# Subjective Expectations and Asset-Return Puzzles

*By* Martin L. Weitzman*

*In textbook expositions of the equity-premium, riskfree-rate and equity-volatility puzzles, agents are sure of the economy's structure while growth rates are normally distributed. But because of parameter uncertainty the thin-tailed normal distribution conditioned on realized data becomes a thick-tailed Student-*t *distribution, which changes the entire nature of what is considered "puzzling" by reversing every inequality discrepancy needing to be explained. This paper shows that Bayesian updating of unknown structural parameters inevitably adds a permanent tail-thickening effect to posterior expectations. The expected-utility ramifications of this for asset pricing are strong, work against the puzzles, and are very sensitive to subjective prior beliefs—even with asymptotically infinite data.* (*JEL* D84, G12)

*The map appears to us more real than the land.*

—D. H. Lawrence

Three major puzzles, described later in this section, have captured the attention of macroeconomic finance: the equity-premium, riskfree-rate and equity-volatility puzzles. A common strand of these three asset-return puzzles is that markets are behaving as if investors fear some unknown hidden randomness that isn't obvious from the data. People are acting in the aggregate like there is much more marginal-utility–weighted subjective variability about future growth rates than past observations seem to support. This paper offers a single unified theory for all three macro-finance puzzles based on the idea that what is learnable about the future stochastic consumption-growth process from any number of past empirical observations must fall far short of full structural knowledge. The main findings can be summarized as follows: (a) the process of discovering structural parameters has significant economic consequences, with parameters controlling the spread of the distribution of future consumption growth rates (like the growth-rate variance) being the most critical for consumption-based asset pricing; (b) integrating out structural-parameter uncertainty by Bayes's rule spreads apart probabilities and thickens the tails of the posterior distribution for predicting the future consumption growth rate, an effect that persists indefinitely if structural parameters are conceptualized as continually evolving; (c) the thickened posterior-predictive left tail represents structural uncertainty about bad events, which for *any* relatively-risk-averse utility function creates a fear-factor effect that can easily dominate quantitative applications of expected-utility theory; (d) such tail-thickened posterior-predictive growth rates have strong repercussions on asset prices that can parsimoniously account for, and even reverse, all three major asset-return puzzles; (e) explanations of macroeconomic asset returns by rational-expectations calibrations and regressions may be illusory because, no matter how much objective data there are, any desired equity premium or riskfree rate can always be reverse-engineered by making tiny, seemingly-innocuous changes in subjective prior beliefs.

This paper begins by noting that macroeconomic asset pricing is dominated by a pervasive subset of rationally formed expectations, which in the literature is sometimes called *REE* for *Rational Expectations Equilibrium*. The key characteristic of *REE* (defining it as a proper subset of the set of all rationally formed Bayesian equilibria) is the imposed extra assumption that the subjective probability

* Department of Economics, Harvard University, Cambridge, MA 02138 (e-mail: mweitzman@harvard.edu). For helpful detailed comments on earlier drafts of this paper, but without implicating them for its remaining defects, I am grateful to Andrew Abel, Daron Acemoglu, Evan Anderson, Roland Benabou, Gary Chamberlain, Vincent Crawford, Xavier Gabaix, Alfred Galichon, John Geweke, Jerry Green, Jerry Hausman, Joseph Kadane, Narayana Kocherlakota, Lawrence Kotlikoff, Stephen LeRoy, Mark Machina, Angelo Melino, Jonathan Parker, Robert Pindyck, Christopher Sims, Jeff Strnad, and Yongli Zhang.

distribution of outcomes believed by agents within an economic system equals the objective frequency distribution actually generated by the system itself. For the purposes of this paper, *REE* is effectively a dynamic stochastic general equilibrium where all reduced-form structural parameters of the data-generating process are known—presumably because they have already been learned previously as some kind of an ergodic limit from a sufficiently large sample. When the *REE* concept of a dynamic stochastic general equilibrium is applied empirically to price assets in a macroeconomic setting, it produces the three major asset-return puzzles described briefly below.

The "equity-premium puzzle" refers to the striking failure of *REE* to explain a historical difference of some six or so percentage points between the average return from a representative stock market portfolio and the average return from a representative portfolio of relatively safe stores of value. Such a large risk premium for equity suggests a fear of the unknown that seems inconsistent with a nonbizarre, comfortably familiar coefficient of relative risk aversion, say with conventional values $\gamma \approx 2 \pm 1$.

The "riskfree-rate puzzle" refers to the five-percentage-point or so discrepancy between the interest rate that is predicted by the *REE* Ramsey formula and what is actually observed. For a plausible risk-aversion coefficient $\gamma \approx 2$ and a plausible rate of pure time preference $\rho \approx 2$ percent, the *REE* Ramsey formula predicts a riskfree interest rate of $r^f \approx 6$ percent, while what people are actually willing to accept to reduce fear of the unknown is $\hat{r}^f \approx 1$ percent.

The term "equity-volatility puzzle" as used here refers to the empirical fact that actual returns on a representative stock market index have a variance some two orders of magnitude larger than the variance of any consumption-dividend-like fundamental in the real economy that might possibly be driving them or that might be relevant for welfare calibration. If comprehensive or representative equity is conceptualized (at a very high level of abstraction) as if acting like a surrogate claim on the consumption dividend produced by the macroeconomy itself, then returns on aggregate equity should (at least very roughly) reflect more-fundamental growth expectations for the underlying real economy. Even allowing, however, for leverage and other actual complications, the way-too-large empirical volatility of equity prices seems badly disconnected from the basic spirit of a real-economy-driven *REE*. Instead of self-confident *REE* investors with sure expectations of objective frequencies generated by an already known stochastic structure (about which nothing further remains to be learned), the whole situation looks and feels more like skittish investors nervously reacting with unsure expectations to unknown deeper forces of shifting structure.

In a *nonergodic* situation where hidden parameters are *evolving*, everyone is perennially uncertain about current structure and learning is *not* converging to a *REE* because no matter how the data are filtered there are not "true" *REE* structural-parameter values to converge to.[1] By postulating known stable structural parameters, *REE* makes the probability density of future growth rates seem more centered and more thin-tailed than it actually is—other things being equal. But integrating out Bayesian uncertainty about parameters controlling the degree of tail-spread of any given "parent distribution" inevitably broadens and thickens the tails of the subjective posterior-predictive "child distribution" that goes into the Euler equation determining asset prices.[2] The point is general, but the particular example carried throughout this paper is of a thin-tailed normal parent distribution that becomes a thick-tailed Student-*t* child distribution from uncertainty about the variance parameter.[3]

A derived implication of the expected-utility hypothesis is that agents having *any* utility function with everywhere-positive relative risk aversion especially dislike uncertainty in the key structural parameters of the stochastic

[1] More technically, this paper shows that when agents are experiencing a dynamically evolving stochastic process that is relevant to asset pricing, the subjective probability measures from Bayesian learning stay uniformly bounded away from the actual data-generating process—even with asymptotically infinite past observations.

[2] An Euler equation is the first-order condition reflecting intertemporal consumption trade-offs that is used to price assets. Euler equations are intended to hold only in subjective expectations, as opposed to holding in large-sample frequencies, a distinction that gets obscured under *REE*.

[3] The Student-*t* density from a large number of observations looks almost exactly like its bell-shaped normal parent, except that the probabilities are somewhat more spread apart, making the tails appear relatively thicker at the expense of a slightly flatter center.

consumption-growth process. An interpretation of why people especially dislike structural ignorance about future consumption is that they dread the thickened-left-tail heightened probability of a negative-growth disaster that they find scary, disruptive, and without precedent. Aversion to structural uncertainty increases both the equity premium and equity volatility, while simultaneously decreasing the riskfree interest rate. The potential influence of tail-thickened growth rates representing structural uncertainty is confirmed just by plugging a Student-*t* distribution into standard asset pricing formulas where a normal usually goes and then noting the reversal of all puzzle-discrepency inequalities requiring explanation. This tail-thickening reversal of what is considered "puzzling" (which is therefore simultaneously a reversal of what needs to be "explained") is a strong force. The same anti-puzzle pattern is shown to occur even with unlimited data from a stochastic growth process whose structural parameters are evolving arbitrarily slowly. Such a degree of nonrobustness means that the usual calibration of asset prices to the standard model of a steady-state-distributed *REE* is questionable because *REE* asset-return outcomes and conclusions are fragile to even the tiniest evolutionary-structural perturbations.

Within *REE*, the financial equilibrium of a small-sample situation having a remote chance of a disastrous out-of-sample happening is dubbed the "peso problem."[4] In a peso problem, possible future occurrences of unlikely bad events that are *not* included in the too-small sample (such as the presumed structure being undermined by a natural or socioeconomic disaster) are taken into account by real-world investors who know the true *REE* data-generating process. Naturally, these rare out-of-sample disaster possibilities are missed by unknowing calibrators simulating past sample frequencies. An artificial *REE* peso premium then appears in the data because to an outside observer it looks like inside investors are being rewarded by an inexplicably high empirical asset return, while actually they are bearing the extra risk of rare disasters in the left tail of the distribution that happen not to have materialized within the limited sample. This paper shows that an asset-pricing equilibrium with a peso problem is not just a hypothetical possibility, but rather it is a generic inevitability that must accompany a learning situation where agents are interchangeable with econometricians trying to infer tail structure from the same incomplete information. A Bayesian translation of a peso problem is that there are insufficient data to construct a reliable posterior distribution based solely upon sample frequencies—i.e., a posterior that is independent of imposed priors. In a Bayesian-learning equilibrium where hidden structural parameters are evolving stochastically, it turns out that asset prices *always* depend critically upon subjective prior beliefs and there are *never* enough data on frequencies of rare tail events for asset prices to depend only upon the empirical distribution of past observations.

The pioneering model of Thomas A. Rietz (1988), later extended by Robert J. Barro (2006), attempts to explain the equity-premium puzzle from within a *REE* framework by thickening the tails of the distribution of growth rates via directly inserting a discrete i.i.d. rare-disaster state having a *known* proportional reduction of consumption occur with *known* probability. This method can be interpreted as essentially arguing through suggestive numerical examples (without abandoning objective-frequency-based *REE*) that a peso problem may apply because the data sample being used in the traditional puzzles literature, which is taken from relatively tranquil historical periods and countries, may be understating the potential for a worst-imaginable-case scenario of large negative future growth rates. A drawback of this approach is the inherent implausibility of being able to meaningfully calibrate *REE* objective frequency distributions of rare disasters (such as world wars, great depressions, global pandemics, geophysical catastrophes, or the like) because *the rarer the event the more uncertain is our estimate of its probability*. I return to the Rietz-Barro model later when, after formally developing the evolutionary-learning apparatus in this paper, a

[4] The name "peso problem" comes from the once puzzlingly high empirical yields on Mexican bonds during a time when the Mexican peso had been pegged to the US dollar at the same fixed exchange rate for decades. Then one day there was a sudden sharp devaluation of the peso against the dollar. After the collapse of the peso, the previous in-sample "peso premium" was explained ex post factum by the small probability of a huge out-of-sample devaluation that investors had understood to be a possibility all along.

more meaningful comparison of the two methods can be made that assesses their very different approaches to statistical inference in the presence of a commonly shared peso problem.

This paper is not the first to investigate the effects of subjective uncertainty on asset pricing. There are several earlier examples having some Bayesian features or overtones.[5] Broadly speaking, these papers explicitly or implicitly suggest that the need for transient Bayesian learning about structural parameters along the path to a *REE* may temporarily reduce the degree of one or another asset-return anomaly. What seems to be missing from previous Bayesian-learning literature, however, is a sense of the sheer power that distribution-spreading structural parameter uncertainty can bring to bear on equilibrium asset pricing, especially when an evolving structure keeps learning relevant forever. In effect, some qualitative implications of structural uncertainty are appreciated in this literature, but not the quantitative magnitude of its permanent dominance over asset-pricing Euler-equation formulas via thickened posterior-predictive tails.

An exception in the vast puzzle-related literature is the admirably terse five-page communication by John Geweke (2001) that applies a Bayesian framework to the most standard model prototypically used to analyze behavior toward risk and then notes the curious fragility of the existence of finite expected utility itself.[6] In a sense the present paper begins by accepting this important nonrobustness insight, but pushes it further to argue that the inherent sensitivity of the standard prototype formulation constitutes a significant clue for unraveling what is driving the asset-return puzzles and for giving them a unified general-equilibrium interpretation that parsimoniously links the stylized facts.

This paper argues that the three macrofinance asset-return puzzles are not nearly so puzzling in a nonergodic Bayesian-learning formulation whose unknown structural parameters are evolving. Instead, the arrow of causality in this unified Bayesian explanation is reversed; the puzzling numbers being observed empirically are trying to tell a parsimoniously consistent story about the subjective revealed-prior distribution of growth-structure uncertainty that investors must implicitly have in their minds to generate such puzzling data patterns. This paper suggests that the "strong force" of evolutionary-structural uncertainty is empirically a far more powerful determinant of asset prices and returns than the "weak force" of known-fixed-structure *REE*-type pure risk. Measured in marginal-utility-weighted units, the subjective probability distribution of tail-thickened posterior-predictive growth prospects is in some critical respects closer to the relatively stormy volatility record of stock market wealth than it is to the far more placid smoothness of past consumption.

## I. The Family of *REE* Asset-Return Puzzles

The core issue for this paper is whether the three asset-return puzzles can be explained by the subjective beliefs of agents regarding structural uncertainty. This section frames the puzzles in a simple *REE* format that is particularly amenable to easing the later transition into a generalization whose nonergodic structure is allowed to evolve over time. For this purpose, a stark endowment-production dual-canonical model is used where everything but the most basic architecture of the model has been set aside. To focus on the big picture, this paper heroically assumes away the details in such diversionary complications as defaults, leverage, illiquidity, taxes, autocorrelation, irrationality, heterogeneous agents, exotic preferences, changing tastes, borrowing constraints, adjustment costs, business cycles, timing frictions, human capital, incomplete markets, idiosyncratic risks, and the like.

Let $t$ denote the present period. From the present perspective, consumption $C_{t+j}$ in future period $t + j$ (with $j \geq 1$) is a random variable, which, for the time being at least, comes from a very general evolutionary stochastic process. The population consists of a large fixed number

[5] Such earlier papers include Robert B. Barsky and J. Bradford DeLong (1993), Allan G. Timmermann (1993), Peter Bossaerts (1995), Michael J. Brennan and Yihong Xia (2001), Andrew Abel (2002), Alon Brav and J. B. Heaton (2002), Jonathan Lewellen and Jay Shanken (2002), and several others.

[6] I am grateful to two readers of an early draft of this paper for informing me of Geweke's pioneering note. Geweke found that a Bayesian formulation similar to what underlies this paper can cause serious convergence problems for indefinite integrals representing expected utility.

of identical people who live forever. The utility $U$ of consumption $C$ is specified by the isoelastic power function

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

with corresponding marginal utility

$$U'(C) = C^{-\gamma}, \tag{2}$$

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

The pure-time-preference multiplicative factor for discounting one-period utility into present utility is $\beta < 1$. At the present time $t$ the representative agent's welfare is

$$V_t = E_t\left[\frac{1}{1-\gamma}\sum_{j=0}^{\infty}\beta^j(C_{t+j})^{1-\gamma}\right], \tag{3}$$

where throughout this paper the expectation operator $E_t$ is understood as being taken over a *subjective* distribution of future growth rates, conditioned on all information available at time $t$. The "stochastic discount factor," or marginal rate of substitution between $C_t$ and $C_{t+1}$, is $M_{t+1} \equiv \beta U'(C_{t+1})/U'(C_t)$, and for any asset $\alpha$ whose gross return in period $t+1$ is $R^{\alpha}_{t+1}$, the relevant Euler equation is

$$\beta E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R^{\alpha}_{t+1}\right] = 1. \tag{4}$$

Later this section will also deal with an *AK*-type linear-production version (with capital $K$ and uncertain aggregate productivity $A$), but first begins with the simplest example of the textbook workhorse formulation of a Lucas-Mehra-Prescott endowment-growth economy, which is ubiquitous as a benchmark point of departure throughout the finance-macroeconomics literature.[7] In this pure exchange model of dynamic general equilibrium, consumption growth is given by an exogenous stochastic process and all asset markets are like phantom entities because no one actually ends up taking a net position in any of them. The paper concentrates on three basic investment vehicles: a "riskfree" asset, "one-period" equity, and "multi-period" equity, all of which are abstractions of reality. In all cases gross returns are asset payoffs divided by asset price, with consumption as numeraire.

The riskfree asset effectively guarantees that this period's consumption will also be paid in the next period, and is approximated in an actual economy by a portfolio of the safest possible stores of value, including hard currency, Swiss bank accounts, US Treasury bills, and inventories of real goods.[8] In the theoretical fruit-tree economy, substituting the payoff of this period's consumption into the Euler equation (4) gives the price of the riskfree asset (normalized for commensurability with equity payoffs to pay out period-$t$ consumption in period $t+1$) as

$$P^f_t = (C_t)^{1+\gamma}\beta E_t[(C_{t+1})^{-\gamma}], \tag{5}$$

while the gross one-period return on the riskfree asset $R^f_{t+1}$ in period $t+1$ is

$$R^f_{t+1} = \frac{C_t}{P^f_t} = \frac{1}{\beta E_t[(C_{t+1}/C_t)^{-\gamma}]}. \tag{6}$$

One-period equity is a hypothetical asset that pays *only* next period's consumption endowment and thereafter expires. The price of this risky asset at time $t$ is

$$P^{1e}_t = (C_t)^{\gamma}\beta E_t[(C_{t+1})^{1-\gamma}], \tag{7}$$

[7] The famous fruit-tree model of asset prices in a growing economy traces back to two seminal articles: Robert E. Lucas Jr. (1978) and Rajnish Mehra and Edward C. Prescott (1985). For applications, see the survey articles of John Y. 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 two review articles above and in the textbook expositions of John H. Cochrane (2001) or Darrell Duffie (2001).

[8] The literature concentrates on very short-term US Treasury bills, but I think this interpretation of a "riskfree" asset is much too narrow. Of course, no asset is completely safe—not even inventories of stored food or medicine. The analysis in Barro (2006), however, seems to suggest that accounting for probabilities of events like defaults on government bonds has little effect on asset prices, at least within his model.

with gross return

$$(8)\qquad R^{1e}_{t+1} = \frac{C_{t+1}}{P^{1e}_t} = \frac{C_{t+1}/C_t}{\beta E_t[(C_{t+1}/C_t)^{1-\gamma}]}.$$

Multi-period equity is approximated in the real world by a broad-based representative index of publicly traded shares of stocks whose aggregation weights mimic the comprehensive wealth portfolio of the entire economy. In the theoretical fruit-tree endowment economy, multi-period equity is modeled abstractly as a claim on the stream of all future consumption dividends. Thus, in period $t$, the ex-dividend price of equity $P^e_t$ is the price of fruit trees claiming ownership of all dividends accruing from time $t + 1$ onward, which by repeated use of the Euler condition can be written as

$$(9)\qquad P^e_t = (C_t)^\gamma \sum_{j=1}^{\infty} \beta^j E_t[(C_{t+j})^{1-\gamma}].$$

The realized gross return on multi-period equity between periods $t$ and $t + 1$ is

$$(10)\qquad R^e_{t+1} = \frac{C_{t+1} + P^e_{t+1}}{P^e_t}.$$

Combining (7), (8) with (9), (10) and rewriting terms gives a tight general connection between the two realized equity returns, expressed symmetrically in welfare-utility fundamentals as

$$(11)\qquad \frac{R^e_{t+1}}{R^{1e}_{t+1}} = \frac{V_{t+1}}{U_{t+1}} \left\{ \frac{E_t[U_{t+1}]}{E_t[V_{t+1}]} \right\}.$$

For any time $t$, multi-period financial wealth in this endowment-exchange economy is

$$(12)\qquad W_t = C_t + P^e_t.$$

Substituting (1), (3), (9) into (12) and cancelling redundant terms gives

$$(13)\qquad \frac{V_t}{U_t} = \frac{W_t}{C_t},$$

which suggests that volatile wealth and volatile consumption have a symmetric relationship to welfare, an important theme that will be pursued further in Section V of the paper.

Everything up to this point works with a very general (*REE* or non-*REE*) stochastic process. For all times $t$, let

$$(14)\qquad X_t = \ln C_{t+1} - \ln C_t$$

be the geometric growth rate of consumption during period $t$. In the rest of this section, I develop the model's implications under the simplifying assumption that growth rates $\{X_t\}$ are i.i.d. with known distribution. In later sections, I relax the assumption of a known fixed structure to show that the conclusions are very different under evolutionary uncertainty.

In the special *REE*-i.i.d. known-distribution case, the riskfree-rate formula (6) in logarithmic form becomes

$$(15)\qquad r^f = \rho - \ln E[\exp(-\gamma X)],$$

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

When the random variable realizations $\{x_t\}$ are i.i.d., it is readily shown from (7) and (9) that the price-earnings ratios $P^{1e}_t/C_t$ and $P^e_t/C_t$ for both forms of risky-asset equity are constants independent of $t$ and (from combining (8), (10), (11), (14)) that

$$(16)\qquad R^e(x) = R^{1e}(x) = \frac{\exp(x)}{\beta E[\exp((1-\gamma)X)]}.$$

Taking the natural logarithm of the expected value of (16) and subtracting (15), the average equity premium in each period (under the i.i.d.-growth assumption) is

$$(17)\qquad \ln E[R^e] - r^f$$
$$= \ln E[R^{1e}] - r^f$$
$$= \ln E[\exp(X)] + \ln E[\exp(-\gamma X)]$$
$$- \ln E[\exp((1-\gamma)X)].$$

Equation (17) is a theoretical formula for calculating the equity risk premium, given any coefficient of relative risk aversion $\gamma$, and, more importantly here, given the known i.i.d. probability distribution of the uncertain future growth rate $X$. Concerning the relative-risk-aversion taste parameter $\gamma$, there seems to be some rough agreement 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.

Preferences are standardly conceptualized as being fixed over time. By contrast, much less is known about what is the appropriate probability distribution to use for representing future growth rates. Even under the best of circumstances (with a known fixed 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 $X$. 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 actually saying about its likely parameter values. The functional form that naturally leaps to mind is the normal distribution

$$X \sim N(\mu, V), \tag{18}$$

which is the ubiquitous benchmark case assumed throughout the asset-pricing literature.

The expository literature proceeds by implicitly presuming that the "true" structural parameters $\mu$ and $V$ are constants already learned by the agents inside the economy (although perhaps not yet learned by an outside observer), and then continues on by substituting the normal distribution (18) into formula (17), which reduces (17) 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 (17) a distribution accounting for structural-parameter sampling error (like the Student-*t*).

Let $\hat{x}$ be the sample mean and $\hat{V}$ be the sample variance of a long time series of past growth rates. Implicitly in the *REE* interpretation, the sample size is presumed large enough to make $\hat{x}$ 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 $(\hat{x}, \hat{V})$ for $(\mu, V)$ in their subjective Euler equations. With (18), using the formula for the expectation of a lognormal random variable and cancelling redundant terms simplifies (17) into the standard expression

$$\ln E[R^e] - r^f = \gamma V[X], \tag{19}$$

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

Considering the United States 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[9] to be $\ln E[R^e] \approx 7$ percent. The historically observed real return on an index of the safest available short-maturity bills is less than 1 percent per annum, implying for the equity premium that $\ln E[R^e] - r^f \approx 6$ percent. The mean yearly growth rate of US per capita consumption over the last century or so is about 2 percent, with a standard deviation taken here to be about 2 percent, meaning $\hat{V} \approx 0.04$ percent. Suppose $\gamma \approx 2$. Plugging these values into the right-hand side of (19) gives $\gamma\hat{V} \approx 0.08$ percent.

Thus, the actually observed equity premium on the left-hand side of equation (17) exceeds the estimate (19) of the right-hand side by some 75 times. If this were to be explained with the data above 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 magnitude. 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

[9] The following numbers are from Mehra and Prescott (2003) and/or Campbell (2003), who also show roughly similar summary statistics based on other time periods and other countries. Too short a time series prevents treatment as a stylized fact of an "overpriced portfolio-insurance puzzle" (empirically, paper profit-returns from selling unhedged out-of-the-money index put options would have been extraordinarily high over the restricted sample period for which data are available), which is consistent with the model of this paper.

premium as is observed. Of course, the underlying model is extraordinarily crude and can be criticized on any number of valid counts. Still, two orders of magnitude seems like an awfully large base-case discrepancy to be explained away ex post facto, even coming from a very primitive model.

Turning to the riskfree-rate puzzle, the meaning given in the asset-pricing literature to equation (15) parallels the interpretation given to the equity premium formula. The expository literature postulates the normal distribution (18), but then imagines that the representative agent ignores the statistical uncertainty inherent in estimating the "true" values of $E[X] = \mu$ and $V[X] = V$. Using in (15) the formula for the expectation of a lognormal distribution gives

$$r^f = \rho + \gamma E[X] - \frac{1}{2}\gamma^2 V[X], \tag{20}$$

which is a familiar generic equation appearing in one form or another throughout equilibrium stochastic-growth interest-rate theory. (Its origins trace back to the famous neoclassical Ramsey optimal-growth model of the 1920s.)

Noncontroversial estimates of the relevant parameters appearing in (20) (calculated on an annual basis) are: $\hat{x} \approx 2$ percent, $\hat{V} \approx 0.04$ percent, $\rho \approx 2$ percent, $\gamma \approx 2$. With these representative parameter values plugged into the right-hand side of (20), the left-hand side of (15) becomes $r^f \approx 5.9$ percent. When compared with an actual real-world riskfree rate $\hat{r}^f \approx 1$ percent, the theoretical formula is too high by $\approx 4.9$ percent. This gross discrepancy is the riskfree-rate puzzle.

As if all of the above were not vexing enough, there is also the enigmatic appearance in the data of what I am calling the "equity-volatility puzzle." From (16) it must hold identically for all $j$ that

$$r^e_{t+j} - E[r^e] = x_{t+j} - E[X], \tag{21}$$

and therefore in this ultra-simplified i.i.d. *REE* economy the entire financial-economic system vibrates in unison. According to (21), the realized deviation from the mean of continuously compounded financial returns on multi-period equity-wealth $r^e - E[r^e]$ should coincide exactly with the realized deviation from the mean of its underlying real fundamental $x - E[X]$, implying that all higher-order moments of the two distributions should match. An exact lock-step coincidence is asking way too much, but it is painfully obvious that even just the two empirical second moments are very badly mismatched in the time-series sample because the standard deviation of equity returns $\hat{\sigma}[r^e] \approx 17$ percent is much bigger than the standard deviation of growth rates $\hat{\sigma}[x] \approx 2$ percent. This is taken to be the "equity-volatility puzzle."

Equity returns are volatile relative to almost anything else in the economy. For the model of this paper, I understand the "equity-volatility puzzle" to be the stylized fact that, contrary to the simple theory, the variance of historical returns to a broad-based stock market index is about two orders of magnitude greater than the variance of the welfare-relevant fundamental of a consumption payout, for which representative equity is supposed to be the surrogate claimant. Conforming once again here with the familiar macro-asset-pricing puzzle pattern, it turns out that substituting alternative formulations (including an equity claim on consumption dividends that is leveraged to an empirically plausible degree) can lessen the initial orders-of-magnitude discrepancy (here of the degree of variance mismatch between the welfare-relevant real-production side of an economy and its dual financial-wealth side), but, as usual, something central of the mystery remains that still seems way off base.

Comprehensive financial wealth $W$ in an endowment-exchange *REE* is mathematically equivalent to comprehensive production capital $K$ in the optimal stochastic growth problem of a linear-production $AK$-type model with uncertain aggregate productivity $A$. Leaving aside details of a rigorous proof, the identification key to this endowment-production duality principle is $R^e_{t+j} \leftrightarrow A_{t+j}$ (or $r^e_{t+j} \leftrightarrow \ln A_{t+j}$) and $W_{t+j} \leftrightarrow K_{t+j}$, where the symbol "$\leftrightarrow$" means mathematical equivalence for all $j \geq 0$. "Comprehensive production capital" $K$ is intended here to represent the capitalized value (at stochastic general equilibrium prices) of returns to *all* factors of production in the economy—not only reproducible capital like equipment and structures, but also human and intangible capital, as well as labor, land, minerals, and so forth. In the $AK$ production version with comprehensive $K$ and stochastic $A$, the control variable $C_{t+j}$ is *chosen* (just before $A_{t+j+1}$ is realized) to maximize $V_{t+j}$ in an expression of the form (3). The system's state-transition equation is

$$(22)\quad K_{t+j+1} = A_{t+j+1}[K_{t+j} - C_{t+j}] \leftrightarrow W_{t+j+1} = R^e_{t+j+1}[W_{t+j} - C_{t+j}],$$

where the dual-equivalent comprehensive-wealth equation of motion in (22) comes from (12), (10). Therefore, it matters not whether stochastic consumption $\{C_{t+j}\}$ is first taken as the primitive driver in the endowment economy while stochastic returns $\{R^e_{t+j}\}$ are derived and subsequently taken as primitive-driver stochastic productivity $\{A_{t+j}\}$ ($= \{R^e_{t+j}\}$) for the production economy, or whether stochastic productivity $\{A_{t+j}\}$ ($= \{R^e_{t+j}\}$) is first taken as the primitive driver in the production economy while stochastic optimal consumption $\{C_{t+j}\}$ is derived and subsequently taken as primitive driver for the endowment economy, because the two stochastic equilibria are not operationally distinguishable to an outside observer.

Why is the duality between endowment and production formulations of the *same* underlying model important for this paper? Because the venerable "discipline imposed by general equilibrium modeling" (which is an important rationale for using the Lucas-Mehra-Prescott fruit-tree format in the first place rather than some partial-equilibrium format) here is practically shouting at us that $W$ and $K$ (as well as $R$ and $A$) are identical under *REE*, or at most they are two sides (financial and real) of the same coin. Therefore, if the *REE* equivalence between comprehensive financial-wealth and aggregate production-capital does *not* show up anywhere in the real economy—because the empirical variance of financial equity-wealth is two orders of magnitude bigger than the empirical variance of practically *anything* in the real economy—then from the "discipline imposed by general equilibrium modeling" it is simply unclear (under *REE*) *which* interpretation (the "comprehensive wealth-capital of consumption" or the "consumption of comprehensive wealth-capital") should take precedence for calibrating welfare—a consequential theme stressed repeatedly throughout the paper.

Summing up the scorecard for this super-simple i.i.d.-normal application of a dual-canonical endowment-production *REE* model, we have three strong orders-of-magnitude contradictions with reality. Some heuristic intuition for what is coming up next in the paper can be gleaned simply by performing the experiment of substituting a Student-*t* distribution from any large (but finite) sample of observations for the normal distribution in formulas (15) and (17). When the limits of the relevant indefinite integrals containing the Student-*t* distribution are evaluated, it is readily seen from formula (15) that $r^f \to -\infty$, while from (17) 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 core question: *what are we supposed to be explaining here?* Should we be trying to explain the *puzzle pattern*: why is the actually observed equity premium 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 *antipuzzle pattern*: why is the actually observed equity premium 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)? It seems difficult not to conclude that something fundamental is deeply wrong in the underlying *REE* formulation when the contradictions are so unsettling from simply recognizing that the distribution implied by the normal *conditioned on finite realized data* is the Student-*t*.

Intuitively, a normal density "becomes" a Student-*t* from a tail-thickening spreading-apart of probabilities caused by the variance of the normal having itself an (inverted gamma) probability distribution. There is then no surprise from expected utility theory that people are more averse qualitatively to a relatively thick-tailed Student-*t* child distribution than they are to the relatively thin-tailed normal parent which begets it. A much more surprising consequence of expected utility theory is the *quantitative strength* of this endogenously derived aversion to the effects of unknown variance-structure. The story behind this quantitative strength is that thickened posterior left tails represent structural uncertainty about rare disasters that terrify people. This fear-factor effect holds for *any* utility function having everywhere-positive relative risk aversion. The next section formalizes the idea that nonergodic parameter uncertainty leads to a permanently tail-thickened distribution of growth rates that can cause expected marginal utility to blow up—and shows a rigorous sense in which "containing the Student-$t$-explosion" necessitates an unavoidable dependence of asset prices upon some form or another of exogenously imposed subjective beliefs.

## II. Hidden-Structure Expectations of Future Growth

Perhaps surprisingly, it turns out for asset-pricing implications that the most critical issue involved in Bayesian learning about the probability density of future growth rates is the unknown variance (whose role in this context is to represent more generally *all* parameters influencing the tail-spread of *any* distribution). The case of the mean unknown but variance known garners the lion's share of attention in the asset-price learning literature, partly because of its greater analytical tractability and partly because of a widespread impression that with large samples in continuous time it is relatively easy to learn the true variance. For simplification, it is convenient here to be able to postulate straightaway a situation 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 analysis when $E[X]$ and $V[X]$ are both unknown involves more notation but is essentially the same.

To indicate where the argument is now and where it is leading, the assumptions behind the core model to be used throughout the rest of the paper are stated formally here. The Euler equation (4) holds for the utility function (1) in *ex ante subjective expectations* (as contrasted with holding in *ex post realized frequencies*—more on this distinction later). The presumed conditional-i.i.d. probability distributions are: $X \sim N(\mu,V)$ and $r^e \sim N(E[r^e],V[r^e])$. Six constants of the model are effectively assumed known: $E[r^e]$, $V[r^e]$, $r^f$, $\rho$, $\gamma$, $\mu$. Only one structural parameter is evolving and must be estimated statistically: $V = V[X]$.

The first order of business in this section is to show that the startling asset-pricing *antipuzzle pattern* (from Student-$t$-distributed growth rates, described at the end of the last section of the paper) *persists when there are tiny variance shocks—even with infinite data*.

If there were an infinite sample of bygone observations, then, at any time $t$,

$$\nu_t = \frac{1}{k}\sum_{j=1}^{\infty}\left(1-\frac{1}{k}\right)^{j-1}(x_{t-j}-\mu)^2 \tag{23}$$

would represent an exponentially weighted average back to the remotest past of all realized variances of previous growth rates, which gives progressively greater influence to more recent events. The parameter $k$ appearing in (23) is treated throughout this paper as a known positive constant called the *effective* sample size. It would be neat if we were able to show that the standardized random variable $(X_t-\mu)/\sqrt{\nu_t}$ is distributed as Student-$t$ with $k$ degrees of freedom, which would make the probability density function of $x_t$ be

$$f(x_t|\nu_t,k) \propto \left(1+\frac{(x_t-\mu)^2}{k\nu_t}\right)^{-(k+1)/2}, \tag{24}$$

because then we might have some literary license to tell a simple story *as if* just before it is observed the random variable $X_t$ is distributed as the Student-$t$ statistic naturally associated with the outcome of "running a regression" on a sample of $k+1$ past realizations to estimate and predict $x_t$. Having (24) hold for *all* periods $t$ is heuristically like randomly losing one of $k+1$ fictitious observations during each period, which is replaced by a new observation at the period's end—thus making the Student-$t$ distribution *always* have $k$ degrees of freedom. Intuitively, the *effective* sample size $k$ might remain constant over time (instead of increasing with the *actual* sample size, thereby forcing the Student-$t$ child distribution to converge to its parent normal) if the information gained from a new realization of $x_t$ in (24) is counterbalanced by the information lost from a hidden shock to the latent variance. What comes next gives a rigorous Bayesian-learning rationale for this intuitively appealing story about Student-*t*-distributed growth rates having an unchanging number of degrees of freedom. The basic underlying analytical strategy here is simple: take advantage of the fact that gamma-normal conjugacy under multiplicative beta shocks to the gamma generates Student-*t* distributions, even with infinite data. However, the detailed architecture of the model that follows is intricate and may be challenging even for someone acquainted with Bayesian methods. A reader willing to take this Student-*t* story on faith may wish to skim over mathematical details in favor of a general impression of how it all hangs together.

Presuming the normal specification (18), for analytical convenience, the Bayesian literature tends to work with the random variable $\theta \equiv 1/V$, called the *precision*. Assume at any time $t$ that, conditional on $\theta_t$, the random-variable growth rate $X_t$ is independently drawn from a normal distribution,

$$X_t|\theta_t \sim N(\mu, 1/\theta_t). \tag{25}$$

The stochastic growth process is conceptualized as if having started at time $\tau$, where $\tau$ is an arbitrarily large negative number. A joint distribution for $\{X_{\tau+1},\dots\}$ is constructed by first conditioning on some latent process $\{\theta_{\tau+1},\dots\}$, which is never observed, and subsequently integrating out the $\theta$'s to get an unconditional distribution for the $X$ process. At any time $t$ the agent can then form a conditional distribution for $\{X_t,\dots\}$ conditioned on realizations $\{\dots,x_{t-1}\}$, which is the reduced-form distribution that ultimately matters for the model.

Let $\underline{\theta}$ and $\bar{\theta}$ be a priori–imposed positive lower and upper bounds on $\theta$. Their values will later be explained, but for now $\underline{\theta} > 0$ is a given arbitrarily *small* number while $\bar{\theta} < \infty$ is a given arbitrarily *large* number. A condition like $\underline{\theta} \le \theta_t \le \bar{\theta}$ is needed for technical reasons to keep expected utility bounded, and this mathematical need to prevent unbounded utility in key asset-pricing formulas necessitates the intricacy of the analysis undertaken here. Given the prior density $p_\tau(\theta_\tau)$ (conceptualized as if imposed at initial time $\tau$, which is positive for $\underline{\theta} \le \theta_t \le \bar{\theta}$ and is dogmatically set to zero elsewhere), the joint probability distribution of the whole sample of $\{\theta_{\tau+1},\dots,\theta_t\}$ taken together is generated recursively from well-defined conditional probability distributions as follows. Suppose that, at the end of period $t-1$ (or the beginning of period $t$), the situation is as if (in classical-frequentist terms) the hidden value of the precision were hit by an unobserved nonnegative multiplicative i.i.d. shock $\zeta_t$, making the transition equation be

$$\theta_t = \zeta_t \theta_{t-1}, \tag{26}$$

except when (26) conflicts with the a priori–restricted (by the Bayesian prior) range of $\theta_t$ being $\underline{\theta} \le \theta_t \le \bar{\theta}$, which always takes precedence. Formally, this is implemented in the model by placing two inward-reflecting barriers at $\theta = \underline{\theta}$ and $\theta = \bar{\theta}$, which prevent excessively extreme values of $\theta_t$ from ever emerging by trimming at both ends the distribution of $\zeta_t|\theta_{t-1}$, thereby imposing the restriction $\zeta_t|\theta_{t-1} \in [\underline{\theta}/\theta_{t-1}, \bar{\theta}/\theta_{t-1}]$. If the probability density function of $\theta_{t-1}$ at time $t-1$ is $p_{t-1}(\theta_{t-1}|x_{\tau+1},\dots,x_{t-1})$ and the i.i.d. density of $\zeta_t$ in (26) is $\pi(\zeta)$, then from Bayesian updating the distribution at time $t$ of $\theta_t|x_{\tau+1},\dots,x_t$ is here

$$p_t(\theta_t|\dots,x_t) \propto \sqrt{\theta_t}\exp\left(-\frac{\theta_t(x_t-\mu)^2}{2}\right) \int_{\underline{\theta}}^{\bar{\theta}} p_{t-1}(\theta_{t-1}|\dots,x_{t-1})\,\pi\left(\frac{\theta_t}{\theta_{t-1}}\right) d\theta_{t-1} \tag{27}$$

for $\underline{\theta} \le \theta_t \le \bar{\theta}$, and is $p_t(\theta_t) = 0$ elsewhere.

For the sake of analytical tractability—in order to be able to take advantage of the convenient closed-form properties of the noncontroversial family of normal-gamma-beta self-conjugate distributions—the i.i.d.-multiplicative-shock density function $\pi(\zeta)$ needs to be a (transformed) beta distribution (because the product of a gamma r.v. with a beta r.v. is a gamma r.v.). But then simply to have notation compatible with equation (23) requires the particular form

$$\pi(\zeta|k) \propto \left[\left(\frac{k-1}{k}\right)\zeta\right]^{\frac{k-3}{2}} \times \left[1-\left(\frac{k-1}{k}\right)\zeta\right]^{-\frac{1}{2}} \tag{28}$$

for $0 < \zeta < k/(k-1)$, where $k$ is some extremely large number interpreted as a known constant parameter. From applying standard textbook formulas it can readily be confirmed that $E[\zeta] = 1$ and $V[\zeta] = 2/(k(k+1) - 2)$, so that $k$ is an inverse measure of the variance $V[\zeta]$ of the multiplicative shock (26), (28).[10]

While it is possible to construct the prior distribution $p_\tau(\theta_\tau)$ around a quite general specification of a priori information and to show that conclusions are robust to a much broader class of priors, it comes at some cost in required mathematical detail. Instead, this paper opts for maximum analytical transparency under the circumstances by working directly with a particularly simple and very intuitive parametric family of prior distributions created as follows. Let $Y$ be a uniformly distributed random variable normalized so that $E[Y] = 0$ and $V[Y] = 1$, meaning the probability density function of $y$ is $\psi(y) = 1/\sqrt{12}$ for $-\sqrt{3} \le y \le +\sqrt{3}$ and $\psi(y) = 0$ elsewhere. For any given positive parameters $\mathcal{I}$ and $\nu$ introduce the random variable $\theta_\tau$ defined by the equation

$$(29) \qquad \ln\theta_\tau = y/\mathcal{I} + \ln(1/\nu),$$

so that $\ln\theta_\tau$ has a uniform distribution of width $\sqrt{12}/\mathcal{I}$ centered on $\ln(1/\nu)$. It is important to note for future reference that $E[\ln\theta_\tau] = \ln(1/\nu)$ and $V[\ln\theta_\tau] = 1/\mathcal{I}^2$. From the Jacobian inverse formula, the probability density of $\theta_\tau$ corresponding to the transformation (29) is

$$(30) \qquad p_\tau(\theta_\tau|\nu,k,\mathcal{I}) = \frac{\mathcal{I}}{\theta_\tau}\psi((\ln\theta_\tau + \ln\nu)\mathcal{I}),$$

which means that $p_\tau(\theta_\tau)$ is a proper reference prior of the well-known form $\propto 1/\theta_\tau$ within its range $[\underline{\theta}(\nu,\mathcal{I}), \bar{\theta}(\nu,\mathcal{I})]$, where $\underline{\theta}(\nu,\mathcal{I}) \equiv \exp(-\sqrt{3}/\mathcal{I})/\nu$ and $\bar{\theta}(\nu,\mathcal{I}) \equiv \exp(+\sqrt{3}/\mathcal{I})/\nu$. If a reader desires to understand on a rigorous mathematical level where the results of this paper are coming from, then it is critical to see clearly that since specification (29), (30) dogmatically restricts the prior $\theta_\tau$ to have positive probability *only* within the closed interval $[\underline{\theta}(\nu,\mathcal{I}), \bar{\theta}(\nu,\mathcal{I})]$, the *same* dogmatic restriction is inherited forever thereafter by all subsequent Bayesian-updated posteriors of $\theta_t | t > \tau$.

For expositional simplicity and without significant loss of generality, the paper pretends that there are infinite past data-observations, meaning here the limiting case $\tau \to -\infty$ is imposed. Priors in standard usage are often taken as given at the beginning of the stochastic process and agents don't later get to change the prior, but this strict temporal order is more of a modeling convention than any kind of fundamental consistency requirement of Bayes's theorem itself—which formally allows "prior" beliefs to be influenced by inspecting "later" data.[11] Especially for large time series, the prior should not be envisioned literally as having once upon a time been carved in stone and rigidly thereafter as being forever immutable. The most relevant use of a prior is for sensitivity analysis: to show the range of possible posterior-predictive inferences or outcomes that the data can support. The prior is best conceptualized as an inherently flexible instrument representing nondata judgements whose primary purpose is to aid the decision maker by helping to answer what-if back-and-forth hypothetical questions of the form: if with my present state of mind I had (in predata times, here in eons long past) hypothesized thus and such parameter values for my prior distribution, then later, conditioned upon the realized data (here a huge sample of past growth rates), what are these hypothesized prior parameter values now saying about what I might expect from the future? In this spirit $\mathcal{I}$

[10] Under the extreme case $\underline{\theta} = 0$, standard limit theory of stochastic processes implies here for $k < \infty$ that $\theta_t \to 0$ almost surely as $t \to \infty$, so that "eventually" an implausibly extreme equilibrium will emerge. However, this stochastic limit result is only pointwise convergent (as opposed to being uniformly convergent for all $k < \infty$), meaning the effect can be postponed indefinitely simply by choosing a sufficiently big value of $k$. (Arbitrarily large $k$ is taken as the base case throughout the paper.) Results in Neil Shephard (1994) indicate how this problem might be eliminated altogether for *any* given $k < \infty$ (when $\underline{\theta} = 0$) by a slight transformation that turns (26) into a random walk in $\ln\theta_t$—a strategy not followed here only because it needlessly complicates the notation and analysis without altering the fundamental insights or conclusions. As will later be elaborated, the strong prior-sensitive results of this paper are robust and they do not depend on specification (26)–(28), which has been chosen primarily for relative ease of manipulation.

[11] In practice of course, this approach is routine because as Sherlock Holmes explained to Watson: "It is a capital mistake to theorize before one has data." Otherwise, Bayesian updating is too mechanically predetermined to be a realistic description of how people actually make inferences.

is a quasi-parameter that, while formally fixed, is here actually intended to be varied to test the robustness of asset prices.

The quasi-constant parameter $\mathcal{I} = 1/\sigma[\ln \theta_{-\infty}]$ is called the *information content of the prior* because (when evaluated at $\nu = \nu_t$) $\mathcal{I}$ quantifies how *informative* (or precise) the currently selected prior distribution is in pinning down the subjective predata prior estimate of $\theta_{-\infty}$ as being "close" to the point estimate of what was subsequently observed in an unboundedly large sample of data. The role of $\mathcal{I}$ here is to guide the decision-making agent by prompting the thought-experimental sensitivity-analysis question: after centering my hypothetical prior on the point estimate $\hat{\theta}_t = 1/\nu_t$ from an infinite time series of past observations, how much does my posterior prediction of future growth rates now depend upon the imposed spread of this prior, which was hypothetically as-if-fixed infinitely long ago? An implicit subtext is that nobody has the foggiest notion in the world about what is actually an a priori–appropriate value of the information content $\mathcal{I}$, which hypothetically reflects underlying nondata prior thoughts at some infinitely remote past time, so that, by the logic of *REE,* its value had better not matter in the slightest. A favorite default setting would be the case $\mathcal{I} \to 0$ representing the well-known textbook statistical situation of a "noninformative" (or "diffuse" or "vague") prior for the precision $\theta_{-\infty}$, which turns (30) into the popular parameterization-invariant improper reference prior $p_{-\infty}(\theta_{-\infty}|\nu, k, \mathcal{I} \to 0) \propto 1/\theta_{-\infty}$ for all $\theta_{-\infty} > 0$ and is then the exact Bayesian counterpart to the standard classical normal-linear regression case (for which it will be shown that the distribution of $X_t$ approaches the Student-*t* form (24) as $\mathcal{I} \to 0$). The extreme opposite situation of infinite informativeness $\mathcal{I} \to \infty$ corresponds to the perfect-knowledge a priori, no-learning-required situation where $X \sim N(\mu, \nu)$. Note that within this Bayesian-learning setup the value of $\theta_t$ is always obscured by hidden uncertainty, but $\nu$, $k$, and $\mathcal{I}$ are treated as if they are known positive parameters.

*Conditioned on* (here infinite) *realized past data* $\{x_{t-j}\}$ for $j \geq 1$, let the posterior subjective probability density function of the precision under Bayesian iterative-updated learning at time *t* be $p_t(\theta_t|\nu, k, \mathcal{I})$ as defined by the stochastic system (25)–(30) (for $\tau \to -\infty$). We next apply a basic recursive property of the family of normal-gamma-beta conjugate distributions (when $\mathcal{I} = 0$) to the particular situation of this paper.

THEOREM 0: *Conditioned on infinite realized past data, at time t the posterior-predictive probability density function* $p_t(\theta_t|\nu, k, \mathcal{I})$ *is continuous in* $\nu, k, \mathcal{I}$ *and as* $\mathcal{I} \to 0$ *has the following limiting distribution for all positive* $\nu$ and $k$:

$$(31)\quad p_t(\theta_t|\nu, k, \mathcal{I} \to 0) \propto \theta_t^{\frac{k}{2}-1} \exp\left(-\frac{k\nu_t}{2}\theta_t\right),$$

*when* $\theta_t > 0$ *(and* $p_t(\theta_t) = 0$ *elsewhere), where the state variable* $\nu_t$ *is defined by (23).*

PROOF:

See the Mathematical Appendix.

The gamma distribution (31) has mean $1/\nu_t$ and variance $2/k\nu_t^2$, so that for $\mathcal{I} \to 0$ the value of $k$ selected as a primitive in (28) ultimately ends up controlling the variability of the posterior distribution of the precision. A nonevolving structure corresponds here to the special situation $k \to \infty$, for which case a conventional application of Bayes's rule allows the "true" precision to be learned exactly (for any given $\mathcal{I} > 0$) as the asymptotic-ergodic limit of the average from an infinite number of past observations, representing an idealization perhaps most easily imagined in continuous time.

After integrating out the precision from the conditional-normal distribution (25), the unconditional or marginal posterior-predictive probability density function of the future growth rate $X_t$ is:

$$(32)\quad g_t(x_t|\nu, k, \mathcal{I}) \propto \int_0^\infty \sqrt{\theta} \exp(-\theta(x_t - \mu)^2/2) p_t(\theta|\nu, k, \mathcal{I})\, d\theta.$$

For any given positive $k$, as $\mathcal{I} \to 0$, straightforward brute-force integration of (32) for the situation (31) shows that $g_t(x_t|\nu, k, \mathcal{I} \to 0)$ converges to the Student-*t* distribution (24) with $k$ degrees of freedom. (Any Bayesian textbook shows that a normal with gamma precision becomes a Student-*t*). Speaking generically, with power utility the formula for "expected future marginal utility" or "expected stochastic discount factor" or "expected pricing kernel" reflects the mathematical properties of the moment generating function of $X$. The moment generating function (m.g.f.) of a Student-$t$ distribution such as (24) is unboundedly large because the defining integral diverges to plus infinity as $\mathcal{I} \to 0$ in (32), which causes the explosion of expected marginal utility, which creates the startling antipuzzle pattern (described at the end of Section I), by reversing dramatically what needs to be explained—only the same thing is happening here with *infinite* data and, when $k$ is indefinitely big, with super-slow-motion evolutionary change unfolding at an *infinitesimal* pace. A situation can therefore always be synthesized where the expected stochastic discount factor is made to become arbitrarily large simply by choosing for (32) a sufficiently small value of $\mathcal{I}$, no matter what value of $k < \infty$ has been given. Note that as $k \to \infty$ *for any given* $\mathcal{I} > 0$, the moment generating function of the distribution $g_t(x_t | \nu, k, \mathcal{I})$ converges *pointwise* to the m.g.f. of $N(\mu, \nu_t)$, but this convergence is *not uniform for all* $\mathcal{I} > 0$, which accounts for the counterintuitiveness of the central finding of this paper that innocuous changes in subjective prior beliefs can have more effect on asset prices than huge samples from a *REE* data-generating process. It is important (for appreciating this paper's generality) to know that the unboundedness potential for $E_t[M_{t+j}] = E_t[\beta^j U'(C_{t+j})/U'(C_t)]$ (with $j \geq 1$) is generic by comprehending that *utility isoelasticity per se is inessential* to the argument, because as $\mathcal{I} \to 0$ the expected stochastic discount factor $E[M] \to +\infty$ *for any relatively risk-averse utility function*—i.e., for *any* $U(C)$ satisfying the minimal curvature requirement: $\inf_{C>0}\{-CU''(C)/U'(C)\} > 0$.

Expressed in Bayesian asset-pricing language, there is one especially crucial bare-minimum prerequisite for the frequentist law-of-large-numbers justification behind calibration or inference to be valid. The critical Bayesian-translated prerequisite behind the classical notion to "just let the data speak for themselves" is that as the number of observations increases without bound, asset-pricing expectation formulas involving marginal utility should become uniformly free of the prior, no matter how uninformative it may be. To have *REE* serve as a robust and trustworthy basis upon which to understand asset returns presupposes that the observed data should asymptotically dominate uniformly (in marginal utility space) *any* reasonable representation of a not-very-informative prior distribution of beliefs—meaning here that the past data information should asymptotically override the influence of *any positive value* of $\mathcal{I}$. Asymptotic dominance of the data over the prior often accompanies an unchanging-structure environment, but such ergodicity does not emerge here, essentially because the stochastic process is evolutionary and learning never "catches up" with the moving target of the unobservable "true" value of $\theta$ (except for the extreme case $k = \infty$). From its very first application to a macroeconomic finance model, therefore, *REE* is a seriously misleading equilibrium concept for pricing assets because it is describing an unstable knife-edge balance in price distributions, having probability-of-existence measure zero, which unravels completely in the presence of even an infinitesimally small bit of evolutionary-structural uncertainty. *For any given* $k < \infty$, the informativeness parameter $\mathcal{I}$ chosen for the prior manifests itself as a smear of background uncertainty that refuses, even with the interdiction of infinite past data, to relinquish its potentially decisive hold on influencing present expectations of future stochastic discount factors.

Because they can be driven to an arbitrary extent by tiny changes in the assumed information content of subjectively imposed prior beliefs, even with infinite data, asset returns are highly reactive to the sentiment fluctuations and mood swings of fickle investors. Very slight shifts in $\mathcal{I}$ can make answers to basic asset-pricing questions come out very differently. Asset prices are in this sense *always* peculiarly vulnerable to subjective judgements about the possibilities of bad future evolutionary mutations of history and can *never* rely solely on the frequency distribution of past events. It follows that classical asset-pricing regressions (and calibrations) trying to fit ex post empirical realizations of an Euler condition in *REE* may be fundamentally misspecified, and perhaps it becomes more understandable that such a stationary-frequency pure-recurrent-risk methodology often ends up effectively rejecting the Euler equation itself by producing pricing errors and paradoxes. The basic message of this paper is that an asset-pricing equilibrium must of necessity be based upon sensitivity to nondata judgements for which a change in the information content of subjective prior beliefs always has the potential to trump objective data-evidence, even with infinite data. This basic message provides the missing link in a unified Bayesian-evolutionary approach capable of connecting parsimoniously the three asset-pricing puzzles. Whether such a theory is better labeled stationary or nonstationary in one or another particular state-space of underlying structure (or meta-structure) is essentially beside the point here. The substantive issue is that no amount of data generated by this model could ever enable a calibrator or econometrician to disconnect the posterior-predictive stochastic discount factor from the effects of subjective prior information in order to recover some hypothetical prior-belief-free, purely data-determined value of $E_t[M_{t+j}] = E_t[\beta^j U'(C_{t+j})/U'(C_t)]$ (with $j \geq 1$).

An outside observer cannot know *directly* what value of $\mathcal{I}$ describes an investor's prior beliefs, as $\mathcal{I}$ can at best be inferred only indirectly from the data (given $k$). At any time $t$ the state variable for the stochastic process (25)–(28) is considered to be $\nu_t$ defined by (23). Throughout the rest of this paper we will "just let the data speak for themselves" in a subjective Bayesian (as opposed to objective frequentist) sense by telling us in effect what is the revealed-prior information content $\mathcal{I}(\nu)$ that real-world investors must *implicitly* be using for their priors in state $\nu = \nu_t$ to replicate (in an artificially simulated data-generating process based upon this model) the empirical pattern of one or another "stylized fact" observed in real-world data. Taking (32) as the representative agent's posterior-predictive probability density function for future growth rates, the next three sections of the paper are devoted to exploring what are intended to be applications to an analytically tractable partial-equilibrium *suggestive example* of the general theme that (contrary to *REE*) subjective prior beliefs inhabit such a large state-space in marginal utility that a data-generating process based on the evolutionary model of this paper *could* readily be made to generate the asset-return patterns observed as stylized facts in the data. The examples are just "suggestive" because many other factors may also be involved in explaining the puzzles. The only position being taken here is that with evolutionary uncertainty the effect of prior beliefs on asset prices is potentially so powerful that it *could by itself* quite easily generate the observed patterns—and therefore it seems highly unlikely that prior beliefs *do not* play a major role. For each such "suggestive example" the sharpest insight comes from having in mind the mental image of a double-limiting situation where $k \to \infty$ and $\mathcal{I} \to 0$ simultaneously, so that the value of $\nu_t$ defined by (23) approaches some known fixed constant and the probability density function $g(x|\nu, k \to \infty, \mathcal{I} \to 0)$ defined by equation (32) converges to the normal distribution $x \sim N(\mu, \nu_t)$. Operationally, this prototype double-limiting situation comes arbitrarily close to the standard familiar textbook-workhorse *REE* case of growth-rate risk appearing to be i.i.d. normal with known parameters—only the model never quite gets to such an i.i.d.-normal distribution because some very small (but nevertheless consequential for asset pricing) new uncertainty evolves whenever $k < \infty$.

Such an extreme thought experiment forces most of the "action" to occur in the farthest reaches of the left tail of the distribution of $x$, and may be literally unbelievable in such a restrictive one-dimensional model while figuratively symbolizing an all-too-real situation of extreme sensitivity to higher-dimensional nonergodic low-utility states of the world that are visited infrequently in history and can be learned about only very gradually. By the time it takes to learn about one region of the higher-dimensional rare-disaster space, the ever-evolving world has wandered off into a different higher-dimensional region with different structural parameter values representing "new," previously unforeseen rare low-utility possibilities that have subsequently evolved. The actual higher-dimensional extreme events, whose occurrence probabilities are controlled by unknown tail-spread parameters, occur exceedingly infrequently. Other things being equal, the rarer is an event the more vague must be the inferred probability of its occurrence. In a truly ergodic world, agents would *eventually* learn everything there is to know about all regions of the higher-dimensional space representing rare tail-possibilities, but it would take a *very* long time—so long that even *very* slow random evolutionary changes would abort any nascent would-be-ergodic knowledge-accumulation process. The net result is thicker tails in an evolutionary world than in the corresponding stationary-distribution world, but without the literal tail extremeness of the double-limiting situation taken here as a prototype where $g(x|\nu, k \to \infty, \mathcal{I} \to 0) \to N(\mu, \nu_t)$. If the kind of evolutionary scenario just described seems counterintuitive or difficult to envision, it is largely because there is no proper analogue in a *REE* world having a constant learnable "true" variance-like parameter controlling the tail frequency of extreme rare events.

## III. The Hidden-Structure Equity Premium

Rewriting (5) in compressed notation that suppresses time subscripts, the price of the riskfree asset (normalized per unit of consumption) is $P^f(\nu, k, \mathcal{I}) = \beta E[\exp(-\gamma X)]$. From (7), the price of one-period equity (normalized per unit of consumption) is $P^{1e}(\nu, k, \mathcal{I}) = \beta E[\exp((1-\gamma)X)]$. In both cases $X$ is a random variable whose probability density function $g(x|\nu, k, \mathcal{I})$ is given by (32) for $x = x_t$. The realized one-period equity premium in ratio form is then

$$\text{(33)} \qquad \frac{R^{1e}(x|\nu, k, \mathcal{I})}{R^f(\nu, k, \mathcal{I})} = \frac{P^f(\nu, k, \mathcal{I})}{P^{1e}(\nu, k, \mathcal{I})} \exp(x),$$

where

$$\text{(34)} \qquad \frac{P^f(\nu, k, \mathcal{I})}{P^{1e}(\nu, k, \mathcal{I})} = \frac{\int_{-\infty}^{\infty} \exp(-\gamma x) g(x|\nu, k, \mathcal{I})\, dx}{\int_{-\infty}^{\infty} \exp((1-\gamma)x) g(x|\nu, k, \mathcal{I})\, dx}.$$

The following proposition contains two related types of results. First, for all $k < \infty$ (and $\nu > 0$) some value of $\mathcal{I}$ matches any given feasible one-period asset-price ratio (34). Second, by choosing carefully the function $\mathcal{I}(k, \nu)$ and then going to the limit $k \to \infty$, essentially any desired one-period equity premium (33) can be replicated in a simulated data-generating process *as if* it came from the super-simple i.i.d.-normal *REE* model of Section II. (In all three theorems that follow, $\nu$ plays the role of representing the current value of the state variable $\nu_t$, while $\nu'$ plays the role of representing future values of $\nu_{t+j}$ for any $j \geq 1$.)

THEOREM 1: *First part: let $\gamma > ½$. Let $\bar{q}$ be any given value of the equity premium needing to be "explained." Then, for every $k < \infty$ and $\nu'$ satisfying $\gamma\nu' < \bar{q}$, there exists a $\mathcal{I}_q(k, \nu') >$ 0 such that*

$$\text{(35)} \qquad \frac{P^f(\nu', k, \mathcal{I}_q(k, \nu'))}{P^{1e}(\nu', k, \mathcal{I}_q(k, \nu'))} = \exp(\bar{q} - \mu - \tfrac{1}{2}\nu').$$

*Second part: suppose $\nu = \nu_t < \bar{q}/\gamma$. Then, for any positive integer $j$, as $k \to \infty$, the random variable $X_{t+j}$ converges to the i.i.d. random variable $\mu + \sqrt{\nu} Z$ with $Z \sim$ i.i.d.$N(0,1)$, where the convergence is uniform for all $\mathcal{I} \geq 0$ and of the same strength as the convergence of a Student-t child distribution to its normal parent when the number of effective observations k approaches infinity. Furthermore, if $\mathcal{I}$ is chosen as $\mathcal{I}_q(k, \nu')$ for $0 < \nu' < \bar{q}/\gamma$ and simultaneously $k \to \infty$, then the limiting realized equity premium $R^{1e}_{t+j}/R^f_{t+j}$ in (33) converges in probability to the i.i.d. lognormal random variable $\exp(\bar{q} - \frac{1}{2}\nu + \sqrt{\nu} Z)$.*

PROOF:

Using (25) and the formula for the expectation of a lognormal random variable, rewrite (34) (after cancelling redundant terms in $\mu$) as

$$\text{(36)} \qquad \frac{P^f(\nu', k, \mathcal{I})}{P^{1e}(\nu', k, \mathcal{I})} = \exp(-\mu) \times \frac{\int_{-\infty}^{\infty} \exp(\gamma^2/2\theta) \psi(\theta|\nu', k, \mathcal{I})\, d\theta}{\int_{-\infty}^{\infty} \exp((1-\gamma)^2/2\theta) \psi(\theta|\nu', k, \mathcal{I})\, d\theta}.$$

As $\mathcal{I} \to 0$, the probability density function $g(x|\nu, k, \mathcal{I})$ defined by (32) approaches the Student-*t* distribution (24), whose moment generating function is unbounded. Consequently, as $\mathcal{I} \to 0$ both integrals in (34) and in (36) approach $+\infty$. Therefore, from (36),

$$\text{(37)} \qquad \lim_{\mathcal{I} \to 0} \frac{P^f(\nu', k, \mathcal{I})}{P^{1e}(\nu', k, \mathcal{I})} = \exp(-\mu) \lim_{\theta \to 0} \frac{\exp(\gamma^2/2\theta)}{\exp((1-\gamma)^2/2\theta)}.$$

Because

$$\text{(38)} \qquad \ln \frac{\exp(\gamma^2/2\theta)}{\exp((1-\gamma)^2/2\theta)} = \frac{\gamma - \frac{1}{2}}{\theta},$$

plugging (38) into (37) for $\gamma > 1/2$ gives

$$\text{(39)} \qquad \lim_{\mathcal{I}\to 0} \frac{P^f(\nu',k,\mathcal{I})}{P^{1e}(\nu',k,\mathcal{I})} = \lim_{\theta\to 0} \frac{\gamma - \frac{1}{2}}{\theta} = +\infty.$$

At the other extreme of $\mathcal{I}$, from (34) it is apparent that as $\mathcal{I} \to \infty$, then $P^f(\nu',k,\mathcal{I})/P^{1e}(\nu',k,\mathcal{I}) \to \exp(\gamma\nu' - \mu - 1/2\nu')$, because the economy is then effectively in the conventional lognormal case (18). The function $P^f(\nu',k,\mathcal{I})/P^{1e}(\nu',k,\mathcal{I})$ defined by (34) is continuous in $\mathcal{I}$. Since

$$\text{(40)} \qquad \frac{P^f(\nu',k,\infty)}{P^{1e}(\nu',k,\infty)} < \exp(\bar{q} - \mu - \tfrac{1}{2}\nu') < \frac{P^f(\nu',k,0)}{P^{1e}(\nu',k,0)}$$

for $\bar{q} > \gamma\nu'$ with $\gamma > 1/2$, condition (35) follows and the first part of the theorem is proved.

Turning to the second part of the theorem, to save space this notation-intensive section of the proof is only sketched here. The fact that as $k \to \infty$ the random variable $X_{t+j}$ converges uniformly (for all $\mathcal{I}$ near zero) to the i.i.d. random variable $\mu + \sqrt{\nu}Z$ with $Z \sim$ i.i.d.$N(0,1)$, in the same mode as a Student-*t* distribution converges to a normal as $k \to \infty$, essentially comes from (24). As $k \to \infty$, from (23) $\nu' \equiv \nu_{t+j} \to \nu = \nu_t$, implying $\nu' < \bar{q}/\gamma$ with probability $\to 1$. If $\mathcal{I}$ is chosen to be $\mathcal{I}_q(k,\nu')$, it then follows (from (33), (35), and $X \to \mu + \sqrt{\nu}Z$) that as $k \to \infty$ the realized one-period (ratio) equity premium converges to the i.i.d.-lognormal random variable $\exp(\bar{q} - 1/2\nu + \sqrt{\nu}Z)$.

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 $\mathcal{I} \to 0$ for any fixed $k < \infty$, $\nu > 0$, the limit of (32) is a Student-*t* distribution of the form (24), the same as would emerge if a regression had been run on $k + 1$ data points. With the presumed prototype case of $k$ huge and $\mathcal{I}$ tiny, the central part of the Student-*t*-like distribution (32) is approximated extremely well by a normal curve with mean $\mu$ and variance $\nu_t$ fitting the data throughout its middle range. However, for applications involving the implications of aversion to uncertain evolving structure (such as calculating the nonergodic equity premium), to ignore what is happening away from the center of the distribution has the potential to wreak havoc on subjective expectation-based asset-price calculations. With these applications in mind, such a normal distribution may be a terrible approximation indeed, because the more-spread-out dampened-*t* distribution (32) is capable of producing an explosion in asset pricing formulas like (34), implying for the limit as $\mathcal{I} \to 0$ an unboundedly large equity premium.

The statistical fact that the moment-generating function of a Student-*t* distribution is infinite has the important economic interpretation that, at least hypothetically, evolving-model-structure uncertainty has the potential in a normal-gamma-beta Bayesian-learning world to be a far more significant determinant of asset prices than the pure risk embodied in a stationary-distributed normal random growth variable with known mean and variance. In the limit, as $\mathcal{I} \to 0$ (for fixed $k < \infty$), 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 fixed structural parameters. The key to understanding the *REE* dilemma concerning how to interpret the "equity-premium puzzle" is that the "premium" is not on pure known-structure risk alone, but rather it is a combined premium on known-structure risk plus (potentially vastly more significant) *recognition of, and adaptation to, unknown evolving structural uncertainty.*

An explosion of the equity premium does not happen in the real world, of course, but a contained potentially 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 backward at their own prior, and what they are seeing there is a spooky reflection of their own 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 when their world has already taken a very bad turn. This eerie sensation of low-$\mathcal{I}$ diffuse background shadow-uncertainty may not be simple to articulate, yet it frightens investors away from taking a more aggressive stance in equities and scares them into a more apprehensive position of wanting to hold instead (on the margin) a portfolio of some much safer stores of value, such as hard-currency cash, inventories of durable real goods, Swiss bank accounts, US or UK short-maturity treasury bills, perhaps precious metals or rare gems, or maybe even stockpiles of storable foods or medicines—as a hedge against unforeseen bad future mutations of history. Consequently, in an evolutionary equilibrium where there is zero net demand for them, these relatively safe assets bear very low, even negative, real rates of return.

Such a Bayesian evolutionary-learning thickened-tail explanation is not easily dismissed. 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 (19) when $\hat{V} \approx 0.04$ percent and $1 \leq \gamma \leq 4$." Unfortunately for this logic, the point-calibrated right-hand side of (19) 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 *REE* formula that makes the future seem far less uncertain than it actually is. Those wishing to downplay this line of reasoning in favor of the *REE* status quo ante might be hard pressed to come up with their own Bayesian rationale for parameterizing tail behavior by calibrating moments of unobservable, subjectively distributed future growth rates with point estimates equal to past sample averages. In essence, the *REE* approach that produces the family of asset-pricing puzzles avoids the consequences (on marginal-utility-weighted asset-pricing stochastic discount factors) of overpowering sensitivity to low values of prior $\mathcal{I}$ only by effectively imposing from the very beginning the fragile pure-recurrent-risk case $k = \infty$ of a normal distribution with *known* structural parameters.

Returning here to the Rietz-Barro *REE* formulation described in the introduction, it generates more-spread-apart growth rates *and simultaneously* imposes a cutoff on the allowed extent of what might be called (in the terminology of this paper) "thickened-tail damage to expected utility" by adding an extra discrete i.i.d. rare-disaster state having a *known* proportional reduction of consumption occur with *known* probability. The Rietz-Barro *REE* modeling approach is greatly appealing because of its simplicity and its tractability in connecting the model directly with data. Such a strategy seemingly circumvents the need to consider the tricky analytical and conceptual issues involved in specifying learning-inference mechanisms or understanding subjective-probability beliefs. But this also comes with the drawback of relying upon a method that does comparative statics on sensitivity to parameter values by assuming that agents always know—immediately and forever—which one of a continuum of possible *REE* steady-state distributions they are currently occupying, and also that transition dynamics are inessential for such comparative-steady-state calibration exercises. However, a serious attempt to specify how agents actually traverse an infinity of possible *REE*s, or even how they know which *REE* they are now in, will involve learning and is likely to have strong tail-thickened transition-dynamics consequences similar to what emerges from the model of this paper—the full implications of which cannot be remotely appreciated by examining only the behavior of *REE* steady-state distributions.

With the Rietz-Barro *REE* formulation, it appears that speculative but not unreasonable parameter values may explain various puzzles. Although these calibration exercises seemingly relate the structural parameters of the model to a century or so of real-world data, such numerical specifications nevertheless rely ultimately upon subjective prior beliefs to a degree perhaps not apparent at first glance. For example, Barro postulates a rare catastrophic contraction of consumption between 15 percent and 64 percent with i.i.d. probability around 1.5 percent per year, which might just as well be interpreted to imply an extraordinarily fuzzy equity premium—via highly nonlinear hypersensitivity to the contraction percentage that is subjectively being assumed—even when this shrinkage factor is arbitrarily limited to be in the range [$-64$ percent, $-15$ percent]. A subjectively imposed contraction possibility of 15 percent has little effect on the equity-premium puzzle but a subjectively imposed contraction possibility of 64 percent implies an equity premium *antipuzzle*. Given that the allowable range itself is uncertain and, as then seems appropriate, a nondogmatic (i.e., positive) subjective probability density needs to be placed on its supports, we are essentially back in the continuously thick-tailed unknown-structure almost-full-support world of Bayesian learning and inference that this paper is attempting to model—along with its inevitable sensitivity to fear-factor tail effects controlled by subjective beliefs about the a priori possibility of bad events.

At the end of the day, the Rietz-Barro *REE* formulation and the $\mathcal{I}$-based subjective-prior evolutionary-learning formulation of this paper both make roughly the same qualitative predictions (from the commonly shared mechanism of thickened tails), while neither model can explain quantitatively such asset-return puzzles as the observed equity premium without relying on marginal utility to become very high at increasingly disastrous consumption levels. Both modeling approaches depend hypersensitively upon ad hoc truncation mechanisms to cut off the thickened-tail damage in low-utility states, which seems unavoidable for this kind of analysis of subjective expectations concerning rare disasters but is very hard to test directly. The Rietz-Barro *REE*-based approach "works" empirically in the sense that disaster-parameter values point-calibrated to plausible data seem to "explain" asset-return puzzles. This is perhaps comforting as a suggestive example which indicates that a thickened-tail explanation is at least capable of passing a first-round plausibility test when one ignores the inherent diffuseness of the uncertainty about rare events. Such an atheoretical point-calibration approach, however, "does not work" at all in a sensitivity-analysis sense because its power to explain puzzle patterns is overwhelmingly fragile to hidden subjective judgements about the underlying distribution of the rare disasters being taken implicitly as massed at a single known point. This paper is saying that the real issue is not so much to find comforting suggestive-example *REE* point calibrations of rare disasters that *resolve* the asset-return puzzles as it is to grasp the more essential idea that the asset-return puzzles *dissolve* when the Bayesian thickened tail issue is confronted directly and rare-disaster Dirac-point-mass distributions are replaced by nondogmatic densities having everywhere-positive probabilities. In both approaches, subjective prior beliefs controlling the allowable variability of growth rates are prominent features of the landscape—it is just more difficult to see them (and the critical role they are playing) in the *REE* map. Contrary to the prevailing impression from consumption-based *REE*-inspired calculations, the actual big picture is that a realized-frequency application of expected utility theory by itself *does not allow us to determine asset prices or returns* to the (tail-specification independent, subjective-judgement free) degree that we might have preferred or that calibration exercises might seem to be suggesting. Explicitly modeling the underlying learning process—which in any case is required to justify *REE* in the first place—just makes this central issue (of thickened-tail posterior expectations depending hypersensitively upon seemingly innocuous subjective prior information) so prominent that it simply cannot be avoided.

This "Bayesian peso problem" means that for asset pricing applications it is not at all unscientific to adhere to the non-*REE* idea that no amount of past data can be nearly large enough to identify the relevant structural uncertainty concerning future economic growth. Moreover, as a corollary, *REE* calibrations ignoring this basic principle of learning about hidden evolving parameters may very easily end up badly underestimating the comparative utility-risk of a real-world gamble on the unknown structural potential for future economic growth, relative to a nearly-safe investment in a nearly-sure thing. Extreme sensitivity to subjective judgments, even with unlimited data, could make some aspects of the conventional *REE*-inspired research program for explaining asset returns seem like overreaching. What might be appropriate is a scaling back to a less sharp research strategy than *REE*, which begins by recognizing that in a nonergodic world asset prices legitimately depend upon subjective fuzziness about the future. In this subjectivity-dominated world, equity returns cannot be fully understood as a function of past data scaled by a randomness term, whose volatility essentially reflects the recurrent volatility of some objective-frequency-based, statistically identifiable distress factor in the real economy.

Translated into classical-frequentist statistical language, the second part of Theorem 1 has the following rigorous interpretation. For given $\nu = \nu_t$, pick any equity premium $\bar{q} > \gamma\nu$, name any sample size $n$, and choose any desired level of statistical confidence relative to the supposedly "true" data-generating process. Then there exists some sufficiently large $k$ and accompanying function $\mathcal{I} = \mathcal{I}_q(k, \nu')$ (where for $\nu'$ will be substituted future realizations of $\nu_{t+j}$, with $1 \leq j \leq n$) such that the empirically observed frequency distribution of the $n$ realized values of the one-period equity premium simulation-generated by *this* hidden-structure model is guaranteed to differ only insignificantly (in terms of the desired level of statistical confidence) from the sampling distribution that *would* be simulation-generated in a sample of size $n$ *if* the "true" equity premium $r^{1e} - r^f$ were i.i.d. $N(\bar{q} - \nu/2, \nu)$. (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 $\nu = \hat{V}[r^e]$—more on this mismatched equity volatility later.)

What is being presented here is but one illustrative normal-gamma-beta example of the economic consequences of an evolving-structure tail-thickening effect, but it seems to be very difficult to get around the moral of this story. For any finite value of $k$, however large, Bayesian distribution-spreading will cause the observed equity premium to be exceedingly sensitive to tiny, seemingly negligible changes in the subjectively assumed information content of the prior distribution when, according to the key assumption behind *REE*, such innocuous changes in the informativeness of prior beliefs should have been rendered irrelevant by the infinite data-evidence. The dominant statistical-economic force behind the puzzles in this paper's way of looking at things is that seemingly thin-tailed probability distributions (like the normal), which are actually only thin-tailed *conditional on known structural parameters* of the model, become tail-thickened (like the Student-*t*) after integrating out the parameter uncertainty. Intuitively, no finite sample of *effective* size $k < \infty$ can accurately assess tail thickness, and therefore the attitudes of a risk-averse Bayesian agent toward investing in various risk-classes of assets may, at least in principle, be driven to an arbitrarily large extent by this unavoidable feature of Bayesian expectational uncertainty.

The important generic result in Michael Schwarz (1999) can be interpreted as saying that with a prior that is scale-invariant to measurement units, the moment-generating function of the posterior distribution is infinite (i.e., the posterior distribution has a "thickened" tail) for essentially any reasonably specified probability density function. This occurs even when the random variable is being sampled an arbitrarily large number of times from a thin-tailed parent distribution whose moment-generating function is finite. Such a result means that there is a very broad sense in which, at least hypothetically-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-recurrent-risk part of stochastic growth rates, than are they averse to the pure recurrent risk itself. When investors are modeled as perceiving and acting upon these inevitably spread-apart subjective posterior-predictive distributions, then a fully rational equilibrium interpretation can weave a parsimonious unifying Bayesian strand through the entire family of asset-return puzzles, as the next three sections of the paper (when combined with this section) will indicate.

## IV. The Hidden-Structure Riskfree Interest Rate

We can use the same mathematical-statistical apparatus to calculate the hidden-evolving-structure riskfree interest rate. (Actually, the last section of the paper and this section might well have been reversed sequentially because the riskfree rate is much easier to calculate and understand than the equity premium.) For all other parameter values fixed, let $f(\nu, k, \mathcal{I})$ be the value of $r^f$ that comes out of formula (15) when the probability density function of $X$ is $g(x|\nu,k,\mathcal{I})$ defined by equation (32) for $x = x_t$. Plugging the subjective posterior-predictive distribution (32) into the right-hand side of equation (15), the result is

$$(41)\quad f(\nu,k,\mathcal{I}) \equiv \rho - \ln\int_{-\infty}^{\infty}\exp(-\gamma x)g(x|\nu,k,\mathcal{I})\,dx.$$

THEOREM 2: *Let $r^f(\nu')$ be any given continuous function of $\nu'$ satisfying $r^f(\nu') < \rho + \gamma\mu - \tfrac{1}{2}\gamma^2\nu'$ for all $\nu' > 0$. Then, for every $k < \infty$, $\nu' > 0$, there exists a $\mathcal{I}_f(k,\nu') > 0$ such that*

$$(42)\qquad r^f(\nu') = f(\nu',k,\mathcal{I}_f(k,\nu')).$$

*Additionally, the limiting realized riskfree rate can be made to converge to the same constant value $\bar{r}^f < \rho + \gamma\mu - \tfrac{1}{2}\gamma^2\nu_t$ if, in every future state $\nu' = \nu_{t+j}$, the value of $\mathcal{I}$ is chosen to be $\mathcal{I}_f(k,\nu')$ for $r^f(\nu') = \bar{r}^f$ while simultaneously the limit $k \to \infty$ is taken.*

PROOF:

As $\mathcal{I} \to 0$, the probability density function $g(x|\nu,k,\mathcal{I})$ defined by (32) approaches the Student-$t$ distribution (24), whose moment generating function is unbounded. From (41), therefore, $f(\nu',k,0) = -\infty$. At the opposite extreme, as $\mathcal{I} \to \infty$, then $f(\nu',k,\infty) = \rho + \gamma\mu - \tfrac{1}{2}\gamma^2\nu'$, because the economy is then effectively in the conventional lognormal case (18). Thus,

$$(43)\qquad f(\nu',k,0) < r^f < f(\nu',k,\infty),$$

and, since $f(\nu',k,\mathcal{I})$ defined by (41) is continuous in $\mathcal{I}$, the conclusion (42) follows. The convergence to a constant value for all future periods follows from the fact that $\nu' = \nu_{t+j}$ effectively becomes constant because $\nu_{t+j} \to \nu_t$ as $k \to \infty$, so that the condition $r^f(\nu') = \bar{r}^f < \rho + \gamma\mu - \tfrac{1}{2}\gamma^2\nu'$ holds on the future trajectory with probability $\to 1$ as $k \to \infty$.

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 like a Student-$t$ distribution in its tails and carries with it a potentially explosive moment generating function reflecting an intense aversion to unforeseen low-precision evolutionary-mutational future histories. The bottom line once more is that a "Bayesian peso problem" can cause incorrect *REE*-based inferences about expected future utility, which are essentially mimicking the observed historical frequency of past growth rates, to underestimate enormously just how relatively much more attractive are relatively safer 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 actually observed and the data-generating process parallels the equity premium version. Pick $r^f = \bar{r}^f < \rho + \gamma\mu - \tfrac{1}{2}\gamma^2\nu_t$, name some number $n$, and choose any desired level of statistical strength, here representing measurement accuracy. Then there exists some (large) $k$ and accompanying $\mathcal{I} = \mathcal{I}_f(k,\nu')$ such that the frequency distribution of the $n$ riskfree-rate realizations simulation-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 $n$ *if* the "true" riskfree rate were the constant value $\bar{r}^f$.

## V. Equity Volatility in an *As-if-REE* Parable

It has already been amply demonstrated that the dynamic evolution of future asset prices and returns is wickedly hyperreactive to nondata subjective prior beliefs, even with infinite past data. Such a complicated family of nonergodic prior-hypersensitive trajectories simply cannot be distilled down into the neat form of a rigorous story about *REE* steady-state distributions. Yet it is only human nature to yearn deeply to be able to capture the essential spirit of a bewildering real-world actuality by reformulating it in the more reassuring language of some familiar—but necessarily oversimplified—paradigm. This section of the paper is different from the others in its methodology and in what it is trying to do. Take as given the almost-inevitable fact that, no matter how much we are warned against doing it, we still can't help but visualize the evolution of an intricate nonergodic stochastic economic process in terms of an oversimplified fable about a *REE* steady-state distribution. The question being addressed here can easily be posed but is very tricky to answer: if everyone is going to think in terms of *REE* anyway, what then is the least distortionary *REE*-type parable for conceptualizing the complicated reality behind the "equity-volatility puzzle"?

For the dual endowment-production i.i.d.-normal *REE* in Section II, equity returns should vibrate consistently with growth rates as prescribed by equation (21). According to (21), for an economy-wide comprehensive wealth index embodying an implicit claim on the future aggregate consumption of the underlying real economy, all higher-order central moments of $r^e$ and $x$ should match subjectively and objectively. Alas, the empirical second moments of $r^e$ and $x$ are not even remotely matched in the time-series data because $\hat{V}[r^e]/\hat{V}[x] \approx 75$. With the evolutionary version of the model, however, *future X is subjectively perceived "as if" it is much more variable than it seems to be from past time series data* in the sense that for $\gamma > 1$ the "true" value of $EU = E[\exp((1-\gamma)X)/(1-\gamma)]$ is "felt" to be much lower than what would appear to be indicated by simply identifying the variance of future as-if-normal $X$ with its past sample average $\hat{V}[x]$, which, when mechanically plugged into the familiar formula for the expectation of a lognormally distributed random variable, would give the welfare value $\exp((1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2\hat{V}[x])/(1-\gamma) \gg EU$. The standard welfare calibration "doesn't work" here because the agent "feels" much worse than if $X \sim N(\mu, \hat{V}[x])$. The issue being addressed now is whether another i.i.d.-normal specification having the *same mean but greater variance*, whose i.i.d.-normal variance is calibrated to be welfare-equivalent to the actual Student-*t*-like distribution of $X$ defined by (32), can be made to "work better"—if not perfectly—in a quick-and-dirty heuristic *as-if-REE* story.

The price-earnings ratio $P^e/C$ of equity implicit in (9) depends on expectations over an infinite future horizon, and is extraordinarily hyperreactive to low values of $\mathcal{I}$ (presumably even more so than the one-period riskfree rate, perhaps identifiable with a "storage technology"). Such extreme hypersensitivity to the information content of subjective prior beliefs suggests very strongly (at least to me) that a fuller, more-complicated model might be built around transcription errors that cause tiny contaminations of $\mathcal{I}$ to become amplified into arbitrarily large animal-spirit-like hyperreactive equity-price fluctuations, thereby introducing into the model extra volatility that could be made to match the stylized-fact data pattern. However that may be, as a practical matter (whatever is the causal mechanism actually producing the large swings in stock-market prices), to proceed further here analytically requires some simplifying assumption about the *reduced form* of equity returns. 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 workhorse reduced-form assumption that *equity returns are independently normally distributed with known mean and variance*. The rest of the general-equilibrium system in the *as-if-REE* parable will now be made to revolve around this centerpiece assumption of known-fixed-structure i.i.d.-normal equity returns.

From the basic duality equivalence between production and endowment versions of the same core *REE* model, and from the critical "discipline imposed by general equilibrium modeling," if the primitive real-productivity-return $r^e$ $(= \ln A)$ in the *AK*-production version is *known* to be i.i.d.-normal, then so too must the derived real growth rate of consumption $X$ $(= \ln A + \{E[X] - E[\ln A]\})$ be i.i.d.-normal, and with the identical variance. In this case, equation (21) holds with the arrow of causal reasoning going from the presumed-known value of $\sigma[r^e]$ to the implied value of $\sigma[X]$ $(= \sigma[r^e])$. Equation (13) seems to be suggesting that volatile wealth is "welfare equivalent" to volatile consumption. This section of the paper is trying to answer the question: between the two observed variability alternatives ($\hat{\sigma}[r_e] \approx 17$ percent standing in for the left-hand side of equation (21) and representing the past variability of comprehensive-wealth returns, or $\hat{\sigma}[x] \approx 2$ percent standing in for the right-hand side and representing the past variability of consumption growth), *which empirical variability* (wealth $\hat{\sigma}[r_e]$ or consumption $\hat{\sigma}[x]$) *better matches the agent's true welfare situation*?

Waving aside the "rationality" of such beliefs, suppose for the sake of the thought-experimental quick-and-dirty heuristics here that (for some given $E[X^N]$ and $\sigma[X^N]$) the random variable

$$X^N(X|\nu,k,\mathcal{I}) \sim N(E[X^N], \sigma^2[X^N]) \tag{44}$$

represents a functional transformation of the random variable $X$ into the normally distributed random variable $X^N(X|\nu,k,\mathcal{I})$ embodying an agent's subjective probability belief that future growth rates are i.i.d.-normal with known parameters $E[X^N]$ and $\sigma[X^N]$. (It can be shown that such a transformation exists for some implicitly defined Jacobian-inverse monotonic function.) 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 consumption corresponding to (44). 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

$$r^N(X^N(X)) - E[r^N] = X^N(X) - E[X^N], \tag{45}$$

which is the normal counterpart here of (21). The operational question now is: *how do we observe that (45) is untrue?* It turns out that i.i.d. as-if-normal growth rates can yield the *same* expected return on equity as the formulation in previous sections of the paper, so that $E[r^N] = E[r^{1e}] = E[r^e]$, which, provided also that $\sigma[r^N] = \hat{\sigma}[r^e]$, signifies here that *observed equity data alone* cannot refute the hypothesis $x \sim$ i.i.d. $N(\mu, \hat{V}[r^e])$, *given* the standard assumption that equity returns are *known* by the agents to be i.i.d. normal in the first place.

The following "calibration 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 $\sigma[X^N] \approx \hat{\sigma}[r^e]$. 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-*REE* parable wherein the representative agent has a subjective normally distributed *welfare equivalent* belief, which is consistent with (45) and the equity-return data, "as if" the future growth rate is $X^N$ with known variance equal to the observed variance of returns on wealth. 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 variance of returns on equity-wealth, rather than by the relatively low variance of realized past growth rates. None of this really "explains" why $\hat{V}[r^e]/\hat{V}[x] \approx 75$ in the first place. But because here $\sigma[r^N] = \sigma[X^N]$ by construction, at least with this artificially synthesized *REE*-like as-if-i.i.d.-normal growth parable, there is no longer a jarring mismatch of variabilities wanting to be explained between equity-wealth returns and underlying *welfare equivalent* growth fundamentals.

THEOREM 3: *Let $\sigma(\nu') > 0$ be any given continuous function of $\nu'$ satisfying $\sigma(\nu') > \sqrt{\nu'}$ for all $\nu' > 0$. Let $r^N(X^N(X))$ and $X^N(X)$ be related by (45) where the distribution of $x = x_t$ at any time $t$ is given by (32). Then for every $k < \infty$, $\nu' > 0$, there exists a $\mathcal{I}_s(k,\nu') > 0$ such that the following four calibration conditions are simultaneously matched:*

$$E[r^N(X^N(X))] = E[r^{1e}(X)], \tag{46}$$

$$\sigma[r^N(X^N(X))] = \sigma[X^N(X)] = \sigma(\nu'), \tag{47}$$

$$E[X^N(X)] = E[X] = \mu, \tag{48}$$

$$\forall C > 0: E[U(C\exp(X^N(X)))] = E[U(C\exp(X))]. \tag{49}$$

*Additionally, the limiting realized value of $\sigma(\nu') = \sigma(\nu_{t+j})$ can be made to converge to the same constant value $\bar{\sigma} > \sqrt{\nu_t}$ if, in every future state $\nu' = \nu_{t+j}$, the value of $\mathcal{I}$ is chosen to be $\mathcal{I}_s(k,\nu')$ for $\sigma(\nu') = \bar{\sigma}$ and simultaneously the limit $k \rightarrow \infty$ is taken.*

PROOF:

Define $s(\mathcal{I})$ to be the implicit solution of the equation

$$(50)\quad \frac{1}{\sqrt{2\pi}\,s(\mathcal{I})} \times \int_{-\infty}^{\infty} \exp\left((1-\gamma)x^N - \frac{(x^N-\mu)^2}{2s(\mathcal{I})^2}\right) dx^N = \int_{-\infty}^{\infty} \exp((1-\gamma)x)\,g(x|\nu',k,\mathcal{I})\,dx,$$

and note for this definition that (49) and (48) are satisfied by construction.

It can readily be shown that

$$(51)\quad r^{1e}(x) = x + \{\rho - \ln E[\exp((1-\gamma)X)]\},$$

and, analogously,

$$(52)\quad r^N(x^N) = x^N + \{\rho - \ln E[\exp((1-\gamma)X^N)]\},$$

so that (46) then follows from (48), (50), (51), (52).

As $\mathcal{I} \to \infty$, the probability density function $g(x|\nu',k,\mathcal{I})$ goes to a normal distribution with variance $\nu'$, and consequently the integral on the right-hand side of equation (50) approaches (by the expected-lognormal formula) $\exp((1-\gamma)\mu + ½(1-\gamma)^2\nu')$, implying $s(\infty) = \sqrt{\nu'}$. As $\mathcal{I} \to 0$, the probability density function $g(x|\nu',k,\mathcal{I})$ defined by (32) approaches the Student-*t* distribution (24), whose moment generating function is unbounded, implying the right-hand side of (50) is also unbounded, meaning $s(0) = \infty$. Thus,

$$(53)\quad s(\infty) < \sigma(\nu') < s(0),$$

and, by continuity of the function $s(\mathcal{I})$, there must exist a $\mathcal{I}_s(k,\nu') > 0$, satisfying

$$(54)\quad s(\mathcal{I}_s) = \sigma(\nu'),$$

which, when combined with (45), proves (47). The convergence to a constant value of $\sigma(\nu') = \sigma(\nu_{t+j}) = \bar{\sigma}$ for all future periods $t+j$ follows from the fact that $\nu' = \nu_{t+j}$ effectively becomes constant over time as $k \to \infty$, so that the condition $\sigma(\nu') = \bar{\sigma} > \sqrt{\nu'}$ holds on the future trajectory with probability $\to 1$ as $k \to \infty$.

The force behind Theorem 3 is the same "strong force" driving the previous two theorems: intense aversion to the structural parameter uncertainty embodied in tail-thickened *t*-distributed subjective future growth rates. Compared with the Student-*t* distribution $X \sim g(x|\nu,k,\mathcal{I} \to 0)$, a representative agent with $\gamma > 1$ 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 $g(x|\nu,k,\mathcal{I} \to 0)$ with a completely uninformative prior is contained by the substitution of $g(x|\nu,k,\mathcal{I} = \mathcal{I}_s)$ with a somewhat informative prior $\mathcal{I}_s(k,\nu) > 0$.

Theorem 3 is effectively saying that if you *must* pressure-contain the wickedly complicated dynamic behavior of prior-sensitive asset prices under nonergodic evolutionary uncertainty to fit within the analytically tractable mold of a prior-free stationary-frequency as-if-i.i.d.-normal fable, then the *REE*-like calibration $\sigma[X^N] = \hat{\sigma}[r^e]$ tells the better welfare parable than the *REE*-like calibration $\sigma[X^N] = \hat{\sigma}[x]$. To an outsider classical-frequentist econometrician thinking in terms of an i.i.d.-normal *REE* specification, however, agent-investors inside this *as-if-REE* economy will appear from way-too-low actual equity prices for $\gamma \approx 2$ to be irrationally incapable of internalizing what the data are clearly saying about $\hat{\sigma}[x] \approx 2$ percent. Instead, with $k \to \infty$, the low actual equity prices portray these agents as if clinging stubbornly to an irrational mental image of a stochastic discount factor that would be consistent only with their future welfare depending upon the realization of some hypothetical much-more-variable normally distributed growth rate having known counterfactual standard deviation $\sigma[X^N] = \hat{\sigma}[r^e] \approx 17$ percent. But when $k \to \infty$ and for, say, one hundred independent observations, the frequentist hypothesis that the observed sample value of $\hat{\sigma}[x^N] = 2$ percent could have been generated by (agents having in their heads) a "true" (welfare equivalent) value of $\sigma[X^N] = 17$ percent is classically rejected by a chi-square test at the 99.99 percent confidence level!

## VI. An Empirical "Test" of the *As-if-REE* Parable

Viewing the three theorems of the paper through the lens of the welfare-equivalent as-if-i.i.d.-normal-growth fable of Theorem 3 delivers the package of a neat closed-form relationship (accompanying the well-known formula for the expectation of a lognormal random variable) among the equity premium $q$, the riskfree rate $f$, and the variability-in-common $s$ of the economic-financial system (where $s$ in the as-if story represents the mutually shared standard deviation of equity returns and *welfare equivalent* growth rates). With this notation, (19) and (20) become $q = \gamma s^2$ and $f = \rho + \gamma\mu - \frac{1}{2}\gamma^2 s^2$. The three theorems themselves are only partial equilibrium statements in the sense that each one matches just one side of the whole asset-returns-puzzle triangle. The $\mathcal{I}(k,\nu)$ function that works for any one theorem will not work for the other two—essentially because a system with just one degree of freedom (the parameter $\mathcal{I}$ in (30)) cannot match three observables simultaneously (which would require at least a three-parameter prior specification). Suppose, however (what at this stage is merely an unproved, but not implausible, conjecture), that a more general higher-dimensional parameterization can be made to deliver a situation "as if" the same vector-valued $\mathcal{I}(k,\nu)$ function works for all three theorems. The following experimental question then arises naturally: does the simple i.i.d.-lognormal relationship among $q$, $f$, and $s$ of the closed form (19), (20) hold *empirically*, conditional upon the same $\mathcal{I}(k,\nu)$ function working for all three theorems? The answer from the experiment is "yes," which conveys at least *some* intuitive feel for the degree to which this heuristic way of looking at things represents a relatively coherent theoretical-empirical mental 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 $E[X^N] = \hat{x}$ and standard deviation $s = \sigma[X^N] = \hat{\sigma}[r^e]$ renders, along with (45), an internally consistent *as-if-REE* story connecting the actual stylized facts of our economic world. In Table 1, parameter settings have been selected that should represent numbers well within the "comfort zone" for most economists. The data in Table 1 are intended to be an overall approximation of stylized facts that have been observed for many countries over long time periods. All rates are per-annum and real.

With any given $k < \infty$, $\nu > 0$, the model explains endogenously three values, written here for simplicity (by suppressing dependence on $k$ and $\nu$) as $q(\mathcal{I})$, $f(\mathcal{I})$, $s(\mathcal{I})$—all three being functions of the one free parameter $\mathcal{I}$. Purely for conceptual-notational convenience, pretend that $\hbar = 6.625 \times 10^{-34}$ represents some tiny-tiny quantum threshold of observability, below which prior information $\mathcal{I}$ in the unboundedly distant past is considered to be "effectively zero" and the relationship among $\mathcal{I}_q$, $\mathcal{I}_f$, $\mathcal{I}_s$ becomes so blurred that the situation can be treated as if the *same* $\mathcal{I}(k,\nu)$ function works for all three theorems (because the priors induced by $\mathcal{I}_q$, $\mathcal{I}_f$, $\mathcal{I}_s$ and $\mathcal{I}(k,\nu)$ are *each* operationally indistinguishable from a zero-information completely uninformative diffuse prior). Under such circumstances of extremely low prior informativeness (combined with extremely slow evolution and an extremely large dataset), we will not be able to observe or calculate the underlying primitive values of $\mathcal{I}_q$, $\mathcal{I}_f$, $\mathcal{I}_s$ *directly* (although we know in theory that there exists some astronomically large number $k$, for which simultaneously $0 < \mathcal{I}_q < \hbar$, $0 < \mathcal{I}_f < \hbar$, $0 < \mathcal{I}_s < \hbar$, because $k \to \infty$ implies that $\mathcal{I}_q \to 0$, $\mathcal{I}_f \to 0$, $\mathcal{I}_s \to 0$). However, and more usefully here, an *indirect* calibration experiment can be performed by setting *any one* of the stylized-fact constants $q|0 < \mathcal{I}_q < \hbar$, $f|0 < \mathcal{I}_f < \hbar$, $s|0 < \mathcal{I}_s < \hbar$ equal to its observed value in Table 1 and then backing out the implied values of the *other two* remaining stylized-fact constants by inverting the two analytically tractable as-if-i.i.d.-lognormal-consumption equations of the closed form (19) and (20). In this way of looking at things, $k$ is considered to be so large that the joint condition $0 < \mathcal{I}_q < \hbar$, $0 < \mathcal{I}_f < \hbar$, $0 < \mathcal{I}_s < \hbar$ implies simultaneously that $\mathcal{I}_q$, $\mathcal{I}_f$, $\mathcal{I}_s$ are each effectively zero, and therefore operationally interchangeable—meaning that in some (admittedly restricted) sense this calibration exercise is "testing" the hypothesis that a sufficiently uninformative prior can here "explain" the stylized facts.

Defining $\mathcal{I}_s | 0 < \mathcal{I}_s < \hbar$ to be an implicit solution of

$$\mathcal{I}_s = s^{-1}(\sigma[r^e]) = s^{-1}(17 \text{ percent}),$$

we then have, from (19) with $V[X] \equiv s^2(\mathcal{I}_s)$,

$$\ln E[R^e] - r^f = \gamma s^2(\mathcal{I}_s) = q|0 < \mathcal{I}_s < \hbar$$

$$= 5.8 \text{ percent},$$

TABLE 1—SOME MACROECONOMIC "STYLIZED FACTS"

| Assumed numbers | Value |
|---|---|
| Mean arithmetic return on equity | $\ln E[R^e] \approx 7$ percent |
| Geometric standard deviation of return on equity | $\sigma[r^e] \approx 17$ percent |
| Riskfree interest rate | $r^f \approx 1$ percent |
| Implied equity premium | $\ln E[R^e] - r^f \approx 6$ percent |
| Mean growth rate of per capita consumption | $E[x] \approx 2$ percent |
| Standard deviation of growth rate of per-capita consumption | $\sigma[x] \approx 2$ percent |
| Rate of pure time preference | $\rho \approx 2$ percent |
| Coefficient of relative risk aversion | $\gamma \approx 2$ |

to be compared with $q|0 < \mathcal{I}_q < \hbar\ = 6$ percent. From (20) with $V[X] \equiv s^2(\mathcal{I}_s)$,

$$r^f = \rho + \gamma E[X] - \tfrac{1}{2}\gamma^2 s^2(\mathcal{I}_s) = f|0 < \mathcal{I}_s < \hbar$$

$$= 0.2 \text{ percent},$$

to be compared with $f|0 < \mathcal{I}_f < \hbar\ = 1$ percent.

Defining $\mathcal{I}_q|0 < \mathcal{I}_q < \hbar$ to be the implicit solution of

$$\mathcal{I}_q = q^{-1}(\ln E[R^e] - r_f) = q^{-1}(6 \text{ percent}),$$

we then have, from (20) and (19),

$$r^f = \rho + \gamma E[X] - \gamma q(\mathcal{I}_q)/2 = f|0 < \mathcal{I}_q < \hbar$$

$$= 0 \text{ percent},$$

to be compared with $f|0 < \mathcal{I}_f < \hbar\ = 1$ percent. From (19) with $V[X] \equiv \sigma^2[r^e]$,

$$\sigma[r^e] = \sqrt{q(\mathcal{I}_q)/\gamma} = s|0 < \mathcal{I}_q < \hbar$$

$$= 17 \text{ percent},$$

to be compared with $s|0 < \mathcal{I}_s < \hbar\ = 17$ percent.

Defining $\mathcal{I}_f|0 < \mathcal{I}_f < \hbar$ to be an implicit solution of

$$\mathcal{I}_f = f^{-1}(r_f) = f^{-1}(1 \text{ percent}),$$

we then have, from (20) and (19),

$$\ln E[R^e] - r^f = 2[\rho + \gamma E[X] - f(\mathcal{I}_f)]/\gamma$$

$$= q|0 < \mathcal{I}_f < \hbar\ = 5 \text{ percent},$$

to be compared with $q|0 < \mathcal{I}_q < \hbar\ = 6$ percent.

From (20) with $V[X] \equiv \sigma^2[r^e]$,

$$\sigma[r^e] = \sqrt{2[\rho + \gamma E[X] - f(\mathcal{I}_f)]/\gamma}$$

$$= s|0 < \mathcal{I}_f < \hbar = 16 \text{ percent},$$

to be compared with $s|0 < \mathcal{I}_s < \hbar\ = 17$ percent.

As a kind of very rough test for the internal consistency and raw fit of the as-if-i.i.d.-normal-growth story (hypothetically conditional on a higher-dimensional version of the same $\mathcal{I}(k, \nu)$ function simultaneously working for all three theorems), the empirical outcomes of these Bayesian calibration experiments fit nearly exactly. Thus, at the very minimum, there is *some* story here about why everything coheres almost perfectly in the bare-bones canonical i.i.d.-normal *REE* model when, by just the simplest substitution, a welfare-equivalent growth variability $s = \sigma[X^N] = \hat{\sigma}[r^e]$ equal to the observed standard deviation of equity-wealth returns replaces the observed real-growth variability $\hat{\sigma}[x]$. Otherwise, such a near-perfect fit must be interpreted as just happening to be a miraculous coincidence in the data.

Continuing on 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 $s$ of consumption. With i.i.d. lognormality of $\{C_{t+1}/C_t\}$, the imaginary deterministic path having the same mean consumption as the stochastic trajectory (14) is

$$\overline{C}_{t+1} = \exp\left(\mu - \frac{1}{2}s^2\right)\overline{C}_t. \tag{55}$$

Using formula (55), it can readily be shown (following Lucas (2003)) that the welfare gain from a mean-preserving shrinkage that compresses the stochastic trajectory $C_{t+1} = C_t \exp(X_t)$ into the deterministic path (55) is equivalent to a change in *each* period's consumption of

$$\Delta C_t = \left(\exp\left(\frac{1}{2}\gamma s^2\right) - 1\right) C_t. \tag{56}$$

When $\gamma \approx 2$ and the historical value of $s = \hat{\sigma}[x] \approx 2$ percent is used in (56), then $\Delta C_t / C_t \approx 0.04$ percent, 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 almost nothing. Such a number, however, captures only the "weak force" of known-fixed-structure 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 $s = \sigma[X^N] = \hat{\sigma}[r^e] \approx 17$ percent in formula (56), for which case $\Delta C_t / C_t \approx 3$ percent. Accounted in this welfare-equivalent metric of shrunken deterministic consumption, therefore, structural uncertainty concerning the evolving future growth process turns out empirically to be more significant by two orders of magnitude than known fixed-structure pure growth-rate risk.

## VII. Conclusion

The hidden-structure evolutionary 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 an equity-volatility 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 *REE* lens. From a nonergodic Bayesian learning perspective, the puzzling numbers being observed in the data are telling a rational story about the implicit revealed-prior subjective distribution of background structural-parameter uncertainty accompanying the evolutionary growth process actually generating such data. While the story is "rational," it is not about any form of "rational expectations" that is like *REE*—i.e., the known stationary-distribution form of "rational expectations" that in routine practice is standardly applied to macrofinance and other areas of economics.

In principle, consumption-based representative agent models provide a complete answer to all macroeconomic asset pricing questions and give a unified theory integrating 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 nonergodic structural uncertainty there is some theoretical justification for treating the subjective variability of the future growth rate *as if* it were equivalent in welfare to the observed variability of a comprehensive economy-wide index of equity-wealth returns. For this *as-if-REE* high-growth-variability interpretation, the simple standard model of asset pricing may have the potential to be a decent shortcut conceptualization of what is actually happening in a complicated ever-changing world where unforeseen bad events—scary, disruptive, and without precedent—may evolve at any future time.

## Mathematical Appendix

PROOF OF THEOREM 0:

In the interest of space, this proof of Theorem 0 is extremely compressed. Further details are available upon request from the author, but a reader with some background in Bayesian statistical decision theory should be able to follow at least the main strands of the argument. (To be realistic, the details of a rigorous proof of Theorem 0 may not be accessible to a reader who has not been exposed to Bayesian theory at all, and therefore such a person perhaps must accept more or less on faith that the result can be proved.) For an interested reader's convenience, I tie the notation as closely as possible to the treatment of a similar (but by no means identical) classical-frequentist state-space local-level formulation of a stochastic volatility process in Shephard (1994). I use Shephard's notation except where it conflicts with the notation of this paper, which then takes precedence.

Theorem 0 is proved as a corollary to the following two lemmas.

Consider the stochastic process

$$X_t|\theta_t \sim N(\mu, 1/\theta_t), \tag{57}$$

$$\theta_{t+1} = \frac{\theta_t \eta_{t+1}}{\omega}, \tag{58}$$

with $\eta_{t+1} \sim$ *i.i.d.* Beta $(\omega a_t, (1-\omega)a_t)$, where $\omega : 0 < \omega < 1$ is a constant controlling the speed at which the precision moves. Define $k = 1/(1-\omega)$. Then with $\zeta = \omega\eta$, equations (26) and (58) are identical, and when $a_t = k/2$, the corresponding density of $\zeta$ is (28).

LEMMA 4: *Suppose the unconditional distribution of $\theta_t$ is Gamma$(a_t, b_t)$ and suppose the joint distribution of $X_{t+1}, \theta_{t+1}$ conditional on $\theta_t$ is determined by (57) (led by one period) and (58), with $\eta_{t+1}|\{\theta_t, x_{t+1}\} \sim$ Beta$(\omega a_t, (1-\omega)a_t)$. Then the distribution of $\{\theta_{t+1}|x_{t+1}\}$ is Gamma $(a_{t+1}, b_{t+1})$ with*

$$a_{t+1} = \omega a_t + ½, \tag{59}$$

$$b_{t+1} = \omega b_t + (x_t - \mu)^2/2. \tag{60}$$

PROOF:

Application of Shephard (1994).

From Lemma 4, the normal-gamma-beta specification is a conjugate form that is recursive by induction. Therefore, if the prior of $\theta$ is a gamma density, then the posterior always is. The following lemma describes the limiting posterior conditional on infinite past data.

LEMMA 5: *Suppose the prior distribution of $\theta_{-\infty}$ is Gamma$(a', b')$ for any nonnegative $(a', b')$. Then, in the limit, conditional on infinite observations,*

$$a_t = k/2, \tag{61}$$

$$b_t = k\nu_t/2, \tag{62}$$

*where $\nu_t$ is defined by (23).*

PROOF:

Make use of the conjugacy properties induced by Lemma 4, and confirm that in the limit (59) and (60) become (61) and (62) irrespective of initial $(a', b')$.

Equation (31) of Theorem 0 then follows as a corollary of Lemma 5 with $a' = b' = \mathcal{I} = 0$. The continuity in $\mathcal{I}$ of $p_t(\theta_t|\nu, k, \mathcal{I})$ is mathematically quite complicated, and so is better treated here as an intuitively plausible technical regularity assumption.

## REFERENCES

**Abel, Andrew B.** 2002. "An Exploration of the Effects of Pessimism and Doubt on Asset Returns." *Journal of Economic Dynamics and Control*, 26(7–8): 1075–92.

**Barro, Robert J.** 2006. "Rare Disasters and Asset Markets in the Twentieth Century." *Quarterly Journal of Economics*, 121(3): 823–66.

**Barsky, Robert B., and J. Bradford DeLong.** 1993. "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*, 108(2): 291–311.

**Bossaerts, Peter.** 1995. "The Econometrics of Learning in Financial Markets." *Econometric Theory*, 11(1): 151–89.

**Brav, Alon, and J. B. Heaton.** 2002. "Competing Theories of Financial Anomalies." *Review of Financial Studies*, 15(2): 575–606.

**Brennan, Michael J., and Yihong Xia.** 2001. "Stock Price Volatility and Equity Premium." *Journal of Monetary Economics*, 47(2): 249–83.

**Campbell, John Y.** 2003. "Consumption-Based Asset Pricing." In *Handbook of the Economics of Finance, Volume 1B,* ed. George M. Constantinides, Milton Harris, and Rene Stulz, 803–87. Amsterdam: Elsevier Science, North-Holland.

**Cochrane, John H.** 2001. *Asset Pricing.* Princeton and Oxford: Princeton University Press.

**Duffie, Darrell.** 2001. *Dynamic Asset Pricing Theory.* Third edition. Princeton and Oxford: Princeton University Press.

**Geweke, John.** 2001. "A Note on Some Limitations of CRRA Utility." *Economics Letters*, 71(3): 341–45.

**Lewellen, Jonathan, and Jay Shanken.** 2002.

"Learning, Asset-Pricing Tests, and Market Efficiency." *Journal of Finance*, 57(3): 1113–45.

**Lucas, Robert E., Jr.** 1978. "Asset Prices in an Exchange Economy." *Econometrica*, 46(6): 1429–45.

**Lucas, Robert E., Jr.** 2003. "Macroeconomic Priorities." *American Economic Review*, 93(1): 1–14.

**Mehra, Rajnish, and Edward C. Prescott.** 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, 15(2): 145–61.

**Mehra, Rajnish, and Edward C. Prescott.** 2003. "The Equity Premium in Retrospect." In *Handbook of the Economics of Finance, Volume 1B*, ed. George M. Constantinides, Milton Harris, and Rene Stulz, 889–938. Amsterdam: Elsevier Science, North-Holland.

**Rietz, Thomas A.** 1988. "The Equity Risk Premium: A Solution." *Journal of Monetary Economics*, 22(1): 117–31.

**Schwarz, Michael.** 1999. "Decision Making under Extreme Uncertainty." Ph.D. diss. Stanford University.

**Shephard, Neil.** 1994. "Local Scale Models: State Space Alternative to Integrated GARCH Processes." *Journal of Econometrics*, 60(1–2): 181–202.

**Timmermann, Allan G.** 1993. "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices." *Quarterly Journal of Economics*, 108(4): 1135–45.