Risk, Uncertainty, and Asset-Pricing 'Antipuzzles'

Martin L. Weitzman*

Harvard University

This Draft: February 13, 2006
Comments Welcomed

Abstract

In textbook expositions of the equity-premium, riskfree-rate and variability-mismatch puzzles, growth rates are typically normally distributed. But then simply recognizing that the implied distribution *conditional on realized data* is Student-t entails a startling *antipuzzle pattern*, which reverses everything needing to be explained. This paper shows generally that the unobservable nature of structural parameters adds to expectation beliefs a permanent thick-tailed background layer of uncertainty, which connects the three puzzles by one parsimonious unifying principle: measured in marginal utiles, future growth prospects are potentially much more uncertain than might be believed from merely simulating past growth rates.

1 Introduction: Structural Uncertainty and Asset Prices

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

*(e-address: mweitzman@harvard.edu) For helpful detailed comments on earlier drafts of this paper, but without blaming them for its remaining defects, I am grateful to Andrew Abel, Evan Anderson, Gary Chamberlain, Xavier Gabaix, Alfred Galichon, John Geweke, Jerry Hausman, Joseph Kadane, Narayana Kocherlakota, Lawrence Kotlikoff, Angelo Melino, Jonathan Parker, and Jeff Strnad.

the standard formulation of the problem in terms of a non-bizarre, comfortably-familiar coefficient of relative risk aversion, say with conventional values $\gamma \approx 2 \pm 1$.

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

If the aggregate stock market can be seen as some kind of a proxy for the portfolio of all wealth in an economy and if its payoffs are a proxy for total consumption, then in principle returns on comprehensive economy-wide equity should reflect more-fundamental growth expectations about the underlying real economy. What is being called here the "variability mismatch puzzle" refers to the counterintuitive empirical fact that actual returns on a representative stock market index appear to be about an order of magnitude excessively more variable than any "fundamental" that might be driving them.

The point of departure for this paper is to note that macroeconomic asset pricing is dominated by the paradigm of so-called 'rational expectations' (which more aptly might have been named 'ergodic expectations'). In this benchmark parable of a stationary-ergodic equilibrium, insider agents have effectively learned their way into knowing the "true" structural parameters of the stochastic growth process. Simultaneously, outsider econometricians have accumulated enough data to justify having a sufficiently high level of statistical confidence to effectively allow substituting sample-frequency moments for "true" population moments when fitting an Euler equation (which, strictly speaking, holds only in ex-ante subjective-belief expectations). While such classical methodology may well be appropriate for many economic applications, the paper will contend that this way of framing the issues – and even just writing an Euler equation in ex-post empirically-realized frequencies – is a fatally flawed procedure for the particular application of analyzing aversion to structural uncertainty, which underlies (or, more accurately, *should* underlie) all asset-pricing calculations.

In a *nonstationary* (or evolutionary) world, insider agents and outsider econometricians are as one in being perennially uncertain about the underlying structural parameters of the future growth process, because learning is *not* converging to an ergodic distribution of growth rates. All other things being equal, by excluding evolutionary change the 'rational expectations' vision of empirical asset pricing makes the distribution of future growth rates seem more thin-tailed and less uncertain than it actually is. When underlying coefficients do not have "true" constant values because the stochastic growth process is evolutionary with hidden structural parameters, then classical-frequentist statistical inference can understate enormously the amount of thick-tailed predictive uncertainty about the future marginal-

utility-weighted stochastic discount factor. This potentially unbounded prediction bias in forecasting expected future marginal utility spills over into severe pricing-kernel errors, which cascade into dramatically incorrect asset valuations, culminating in the bedeviling family of asset-return puzzles. Such type of effect is confirmed dramatically by just plugging into the relevant asset-pricing formulas a Student-t distribution where a normal standardly goes, and then noting the startling *antipuzzle* reversal of all inequalities needing to be explained.

This paper will show that 'rational expectations' is an inappropriate and misleading equilibrium concept for pricing assets because it represents an *unstable* knife-edge balance, which jumps discontinuously into a very different kind of asset-pricing equilibrium with radically dissimilar properties whenever just the tiniest amount of non-ergodic evolutionary-structural uncertainty is appended to the core model. Understood in a context of Bayesian statistical learning-inference, the fact that structural growth parameters of the model are unknown hidden variables introduces an extra layer of posterior uncertainty, which derives ultimately from the refusal of nondogmatic prior beliefs to preclude the evolution of unforeseen bad future histories, and whose repercussions do not dampen down to zero for any number of sample observations. Such omnipresent background uncertainty spreads out critically the probability distribution of future growth rates and is capable of acting strongly upon asset prices to increase significantly the values of both the equity premium and variability mismatch, while simultaneously decreasing markedly the riskfree interest rate.

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

A sole possible exception in the vast sea of equity-puzzle literature is a terse six-page

communication by Geweke (2001), which applies a Bayesian framework to the most standard model prototypically used to analyze behavior towards risk and then notes the extraordinary fragility of the existence of finite expected utility itself.[1] In a sense the present paper begins by accepting this shattering non-robustness insight, but pushes it further to argue that the inherent sensitivity of the standard prototype formulation constitutes a significant clue for unraveling what is driving the asset-pricing puzzles and for giving them a unified general-equilibrium interpretation that parsimoniously links together the stylized time-series facts.

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

2 The Three Macro Puzzles in Dual-Canonical Form

The critical issue for this paper is whether the appearance of the three related asset-return "macro puzzles" might essentially be attributable to background evolutionary-structural uncertainty. To cut sharply to the analytical essence of this central issue, a super-stark dual-canonical model is used where everything else except the most basic architecture of the model has been set aside. Heroically assumed away are the details in such diversionary (for this paper) complications as leverage, illiquidity, defaults, taxes, autocorrelation, irrationality, heterogeneous agents, exotic preferences, borrowing constraints, adjustment costs, business cycles, timing frictions, incomplete markets, idiosyncratic risks, and the like.

Let t denote the present period. From the present perspective, consumption C_{t+i} in

[1] I am grateful to two readers of the first draft of this paper for informing me of Geweke's pioneering earlier note, after seeing that I had independently derived results with a similar flavor.

future period $t+i$ (here with $i \geq 1$) 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 of identical people. 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 γ.

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

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

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 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 α whose gross return in period $t+1$ is R^{α}_{t+1}, the relevant Euler equation is

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

The paper will also soon treat an AK-type production version of a dynamic stochastic general equilibrium (with comprehensive K and uncertain A), but first begins with the simplest most-heroic version 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-economics literature.[2] In this endowment-exchange model of 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-

[2]The famous fruit-tree model of asset prices in a growing economy traces back to two seminal articles: Lucas (1978) and Mehra-Prescott (1985). For applications, see the survey articles of Campbell (2003) or Mehra and Prescott (2003), both of which also give due historical credit to the other pioneering originators of the important set of ideas and the stylized empirical facts used throughout this paper. Citations for the many sources of these (and related) seminal asset-pricing ideas are omitted here only to save space, and because they are readily available, e.g., in the above two review articles and in the textbook expositions of Cochrane (2001), Duffie (2001), or Gollier (2001).

period-ahead" equity, and "comprehensive" equity, all of which are abstractions of reality. 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, U.S. treasury bills, and inventories of real goods. 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 at time t as

$$P_t^f \;=\; (C_t)^{1+\gamma}\beta E_t\left[(C_{t+1})^{-\gamma}\right], \tag{5}$$

while the realized gross one-period return on the riskfree asset R_{t+1}^f in period $t+1$ is

$$R_{t+1}^f \;=\; \frac{C_t}{P_t^f}. \tag{6}$$

"One-period-ahead" 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

$$P_t^{1e} \;=\; (C_t)^{\gamma}\beta E_t\left[(C_{t+1})^{1-\gamma}\right], \tag{7}$$

while its realized gross return is

$$R_{t+1}^{1e} \;=\; \frac{C_{t+1}}{P_t^{1e}}. \tag{8}$$

"Comprehensive" equity is approximated in the real world by a broad-based representative index of publicly-traded shares of stocks whose aggregation weights mimic the wealth portfolio of the entire economy. In the theoretical fruit-tree endowment economy, "comprehensive" 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_t^e 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

$$P_t^e \;=\; (C_t)^{\gamma}\sum_{i=1}^{\infty}\beta^i E_t\left[(C_{t+i})^{1-\gamma}\right]. \tag{9}$$

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

$$R_{t+1}^e \;=\; \frac{C_{t+1}+P_{t+1}^e}{P_t^e}. \tag{10}$$

Combining (7), (8) with (9), (10) and rewriting terms gives a tight connection between

the two realized equity returns, expressed symmetrically in welfare-utility fundamentals as

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

For any time t, comprehensive financial wealth in this endowment-exchange economy is

$$W_t \;=\; C_t + P^e_t. \tag{12}$$

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

$$\frac{V_t}{U_t} \;=\; \frac{W_t}{C_t}, \tag{13}$$

which suggests that volatile wealth and volatile consumption have a symmetric relationship to welfare, an important theme that will be pursued further in Section 6 of the paper.

Turning now to the dual production side (but for brevity omitting details of a rigorous proof), it should not be difficult to grasp intuitively that comprehensive financial wealth in the pure endowment-exchange dynamic stochastic general equilibrium is dual-isomorphic to comprehensive production capital in the optimal stochastic growth problem of a linear-production AK-type model with uncertain productivity. The identification key to this full duality of endowment-production equivalence is $A_{t+i} \leftrightarrow R^e_{t+i}$ and $K_{t+i} \leftrightarrow W_{t+i}$, where the symbol "$\leftrightarrow$ " means mathematical isomorphism for all $i \geq 0$. "Comprehensive production capital" is intended here to represent the *capitalized value* (at stochastic general equilibrium prices) *of returns to all factors of production*, including labor, land, minerals, human and intangible (as well as reproducible) capital. In the AK production version with comprehensive K and stochastic A, the control variable C_{t+i} is *chosen* (just before A_{t+i+1} is realized) to maximize V_{t+i} in an expression of the form (3). The system's state-transition equation is

$$K_{t+i+1} = A_{t+i+1}[K_{t+i} - C_{t+i}] \;\leftrightarrow\; W_{t+i+1} = R^e_{t+i+1}[W_{t+i} - C_{t+i}], \tag{14}$$

where the dual-isomorphic comprehensive-wealth equation of motion in (14) comes from (12), (10). Therefore, it matters not whether stochastic consumption $\{C_{t+i}\}$ is taken as primitive in the endowment economy while stochastic returns $\{R^e_{t+i}\}$ are derived and subsequently taken as primitive stochastic productivity $\{A_{t+i}\}(= \{R^e_{t+i}\})$ for the production economy, or whether stochastic productivity $\{A_{t+i}\}(= \{R^e_{t+i}\})$ is taken as primitive in the production economy while optimal consumption $\{C_{t+i}\}$ is derived and subsequently taken as primitive for the endowment economy, because the two equilibria are not operationally distinguishable to an outside observer. The reason for stressing here this mathematical isomorphism of the

financial-wealth side with the production-capital side is to lay the groundwork for claiming later that the venerated "discipline imposed by general equilibrium modeling" does not allow one interpretation to take priority over the other, since at the level of abstraction of this macro-model comprehensive capital is comprehensive wealth and both must work in unison.

For all times t, define x_t by the equation

$$x_t \;=\; \ln C_{t+1} - \ln C_t, \tag{15}$$

which means that x_t represents the geometric growth rate of consumption during period t. At about this point in developing the argument, which up to now applies for a very general evolutionary stochastic process, the expository literature introduces the assumption of a stationary-ergodic 'rational expectations' structure. Consistent with the spirit of using a simple formulation, here the super-stationary postulate is now imposed that the random variables $\{x_t\}$ are i.i.d. with a *known* distribution – but this assumption is intended to apply *only* for expository purposes throughout the remainder of *this* section of the paper. In this special i.i.d. case, the riskfree-rate formula (6), (5) in logarithmic form becomes

$$r^f \;=\; \rho - \ln E[\exp(-\gamma x)], \tag{16}$$

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 variables $\{x_t\}$ 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

$$R^e(x) \;=\; R^{1e}(x) \;=\; A(x) \;=\; \frac{\exp(x)}{\beta E[\exp((1-\gamma)x)]}. \tag{17}$$

Taking the natural logarithm of the expected value of (17) and subtracting (16), the ergodic-average equity premium in each period (under the i.i.d.-growth assumption) is

$$\ln E[R^e] - r^f = \ln E[R^{1e}] - r^f = \ln E[\exp(x)] + \ln E[\exp(-\gamma x)] - \ln E[\exp((1-\gamma)x)]. \tag{18}$$

Equation (18) is a theoretical formula for calculating the equity risk premium, given any coefficient of relative risk aversion γ, and, more importantly here, given the i.i.d. probability distribution of the uncertain future growth rate x. Concerning the relative-risk-aversion taste parameter γ, there seems to be some rough agreement within the economics profession as a whole that an array of evidence from a variety of sources suggests that it is somewhere between about one and about three. More precisely stated, any proposed solution which

does *not* explain the equity premium for $\gamma \leq 4$ would likely be viewed suspiciously by most members of the broadly-defined community of professional economists as being dependent upon an unacceptably high degree of risk aversion.

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

$$x \sim N(\mu, V), \tag{19}$$

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

The equity premium literature generally proceeds by implicitly presuming that the "true" structural parameters μ and V are constants known by the agents inside the economy (although perhaps not known to an outside observer), and then continues on by substituting the normal distribution (19) into formula (18), which reduces (18) to a simple analyzable expression. Instead of allowing representative agents in the economy to be aware that μ 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 (18) the relevant Student-t statistic to account for structural-uncertainty sampling error).

Let $\widehat{x}$ be the sample mean and $\widehat{V}$ be the sample variance of a long time series of past growth rates. Implicitly in the 'rational expectations' interpretation, the sample size is presumed large enough to make $\widehat{x}$ and $\widehat{V}$ be sufficiently accurate estimates of their underlying "true" values μ and V so that agents inside the economy can be imagined as having substituted $(\widehat{x}, \widehat{V})$ for (μ, V) in their subjective Euler equations. However, little formal attempt is made either to define carefully for this context "sufficiently accurate" or to confirm just exactly what happens to formula (18) if the estimates, and therefore the approximations, are not "sufficiently accurate." When (19) is assumed along with the extreme point-mass dogmatic-prior case $E[x] = \widehat{x}$, $V[x] = \widehat{V}$, then using the formula for the expectation of a lognormal random variable and cancelling the many redundant terms simplifies (18) into the standard expression

$$\ln E[R^e] - r^f \;=\; \gamma V[x], \tag{20}$$

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

Considering the U.S. as a prime example, in the last century or so the average annual real arithmetic return on the broadest available stock market index is taken to be $\ln E[R^e] \approx 7\%$.[3] The historically observed real return on an index of the safest available short-maturity bills is less than 1% per annum, implying for the equity premium that $\ln E[R^e] - r^f \approx 6\%$. The mean yearly growth rate of U.S. per capita consumption over the last century or so is about 2%, with a standard deviation taken here to be about 2%, meaning $\widehat{V} \approx 0.04\%$. Suppose $\gamma \approx 2$. Plugging these values into the right hand side of (20) gives $\gamma\widehat{V} \approx 0.08\%$.

Thus, the actually observed equity premium on the left hand side of equation (18) exceeds the estimate (20) of the right hand side by some seventy-five times. If this were to be explained with the above data by a different value of γ, it would require the coefficient of relative risk aversion to be 150, which is away from acceptable reality by about two orders of magnitude. This is the form or variant of the equity premium puzzle applicable to the above dual-isomorphic endowment-production general-equilibrium model, and it is apparent why characterizing such a result as "embarrassing" may be putting it very mildly. Plugging in some reasonable alternative specifications or different parameter values can have the effect of chipping away at the puzzle, but the overwhelming impression is that the equity premium is off by at least an order of magnitude. There just does not seem to be enough variability in the recent past historical growth record of advanced capitalist countries to warrant such a high risk premium as is observed. Of course, the underlying model is extremely crude and can be criticized on any number of valid counts. Economics is not physics, after all, so there is plenty of wiggle room for a paradigm aspiring to be the "standard economic model." Still, two orders of magnitude seems like an awfully large base-case discrepancy to be explained away *ex post factum*, even coming from a very primitive model.

Turning to the riskfree rate puzzle, the meaning given in the asset-pricing literature to equation (16) parallels the interpretation given to the equity premium formula. The literature typically proceeds from (16) by postulating the normal distribution (19), but then imagines that the representative agent ignores the statistical uncertainty inherent in estimating the "true" values of μ and V. Substituting the deterministic-structure point-mass-parameter dogmatic-prior version $E[x] = \widehat{x}$, $V[x] = \widehat{V}$ into (16), and then using the formula for the expectation of a lognormal distribution, gives

$$r^f \;=\; \rho + \gamma E[x] - \frac{1}{2}\gamma^2 V[x], \tag{21}$$

[3] These numbers are from Mehra and Prescott (2003) and/or Campbell (2003), who also show essentially similar summary statistics based on other time periods and other countries (but most of which naturally have somewhat lower values of $\ln E[R_e]$ than "America in the American century").

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

Non-controversial estimates of the relevant parameters appearing in (21) (calculated on an annual basis) are: $\widehat{x} \approx 2\%$, $\widehat{V} \approx .04\%$, $\rho \approx 2\%$, $\gamma \approx 2$. With these representative parameter values plugged into the right hand side of (21), the left hand side of (16) becomes $r^f \approx 5.9\%$. When compared with an actual real-world riskfree rate $\widehat{r}^f \approx 1\%$, the theoretical formula is too high by $\approx 4.9\%$. This gross discrepancy is the riskfree rate puzzle. With the other base-case parameters set at the above values, the value of γ required to explain the riskfree interest rate discrepancy is essentially $\gamma \approx 0$, whereas $\gamma \approx 150$ is required to explain the equity risk premium. Choosing a coefficient of relative risk aversion to ease the riskfree rate puzzle exacerbates the equity premium puzzle, and vice versa. The simultaneous existence of two strong contradictions with reality, which, in addition, seem to be strongly contradicting each other, might be characterized as being "disturbing times three."

Yet another enigma is the "variability mismatch puzzle." From duality isomorphism, the assumption of i.i.d. primitive x $(= \ln A)$ for the endowment economy is symmetrically isomorphic to the assumption of i.i.d. primitive r^e $(= \ln A + \{E[r^e] - E[\ln A]\})$ for the production economy and, whatever the direction of causality in (14), from (17) it must hold identically for all i that

$$r^e_{t+i} - E[r^e] \;=\; x_{t+i} - E[x]. \tag{22}$$

In this i.i.d. economy, the entire financial-economic system vibrates in unison. According to (22), the realized deviation from the mean of continuously-compounded financial returns on comprehensive equity-wealth $r^e - E[r^e]$ should, for this simple 'rational expectations' equilibrium, 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. However, it is painfully obvious in the time-series sample that even just the two empirical second-moment variabilities are badly mismatched because the (geometrically calculated) standard deviation of equity returns $\widehat{\sigma}[r^e] \approx 17\%$ is much bigger than the (geometrically calculated) standard deviation of growth rates $\widehat{\sigma}[x] \approx 2\%$.

The relevant macroeconomic form of the "variability mismatch puzzle" is understood here to be the stylized fact that, contrary to the simple theory, in actuality the historical returns to a broad-based stock market index counterintuitively appear to be about an order of magnitude more variable than the underlying fundamental of an aggregate-output real-growth payout, for which representative equity is supposed to be the surrogate claimant. Conforming once again with the all-too-familiar quantitative asset-pricing macro-puzzle family pattern, it turns out that substituting alternative specifications or different parameter values

can lessen the initial order-of-magnitude discrepancy (here of the degree of variability mismatch between the real-production side of an economy and its dual financial-wealth side), but something central of the mystery remains that still seems way off base. Note for later reference that endowment-production duality is perfectly symmetric (and completely silent) on the critical question: *which sample variability* ($\widehat{\sigma}[x]$ or $\widehat{\sigma}[r^e]$) should be taken to calibrate welfare in the i.i.d.-normal case when it turns out empirically that $\widehat{\sigma}[x] \neq \widehat{\sigma}[r^e]$?

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

A decent heuristic intuition for what is coming up next can be gotten simply by substituting a Student-t distribution from an arbitrarily large (but finite) sample of observations for the normal distribution in formulas (16) and (18). When the limits of the relevant indefinite integrals containing the Student-t distribution are evaluated, it is readily seen from formula (16) that $r^f \rightarrow -\infty$, while from (18) careful limit calculations show that $\ln E[R^e] - r^f \rightarrow +\infty$. 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)? The luxury of ignoring these critical questions is not a viable option when the contradictions are so staggering from simply recognizing that the distribution implied by the normal *conditional on any finite sample of realized data* is the Student-t, and therefore something is seriously wrong here.

The next section of the paper shows how to pose and answer such questions rigorously in a non-ergodic general-equilibrium setting by using a hybrid distribution whose tail properties are midway between a normal and a Student-t. The formulation will "contain the t-explosion" by introducing an *ad hoc* bound on the prior (to conveniently get rid of the infinities), and then, by artificially manipulating the bound, the model will force the resulting

hybrid probability density function of x to go to the i.i.d. limit of a "Bayesian-t learning distribution" when the number of degrees of freedom is effectively made to approach infinity. This limiting hybrid "Bayesian-t learning distribution" looks like an i.i.d.-normal but in its impact on investor sentiments acts more like an i.i.d.-t, thereby generating time series data that seem highly anomalous when mistakenly judged to be the outcome of a stationary-ergodic 'rational expectations' stochastic equilibrium.

3 Hidden-Structure Expectations of Future Growth

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

To indicate where the argument is now and where it is heading, the assumptions behind the model to be used throughout the rest of the paper are stated formally here. The Euler equation (4) is presumed to hold 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 $r^e \sim N(E[r^e], V[r^e])$. The following six quasi-constants of the model are effectively assumed known: $E[r^e]$, $V[r^e]$, r^f, ρ, γ, μ. Only one structural parameter is unknown and must be estimated statistically: $V = V[x]$.

The point of what next follows is to show that the startling asset pricing *antipuzzle pattern* (from Student-t-distributed growth rates) *persists whenever there are variance shocks – even with unlimited data .* If there were an infinite record of bygone observations, then, at any time t,

$$\nu_t \;=\; \frac{1}{n}\sum_{i=1}^{\infty}(1-\frac{1}{n})^{i-1}\,(x_{t-i}-\mu)^2 \tag{23}$$

is an exponentially-weighted average back to the remotest past of all previous realized variances, which gives progressively greater influence to more recent events. The hyperparameter n appearing in (23) is a known positive integer called the *effective* sample size. It would be

neat if the standardized random variable $(\widetilde{x}_t - \mu)/\sqrt{\nu_t}$ were distributed as a Student-t with $n-1$ degrees of freedom, which would make the probability density function of $x = \widetilde{x}_t$ with $\nu = \nu_t$ be

$$\phi(x \mid \nu) \;=\; \frac{1}{\sqrt{(n-1)\pi\nu}} \, \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \, \left[1 + \frac{(x-\mu)^2}{(n-1)\nu}\right]^{-n/2} , \tag{24}$$

because then we might have some heuristic-poetic license to tell a story *as if* just before x_t is observed the random variable $\widetilde{x}_t$ is distributed as the Student-t statistic naturally associated with the outcome of "running a regression" to determine its mean (at time $t-1$, on a sample of n past realizations). Having (24) hold for *all* periods t is intuitively analogous to randomly losing one of n fictitious observations during each period, which is replaced by a new observation at the period's end – thus forcing the Student-t distribution to *always* have $n-1$ degrees of freedom. The evolutionary stochastic volatility process (23), (24) is well defined. Heuristically, n might remain constant over time (instead of increasing with the actual number of observations, thereby making the Student-t distribution converge to its underlying normal density) if the information gained from a new realization of x_t in (24) is counterbalanced by the information lost from a variance shock that changes ν_{t+1}. What next follows gives an evolutionary-Bayesian rationale for this intuitively-appealing story about Student-t-distributed growth rates having an unchanging degree of freedom.

History testifies that over time new uncertainties evolve, concerning events or situations not previously encountered. If we wish to model perceptions of new forms of uncertainty as *nonstationary* phenomena, then posterior probability updating must go beyond the standard familiar application of Bayes's rule in a *stationary* environment. The most well-known formalization of such kind of ongoing evolutionary uncertainty is the simple optimal-forecasting adaptive-expectations exponentially-distributed-lag model of a hidden-structure i.i.d.-normal stochastic process, the latent mean of which is a random walk whose nonobservable "true" current value is a moving target that can never be fully learned because the distribution-narrowing new information from one more noisy observation is just offset in informational equilibrium by the distribution-spreading new uncertainty from one more hidden random step. In that familiar model, the reduced form of the information-evaporation mechanism is an ad hoc, analytically tractable, mean-preserving linear spread per forecast-period ahead of the (known) future variance. Analogously, the setup of this paper (where the expected growth rate is always a known constant but the latent variance is unknown) wants some similarly-tractable distribution-spreading loss-of-information mechanism for making the nonobservable "true" current value of the variance be an evolving not-fully-learnable moving target, even as more information is being gathered over time.

Presuming the normal specification (19), for notational tractability the Bayesian litera-

ture tends to work with the random variable $\theta \equiv 1/V$, called the *precision*. Suppose for the sake of argument that the posterior probability density function of the precision θ takes the gamma form

$$\varphi(\theta \mid a, b) \;=\; \frac{b^a}{\Gamma(a)}\, \theta^{a-1}\, \exp(-b\,\theta) \tag{25}$$

for some yet-to-be-derived data-dependent parameters a and b. (Later it will be proved by induction on the number of past observations that (25) *must* be the relevant functional form of the distribution of θ, but meanwhile it is just taken as given.) Assume at time $t-1$ that, *conditional on* $\theta = \theta_t$, the one-period-ahead random-variable growth rate $\widetilde{x}_t$ is independent-normal:

$$\widetilde{x}_t \mid \theta_t \;\sim\; N(\mu, 1/\theta_t). \tag{26}$$

Divide any period t into two sub-periods: t° symbolizing "morning," and $t^\bullet$ symbolizing "evening." The "morning" sub-period t° represents the time segment just *after* the information contained in x_{t-1} has been observed at the conclusion of period $t-1$, but just *before* any newly-evolved uncertainty has materialized in period t. The "evening" sub-period $t^\bullet$ represents the time segment of period t just *after* the newly-evolved distribution-spreading signal has been perceived (of additional future growth uncertainty from a random change in the "true" θ), but just *before* the new information contained in a fresh value of x_t has arrived. (Arrival of x_t occurs at the "midnight instant" exactly ending period t).

With the gamma distribution (25), $E[\theta] = a/b$ and $V[\theta] = a/b^2$. Therefore, the only way to engineer a mean-preserving spread of θ that maintains the tractable form (25) is to have a and b decrease by the same proportionality factor. Towards this goal, suppose that the relationship transforming the gamma parameters $a, b = a_t^\circ, b_t^\circ$ (for the less-uncertain "morning") into the gamma parameters $a, b = a_t^\bullet, b_t^\bullet$ (for the more-uncertain "evening") always takes the reduced form[4]

$$a_t^\bullet \;=\; (1 - \frac{1}{n})\, a_t^\circ, \tag{27}$$

$$b_t^\bullet \;=\; (1 - \frac{1}{n})\, b_t^\circ. \tag{28}$$

The hyperparameter $1/n$ in (27), (28) is the fractional increase of the variance of the precision in sub-period $t^\bullet$ relative to the variance of the precision in sub-period t°. Since (27), (28) is a mean-preserving spread of θ_t, from (26) it is also a mean-preserving spread of $\widetilde{x}_t$. A *stationary* environment corresponds here to the special situation $n = \infty$, for which case

[4]Let θ^* be the hidden "true" precision. The random meander of $\{\theta_t^*\}$ that justifies the reduced form (27), (28) is of the martingale form $\theta_t^* = \eta_t \theta_{t-1}^* n/(n-1)$, having multiplicative i.i.d. shocks $\{\eta_t\}$ with beta density $\propto \eta^{\alpha-1}(1-\eta)^{\beta-1}$ for $0 < \eta < 1$, where $\alpha = a_t^\circ(1-1/n)$, $\beta = a_t^\circ/n$. (It will later be shown that with infinite data $a_t^\circ = n/2$.) Shephard (1994) used such kind of gamma-beta conjugacy in a parallel context.

a conventional application of Bayes's rule allows the "true" variance to be learned exactly 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).

The basic reason for modeling evolutionary uncertainty here by the particular ad hoc "leaky posterior information" reduced form (27), (28) is essentially the Markovian tractability it will deliver by compressing all of an infinite past into just *one* sufficient-statistic distributed-lag state variable ν_t satisfying (23). Note that objections against (27), (28) are not operational when n is a free parameter allowed in perturbation exercises to be larger than a googol: $G = 10^{100}$. The overarching message of the paper is that asset pricing is critically dependent upon the subjectively-chosen prior distribution *even* when $n > G$, and therefore 'rational expectations' cannot be relied upon to enable asset prices to learn their way out of the unavoidable influence of subjective prior beliefs. The simple reduced-form specification (27), (28) merely represents a convenient way to transmit this core message.

Each x_t is the independent realization of a normal random variable whose mean is known to be μ, but whose precision is unknown. Conditional on the precision being θ_t, the distribution of $\widetilde{x}_t$ just before it is realized or observed is given by (26), while the probability density function of θ_t is given by (25) for $a = a_t^\bullet$ and $b = b_t^\bullet$. The joint probability (just before the realization x_t is observed) of the pair $\{\theta_t, \widetilde{x}_t\}$ occurring together in this nonstationary setup is

$$p_t(\theta_t, \widetilde{x}_t) \;\propto\; \sqrt{\theta_t}\, \exp\left(-(\widetilde{x}_t - \mu)^2 \theta_t/2\right)\, \theta^{a_t^\bullet - 1}\, \exp(-b_t^\bullet\, \theta_t), \tag{29}$$

and regrouping terms in (29) just after $x = x_t$ has been observed gives (via the relevant application to this nonstationary setup of Bayes's rule) a new probability density function

$$\varphi(\theta_{t+1} \mid a_{t+1}^\circ, b_{t+1}^\circ) \;\propto\; \theta_{t+1}^{a_t^\bullet - \frac{1}{2}}\, \exp\left(-\left[b_t^\bullet + (x_t - \mu)^2/2\right] \theta_{t+1}\right). \tag{30}$$

From matching (27), (28) with (25), (30), we have thus derived the recursions

$$a_{t+1}^\circ \;=\; a_t^\bullet + \frac{1}{2}, \tag{31}$$

$$b_{t+1}^\circ \;=\; b_t^\bullet + \frac{(x_t - \mu)^2}{2}. \tag{32}$$

Turning to the prior distribution of θ (conceptualized as having been imposed *a priori* at some time $\tau < t$), for analytical tractability it is also assumed to be of the gamma form

$$\varphi(\theta \mid a_\tau, b_\tau) \;=\; \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)}\, \theta^{a_\tau - 1}\, \exp(-b_\tau\, \theta), \tag{33}$$

which, coming after (31), (32) and (27), (28), completes the induction argument in the proof that the posterior must be of the gamma form (25). The mean of the gamma prior (33) is a_τ/b_τ, while its variance is $a_\tau/{b_\tau}^2$. Thus, the prior mean and variance of θ can be assigned *any* values just by judiciously selecting a_τ and b_τ. When the simultaneous limits $b_\tau \to \infty$ and $a_\tau/b_\tau \to \theta^*$ are taken, the prior is describing a known (point mass) precision θ^*, while it turns out that classical statistical analysis essentially parallels the diffuse-prior case where a_τ and b_τ are both arbitrarily small.

It is analytically convenient to assume an infinite number of past observations, for which case $\tau = -\infty$, and any remnants of prior $a_{-\infty}$ or $b_{-\infty}$ are then completely overridden in the posterior by the data-evidence because (27), (28) (with (31), (32)) converges from *any* initial conditions $a_{-\infty}$, $b_{-\infty}$ to

$$a_t^\bullet \;=\; \frac{n-1}{2}, \tag{34}$$

$$b_t^\bullet \;=\; \frac{n-1}{2}\,\nu_t, \tag{35}$$

where ν_t is given by equation (23).

The fact that the unconditional or marginal probability of the random variable $\widetilde{x}_t$ must be a (non-standardized) Student-t distribution with $n-1$ degrees of freedom of the form (24) then follows directly from integrating out of (26) the dummy variable θ_t, whose gamma probability density function is $\varphi(\theta_t \mid a_t^\bullet, b_t^\bullet)$ with parameter values given by (34), (35). The moment generating function of a Student-t distribution is unboundedly large, thereby causing the explosion of expected marginal utility that creates the startling *antipuzzle* pattern described at the end of Section 2 (where all of the puzzling inequalities needing to be explained are dramatically reversed). To rectify such an embarrassing situation, there are any number of conceivable *ad hoc* "dampening specifications" that might be used to contain the effects on expected marginal utility of a Student-t-induced explosion. In one way or another, all such "dampening specifications" impose arbitrary *a priori* restrictions on probability distributions or utility functions – and they all give essentially the same final message. As a compromise between generality and tractability, here one particularly neat conjugate-parametric family structure is created as follows. Let δ be a non-negative hyperparameter representing an imposed *a priori* lower-bound support for the distribution of θ. Assume that in place of (33), the Bayesian prior distribution of the precision is now a *truncated* gamma probability density function (with truncation hyperparameter δ) of the form:

$$\varphi_\delta(\theta_\tau \mid a_\tau, b_\tau) \;=\; \frac{\theta_\tau^{a_t-1}\,\exp(-b_\tau\,\theta_\tau)}{\int_\delta^\infty \theta_\tau^{a_t-1}\,\exp(-b_\tau\,\theta_\tau)\,d\theta_\tau} \tag{36}$$

for $\theta_\tau \geq \delta$, while $\varphi_\delta(\theta_\tau \mid a_\tau, b_\tau) = 0$ for $\theta_\tau < \delta$. Because the entire conjugate-recursive normal-gamma family structure is preserved for $\delta > 0$, all of the results that were previously derived continue to hold with the obvious modification now that for all periods t the posterior gamma distribution of θ_t is truncated at $\theta_t = \delta$. Notice, significantly, that unlike $a_{-\infty}$ and $b_{-\infty}$, whose remnants in the posterior vanish without a trace, the *remnant of prior* δ *remains completely intact* in the posterior because the condition $\theta \geq \delta$ was imposed *a priori*.

When choosing δ to be positive, the model is dogmatically banning forever all variances above $1/\delta$. The mathematical reason for declaring permanently impermissible future worlds of unboundedly high variance is to make the integral defining the moment generating function of x converge to a finite value (because the Student-t does not). However, the implicit subtext is that nobody has the slightest idea about what is actually an appropriate value of δ, which theoretically reflects underlying prior thoughts at the infinitely-remote past time $\tau = -\infty$ but might appear to exist only for the seemingly minor mathematical purpose of placing some arbitrary finite upper bound above expected future marginal utility.

The model is made consistently recursive henceforth by now defining for any time t the fundamental state variable of the economy to be the distributed-lag sample variance ν_t defined by (23). Note that the *only* dependence on time t of the above system of equations enters via the state-variable value ν_t. Hereafter throughout the rest of the paper the subscript t is dropped from notation in this now-fully-time-autonomous hidden-structure dynamic system. Within this setup the "true" value of θ is obscured by hidden uncertainty, but the parameters $(n,\, \delta)$ and the state variable ν are known with complete certainty.

To summarize here, the asymptotic posterior-predictive probability density function of the precision is

$$\psi(\theta \mid \nu, n, \delta) \;=\; \frac{\theta^{\frac{n-1}{2}-1}\, \exp(-\frac{n-1}{2}\nu\,\theta)}{\int\limits_{\delta}^{\infty} \theta^{\frac{n-1}{2}-1}\, \exp(-\frac{n-1}{2}\nu\,\theta)\, d\theta} \tag{37}$$

for $\theta \geq \delta$, while $\psi(\theta \mid \nu, n, \delta) = 0$ for $\theta < \delta$.

From combining (37) with (26) and integrating out $\theta = \theta_t$, the unconditional (or marginal) posterior-predictive probability density function of the future growth rate x is

$$g(x \mid \nu, n, \delta) \;=\; \frac{\int\limits_{\delta}^{\infty} \exp(-\frac{\theta(x-\mu)^2}{2})\; \theta^{\frac{n}{2}-1}\, \exp(-\frac{n-1}{2}\nu\,\theta)\, d\theta}{\int\limits_{-\infty}^{\infty} \{\int\limits_{\delta}^{\infty} \exp(-\frac{\theta(x-\mu)^2}{2})\; \theta^{\frac{n}{2}-1}\, \exp(-\frac{n-1}{2}\nu\,\theta)\, d\theta\}\, dx}. \tag{38}$$

Of course an outside observer cannot know *directly* what value of δ describes an investor's prior beliefs, as δ can only be inferred indirectly from the data. A favorite default setting

would be the case $\delta = 0$, which effectively corresponds to the familiar classical normal-linear regression case because (38) approaches the Student-t form (24) as $\delta \to 0$. Speaking generally, 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 of a Student-t distribution such as (24) is unboundedly large because the defining integral diverges to plus infinity as $\delta \to 0$ in (38). A situation can therefore always be synthesized where expected pricing kernels or stochastic discount factors are made arbitrarily large simply by choosing for (38) a sufficiently small value of δ, no matter what value of $n < \infty$ has been given.

Translated into Bayesian terms here, a bare-minimum necessary prerequisite for the validity of the frequentist law-of-large-numbers justification behind calibration (the notion to "just let the data speak for themselves") is that as the number of observations approaches infinity, asset-pricing expectation formulas involving marginal utility should become uniformly free of the initial or prior state. To have 'rational expectations' 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 prior distribution of beliefs, which here means that the past data-information contained in ν should asymptotically override the influence of *any* positive value of δ – just like ν overrides *any* positive initial values of $a_{-\infty}$ and $b_{-\infty}$. Asymptotic dominance of the data over the prior often accompanies a Markov-stationary 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 θ. 'Rational expectations' is a deeply flawed equilibrium concept for pricing assets because it is describing an *unstable* razor-thin equilibrium in prices, having probability-of-existence measure zero, which unravels completely in the presence of even an infinitesimally-small bit of structural uncertainty. *For any given* $n < \infty$, the value of δ chosen for the prior manifests itself as a smear of background uncertainty that refuses, even with the interdiction of an arbitrarily large amount of past data, to relinquish its potentially decisive hold on influencing present expectations of future asset-pricing kernels.

From a Bayesian perspective, we "just let the data speak for themselves" in a different sense from the classical frequentist law-of-large-numbers interpretation of this phrase. The remaining sections of the paper will each "just let the data speak for themselves" by telling us, for the particular *ad hoc* dampening specification chosen for this paper, what are the implied values of $\delta(\nu, n)$ that real-world investors must *implicitly* be using for their priors in state-of-mind ν to be consistent with one or another stylized fact about asset returns appearing as an empirical pattern in the actual asset-return sample.

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

Taking (37) and (38) as the representative agent's subjective probability density functions, the remainder of this paper is devoted to exploring what are intended merely to be some analytically-tractable partial-equilibrium *suggestive examples* of the general theme that (contrary to 'rational expectations') subjective prior beliefs can (and presumably must) play a critical role in generating the asset-return patterns observed as stylized facts in the data. For each such *suggestive example*, the sharpest insight comes from having in mind the mental image of a double-limiting situation where simultaneously $n \to \infty$ *and* $\delta \to 0$, so the value of ν defined by (23) approaches some known constant that is unchanging over time, and also the probability density function $g(x \mid \nu, n, \delta)$ defined by equation (38) converges to the normal distribution $x \sim N(\mu, \nu)$. This prototype double-limiting situation comes arbitrarily close to the standard familiar textbook case of growth-rate risk being i.i.d. normal with known parameters, only the model never quite gets to such a stationary-ergodic normal distribution because some very small (but nevertheless consequential for asset pricing) new uncertainty evolves and subsequent learning occurs whenever $n < \infty$. Such an extreme thought experiment creates a situation that is operationally indistinguishable from the textbook workhorse 'rational expectations' i.i.d.-normal specification, and therefore it is ideal for focusing the mind very sharply on understanding intuitively a core Bayesian structural model-uncertainty mechanism common to the entire family of asset-return puzzles.

4 The Hidden-Structure Equity Premium

Rewriting (5) in state notation, the price of the riskfree asset (normalized per unit of consumption) is $P^f(\nu, n, \delta) = \beta E[\exp(-\gamma x)]$. From (7), the price of one-period-ahead equity (normalized per unit of consumption) is $P^{1e}(\nu, n, \delta) = \beta E[\exp((1-\gamma)x)]$. In both cases x is a random variable whose probability density function $g(x \mid \nu, n, \delta)$ is given by (38). The *realized* one-period ahead equity premium in ratio form is then

$$\frac{R^{1e}(x \mid \nu, n, \delta)}{R^f(\nu, n, \delta)} \;=\; \frac{P^f(\nu, n, \delta)}{P^{1e}(\nu, n, \delta)} \exp(x), \tag{39}$$

where

$$\frac{P^f(\nu, n, \delta)}{P^{1e}(\nu, n, \delta)} \;=\; \frac{\int_{-\infty}^{\infty} \exp(-\gamma x)\, g(x \mid \nu, n, \delta)\, dx}{\int_{-\infty}^{\infty} \exp((1-\gamma)x)\, g(x \mid \nu, n, \delta)\, dx}. \tag{40}$$

The following proposition contains two related types of results. First, for all $n < \infty$ (and $\nu > 0$) some value of δ matches any given feasible one period ahead asset-price ratio (40). Second, by choosing carefully $\delta(n, \nu)$ and then going to the limit $n \to \infty$, essentially any desired one-period-ahead equity premium can be replicated in a simulated data generating process *as if* it came from the super-simple i.i.d.-normal model of Section 2. (In what follows, ν plays the role of representing the current value of the state variable ν_t, while ν' plays the role of representing future values of ν_{t+i} for any $i \geq 1$.)

Theorem 1 *First part: let $\gamma > \frac{1}{2}$. Let $\overline{q}$ be any given value of the equity premium needing to be "explained." Then for every $n < \infty$ and ν' satisfying $\gamma\nu' < \overline{q}$, there exists a $\delta_q(n, \nu') > 0$ such that*

$$\frac{P^f(\nu', n, \delta_q(n, \nu'))}{P^{1e}(\nu', n, \delta_q(n, \nu'))} \;=\; \exp(\overline{q} - \mu - \frac{1}{2}\nu'). \tag{41}$$

Second part: Suppose $\nu = \nu_t < \overline{q}/\gamma$. Then for any positive integer i, as $n \to \infty$, the random variable x_{t+i} 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 $\delta \geq 0$ and of the same strength as the convergence of a Student-t distribution to a normal when the number of effective observations n approaches infinity. Furthermore, if δ is simultaneously chosen as $\delta_q(n, \nu')$ for $0 < \nu' < \overline{q}/\gamma$, then the limiting realized equity premium R^{1e}_{t+i}/R^f_{t+i} in (39) converges to the i.i.d. lognormal random variable $\exp(\overline{q} - \frac{1}{2}\nu + \sqrt{\nu} z)$ as $n \to \infty$.

Proof. Using (26) and the formula for the expectation of a lognormal random variable

rewrite (40) (after cancelling redundant terms in μ) as

$$\frac{P^f(\nu', n, \delta)}{P^{1e}(\nu', n, \delta)} \;=\; \exp(-\mu)\,\frac{\int_0^\infty \exp(\gamma^2/2\theta)\,\psi(\theta \mid \nu', n, \delta)\,d\theta}{\int_0^\infty \exp((1-\gamma)^2/2\theta)\,\psi(\theta \mid \nu', n, \delta)\,d\theta}. \tag{42}$$

As $\delta \to 0$, the probability density function $g(x \mid \nu, n, \delta)$ defined by (38) approaches the Student-t distribution (24), whose moment generating function is unbounded. Consequently, as $\delta \to 0$ both integrals in (40) and in (42) approach $+\infty$. Therefore, from (42),

$$\lim_{\delta \to 0} \frac{P^f(\nu', n, \delta)}{P^{1e}(\nu', n, \delta)} \;=\; \exp(-\mu)\,\lim_{\theta \to 0} \frac{\exp(\gamma^2/2\theta)}{\exp((1-\gamma)^2/2\theta)}. \tag{43}$$

Because

$$\ln \frac{\exp(\gamma^2/2\theta)}{\exp((1-\gamma)^2/2\theta)} \;=\; \frac{\gamma - \frac{1}{2}}{\theta}, \tag{44}$$

plugging (44) into (43) for $\gamma > \frac{1}{2}$ gives

$$\lim_{\delta \to 0} \frac{P^f(\nu', n, \delta)}{P^{1e}(\nu', n, \delta)} \;=\; \lim_{\theta \to 0} \frac{\gamma - \frac{1}{2}}{\theta} \;=\; +\infty. \tag{45}$$

At the other extreme of δ, from (40) it is apparent that as $\delta \to \infty$, then $P^f(\nu', n, \delta)/P^{1e}(\nu', n, \delta) \to \exp(-\mu)$, because the economy is then effectively in the deterministic growth case. The function $P^f(\nu', n, \delta)/P^{1e}(\nu', n, \delta)$ defined by (40) is continuous in δ. Since

$$\frac{P^f(\nu', n, \infty)}{P^{1e}(\nu', n, \infty)} \;<\; \exp(\overline{q} - \mu - \frac{1}{2}\nu') \;<\; \frac{P^f(\nu', n, 0)}{P^{1e}(\nu', n, 0)} \tag{46}$$

for $\overline{q} > \gamma\nu'$ with $\gamma > \frac{1}{2}$, condition (41) 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 $n \to \infty$ the random variable x_{t+i} converges uniformly (for all $\delta \geq 0$) to an 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 $n \to \infty$, essentially comes from (24). As $n \to \infty$, from (23) $\nu' \equiv \nu_{t+i} \to \nu$, implying $\nu' < \overline{q}/\gamma$ with probability $\to 1$. If δ is chosen to be $\delta_q(n, \nu')$, it then follows (from (39), (41), and $x \to \mu + \sqrt{\nu}z$) that as $n \to \infty$ the realized one-period-ahead (ratio) equity premium converges to the i.i.d.-lognormal random variable $\exp(\overline{q} - \frac{1}{2}\nu + \sqrt{\nu}z)$. ■

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 $\delta \to 0$ for any fixed $n < \infty$,

$\nu > 0$, the limit of (38) is a Student-t distribution of the form (24), the same as would emerge if a regression had been run on n data points. With the presumed prototype case n huge and δ tiny, the central part of the t-like distribution (38) is approximated extremely well by a normal curve with mean μ and variance ν fitting the data throughout its middle range. However, for applications involving the implications of uncertainty-aversion, such as calculating the equity premium, to ignore what is happening away from the center of the distribution has the potential to wreak havoc on subjective-expectation-based asset-price calculations. For these applications, such a normal distribution may be a very bad approximation indeed, because the more-spread-out dampened-t distribution (38) is capable in principle of producing an explosion in asset pricing formulas like (40), implying in the limit as $\delta \to 0$ an unboundedly large equity premium. Within the Bayesian framework of this model, therefore, the statistical fact that the moment generating function of a Student-t distribution is infinite has the important economic interpretation that, at least hypothetically, evolving model-structure uncertainty has the potential in such a normal-gamma world to be a far more significant determinant of asset prices than pure stationary-ergodic risk. In the limit as $\delta \to 0$ (for fixed $n < \infty$), 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* structural parameters. The key to understanding the 'rational expectations' dilemma concerning how to interpret the "equity premium puzzle" is that the "premium" is not on pure *ergodic risk* alone, but rather it is a *combined premium* on ergodic risk *plus* (potentially vastly more significant) *structural uncertainty*.

An explosion of the equity premium does not happen in the real world, of course, but a contained near-explosive outcome remains the mathematical driving force behind the scene, which imparts the statistical illusion of an enormous equity premium incompatible with the standard neoclassical paradigm. When people are peering forward into the future they are also looking back at their own prior, and what they are seeing 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 just when their world has already taken a very bad turn. This eerie sensation of low-δ diffuse background shadow-risk may not be simple to articulate, yet it frightens investors away from taking a more aggressive stance in equities and scares them into a more apprehensive position of wanting to hold instead (on the margin) a portfolio of some safer stores of value, such as hard-currency cash, inventories of real goods, Swiss bank accounts, U.S. or U.K. short-maturity treasury bills, perhaps precious

metals, or even stockpiles of food – as a hedge against unforeseen bad future evolutions of history. Consequently, in an evolutionary equilibrium where there is zero net demand for them, these relatively-safe assets bear very low, even negative, real rates of return.

I do not believe that such type of Bayesian statistical explanation is easily dismissable. The equity premium puzzle is the *quantitative* paradox that the observed value of $\ln E[R^e]-r^f$ is too big to be reconciled with the standard neoclassical stochastic growth paradigm having familiar parameter values. But *compared to what* is the observed value of $\ln E[R^e]-r^f$ "too big"? Essentially, the answer given in the equity-premium literature is: "compared to the right hand side of formula (20) when $\widehat{V} \approx 0.04\%$ and $1 \leq \gamma \leq 4$." Unfortunately for this logic, the point-calibrated right hand side of (20) gives a terrible prediction for the observed realizations of R^e/R^f because in the underlying calculation all assets have been priced by a 'rational expectations' formula that makes the future seem far less uncertain than it actually is. Anyone wishing to downplay this line of reasoning in favor of the *status quo ante* would be hard pressed to come up with their own Bayesian rationale for calibrating variances of nonobservable subjectively-distributed future growth rates by point estimates equal to past sample averages. In essence, the 'rational expectations' approach that produces the family of asset-pricing puzzles avoids the consequences (on marginal-utility-weighted asset-pricing kernels) of overpowering sensitivity to low values of prior δ only by effectively imposing from the very beginning the extremely-brittle deterministic-structure pure-ergodic-risk case $n = \infty$ of a normal distribution with known parameters.

An early attempt to explain the equity premium puzzle by Rietz (1988) can be interpreted as essentially arguing through numerical examples that either the sample or the imposed structure of the model (or both) may not be adequately representing a worst-imaginable-case scenario of large negative future growth rates. The impact on financial equilibrium of a situation where there is a tiny probability of a catastrophic out-of-sample or wrong-structure-indicating event has been dubbed the "peso problem." In a peso problem, the small probability of a disastrous future happening (such as a collapse of the presumed structure from a natural or socio-economic catastrophe) is taken into account by real-world investors (in the form of a "peso premium") but not necessarily by the calibrated model, because such an event may not be in the sample being used for the calibration. In a sense this paper is providing a statistical-decision-theoretic microfoundation for explaining why a form of peso-problem logic is unavoidable in asset pricing because it is genetically ingrained.

Theorem 1 is trying to tell us that the statistical architecture of something conceptually akin to a peso problem is inescapably hardwired into the "deep structure" of how Bayesian inferences (about future economic growth, at unknown rates) interact with *any concave utility function* $U(C)$ having the innocuous curvature property $-CU''(C)/U'(C) > \frac{1}{2}$. Bayesian

inferences about the unknown hidden-structure variance fatten the posterior tails of probability density functions with dramatic consequences when expressed in subjective-expectation units of future marginal utility – as the example of replacing the workhorse normal distribution by its t-like posterior distribution demonstrates. This "Bayesian-statistical peso problem" means that for asset pricing applications it is not the least bit absurd or unscientific to adhere to the non-rational-expectations non-ergodic idea that *no amount of data may be large enough to identify all of the relevant structural uncertainty concerning future economic growth.* The Bayesian peso problem is essentially saying that to calibrate an exponential evolutionary process having an uncertain future growth rate by plugging the sample variance of observed past growth rates into an "extremely bad" approximation of the subjectively-distributed stochastic discount factor, is to underestimate "extremely badly" the comparative utility-risk of a real-world gamble on the unknown structural potential for future economic growth, relative to a safe investment in a near-money sure thing.

Translated into classical-frequentist statistical language, the second part of the theorem has the following rigorous interpretation. For given $\nu = \nu_t$, pick *any* equity premium $\overline{q} > \gamma\nu$, name some sample size k, and choose *any* desired level of statistical confidence relative to the supposedly "true" data generating process. Then there exists some sufficiently large n and accompanying function $\delta = \delta_q(n, \nu')$ (where for ν' will be substituted future realizations of ν_{t+i}, with $1 \leq i \leq k$) such that the empirically observed frequency distribution of the k realized values of the one-period-ahead 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 k *if* the "true" equity premium $r^{1e} - r^f$ were i.i.d. $N\left(\overline{q} - \frac{\nu}{2}, \nu\right)$. (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 $\nu = \widehat{V}[x]$ instead of $\widehat{V}[r^e]$ – more on this variance mismatch later.)

Of course, what is being presented here is but one illustrative example of the economic consequences of a hidden-structure tail-fattening effect, but I believe that it is very difficult to get around the moral of this story. For any finite value of n, however large, the results of Bayesian distribution-spreading will cause the equity premium to be very sensitive to seemingly negligible changes in the assumed prior distribution of the variance, when, according to the key stationary-ergodic assumption behind 'rational expectations,' such seemingly innocuous prior-belief changes should have been learned away by the data-evidence long ago. This kind of extraordinary fragility to subjective prior beliefs even with unlimited data effectively renders unbelievable the standard 'rational expectations' parable of asset pricing. The dominant statistical-economic force behind the puzzles is that seemingly thin-tailed

probability distributions (like the normal), which actually are only thin-tailed *conditional on known structural parameters* of the model, become thick-tailed (like the Student-t) after integrating out the parameter uncertainty. Intuitively, no finite sample of effective size $n < \infty$ can accurately assess tail thickness, and therefore the attitude of a risk-averse Bayesian agent towards investing in various risk-classes of assets may be driven to an arbitrarily large extent by this unavoidable feature of Bayesian expectational uncertainty.

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

5 The Hidden-Structure Riskfree Interest Rate

We can use the same mathematical-statistical apparatus to calculate the hidden-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 $f(\nu, n, \delta)$ be the value of r^f that comes out of formula (16) when the probability density function of x is $g(x \mid \nu, n, \delta)$ defined by equation (38). Plugging the subjective posterior-predictive distribution (38) into the right hand side of equation (16), the result is

$$f(\nu, n, \delta) \;\equiv\; \rho - \ln \int_{-\infty}^{\infty} \exp(-\gamma x)\, g(x \mid \nu, n, \delta)\, dx. \tag{47}$$

Theorem 2 *Let $r^f(\nu')$ be any given continuous function of ν' satisfying $r^f(\nu') < \rho + \gamma\mu -$*

$\frac{1}{2}\gamma^2\nu'$ *for all* $\nu' > 0$. *Then for every* $n < \infty$, $\nu' > 0$, *there exists a* $\delta_f(n, \nu') > 0$ *such that*

$$r^f(\nu') = f(\nu', n, \delta_f(n, \nu')). \tag{48}$$

Furthermore, the limiting realized riskfree rate can be made to converge to the same constant value $\overline{r}^f$ *if, in every future state* $\nu' = \nu_{t+i}$, *the value of* δ *is chosen to be* $\delta_f(n, \nu')$ *for* $r^f(\nu') = \overline{r}^f$ *and then the limit* $n \to \infty$ *is taken.*

Proof. As $\delta \to 0$, the probability density function $g(x \mid \nu, n, \delta)$ defined by (38) approaches the Student-t distribution (24), whose moment generating function is unbounded. From (47) therefore, $f(\nu', n, 0) = -\infty$. It is apparent that as $\delta \to \infty$, then $f(\nu', n, \infty) = \rho + \gamma\mu$, because the economy is then effectively in a situation of deterministic growth. Thus,

$$f(\nu', n, 0) < r^f < f(\nu', n, \infty), \tag{49}$$

and, since $f(\nu', n, \delta)$ defined by (47) is continuous in δ, the conclusion (48) follows. The convergence to a constant value for all future periods follows from the fact that ν effectively becomes constant over time as $n \to \infty$, so that the condition $r^f(\nu') = \overline{r}^f < \rho + \gamma\mu - \frac{1}{2}\gamma^2\nu'$ holds on the future trajectory with probability $\to 1$ as $n \to \infty$. ■

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 somewhat like a Student-t statistic 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" causes incorrect stationary-ergodic 'rational expectations' inferences about expected future utility, which are based upon mimicking the observed historical frequency of past growth rates, to underestimate enormously just how relatively much more attractive are safe stores of value when compared with a real-world Bayesian gamble on the uncertain growth-structure of an unknown future economy.

The relevant classical-frequentist statistical statement here about the relationship between the riskfree rate that is observed in the data and the supposedly "true" data generating process parallels the equity premium version. Pick $r^f = \overline{r}^f < \rho + \gamma\mu - \frac{1}{2}\gamma^2\nu$, name some number k, and choose any desired level of statistical strength, here representing measurement accuracy. Then there exists some (large) n and accompanying $\delta = \delta_f(n, \nu')$ such that the frequency distribution of the k riskfree-rate realizations generated by *this* hidden-structure model is guaranteed statistically to differ only within measurement error from what *would* be generated in a sample of size k *if* the "true" riskfree rate were the constant value $\overline{r}^f$.

6 Welfare-Equivalent *as if Normal* Growth Variability

It has already been amply demonstrated that the dynamic evolution of future asset prices is wickedly sensitive to prior beliefs, even with infinite past data. Such a complicated story about asset-return trajectories just cannot be distilled down into the neat form of a tightly-compressed parable about a stationary-ergodic world. Since it seems to be hardwired into the human brain to want desperately to cling to any kind of simplifying mental image conceivably available to help comprehend the evolution of an extraordinarily complicated universe, this section of the paper is openly heuristic in pursuing that aim. It is offering nothing more than a quick-and-dirty intuitive interpretation of the "variability mismatch puzzle" expressed in subjective as-if-stationary expected-utility-welfare terms. Its purpose is merely to convey some feel for the magnitude of the cost of structural uncertainty by phrasing it in a user-friendly welfare-equivalent version of the familiar i.i.d.-normal distribution.

For the dual-equivalent endowment-production i.i.d. equilibrium in Section 2 of this paper, equity returns vibrate consistently with growth rates as prescribed by equation (22). According to (22), 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 x $(= \ln A)$ and r^e $(= \ln A + \{E[r^e] - E[\ln A]\})$ should match subjectively and objectively under 'rational expectations.' Alas, the empirical second moments of x and r^e are not even remotely matched in the time-series data because $\widehat{V}[r^e]/\widehat{V}[x] \approx 75$. In the evolutionary version of the model, however, *future x is subjectively perceived as being much more variable than it seems to be from past time series data* in the sense that the relevant stochastic discount factor $M = \beta \exp(-\gamma x)$ is subjectively expected to be much larger in expected utiles than what would appear to be indicated by simply identifying the variance of future as-if-normal x with its past sample average ν, which would give the plug-in lognormal expected-utile value $\beta \exp(-\gamma\mu + \frac{1}{2}\gamma^2\nu) \ll E[M]$.

The price-earnings ratio P^e/C of comprehensive equity in (9) depends on expectations over an infinite future horizon, and is extraordinarily sensitive to δ (presumably even more so than the riskfree rate, which depends only on one-period-ahead expectations). Such extreme sensitivity to subjective prior beliefs suggests very strongly that a fuller more-complicated model could be built around a "trembling hand" inducing measurement errors that cause tiny contaminations of δ to become amplified into large animal-spirit-like price fluctuations, thereby potentially introducing into the model elements of what might look like "predictability." However that may be, as a practical matter (whatever is the causal mechanism producing large swings in stock-market prices), to proceed further here analytically requires some simplifying assumption about the *reduced form* of equity returns. The textbook bench-

mark 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 will now be made to revolve around this centerpiece assumption.

From the basic duality isomorphism between production and endowment versions of the core dynamic stochastic model and from the "discipline imposed by general equilibrium modeling," if the productivity-return $r^e\,(=\ln A+\{E[r^e]-E[\ln A]\})$ is *known* to be i.i.d.-normal then so too must the growth rate x $(=\ln A)$ be i.i.d.-normal, and with the same standard deviation. In this case equation (22) 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. In a sense, this section of the paper is trying to answer the critical question: between the two observed variability alternatives ($\widehat{\sigma}[r_e]$ standing in for the left hand side of equation (22) and representing the past variability of wealth returns or $\widehat{\sigma}[x]$ standing in for the right hand side and representing the past variability of consumption growth), *which empirical variability* (wealth or consumption) *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 in this section of the paper that

$$x^N(x\mid \nu,n,\delta)\ \sim\ N(E[x^N],\sigma^2[x^N]) \tag{50}$$

is a random variable function of the random variable x representing the agent's *subjective probability belief* that future growth rates are i.i.d.-normal with known parameters $E[x^N]$ and $\sigma[x^N]$. Let this agent also have a *subjective probability belief* in a stock-market payoff implicitly representing a unit claim on the lognormally-i.i.d. future aggregate consumption corresponding to (50). 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

$$r^N(x^N(x))-E[r^N]\ \ =\ \ x^N(x)-E[x^N], \tag{51}$$

which is exactly the normal counterpart here of equation (22). It will turn out that i.i.d. as-if-normal growth rates can be made to yield the same expected one-period return on equity as the formulation in previous sections of the paper, so that $E[r^N]=E[r^{1e}]$, which, provided also that $\sigma[r^N]=\widehat{\sigma}[r^e]$, signifies here that *observed equity data alone* cannot refute

this as-if-normality hypothesis about subjective future growth beliefs, given the standard assumption that equity returns are *known* by the agents to be i.i.d. normal in the first place.

The following representation theorem establishes the existence of a conceptually-useful consequence of expected-utility indifference between x^N and x. In the framework of this model, it turns out that forcing x^N by construction to give the same expected utility as x is intimately connected with the important implication for welfare calibration that $\sigma[x^N] \approx \widehat{\sigma}[r^e]$. This third proposition of the paper can therefore be interpreted as providing at least a sense in which there might be *some* rationale for telling an as-if parable wherein the representative agent has a subjective normally-distributed welfare-equivalent belief, which is consistent with (51) and the equity-return data, "as if" the future growth rate is x^N with known variability equal to the observed variability of returns on 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 variability of returns on equity-wealth, rather than by the relatively low variability of realized past growth rates. Because here $\sigma[r^N] = \sigma[x^N]$, at least from within the framework of this artificially constructed welfare-equivalent as-if-i.i.d.-normal-growth parable, there is no longer a jarring mismatch of variabilities wanting to be explained between equity returns and underlying fundamentals.

Theorem 3 *Let $\sigma > 0$ be any given continuous function of ν satisfying $\sigma^2 > \nu$ for all $\nu > 0$. Let $r^N(x^N(x))$ be a solution of (50), (51). Then for every $n < \infty$, $\nu > 0$, there exists a $\delta_s(n, \nu) > 0$ such that the following four conditions are simultaneously satisfied:*

$$E[r^N(x^N(x))] \;=\; E[r^{1e}(x)], \tag{52}$$

$$\sigma[r^N(x^N(x))] \;=\; \sigma[x^N(x)] \;=\; \sigma, \tag{53}$$

$$E[x^N(x)] \;=\; E[x] \;=\; \mu, \tag{54}$$

$$\forall C > 0: \; E[U(C\exp(x^N(x)))] \;=\; E[U(C\exp(x))]. \tag{55}$$

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

$$\frac{1}{\sqrt{2\pi}s(\delta)}\int_{-\infty}^{\infty}\exp\left((1-\gamma)x^N - \frac{(x^N-\mu)^2}{2s(\delta)^2}\right)dx^N \;=\; \int_{-\infty}^{\infty}\exp((1-\gamma)x)\,g(x \mid \nu, n, \delta)\,dx, \tag{56}$$

and note for this definition that (55) and (54) are satisfied by construction.

It can readily be shown that

$$r^{1e}(x) \;=\; x + \{\rho - \ln E[\exp((1-\gamma)x)]\}, \tag{57}$$

and, analogously,

$$r^N(x^N) \;=\; x^N + \{\rho - \ln E[\exp((1-\gamma)x^N)]\}, \tag{58}$$

so that (52) then follows from (54), (56), (57), (58).

As $\delta \to \infty$, the probability density function $g(x \mid \nu, n, \delta)$ goes to the deterministic point distribution $x = \mu$, so consequently the integral on the right hand side of equation (56) approaches $\exp((1-\gamma)\mu$, implying $s(\infty) = 0$. As $\delta \to 0$, the probability density function $g(x \mid \nu, n, \delta)$ defined by (38) approaches the Student-t distribution (24), whose moment generating function is unbounded, implying the right hand side of (56) is also unbounded, meaning $s(0) = \infty$. Thus

$$s(\infty) \;<\; \sigma \;<\; s(0), \tag{59}$$

and, by continuity of the function $s(\delta)$, there must exist a $\delta_s(n, \nu) > 0$ satisfying

$$s(\delta_s) \;=\; \sigma, \tag{60}$$

which, when combined with (51), proves (53) and concludes the proof. ■

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

Theorem 3 is effectively saying that if you want to force the wickedly complicated dynamic behavior of super-prior-sensitive asset prices under evolutionary uncertainty into the analytically tractable image conveyed by a prior-free 'rational expectations' as-if-i.i.d.-normal-growth story, then the calibration $\sigma[x^N] = \widehat{\sigma}[r_e]$ tells the better as-if welfare parable than the calibration $\sigma[x^N] = \widehat{\sigma}[x]$. To an outsider classical-frequentist statistician imposing a stationary specification, however, agent-investors inside this as-if-rational-expectations economy will appear to be irrationally incapable of internalizing what the data are clearly saying about $\widehat{\sigma}[x] \approx 2\%$. Instead, with $n \to \infty$ these agents seem to be clinging stubbornly in their mind's eye to an unshakably-consistent, but highly irrational, mental image *as if* their future welfare depends via the stochastic discount factor upon the realization of some hypothetical much-more-variable normally-distributed growth rate whose counterfactual standard deviation is $\sigma[x^N] = \widehat{\sigma}[r^e] \approx 17\%$. With, say, one hundred independent observations, however, the frequentist hypothesis that the observed sample value of $\widehat{\sigma}[x^N] = 2\%$ could have been gen-

erated by (agents having in their heads) a "true" (welfare equivalent) value of $\sigma[x^N] = 17\%$ is classically rejected by a chi-square test at the 99.99% confidence level!

7 Some Bayesian *as if Normal* Calibration Exercises

Viewing the three theorems of the paper through the lens of the welfare-equivalent as-if-i.i.d.-normal-growth story of Theorem 3 delivers the package of a neat analytically-tractable relationship among q, f, and s of the closed form (20), (21), which accompanies the well-known formula for the expectation of a lognormal random variable. The three theorems themselves are only partial equilibrium statements in the sense that each one matches just one side of the whole asset-pricing-puzzle triangle. The $\delta(n, \nu)$ function that works for any one theorem will not work for the other two – essentially because a system parameterized with just one degree of freedom cannot match three observables simultaneously. 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 $\delta(n, \nu)$ function works for all three theorems. The following question then arises naturally: does the simple relationship among q, f, and s of the closed form (20), (21) hold *empirically*, conditional upon the same $\delta(n, \nu)$ function working for all three theorems? The answer is "yes." The experimental outcome that all three stylized-fact values of the equity premium, riskfree rate, and equity variability "fit," in the sense that they come close to matching simultaneously the theoretically-predicted as-if-lognormal relationship among themselves, conveys at least *some* intuitive feel for the degree to which this heuristic way of looking at things 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 $E[x^N] = \widehat{x}$ and standard deviation $\sigma[x^N] = \widehat{\sigma}[r^e]$ renders, along with (51), an internally-consistent as-if story connecting together the actual stylized facts of our economic world. In Table 1, quasi-constant parameter settings have been selected that, I think, represent stylized-fact numbers well within the "comfort zone" for most economists. All rates are real and given by annual values. The data are intended to be an overall approximation of what has been observed for many countries over long time periods.

With any given $n < \infty$, $\nu > 0$, the model "explains" endogenously three quasi-constants, written here for simplicity (by suppressing dependence on n and ν) as $q(\delta)$, $f(\delta)$, $s(\delta)$ – all three being functions of the one free parameter δ. For conceptual-notational convenience, pretend that $h = 6.625 \times 10^{-34}$ represents some arbitrarily small quantum threshold of observability, below which δ is considered to be "effectively zero" and the relationship between

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

Table 1: Some Economic "Stylized Facts"

δ_q, δ_f, and δ_s becomes so blurred by indeterminacy that the situation is treated as if the *same* $\delta(n, \nu)$ function (being indistinguishable from zero) works for all three theorems. Under such circumstances, we will not be able to observe or calculate the underlying primitive values of δ_q, δ_f, and δ_s *directly* (although we know in theory that there exists some astronomically-large value of $n > G = 10^{100}$, for which simultaneously $\delta_q < h$, $\delta_f < h$, and $\delta_s < h$, because $n \to \infty$ implies that $\delta_q \to 0$, $\delta_f \to 0$, and $\delta_s \to 0$). However, and more usefully here, an *indirect* calibration experiment can be performed by setting any one of the three quasi-constants $q \mid \{\delta_q < h\}$, $f \mid \{\delta_f < h\}$, and $s \mid \{\delta_s < h\}$ equal to its observed value in Table 1 and then backing out the implied values of the other two remaining quasi-constants by inverting the two analytically-tractable as-if-i.i.d.-lognormal-consumption equations of the closed form (20) and (21). Because there are two equations ((20) and (21)) uniformly-continuous in three unknown variables, if *any one* of $\{\widehat{q},\ \widehat{f},\ \widehat{s}\}$ is "sufficiently near" to explaining the other two, then *each* of $\{\widehat{q},\ \widehat{f},\ \widehat{s}\}$ must also be "sufficiently near" to explaining the other two. The following calibration exercise shows empirically that the entire i.i.d.-lognormal system is "sufficiently near" to $\{\widehat{q},\ \widehat{f},\ \widehat{s}\}$ in the distance-metric of what might be considered on intuitive grounds to be the most natural topology to use here.

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

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

we then have, from (20) with $V[x] \equiv s^2(\delta_s)$,

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

to be compared with $q \mid \{\delta_q < h\} \ = \ 6\%$. From (21) with $V[x] \equiv s^2(\delta_s)$,

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

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

Defining $\delta_q < h$ to be the implicit solution of

$$\delta_q \;=\; q^{-1}(\ln E[R^e] - r_f) \;=\; q^{-1}(6\%),$$

we then have, from (21) and (20),

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

to be compared with $f \mid \{\delta_f < h\} \;=\; 1\%$. From (20) with $V[x] \equiv \sigma^2[r^e]$,

$$\sigma[r^e] \;=\; \sqrt{q(\delta_q)/\gamma} \;=\; s \mid \{\delta_q < h\} \;=\; 17\%,$$

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

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

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

we then have, from (21) and (20),

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

to be compared with $q \mid \{\delta_q < h\} \;=\; 6\%$. From (21) with $V[x] \equiv \sigma^2[r^e]$,

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

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

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

Continuing 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 σ 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 (15) is

$$\overline{C}_{t+1} \;=\; \exp(\mu - \frac{1}{2}\sigma^2)\,\overline{C}_t. \tag{61}$$

Using formula (61), 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 (61) is equivalent to a change in each period's consumption of

$$\Delta C_t \;=\; (\exp(\frac{1}{2}\gamma\sigma^2) - 1)\,C_t. \tag{62}$$

When $\gamma \approx 2$ and the historical value of $\sigma = \widehat{\sigma}[x] \approx 2\%$ is used in (62), then $\Delta C_t/C_t \approx 0.04\%$, which is the kind of magnitude sometimes used to argue that the cost of growth variability is so counterintuitively low that even a complete removal of all conceivable macroeconomic uncertainty would be worth very little. Such a number, however, captures only the "weak force" of stationary-ergodic growth-rate risk. The welfare equivalent of a magic-trick elimination of *all* uncertainty about future growth, including the "strong force" of structural uncertainty, is better assessed by using the subjective value $\sigma = \widehat{\sigma}[x_N] = \widehat{\sigma}[r^e] \approx 17\%$ in formula (62), for which case $\Delta C_t/C_t \approx 2.9\%$. Accounted in this welfare-equivalent metric of shrunken consumption therefore, structural *uncertainty* about the evolving future growth process turns out empirically to be far more significant than pure growth-rate *risk*.

8 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 a "variability mismatch puzzle," whose magnitudes of discrepancy are close numerically to what is observed empirically. This paper argues that such numerical "discrepancies" are puzzles, however, only when seen through a 'rational expectations' lens. From a Bayesian learning perspective, the "puzzling" numbers being observed in the data are telling a rational (but not 'rational expectations' in the conventional stationary-recurrent-ergodic sense) story about the implicit subjective distribution of background structural-parameter uncertainty accompanying the uncertain evolutionary growth process actually generating such data.

In principle, consumption-based representative agent models provide a complete answer to all macroeconomic asset pricing questions and give a unified theory integrating together the economics of finance with the real economy. In practice, consumption-based representa-

tive agent models with standard preferences and a traditional degree of relative risk aversion work poorly when the variance of the growth of future consumption is point-calibrated to the sample variance of its past values. The theme of this paper is that with evolutionary-structural uncertainty there is some theoretical justification for treating the 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 which *as if* interpretation the simple standard model may have the potential to work well in practice.

References

[1] Abel, Andrew B. (2002), "An Exploration of the Effects of Pessimism and Doubt on Asset Returns." *Journal of Economic Dynamics and Control*, 26, 1075-1092.

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

[3] Bossaerts, Peter (1995), "The Econometrics of Learning in Financial Markets." *Econometric Theory*, 11, 151-89.

[4] Brav, Alon, and J. B. Heaton (2002), "Competing Theories of Financial Anomalies." *The Review of Financial Studies*, 15, 575-606.

[5] Brennan, Michael J. and Yihong Xia (2001), "Stock Return Volatility and Equity Premium." *Journal of Monetary Economics*, 47, 249-283.

[6] Campbell, John Y. (2003), "Consumption-based asset pricing." Chapter in Constantinides, Harris, and Stulz, *Handbook of the Economics of Finance*. Elsevier B.V.

[7] Cecchetti, S. G., P. S. Lam, and N. C. Mark (2000), "Asset Pricing with Distorted Beliefs: Are Equity Returns too Good to be True?" *American Economic Review,* 90, 787-805.

[8] Cochrane, John H. (2001), *Asset Pricing*. Princeton University Press, NJ.

[9] Duffie, Darrell (2001), *Dynamic Asset Pricing Theory* (third edition). Princeton, NJ: Princeton University Press.

[10] Gollier, Christian (2001), *The Economics of Risk and Time*. Cambridge, MA: The MIT Press.

[11] Geweke, John (2001), "A note on some limitations of CRRA utility." *Economics Letters*, 71, 341-345.

[12] Lewellen, Jonathan, and Jay Shanken (2002), "Learning, Asset-Pricing Tests, and Market Efficiency." *Journal of Finance*, 57, 1113-1145.

[13] Lucas Jr., Robert E. (1978), "Asset Prices in an Exchange Economy." *Econometrica,* 46, 1429-1445.

[14] Lucas Jr., Robert E. (2003), "Macroeconomic Priorities." *American Economic Review,* 93(1), 1-14.

[15] Mehra, Rajnish, and Edward C. Prescott (1985), "The Equity Premium: a Puzzle." *Journal of Monetary Economy,* 15, 145-161.

[16] Mehra, Rajnish, and Edward C. Prescott (2003), "The equity premium in retrospect." Chapter 14 in Constantinides, Harris, and Stulz, *Handbook of the Economics of Finance.* Elsevier B.V.

[17] Rietz, Tom (1988), "The equity risk premium: a solution." *Journal of Monetary Economics*, 21, 117-132.

[18] Schwarz, Michael (1999), "Decision Making Under Extreme Uncertainty." Stanford University Graduate School of Business: Ph.D. Dissertation.

[19] Shephard, Neil (1994), "Local Scale Models." *Journal of Econometrics*, 60, 181-202.

[20] Timmermann, Allan G. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices." *Quarterly Journal of Economics*, 108, 1135-1145.

[21] Veronesi, Pietro (2000), "How Does Information Quality Affect Stock Returns?" *Journal of Finance*, 55, 807-837.