Risk, Uncertainty, and Asset-Pricing ‘Puzzles’

Martin L. Weitzman*
Harvard University

This Draft: September 6, 2005
Comments Welcomed

Abstract

In conventional rational expectations expositions of the “equity premium puzzle,” “riskfree rate puzzle,” and “variability mismatch puzzle,” the subjective distribution of future growth rates is essentially made to mimic its past sample moments. This paper shows that the unobservable nature of structural growth parameters adds to expectation beliefs a permanent thick-tailed background layer of uncertainty that never converges to a stationary-ergodic rational expectations equilibrium, yet which explains the three ‘puzzles’ parsimoniously by one unifying principle. The posterior data generating process of a Bayesian evolutionary-learning equilibrium with a vague prior is consistent with all three values of the ‘puzzles’ observed as stylized facts.

1 Introduction: Structural Uncertainty and Asset Prices

The “equity premium puzzle” refers to the spectacular failure of the standard neoclassical representative-agent model of stochastic general-equilibrium growth to explain a historical difference of some six or so percentage points between the average return to a representative stock market portfolio and the average return from a representative portfolio of relatively safe stores of value. Such a large premium on risk *cum* uncertainty suggests either that people are perceiving more uncertainty about the future than past data would at first glance appear to indicate, or else that something is fundamentally wrong with the standard formulation of the

*(e-address: mweitzman@harvard.edu) For helpful detailed comments on earlier drafts of this paper, but without blaming them for its remaining defects, I am grateful to Andrew Abel, Gary Chamberlain, Xavier Gabaix, Alfred Galichon, Jerry Hausman, Joseph Kadane, Angelo Melino, Jonathan Parker, Jeff Strnad, and, especially, Evan Anderson, John Geweke, and Narayana Kocherlakota.
problem in terms of a non-bizarre, comfortably-familiar coefficient of relative risk aversion, say with conventional values $\gamma \approx 2 \pm 1$.

For this same risk-aversion coefficient of $\gamma \approx 2$, the stochastic generalization of the basic Ramsey formula from equilibrium growth theory predicts a riskfree interest rate in the approximate neighborhood $r^f \approx 5 - 6\%$, while what is actually observed is more in the range $\hat{r}^f \approx 0 - 1\%$. The large discrepancy between these two values is the “riskfree rate puzzle,” which represents another big disappointment with the standard neoclassical model.

In principle, aggregate returns on comprehensive economy-wide equity should reflect more-fundamental growth expectations about the underlying real economy. What is being called here the “variability mismatch puzzle” refers to the counterintuitive empirical fact that actual returns on a representative stock market index appear to be about an order of magnitude excessively more variable than any “fundamental” that might be driving them.

Taken together, this glaring trio of theoretical predictions that are away from reality by entire orders of magnitude seems devastating to the credibility of the standard economic approach to analyzing behavior under uncertainty. The problem for the theory is not the qualitative fact that it misses three empirical targets (this is economics after all), but rather the quantitative fact that in each case the theory is off-target by so much that it is not easy to rectify the discrepancies, even with generous supplements of ex-post-factum amendments. The proper interpretation of these three asset-pricing macro-puzzles has important ramifications throughout all branches of economics because it goes centrally to the core issue of discounting for risk, uncertainty, and time. These intuitively-related paradoxes are fairly crying out that something is deeply wrong with the standard paradigm – something which is unlikely to be corrected by tinkering with small modifications of the basic model. Some big critical element, which would capture the characteristic that makes the reward for bearing risk so high, seems to be missing from the formulation. At least for asset pricing applications, a consensus has developed among economists that the standard model is seriously flawed.

Perhaps not surprisingly, this bedeviling family of asset-pricing puzzles has spawned a huge research endeavor. In attempting to explain the paradoxes, an enormous post-puzzles literature has developed, which is filled with some imaginatively fruitful variations on the standard model. Many valuable insights have come out of the plethora of recent puzzle-driven models, but it still seems fair to say that no new consensus has yet emerged from within the economics profession as a whole that the puzzles have been satisfactorily resolved.

The point of departure for this paper is to note that the macroeconomic asset-pricing literature is dominated by the conventional model of rational expectations. In this stationary-ergodic paradigm, insider agents have effectively learned their way into knowing the “true” structural parameters of the stochastic growth process and simultaneously outsider econome-
tricians have accumulated enough data to justify having a high level of statistical confidence when substituting sample-frequency moments for “true” population moments in the Euler equation. While this classical methodology may well be appropriate for some economic applications, the paper will argue that such a way of framing the issues – and even just writing an Euler equation in ex-post empirically-realized frequencies – can be a badly flawed procedure for the particular application of analyzing aversion to structural uncertainty, which underlies (or, more accurately, should underlie) all asset-pricing calculations.

In a nonstationary (or evolutionary) world, insider agents and outsider econometricians are as one in being perennially uncertain about the underlying structural parameters of the future growth process, because learning does not converge to an ergodic distribution of growth rates. The flaw in the rational-expectations vision of empirical asset pricing is its generic bias towards making the future seem a lot less uncertain than it actually is by excluding evolutionary change. When a growth-coefficient does not have a “true” constant value because the stochastic growth process is evolutionary with hidden structural parameters, then classical-frequentist statistical inference can understate enormously the amount of predictive uncertainty about the future marginal-utility-weighted stochastic discount factor. This potentially unbounded prediction bias in forecasting expected future marginal utility spills over into severe pricing-kernel errors, which cascade into dramatically incorrect asset valuations, culminating in the bedeviling family of asset-return ‘puzzles.’

This paper attempts to shed light on the equity-premium, riskfree-rate, and variability-mismatch puzzles by simultaneously rooting all three issues together deeply into the common ground of Bayesian statistical inference. The paper is written in the spirit of a reconnaissance foray to see whether the “big picture” of the standard model begins to make sense of the data when just the simplest structural uncertainty is included in the formation of subjective-belief expectations. The fact that structural growth parameters of the model are hidden variables will introduce an extra layer of background uncertainty, which derives ultimately from the refusal of nondogmatic prior beliefs to preclude the evolution of unforeseen bad future histories, and whose repercussions do not dampen down to zero even for an infinite number of sample observations. Such omnipresent background uncertainty spreads out critically the probability distribution of future growth rates and is capable of acting strongly upon asset prices to increase significantly the values of both the equity premium and variability mismatch, while simultaneously decreasing markedly the riskfree interest rate.

To convey the essential statistical-economic insights as sharply as possible, only the most basic specification of the interplay between Bayesian learning and stochastic general equilibrium growth is modeled. Thus, to ease the computational burdens from delivering its core message, the model analyzes a very stark competitive equilibrium, with a single represen-
tative agent whose utility function is isoelastic, for a pure endowment-exchange fruit-tree economy (having no genuine production or investment), where the growth of aggregate output and the return to representative equity are both i.i.d. normal, and so on. The big-picture issue here, addressed at a very high level of abstraction, is whether the dark cloud hanging over the standard paradigm (in the form of three glaring, embarrassing, order-of-magnitude empirical discrepancies) is essentially lifted when just the tiniest “epsilon” of non-ergodic evolutionary-structural uncertainty is added to the core rational-expectationist model. For analytical sharpness the only change made from the simplest stationary specification (wherein all parameter values are constants assumed to be known by insider agents) is to have the model include Bayesian learning about just one structural parameter representing the unknown evolving variance of the normally distributed future growth rate. Although the model employs only familiar, analytically tractable, garden-variety specifications to be able to derive a relatively transparent expression for the family of asset-pricing discrepancies, it will become apparent that the basic insights have much broader applicability.

This paper is far from being the first to investigate the effects of Bayesian statistical uncertainty on asset pricing. Earlier examples having some Bayesian features or overtones include Barsky and DeLong (1993), Timmerman (1993), Bossaerts (1995), Cecchetti, Lam and Mark (2000), Veronesi (2000), Brennan and Xia (2001), Abel (2002), Brav and Heaton (2002), Lewellen and Shanken (2002), and several others. Broadly speaking, these papers indicate or hint, either explicitly or implicitly, that the need for (transient) Bayesian learning about structural parameters (along the path to a rational expectations equilibrium) may (temporarily) reduce the degree of one or another equity anomaly. What has been utterly missing from this literature, however, is any sense of the potentially unlimited power of the permanent strong force that distribution-spreading structural parameter uncertainty can bring to bear on asset pricing equations when an evolutionary stochastic process derails the ergodic convergence required to underpin rational expectations. In effect, the direction of this Bayesian-learning component of structural model uncertainty is (somewhat) appreciated in (some of) the literature, but not the stunning magnitude of the sustained “strong force” that it is capable of unleashing via its overwhelming ability to dominate so completely the numerical evaluation of standard expectation formulas involving stochastic discount factors.

A sole possible exception in the vast sea of equity-puzzle literature is an important note by Geweke (2001), who applies a Bayesian framework to the most standard model prototypically used to analyze behavior towards risk and then demonstrates the extraordinary fragility of the existence of finite expected utility itself.\footnote{I am grateful to two readers of the first draft of this paper for informing me of Geweke’s pioneering earlier article, after noticing that I had independently derived results with a similar flavor.} In a sense the present paper begins by accepting
this non-robustness insight, but pushes it further to argue that the inherent sensitivity of the standard prototype formulation constitutes a significant clue for unraveling what is causing the asset-pricing puzzles and for giving them a unified general-equilibrium interpretation that simultaneously fits the stylized time-series facts.

This paper will end up arguing that the three equity macro-puzzles are not nearly so ‘puzzling’ in a Bayesian evolutionary-learning framework that includes hidden-structure model-parameter uncertainty. Instead, the arrow of causality in a unified Bayesian explanation is reversed: the ‘puzzling’ numbers being observed empirically are trying to tell us economics researchers a revealing story about the implicit background subjective distribution of future growth-structure uncertainty that people in the real world actually have, and which is generating such data. In the final section of the paper the three ‘puzzling’ time-series sample averages of the equity premium, riskfree rate and variability mismatch are inverted to back out the implicit subjective probability distribution of the future growth rate. The paper shows that the “strong force” of evolutionary-structural uncertainty is a far more powerful determinant of asset prices and returns than the “weak force” of rational-expectations stationary-ergodic risk. Measured in the appropriate welfare-equivalent state space of expected utility, a world view about the subjective uncertainty of future growth prospects emerges that is much closer in expected-utility terms to what is being suggested by the relatively stormy volatility record of stock market wealth than it is to the far more placid smoothness of past consumption.

2 The Three Macro-Puzzles in Canonical Form

The critical issue for this paper is whether the appearance of the three related asset-return ‘macro puzzles’ might essentially be attributable to background evolutionary-structural uncertainty. To cut sharply to the analytical essence of this central issue, a super-stark model is used where everything else except the most basic architecture of the model has been set aside. Heroically assumed away are the details in such diversionary (for this paper) complications as production, technology, storage, leverage, illiquidity, defaults, taxes, autocorrelation, irrationality, heterogeneous agents, exotic preferences, borrowing constraints, labor income, human capital, business cycles, timing frictions, incomplete markets, idiosyncratic risks, and the like. The prototype specification is the simplest most-heroic version of the textbook workhorse dynamic-stochastic-general-equilibrium formulation of a fruit-tree endowment-exchange growth economy, which is ubiquitous as a benchmark point of depar-
ture throughout the finance-economics literature.\footnote{The famous fruit-tree model of asset prices in a growing economy traces back to two seminal articles: Lucas (1978) and Mehra-Prescott (1985). For applications, see the survey articles of Campbell (2003) or Mehra and Prescott (2003), both of which also give due historical credit to the other pioneering originators of the important set of ideas and the stylized empirical facts used throughout this paper. Citations for the many sources of these (and related) seminal asset-pricing ideas are omitted here only to save space, and because they are readily available, e.g., in the above two review articles and in the textbook expositions of Cochrane (2001), Duffie (2001), or Gollier (2001).}

The primitive “fundamental” in this economy is the growth rate $x$. Consumption-output endowment $C_t$ at time $t$ is given by the recursive equation

$$\ln C_t = x_t + \ln C_{t-1}, \quad (1)$$

where, for the time being at least, the random variables $\{x_t\}$ are allowed to come from a very general \textit{evolutionary} stochastic process.

The population consists of a large fixed number of identical people. Consumption in period $t$ is $C_t$. The utility $U$ of consumption $C$ is specified by the isoelastic power function

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma} \quad (2)$$

with corresponding marginal utility

$$U'(C) = C^{-\gamma}, \quad (3)$$

where the coefficient of relative risk aversion is the positive constant $\gamma$.

The pure-time-preference multiplicative factor for discounting one-period-ahead utility into present utility is $\beta$. At time $t$ the representative agent’s welfare is

$$W_t = E_t \left[ \frac{1}{1-\gamma} \sum_{j=0}^{\infty} \beta^j (C_{t+j})^{1-\gamma} \right], \quad (4)$$

where the expectation operator $E_t$ is conditioned on all information available at time $t$. It follows from (3), (1), and (4) that for all non-negative integers $i$ the marginal rate of substitution between $C_{t+i}$ and $C_{t+i+1}$ is

$$M_{t+i+1} \equiv \frac{\beta U'(C_{t+i+1})}{U'(C_{t+i})} = \beta \exp(-\gamma x_{t+i+1}), \quad (5)$$

and for any asset whose gross return in period $t+i+1$ is $R_{t+i+1}$, the relevant Euler equation
is

$$\beta E_{t+i}[\exp(-\gamma x_{t+i+1})R_{t+i+1}] = 1. \quad (6)$$

Within this model, all asset markets are in some sense phantom entities because no one actually ends up taking a net position in any of them. They exist as shadow exchange possibilities, but in this pure fruit-tree economy where fruit spoils after one period there is no avoiding the ultimate reality that everyone’s consumption must equal the yield on their endowment of one fruit tree per agent, no matter how the asset markets equilibrate. The paper follows tradition by concentrating on three basic investment vehicles: a “riskfree” asset, “one-period-ahead” equity, and “comprehensive” equity, all of which are abstractions of reality.

The “riskfree” asset is approximated in an actual economy by a portfolio of the safest possible stores of value, including hard currency, very-short-term U.S. or U.K. treasury bills, and inventories of real goods. In the theoretical fruit-tree economy, substituting a constant for \( R_{t+i+1} \) into the Euler equation (6) gives the price of the riskfree asset at time \( t+i \) as

$$P^f_{t+i} = \beta E_{t+i}[\exp(-\gamma x_{t+i+1})], \quad (7)$$

while the realized gross one-period return on a riskfree asset \( R^f_{t+i+1} \) is

$$R^f_{t+i+1} = \frac{1}{P^f_{t+i}}, \quad (8)$$

whose more familiar logarithmic form is

$$r^f_{t+i+1} = \rho - \ln E_{t+i}[\exp(-\gamma x_{t+i+1})], \quad (9)$$

where \( \rho \equiv -\ln \beta \) is the instantaneous rate of pure time preference and \( r^f_{t+i+1} \equiv \ln R^f_{t+i+1} \).

“One-period-ahead” equity is an asset that pays only next period’s consumption-output-endowment, and thereafter expires. Its price at time \( t+i \) is readily shown to be

$$P^{le}_{t+i} = \beta E_{t+i}[\exp((1-\gamma)x_{t+i+1})], \quad (10)$$

from which it follows that its realized gross return is

$$R^{le}_{t+i+1}(x_{t+i+1}) = \frac{\exp(x_{t+i+1})}{P^{le}_{t+i}}. \quad (11)$$

“Comprehensive” equity is approximated in the real world by a broad-based index of
publicly-traded shares of stocks whose aggregation weights represent the weights of the entire economy it is representing. In the theoretical fruit-tree economy, “comprehensive” equity is modeled abstractly as a claim on the stream of all future consumption-output-endowment dividends accruing to a fruit tree. Thus, in period $t + i$ the ex-dividend price of equity $P^e_{t+i}$ is the price of a claim on all dividends accruing to a fruit tree from time $t + i + 1$ onward, which by repeated use of the Euler condition can be written as

$$P^e_{t+i} = \sum_{j=1}^{\infty} \beta^j C_{t+i}E_{t+i} \left[ \left( \frac{C_{t+i+j}}{C_{t+i}} \right)^{1-\gamma} \right]. \quad (12)$$

The realized gross return on comprehensive equity between periods $t + i$ and $t + i + 1$ is

$$R^e_{t+i+1} = \frac{C_{t+i+1}}{P^e_{t+i}} + \frac{P^e_{t+i+1}}{P^e_{t+i}}. \quad (13)$$

Define the price-earnings ratio in period $t + i$ to be

$$\omega_{t+i} \equiv \frac{P^e_{t+i}}{C_{t+i}}, \quad (14)$$

and then write equation (13) as

$$R^e_{t+i+1} = \frac{1 + \omega_{t+i+1}}{\omega_{t+i}} \exp(x_{t+i+1}). \quad (15)$$

Dividing (15) by (8) the realized equity premium in ratio form for period $t + i + 1$ (conditioned all information available at time $t + i$) is

$$\frac{R^e_{t+i+1}}{R^f_{t+i+1}} = \beta E_{t+i} \left[ \exp(-\gamma x_{t+i+1}) \right] \frac{1 + \omega_{t+i+1}}{\omega_{t+i}} \exp(x_{t+i+1}), \quad (16)$$

The price-earnings ratio itself can be expressed in a useful closed form, which reveals its relation to expectations of future growth, by plugging (1) into (12) and then rewriting (14) as

$$\omega_{t+i} = \sum_{j=1}^{\infty} \beta^j E_{t+i} \left[ \exp \left( (1 - \gamma) \sum_{k=1}^{j} x_{t+i+k} \right) \right]. \quad (17)$$

At about this point in developing the argument, which up to now applies for a very general evolutionary stochastic process, the expository literature introduces the assumption of stationary-ergodic rational-expectations growth rates. Consistent with the spirit of using the simplest possible formulation, here the super stationary postulate is now imposed that the random variables $\{x_{t+i}\}$ are i.i.d. with a known distribution – but this assumption is
intended to apply only for expository purposes throughout the remainder of this section of the paper. In this special i.i.d. case, the riskfree-rate formula (9) becomes

\[ r^f = \rho - \ln E[\exp(-\gamma x)]. \]  

(18)

When the random variables \( \{x_t\} \) are i.i.d., it is readily shown from (17) that the price-earnings ratio is always the constant value

\[ \omega = \frac{\beta E[\exp((1 - \gamma)x)]}{1 - \beta E[\exp((1 - \gamma)x)]}. \]  

(19)

Then plugging (19) into (16), taking the natural log of the expected value, and subtracting (18), the equity premium in each period (under the i.i.d.-growth assumption) is

\[ \ln E[R^e] - r^f = \ln E[\exp(x)] + \ln E[\exp(-\gamma x)] - \ln E[\exp((1 - \gamma)x)]. \]  

(20)

Equation (20) is a theoretical formula for calculating the equity risk premium, given any coefficient of relative risk aversion \( \gamma \), and, more importantly here, given the probability distribution of the uncertain future growth rate \( x \). Concerning the relative-risk-aversion taste parameter \( \gamma \), there seems to be some rough agreement within the economics profession as a whole that an array of evidence from a variety of sources suggests that it is somewhere between about one and about three. More precisely stated, any proposed solution which does not explain the equity premium for \( \gamma \leq 4 \) would likely be viewed suspiciously by most members of the broadly-defined community of professional economists as being dependent upon an unacceptably high degree of risk aversion.

By way of contrast with preferences, which are standardly conceptualized as being fixed over time, much less is known about what is the appropriate probability distribution to use for representing future growth rates. The reason for this traces back to the unavoidable truth that, even under the best of circumstances (with a given, stable, stationary stochastic specification that can accurately be extrapolated from the past onto the future), no one can know with certainty the critical structural parameters of the distribution of \( x \). At this juncture in the story, the best that anyone can do is to infer from the past some estimate of the probability distribution of \( x \). The rest of the story hinges on specifying the form of the assumed probability density function of \( x \), and then looking to see what the data are saying about its likely parameter values. The functional form that naturally leaps to mind is the normal distribution

\[ x \sim N(\mu, V), \]  

(21)
The equity premium literature generally proceeds by implicitly presuming that the “true” structural parameters $\mu$ and $V$ are constants known by the agents inside the economy (although perhaps not known by an outside observer), and then continues on by substituting the normal distribution (21) into formula (20), which reduces (20) to a simple analyzable expression. Instead of allowing representative agents in the economy to be aware that $\mu$ and $V$ are unknown random variables, the standard practice essentially uses the first two sample moments and then goes on pretending that normality still holds (in place of substituting into (20) the relevant Student-\(t\) statistic to account for structural-uncertainty sampling error).

Let $\bar{x}$ be the sample mean and $\hat{V}$ be the sample variance of a long time series of past growth rates. Implicitly in the rational expectations interpretation, the sample size is presumed large enough to make $\bar{x}$ and $\hat{V}$ be sufficiently accurate estimates of their underlying “true” values $\mu$ and $V$ so that agents inside the economy can be imagined as having substituted $(\bar{x}, \hat{V})$ for $(\mu, V)$ in their Euler equations, but little formal attempt is made either to define carefully for this context “sufficiently accurate” or to confirm just exactly what happens to formula (20) if the estimates, and therefore the approximations, are not “sufficiently accurate.” When (21) is assumed along with the extreme point-mass dogmatic-prior case $E[x] = \bar{x}, V[x] = \hat{V}$, then using the formula for the expectation of a lognormal random variable and cancelling the many redundant terms simplifies (20) into the standard expression

$$\ln E[R_e] - r_f = \gamma V[x],$$

and for this special deterministic-structure case the equity premium puzzle is readily stated.

Considering the U.S. as a prime example, in the last century or so the average annual real arithmetic return on the broadest available stock market index is taken to be $\ln E[R_e] \approx 7\%$.\(^3\) The historically observed real return on an index of the safest available short-maturity bills is less than 1\% per annum, implying for the equity premium that $\ln E[R_e] - r_f \approx 6\%$. The mean yearly growth rate of U.S. per capita consumption over the last century or so is about 2\%, with a standard deviation taken here to be about 2\%, meaning $\hat{V} \approx 0.04\%$. Suppose $\gamma \approx 2$. Plugging these values into the right hand side of (22) gives $\gamma \hat{V} \approx 0.08\%$.

Thus, the actually observed equity premium on the left hand side of equation (20) exceeds the estimate (22) of the right hand side by some seventy-five times. If this were to be explained with the above data by a different value of $\gamma$, it would require the coefficient of relative risk aversion to be 150, which is away from acceptable reality by about two orders of

\(^3\)These numbers are from Mehra and Prescott (2003) and/or Campbell (2003), who also show essentially similar summary statistics based on other time periods and other countries (but most of which naturally have somewhat lower values of $\ln E[R_e]$ than “America in the American century”).
magnitude. This is the form or variant of the equity premium puzzle applicable to the above fruit-tree macro-model, and it is apparent why characterizing such a result as “embarrassing” (for the standard neoclassical paradigm) may be putting it very mildly. Plugging in some reasonable alternative specifications or different parameter values can have the effect of chipping away at the puzzle, but the overwhelming impression is that the equity premium is off by at least an order of magnitude. There just does not seem to be enough variability in the recent past historical growth record of advanced capitalist countries to warrant such a high risk premium as is observed. Of course, the underlying model is extremely crude and can be criticized on any number of valid counts. Economics is not physics, after all, so there is plenty of wiggle room for a paradigm aspiring to be the “standard economic model.” Still, two orders of magnitude seems like an awfully large base-case discrepancy to be explained away \textit{ex post factum}, even coming from a very primitive model.

Turning to the riskfree rate puzzle, the meaning given in the asset-pricing literature to equation (18) parallels the interpretation given to the equity premium formula. The literature typically proceeds from (18) by postulating the normal distribution (21), but then imagines that the representative agent ignores the statistical uncertainty inherent in estimating the “true” values of $\mu$ and $V$. Instead, these two structural parameters are standardly treated by plugging in their sample values as proxies for their “true” values, and then pretending that normality still holds. Substituting the deterministic-structure point-mass-parameter dogmatic-prior version $E[x] = \hat{x}$, $V[x] = \hat{V}$ into (18), and then using the formula for the expectation of a lognormal distribution, gives

$$r_f = \rho + \gamma E[x] - \frac{1}{2} \gamma^2 V[x],$$

which is a familiar generic equation appearing in one form or another all throughout equilibrium stochastic-growth theory. (Its origins trace back to the famous neoclassical Ramsey model of the 1920’s.) Non-controversial estimates of the relevant parameters appearing in (23) (calculated on an annual basis) are: $\hat{x} \approx 2\%$, $\hat{V} \approx .04\%$, $\rho \approx 2\%$, $\gamma \approx 2$. With these representative parameter values plugged into the right hand side of (23), the left hand side of (18) becomes $r_f \approx 5.9\%$. When compared with an actual real-world riskfree rate $\hat{r}_f \approx 1\%$, the theoretical formula is too high by $\approx 4.9\%$. This gross discrepancy is the riskfree rate puzzle. With the other base-case parameters set at the above values, the value of $\gamma$ required to explain the riskfree interest rate discrepancy is essentially $\gamma \approx 0$, whereas $\gamma \approx 150$ is required to explain the equity risk premium. Choosing a coefficient of relative risk aversion to ease the riskfree rate puzzle exacerbates the equity premium puzzle, and vice versa. The simultaneous
existence of two strong contradictions with reality, which, in addition, seem to be strongly contradicting each other, might be characterized as being “disturbing times three.”

As if all of the above were not vexing enough, there is also the enigmatic appearance in the data of what is being called throughout this paper the “variability mismatch puzzle.” In the ultra-simple fruit-tree model with i.i.d. growth of fruit and power utility, the same variability characterizes consumption, output, endowments, dividends, capital gains, one-period-ahead equity returns, and comprehensive equity returns. The sole genuine “fundamental” at time $t$ is the random variable $x_t$, at whose frequency the entire economic-financial system vibrates. From (11), (15), and (19), the i.i.d. case implies $R_t^e = R_t^1e$ and, with $r_t^e \equiv \ln R_t^e$,

$$r_t^e - E[r_t^e] = x_t - E[x].$$

According to (24), the deviation from the mean of continuously-compounded returns on equity $r^e - E[r^e]$ should, for this simplest rational-expectations i.i.d. model, coincide exactly with the deviation from the mean of its underlying “fundamental” $x - E[x]$, implying that all higher-order moments of the two distributions should match. From just glancing at the time-series sample, however, it is painfully obvious that the two empirical second-moment variabilities are badly mismatched because the (geometrically-calculated) standard deviation of equity returns $\hat{\sigma}[r^e] \approx 17\%$ is much bigger than the (geometrically-calculated) standard deviation of growth rates $\hat{\sigma}[x] \approx 2\%$. Relative to the highly abstract super-simple fruit-tree model of this section of the paper (with i.i.d. growth of consumption-endowment fruit and isoelastic utility), the relevant macroeconomic form of the “variability mismatch puzzle” is understood to be the stylized fact that, in actuality, the historical returns to a broad-based stock market index counterintuitively appear to be about an order of magnitude more variable than the underlying fundamental of an aggregate-output real-growth payoff, for which representative equity is supposed to be the surrogate claimant. Conforming once again here with the all-too-familiar quantitative asset-pricing macro-puzzle family pattern, it turns out that substituting alternative specifications or different parameter values can lessen the initial order-of-magnitude discrepancy (here of the degree of variability mismatch), but something central of the mystery remains that still seems way off base.

Summing up the scorecard for this super-simple i.i.d. fruit-tree version of the standard dynamic-stochastic-general-equilibrium growth model, we have three strong contradictions with reality and one serious internal contradiction, making the grand total add up to a conundrum that is disturbing times four. The next section of the paper examines what happens to the family of asset-return puzzles when the relevant structural parameters take on the familiar $t$-type sampling distributions that arise naturally in a non-stationary learning
environment when sample points are drawn independently from a normal population.

A strong heuristic intuition for what is coming up next can be gotten simply by substituting a Student-\(t\) distribution (from an arbitrarily large, but finite, sample of observations) for the normal distribution in formulas (18) and (20). When the limits of the relevant indefinite integrals containing the Student-\(t\) distribution are evaluated, it is readily seen from formula (18) that \(r^f \to -\infty\), while from (20) careful limit calculations show that \(\ln E[R^e]-r^f \to +\infty\). These extreme limiting values hint at the potentially enormous power of the “strong force” of structural parameter uncertainty to reverse categorically the asset-pricing puzzles, thereby raising into sharp prominence the critical question: what are we supposed to be explaining here? Should we be trying to explain why the actually-observed equity premium is so embarrassingly high while the actually-observed riskfree rate is so embarrassingly low (relative to a theoretical formula based on the normal distribution)? Or should we be trying to explain the opposite pattern: why the actually-observed equity premium is so embarrassingly low while the actually-observed riskfree rate is so embarrassingly high (relative to a theoretical formula based on a Student-\(t\) distribution that is operationally indistinguishable from the normal for which it is a sufficient statistic)?

The next section of the paper shows how to pose and answer such questions rigorously in a non-ergodic general-equilibrium setting by using a hybrid distribution whose tail properties are midway between a normal and a Student-\(t\). The formulation will firstly “contain the t-explosion” by introducing a bound on the diffuseness of the prior (to get rid of the infinities), and then the model will force the resulting hybrid probability density function of \(x\) to go to the i.i.d. limit of a “Bayesian-\(t\) learning distribution” when the number of degrees of freedom is made to approach infinity. The limiting hybrid “Bayesian-\(t\) learning distribution” looks like an i.i.d.-normal but in its effects on investor sentiments acts more like an i.i.d.-\(t\), thereby generating via this mechanism time series data that seem highly anomalous when mistakenly judged to be the outcome of a stationary-ergodic rational-expectations stochastic process.

3 Hidden-Structure Expectations of Future Growth

Perhaps surprisingly, it turns out for asset-pricing implications that the most critical single issue involved in Bayesian learning about the probability distribution of future growth rates concerns the variance being unknown. (The case of the mean being unknown garners the lion’s share of attention in the asset-price-learning literature, in part because of analytical tractability and in part because of a widespread perception that with large samples available in continuous time it is relatively easy to learn the “true” variance.) For notational and conceptual simplification, it is very convenient to be able to postulate straightaway a situa-
tion where \( E[x] \) is a given known constant \( \mu \), so that the only genuine statistical uncertainty in the system concerns the estimation of the hidden value of the variance \( V[x] \). The case where \( E[x] \) and \( V[x] \) are both unknown is less neat, but gives essentially the same results.

To indicate where the argument is now and where it is heading, the assumptions behind the model to be used throughout the rest of the paper are stated formally here. The Euler equation (6) is presumed to hold for the utility function (2) in \textit{ex-ante subjective expectations} (as contrasted with holding in \textit{ex-post realized frequencies} – more on this distinction later). The presumed probability distributions are: \( x \sim N(\mu, V) \) and \( r^e \sim N(E[r^e], V[r^e]) \). The following six quasi-constants of the model are effectively assumed to be known by insider agents: \( E[r^e], V[r^e], r^f, \rho, \gamma, \mu \). Only one structural parameter is unknown and must be estimated statistically: \( V = V[x] \). This section effectively derives the Bayesian subjective distribution of the critical hidden-structure evolutionary parameter \( V \), which will then be applied to the model throughout the remainder of the paper.

Assuming the normal specification (21), define the random variable

\[
\theta \equiv \frac{1}{V},
\]

which is commonly called the \textit{precision} of a probability distribution. It is assumed for all times \( t \) that, \textit{conditional on} \( \theta = \theta_t \), the growth rate \( x = x_t \) is i.i.d.-normal:

\[
x | \theta_t \sim N(\mu, 1/\theta_t).
\]

Bayes’s theorem tells us that the posterior of \( \theta_t \) is its likelihood times its prior. The prior gives voice to the subjective “what if” side of a debate inside the head of the representative agent-investor at time \( t \) about how to predict \( \theta_t \). The likelihood gives voice to the side of objective data-evidence. For this setup, the relevant \textit{likelihood function} \( \ell_t(\theta_t) \) will later be proved (by induction on the number of past observations) to be of the gamma density form

\[
\ell_t(\theta_t) = \frac{b_t^{\alpha_t}}{\Gamma(\alpha_t)} \theta_t^{\alpha_t-1} \exp(-b_t \theta_t).
\]

The coefficients \( \alpha_t > 0, b_t > 0 \) appearing in (27) will turn out to be functions of past data, but for now are treated as given parameters. Temporarily assuming on faith here the gamma form (27), the mean value of \( \theta_t \) in the likelihood function is

\[
E_t^e[\theta_t] = \frac{\alpha_t}{b_t}
\]
while its variance is

\[ V_t[\theta_t] = \frac{a_t}{b_t^2}. \]  (29)

The distribution of \( \theta_t \) in (27) evolves over time with \( a_t \) and \( b_t \) according to the following logic. Suppose someone was trying to predict future values of the precision at time \( t - 1 \). It seems natural to presume that the passage of time always carries with it the potential for new forms of uncertainty to evolve, which have not previously been encountered, and that make us more unsure about the future. In this setup where the expected growth rate is always a known constant, the introduction of newly-evolved growth uncertainties is modeled as a mean-preserving spread of the gamma likelihood function (27).

In between period \( t - 1 \) and period \( t \), after the new sources of uncertainty have arrived but before \( x_t \) has been realized, suppose that the one-period-ahead values of \( a_t \) and \( b_t \) become

\[ a_t' = (1 - \epsilon)a_{t-1}, \]  (30)

\[ b_t' = (1 - \epsilon)b_{t-1}, \]  (31)

where \( \epsilon \) is conceptualized as being a very small positive number. It follows from (28), (29) that (30), (31) constitutes a mean-preserving spread of the likelihood of next period’s precision. The parameter \( \epsilon \) quantifies the fractional increase in the variance of next period’s precision relative to the variance of this period’s precision – after the not-previously-encountered newly-evolved forms of future growth uncertainty have appeared, but just before the new information contained in a fresh value of \( x_t \) has arrived. (If \( V = 1/\theta \) is constant, then its value could be inferred exactly in continuous time were an infinity of observations available, which, as will become apparent, corresponds here to the special limiting case \( \epsilon \rightarrow 0 \).)

Each \( x_t \) is the independent realization of a normal random variable whose mean is known to be \( \mu \), but whose precision is unknown. Conditional on the precision being \( \theta_t \), the distribution of \( x \) just before it is realized or observed is given by (26). The likelihood (just before the realization \( x_t \) is observed) of the pair \( \{\theta, x\} \) occurring together is therefore

\[ \ell_t'(\theta; x) \propto \sqrt{\theta} \exp(-(x - \mu)^2/2) \theta^{a_t'-1} \exp(-b_t' \theta). \]  (32)

Applying Bayes’s theorem to (32) just after \( x = x_t \) has been observed and regrouping terms gives the new likelihood function

\[ \ell_t(\theta_t) \propto \theta_t^{a_t'-1/2} \exp \left( - \left[ b_t' + (x_t - \mu)^2/2 \right] \theta_t \right). \]  (33)

Plugging (30), (31) into (33) and comparing the result with (27), we have derived for this
setup the recursive relations

\[ a_t = (1 - \epsilon) a_{t-1} + \frac{1}{2}, \quad (34) \]

\[ b_t = (1 - \epsilon) b_{t-1} + \frac{(x_t - \mu)^2}{2}, \quad (35) \]

which completes the induction step in the proof of the gamma form (27).

It is analytically very convenient to assume an infinite number of past observations, in which case the solutions of (34), (35) converge from any initial condition to

\[ a_t = \frac{1}{2\epsilon}, \quad (36) \]

\[ b_t = \frac{\nu_t}{2\epsilon}, \quad (37) \]

where

\[ \nu_t \equiv \epsilon \sum_{i=0}^{\infty} (1 - \epsilon)^i (x_{t-i} - \mu)^2 \quad (38) \]

is the (exponentially weighted) sample variance of past growth rates.

The model is made consistently recursive henceforth by now defining for any time \( t \) the fundamental state variable of the economy to be the (exponentially weighted) sample variance \( \nu_t \), whose equation of evolutionary motion is

\[ \nu_t = (1 - \epsilon) \nu_{t-1} + \epsilon (x_t - \mu)^2, \quad (39) \]

and where it should be noted that the only dependence on time \( t \) of the above system of equations enters via the state-variable value \( \nu_t \). Hereafter throughout the rest of the paper the subscript \( t \) is dropped from notation in this now-fully-time-autonomous hidden-structure dynamic system. Note importantly that (28) and (29) now become \( E^\ell[\theta \mid \nu] = 1/\nu \) and \( V^\ell[\theta \mid \nu] = 2\epsilon/\nu^2 \), which seemingly (but mistakenly, as it turns out) implies that even a vague prior cannot undo absolute convergence to an ergodic rational-expectations equilibrium for sufficiently small \( \epsilon \).

Turning to the prior distribution, some minimal structure is imposed on it as follows. Let \( y \) be any random variable defined on \((-\infty, +\infty)\) having a continuous probability density function \( \phi(y) \) with \( \phi(0) > 0 \), which is normalized so that

\[ \int_{-\infty}^{\infty} y \phi(y) \, dy = 0, \quad (40) \]
\[ \int_{-\infty}^{\infty} y^2 \phi(y) \, dy = 1, \]  
(41)

and, for reasons that will soon become apparent, satisfies

\[ \int_{-\infty}^{\infty} \exp(\lambda \exp(-y/\delta)) \phi(y) \, dy < \infty \]  
(42)

for all \( \lambda > 0 \) and \( \delta > 0 \). Condition (42) has the effect of preventing “too much” probability mass from accumulating in the left tail of \( \phi(y) \) as \( y \to -\infty \), and it will be satisfied by a large family of thin-left-tailed distributions, including \( \text{any} \) \( \phi(y) \) that has a finite lower support.

Now introduce the random variable \( \theta_{-\infty} \) defined by the equation

\[ \ln \theta_{-\infty} = y/\delta + \ln(1/\nu). \]  
(43)

The prior probability density function \( p(\theta_{-\infty}) \) is then taken to be what corresponds mathematically to the transformation (43), meaning from the Jacobian inverse formula that

\[ p(\theta_{-\infty}) = \frac{\delta}{\theta_{-\infty}} \phi(\delta(\ln \theta_{-\infty} + \ln \nu)), \]  
(44)

where it should be kept in mind throughout what follows that \( E[\ln \theta_{-\infty}] = \ln(1/\nu) \) and \( V[\ln \theta_{-\infty}] = 1/\delta^2 \). The hyper-parameter \( \delta = 1/\sigma[\ln \theta_{-\infty}] \) is called the coefficient of prior information because it quantifies how informative is the prior distribution. (Note that the extreme situation of infinite informativeness \( \delta \to \infty \) is the rational-expectations i.i.d. case familiar from the asset-pricing expository literature where \( x \sim N(\mu, \nu). \))

Multiplying the likelihood (27), (36), (37) times the prior (44), the posterior-predictive probability density function of the precision \( \theta \) is

\[ \psi(\theta \mid \nu, \epsilon, \delta) = \frac{\phi((\ln \theta + \ln \nu)\delta) \theta^{1/2-2} \exp(-\nu \theta/2\epsilon)}{\int_{0}^{\infty} \phi((\ln \theta + \ln \nu)\delta) \theta^{1/2-2} \exp(-\nu \theta/2\epsilon) \, d\theta}, \]  
(45)

From combining (45) with (26) and integrating out \( \theta \), the unconditional (or marginal) posterior-predictive probability density function of the future growth rate \( x \) is

\[ g(x \mid \nu, \epsilon, \delta) = \frac{\int_{0}^{\infty} \int_{0}^{\infty} \exp(-\theta(x-\mu)^2/2) \phi(\delta \ln(\nu \theta)) \theta^{1/2-2} \exp(-\nu \theta/2\epsilon) \, d\theta \, dx}{\int_{-\infty}^{\infty} \int_{0}^{\infty} \exp(-\theta(x-\mu)^2/2) \phi(\delta \ln(\nu \theta)) \theta^{1/2-2} \exp(-\nu \theta/2\epsilon) \, d\theta \, dx}. \]  
(46)
Of course an outside observer cannot know directly what value of \( \delta \) describes an investor’s prior beliefs, as \( \delta \) can only be inferred indirectly from the data. A favorite default setting would be the case \( \delta \to 0 \) representing the standard textbook statistical situation of a “diffuse” or “vague” or “noninformative” prior for the precision \( \theta \), which turns (44) into the standard Bayesian reference prior \( p(\theta) \propto 1/\theta \) and effectively corresponds to the familiar classical normal-linear regression case. For any given positive \( \epsilon \), straightforward (if tedious) integration, after replacing \( \phi(\delta \ln(\nu \theta)) \) in (46) by \( \phi(0) \) when passing to the limit as \( \delta \to 0 \), shows that \( g(x \mid \nu, \epsilon, \delta) \) then reduces to a Student-\( t \) distribution with \( 1/\epsilon - 2 \) degrees of freedom, location parameter \( \mu \), and precision \( (1 - 2\epsilon)/\nu \), whose probability density function is

\[
g(x \mid \nu, \epsilon, \delta = 0) = \frac{\sqrt{\epsilon} \Gamma(\frac{1+2}{2})}{\sqrt{\pi} \nu \Gamma(\frac{1-2\epsilon}{2})} \left[ 1 + \frac{\epsilon}{\nu} (x - \mu)^2 \right]^{\frac{1}{2} - \frac{1}{2}}.
\]  

(Note from (47) that henceforth we are given literary license to conceptualize the case of a diffuse prior \( \delta \to 0 \) “as if” a regression had been run on a sample of size \( n = 1/\epsilon - 1 \).

Speaking generally, with power utility the formula for “expected future marginal utility” or “expected stochastic discount factor” or “expected pricing kernel,” (all of which names are essentially interchangeable) reflects the mathematical properties of the moment generating function of \( x \). The moment generating function of a Student-\( t \) distribution such as (47) is unboundedly large because the defining integral diverges to plus infinity as \( \delta \to 0 \). A situation can therefore always be synthesized where expected pricing kernels or stochastic discount factors are made arbitrarily large simply by choosing for (46) a sufficiently small value of \( \delta \), no matter what positive value of \( \epsilon \) has been given.

Translated into Bayesian terms here, a bare-minimum necessary prerequisite for the validity of the frequentist law-of-large-numbers justification behind calibration (the notion to “just let the data sample speak for itself”) is that asset-pricing expectation formulas involving marginal utility should become independent of the prior as the sample size approaches infinity. For rational expectations to serve as a robust and trustworthy basis on which to understand asset returns requires that the data of the likelihood function should asymptotically dominate (in marginal utility space) any reasonable representation of a not-very-informative prior distribution. Such asymptotic dominance of the data over a vague prior often accompanies a Markov-stationary environment, but ergodicity does not emerge here, essentially because the stochastic process is evolutionary and learning never “catches up” with the “true value” of \( \theta \). For any given \( \epsilon > 0 \), the value of \( \delta \) chosen for the prior manifests itself as a smear of posterior-predictive background uncertainty that refuses, even with the interdiction of an infinite amount of past data, to relinquish its decisive hold on influencing present expectations of future asset-pricing kernels.
From a Bayesian perspective, we “just let the data sample speak for itself” in a different sense from the classical frequentist law-of-large-numbers interpretation of this phrase. The remaining sections of the paper will each “just let the data sample speak for itself” by telling us what is the information content $\delta$ that real-world investors must implicitly be using in their priors, in order to be consistent with one or another stylized fact that we researchers seem to be observing in the actual asset-return sample coexisting with realizations of the data generating process (27), (26). For these purposes it does not matter whether investors arrived at such a prior through their own imagination or through some transmission process connecting them with beliefs in the infinitely-remote past. The prior, after all, just represents subjective opinion open to modification by the evidence, whose most basic role here is to help a representative agent-investor answer what-if hypothetical questions of the form: if I had believed this a priori, then, after seeing the (here infinite sample of) data, what might I expect from the future?

Because they can be driven to an arbitrary extent by the fickle whimsicality of investor-agents concerning what value of the information coefficient $\delta$ to select for the prior, asset prices always have the potential of depending critically upon prior beliefs (regardless of the amount of data accumulated) and, at least theoretically, they might swing wildly in reaction to just the tiniest changes in $\delta$. It follows that classical asset-pricing-kernel regressions trying to fit rational-expectations ex-post-empirical realizations of an Euler condition are fundamentally misspecified from the beginning, and perhaps it is then of little wonder that such a stationary-frequentist pure-ergodic-risk methodology typically ends up effectively rejecting the Euler equation itself by producing pricing errors and paradoxes. The message of this paper that an asset-pricing equilibrium may plausibly be based upon a permanent “strong force” of structural uncertainty, in which prior beliefs have the power to trump data-evidence every time, is the crucial missing link in a unified Bayesian approach capable of rationalizing the three so-called equity ‘puzzles.’ (Whether such a theory is better labeled stationary or non-stationary is essentially beside the point here, the substantive issue being that no amount of data generated by this model enables a statistician to disconnect the posterior-predictive stochastic discount factor from the effects of prior information in order to get at some hypothetical prior-belief-free purely-data-driven “objective” structure.)

Taking (45) and (46) as the representative agent’s subjective probability density functions, the paper is now ready to compute the hidden-structure equity premium, the hidden-structure riskfree interest rate, and the hidden-structure variability mismatch. The next three sections do these partial-equilibrium calculations, in turn. In the last section of the paper, implicit parameter values of the subjective probability distribution of future growth rates are backed out of the data by Bayesian inverse calibration. For each application, the
The sharpest insight comes from having in mind the mental image of a double-limiting situation where simultaneously \( \epsilon \to 0 \) and \( \delta \to 0 \), so that the value of \( \nu \) defined by (39) approaches some known constant that is unchanging over time, and also the probability density function \( g(x \mid \nu, \epsilon, \delta) \) defined by equation (46) converges to the normal distribution \( x \sim N(\mu, \nu) \) because \( V[\theta \mid \nu] = \epsilon/\nu \) goes to zero. This prototype double-limiting situation comes arbitrarily close to the standard familiar textbook case of growth-rate risk being i.i.d. normal with known parameters, only the model never quite gets to such a stationary-ergodic normal distribution because some very small (but nevertheless consequential for asset pricing) new uncertainty and subsequent learning occur whenever \( \epsilon > 0 \). Such an extreme thought experiment creates a situation that is operationally indistinguishable from the textbook workhorse rational-expectations i.i.d.-lognormal specification, and therefore it is ideal for focusing the mind very sharply on understanding intuitively the core Bayesian structural model-uncertainty mechanism driving the entire family of asset-return ‘puzzles.’

4 The Hidden-Structure Equity Premium

Rewriting (7) in state-notation, the price of the riskfree asset is \( Pf(\nu, \epsilon, \delta) = \beta E[\exp(-\gamma x)] \). From (10), the price of one-period ahead equity is \( P1e(\nu, \epsilon, \delta) = \beta E[\exp((1-\gamma) x)] \), where in both cases \( x \) is a random variable whose probability density function \( g(x \mid \nu, \epsilon, \delta) \) is given by (46). The realized one-period ahead equity premium in ratio form is then

\[
\frac{R1e(x \mid \nu, \epsilon, \delta)}{Rf(\nu, \epsilon, \delta)} = \frac{Pf(\nu, \epsilon, \delta)}{P1e(\nu, \epsilon, \delta)} \exp(x),
\]

where

\[
\frac{Pf(\nu, \epsilon, \delta)}{P1e(\nu, \epsilon, \delta)} = \frac{\int_{-\infty}^{\infty} e^{-\gamma x} g(x \mid \nu, \epsilon, \delta) \, dx}{\int_{-\infty}^{\infty} e^{(1-\gamma) x} g(x \mid \nu, \epsilon, \delta) \, dx}.
\]

The following proposition contains two results. First, for all \( \epsilon > 0 \), some value of \( \delta \) matches any given one period ahead asset-price ratio (49). Second, by choosing carefully \( \delta(\epsilon, \nu) \) and then going to the limit \( \epsilon \to 0 \), any desired equity premium can be replicated in the data as if it came from the super-simple i.i.d.-normal model of Section 2.

**Theorem 1** First part: let \( \gamma > \frac{1}{2} \). Let \( \bar{\pi} \) be any given value satisfying \( \bar{\pi} > \gamma \nu \). Then for every \( \epsilon > 0 \), \( \nu' < \bar{\pi}/\gamma \), there exists a \( \delta_{\bar{\pi}}(\epsilon, \nu') > 0 \) such that

\[
\frac{Pf(\nu', \epsilon, \delta_{\bar{\pi}}(\epsilon, \nu'))}{P1e(\nu', \epsilon, \delta_{\bar{\pi}}(\epsilon, \nu'))} = \exp(\bar{\pi} - \mu - \frac{1}{2}\nu').
\]
Second part: for any non-negative integer \( i \), as \( \epsilon \to 0 \), the random variable \( x_{t+i+1} \) converges in probability to an i.i.d. random variable \( z \sim N(\mu, \nu) \) where the convergence is uniform for all \( \delta \geq 0 \) and of the same strength as the convergence of a Student-t distribution to a normal when the number of observations approaches infinity. Furthermore, if \( \delta \) is chosen as \( \delta_{\epsilon}(\epsilon, \nu') \) for \( 0 < \nu' < \pi / \gamma \), then the limiting realized equity premium \( R_{t+i+1}^c / R_{t+i+1}^f \) in (16) converges to the i.i.d. lognormal random variable \( \exp(z + \mu - \frac{1}{2} \nu') \) as \( \epsilon \to 0 \).

**Proof.** Using (26) and the formula for the expectation of a lognormal random variable rewrite (49) (after cancelling terms in \( \mu \)) as

\[
\frac{P^{f}(\nu', \epsilon, \delta)}{P^{1e}(\nu', \epsilon, \delta)} = e^{-\frac{\mu}{2}} \frac{\int_{0}^{\infty} e^{\gamma^2/2\theta} \psi(\theta | \nu', \epsilon, \delta) \, d\theta}{\int_{0}^{\infty} e^{(1-\gamma)^2/2\theta} \psi(\theta | \nu', \epsilon, \delta) \, d\theta}.
\]

As \( \delta \to 0 \), the probability density function \( g(x | \nu, \epsilon, \delta) \) defined by (46) approaches the Student-t distribution (47), whose moment generating function is unbounded. Consequently, as \( \delta \to 0 \) both integrals in (49) and in (51) approach \( +\infty \). Therefore, from (51),

\[
\lim_{\delta \to 0} \frac{P^{f}(\nu', \epsilon, \delta)}{P^{1e}(\nu', \epsilon, \delta)} = e^{-\frac{\mu}{2}} \lim_{\theta \to 0} \frac{\exp(\gamma^2/2\theta)}{\exp((1-\gamma)^2/2\theta)}.
\]

Because

\[
\ln \frac{\exp(\gamma^2/2\theta)}{\exp((1-\gamma)^2/2\theta)} = \frac{\gamma - \frac{1}{2}}{\theta},
\]

plugging (53) into (52) gives

\[
\lim_{\delta \to 0} \frac{P^{f}(\nu', \epsilon, \delta)}{P^{1e}(\nu', \epsilon, \delta)} = \lim_{\theta \to 0} \frac{\gamma - \frac{1}{2}}{\theta} = +\infty.
\]

At the other extreme of \( \delta \), it is apparent that as \( \delta \to \infty \), then \( P^{f}(\nu', \epsilon, \delta)/P^{1e}(\nu', \epsilon, \delta) \to \exp(\gamma \nu' - \mu - \frac{1}{2} \nu') \), because the economy is then effectively in the conventional lognormal case (22). The function \( P^{f}(\nu', \epsilon, \delta)/P^{1e}(\nu', \epsilon, \delta) \) defined by (49) is continuous in \( \delta \). Since

\[
\frac{P^{f}(\nu', \epsilon, 0)}{P^{1e}(\nu', \epsilon, 0)} < \exp(\pi - \mu - \frac{1}{2} \nu') < \frac{P^{f}(\nu', \epsilon, \infty)}{P^{1e}(\nu', \epsilon, \infty)},
\]

condition (50) follows and the first part of the theorem is proved.

Turning to the second part of the theorem, to save space the notation-intensive proof is only sketched here. The fact that as \( \epsilon \to 0 \) the random variable \( x_{t+i+1} \) converges in probability to an i.i.d. random variable \( z \sim N(\mu, \nu) \), in the same mode as a Student-t
distribution converges to a normal as \( n \to \infty \), essentially comes from (47). As \( \epsilon \to 0 \), from (39) \( \nu' \equiv \nu_{t+i} \to \nu \), implying \( \nu' < \pi/\gamma \). If \( \delta \) is chosen as \( \delta_\pi(\epsilon, \nu') \), from (50) it then follows that as \( \epsilon \to 0 \) the realized one-period-ahead equity premium converges in probability to the i.i.d. lognormal random variable \( \exp(z + \pi - \mu - \frac{1}{2}\nu) \). It is intuitively obvious that if in every future period \( t + i + 1 \) the one-period-ahead realized equity premium \( R_{t+i+1}^{le} / R_{t+i+1}^f \) is always the i.i.d. lognormal random variable \( \exp(z + \pi - \mu - \frac{1}{2}\nu) \), then this must also be the realized premium on comprehensive equity \( R_{t+i+1}^{le} / R_{t+i+1}^f \). (To save space, this last part of the proof of this is left as an exercise, but basically it just comes from repeatedly applying the law of iterated expectations to (17), (16)).

The essence of the Bayesian statistical mechanism driving the first part of Theorem 1 can be intuited by examining what happens in the limiting case. As \( \delta \to 0 \) for fixed \( \epsilon > 0 \), \( \nu > 0 \), the limit of (46) is distributed like a Student-\( t \) statistic of the form (47), the same as if a regression had been run on \( n = 1/\epsilon - 1 \) data points. With the presumed prototype case of small positive \( \epsilon \) and \( \delta \), the central part of the \( t \)-like distribution (46) is approximated extremely well by a normal curve with mean \( \mu \) and variance \( \nu \) fitting the data throughout its middle range. However, for applications involving the implications of uncertainty-aversion, such as calculating the equity premium, to ignore what is happening away from the center of the distribution has the potential to wreak havoc on the calculations. For these applications, such a normal distribution may be a very bad approximation indeed, because the more-spread-out dampened-\( t \) distribution (46) is capable in principle of producing an explosion in asset pricing formulas like (49), implying in the limit as \( \delta \to 0 \) an unboundedly large equity premium. Within the Bayesian framework of this model, therefore, the statistical fact that the moment generating function of a Student-\( t \) distribution is infinite has the important economic interpretation that, at least as a matter of principle, model-structure uncertainty has the potential in such a normal-gamma world to be a far more significant determinant of asset prices than ergodic risk. In the diffuse-prior limit as \( \delta \to 0 \), the representative agent becomes explosively more averse to the “strong force” of statistical uncertainty about the future growth process, whose structural parameters are unknown and must be estimated, than is this agent averse to the “weak force” of the pure risk per se of being exposed to the same underlying stochastic growth process, except with known structural parameters. The key to understanding the rational-expectations misunderstanding of the “equity premium puzzle” is that the “premium” is not on pure ergodic risk alone, but rather it is a combined premium on ergodic risk plus (potentially vastly more significant) structural uncertainty.

An explosion of the equity premium does not happen in the real world, of course, but a tamed near-explosive outcome remains the mathematical driving force behind the scene, which imparts the statistical illusion of an enormous equity premium incompatible with the
standard neoclassical paradigm. When people are peering forward into the future they are also looking back at their own prior, and what they are seeing is a spooky reflection of their present insecurity in not being able to judge accurately the possibility of unforeseen bad evolutionary mutations of future history that might conceivably ruin equity investors by wiping out their stock market holdings at a time just when their world has already taken a bad turn. This eerie sensation of low-$\delta$ diffuse background shadow-risk may not be simple to articulate, yet it frightens investors away from taking a more aggressive stance in equities and scares them into a position of wanting to hold instead (on the margin) a portfolio of some safer stores of value, such as cash, inventories of real goods, government treasury bills, perhaps precious metals, or even stockpiles of food – as a hedge against unforeseen bad future evolutions of history. Consequently, in an evolutionary equilibrium where there is zero net demand for them, these relatively-safe assets bear very low, even negative, rates of return.

Such type of Bayesian statistical explanation is not easily dismissable. The equity premium puzzle is the quantitative paradox that the observed value of $\ln E[R^e] - r^f$ is too big to be reconciled with the standard neoclassical stochastic growth paradigm having familiar parameter values. But compared to what is the observed value of $\ln E[R^e] - r^f$ “too big”? Essentially, the answer given in the equity-premium literature is: “compared to the right hand side of formula (22) when $\hat{V} \approx 0.04\%$ and $\gamma \leq 4$.” Unfortunately for this logic, the point-calibrated right hand side of (22) gives a terrible prediction for the observed realizations of $R^e/R^f$ because in the underlying calculation all assets have been priced by a rational-expectations formula that makes the future seem far less uncertain than it actually is. Anyone wishing to downplay this line of reasoning in favor of the status quo ante would be hard pressed to come up with their own Bayesian rationale for calibrating variances of non-observable subjectively-distributed future growth rates by point estimates equal to past sample averages. In essence, the rational-expectations approach that produces the family of asset-pricing puzzles avoids the consequences (on marginal-utility-weighted asset-pricing kernels) of overpowering sensitivity to low values of prior information $\delta$ only by effectively imposing from the very beginning the extreme deterministic-structure pure-ergodic-risk case $\delta = \infty$ (or $\epsilon = 0$) of a normal distribution with known parameters.

An early attempt to explain the equity premium puzzle by Rietz (1988) can be interpreted as essentially arguing through numerical examples that the sample may not be adequately representing a worst-imaginable-case scenario of large negative future growth rates. The impact on financial equilibrium of a situation where there is a tiny probability of a catastrophic out-of-sample event has been dubbed the “peso problem.” In a peso problem, the small probability of a disastrous future happening (such as a collapse of the presumed structure from a natural or socio-economic catastrophe) is taken into account by real-world investors.
(in the form of a “peso premium”) but not by the calibrated model, because such an event is not in the sample being used for the calibration.

Theorem 1 is trying to tell us that the statistical architecture of something conceptually akin to a peso problem is inescapably hardwired into the “deep structure” of how Bayesian inferences (about future economic growth, at unknown rates) interact with any curved utility function having a positive infimum of curvature (as measured locally by the elasticity of marginal utility). Bayesian inferences about the unknown hidden-structure variance fatten the posterior tails of probability density functions with dramatic consequences when expressed in subjective-expectation units of future marginal utility – as the example of replacing the workhorse normal distribution by its t-like posterior distribution demonstrates. This “Bayesian-statistical peso problem” means that for asset pricing applications it is not the least bit absurd to adhere to the non-rational-expectations non-ergodic idea that no amount of data may be large enough to identify all of the relevant structural uncertainty concerning future economic growth. The Bayesian peso problem is essentially saying that to calibrate an exponential evolutionary process having an uncertain future growth rate by plugging the sample variance of observed past growth rates into an “extremely bad” approximation of the subjectively-distributed stochastic discount factor, is to underestimate “extremely badly” the comparative utility-risk of a real-world gamble on the unknown structural potential for future economic growth, relative to a safe investment in a near-money sure thing.

Translated into classical-frequentist statistical language, the second part of the theorem has the following important interpretation. For any given $\nu$, pick $\pi > \gamma \nu$, name some number $k$, and choose any desired level of statistical strength relative to the supposedly “true” data generating process. Then there exists some sufficiently small $\epsilon > 0$ and accompanying values of $\delta = \delta_{\pi}(\epsilon, \nu')$ (where $\nu' = \nu_{t+i+1}, i < k$) such that the empirically observed frequency distribution of the $k$ realized values of the equity premium simulation-generated by this hidden-structure model is guaranteed to differ only insignificantly (in terms of the desired level of statistical strength) from the sampling distribution that would be simulation-generated in a sample of size $k$ if the “true” equity premium were i.i.d. $N(\pi, \nu)$. (Note that the data generating process being described here makes the first moment of the equity premium match statistically the empirical data, but it counterfactually makes the second moment be $\nu = \hat{V}[x]$ instead of $\hat{V}[r^e]$ – more on this variance mismatch later.)

Of course, what is being presented here is but one illustrative example of the economic consequences of a hidden-structure tail-fattening effect, but I believe that it is very difficult to get around the moral of this story. For any positive value of $\epsilon$, however small, the results of Bayesian distribution-spreading will cause the equity premium to be very sensitive to seemingly negligible changes in the assumed information content of prior beliefs, when
such innocuous prior-belief changes, by classical-frequentist logic, should have been washed away by the data-evidence long ago. Such fragility to subjective prior informativeness even with unlimited data effectively renders unbelievable the standard stationary-ergodic rational-expectations parable of asset pricing. The dominant statistical-economic force behind the puzzles is that seemingly thin-tailed probability distributions (like the normal), which actually are only thin-tailed conditional on known structural parameters of the model, become thick-tailed (like the t) after integrating out the parameter uncertainty. Intuitively, no finite sample of effective size $n(=1/\epsilon -1) < \infty$ can accurately assess tail thickness, and therefore the attitude of a risk-averse Bayesian agent towards investing in various risk-classes of assets may be driven to an arbitrarily large extent by this unavoidable feature of Bayesian expectational uncertainty.

The important result in Schwarz (1999) can be interpreted as saying that for essentially any reasonably-specified non-dogmatic probability density functions, the conclusions from which are invariant to measurement units, the moment generating function of the posterior distribution is infinite (i.e., the posterior distribution has a “thick” tail) even when the random variable is being drawn from a thin-tailed parent distribution whose moment generating function is finite. Such a result means that there is a generic sense in which, at least potentially, people are significantly more afraid of not knowing what are the structural-parameter settings inside the black box, whose data generating process drives the pure-risk part of stochastic growth rates, than are they averse to the pure risk itself. When investors are modeled as perceiving and acting upon these inevitably-spread-out subjective posterior-predictive distributions, then a fully-rational equilibrium interpretation integrates together a unified Bayesian treatment of the entire family of equity puzzles, as the next three sections of the paper (when combined with this section) will show, in turn.

5 The Hidden-Structure Riskfree Interest Rate

We can use the same mathematical-statistical apparatus to calculate the hidden-structure riskfree interest rate. For all other parameter values fixed, let $f(\delta)$ be the value of $r^f$ as a function of the prior’s information content $\delta$ that comes out of formula (18) when the probability density function of $x$ is $g(x \mid \nu, \epsilon, \delta)$ defined by equation (46). Plugging the subjective posterior-predictive distribution (46) into the right hand side of equation (18), the result is

$$f(\delta) \equiv \rho - \ln \int_{-\infty}^{\infty} \exp(-\gamma x) g(x \mid \nu, \epsilon, \delta) \, dx.$$  

**Theorem 2** Let $r^f$ be any given continuous function of $\nu$ satisfying $r^f < \rho + \gamma \mu - \frac{1}{2} \gamma^2 \nu$ for
all \( \nu > 0 \). Then for every \( \epsilon > 0, \nu > 0 \), there exists a \( \delta f(\epsilon, \nu) > 0 \) such that

\[ r^f = f(\delta f). \tag{57} \]

Furthermore, the limiting realized riskfree rate can be made to converge to the same constant value \( \bar{f} \) in every future period if \( \delta \) is chosen as \( \delta f(\epsilon, \nu) \) for \( f(\delta) = \bar{f} \) while \( \epsilon \to 0 \).

**Proof.** As \( \delta \to 0 \), the probability density function \( g(x \mid \nu, \epsilon, \delta) \) defined by (46) approaches the Student-\( t \) distribution (47), whose moment generating function is unbounded. From the definition (56) therefore, \( f(0) = -\infty \). At the other extreme of \( \delta \), it is apparent that as \( \delta \to \infty \), then \( f(\delta) \to \rho + \gamma \mu - \frac{1}{2} \gamma^2 \nu \), because the economy is then effectively in the conventional lognormal case (23). Thus,

\[ f(0) < r^f < f(\infty), \tag{58} \]

and, since \( f(\delta) \) defined by (56) is continuous in \( \delta \), the conclusion (57) follows. The convergence to a constant value for all future periods follows from the fact that \( \nu \) effectively becomes constant over time as \( \epsilon \to 0 \), so that the condition \( r^f = \bar{f} < \rho + \gamma \mu - \frac{1}{2} \gamma^2 \nu \) always holds in the future. \( \blacksquare \)

The discussion of Theorem 2 so closely parallels the discussion of Theorem 1 that it is largely omitted in the interest of space. The driving mechanism again is that the random variable of subjective future growth rates behaves somewhat like a Student-\( t \) statistic in its tails and carries with it a potentially explosive moment generating function reflecting an intense aversion to low-precision evolutionary-mutational future histories. The bottom line once more is that a “Bayesian peso problem” causes false rational-expectations inferences about expected future utility, which are based upon mimicking the observed historical frequency of past growth rates, to underestimate enormously just how relatively much more attractive are safe stores of value when compared with a real-world Bayesian gamble on the uncertain growth-structure of an unknown future economy.

The relevant classical-frequentist statistical statement here about the relationship between the riskfree rate that is observed in the data and the supposedly “true” data generating process parallels the equity premium version. Pick \( r^f = \bar{f} < \rho + \gamma \mu - \frac{1}{2} \gamma^2 \nu \), name some number \( k \), and choose any desired level of statistical strength, here representing measurement accuracy. Then there exists some small \( \epsilon > 0 \) and accompanying \( \delta = \delta f(\epsilon, \nu) \) such that the frequency distribution of the \( k \) riskfree-rate realizations generated by this hidden-structure model is guaranteed statistically to differ only within measurement error from what would be generated in a sample of size \( k \) if the “true” riskfree rate was the constant value \( \bar{f} \).
6 Welfare-Equivalent as if Normal Growth Variability

This section of the paper is openly heuristic. It suggests an intuitive interpretation of the “variability mismatch puzzle” in subjective expected-utility-welfare terms (as contrasted with objective realized-growth-frequency terms). The main purpose is to convey some sense of the magnitude of the cost of structural uncertainty by expressing it in a user-friendly welfare-equivalent version of the familiar i.i.d.-normal distribution.

For the ultra-stark canonical form of the tree model in Section 2 of this paper, with growth of future consumption-endowment fruit known to be i.i.d. and isoelastic utility, the only genuine “fundamental” of the system was the growth rate \( x \). The idea that equity returns should vibrate consistently with the vibrations of their underlying “fundamentals” found expression for the i.i.d. rational-expectations macro-model of Section 2 in equation (24). According to (24), for an economy-wide comprehensive stock index embodying an implicit claim on the future aggregate output of the underlying real economy, all higher-order central moments of \( x \) and \( r^e \) should match subjectively and objectively, given the assumption of i.i.d. rational expectations. Alas, the empirical second moments of \( x \) and \( r^e \) are not even remotely matched in the time-series data because \( \tilde{V}[r^e]/\tilde{V}[x] \approx 75 \). Subjectively, however, the underlying “fundamental” \( x \) is perceived as if it is much more variable than it seems in the time-series data, because the marginal-utility-weighted pricing-kernel part of the stochastic discount factor \( \beta \exp(-\gamma x) \), is subjectively expected to be much larger than what would appear to be indicated by identifying the variance of future \( x \) with its past sample average \( \nu \), which would give the plug-in expected value \( E[\beta \exp(-\gamma x)] = \beta \exp(-\gamma \mu + \frac{1}{2} \gamma^2 \nu) \).

From Theorem 1, when we force the model to match the equity premium in the data, then the variability of realized equity returns counterfactually matches the variability of growth rates, instead of matching the variability of equity returns actually observed in the data. This section reverses the causality of Theorem 1 and instead asks: what happens to everything else if we simply force the model to match the variability of equity returns observed in the data? To proceed further here analytically, some simplifying assumption about the reduced form of equity returns is needed. The textbook benchmark assumption (which is ubiquitous throughout expository finance economics and which is consistent with the time-series data for low-frequency periods of a year or more) is that continuously-compounded equity returns are i.i.d.-normal. For the purposes of analytical tractability, this section of the paper merely follows the literature blindly by accepting as a given point of departure the standard reduced-form assumption (around which centerpiece the rest of the general equilibrium system will now be made to revolve) that equity returns are independently normally distributed with known mean and variance – without ever specifying from first principles the exact payoff-
transmission mechanism through which this normal equity-return distribution with known parameters emerges in the first place. This workhorse reduced-form methodology is used throughout the literature for various purposes, but it never really explains why equity returns are i.i.d.-normal with known parameters. Here it is used to examine the welfare consequences on the rest of the general equilibrium system of imposing this assumption.

Suppose, for the sake of argument, that the representative agent feels more comfortable conceptualizing (or the modeler feels more comfortable telling a story about) uncertain future growth prospects in terms of a familiar user-friendly i.i.d.-normal probability distribution. The representative agent here understands that, due to the “Bayesian peso problem,” using a point estimate of the standard deviation of growth rates from the past sample understates significantly in units of subjectively-expected utility the relevant future variability, which is some probability smear over a hidden-structure parameter. It then follows that a subjective future growth rate known to be i.i.d.-normally distributed, which is specifically constructed to have the same expected utility $E[U]$ and mean $E[x]$ as the subjective distribution $g(x \mid \nu, \epsilon, \delta)$, must necessarily have a significantly higher known variance than the realized sample variance $\nu$ from past data. We now examine the important connection between the known given variability of equity returns (already postulated to be i.i.d.-normally distributed with known parameters) and the variability of a welfare-equivalent subjective belief in i.i.d.-normally-distributed growth rates. In a sense, this section of the paper is trying to answer the question: between the two observed variability alternatives, $\tilde{\sigma}[r_e]$ standing in for the left hand side of equation (24) or $\tilde{\sigma}[x]$ standing in for the right hand side, which variability better matches the agent’s true welfare situation?

Leaving aside the “rationality” of such beliefs, suppose in this section of the paper that

$$x^N(x \mid \nu, \epsilon, \delta) \sim N(E[x^N], \sigma^2[x^N])$$

is a random variable function of the random variable $x$ representing the agent’s subjective probability belief that future growth rates are i.i.d.-normal with known parameters $E[x^N]$ and $\sigma[x^N]$. Let this agent also have a subjective probability belief in a stock-market payoff implicitly representing a unit claim on the lognormally-i.i.d. future aggregate output corresponding to (59). Such a payoff claim gives rise to the subjective probability belief of a (geometrically measured) return on comprehensive economy-wide equity $r^N(x^N)$ satisfying

$$r^N(x^N(x)) - E[r^N] = x^N(x) - E[x^N],$$

which is exactly the normal counterpart here of equation (24). It will turn out that i.i.d. as-if-normal growth rates can be made to yield the same expected one-period return on
equity as the formulation in previous sections of the paper, so that $E[r^N] = E[r^1e]$, which, provided also that $\sigma[r^N] = \tilde{\sigma}[r^e]$, signifies here that observed equity data alone cannot refute this as-if-normality hypothesis about subjective future growth beliefs, given the standard assumption that equity returns are known by the agents to be i.i.d. normal in the first place.

The following representation theorem establishes the existence of a conceptually-useful consequence of expected-utility indifference between $x^N$ and $x$. In the framework of this model, it turns out that forcing $x^N$ by construction to give the same expected utility as $x$ is intimately connected with the important implication for welfare calibration that $\sigma[x^N] \approx \tilde{\sigma}[r^e]$. This third proposition of the paper can therefore be interpreted as providing at least a sense in which there might be some rationale for telling an as-if parable wherein the representative agent has a subjective normally-distributed welfare-equivalent belief, which is consistent with (60) and the equity-return data, “as if” the future growth rate is $x^N$ with known variability equal to the observed variability of returns on equity. In this subjective interpretation (“as if” growth rates are i.i.d.-normal with known mean and variance), the welfare situation of the agent is represented by the relatively high variability of returns on equity, rather than by the relatively low variability of realized past growth rates. Because here $\sigma[r^N] = \sigma[x^N]$, at least from within the framework of this artificially-constructed welfare-equivalent as-if-i.i.d.-normal parable, there is no longer a jarring mismatch of variabilities wanting to be explained between equity returns and underlying fundamentals.

**Theorem 3** Let $\sigma > 0$ be any given continuous function of $\nu$ satisfying $\sigma^2 > \nu$ for all $\nu > 0$. Let $r^N(x^N(x))$ be a solution of (59), (60). Then for every $\epsilon > 0$, $\nu > 0$, there exists a $\delta_s(\epsilon, \nu) > 0$ such that the following four conditions are simultaneously satisfied:

1. $E[r^N(x^N(x))] = E[r^1e(x)]$, \hspace{1cm} (61)
2. $\sigma[r^N(x^N(x))] = \sigma[x^N(x)] = \sigma$. \hspace{1cm} (62)
3. $E[x^N(x)] = E[x] = \mu$, \hspace{1cm} (63)
4. $\forall C > 0 : E[U(C \exp(x^N(x)))] = E[U(C \exp(x))]$, \hspace{1cm} (64)

**Proof.** Define $s(\delta)$ to be the implicit solution of the equation

$$
\frac{1}{\sqrt{2\pi s(\delta)}} \int_{-\infty}^{\infty} \exp\left((1 - \gamma)x^N - \frac{(x^N - \mu)^2}{2s(\delta)^2}\right) \, dx^N = \int_{-\infty}^{\infty} \exp((1 - \gamma)x) g(x \mid \nu, \epsilon, \delta) \, dx, \hspace{1cm} (65)
$$

and note for this definition that (64) and (63) are satisfied by construction.
It can readily be shown that
\[ r^{1e}(x) = x + \{\rho - \ln E[\exp((1 - \gamma)x)]\}, \]  
and, analogously,
\[ r^N(x^N) = x^N + \{\rho - \ln E[\exp((1 - \gamma)x^N)]\}, \]  
so that (61) then follows from (63), (65), (66), (67).

As \( \delta \to \infty \), the probability density function \( g(x | \nu, \epsilon, \delta) \) goes to a normal distribution with variance \( \nu \), and consequently the integral on the right hand side of equation (65) approaches (by the expected-lognormal formula) \( \exp((1-\gamma)(1+\frac{1}{2}(1-\gamma)^2\nu)) \), implying \( s(\infty) = \sqrt{\nu} \). As \( \delta \to 0 \), the probability density function \( g(x | \nu, \epsilon, \delta) \) defined by (46) approaches the Student-t distribution (47), whose moment generating function is unbounded, implying the right hand side of (65) is also unbounded, meaning \( s(0) = \infty \). Thus
\[ s(\infty) < \sigma < s(0), \]  
and, by continuity of the function \( s(\delta) \), there must exist a \( \delta_s(\epsilon, \nu) > 0 \) satisfying
\[ s(\delta_s) = \sigma, \]  
which, when combined with (60), proves (62) and concludes the proof. \( \blacksquare \)

The force behind Theorem 3 is the same “strong force” that is driving the previous two theorems: intense aversion to the structural parameter uncertainty embodied in fat-tailed \( t \)-distributed subjective future growth rates. Compared with the Student-\( t \) distribution \( x \sim g(x | \nu, \epsilon, 0) \), a representative agent will always prefer, for any finite \( s \), the normal distribution \( x \sim N(\mu, s^2) \). Theorem 3 results when the limiting explosiveness of the moment generating function of \( g(x | \nu, \epsilon, \delta = 0^+) \) with a diffuse prior is contained by the substitution of \( g(x | \nu, \epsilon, \delta = \delta_s) \) with a somewhat-informative prior \( \delta_s(\epsilon, \nu) > 0 \).

Theorem 3 is effectively saying that if you must compress the complicated reality of future evolutionary uncertainty into a simple stationary-ergodic as-if-i.i.d.-normal-growth story, then \( \sigma[x^N] = \tilde{\sigma}[r_e] \) tells the better welfare parable than \( \sigma[x^N] = \tilde{\sigma}[x] \). To an outsider classical-frequentist statistician, however, agent-investors inside this economy appear to be irrationally incapable of internalizing what the data are clearly saying about \( \tilde{\sigma}[x] \approx 2\% \). Instead, these agents seem to be clinging stubbornly in their mind’s eye to an unshakably-consistent, but highly irrational, mental image as if their future welfare depends via the stochastic discount factor upon the realization of some hypothetical much-more-variable normally-distributed growth rate whose counterfactual standard deviation is \( \sigma[x^N] = \tilde{\sigma}[r_e] \approx 17\% \). With one
hundred independent observations, however, the frequentist hypothesis that the observed sample value of $\hat{\sigma}[x^N] = 2\%$ could have been generated by (agents having in their heads) a “true” welfare-equivalent value of $\sigma[x^N] = 17\%$ is classically rejected by a chi-squared test at the 99.99% confidence level!

7 Some Bayesian as if Normal Calibration Exercises

Viewing the three theorems of the paper through the lens of the welfare-equivalent as-if-i.i.d.-normal-growth story of Theorem 3 delivers a neat analytically-tractable relationship among $\pi$, $f$, and $s$ of the closed form (22), (23), which accompanies the well-known formula for the expectation of a lognormal random variable. The three theorems themselves are only partial equilibrium statements in the sense that the $\delta(\epsilon, \nu)$ function that works for any one theorem will not work for the other two – essentially because a system parameterized with just one degree of freedom cannot explain three observables simultaneously. Suppose however (what at this stage is just an unproved conjecture) that a more general three-dimensional parameterization can be made to deliver a situation “as if” the same $\delta(\epsilon, \nu)$ function works for all three theorems. The following question then arises naturally: does the simple relationship among $\pi$, $f$, and $s$ of the closed form (22), (23) hold empirically, conditional on the same $\delta(\epsilon, \nu)$ function working for all three theorems? The answer is “yes.” The experimental outcome that all three stylized-fact values of the equity premium, riskfree rate, and equity variability “fit,” in the sense that they come close to matching simultaneously the theoretically-predicted as-if-lognormal relationship among themselves, conveys at least some intuitive feel for the degree to which this heuristic way of looking at things is a relatively coherent theoretical-empirical construct.

The proposed exercise will test whether the welfare-equivalent interpretation of Theorem 3 that the future growth rate $x$ is subjectively distributed as if it were the i.i.d.-normal random variable $x^N$ with mean $E[x^N] = \bar{x}$ and standard deviation $\sigma[x^N] = \tilde{\sigma}[x^e]$ renders, along with (60), an internally-consistent as-if story connecting together the actual stylized facts of our economic world. In Table 1, quasi-constant parameter settings have been selected that, I think, represent stylized-fact numbers well within the “comfort zone” for most economists. All rates are real and given by annual values. The data are intended to be an overall approximation of what has been observed for many countries over long time periods.

The model is explaining endogenously three quasi-constants $\pi(\delta)$, $f(\delta)$, and $s(\delta)$ as functions of the one free informativeness parameter $\delta$. As a conceptual-notational convenience, pretend that $h > 0$ represents some arbitrarily small quantum threshold of observability below which any prior distribution is considered to be “effectively diffuse” and the rela-
Quasi-Constant Parameter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean arithmetic return on equity</td>
<td>$\ln E[R^e] \approx 7%$</td>
</tr>
<tr>
<td>Geometric standard deviation of return on equity</td>
<td>$\sigma[r^e] \approx 17%$</td>
</tr>
<tr>
<td>Riskfree interest rate</td>
<td>$r^f \approx 1%$</td>
</tr>
<tr>
<td>Implied equity premium</td>
<td>$\ln E[R^e] - r^f \approx 6%$</td>
</tr>
<tr>
<td>Mean growth rate of per-capita consumption</td>
<td>$E[x] \approx 2%$</td>
</tr>
<tr>
<td>Standard deviation of growth rate of per-capita consumption</td>
<td>$\sigma[x] \approx 2%$</td>
</tr>
<tr>
<td>Rate of pure time preference</td>
<td>$\rho \approx 2%$</td>
</tr>
<tr>
<td>Coefficient of relative risk aversion</td>
<td>$\gamma \approx 2$</td>
</tr>
</tbody>
</table>

Table 1: Some Economic "Stylized Facts"

The relationship between $\delta_\pi$, $\delta_f$, and $\delta_s$ becomes so blurred by indeterminacy that the situation is considered to be as if the same $\delta(\epsilon, \nu)$ function works for all three theorems. (Just to give a specific number here, let $h = 6.6 \times 10^{-27}$.) Under such circumstances, we will not be able to observe or calculate the underlying primitive values of $\delta_\pi$, $\delta_f$, and $\delta_s$ directly (although we know that in theory there exists some tiny-tiny $\epsilon^+ > 0$, for which simultaneously $\delta_\pi < h$, $\delta_f < h$, and $\delta_s < h$, because $\epsilon \to 0$ implies that simultaneously $\delta_\pi \to 0$, $\delta_f \to 0$, and $\delta_s \to 0$). However, and more usefully here, an indirect calibration experiment can be performed by setting any one of the three quasi-constants $\pi \mid \{\delta_\pi < h\}$, $f \mid \{\delta_f < h\}$, and $s \mid \{\delta_s < h\}$ equal to its observed value in Table 1 and then backing out the implied values of the other two remaining quasi-constants by inverting the two analytically-tractable as-if-normal-growth equations of the closed form (22) and (23). Because there are two equations ((22) and (23)) absolutely-continuous in three unknown variables, if any one of $\{\hat{\pi}, \hat{f}, \hat{s}\}$ is “sufficiently near” to explaining the other two, then each of $\{\hat{\pi}, \hat{f}, \hat{s}\}$ must be “sufficiently near” to explaining the other two. The following calibration exercise shows empirically that the entire system is “sufficiently near” to $\{\hat{\pi}, \hat{f}, \hat{s}\}$ in the distance-metric of what might be considered on intuitive grounds to be the most natural topology to use here.

Defining $\delta_s < h$ to be an implicit solution of

$$
\delta_s = s^{-1}(\sigma[r^e]) = s^{-1}(17\%),
$$

we then have, from (22) with $V[x] \equiv s^2(\delta_s)$,

$$
\ln E[R^e] - r^f = \gamma s^2(\delta_s) = \pi \mid \{\delta_s < h\} = 5.8\%,
$$

to be compared with $\pi \mid \{\delta_\pi < h\} = 6\%$. From (23) with $V[x] \equiv s^2(\delta_s)$,

$$
r^f = \rho + \gamma E[x] - \frac{1}{2} \gamma^2 s^2(\delta_s) = f \mid \{\delta_s < h\} = 0.2\%,
$$

32
to be compared with $f \{ \delta_f < h \} = 1\%$.

Defining $\delta_\pi < h$ to be the implicit solution of

$$\delta_\pi = \pi^{-1}(\ln E[R^e] - r_f) = \pi^{-1}(6\%)$$

we then have, from (23) and (22),

$$r^f = \rho + \gamma E[x] - \gamma \pi(\delta_\pi)/2 = f \{ \delta_\pi < h \} = 0\%,$$

to be compared with $f \{ \delta_f < h \} = 1\%$. From (22) with $V[x] \equiv \sigma^2[r^e]$, 

$$\sigma[r^e] = \sqrt{\pi(\delta_\pi)/\gamma} = s \{ \delta_\pi < h \} = 17\%,$$

to be compared with $s \{ \delta_s < h \} = 17\%$.

Defining $\delta_f < h$ to be an implicit solution of

$$\delta_f = f^{-1}(r_f) = f^{-1}(1\%),$$

we then have, from (23) and (22),

$$\ln E[R^e] - r^f = 2[\rho + \gamma E[x] - f(\delta_f)]/\gamma = \pi \{ \delta_f < h \} = 5\%,$$

to be compared with $\pi \{ \delta_\pi < h \} = 6\%$. From (23) with $V[x] \equiv \sigma^2[r^e]$, 

$$\sigma[r^e] = \sqrt{2[\rho + \gamma E[x] - f(\delta_f)]/\gamma} = s \{ \delta_f < h \} = 16\%,$$

to be compared with $s \{ \delta_s < h \} = 17\%$.

As a kind of a test for the internal consistency and raw fit of the as-if-i.i.d.-normal-growth story (hypothetically conditional on a three-dimensional version of the same $\delta(\epsilon, \nu)$ function working for all three theorems), the results of these Bayesian calibration exercises fit nearly exactly. At the very minimum, therefore, this model provides some story about why everything coheres almost perfectly in the bare-bones canonical i.i.d.-normal model when, by just the simplest substitution, a welfare-equivalent growth variability $\sigma[x_N] = \tilde{\sigma}[r^e]$ equal to the observed standard deviation of equity returns replaces the observed growth variability $\bar{\sigma}[x]$. Otherwise, such a near-perfect fit must be interpreted as merely happening to be some kind of a miraculous coincidence in the data.

Continuing with the above as-if-i.i.d.-normal-growth scenario, consider next a purely hypothetical thought experiment in which the magic trick is performed of eliminating all
future variability $\sigma$ of consumption. With i.i.d. lognormality, the imaginary deterministic path having the same mean consumption as the stochastic trajectory (1) is

$$C_{t+1} = \exp(\mu - \frac{1}{2} \sigma^2) C_t. \quad (70)$$

Using formula (70), it can readily be shown (following Lucas (2003)) that the welfare gain from a mean-preserving shrinkage that compresses the stochastic trajectory (1) into the deterministic path (70) is equivalent to a change in each period’s consumption of

$$\Delta C_t = (\exp(\frac{1}{2} \gamma \sigma^2) - 1) C_t. \quad (71)$$

When $\gamma \approx 2$ and the historical value of $\sigma = \tilde{\sigma}[x] \approx 2\%$ is used in (71), then $\Delta C_t/C_t \approx 0.04\%$, which is the kind of magnitude sometimes used to argue that the cost of growth variability is so counterintuitively low that even a complete removal of all conceivable macroeconomic uncertainty would be worth very little. Such a number, however, captures only the “weak force” of stationary-ergodic growth-rate risk. The welfare equivalent of a magic-trick elimination of all uncertainty about future growth, including the “strong force” of structural uncertainty, is better assessed by using the subjective value $\sigma = \tilde{\sigma}[x_N] = \tilde{\sigma}[r^e] \approx 17\%$ in formula (71), for which case $\Delta C_t/C_t \approx 2.9\%$. Accounted in this welfare-equivalent metric of shrunken consumption therefore, structural uncertainty about the future growth process turns out empirically to be far more significant than stationary-ergodic growth-rate risk.

8 Conclusion

The hidden-structure model of this paper is predicting that a classical story based upon a misspecified ex-post-realized-frequency interpretation of the Euler equation will generate data appearing to show an “equity premium puzzle,” a “riskfree rate puzzle,” and a “variability mismatch puzzle,” whose magnitudes of discrepancy are close numerically to what is observed empirically. This paper argues that such numerical “discrepancies” are puzzles, however, only when seen through a rational-expectations lens. From a Bayesian learning perspective, the “puzzling” numbers being observed in the data are telling a rational (but not “rational expectations” in the conventional stationary-ergodic sense) story about the implicit subjective distribution of background structural-parameter uncertainty arising from the uncertain evolutionary-growth process that is generating such data.

In principle, consumption-based representative agent models provide a complete answer to all asset pricing questions and give a unified theory integrating together the economics of
finance with the real economy. In practice, consumption-based representative agent models with standard preferences and a traditional degree of relative risk aversion work poorly when the variance of the growth of future consumption is point-calibrated to the sample variance of its past values. The theme of this paper is that with evolutionary-structural uncertainty there is some theoretical justification for treating the non-observable variability of the subjective future growth rate as if it were equivalent in welfare to the observed variability of a comprehensive economy-wide index of equity returns, for which as if interpretation the simple standard neoclassical model may have the potential to work well in practice.

References


