac{Cov(\hat{x}_{c,t+1}, \hat{x}_{d,t+1}|\mathcal{F}_t)}{Var(\hat{\mu}_{c,t+1}|\mathcal{F}_t)} - \frac{1}{\psi}\right], \tag{A.73}$$

$$RP_{sg} \triangleq \boldsymbol{\beta}_s \mathbf{J}_{sg}^* \boldsymbol{\lambda}_g^\top = \kappa_{1d}^{**}\gamma\left[Cov(\hat{x}_{d,t+1}, g_{c,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{x}_{c,t+1}, g_{c,t+1}|\mathcal{F}_t)\right], \tag{A.74}$$

$$RP_{gs} \triangleq \boldsymbol{\beta}_g \mathbf{J}_{gs}^* \boldsymbol{\lambda}_s^\top = \kappa_{1c}^{**}\left[\gamma - \frac{1}{\psi}\right] Cov(\hat{x}_{c,t+1}, g_{d,t+1}|\mathcal{F}_t). \tag{A.75}$$

Therefore, the risk premium expressions for the long-run risk model are identical to those of the i.i.d. model except for the constant terms κ_{1c}^{**} and κ_{1d}^{**}.

A.5.5 Equity Volatility and Risk-free Rate

From (A.51) and (A.71), we can decompose the conditional variance of the risky asset's returns, $Var[r_{m,t+1}|\mathcal{F}_t] = \boldsymbol{\beta}\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\beta}^\top$, as follows.

$$\begin{aligned} Var[r_{m,t+1}|\mathcal{F}_t] =& \underbrace{Var(g_{d,t+1}|\mathcal{F}_t)}_{\text{from growth}} \\ &+ \underbrace{(\kappa^{**}_{1d})^2 Var\left(\hat{x}_{d,t+1} - \frac{1}{\psi}\hat{x}_{c,t+1}|\mathcal{F}_t\right)}_{\text{from signal}} \\ &+ \underbrace{\kappa^{**}_{1d}\left(Cov(\hat{x}_{d,t+1}, g_{d,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{x}_{d,t+1}, g_{c,t+1}|\mathcal{F}_t)\right)}_{\text{from signal}\times\text{growth}} \\ &+ \underbrace{\kappa^{**}_{1d}\left(Cov(\hat{x}_{d,t+1}, g_{d,t+1}|\mathcal{F}_t) - \frac{1}{\psi}Cov(\hat{x}_{c,t+1}, g_{d,t+1}|\mathcal{F}_t)\right)}_{\text{from growth}\times\text{signal}}. \end{aligned} \tag{A.76}$$

On the other hand, the log risk-free rate is

$$\begin{aligned} r_{f,t} &= -E[m_{t+1}|\mathcal{F}_t] - \frac{1}{2}Var(m_{t+1}|\mathcal{F}_t) \\ &= -\theta\log\delta - (\theta-1)\left(\kappa_{0c} + \kappa_{1c}A_{0c} - A_{0c}\right) + \gamma\mathbf{I}_c\boldsymbol{\mu} + \frac{1}{\psi}\mathbf{I}_c\hat{\mathbf{x}}_t - \frac{1}{2}\boldsymbol{\lambda}\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\lambda}^\top \\ &= -\log\delta + \frac{1}{\psi}(\mu_c + \hat{x}_{c,t}) - \frac{1-\theta}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top - \frac{1}{2}\boldsymbol{\lambda}\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\lambda}^\top \\ &= -\log\delta + \frac{1}{\psi}E[g_{c,t+1}|\mathcal{F}_t] - \frac{1-\theta}{2\theta}\boldsymbol{\lambda}^*\boldsymbol{\Sigma}^*_{gs}\left(\boldsymbol{\lambda}^*\right)^\top - \frac{1}{2}\boldsymbol{\lambda}\boldsymbol{\Sigma}^*_{gs}\boldsymbol{\lambda}^\top. \end{aligned}$$

A.5.6 Systematic and Idiosyncratic Long-Run Risks

Table A.3 shows the effects of parameter ϕ_{xd} and χ_{xd} extending Table 2. As shown in the table, both systematic and idiosyncratic long-run risks in the dividend growth process increase the intertemporal co-uncertainty and the risk premium.

Table A.3: Systematic and Idiosyncratic Long-Run Risks and Asset Moments

Exposure to long-run growth $x_{c,t}$ ϕ_{xd}	Dividend long-run growth volatility χ_{xd}	Annualized equity premium (%) $E(R_m - R_f)$	Annualized equity volatility (%) $SD(r_{m,t})$	Annualized risk-free rates (%) $E(R_f)$	Intertemporal co-uncertainty ($\times 10^7$) $Cov(\hat{\mathbf{x}}_{\mathbf{t+1}}\|\mathcal{F}_t)$
Panel A: $\rho_{gs} = 0.3$ *and* $\rho_s = -0.9$					
0	3.0	3.50	16.48	0.87	0.76
0	4.5	4.81	22.04	0.94	1.16
0	6.0	6.24	28.79	1.00	1.46
1	3.0	5.29	16.86	0.80	1.45
1	4.5	6.58	21.76	0.88	1.91
1	6.0	7.98	27.78	0.95	2.26
3	3.0	8.26	17.67	0.68	2.67
3	4.5	9.62	21.54	0.75	3.29
3	6.0	11.03	26.49	0.83	3.78
Panel B: $\rho_{gs} = 0.2$ *and* $\rho_s = -0.95$					
0	3.0	3.76	16.26	0.89	0.89
0	4.5	5.20	21.85	0.98	1.35
0	6.0	6.71	28.70	1.05	1.67
1	3.0	5.49	16.72	0.82	1.61
1	4.5	6.90	21.62	0.91	2.11
1	6.0	8.37	27.69	0.99	2.49
3	3.0	8.31	17.69	0.69	2.88
3	4.5	9.80	21.48	0.78	3.54
3	6.0	11.29	26.43	0.87	4.05
Panel C: $\rho_{gs} = 0$ *and* $\rho_s = -1.0$					
0	3.0	1.47	14.88	1.44	1.22
0	4.5	2.93	20.27	1.49	1.80
0	6.0	4.00	27.59	1.53	2.18
1	3.0	3.02	15.00	1.40	2.01
1	4.5	4.44	19.88	1.46	2.61
1	6.0	5.59	26.38	1.50	3.03
3	3.0	5.42	15.93	1.32	3.38
3	4.5	7.01	19.65	1.39	4.11
3	6.0	8.33	24.93	1.44	4.66