A Unified Bayesian Theory of Equity 'Puzzles'

Martin L. Weitzman*
Harvard University

This Draft: April 1, 2005
Comments Appreciated

Abstract

In expositions of the equity premium, risk-free rate, and excess volatility puzzles, the subjective distribution of future growth rates has its mean and variance calibrated to past sample averages. This paper shows that proper Bayesian estimation of uncertain structural growth parameters adds an irreducible fat-tailed background layer of uncertainty that can explain all three puzzles parsimoniously by one unified theory. The Bayesian statistical-economic equilibrium has just one degree of freedom, yet the data generating process of the model matches simultaneously all three empirically-observed values of the equity premium, risk-free rate, and excess volatility.

1 Introduction: Structural Uncertainty and Asset Prices

The "equity premium puzzle" refers to the spectacular failure of the standard neoclassical representative-agent model of stochastic economic growth to explain a historical difference of some six 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. To justify such a large risk premium suggests that either there is more uncertainty out there than the data appear to indicate or else something is fundamentally wrong with the standard formulation of the problem in terms of a non-bizarre, comfortably-familiar coefficient of relative risk aversion, say with values $\gamma \approx 2 \pm 1$.

The "risk-free rate puzzle" represents another big disappointment with the standard neoclassical model. With $\gamma \approx 2$, the stochastic generalization of the basic Ramsey formula

*(e-address: mweitzman@harvard.edu) For helpful detailed comments on an earlier draft of this paper, but without blaming them for remaining defects, I am grateful to Andrew Abel, Gary Chamberlain, Xavier Gabaix, Alfred Galichon, John Geweke, Joseph Kadane, Naranya Kocherlakota, and Jonathan Parker.

from equilibrium growth theory predicts a risk-free interest rate in the range $r_f \approx 5 - 6\%$, while what is actually observed is in the range $\widehat{r}_f \approx 0 - 1\%$.

The "excess volatility puzzle" stems from the generic idea that returns on stocks should reflect expectations of their underlying fundamentals. In the form most relevant to the sparse macro framework of this paper, aggregate returns on comprehensive economy-wide equity should mirror fundamental growth expectations about the underlying real economy. But this seemingly-intuitive idea appears from a first glance at the data to be way off, because actual returns on a representative stock market index are an order of magnitude excessively more volatile than any "fundamental" that might be driving them.

Taken together, this unholy trinity of theoretical predictions that miss their empirical targets by orders of magnitude is devastating for the credibility of the neoclassical paradigm. The proper interpretation of these equity macro-puzzles has important ramifications throughout all branches of economics because it goes centrally to the core issue of how to discount time and risk. The three intuitively-related paradoxes are fairly crying out that something is deeply wrong with the standard paradigm, which is unlikely to be corrected by tinkering with small modifications of the basic model. Some big critical element, which would capture the characteristic that appears to make stocks comparatively so risky, seems to be missing from the formulation. At least for asset pricing applications, a consensus has developed among economists that the standard model is seriously flawed.

Not surprisingly therefore, this family of equity macro-puzzles has stimulated a lot of economic research. In attempting to explain the paradoxes, an enormous post-puzzles literature has developed, which is filled with some imaginatively fruitful variations on the standard model. For example, to overcome one or another equity puzzle many new models feature exotic (and complicated) reverse-engineered formal (or behavioral informal) preferences having aggregated coefficients of relative risk aversion that are typically very high, time-varying, and correlated with the real economy. Some valuable insights have come out of these recent models, but it still seems fair to say that no new consensus has yet emerged from within the economics profession as a whole that the puzzles have been satisfactorily resolved.

The point of departure for this paper is to note that, throughout the expository literature, the equity premium and the risk-free rate are typically calibrated by plugging into the relevant formulas the sample mean and sample variance of past growth rates. But strictly speaking, the correct procedure requires inserting the full subjective probability distributions of uncertain structural parameters of the model, not just their point estimates. Missing from the standard framework is a formal incorporation of the decision-theoretic specification required to make a rigorous statistical-economic general-equilibrium growth model. In effect, the implicit statistical methodology assumes that the time series are long enough that the law of large numbers allows substituting the sample moments of past growth rates for the population moments of future growth rates. This intuitive methodology may well be justified for some economic applications. But the paper will show that such point calibration is a fatally flawed procedure for the particular application of analyzing aversion to model-structure uncertainty, which underlies (or, more accurately, *should* underlie) all asset-pricing calculations. The core problem is that calibrating population moments to sample frequencies can understate enormously the researcher's (and the investor's) predictive uncertainty about the marginal-utility-weighted stochastic discount factor. This forecast-expectation bias spills over into severe pricing-kernel errors, which in turn cascade into the dramatically incorrect asset valuations that finally culminate in the equity family of 'puzzles.'

This paper attempts to shed light on the equity-premium, risk-free-rate, and excess-volatility puzzles by rooting all three issues together deeply into the common ground of Bayesian statistical inference. The fact that structural parameters of the model are unknown introduces an extra layer of Bayesian posterior background uncertainty, which is inherited from the prior and which, counter-intuitively, does not converge uniformly to zero as the number of subsequent observations increases to infinity. Such ubiquitous background uncertainty fattens critically the tails of the posterior distribution of future growth rates and acts strongly upon asset prices to increase significantly the values of both the equity premium and excess volatility, while simultaneously decreasing markedly the risk-free interest rate.

To convey the essential statistical insights as sharply as possible, the simplest imaginable specification of the interplay between Bayesian statistical inference and stochastic general equilibrium growth is modeled. Thus, to ease the computational burdens from delivering its basic message the model analyzes a very stark competitive equilibrium over just two time periods, with a single representative agent whose utility function is isoelastic, for a pure endowment-exchange economy (no genuine production or investment), where the return to representative equity equals the growth of aggregate consumption and both are *i.i.d.* normal, and so on. Abstracted away are such diversionary complications as leverage effects, labor income, distortionary taxes, and the like. For analytical clarity the *only* change made from the simplest stochastic specification (wherein *all* parameter values are assumed known) is to have the model include a consistent Bayesian treatment of *just two* of its structural parameters: the mean and the variance of the normally distributed future growth rate.

In this model the prior probability density of growth rates is essentially characterized by a single critical positive number δ, whose inverse $1/\delta$ quantifies the amount of background uncertainty that later shows up in the Bayesian posterior distribution. As the modeler (or the representative investor) decreases this δ-coefficient continuously (which amounts to moving from a normal distribution of future growth rates towards a fatter-tailed t distribution), the equity premium and excess volatility both increase without limit while the risk-free rate simultaneously decreases, also without limit. Furthermore, the *same* numerical value δ^* simultaneously generates, almost exactly, the equity premium, risk-free rate, and excess volatility that are observed in the time-series data. Although the formal model employs only familiar, analytically tractable, garden-variety specifications in order to be able to derive a relatively transparent expression for the family of equity discrepancies, it will become apparent that the basic insights have much broader applicability.

This paper is far from being the first to investigate the effects of Bayesian statistical uncertainty on asset pricing. Earlier examples having some Bayesian features include Barsky and DeLong (1993), Timmerman (1993), Bossaerts (1995), Cecchetti, Lam and Mark (2000), Veronesi (2000), Brennan and Xia (2001), Abel (2002), Brav and Heaton (2002), Lewellen and Shanken (2002), and several others. Broadly speaking, these papers indicate or hint, either explicitly or implicitly, that the need for Bayesian learning about structural parameters tends to reduce the degree of one or another equity anomaly. What has been utterly missing from this literature, however, is any sense of the *strong force* that tail-fattening structural parameter uncertainty brings to bear on asset pricing equations by its overwhelming ability to dominate the numerical evaluation of standard theoretical formulas involving stochastic discount factors. In effect, the direction of this Bayesian component of structural model uncertainty is (somewhat) appreciated in (some of) the literature, but not the stunning magnitude of the "strong force" it exerts on expected-marginal-utility calculations.

The single possible exception is an important note by Geweke (2001), who applies a Bayesian framework to the most standard model prototypically used to analyze behavior towards risk and then demonstrates the extraordinary fragility of the existence of finite expected utility itself.[1] In a sense the present paper begins by accepting this non-robustness insight, but pushes it further to argue that the inherent sensitivity of the standard prototype formulation constitutes an important clue for unraveling what may be causing the equity puzzles and for giving them a unified general-equilibrium interpretation that simultaneously fits the stylized time-series facts.

This paper will end up arguing that there are no equity 'puzzles' as such arising from within a Bayesian framework. Instead, the arrow of causality in a unified Bayesian explanation is reversed: the 'puzzling' numbers being observed empirically are trying to tell us something very important about the implicit background prior distribution of structural model-parameter uncertainty that is generating such data. In the final section of the paper the three 'puzzling' time-series sample averages of the equity premium, risk-free rate and ex-

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

cess volatility are inverted to back out the implicit subjective probability distribution of the future growth rate. Measured in the appropriate welfare-equivalent state space of expected utility, a world view about the subjective uncertainty of future growth prospects emerges from this Bayesian calibration exercise that is much closer 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 Family of Equity Puzzles

To cut sharply to the analytical essence of the equity macro-puzzles, a super-stark model is used here. This prototype model is a pruned-down version of the textbook workhorse formulation used throughout the financial economics literature. Here everything else except the core structure will be set aside. Essentially, it is fair to say that the models used in this literature are generically isomorphic to the ultra-bare reduced form model presented here.[2]

In this core prototype model, there are two periods, the present and the future. (Mathematically, the first-order conditions from a multi-period version must decompose anyway into an overlapping sequence of first-order conditions from the two-period model, with the details being inessential to the main message of this paper while the resulting clutter of notation is distracting.) The population consists of a large fixed number of identical people. Present per-capita consumption is given as C_0, while future per-capita consumption is the random variable C_1. The utility U of consumption C is specified by the Von Neumann-Morgenstern utility function $U(C)$. The pure-time-preference multiplicative factor for discounting future utility into present utility is β.

Future consumption C_1 is a random variable having some present subjective probability distribution, but whose future realization is presently unknown. The growth rate of everything in this simple endowment-exchange economy is the random variable

$$g \equiv \ln C_1 - \ln C_0, \tag{1}$$

while the expected growth rate is calculated as $E[g] = E[\ln C_1] - \ln C_0$.

The primitive driving force throughout this model is the unknown growth rate g, which quantifies the future state of the system and serves as its sole "fundamental." A one-period

[2]See, for example, the survey articles of Campbell (2003) or Mehra and Prescott (2003), both of which also give due historical credit to the pioneering originators of the important set of ideas 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 Duffie (2001), Cochrane (2001), or Gollier (2001).

asset α is viewed abstractly here as being a security promising to pay a contingent claim in future state g equal to the gross payoff $\Pi_\alpha(g)$, denominated in units of consumption. The expression $\Pi_\alpha(g)$ symbolizes the unit payoff function for asset α. Let the price of this asset be P_α. Then the corresponding asset return function is

$$R_\alpha(g) \;=\; \frac{\Pi_\alpha(g)}{P_\alpha}. \tag{2}$$

Within this model all asset markets are in some sense phantom entities, because no one actually ends up taking a net position in any of them. They exist as shadow exchange possibilities, but in this pure endowment economy there is no avoiding the ultimate reality that everyone's future consumption will end up being the future endowment, no matter how the asset markets equilibrate. The fundamental Euler equation of asset-pricing equilibrium for this economy is

$$U'(C_0) \;=\; \beta E[U'(C_1)R_\alpha(g)]. \tag{3}$$

For practical purposes of analysis, throughout the paper expressions like (3) will be further simplified by following the literature in choosing the utility function to be of the standard iso-elastic form

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma} \tag{4}$$

with corresponding marginal utility

$$U'(C) = C^{-\gamma}, \tag{5}$$

where the coefficient of relative risk aversion is the positive constant γ. Substituting (1), (2), and (5) into (3) and rearranging terms gives the expression

$$P_\alpha \;=\; \beta E[\exp(-\gamma g)\Pi_\alpha(g)]. \tag{6}$$

Plugging (6) into (2) and taking expected values results in the basic relationship

$$E[R_\alpha(g)] \;=\; \frac{E[\Pi_\alpha(g)]}{\beta E[\exp(-\gamma g)\Pi_\alpha(g)]}. \tag{7}$$

An immediate application of (7) is to derive the risk-free interest rate. For this situation we use the standard notation $\alpha = f$ to indicate that we are treating the special case of a deterministic asset paying one unit of consumption in the future. The corresponding unit payoff function here is

$$\Pi_f(g) \;=\; 1, \tag{8}$$

for which case (7) becomes

$$R_f \; = \; \frac{1}{\beta E[\exp(-\gamma g)]}. \tag{9}$$

Another important application of formula (7) is to the special case of a comprehensive broad-based equity index representing the entire economy. Throughout this paper, the abstraction is made that the one-period-ahead payoff to comprehensive (or representative) equity is essentially a unit claim on the future aggregate output of the underlying real economy. In this case we use the standard notation $\alpha = e$ to indicate that we are treating the situation of economy-wide one-period equity. The corresponding unit payoff function here is

$$\Pi_e(g) \; = \; \exp(g), \tag{10}$$

for which case equation (7) becomes

$$E[R_e] \; = \; \frac{E[\exp(g)]}{\beta E[\exp((1-\gamma)g)]}. \tag{11}$$

Dividing (11) by (9), the equity premium in ratio form is

$$\frac{E[R_e]}{R_f} \; = \; \frac{E[\exp(g)] \; E[\exp(-\gamma g)]}{E[\exp((1-\gamma)g)]}. \tag{12}$$

The meaning given in the literature to (12) goes along the following lines. Interpret the left hand side of the equation as the *actual* risk premium ratio that is observed historically in the real world. Interpret the right hand side of (12) as a theoretical *formula* for calculating this risk premium, given any coefficient of relative risk aversion γ, and, more importantly here, given the subjective probability distribution of the uncertain future growth rate g.

Concerning the relative-risk-aversion parameter γ, there seems to be some 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 accurately stated, any proposed solution which does not explain the equity premium for $\gamma \leq 3$ 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, there is much less consensus 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 known, stable, stationary stochastic specification that can accurately be extrapolated from the past onto the future), we cannot know the critical structural parameters of the distribution of g unless there is an infinitely long time series of historical growth rates.

At this point in the story, the best anyone can do is to *infer* from the past some *estimate* of the probability distribution of g . The rest of the story hinges on specifying the form of the assumed density function of g, 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 probability density function

$$g \sim N(\mu, V), \tag{13}$$

which is the ubiquitous benchmark case assumed throughout the asset-pricing literature. The only difference in this paper is that μ and V will be treated here as unknown parameter values that must be estimated statistically from past data.

When (13) is assumed but μ and V are viewed as random variables, then using the formula for the expectation of a lognormal distribution, cancelling the many redundant terms, and taking logarithms transforms the theoretical equity premium formula (12) into the formula

$$\ln E[R_e] - r_f \; = \; \ln E[\exp(\gamma V)], \tag{14}$$

where the usual definition $r_f \equiv \ln R_f$ is applied and the expectation operator on the right hand side of equation (14) is understood here as being taken over V.

From the exponential function on the right hand side of equation (14) being convex in V, a mean-preserving spread of V increases the theoretically-calculated equity premium. Therefore, plugging the mean of V into (14) biases this formula in the direction that *point calibration always under-predicts the equity premium.* The equity-premium literature generally proceeds from (14) by ignoring the bias-producing uncertainty inherent in using point estimates of V. Instead, the usual practice calibrates (14) by plugging in the sample variance from n previous observations on growth rates, and then proceeds as if normality still holds (in place of substituting into (14), e.g., the relevant inverted-gamma distribution to account for the sampling error from estimating the unknown structural parameter V).

The observed sample variance is

$$\widehat{V} \; = \; \frac{1}{n}\sum_{i=1}^{n}(g_i - \widehat{g})^2, \tag{15}$$

where

$$\widehat{g} \; = \; \frac{1}{n}\sum_{i=1}^{n} g_i \tag{16}$$

is the sample mean. Implicitly in the equity premium literature, the sample size n is presumed large enough to make (16) and (15) sufficiently accurate estimates of their underlying true values, but no formal attempt is made to define "sufficiently accurate" or to confirm exactly what happens to formula (14) in this model if the estimates, and therefore the approximations, are not "sufficiently accurate." In the expository literature the value of (14) is calculated to be what it reduces to when there is no structural uncertainty and V is known exactly to be equal to $\widehat{V}$. The as-if-deterministic $V = \widehat{V}$ point-mass version of the theoretical formula (14) is the familiar equation from the literature

$$\ln E[R_e] - r_f \;=\; \gamma \widehat{V}, \tag{17}$$

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

Taking 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 $E[R_e] \approx 7\%$.[3] The historically observed real return on an index of the safest available short-maturity bills is about 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 standard deviation taken to be about 2%, meaning $\widehat{V} \approx 0.04\%$. Suppose $\gamma \approx 2$. Plugging these values into (17) gives a calculated value $\gamma \widehat{V} \approx 0.08\%$.

Thus, the actually observed equity premium on the left hand side of equation (14) exceeds the estimate (17) 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 present model, and it is apparent why characterizing such a result as "disturbing" (for the standard neoclassical paradigm) may be putting it very mildly. Plugging in some reasonable alternative 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 equity 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*.

Turning to the risk-free rate puzzle, the meaning given in the asset-pricing literature to equation (9) parallels the interpretation given to the equity premium formula. Interpret the

[3] These numbers are from Prescott and Mehra (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 $E[R_e]$ than "America in the American century").

left hand side of equation (9) as the actual risk-free interest rate that is observed historically in the real world, while the right hand side represents a theoretical formula for calculating this risk-free interest rate, given γ and the subjective probability distribution of the future growth rate g. Concerning the behavioral risk-aversion parameter γ, a value that would be accepted by the economics profession as a whole is about two, roughly. By contrast, the true subjective probability distribution of the future growth rate g is unknown and the best that can be done is to make some statistical inference about the likely probabilities of g from observing its past realizations.

When the normality assumption (13) is made but μ and V are viewed as random variables, then using the formula for the expectation of a lognormal distribution, cancelling redundant terms, and taking logarithms transforms the theoretical risk-free interest rate formula (9) into

$$r_f \;=\; \rho + \gamma E[\mu] - \ln E[\exp(\gamma^2 V/2)], \tag{18}$$

where

$$\rho \;\equiv\; -\ln \beta \tag{19}$$

is the instantaneous rate of pure time preference and the expectations on the right hand side of equation (18) are understood as being taken over μ and V.

From the exponential function on the right hand side of equation (18) being convex in V, a mean-preserving spread of V decreases the theoretically-calculated risk-free interest rate. Therefore, plugging the mean of V into (18) biases this formula in the direction that *point calibration always over-predicts the risk-free rate.*

The literature typically proceeds from (18) by ignoring the statistical uncertainty inherent in estimating μ and V. Instead, these two structural parameters are usually treated by plugging into (18) their sample values and then pretending that normality still holds. Substituting the sample mean $\widehat{g}$ and the sample variance $\widehat{V}$ into (18) gives

$$r_f \;=\; \rho + \gamma \widehat{g} - \frac{1}{2}\gamma^2 \widehat{V}, \tag{20}$$

which is a ubiquitous generic formula appearing in one form or another throughout stochastic growth theory. (Its origins trace back to the famous neoclassical Ramsey model of the 1920's.)

Non-controversial estimates of the relevant parameters appearing in (20) (calculated on an annual basis) are: $\widehat{g} \approx 2\%$, $\widehat{V} \approx .04\%$, $\rho \approx 2\%$, $\gamma \approx 2$. With these representative parameter values plugged into the right hand side of (20), the left hand side becomes $r_f \approx 5.9\%$. When compared with an actual real-world risk-free rate $\widehat{r}_f \approx 1\%$, the theoretical formula is too high by $\approx 4.9\%$.

This gross discrepancy is the risk-free rate puzzle. With the other base-case parameters set at the above values, the coefficient of relative risk aversion required to explain the risk-free interest rate discrepancy is essentially negative, while the coefficient of relative risk aversion required to explain the equity-premium discrepancy estimated from (17) is $\gamma = 150$. The simultaneous existence of two strong contradictions with reality, which, in addition, are strongly contradicting each other, might be characterized as being disturbing "times three."

As if all of the above were not vexing enough, there is also the enigmatic appearance in the data of an "excess volatility puzzle" (of a form corresponding to this stark macro-model). In the ultra-simple *i.i.d.* endowment-exchange economy of this paper, which is a modification of the seminal Mehra and Prescott (1986) model (itself a modification of the seminal Lucas (1978) fruit-tree model), there is no genuine distinction between: 1: consumption; 2: output; 3: endowments; 4: dividends; 5: payoffs. All five "fundamentals" here have the same continuously-compounded growth rate, given by the random variable g, which therefore represents the sole genuine "fundamental" of the system. From combining (2) with (10), the continuously-compounded return on one-period equity r_e should have the same variability as its underlying "fundamental," which for this austere representation means g. In the data, however, 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 $\widehat{\sigma}[g] \approx 2\%$. For the super-sparse fruit-tree model of this paper, I take some slight literary license by interpreting the relevant macroeconomic form of an "excess volatility puzzle" to be the empirical fact that $\widehat{\sigma}[r_e] \gg \widehat{\sigma}[g]$. Thus, the "excess volatility puzzle" is taken here to be the stylized fact that the continuously-compounded one-period return to comprehensive equity counterintuitively appears to be about an order of magnitude more volatile than the underlying fundamental of an aggregate-output growth payoff, from which it is conceptually constituted.

Summing up the scorecard for the standard neoclassical model, all in all we have three strong contradictions with reality and at least one serious internal contradiction, making the grand total add up to being a conundrum that is disturbing times four. It was previously noted that uncertainty in V has the qualitative effect of diminishing simultaneously the magnitude of both the equity-premium and risk-free-rate discrepancies. We next examine what happens quantitatively to the family of equity puzzles when the structural parameters μ and V take on the standard familiar sampling distributions that arise naturally when n sample points are drawn randomly from a normal population.

3 Subjective Expectations of Future Growth

As a preliminary guide to indicate roughly where the argument is now and where it is heading, the outline of the ultimate full model is here sketched. The Euler equation (3) is presumed to hold *in subjective expectations* for the utility function (4). The assumed probability distributions are: $g \sim N(E[g], V[g])$ and $r_e \sim N(E[r_e], V[r_e])$. The following five quasi-constants of the model are effectively assumed known: $E[r_e]$, $V[r_e]$, r_f, ρ, γ. The following two structural parameters are unknown and must be estimated statistically: $E[g]$, $V[g]$. This section derives the Bayesian subjective distributions of $E[g]$ and $V[g]$, which will then be applied to the model throughout the remainder of the paper.

Assuming the normal specification (13), define the random variable

$$w \equiv 1/V, \tag{21}$$

which is commonly called the *precision* of a normal probability distribution. Let the random variable $\widetilde{\mu}$ denote the unknown mean of g. Conditional on known values of w and $\widetilde{\mu}$,

$$g = \widetilde{\mu} + \varepsilon, \tag{22}$$

where $\varepsilon \sim N(0, 1/w)$.

Purely for simplicity here suppose that initially, before any observations are made, the Bayesian pre-sample estimate of the random variable $\widetilde{\mu}$ is distributed as a non-informative diffuse prior. Let $g_1, \ldots, g_n$ be a random *i.i.d.* sample corresponding to the normal probability structure (22), which is drawn from a normal distribution with *known* precision w, but whose Bayesian pre-sample estimate of $\widetilde{\mu}$ is a diffuse-normal distribution. With a known variance, the posterior distribution of $\widetilde{\mu}$ after n independent sample observations is

$$\widetilde{\mu} \sim N(\widehat{g}, 1/nw). \tag{23}$$

From (22) and (23), $E[g] = E[\widetilde{\mu}] = \widehat{g}$. For *given* values of *both* w *and* $\widetilde{\mu}$, the random variable g is distributed according to (22) as

$$g \sim N(\widetilde{\mu}, 1/w), \tag{24}$$

while, for any given value of w *alone*, the random variable $\widetilde{\mu}$ is distributed according to (23). Combining these two quasi-independent realizations of normal processes, the random variable g must be distributed normally with mean $\widehat{g}$ and variance equal to the sum of the variance of the normal process (23) plus the variance of the conditionally-independent normal process (24). After adding together the two variances ($1/nw$ from (23) plus $1/w$ from (24)), the posterior distribution of g comes out to be

$$g \sim N(\widehat{g}, (n+1)/nw). \tag{25}$$

Thus far, the specification has proceeded as if w were known. When w is not known, Bayesian statistical theory has developed a rigorous and elegantly symmetric isomorphism with the classical statistics of the familiar linear-normal regression setup.[4] The Bayesian dual counterpart to classical statistics works with a normal-gamma family of so-called "natural conjugate" distributions. For reasons that later will become apparent, we work here with a three-parameter generalization of the two-parameter gamma distribution, which forms a *normal-truncated-gamma* family of natural conjugate distributions.

Consider a non-negative random variable w representing the precision. Let δ be a non-negative parameter representing an arbitrarily imposed lower-bound support for the Bayesian prior distribution of w. For the time being, think of δ as an arbitrarily-small positive number. Assume that the Bayesian prior distribution of the precision is a truncated-gamma probability density function (with truncation parameter δ) of the form:

$$\varphi_0(w) \ = \ \frac{w^{a_0-1} \, e^{-b_0 w}}{\int_\delta^\infty w^{a_0-1} \, e^{-b_0 w} \, dw} \tag{26}$$

for $w \geq \delta$,while $\varphi_0(w) = 0$ for $w < \delta$.

When choosing δ to be positive, the model is dogmatically eliminating a priori all variances above $1/\delta$. The technical reason for declaring impermissible worlds of unboundedly high variance is to make the integral defining the moment generating function of g converge to a finite value. An economic rationale presumably has to do with the difficulty of envisioning the unbounded loss function arising from possibly unlimited variability in growth rates. Whatever the story, the subliminal accompanying message is that nobody has the slightest idea about what is actually an appropriate value of δ, which exists in the first place only to put *some* finite upper bound above expected marginal utility.

The three non-negative parameters δ, a_0, b_0 of the truncated gamma distribution (26) represent prior beliefs about the precision. In the limit as $\delta \to 0^+$, the mean of the truncated-gamma prior (26) approaches a_0/b_0, while the variance approaches $a_0/{b_0}^2$. Thus, at least for small δ, the prior mean and variance can be assigned any values just by judiciously selecting a_0 and b_0. Classical statistical analysis is exactly isomorphic to the limiting case of a non-

[4] Among several other places, clear expositions of Bayesian-classical duality are contained in DeGroot (1970), Zellner (1971), Leamer (1978), Hamilton (1994), Poirier (1995), and Koop (2003).

informative diffuse prior: $\delta \to 0^+$, $a_0 \to 0^+$, $b_0 \to 0^+$. Therefore, the analysis presented here can be viewed as paralleling the classical specification very closely, except that it is slightly more general by allowing positive parameter values other than the limit 0^+.

Let $\varphi_n(w)$ be the *posterior* distribution of the precision w at a time just after observing the n independent realizations $g_1, \ldots, g_n$. When $\delta = 0$, it is well known (see any of the references cited in footnote 4) that the normal-gamma distribution constitutes a conjugate family of priors. When $\delta > 0$, we have the same conjugate family of priors, except that w is subject to a lower-bound constraint. Therefore, the posterior is of the same form as the prior, and subject to the same bounding constraint. The modification of a basic conjugate-prior result in the Bayesian statistical literature needed for this paper is the following **lemma** (stated here without proof):

$$\varphi_n(w) \; = \; \frac{w^{a-1}\, e^{-bw}}{\int_\delta^\infty w^{a-1}\, e^{-bw}\, dw}, \tag{27}$$

for $w \geq \delta$,while $\varphi_n(w) = 0$ for $w < \delta$. The parameters a and b are defined by the equations

$$a \; = \; \frac{n}{2} + a_0 \tag{28}$$

and

$$b \; = \; \frac{1}{2}\sum_{i=1}^{n}(g_i - \widehat{g})^2 + b_0. \tag{29}$$

It is analytically very convenient (and, in the context of this model, comes at the cost of only an insignificant loss of generality) to presume that there is insufficient reason to believe that the distribution of the imposed prior precision is biased relative to the observed sample precision, meaning $E_0[w] \; \approx \; 1/\widehat{V}$. Since (for the small-$\delta$ case of a truncated-gamma prior distribution) $E_0[w] \approx a_0/b_0$, this unbiasedness assumption allows compressing the two parameters a_0 and b_0 into just one parameter, by effectively imposing the additional conditions

$$a_0 = \frac{m}{2} \tag{30}$$

and

$$b_0 = \frac{m}{2}\widehat{V}, \tag{31}$$

where the single parameter m now quantifies the one remaining degree of freedom. With the above unbiasedness specification, m has a natural interpretation "as if" $\widehat{V}$ were the sample variance calculated from a pre-observation fictitious earlier sample of size m drawn from the same underlying population that generated the data. Under this interpretation, m quantifies the "degree of prior confidence" in the value $\widehat{V}$ (of V), which was in fact calculated from the n "real" sample points that were actually observed. The overall situation is then "as if" $\widehat{V}$ were the sample variance from a total sample of size $m + n$. With this unbiasedness simplification, the prior distribution of the precision is now characterized by just two non-negative parameters: δ and m.

To summarize here, with the unbiased-prior assumption (30), (31) the posterior distribution of the precision w is given by the probability density function

$$\psi_n(w \mid \delta, m) \;=\; k_\delta \, w^{a-1} \, e^{-bw}, \tag{32}$$

for $w \geq \delta$, while $\psi_n(w \mid \delta, m) = 0$ for $w < \delta$. The constants k_δ, a and b in (32) are defined by the equations

$$1/k_\delta \;=\; \int_\delta^\infty w^{a-1} \, e^{-bw} \, dw, \tag{33}$$

$$a \;=\; \frac{n+m}{2}, \tag{34}$$

$$b \;=\; \frac{n+m}{2} \widehat{V}. \tag{35}$$

From combining (32) with (25), the unconditional or marginal posterior probability density function of g is

$$f_n(g \mid \delta, m) \;=\; k'_\delta \int_\delta^\infty \exp(-(g-\widehat{g})^2 nw/2(n+1)) \, w^{a-\frac{1}{2}} \, e^{-bw} \, dw, \tag{36}$$

where a and b are defined by (34) and (35), while k'_δ is a constant of integration satisfying

$$1/k'_\delta \;=\; \int_{-\infty}^\infty dx \int_\delta^\infty \exp(-x^2 nw/2(n+1)) \, w^{a-\frac{1}{2}} \, e^{-bw} \, dw. \tag{37}$$

The two non-negative parameters δ and m are highlighted in formulas (32) and (36) merely to remind us that (among many other things, such as $\widehat{V}$ and n, which offhand seem like they should end up being far more important in practice) the posterior probability density function of the future growth rate depends in principle on the lower bound δ and the fictitious-sample size m that are conceptualized by us today as characterizing the prior distribution of the precision prescribed ("by them") at a time n periods ago. Of course nobody today has the slightest notion about what reasonable values of δ or m might have been way back then, before anyone looked at any data. It is for just this reason that everyone's favorite candidate today is the non-informative diffuse prior $\delta \to 0^+$ and $m \to 0^+$, which corresponds exactly to familiar dual-classical statistical regression analysis. In this dual-classical case, straightforward integration (after taking the limits $\delta \to 0^+$ and $m \to 0^+$) shows that (36), (37) reduces to the t distribution

$$f_n(g \mid 0,0) \;=\; \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\,\sqrt{\pi \widehat{V} n}} \left[1 + \frac{(g-\widehat{g})^2}{(n+1)\widehat{V}}\right]^{-\frac{n+1}{2}}, \tag{38}$$

whose moment generating function is unboundedly large because the relevant integral diverges to plus infinity. (It is essentially in order to make this moment-generating integral converge that the condition $\delta > 0$ is imposed in the first place.)

Of course, this entire preliminary discussion of the future consequences of what people now think that people long ago "might have been thinking" about an upper bound on V (of $1/\delta$) or a degree of prior confidence in $\widehat{V}$ (of m) has an unreal tone about it. In practice such issues ought to be non-operational – and therefore not worth contemplating – because the intervening n observations should have effectively bleached the prior parameters out of the posterior distribution. Thus, if the number of data points n is large enough, "it should not matter now what values we select to represent prior beliefs" about δ or m.

The formula for expected utility here essentially reflects the mathematical properties of the moment generating function of g. The intuition that "it should not matter now what values we select to represent prior beliefs" is true, it turns out, for the parameter m, whose effects on expected utility *converge uniformly in n for all $m > 0$*. However, the parameter δ behaves fundamentally differently, because its effects on the moment generating function of g – and therefore on expected utility – *do not converge uniformly in n for all $\delta > 0$*. In this sense there is a critical distinction, which is crucial for all expected marginal-utility-weighted asset-pricing implications, between not knowing what value to assign now to the prior parameter m and not knowing what value to assign now to the prior parameter δ.

The fact that expected utility is not uniformly convergent in n for all positive δ has great significance for the interpretation of this paper. A prior distribution is *our imputation now* of what "they might have imposed" n periods ago during the pre-data past. It is essentially a mental artifice for framing a subjective thought-experimental back-and-forth dialogue between the present and the past about what to expect from the future. In such a setting, *pointwise* convergence of expected utility in n for a given δ is not nearly enough to guarantee a robust prior, because the prior is a subjective creature of *our imagination now*, not an objective unchangeable reality that a real person carved in stone n periods ago to represent some intrinsic characteristic of the then-observable world.

To have faith in the standard practice of calibrating means and variances of normal distributions to past historical averages presupposes a robustness in the interpretation of observable data with respect to whatever values of δ or m are chosen. Therefore, a necessary precondition for the validity here of the classical statistical idea to "just let the data speak for themselves" is that the effects of δ or m should be negligible for sufficiently large n. This condition holds (in the space of subjectively-expected utility) for m, but such a robustness condition *does not hold* (in the space of subjectively-expected utility) *for* δ. The value of δ that has been chosen now to represent the past manifests itself as a piece of current background risk that refuses to go away with the passage of time. From a Bayesian perspective, we "just let the data speak for themselves" in a different sense from the classical statistical interpretation of this phrase. Here, data "speak for themselves" by telling us what is the implied value of δ that real-world investors must *implicitly* be using in their priors, in order to be compatible with what we researchers are observing in the data.

To summarize, within the Bayesian framework appropriate for thinking about basic issues of risk aversion and asset pricing (which underlies the entire family of equity 'puzzles'), the subjective element involved in choosing a prior distribution of structural parameters cannot be disentangled from the calibration process. Non-uniform convergence in stochastic-discount-factor space means that the fickle whimsicality of current investors concerning what value of the structural parameter δ to select for representing the model's initial configuration never loses its critical impact on subsequent behavior under risk, regardless of the amount of data accumulated during the interim. This tremendous sensitivity to the "background shadow of δ" permeates every aspect of asset pricing and represents the critical component of a unified Bayesian theory capable of resolving simultaneously all three of the so-called equity puzzles.

Taking (32) and (36) as our subjective posterior probability density functions, we are now ready to compute the Bayesian equity premium, the Bayesian risk-free interest rate, and Bayesian excess volatility. The next three sections of the paper do these calculations, in turn. In the last section of the paper, implicit parameter values of the subjective probability distribution of future growth rates are backed out of the data by Bayesian inverse calibration. For each application, the sharpest insight comes from having in mind the mental image of a limiting situation where *m is very big, while simultaneously δ is very small.* This limiting situation comes arbitrarily close to the standard familiar textbook case of normally distributed purely-stochastic growth-rate risk, but, as we shall see, it never quite gets there. When m is "very big" (but less than infinity), the subjective Bayesian distribution of future growth rates is essentially unchanged by the arrival of a new datum point. Such a limiting situation nullifies sampling error and focuses the mind sharply on understanding the core Bayesian structural model-uncertainty mechanism driving the entire family of equity 'puzzles.'

4 The Bayesian Equity Premium

We now use the statistical apparatus developed in the last section of the paper to compute the Bayesian equity premium. For fixed m and n, let $\Psi(\delta)$ represent the value of $\ln E[R_e]-r_f$ as a function of δ that is obtained from formula (14) with $V = 1/w$ when the probability density function of w is $\psi_n(w \mid \delta, m)$ defined by equation (32). Plugging the subjective posterior distribution (32) into formula (14) when $V = 1/w$, we obtain

$$\Psi(\delta) \; = \; \ln \int_{\delta}^{\infty} \exp(\gamma/w) \; k_\delta \; w^{a-\frac{1}{2}} \, e^{-bw} \, dw. \tag{39}$$

We then have the following proposition.

Theorem 1 *Suppose that $\gamma > 0$ and $m + n < \infty$. Let $\ln E[R_e] - r_f$ be any given positive value of the equity premium. Then there exists some positive δ_e such that*

$$\ln E[R_e] - r_f \; = \; \Psi(\delta_e). \tag{40}$$

Proof. It is readily apparent from examining (39) that $\Psi(0) = +\infty$. At the other extreme of δ, it is apparent that $\Psi(\infty) = 0^+$, because there is no equity premium when there is no uncertainty. The function $\Psi(\delta)$ defined by (39) is continuous in δ. Since

$$\Psi(\infty) \; < \; \ln E[R_e] - r_f \; < \; \Psi(0), \tag{41}$$

condition (40) follows. ■

The essence of the Bayesian statistical mechanism driving the theorem can be intuited by examining what happens in the limiting case. As $\delta \rightarrow 0^+$, the limit of (36) is a t distribution of the form (38) – except that $m+n$ essentially replaces n. With the presumed case of large $m+n$ and small δ, the central part of the t-like distribution (36) is approximated well by a normal curve with mean $\widehat{g}$ and standard deviation $\widehat{\sigma}[g]$ fitting the data in its middle range. However, for applications involving the implications of risk aversion, such as calculating the equity premium, to ignore what is happening away from the middle of the distribution has the potential of wreaking havoc on the calculations. For these applications, such a normal distribution may be a very bad approximation indeed, because the relatively fatter tail of the dampened-t distribution (36) is capable of producing an explosion in formulas like (9) or (12), implying in the limit as $\delta \to 0^+$ an unboundedly large equity premium. In this framework, therefore, the statistical fact that the moment generating function of a t-distribution is infinite has the extremely important economic interpretation that model-structure uncertainty here is potentially far more important for asset pricing than purely-stochastic risk. As $\delta \to 0^+$, the representative agent becomes incomparably more averse to the "strong force" of statistical uncertainty about the future growth process (with *unknown* structural parameters) 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 (but with *known* structural parameters).

An explosion of the equity premium does not happen in the real world, of course, but a tamed near-explosive outcome remains the mathematical driving force behind the scene, which imparts the statistical illusion of an enormous equity premium incompatible with the standard neoclassical paradigm. When people are peering into the future they are also peering into the past, and they are intuitively sensing there the spooky background presence of a low-δ prior volatility that could leave them holding the bag by wiping out their stock-market investments. This eerie sensation of low-δ background structural shadow-risk cannot easily be articulated, yet it frightens investors away from taking a more aggressive stance in equities and scares them into a position of wanting to hold instead a portfolio of some safer stores of value, such as cash, inventories of real goods, precious metals, or government treasury bills – as a hedge against low-consumption states. Consequently, these relatively-safe assets bear very low, even negative, rates of return.

I do not believe that it will be easy to dismiss such type of Bayesian statistical explanation. 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. But *compared with what* is the observed value of $\ln E[R_e] - r_f$ "too big"? The answer given in the equity-premium literature is: "compared with the right hand side of formula (17)." Unfortunately for this logic, the point-calibrated right hand side of (17) is in practice a terrible estimate of the true value of $\ln E[R_e] - r_f$ as given by equation (14). Anyone wishing to downplay this line of reasoning in favor of the *status quo ante* would be hard pressed to come up with their own Bayesian rationale for calibrating variances of non-observable subjectively-distributed future growth rates by point estimates equal to past sample averages. In essence, the frequentist-inspired approach that produces the family of equity puzzles avoids the consequences on marginal-utility-weighted asset-pricing kernels of non-uniform convergence in δ only by effectively imposing from the very beginning the pointwise-convergent extreme case $m = \infty$.

In an important early attempt to explain the equity premium puzzle, Rietz (1988) argued that we cannot exclude the possibility that our sample size is not large enough to describe adequately the full macro-risk of unknown future growth rates. The impact on financial equilibrium of a situation where there is a tiny probability of a catastrophic out-of-sample event has been dubbed the "peso problem." In a peso problem, the small probability of a disastrous future happening (such as a collapse of the presumed structure from a natural or socio-economic catastrophe) is taken into account by real-world investors (in the form of a "peso premium") but not by the model, because such an event is not in the sample.

I think Theorem 1 is trying to tell us that the statistical architecture of a peso problem is genetically hardwired into the "deep structure" of how Bayesian inferences about exponential processes (of future economic growth, at unknown rates) interact with a curved utility function. Bayesian inferences from finite data fatten the posterior tails of probability density functions with dramatic consequences when expressed in 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 it may not be so absurd to believe that *no* finite sample size is large enough to capture all of the relevant structural model uncertainty concerning future economic growth. I think the Bayesian peso problem is trying to tell us that to calibrate an exponential process having an unknown growth rate, which is essentially intended to describe future worldwide economic prospects, by plugging the sample variance of observed past growth rates into a "very bad" approximation of the subjectively-distributed stochastic discount factor, is to underestimate "very badly" the comparative risk of a real world gamble on the state of the future world's economy relative to a safe investment in a near-money sure thing.

Of course, what is being presented here is just one illustrative example of the economic consequences of such a 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 effects of Bayesian tail-fattening will cause the equity premium to be highly sensitive to seemingly innocuous and negligible changes in the assumed prior of the precision – within a very broad class of non-dogmatic probability distributions obeying standard regularity conditions. The driving statistical-economic force is that seemingly thin-tailed probability distributions, which actually are only thin-tailed *conditional on known structural parameters* of the model, become thick-tailed after integrating out the parameter uncertainty. Intuitively, no finite sample can eliminate the possibility of fat tails, 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 uncertainty. The very important result in Schwarz

(1999) can be interpreted as saying that, for essentially any model whose conclusions are invariant to measurement units, the moment generating function of the posterior distribution is infinite (i.e., the posterior distribution has a "thick" tail), even when the random variable is being drawn from a thin-tailed parent distribution whose moment generating function is finite. People are potentially much more afraid of not knowing what are the structural-parameter settings inside the black box, whose data generating process drives the purely-stochastic growth rates, than they are averse to the purely-stochastic growth-rate risk itself. When investors are modeled as perceiving and acting upon these inevitably-thick-tailed subjective posterior distributions, then a fully-rational general equilibrium interpretation welds together seamlessly a unified Bayesian theory of the entire family of equity 'puzzles,' as the next three sections of the paper will show in turn.

5 The Bayesian Risk-Free Rate

We can use the same mathematical-statistical apparatus to calculate the Bayesian risk-free interest rate. For fixed m and n, let $\Phi(\delta)$ be the value of r_f as a function of δ that comes out of formula (18) when $\widehat{g}$ is substituted for $E[\mu]$ and the probability density function (32) is used to evaluate the distribution of $V = 1/w$. Plugging $\widehat{g}$ and the subjective posterior distribution (32) into the right hand side of equation (18), the result is

$$\Phi(\delta) \ = \ \rho + \gamma\widehat{g} - \ln \int_{\delta}^{\infty} \exp(\gamma^2/2w) \ k_\delta \ w^{a-\frac{1}{2}} \ e^{-bw} \, dw. \tag{42}$$

We then have the following proposition.

Theorem 2 *Suppose* $\gamma > 0$ *and* $m + n < \infty$ *. Let* r_f *be any given value of the risk-free interest rate that satisfies* $r_f < \rho + \gamma\widehat{g}$*. Then there exists a positive* δ_f *such that*

$$r_f = \Phi(\delta_f). \tag{43}$$

Proof. As δ is made to approach zero, the integral on the right hand side of (42) becomes unbounded. Therefore, $\Phi(0) = -\infty$. At the other extreme of δ is the deterministic Ramsey formula $\Phi(\infty) = \rho + \gamma\widehat{g}$. Thus,

$$\Phi(0) < r_f < \Phi(\infty), \tag{44}$$

and, since $\Phi(\delta)$ defined by (42) is continuous in δ, the conclusion (43) follows. ■

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 t statistic in its tails and carries with it a potentially explosive moment generating function reflecting a strong aversion to high-volatility low-precision situations. The bottom line once more is that a "Bayesian peso problem" causes classical-like rational-expectations inferences, which are based upon the observed historical behavior 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 growth-structure of an unknown future economy.

6 Bayesian Excess Volatility

The methodology in this section of the paper unavoidably stretches the mind more than what was previously encountered, because we are forced now to flush out logical inconsistencies and to confront conceptual modeling problems previously swept under the rug. We begin with the standard definition

$$r_e(g) \;\equiv\; \ln R_e(g), \tag{45}$$

and then apply (2) and (10) to obtain

$$r_e - E[r_e] \;\;=\;\; g - E[g]. \tag{46}$$

In the ultra-stark macro-economy being modeled in this paper, the only genuine "fundamental" of the system is the growth rate g. Stocks are being understood here as effectively representing one-period-ahead claims on future $\{\exp(g)\}$. The idea that stock-market returns should reflect underlying "fundamentals" finds expression in this bare-bones macro-model by equation (46). (A further decomposition of stock returns into conventional dividends plus capital gains fills in some interesting and useful supplementary details, but otherwise obscures the intended focus of this paper on the core structural-uncertainty message.) According to (46), for an economy-wide comprehensive stock index, which embodies a claim on the future aggregate output of the underlying real economy, the representative agent should be *subjectively perceiving* the distribution of continuously compounded future growth rates as being equal to the distribution of continuously compounded one-period equity returns. Yet this "perception" appears to be strongly violated in the data. For the super-simple fruit-tree model of this paper, the "excess volatility puzzle" is taken to be the empirical fact that $\widehat{\sigma}[r_e]$ is about an order of magnitude larger than $\widehat{\sigma}[g]$. This section of the paper essentially poses and answers the basic question: *which one of these two volatilities* ($\widehat{\sigma}[r_e]$ representing the left hand side of equation (46), or $\widehat{\sigma}[g]$ representing the right hand side) embodies more appropriately the not-directly-observable subjective volatility of future growth rates?

Because $\widehat{\sigma}[r_e] \gg \widehat{\sigma}[g]$, clearly expression (46) cannot be interpreted as representing equality between random variables in a realized-frequency sense. However, what is required by the relevant theory is not this *ex-post* frequency interpretation per se, but rather that the equality in (46) be *ex-ante subjectively perceived* as holding in expectations. The Euler equation imposes restrictions upon *perceived expectations* of stochastic-discount-factor-weighted equity returns, not upon *past realizations* of equity returns, growth rates, or pricing-kernel errors per se. The whole point of the paper is that g is a subjectively-distributed random variable, whose perceived "strong force" of marginal-utility-weighted subjective future variability is much greater than what might appear to be indicated by point calibration to its un-weighted past sample variability. The representative agent here understands that, due to the "Bayesian peso problem," the realized sample variance understates significantly *in subjective-discount-factor units* the true parent-population variance, which includes the out–of-sample low-g (or high-g) extreme outliers in the tails that have not yet been experienced. Throughout what follows, it is essential to realize that there is no excess volatility 'puzzle' in (46) if the variability of returns on a comprehensive equity index, which is being *subjectively perceived* as representing the entire economy, matches consistently the variability of that same economy's future growth rate, which is being drawn from the full parent-population distribution of subjectively-expected future growth rates.

Let

$$g_N \sim N(E[g_N], V[g_N]) \tag{47}$$

be a random variable representing the *subjective perception* of a normally-distributed future growth rate with known parameters $E[g_N]$ and $V[g_N]$. Let

$$\Pi_N(g_N) \equiv \exp(g_N) \tag{48}$$

be the *subjective perception* of a stock-market payoff representing a unit claim on the log-normally distributed future aggregate output corresponding to (47). Such a unit claim gives rise to the *subjective perception* of a (geometrically measured) return on equity $r_N(g_N)$ satisfying the equation

$$r_N - E[r_N] = g_N - E[g_N], \tag{49}$$

which is exactly the mathematical counterpart here of (46).

The following proposition establishes an important type of observational and welfare equivalence between g_N and g. This last theorem of the paper can be interpreted as providing a rationale for telling an as-if parable wherein the representative agent has a subjective perception, which is consistent with (49), "as if" the future growth rate is g_N with known variability equal to the variability of returns on equity.

Theorem 3 *Suppose $m + n < \infty$. For any given positive value of $\widehat{\sigma}[r_e]$, there exists some positive δ_N such that the following four conditions are simultaneously satisfied:*

$$E[U(C_0 \exp(g_N))] \;=\; E[U(C_0 \exp(g))], \tag{50}$$

$$E[g_N] \;=\; E[g] \;=\; \widehat{g}, \tag{51}$$

$$E[r_N] \;=\; E[r_e], \tag{52}$$

$$\sigma[r_N] \;=\; \sigma[g_N] \;=\; \widehat{\sigma}[r_e]. \tag{53}$$

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

$$\frac{1}{\sqrt{2\pi}S(\delta)} \int_{-\infty}^{\infty} \exp((1-\gamma)g_N - (g_N - \widehat{g})^2/2S(\delta))dg_N = \int_{-\infty}^{\infty} \exp((1-\gamma)g) f_n(g \mid \delta, m)dg \tag{54}$$

and note that for this definition (50) and (51) are satisfied by construction.

It is straightforward to prove, from plugging from (10) into (7), that

$$E[r_e] \;=\; \rho - \ln E[\exp((1-\gamma)g)], \tag{55}$$

and, from plugging (48) into (7), that

$$E[r_N] \;=\; \rho - \ln E[\exp((1-\gamma)g_N)], \tag{56}$$

so that (52) then follows from (54), (55), (56).

For $\delta = \infty$, the integral on the right hand side of equation (54) becomes $\exp((1-\gamma)\widehat{g})$, implying $S(\infty) = 0$. As $\delta \to 0$, the integral on the right hand side of equation (54) becomes unbounded, implying $S(0) = \infty$. Thus,

$$S(\infty) \;<\; \widehat{\sigma}[r_e] \;<\; S(0), \tag{57}$$

and, by continuity of the function $S(\delta)$, there must exist a $\delta_N > 0$ satisfying

$$S(\delta_N) \;=\; \widehat{\sigma}[r_e], \tag{58}$$

which, when combined with (49), 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 t-distribution $g \sim f_n(g \mid 0, m)$, a representative agent will always prefer – *for any finite* S – the normal distribution $g \sim N(\widehat{g}, S^2)$. Theorem 3 results when the limiting explosiveness of the moment generating function of $f_n(g \mid 0^+, m)$ is contained by the substitution of $f_n(g \mid \delta_N, m)$ with $\delta_N > 0$.

In this model, the standard deviation of normally-distributed equity returns is presumed known, while the standard deviation of the growth rate is unknown, but has been observed in the sample to be $\widehat{\sigma}[g]$. To convey a very sharp mental image of what Theorem 3 is saying here, picture the following thought experiment. Imagine drawing a future time-series data sample from the prototype limiting case of the model where *m is extremely big* (but less than infinity), while simultaneously *δ is extremely small* (but greater than zero). Being arbitrarily close to (but not quite at) the pure-stochastic-risk stationary limit where all parameter values are known exactly, the subjective distribution of the precision of future growth rates is arbitrarily close to a point mass and remains almost unchanged as new data arrive over time.

With the above near-stationary setup, the data generating process for future g has the t-like properties of (36), meaning that with very large m the actual growth rates being generated are statistically indistinguishable from a normal random variable with standard deviation $\widehat{\sigma}[g]$, so that $\sigma[g] \approx \widehat{\sigma}[g]$. Simultaneously in this thought experiment, the observed time series of equity returns is reconfirming (up to sampling error) that the distribution of r_N appears to be normal with standard deviation $\sigma[r_N] \approx S(\delta_N) = \widehat{\sigma}[r_e]$. Therefore, since

$$\widehat{\sigma}[r_e] - \widehat{\sigma}[g] \quad \approx \quad \sigma[r_N] - \sigma[g], \tag{59}$$

the observed excess volatility of equity matches statistically the predictions of the theory. Nevertheless, what is actually being observed in this thought experiment seems almost mind-boggling, because year after year of new realizations from this data generating process are confirming that g appears *as if* it is normally distributed with standard deviation $\widehat{\sigma}[g]$, yet fully-rational investors nevertheless continue stubbornly to maintain in their mind's eye an unshakably-consistent welfare-equivalent subjective mental image *as if* the relevant variability of consumption growth rates is $\sigma[g_N]$ $(= \widehat{\sigma}[r_e] \gg \widehat{\sigma}[g])$.

Quasi-Constant Parameter	Value
Mean arithmetic return on equity	$E[R_e] \approx 7\%$
Geometric standard deviation of return on equity	$\sigma[r_e] \approx 17\%$
Risk-free interest rate	$r_f \approx 1\%$
Implied equity premium	$\ln E[R_e] - r_f \approx 6\%$
Mean growth rate of per-capita consumption	$E[g] \approx 2\%$
Standard deviation of growth rate of per-capita consumption	$\sigma[g] \approx 2\%$
Rate of pure time preference	$\rho \approx 2\%$
Coefficient of relative risk aversion	$\gamma \approx 2$

Table 1: Some Stylized Economic "Facts"

7 Some Bayesian Calibration Exercises

Given the free parameter δ, the model endogenously derives theoretical partial-equilibrium formulas for three economic quasi-constants of interest: the equity premium as the function $\Psi(\delta)$, the risk-free rate as the function $\Phi(\delta)$, and equity-return variability as the function $S(\delta)$. Theorem 1 proves the existence of a δ_e that makes $\Psi(\delta_e)$ match the empirically-observed equity premium. Theorem 2 proves the existence of a δ_f that makes $\Phi(\delta_f)$ match the empirically-observed risk-free rate. And Theorem 3 proves the existence of a δ_N that makes $S(\delta_N)$ match the empirically-observed variability of equity returns – all in the context of an internally-consistent fully-rational welfare-equivalent story about an as-if-normally-distributed subjective future growth rate g_N with $\sigma[g_N] = \sigma[r_N] = S(\delta_N) = \widehat{\sigma}[r_e] \gg \widehat{\sigma}[g]$.

The following empirical question then arises naturally from the three partial-equilibrium theorems. Can the *same* value of the exogenous primitive $\delta = \delta^*$ match *simultaneously* the actually-observed values of the three economic-financial variables, so that $\ln \widehat{E}[R_e] - \widehat{r}_f \approx \Psi(\delta^*)$, $\widehat{r}_f \approx \Phi(\delta^*)$, and $\widehat{\sigma}[r_e] \approx S(\delta^*)$? In other words, can the *three* degrees of freedom represented by $\Psi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ be explained empirically by the *one* degree of freedom represented parsimoniously by $\delta = \delta^*$ in this theory? It should be appreciated that such a question represents a quite demanding test of the overall validity of the general-equilibrium version of this model. The answer is "yes," which we now proceed to show.

We are testing whether the welfare-equivalent interpretation that the future growth rate g is subjectively distributed *as if* it were the normal random variable g_N with mean $E[g_N] = \widehat{g}$ and standard deviation $\sigma[g_N] = \widehat{\sigma}[r_e]$ renders, along with (49), an internally-consistent as-if story connecting together the actual parameter values of our economic world. In Table 1, parameter settings have been selected that, I think, represent values well within the "comfort zone" for most economists. All rates are real and given by annual values. The data are intended to be a stylized approximation of what has been observed for many countries over long time periods.

The model is explaining endogenously three quasi-constants $\Psi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ as functions of the one free parameter δ. We do not observe the underlying primitive value of δ directly, although we know that it is operationally indistinguishable from zero when m is conceptualized as being very large because $m \to \infty$ implies $\delta \to 0^+$. However, and more usefully, δ can be calibrated indirectly by setting any one of the three quasi-constants $\Psi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ equal to its observed value in Table 1 and then backing out the implied values of the other two by using the as-if-lognormal formulas (17) and (20).

Defining δ_N to be the implicit solution of

$$\delta_N = S^{-1}(\widehat{\sigma}[r_e]) = S^{-1}(17\%),$$

we then have, from (17) with $\widehat{V} \equiv S(\delta_N)^2$,

$$\ln E[R_e] - r_f = \Psi(\delta_N) = \gamma S(\delta_N)^2 = 5.8\%,$$

and, from (20) with $\widehat{V} \equiv S(\delta_N)^2$,

$$r_f = \Phi(\delta_N) = \rho + \gamma \widehat{g} - \frac{1}{2}\gamma^2 S(\delta_N)^2 = 0.2\%.$$

Defining δ_e to be the implicit solution of

$$\delta_e = \Psi^{-1}(\ln \widehat{E}[R_e] - \widehat{r}_f) = \Psi^{-1}(6\%),$$

we then have, from (20) and (17),

$$r_f = \Phi(\delta_e) = \rho + \gamma \widehat{g} - \gamma \Pi(\delta_e)/2 = 0\%,$$

and, from (17) with $\widehat{V} \equiv S(\delta_N)^2$,

$$\sigma[r_e] = S(\delta_e) = \sqrt{\Psi(\delta_e)/\gamma} = 17\%.$$

Defining δ_f to be the implicit solution of

$$\delta_f = \Phi^{-1}(\widehat{r}_f) = \Phi^{-1}(1\%),$$

we then have, from (20) and (17),

$$\ln E[R_e] - r_f = \Psi(\delta_f) = 2[\rho + \gamma \widehat{g} - \Phi(\delta_f)]/\gamma = 5\%,$$

and, from (20) with $\widehat{V} \equiv S(\delta_N)^2$,

$$\sigma[r_e] = S(\delta_f) = \sqrt{2[\rho + \gamma \widehat{g} - \Phi(\delta_f)]/\gamma} = 16\%.$$

As a rough test for the overall consistency and raw fit of this general-equilibrium theory, the results of these Bayesian as-if-lognormal calibration exercises speak for themselves.

Continuing with the above as-if-lognormal scenario, next consider a purely hypothetical thought experiment in which the magic trick is performed of eliminating *all* future macroeconomic variability about the trend growth rate. Applying the formula for the expectation of a lognormal random variable to (4), the resulting welfare gain is then equivalent to a change in the trend growth rate of

$$\Delta g = \frac{\gamma - 1}{2}\, \sigma^2[g]. \tag{60}$$

When $\gamma \approx 2$ and the historical value of $\sigma[g] \approx 2\%$ is used in (60), then $\Delta g \approx 0.02\%$, which is roughly the order of magnitude of numbers widely cited as indicating that even a complete removal of macroeconomic uncertainty might be worth very little. Such a number, however, captures only the "weak force" of purely stochastic growth-rate risk. The trend-growth welfare equivalent of a strictly hypothetical elimination of *all* uncertainty about the future growth process, which includes the "strong force" of structural parameter uncertainty, is more accurately assessed by using the subjective value $\sigma[g] = \widehat{\sigma}[r_e] \approx 17\%$ in formula (60), which paints a very different picture since in this case $\Delta g \approx 1.45\%$.

8 Conclusion

The δ-theory model of this paper is predicting that, when viewed through the lens of the standard frequentist calibration paradigm, there will simultaneously appear to be an "excess volatility puzzle," a "risk-free rate puzzle," and an "equity premium puzzle," whose magnitudes of discrepancy are very close numerically to what is actually observed in the data. This paper shows that such numerical "discrepancies" are puzzles, however, only when seen through a non-Bayesian lens. From a Bayesian perspective, the "puzzling" numbers being observed in the data are telling an internally-consistent fully-rational story about the implicit prior distribution of background structural-parameter uncertainty stemming from the unknown future growth process that is generating such data.

In principle, consumption-based representative agent models provide a complete answer to all asset pricing questions and give a unified theory integrating together the economics of finance with the real economy. In practice, consumption-based representative agent models with standard preferences and a traditional degree of relative risk aversion work poorly when the variance of the growth of future consumption is point-calibrated to the sample variance of its past values. The theme of this paper is that there is an internally consistent theoretical justification for treating the non-observable variance of the subjective future growth rate *as if* it were equal to the observed variance of a comprehensive economy-wide index of equity returns, for which as-if interpretation the simple standard neoclassical model has 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] DeGroot, Morris H. (1970), *Optimal Statistical Decisions*. New York: McGraw-Hill.

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

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

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

[13] Hamilton, James D. (1994), *Time Series Analysis.* Princeton University Press, NJ.

[14] Koop, Gary (2003), *Bayesian Econometrics.* New York: Wiley.

[15] Leamer, Edward E. (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York: Wiley.

[16] Lucas Jr., Robert E. (1978), "Asset Prices in an Exchange Economy." Econometrica, 46(6), 1429-1445.

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

[18] Mehra, Rajnish, and Edward C. Prescott (1985), "The equity premium: a puzzle." *Journal of Monetary Economics* 15: 145-161.

[19] 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.

[20] Poirier, Dale J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach.* Cambridge MA: The MIT Press.

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

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

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

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

[25] Zellner, Arnold (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: J. Wiley and Sons, Inc.