January 13, 2005
\unified

A Unified Bayesian Theory of Equity 'Puzzles'

Abstract

In the equity premium, risk-free rate and excess volatility puzzles, the subjective distribution of future growth rates typically has its mean and variance point-calibrated to past sample averages. This paper shows that Bayesian estimation of uncertain structural growth parameters introduces an irreducible fat-tailed background uncertainty that can explain all three puzzles parsimoniously by one unified theory. The Bayesian statistical-economic equilibrium has essentially one degree of freedom, yet all three values of the equity premium, risk-free rate, and excess volatility derived from the model match simultaneously the stylized facts observed empirically in the time-series data.

Martin L. Weitzman
Harvard University
mweitzman@harvard.edu
Comments Appreciated

For detailed useful comments on an earlier draft of this paper, but without implicating them in its remaining defects, I am grateful to Andrew Abel, Gary Chamberlain, Alfred Galichon, John Geweke, Joseph Kadane, Naranya Kocherlakota, and Jonathan Parker.

1. Introduction to the Role of Structural Uncertainty

The "equity premium puzzle" refers to the striking failure of the standard neoclassical consumption-based representative-agent model of stochastic economic growth to explain the large historical difference between the average return to a representative stock market portfolio and the average return from a representative portfolio of relatively safe stores of value. The neoclassical general-equilibrium paradigm predicts an equity risk premium that is orders of magnitude lower than what is observed. The discrepancy is so large and so pervasive as to suggest strongly that 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 $\theta \ . \ 2\pm1$.

The "risk-free rate puzzle" represents another big disappointment with the standard neoclassical model. The stochastic generalization of the basic Ramsey formula from equilibrium growth theory predicts a risk-free interest rate far higher than what is actually observed, hence the puzzle. To further compound the conundrum, alterations of the model that might lessen the discrepancy in the risk-free rate anomaly tend to increase the discrepancy in the equity premium anomaly. Thus, for example, to eliminate the equity premium puzzle requires an astronomically *high* rate of relative risk aversion, while to eliminate the risk-free rate puzzle calls for a microscopically *low* rate of relative risk aversion.

The third major puzzle for consumption-based neoclassical theory is the "excess volatility puzzle." In principle, comprehensive asset returns should be determined by fundamental expectations about aggregate future dividends, which in turn should be determined by fundamental expectations about the future growth prospects of the real economy. But in practice, observed stock market returns are vastly more volatile than the more-primitive real growth rates that are supposedly driving them.

Taken together, this unholy trinity of puzzles is more than just disturbing. The proper interpretation of these equity macro-puzzles has important ramifications throughout all of economics. At stake is the central issue of whether or not the standard representative-agent consumption-based stochastic-general-equilibrium paradigm is realistic enough to be trusted as a reliable model for understanding the most basic discounting of time and risk. The three intuitively-related puzzles are devastating for the credibility of the neoclassical paradigm because they are fairly crying out that something is deeply wrong with the formulation. Some critical element, which would capture the characteristic that appears to make stocks comparatively so risky, seems to be missing from the standard model. At least for equity pricing applications, a consensus has developed among economists that the standard model is seriously flawed.

Not surprisingly therefore, the family of equity puzzles has stimulated a lot of economic research. In attempting to explain the paradoxes, an impressive literature has developed, which is filled with some imaginatively fruitful variations on the standard model. To overcome one or another equity puzzle, many new models feature exotic 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 from these new 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 existing literature, the risk premium and the risk-free rate are routinely calibrated by plugging into the relevant formulas the sample mean and sample variance of past growth rates. But strictly speaking, the correct procedure requires the full subjective probability distributions of uncertain structural parameters of the model, not just their point estimates. Missing from the framework is a formal incorporation of the decision-theoretic specification required to make a rigorously unified statistical-economic 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. For many economic usages this intuitive methodology may be justified, but, as will be shown, point calibration is a fatally flawed procedure for the particular application of analyzing aversion to model uncertainty, which underlies (or, more accurately, *should* underlie) all asset-pricing calculations. The core problem is that calibrating population moments to sample frequencies understates significantly the investor's utility-weighted predictive uncertainty, which spills over into dramatically biased asset valuations.

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 basic idea is that structural-parameter model uncertainty introduces a form of Bayesian posterior background risk, 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 risk fattens critically the tails of the posterior distribution of future growth rates and increases significantly the value of both the equity premium and excess volatility, while simultaneously decreasing sharply the risk-free interest rate.

To convey the essential statistical insights as crisply 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 parsimonious competitive equilibrium over just two time periods, with a representative agent, for a pure endowment-exchange economy (no genuine production or investment), where returns to equity equals the growth of consumption and both are *i.i.d.* normal, where the utility function is isoelastic, and so forth. For analytical tractability the *only* change made from this standard stochastic specification (wherein *all* parameter values are 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, whose uncertain values represent the primitive distribution of interest driving the entire system.

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 decreases this δ-parameter 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 $\delta^{(}$ 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 include Barsky and DeLong (1993), Timmerman (1993), Bossaerts (1995), Cechetti, Lam and Mark (2000), Veronesi (2000), Brennan and Xia (2001), Abel (2002), Brav and Heaton (2002), Lewellan 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 missing from this literature, however, is a generic appreciation of the *overwhelming force* that tail-fattening structural parameter uncertainty brings to bear on asset pricing by its dominating influence over any calculation involving expected marginal utility. In effect, the direction of this Bayesian force is appreciated in the literature, but not its magnitude.

The one noteworthy exception is an important paper by John Geweke (2001), who 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 non-robustness insight, but pushes it further to argue that the inherent fragility 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 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 averages of the equity premium, risk-free rate and excess volatility are inverted to back out the implicit subjective probability distribution of the future growth rate. Measured in the appropriate state space of expected marginal utility, a world view about the subjective riskiness of future utility-growth prospects emerges from this Bayesian calibration exercise, which is

---

[1] I wish to express my gratitude to two readers of a previous version of this paper for making me aware of Geweke's earlier article after noticing that I had derived a similar result.

operationally 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 quickly to the analytical essence of the equity macro-puzzles, an ultra-parsimonious model is used here. This prototype model is a drastically 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 super-stark reduced-form model presented here.[2]

In this basic prototype model, there are two periods, the present and the future. (The model generalizes to a multi-period dynamic version, but the details are not really essential for the main message of this paper and the resulting clutter of notation is distracting.) The population consists of a large fixed number of identical people normalized to unity. Present consumption is given as C_0, while future 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 with some known subjective probability distribution, but whose future realization is presently unknown. For convenience and consistency throughout the paper, all growth rates, interest rates, yields, and rates of return are computed as continuously-compounded geometric averages. Thus, the growth rate of this simple endowment-exchange economy is the random variable

$$g \equiv \ln C_1 - \ln C_0 , \qquad \textbf{(1)}$$

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

---

[2] See, for example, the survey articles of Campbell (2003) or Mehra and Prescott (2003). The original specification of this form was introduced by Abel (1988).

The primitive driving force throughout this paper is the unknown future growth rate *g*. An asset *a* is a contingent claim in future state *g* to the payoff $\exp(h_a(g))$, expressed in units of consumption. The expression $h_a(g)$ is the (geometrically calculated) *payoff function* for asset *a*. Let the price of this asset be P_a. Then the corresponding (real, geometrically calculated) *asset return function* is

$$r_a(g) \;=\; h_a(g) - \ln P_a \; , \tag{2}$$

from which it immediately follows that asset returns must be distributed as their payoff, plus some constant.

Within this model all asset markets are in some sense phantom entities, because no one actually ends up taking a 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

$$U'(C_0) \;=\; \beta \; \mathrm{E}[U'(C_1) \; \exp(r_a)] \; . \tag{3}$$

For practical purposes of analysis, throughout the paper equations like (3) will be enormously simplified by choosing the utility function to be of the standard iso-elastic form

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

with corresponding marginal utility

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

where the coefficient of relative risk aversion is the positive constant θ. (The Bernoulli logarithmic utility function is a special limiting case of (4), (5) corresponding to $\theta = 1$.) Plugging (5) into (3) and rearranging terms yields, after taking natural logarithms, the expression

$$\ln P_a = \ln E[\exp(h_a - \theta g)] - \rho \ , \qquad \textbf{(6)}$$

where

$$\rho \equiv -\ln \beta \qquad \textbf{(7)}$$

is the instantaneous rate of pure time preference.

An elegant and extremely useful route for proceeding further is by way of expressing all random variables in units of deviation from their mean. In this spirit, define

$$x \equiv g - E[g] \qquad \textbf{(8)}$$

and

$$y_a \equiv h_a - E[h_a] \ , \qquad \textbf{(9)}$$

and then substitute (8), (9) and (2) into (6). After canceling redundant terms and rearranging, we have derived the fundamental relationship

$$r_a = y_a + E[r_a] \ , \qquad \textbf{(10)}$$

where

$$E[r_a] = \rho + \theta E[g] - \ln E[\exp(y_a - \theta x)] \ . \qquad \textbf{(11)}$$

Expressions (10) and (11) are the workhorse equations of this paper. They show clearly that the derived equilibrium distribution of the returns on any asset is a simple linear function of its more-fundamental payout process, with a slope of one and a value of the intercept given by the reduced-form expression (11).

An immediate application of (11) is to derive the risk-free interest rate. In this situation we use the standard notation $a = f$ to indicate that we are treating here the special case of a deterministic asset $y_a = y_f = 0$, for which (11) becomes

$$r_f \;=\; \rho + \theta \mathrm{E}[g] + \ln\, \mathrm{E}[\exp(-\theta x)] \;. \qquad \textbf{(12)}$$

Another immediate application of formula (11) is for the special case of a comprehensive broad-based equity index representing the entire economy. Here we use the standard notation $a = e$ to indicate that we are treating the situation of economy-wide equity $y_a = y_e = x$, for which equation (11) yields

$$\mathrm{E}[r_e] \;=\; \rho + \theta \mathrm{E}[g] - \ln\, \mathrm{E}[\exp((1-\theta)x)] \;. \qquad \textbf{(13)}$$

Subtracting (12) from (13), the equity premium here is

$$\mathrm{E}[r_e] - r_f \;=\; \ln\, \mathrm{E}[\exp(-\theta x)] - \ln\, \mathrm{E}[\exp((1-\theta)x)] \;. \qquad \textbf{(14)}$$

The meaning given in the literature to result (14) goes along the following lines. Interpret the left hand side of equation (14) as the *actual* risk premium that is observed historically in the real world. Interpret the right hand side of equation (14) as a theoretical *formula* for calculating this risk premium, given any coefficient of relative risk aversion θ, and, more importantly here, given the true subjective probability distribution of deviations of the random future growth rate g from its mean value $\mathrm{E}[g]$.

Concerning the risk-aversion parameter θ, there seems to be some agreement within the economics profession 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 $\theta \leq 3$ 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 the true probability distribution of 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 $g-\mathrm{E}[g]$ 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 $g-\mathrm{E}[g]$. The rest of the story hinges on specifying the form of the assumed density function of $x = g-\mathrm{E}[g]$, 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

$$g \sim \mathrm{N}(\mu, V) \ , \tag{15}$$

where μ and V are unknown parameter values that must be estimated statistically from past data.

When V in (15) is treated as a random variable, then using the formula for the expectation of a lognormal distribution transforms the theoretical equity premium formula (13) into

$$\mathrm{E}[r_e]-r_f = \ln\ \mathrm{E}[\exp(\tfrac{1}{2}\theta^2 V)]\ -\ \ln\ \mathrm{E}[\exp(\tfrac{1}{2}(\theta-1)^2 V)] \ , \tag{16}$$

where the expectation operator is understood here as being taken over V. Note for the exponential coefficients multiplying V in the right hand side of (16) that $\theta^2 > (\theta-1)^2$ whenever $\theta>\tfrac{1}{2}$. It is then relatively straightforward to show that point calibration to the mean of V for $\theta>\tfrac{1}{2}$ biases formula (16) in the direction of predicting a theoretical value of the equity premium that is too low. The equity-premium literature generally proceeds from (13) or (16) by ignoring the bias-producing uncertainty inherent in point estimates of V. Instead, the usual practice calibrates V by plugging in the sample variance from n previous observations on growth rates, and then proceeds as if normality still holds, instead of substituting into (16) the relevant inverted-gamma distribution to account for the sampling error from estimating the unknown structural parameter V.

The observed sample variance is

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} (g_i - \hat{g})^2 \ , \tag{17}$$

where

$$\hat{g} = \frac{1}{n} \sum_{i=1}^{n} g_i \tag{18}$$

is the sample mean. Implicitly in the equity-premium literature, the sample size n is presumed large enough to make (18) and (17) 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 (16) in this model if the estimates, and therefore the approximations, are *not* "sufficiently accurate." In this literature the value of (16) is calculated to be what it reduces to when there is no structural uncertainty and V is known exactly to be equal to $\hat{V}$. After canceling terms, the as-if-deterministic-V version of the theoretical formula (16) then becomes

$$\mathrm{E}[r_e] - r_f = (\theta - \tfrac{1}{2}) \ \hat{V} \ , \tag{19}$$

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 arithmetic return on the broadest available stock market index was $\mathrm{E}[R_e] \approx 7\%$, with an arithmetic standard deviation $\sigma[R_e] \approx 18\%$.[3] Converting to continuously compounded rates gives a geometric mean $\mathrm{E}[r_e] \approx 5.5\%$ and a geometric standard deviation $\sigma[r_e] \approx 17\%$. The historically observed return on an index of the safest available most-liquid short-maturity bills is about 1% per annum, implying for the equity premium that $\mathrm{E}[r_e] - r_f \approx 4.5\%$. The mean yearly growth rate of U.S. per capita consumption over the last century or so is about 2%, with standard deviation about 2%, meaning $\hat{V} \approx .04\%$. Suppose $\theta \approx 2$. Plugging these values into (19) gives $\hat{\Pi} \approx .06\%$.

---

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

Thus, the actually observed equity premium on the left hand side of equation (13) exceeds the estimate (19) 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 θ, it would require the coefficient of relative risk aversion to be 113, which is away from acceptable reality by well over an order of magnitude. This is the equity premium puzzle, and it is apparent why characterizing this 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 volatility 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, a factor of seventy-five 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 literature to equation (12) parallels the interpretation given to the equity premium formula. Interpret the left hand side of equation (12) as the actual risk-free interest rate that is observed historically in the real world. Interpret the right hand side of equation (12) as a theoretical formula for calculating this risk-free interest rate, given θ and the true subjective probability distribution of the future growth rate *g*. Concerning the behavioral risk-aversion parameter θ, a value that would be accepted by the economics profession as a whole is about two, roughly. By contrast, nobody knows what is the true subjective probability distribution of the future growth rate *g*. The best that can be done here is to make some statistical *inference* about the likely probability distribution of *g* from observing past realizations of growth rates.

When the normality specification (15) is made and *V* is treated as a random variable, then using the formula for the expectation of a lognormal distribution transforms the theoretical risk-free rate formula (12) into

$$r_f = \rho + \theta E[g] - \ln E[\exp(\tfrac{1}{2}\theta^2 V)] , \qquad \textbf{(20)}$$

where the expectation of the third term on the right hand side of (20) is taken over V. From the exponential function in (20) being convex in V, a mean-preserving spread of V decreases the theoretically predicted risk-free rate of interest. Therefore, point calibration to the mean of V biases formula (20) in the direction of giving a value of the risk-free rate that is always too high.

The risk-free rate literature typically proceeds from (20) by ignoring the statistical uncertainty inherent in measuring V. Instead, this literature calibrates V by essentially plugging into (20) the point estimate $\hat{V}$ and then proceeds as if normality still holds. Substituting the sample mean $\hat{g}$ and the sample variance $\hat{V}$ into (20) then further transforms the theoretically-calculated value of r_f into

$$r_f = \rho + \theta\hat{g} - \tfrac{1}{2}\theta^2\hat{V} , \tag{21}$$

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 (21) (calculated on an annual basis) are: $\hat{g} \approx 2\%$, $\hat{V} \approx .04\%$, $\rho \approx 2\%$, $\theta \approx 2$. With these representative parameter values plugged into the right hand side of (21), the left hand side becomes $\hat{\Lambda} \approx 5.9\%$. When compared with an actual real-world risk-free rate $r_f \approx 1\%$, 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 negative, while the coefficient of relative risk aversion required to explain the equity-premium discrepancy estimated from (19) is $\theta \approx 113$. The simultaneous existence of two strong contradictions with reality, which, in addition, are strongly contradicting each other, is disturbing times three!

As if all this were not vexing enough, we have the additional enigma of the excess volatility puzzle. The observed time-series standard deviation of real equity returns $\sigma[r_e] \approx 17\%$ is much bigger than the observed time-series standard deviation of real consumption growth rates $\sigma[g] \approx 2\%$. But the return to equity in a comprehensive stock market should essentially reflect the more fundamental growth rate of the real economy it is capitalizing. According to equation (2), for the case $a = e$ of economy-wide equity returns we should be observing approximately that $\sigma[g] \approx \sigma[r_e]$, yet this approximation is off by a factor of about 8.5. In other words, the stock-market "forecast" is about an order of magnitude more volatile than the fundamental underlying consumption dividend that it is supposed to be forecasting.

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 $\mathrm{E}[g]$ and $V[g]$ take on the standard familiar sampling distributions that arise naturally when *n* sample points are drawn randomly from a normal population.

3. The Bayesian Subjective Distribution of Future Growth Rates

To be perfectly clear throughout the rest of the paper, we summarize here the precise specification of the model to be used. The assumed structure is: $r_e \sim \mathrm{N}(\mathrm{E}[r_e], V[r_e])$, $g \sim \mathrm{N}(\mathrm{E}[g], V[g])$. The following structural parameters are assumed to be effectively known and fixed: $\mathrm{E}[r_e]$, $V[r_e]$, r_f, ρ, θ. The following structural parameters are unknown and must be estimated: $\mathrm{E}[g], V[g]$. The Bayesian statistical estimation of $\mathrm{E}[g]$ and $V[g]$ proceeds as follows.

Assuming the normal specification (15), define the random variable

$$W \equiv 1/V \,, \tag{22}$$

which is commonly called the *precision* of a normal probability distribution. Given any *W*, and given any random variable $\tilde{\mu}$, which represents the unknown mean of *g*, we can then write

$$g = \tilde{\mu} + \varepsilon \,, \tag{23}$$

where $\varepsilon \sim \mathrm{N}(0, 1/W)$.

Purely for simplicity here suppose that initially, before any observations are made, the Bayesian pre-sample estimate of the random variable $\tilde{\mu}$ 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 (23), which is drawn from a normal distribution with known precision W, but whose Bayesian pre-sample prior estimate of $\tilde{\mu}$ is a diffuse-normal distribution. With a known variance, the posterior distribution of $\tilde{\mu}$ after n independent sample observations is

$$\tilde{\mu} \sim \mathrm{N}(\hat{g},\ 1/nW) \ . \qquad \textbf{(24)}$$

From (24) and (23), $\mathrm{E}[g] = \mathrm{E}[\tilde{\mu}] = \hat{g}$. Therefore, $g - \mathrm{E}[g] = g - \hat{g}$, and, from applying definition (8) to this situation, $x = g - \hat{g}$ and $x_i = g_i - \hat{g}$.

For *given* values of W *and* $\tilde{\mu}$, the random variable g is distributed according to (23) as

$$g \sim \mathrm{N}(\tilde{\mu},\ 1/W) \ , \qquad \textbf{(25)}$$

whereas, for any given value of W alone, the random variable $\tilde{\mu}$ is distributed according to (24). Combining these two quasi-independent realizations of normal processes, the random variable $x = g - \hat{g}$ must be distributed normally with mean zero and variance equal to the sum of the variance of the normal process (24) plus the variance of the conditionally-independent normal process (25). After adding together the two variances ($1/nW$ from (24) plus $1/W$ from (25)), the posterior distribution of $x = g - \hat{g}$ comes out to be

$$x \sim \mathrm{N}(0,\ (n+1)/nW) \ . \qquad \textbf{(26)}$$

Thus far, the specification has proceeded as if W were known. When W is uncertain, Bayesian statistical theory has developed a rigorous and elegantly symmetric counterpart to 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 conjugate distributions. For reasons that will later become apparent, we work here with a three-parameter generalization of the two-parameter gamma distribution, which forms a *normal-truncated-gamma* family of conjugate distributions.

Consider a non-negative random variable w representing the precision. Let δ be a non-

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

negative parameter representing an arbitrarily imposed lower support for the Bayesian prior distribution of the precision *w*. Assume that the Bayesian prior distribution of the precision *w* is a truncated-gamma probability density function (with truncation parameter δ) of the form

$$\varphi_0^\delta(w) = w^{a_0-1}\, e^{-b_0 w} \Big/ \int_\delta^\infty w^{a_0-1}\, e^{-b_0 w}\, dw \tag{27}$$

for $w \geq \delta$, while $\varphi_0^\delta(w) = 0$ for $w < \delta$. When choosing δ to be positive, the model is effectively 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 *x* converge to a finite value. An economic rationale presumably has to do with the difficulty of envisioning the unbounded loss function arising from unlimited variability in growth rates. The implicit message is that the appropriate value of δ is far from being known *a priori*.

The three non-negative parameters $\delta,\ a_0,\ b_0$ of the truncated gamma distribution (27) represent prior beliefs about the precision. In the limit as $\delta \to 0^+$, the mean of the gamma prior approaches a_0/b_0, while the variance of the gamma prior approaches a_0/b_0^2. Thus, at least for small δ, the prior mean and prior variance of the precision 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 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 limiting value 0^+.

Let $\varphi_n^\delta(w)$ 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 in 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 here is the following lemma, which is stated without proof:

$$\varphi_n^{\delta}(w) = \frac{w^{a-1}\, e^{-bw}}{\int_{\delta}^{\infty} w^{a-1}\, e^{-bw}\, dw}\,, \tag{28}$$

for $w \geq \delta$, while $\varphi_n^{\delta}(w) = 0$ for $w < \delta$, where the parameters a and b are defined by the equations

$$a = \frac{n}{2} + a_0 \tag{29}$$

and

$$b = \frac{1}{2} \sum_{i=1}^{n} x_i^2 + b_0\ . \tag{30}$$

It is analytically very convenient (and, in the context of this model, comes at the cost of only an insignificant loss of generality) to compress the two parameters a_0 and b_0 of the prior truncated-gamma distribution into just one parameter by imposing the additional conditions

$$a_0 = m/2 \tag{31}$$

and

$$b_0 = \hat{V}m/2\ , \tag{32}$$

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

From combining (28) with (26), the unconditional or marginal probability density function of x is

$$f_n(x|\delta,m) = k_n(\delta,m) \int_{\delta}^{\infty} \exp(-x^2 nw/2(n+1))\ w^{a-\frac{1}{2}}\ e^{-bw}\ dw\ , \tag{33}$$

where $k_n(\delta,m)$ is just the constant of integration satisfying

$$1/k_n(\delta,m) = \int_{-\infty}^{\infty} dx \int_{\delta}^{\infty} \exp(-x^2 nw/2(n+1))\ w^{a-\frac{1}{2}}\ e^{-bw}\ dw\ . \tag{34}$$

The two non-negative parameters δ and m are highlighted in formulas (33) and (34) just to remind us that (among many other things, such as $\hat{V}$ 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" n years 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. For just this reason, 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 shows that (33), (34) reduces to the (non-standardized) t distribution

$$f_n(x|0,0) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi \hat{V} n}\ \Gamma(n/2)}\ \left[1 + \frac{x^2}{(n+1)\hat{V}}\right]^{-\frac{n+1}{2}}\ , \tag{35}$$

whose moment generating function is unboundedly large because the relevant integral diverges.

(It is essentially in order to make this moment-generating integral converge that the condition $\delta > 0$ is imposed in the first place.)

This entire preliminary discussion of the future consequences of what people now think that people long ago "might have been thinking" about such things as an upper bound on V (of $1/\delta$) or a degree of prior confidence in $\hat{V}$ (of m) has an unreal tone about it. In practice this issue ought to be non-operational – and therefore not worth contemplating – because the intervening n observations should have bleached the prior parameters out of the posterior distribution. Thus, if the number of data points n is large enough, it "should not matter" what values of δ or m we select now to represent past beliefs. This "should not matter" intuition 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 expected utility *do not* converge uniformly in n for all $\delta > 0$. In this sense there is a distinction, which is critical for all expected-utility 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 years ago during the pre-data past. It is essentially a mental artifice for framing a subjective thought-experimental dialogue between the present and the past concerning how to predict 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 years 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 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 expected utility) for m, but such a robustness condition *does not* hold (in the space of expected utility) for δ . The value of δ that has now been chosen 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 viewpoint, we "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 is being observed.

To summarize, in the Bayesian setting appropriate for thinking about basic issues of risk aversion and asset pricing (which underlie the entire family of equity puzzles), the subjective element involved in choosing a prior distribution of structural parameters cannot be separated from the calibration process. Non-uniform convergence in expected-utility space means that the fickle whimsicality of current investors about 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 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 (33) as our subjective posterior probability density function, 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, *seriatum*. 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.* When *m* is "very big," 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 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 $\Pi(\delta)$ represent the value of $E[r_e]-r_f$ as a function of δ that is obtained from formula (13) when the probability density function is $f_n(x|\delta,m)$ defined by equation (33). Plugging (33) into (13), we obtain

$$\Pi(\delta) \;\equiv\; \ln \int_{-\infty}^{\infty} \exp(-\theta x)\, f_n(x|\delta,m)\; dx \;-\; \ln \int_{-\infty}^{\infty} \exp((1-\theta)x)\, f_n(x|\delta,m)\; dx \;. \tag{36}$$

We then have the following proposition.

Theorem 1 Suppose that $\theta>½$ and $m+n<\infty$. Let $E[r_e]-r_f$ be *any* positive value of the equity premium. Then there exists some positive δ_e such that

$$E[r_e]-r_f \;=\; \Pi(\delta_e) \;. \tag{37}$$

Proof: Conditional on any given precision w, from (26) the random variable x is normally distributed with mean zero and variance $(n+1)/nw$, implying

$$E[\exp(-\theta x)|w] \;=\; \exp((n+1)\theta^2/2nw) \;. \tag{38}$$

Making use of (33) then implies

$$E[\exp(-\theta x)] \;=\; k_n(\delta,m) \int_{\delta}^{\infty} \exp((n+1)\theta^2/2nw)\; w^{a-½}\; e^{-bw}\; dw \;. \tag{39}$$

It is readily apparent that as δ is made to approach zero, the right hand side of (39) approaches infinity. Essentially the same argument holds for $E[\exp((1-\theta)x)]$. Thus, from (36),

$$\lim_{\delta\to 0} \Pi(\delta) \;=\; \ln \lim_{\delta\to 0} \frac{\int_{\delta}^{\infty} \exp((n+1)\theta^2/2nw)\; w^{a-1}\; e^{-bw}\; dw}{\int_{\delta}^{\infty} \exp((n+1)(1-\theta)^2/2nw)\; w^{a-1}\; e^{-bw}\; dw} \;. \tag{40}$$

Because $\theta > \frac{1}{2}$, the ratio on the right hand side of (40) approaches infinity as δ is made to approach zero, implying $\Pi(\delta) = +\infty$. At the other extreme of δ, it is apparent that $\Pi(+\infty) = 0^+$, because there is no equity premium when there is no uncertainty.

The function $\Pi(\delta)$ defined by (36) is continuous in δ. Since

$$\Pi(\infty) \quad < \quad E[r_e] - r_f \quad < \quad \Pi(0) \ , \tag{41}$$

the result (37) follows. $\blacksquare$

The essence of the Bayesian statistical mechanism driving the theorem can be intuited by examining what happens in the limiting case. As $\delta \to 0^+$, the limit of (33) is a (non-standard) *t* distribution of the form (35) – except that $m+n$ replaces *n*. With the presumed case of large $m+n$ and small δ, the central part of the *t*-like distribution (33) is approximated well by a normal 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, the normal distribution may be a very bad approximation indeed, because the relatively fatter tail of the dampened-*t* distribution (33) is capable of producing an explosion in formulas like (13) or (14), implying in the limit as $\delta \to 0^+$ an unboundedly large equity premium. Properly construed, such kinds of explosions are essentially giving an economic interpretation (in terms of pervasive structural background uncertainty about the possibility of taking a serious hit in equities just when consumption is abnormally low) to the statistical fact that the moment generating function of a *t*-distribution is infinite.

An explosion of the equity premium does not happen in the real world, of course, but a tamed near-explosive outcome remains the 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 shadow-risk cannot easily be articulated, yet it frightens people away from taking a more aggressive stance in equities and scares them into a position of wanting to hold instead some safer stores of value such as gold, cash, real goods, or government treasury bills – as a hedge against low-consumption states. Consequently, such 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 for the equity premium puzzle. After all, the *qualitative* fact that $\mathrm{E}[r_e]-r_f$ is positive comes as no surprise, just from first principles of risk aversion. The equity premium puzzle is the *quantitative* paradox that the observed value of $\mathrm{E}[r_e]-r_f$ is too big to be reconciled with the standard neoclassical stochastic growth paradigm. But *compared with what* is the observed value of $\mathrm{E}[r_e]-r_f$ "too big"? The answer given in the equity-premium literature is: "compared with the right hand side of formula (19)." Unfortunately for this logic, the right hand side of (19) is in practice a very bad estimate of the true value of $\mathrm{E}[r_e]-r_f$ as given by equations (13) or (16). 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 effect, the frequentist-inspired literature that produces the family of equity puzzles avoids the consequences on expected utility of non-uniform convergence (in *n*, for any positive δ) only by imposing the pointwise-convergent extreme case $m \to \infty$ right from the beginning. Given any model of utility, it is well known that in principle there exist subjective probabilities that can produce the necessary marginal-utility state shadow-weights to "explain" the observed prices of traded financial assets. The interesting question then becomes: are these subjective probabilities sufficiently close to objective probabilities to be plausible? The existing literature errs by attempting to address this question in the non-relevant space of observed past consumption, where the answer is negative, instead of in the relevant space of non-observed expected utility of subjectively-distributed future consumption, where the answer is positive.

We are witnessing growth data from the past that look as if they are normally distributed with mean $\hat{g}$ and variance $\hat{V}$. But the corresponding Bayesian posterior distribution, which is required to evaluate properly the true impact of risk aversion embodied in Theorem 1, indicates that the all-important difference is an unnoticeable (for large $m+n$) upward adjustment in the probabilities of the higher-variance scenarios. Theorem 1 says that once we compute the Bayesian equity premium rigorously, then the paradox recedes. The underlying statistical reason is that, with less than an infinite amount of data, the *t*-like part of the probability density function (33) has a sufficiently plump tail to make very dramatic the Bayesian expected-utility implications of *model uncertainty* – as captured here by uncertain structural parameter values. The theme of this paper is that model uncertainty drives the entire family of equity 'puzzles,' which gives rise to a very different world view than the picture that emerges, e.g., from just plugging into the equity-premium formula (13) a normal distribution whose variance is point-calibrated to past values.

In a well known attempt to explain the equity premium puzzle, Rietz (1988) argues that we cannot exclude the possibility that our sample size is not large enough to describe adequately the full macroeconomic risk. 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 (like a collapse of the presumed structure) 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. The fault is not really with the model, so to speak, but rather the fault is that the modeler is "forcing" the objective rational-expectations sample variance of past growth rates to stand in for the overall risk-adjusted effects on expected utility of a subjectively uncertain future growth process.

For the model of this paper, such "forcing" by the modeler is a harmless approximation for the near-center of the posterior distribution, which represents the $m+n=\infty$ pure rational-expectations stochastic scenario familiar from the literature. But as one moves ever further away from mid-range point estimates of structural parameters into the tails of a rigorous Bayesian treatment of structural parameter uncertainty, the normal approximation for the case $m+n<\infty$ becomes increasingly untenable. Concerning applications to calculating the equity premium or the risk-free interest rate, a normal distribution of past growth rates is a terrible approximation for evaluating the appropriate *t*-like subjective Bayesian posterior distribution of future growth rates.

I think Theorem 1 is trying to tell us that a statistical analogue of the peso problem may be generically ingrained in 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, as the example of replacing the workhorse normal distribution by its *t*-like posterior distribution demonstrates with dramatic consequences (when expressed in units of expected marginal utility). This "Bayesian-statistical peso problem" means that it is not 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 uncertain growth rate, which is essentially intended to describe future worldwide economic prospects, by plugging the sample variance of observed growth rates from the past into a "very bad" approximation of the subjectively-distributed future growth rate, is to underestimate "very badly" how much more risky is a real world gamble on the state of the future economy, when compared with 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. Other examples with other probability distributions may have less (or more) dramatic consequences, but I believe that it is very difficult to get around the moral of this story. For any chosen value of m, 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 reasonable probability distributions obeying standard regularity conditions.[5] 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, tend to become thick-tailed after integrating out the prior parameter uncertainty. Furthermore, such thick-tailed subjective posterior distributions are decisively important in influencing behavior towards risk (as embodied in expected utility calculations). When investors are modeled as perceiving and acting upon these thick-tailed subjective posterior distributions, a fully-rational general-equilibrium interpretation then has sufficient explanatory power to be able to weave together a unified Bayesian theory of the entire family of equity 'puzzles,' as the next three sections of the paper will demonstrate.

[5] The results in Schwarz (2000) can be interpreted as essentially characterizing the class of such fat-tailed posterior distributions under minimally restrictive assumptions on the prior.

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 formulas (12) or (20) when the probability density function is given by equation (33), which is the Bayesian posterior distribution that is consistent with the rest of the model. Plugging (33) into the right hand side of equation (12), the result is

$$\Phi(\delta) \;\equiv\; \rho + \theta E[g] - \ln \int_{-\infty}^{\infty} \exp(-\theta y)\, f_n(y|\delta,m)\; dy \;. \tag{42}$$

We then have the following proposition.

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

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

Proof: It was already shown in the course of proving Theorem 1 that, from (39), as δ is made to approach zero, the integral in the right hand side of (42) becomes unbounded. Therefore, $\Phi(0)=-\infty$. At the other extreme of δ is the deterministic Ramsey formula $\Phi(\infty)=\rho+\theta E[g]$. 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 very strong aversion to high-volatility low-precision situations. The bottom line once more is that the "Bayesian peso problem" causes classical-like rational-expectations inferences, which are based upon the observed historical behavior of past growth rates, to underestimate greatly just how much more uncertain than riskless stores of value is a real-world Bayesian gamble on the state of the future world 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 confront critical modeling issues previously evaded. The empirical existence of a significant equity *volatility* puzzle signals that the standard model of this paper may be mis-specified right from its first postulate. To see this, rewrite the basic starting equation (2) of the model for the case a=e of a comprehensive economy-wide equity index where $h_a(g) = g$. In this special case, (2) becomes

$$r_e - \mathrm{E}[r_e] \quad = \quad g - \mathrm{E}[g] \ . \qquad \textbf{(45)}$$

Clearly, equation (45) cannot be interpreted as being literally correct in a frequentist sense because of the substantial mismatch between the observed variances of the two random variables r_e and g appearing on alternate sides of the same equality sign. We are now forced to unscramble the precise Bayesian interpretation of what it means operationally for equation (45) to be "true" or "false" in a general equilibrium setting where the growth rate of future consumption is subjectively distributed. The excess volatility puzzle says that, empirically, $\sigma[r_e]$ is about an order of magnitude larger than $\sigma[g]$. How can equation (45) possibly be "true" in the presence of such a seemingly irreconcilable observational disparity? The answer here lies in the fact that, for this model, g is an *unobservable* random variable, whose risk-transformed utility-adjusted subjective variability may be very different from its observed sample variability. The Euler equation, after all, imposes restrictions upon expectations of future marginal-utility-weighted equity returns, not upon past realizations of growth rates or equity returns per se.

I now present two not-directly-verifiable subjective "stories" about future economic growth prospects, both of which will turn out to "fit" the stylized facts equally well. The backdrop for both stories begins with the fact that g is subjective and unobservable. So far as the *first* moment of g is concerned, we know that $E[g] = \hat{g}$ from (24) and (25). However, we are quite unsure about how to represent the *second* moment of g. We know empirically that the variability in equity returns greatly exceeds the variability in growth rates, thereby causing an excess-volatility puzzle concerning how to interpret (45). What we do not yet know or understand is how to conceptualize the second moment of g in a way that resolves the paradox of (45). The question is: can we tell a rigorously-consistent Bayesian story (here it will actually be two stories) that might "explain" the connection described by (45) between the non-observable variability of subjective future growth rates and the observed frequency-measured variability of past equity returns?

Subjective story #1 might be called the *engineered equity variability* story about why the observed variability of equity returns might be so much larger than the observed variability of growth rates. Subjective story #1 begins with the idea that, since nobody in this endowment-exchange economy actually ends up taking a position in equity anyway, at least in principle we are free to posit exogenously any state-contingent payoff-producing mechanism we like for stock-market shares. Behind the scene, such a payoff process can be engineered and priced to yield returns that are distributed as the payoff function, plus some constant. With this engineered equity variability story, stocks are no longer direct one-for-one claims on future consumption. In this subjective story #1, stocks are hypothetical claims, which are available in zero net supply, on future payoffs with a known engineered equity variability.

Continuing with subjective story #1, let the payoff function be $h_1(g)$. Applying the fundamental equations (10) and (11) to this situation generates an equilibrium distribution of the return on equity r_1, which is given as some probability density function $\psi_1(r_1)$ defined by

$$r_1 = y_1 + \{\rho + \theta E[g] - \ln E[\exp(y_1 - \theta x)]\} , \tag{46}$$

where $x = g - E[g]$ and $y_1 = h_1 - E[h_1]$. In this story # 1, therefore, the variability of equity returns $\sigma[r_1]$ is derived by construction from the more basic variability of the engineered payoff process $\sigma[h_1]$.

An alternative interpretation, subjective story #2, might be called the *inherent growth* *variability* story about why the observed variability of equity returns might be so much larger than the observed variability of growth rates. Subjective story #2 begins by having the non-observable future growth rate be denoted by the random variable g_2, where the relationship between the random variables g_2 and g remains yet to be determined. In this subjective story #2, the equity market is perceived by investors as being a broad-based representation of the entire economy, meaning the corresponding payoff function is $h_2(g_2) = g_2$. Applying the fundamental equations (10) and (11) to this situation generates an equilibrium distribution of the return on equity r_2, which is given as some probability density function $\psi_2(r_2)$ defined by

$$\mathrm{r}_2 \;=\; x_2 + \{\rho + \theta \mathrm{E}[g_2] - \ln \mathrm{E}[\exp((1-\theta)x_2)]\} \;, \tag{47}$$

where, in this inherent growth variability story, $x_2 = g_2 - \mathrm{E}[g_2]$. From examining (47), it follows at once that (45) must hold with $r_e = r_2$, $g = g_2$. Therefore, and importantly, in subjective story #2 the variability of returns on equity $\sigma[r_2]$ *equals* the variability of the non-observable subjectively-distributed future growth rate $\sigma[g_2]$ by construction.

We will say that the above two stories about the not-directly-observable variability of subjectively-distributed future growth process are *observationally equivalent* if they generate the same expected growth rate, so that

$$\mathrm{E}[g_2] \;=\; \mathrm{E}[g] \;, \tag{48}$$

and if they generate the same distribution of equity returns, so that, for all *r*,

$$\psi_2(r) \;=\; \psi_1(r) \;. \tag{49}$$

From comparing (46) with (47) in the light of (48) and (49), it is relatively straightforward to show that the engineered equity variability story is observationally equivalent to the inherent growth variability story whenever $x_2 = y_1$ and

$$\mathrm{E}[\exp(y_1 - \theta x_1)]\} \quad = \quad \mathrm{E}[\exp((1-\theta)y_1)] \ . \tag{50}$$

From this point forward, considerable conceptual advantage, not to mention computational convenience, come from further restricting the unconditional distribution of (continuously compounded) equity returns to be a normal random variable with known standard deviation *S*. Such normality of equity returns is the benchmark case assumed in most studies of asset pricing (including almost all expositions of the equity family of puzzles) and it is consistent with the empirically observed low-frequency behavior of a comprehensive stock market index measured over discrete time periods of a year or more. We now engineer a state-contingent payoff process via change of variables so that the probability density function of the growth rate *g* is transformed into a payoff $h_1(g|S)$, which is normally distributed with standard deviation *S*. The corresponding engineered relationship between y_1 and *x* is then described by the equation

$$y_1(x|S) \quad = \quad h_1(x+\mathrm{E}[g]|S) \ - \ \mathrm{E}[h_1(x+\mathrm{E}[g]|S)] \tag{51}$$

for some payoff function $h_1(x+\mathrm{E}[g]|S)$ where

$$y_1 \quad \sim \quad \mathrm{N}(0,\ S^2) \ , \tag{52}$$

and *S* is the standard deviation, to be subsequently determined endogenously as a function of δ.

Given $f_n(x|\delta,m)$ and the normal specification (52), the Jacobian inverse-function formula applied to (51), (52) implies that $y_1(x|S)$ is the unique solution of the differential equation

$$\frac{dy_1}{dx} \quad = \quad \frac{f_n(x|\delta,m)\ \sqrt{2\pi}\ S}{\exp(-y_1^2/2S^2)} \tag{53}$$

with the initializing boundary condition $y_1(0|S) \ = \ 0$.

The third and final theorem of the paper shows that there is a value of δ that causes the standard deviation of the subjectively-estimated future growth rate in story #2 to equal *any given* standard deviation of equity returns, and which is operationally indistinguishable from what financial engineering can accomplish in story #1. Thus, not only are parameter values of δ capable in principle of explaining the equity premium and the risk-free rate for this model, but, with the unconditional-normality assumption for equity returns, the following possibility theorem shows that δ is also capable of explaining rigorously the observed volatility of the stock market itself in a way that is consistent with the centerpiece equation (45) of the theory being satisfied.

Theorem 3: Suppose $m+n<\infty$ and assume that the distribution of returns on equity r_e is normally distributed with *any* known positive standard deviation $\sigma[r_e]$. Define the function $S(\delta)$ to be the particular value of S that implicitly satisfies equation (50) for the normal distribution (52) (conditional on a given value of δ), i.e.,

$$\int_{-\infty}^{\infty} \exp(y_1(x|S(\delta))-\theta x)\, f_n(x|\delta,m)\, dx = \exp(\tfrac{1}{2}(1-\theta)^2 S(\delta)^2) \ . \qquad \textbf{(54)}$$

Then there exists some positive δ_v such that

$$S(\delta_v) = \sigma[r_e] \ , \qquad \textbf{(55)}$$

and subjective story #1 is observationally equivalent to subjective story #2 with

$$g_2 \sim \mathrm{N}(\hat{g},\ \sigma^2[r_e]) \ . \qquad \textbf{(56)}$$

Proof: Setting $\delta=\infty$ corresponds to a deterministic economy, in which case $S(\infty)=0$. At the opposite extreme, setting $\delta=0$ causes the usual explosion of the integral on the left hand side of (54), implying for this case that $S(0)=\infty$. The implication is that

$$S(\infty) < \sigma[r_e] < S(0) \ , \qquad \textbf{(57)}$$

and conclusion (55) then follows from continuity of the function $S(\delta)$. ~

The force behind Theorem 3 is the same 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 of consumption. Compared with the *t*-distribution $x - f_n(x*0,m)$, a representative agent will always prefer – *for any finite S* – the normal distribution $x_2 \,'\, y_1 - \mathrm{N}(0,S^2)$. Theorem 3 results when the limiting explosiveness of the moment generating function of $f_n(x*0,m)$ is contained by the substitution of $f_n(x*\delta_v,m)$ with $\delta_v>0$.

To get a sharp mental image of what is happening here, perform 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). In this limiting situation, 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. The subjective distribution of *x* has the *t*-like properties of (33), meaning that the data being generated are statistically indistinguishable from a normal random variable with standard deviation $\hat{\sigma}[g]$. Simultaneously in this thought experiment, the observed time series of equity returns is reconfirming (up to sampling error) that the distribution of r_e appears to be normal with standard deviation $s(\delta_v) \,'\, \sigma[r_e]$. The *excess volatility of equity* being explained theoretically by the model is $S(\delta_v)\&\sigma[x]$. Therefore, since

$$\hat{\sigma}[r_e]\&\hat{\sigma}[g] \quad . \quad S(\delta_v)\&\sigma[x] \ , \tag{58}$$

the observed excess volatility of equity matches what the theory predicts.

7. Calibrating the Bayesian Model

Given δ, the model endogenously derives theoretical partial-equilibrium formulas for three economic quasi-constants of interest: the equity premium as the function $\Pi(\delta)$, the risk-free rate as the function $\Phi(\delta)$, and equity volatility as the function $S(\delta)$. Theorem 1 proves the existence of a δ_e that makes $\Pi(\delta_e)$ 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.

Theorem 3 proves the existence of a δ_v that makes $S(\delta_v)$ match the empirically-observed volatility of equity returns – in the context of an internally consistent observationally-equivalent story about as-if-normally-distributed subjective future growth rates.

The following empirical question then arises naturally from the three partial-equilibrium theorems. Can the *same* value of the exogenous primitive, $\delta = \delta^*$, explain *simultaneously* the actually-observed values of the three economic-financial variables, so that $\mathrm{E}[r_e]-r_f \approx \Pi(\delta^*)$, $r_f \approx \Phi(\delta^*)$, and $\sigma[r_e] \approx S(\delta^*)$? In other words, can the *three* degrees of freedom represented by $\Pi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ be explained empirically by the *one* degree of freedom represented parsimoniously by δ in this theory? The answer is “yes,” which we now proceed to show.

As was explained in the previous section of the paper, equation (45) cannot literally be a true frequency description because of the huge mismatch between the observed sample variances of the two random variables g and r_e. Purely for analytical tractability, the model of this paper has treated structural parameter uncertainty *only* in its primitive driver, the growth rate. To make sense of (45) in such a model (whose stochastic equity returns are treated as being normally distributed with known standard deviation $\sigma[r_e]$), we are allowed by a basic principle of operationalism to choose the as-if interpretation of story #2 over the more conventional interpretation of story #1. If we want to *conceptualize* the non-observable subjective growth rate of future consumption *as if* it is normally distributed, then in order to mesh seamlessly with equation (45) its as-if standard deviation $\sigma[g_2]$ should be calibrated so that $\sigma[g_2] = S(\delta) = \sigma[r_e]$. Such a state-price-deflated calibration to equity-lognormal units of subjective future consumption is harmless, since it has no objectively-measurable consequences within this model-world. Choosing the interpretation of story #2 merely creates a convenient mental image for telling an operationally equivalent as-if parable about (45) holding in terms of the universally familiar normal probability distribution.

We now inquire whether the observationally-equivalent interpretation that the subjective future growth rate g is distributed *as if* it were normal with standard deviation $S(\delta)$ renders along with (45) a consistent as-if story connecting together the actual parameter values of our economic world. In the following table, parameter settings have been selected that, I think, represent values well within the “comfort zone” for most economists. All rates are real and represented by annual values. The data are intended to be a stylized approximation of what has been observed for many countries over long periods of time.

Table 1. Some Stylized Economic Facts

Quasi-Constant Parameter	Value
Mean arithmetic return on equity	$\mathrm{E}[R_e] . 7\%$
Arithmetic standard deviation of return on equity	$\sigma[R_e] . 18\%$
Implied geometric standard deviation of return on equity	$\sigma[r_e] . 17\%$
Implied mean geometric return on equity	$\mathrm{E}[r_e] . 5.5\%$
Risk-free interest rate	$r_f . 1\%$
Implied equity premium	$\mathrm{E}[r_e] \& r_f . 4.5\%$
Mean growth rate of per-capita consumption	$\mathrm{E}[g] . 2\%$
Rate of pure time preference	$\rho . 2\%$
Coefficient of relative risk aversion	$\theta . 2$

The model is explaining endogenously three quasi-constants $\Pi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ as functions of the one positive parameter δ. We do not observe the underlying primitive value of δ directly, although we know that it is operationally indistinguishable from zero since $m64$ implies $\delta 60^{\%}$. However, and more usefully, δ can be calibrated indirectly by setting any one of the three quasi-constants $\Pi(\delta)$, $\Phi(\delta)$, 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 (19) and (25).

Defining δ_v to be the implicit solution of

$$S(\delta_v) \quad ' \quad 17\% \ , \tag{59}$$

we then have, from (19) with $\hat{V} \, ' \, S^2[\delta_v]$,

$$\Pi(\delta_v) \quad ' \quad (\theta \ \& \ ½) \ S^2(\delta_v) \quad ' \quad 4.4\% \ , \tag{60}$$

and, from (25) with $\hat{V} = S^2[\delta_v]$,

$$\Phi(\delta_v) = \rho + \theta \mathrm{E}[X] - \tfrac{1}{2}\theta^2 S^2(\delta_v) = 0.2\% \ . \tag{61}$$

Defining δ_e to be the implicit solution of

$$\Pi(\delta_e) = 4.5\% \ , \tag{62}$$

we then have, from (25) and (19),

$$\Phi(\delta_e) = \rho + \theta \mathrm{E}[X] - \theta^2 \Pi(\delta_e)/(2\theta - 1) = 0\% \ , \tag{63}$$

and, from (19) with $\hat{V} = S^2[\delta_e]$,

$$S(\delta_e) = \sqrt{2\Pi(\delta_e)}/\sqrt{2\theta - 1} = 17\% \ . \tag{64}$$

Defining δ_f to be the implicit solution of

$$\Phi(\delta_f) = 1\% \ , \tag{65}$$

we then have, from (25) and (19),

$$\Pi(\delta_f) = (2\theta - 1)/\theta^2 \ [\rho + \theta \mathrm{E}[g] - \Phi(\delta_f)] = 3.8\% \ , \tag{66}$$

and, from (25) with $\hat{V} = S^2[\delta_f]$,

$$S(\delta_f) = \sqrt{2}/\theta \ \sqrt{\rho + \theta \mathrm{E}[g] - \Phi(\delta_f)} = 16\% \ . \tag{67}$$

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

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 rational story about the implicit prior distribution of background structural-parameter uncertainty 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 asset returns, for which interpretation the simple standard neoclassical model has the potential to work well in practice.

References

Abel, Andrew B. (1988), "Stock prices under time varying dividend risk: an exact solution in an infinite horizon general equilibrium model." *Journal of Monetary Economics* 22:375-394.

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.

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

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

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

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

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

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Cechetti, 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.

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

DeGroot, Morris H. (1970), *Optimal Statistical Decisions*. New York: McGraw-Hill.

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

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

Hamilton, James D. (1995), *Time Series Analysis*. Princeton University Press, NJ.

Kocherlakota, Naranya R. (1996), "The equity premium: it's still a puzzle." *Journal of Economic Literature* 34: 42-71.

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

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

Malkiel, Burton G. (2003), "The efficient market hypothesis and its critics." *The Journal of Economic Perspectives* 17: 59-82.

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

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.

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

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

Schwarz, Michael (2000), "Decision Making Under Extreme Uncertainty." Mimeo, Harvard University Department of Economics, November 30.

Shiller, Robert J. (2003), "From efficient markets theory to behavioral finance." *The Journal of Economic Perspectives* 17: 83-104.

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

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

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