A Unified Bayesian Theory of Equity 'Puzzles'

Martin L. Weitzman*
Harvard University

This Draft: May 2, 2005
Comments Appreciated

Abstract

In expositions of the equity premium, risk-free rate, and equity variability puzzles, the subjective distribution of future growth rates has its mean and variance calibrated to known past sample averages. This paper shows that Bayesian estimation of unknown structural growth parameters adds to future expectations an irreducible fat-tailed background layer of uncertainty, which can explain all three puzzles parsimoniously by one unified theory. The Bayesian statistical-economic equilibrium has essentially 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 equity variability.

1 Introduction: Structural Uncertainty and Asset Prices

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

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

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

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

In principle, aggregate returns on comprehensive economy-wide equity should reflect fundamental growth expectations about the underlying real economy. What I am calling in this paper the "equity variability puzzle" refers to the counterintuitive empirical fact that actual returns on a representative stock market index appear to be about an order of magnitude excessively more variable than any "fundamental" that might be driving them.

Taken together, this unholy trinity of theoretical predictions that miss their empirical targets by entire 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 discounting for risk and time. These three intuitively-related macroeconomic paradoxes are fairly crying out that something is deeply wrong with the standard paradigm – something which is unlikely to be corrected by tinkering with small modifications of the basic model. Some big critical element, which would capture the characteristic that 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, 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 the observation that, throughout the expository literature, the equity premium and the risk-free rate are calibrated by plugging into the relevant expectation formulas the sample mean and sample variance of past growth rates. But the agent-investor presumably is like the Bayesian econometrician in having a full subjective

probability distribution for uncertain structural parameters of the model, not just a point estimate. 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. While this intuitive methodology may well be justified for some economic applications, the paper will show that such point calibration can be 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 essence of the problem is that, *for any finite number of observations*, mimicking the sample frequencies of past growth rates can understate enormously the researcher's (and the agent-investor's) predictive uncertainty about the future marginal-utility-weighted stochastic discount factor. This forecast-expectation bias spills over into severe pricing-kernel errors, which in turn cascade into 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 equity-variability puzzles by simultaneously rooting all three issues together deeply into the common ground of Bayesian statistical inference. The goal here is to ascertain whether the most basic specification of the standard model can be rehabilitated when only the simplest imaginable structural uncertainty is included in the formation of expectations. The fact that structural parameters of the model are unknown will introduce 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 equity variability, while simultaneously decreasing markedly the risk-free interest rate.

To convey the essential statistical insights as sharply as possible, only most basic conceivable specification of the interplay between Bayesian statistical inference and stochastic general equilibrium growth is modeled. Thus, to ease the computational burdens from delivering its core message the model analyzes a very stark competitive equilibrium over just two time periods, with a single representative agent whose utility function is isoelastic, for a pure endowment-exchange fruit-tree economy (having no genuine production or investment), where the return to representative equity and the growth of aggregate consumption are both *i.i.d.* normal, and so on. Abstracted away are such diversionary complications as leverage effects, labor income, human capital, distortionary taxes, heterogeneous beliefs, idiosyncratic

risk, and the like. The central issue, to be addressed by this paper at a very high level of abstraction, is whether the three equity macro-puzzles are "essentially explained" when only the most basic form of structural uncertainty is added to the core macro-model. For analytical clarity the *only* change made from the simplest possible 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 agent-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 equity variability both increase without limit while the risk-free rate simultaneously decreases, also without limit. Furthermore, essentially the same value of δ comes very close to matching simultaneously the equity premium, risk-free rate, and equity variability that are observed empirically 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 may reduce the degree of one or another equity anomaly. What has been utterly missing from this literature, however, is any sense of the generic *strong force* that tail-fattening structural parameter uncertainty can bring to bear on asset pricing equations by its overwhelming ability to dominate the numerical evaluation of standard expectation 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 can exert 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 the three equity macro-puzzles are not 'puzzling' in a Bayesian framework that includes structural model-parameter uncertainty. Instead, the arrow of causality in a unified Bayesian explanation is reversed: the 'puzzling' numbers being observed empirically are trying to tell us a consistent and revealing story 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 equity variability 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 Macro-Puzzles

The critical issue for this paper is whether or not the appearance of the three related equity 'macro-puzzles' might essentially be attributable to the presence of Bayesian structural uncertainty. To cut sharply to the analytical essence of this central issue, a super-stark model is used where everything else except the most basic abstract structure has been set aside. The prototype model used here is a pruned-down reduced-form version of the textbook workhorse formulation employed throughout the financial economics literature.[2]

In this core prototype model there are two periods, the present and the future. By limiting the analysis to just two periods we do not have to take a stand on any critically-consequential properties of a fully-dynamic version. Thus, we need not say whether a fully-specified formulation involves overlapping generations or infinitely-long-lived agents.

[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.

[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 and the stylized empirical facts used throughout this paper. Citations for the many sources of these (and related) seminal asset-pricing ideas are omitted here only to save space, and because they are readily available, e.g., in the above two review articles and in the textbook expositions of Cochrane (2001), Duffie (2001), or Gollier (2001).

Nor must we commit to whether the underlying dynamics are stationary or non-stationary. With just two periods, we also finesse the controversial unresolved issue of whether "rational bubbles" are a substantive component of equity price variability, or not. The one-period-ahead properties of all such multi-period models must reduce to the two-period form analyzed here.

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 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 $H_\alpha(g)$, denominated in units of consumption. The expression $H_\alpha(g)$ symbolizes the unit payoff function for asset α. Let the price of this asset be P_α. Then the corresponding asset return function is

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

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), (5) into (3) and rearranging terms gives the expression

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

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

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

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

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

for which case (7) becomes

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

whose more familiar logarithmic form is

$$r_f \;=\; \rho - \ln E[\exp(-\gamma g)], \tag{10}$$

where $\rho \equiv -\ln \beta$ is the instantaneous rate of pure time preference and the usual definition $r_f \equiv \ln R_f$ is applied.

Another important application of formula (7) is to the special case of a comprehensive broad-based equity index representing the entire economy. In this very simple two-period model, the relevant macroeconomic abstraction is that the one-period-ahead total payoff to comprehensive equity (what would be aggregate dividends plus aggregate capital gains in a multi-period version) essentially reduces in this two-period world to a unit claim on the future aggregate output of the underlying real economy. For this case we use the notation $\alpha = e$ to indicate that we are treating the situation of economy-wide one-period-ahead equity. The corresponding unit payoff function here is then

$$H_e(g) \;=\; \exp(g), \tag{11}$$

for which case equation (7) becomes

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

Dividing (12) by (9) and taking logarithms, the equity premium is

$$\ln E[R_e] - r_f \;=\; \ln E[\exp(g)] + \ln E[\exp(-\gamma g)] - \ln E[\exp((1-\gamma)g)]. \tag{13}$$

Equation (13) gives a theoretical formula for calculating the equity 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 $\gamma \leq 4$ would likely be viewed suspiciously by most members of the broadly-defined community of professional economists as being dependent upon an unacceptably high degree of risk aversion.

By way of contrast, much less is known about what is the appropriate probability distribution to use for representing future growth rates. The reason for this traces back to the unavoidable truth that, even under the best of circumstances (with a given, stable, stationary stochastic specification that can accurately be extrapolated from the past onto the future), 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

$$g \sim N(\mu, V), \tag{14}$$

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

The equity premium literature generally proceeds by implicitly presuming that the structural parameters μ and V are constants known by the agents inside the economy (but perhaps not known by an outsider econometrician), and then continues on by substituting the normal distribution (14) into formula (13), which reduces (13) to a simple well-known expression. Instead of allowing the representative agent in the economy to recognize that μ and V are

themselves unknown random variables, the standard practice essentially plugs in the first two sample moments and then goes on pretending that normality still holds (in place of substituting into (13) the relevant t-distribution to account for sampling error).

Throughout the paper, it is assumed for simplicity that the n data points $\{g_i\}$ are *i.i.d.* realizations of (14). 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 either to define "sufficiently accurate" or to confirm exactly what happens to formula (13) in this model if the estimates, and therefore the approximations, are not "sufficiently accurate." In the expository literature the value of (13) is calculated to be what it reduces to when there is no structural uncertainty because μ is known precisely to be $\widehat{g}$ and V is known precisely to be $\widehat{V}$. When (14) is assumed along with the extreme point-mass case $\mu = \widehat{g}$, $V = \widehat{V}$, then using the formula for the expectation of a lognormal random variable and cancelling the many redundant terms simplifies (13) into the standard expression

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

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

Considering the U.S. as a prime example, in the last century or so the average annual real arithmetic return on the broadest available stock market index is taken to be $\ln E[R_e] \approx 7\%$.[3] The historically observed real return on an index of the safest available short-maturity bills is less than 1% per annum, implying for the equity premium that $\ln E[R_e] - r_f \approx 6\%$. The mean yearly growth rate of U.S. per capita consumption over the last century or so is about 2%, with standard deviation taken to be about 2%, meaning $\widehat{V} \approx 0.04\%$. Suppose $\gamma \approx 2$. Plugging these values into the right hand side of (17) gives $\gamma\widehat{V} \approx 0.08\%$.

Thus, the actually observed equity premium on the left hand side of equation (13) 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

[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 $\ln E[R_e]$ than "America in the American century").

relative risk aversion to be 150, which is away from acceptable reality by about two orders of magnitude. This is the form or variant of the equity premium puzzle applicable to the above macro-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 specifications or parameter values can have the effect of chipping away at the puzzle, but the overwhelming impression is that the equity premium is off by at least an order of magnitude. There just does not seem to be enough variability in the recent past historical growth record of advanced capitalist countries to warrant such a high risk premium as is observed. Of course, the underlying model is extremely crude and can be criticized on any number of valid counts. Economics is not physics, after all, so there is plenty of wiggle room for a paradigm aspiring to be the "standard economic model." Still, two orders of magnitude seems like an awfully large base-case discrepancy to be explained away *ex post factum*.

Turning to the risk-free rate puzzle, the meaning given in the asset-pricing literature to equation (10) parallels the interpretation given to the equity premium formula. Equation (10) is a theoretical formula for calculating the 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.

The literature typically proceeds from (10) by postulating the normal distribution (14), but then imagines that the representative agent ignores the statistical uncertainty inherent in estimating μ and V. Instead, these two structural parameters are usually treated by plugging in their sample values and then pretending that normality still holds. Substituting the deterministic-structure point-mass-parameter version $\mu = \widehat{g}$, $V = \widehat{V}$ into (10), and then using the formula for the expectation of a lognormal distribution gives

$$r_f \;=\; \rho + \gamma\widehat{g} - \frac{1}{2}\gamma^2\widehat{V}, \tag{18}$$

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

Non-controversial estimates of the relevant parameters appearing in (18) (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 (18), the left hand side of (10) 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 value of γ required to explain the risk-free interest rate discrepancy is essentially $\gamma \approx 0$, whereas $\gamma \approx 150$ is required to explain the equity risk premium. Choosing a coefficient of relative risk aversion to ease the risk-free rate puzzle exacerbates the equity premium puzzle, and vice versa. The simultaneous existence of two strong contradictions with reality, which, in addition, seem to be strongly contradicting each other, might be characterized as being disturbing "times three."

As if all of the above were not vexing enough, there is also the enigmatic appearance in the data of what is being called throughout this paper the "equity variability puzzle." In this ultra-simple *i.i.d.* endowment-exchange model, the sole genuine "fundamental" of the system is the random variable g. From combining (2) with (11) and then taking logs, the continuously-compounded deviation of returns on one-period comprehensive equity $r_e - E[r_e]$ should, for this austere representation, have the same distribution as the deviation of its underlying "fundamental" $g - E[g]$. 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 highly abstract tree model of this paper (with *i.i.d.* fruit and isoelastic utility), the relevant macroeconomic form of an "equity variability puzzle" is taken to be the empirical observation that $\widehat{\sigma}[r_e] \gg \widehat{\sigma}[g]$. Thus, the "equity variability puzzle" is understood in this paper to be the stylized macroeconomic fact that the one-period return to representative equity counterintuitively appears to be about an order of magnitude more variable than the underlying fundamental of an aggregate-output real-growth payoff, for which it is the implicit claimant. Conforming here with the usual equity macro-puzzle family pattern, it turns out that substituting alternative specifications or parameter values can lessen somewhat the initial order-of-magnitude discrepancy (here of excess equity volatility), but something central still seems way off base.

Summing up the scorecard for this super-simple version of the standard neoclassical general-equilibrium growth 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. The next section of the paper examines what happens 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. Only the following two structural parameters are unknown and must be estimated statistically: $E[g]$, $V[g]$. This section effectively 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 (14), define the random variable

$$w \equiv 1/V, \tag{19}$$

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 $\widetilde{\mu}$,

$$g = \widetilde{\mu} + \varepsilon, \tag{20}$$

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 (20), which is drawn from a normal distribution with *known* precision w, but whose Bayesian pre-sample prior 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

$$\widetilde{\mu} \sim N(\widehat{g}, 1/nw). \tag{21}$$

From (20) and (21), $E[g] = E[\widetilde{\mu}] = \widehat{g}$. For *given* values of *both* w *and* $\widetilde{\mu}$, the random variable g is distributed according to (20) as

$$g \sim N(\widetilde{\mu}, 1/w), \tag{22}$$

while, for any given value of w *alone*, the random variable $\widetilde{\mu}$ is distributed according to (21). 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 (21) plus the variance of the conditionally-independent normal

process (22). After adding together the two variances ($1/nw$ from (21) plus $1/w$ from (22)), the posterior distribution of g comes out to be

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

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 hyperparameter 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 hyperparameter δ) of the form:

$$\varphi_0(w) \;=\; \frac{w^{a_0-1}\, e^{-b_0 w}}{\int_{\delta}^{\infty} w^{a_0-1}\, e^{-b_0 w}\, dw} \tag{24}$$

for $w \geq \delta$, while $\varphi_0(w) = 0$ for $w < \delta$. Hyperparameters a_0, b_0, δ are treated as quasi-fixed.

When choosing δ to be positive, the model is dogmatically eliminating a priori all variances above $1/\delta$. As will later become clear, 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 implicit subtext is that nobody has the slightest idea about what is actually an appropriate value of δ, which would appear to exist in the first place only for the seemingly minor mathematical technicality of placing *some* finite upper bound above expected marginal utility.

The three non-negative hyperparameters δ, a_0, b_0 of the truncated gamma distribution (24) represent prior beliefs about the precision. In the limit as $\delta \to 0^+$, the mean of the truncated-gamma prior (24) 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

[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).

selecting a_0 and b_0. Classical statistical analysis is exactly isomorphic to the limiting case of a non-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 hyperparameter 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 self-regenerating 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):

$$\varphi_n(w) \;=\; \frac{w^{a-1}\, e^{-bw}}{\int_{\delta}^{\infty} w^{a-1}\, e^{-bw}\, dw}, \tag{25}$$

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

$$a \;=\; \frac{n}{2} + a_0 \tag{26}$$

and

$$b \;=\; \frac{1}{2}\sum_{i=1}^{n}(g_i - \widehat{g})^2 + b_0. \tag{27}$$

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 hyperparameters a_0 and b_0 into just one hyperparameter, by effectively imposing the additional conditions

$$a_0 = \frac{m}{2} \tag{28}$$

and

$$b_0 = \frac{m}{2}\widehat{V}, \tag{29}$$

where the single hyperparameter 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 hyperparameters: δ and m.

To summarize here, with the unbiased-prior assumption (28), (29) the posterior distribution of the precision w is given by the probability density function

$$\psi_n(w \mid \delta, m) \;=\; k_\delta \, w^{a-1} \, e^{-bw}, \tag{30}$$

for $w \geq \delta$, while $\psi_n(w \mid \delta, m) = 0$ for $w < \delta$. The constants k_δ, a and b in (30) are defined by the equations

$$1/k_\delta \;=\; \int_\delta^\infty w^{a-1} \, e^{-bw} \, dw, \tag{31}$$

$$a \;=\; \frac{n+m}{2}, \tag{32}$$

$$b \;=\; \frac{n+m}{2}\widehat{V}. \tag{33}$$

From combining (30) with (23), the unconditional or marginal posterior probability density function of g is

$$f_n(g \mid \delta, m) \;=\; k'_\delta \int_\delta^\infty \exp(-(g-\widehat{g})^2 nw/2(n+1)) \, w^{a-\frac{1}{2}} \, e^{-bw} \, dw, \tag{34}$$

where a and b are defined by (32) and (33), while k'_δ is a constant of integration satisfying

$$1/k'_\delta \;=\; \int_{-\infty}^\infty dx \int_\delta^\infty \exp(-x^2 nw/2(n+1)) \, w^{a-\frac{1}{2}} \, e^{-bw} \, dw. \tag{35}$$

The two non-negative hyperparameters δ and m are highlighted in formulas (30) and (34) merely to remind us that (among many other things, such as $\widehat{V}$ and n, which offhand seem intuitively 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 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 (34), (35) reduces to the t distribution

$$f_n(g \mid 0,0) \;=\; \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\,\sqrt{\pi \widehat{V} n}} \left[1 + \frac{(g-\widehat{g})^2}{(n+1)\widehat{V}}\right]^{-\frac{n+1}{2}}, \tag{36}$$

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 hyperparameters out of the posterior distribution. Intuitively, we might say that if the number of data points n is large enough, "it should not matter now what values of δ or m we select to represent prior beliefs."

In the setup of this model, the formula for expected marginal utility 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 hyperparameter m, whose effects on expected marginal utility *converge uniformly in n for all $m > 0$*. However, the hyperparameter δ behaves fundamentally differently, because its effects on the moment generating function of g – and therefore on expected marginal utility – *do not converge uniformly in n for all $\delta > 0$* (essentially because the moment generating function of (36) is infinite). In this sense there is an absolutely critical distinction, which is crucial for all marginal-utility-weighted asset-pricing expectation formulas, between not knowing what value to assign now to the prior hyperparameter m and not knowing what value to assign now to the prior hyperparameter δ.

The fact that expected utility is not uniformly convergent in n for all positive δ has enormous significance for the family of equity puzzles. 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 a hypothetical 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 posterior robustness, 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. Non-uniform convergence in stochastic-discount-factor space means that the fickle whimsicality of current investors concerning what value of the structural hyperparameter δ to select for representing the model's initial configuration never loses its potential ability to dominate posterior 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.

A necessary precondition for the validity here of the classical frequentist idea to "just let the data speak for themselves" is that the effects of δ or m on the posterior distribution of g should be negligible for sufficiently large n. This condition holds (in the space of subjectively-expected utility) for m, but such a posterior 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 become invisible 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.

Taking (30) and (34) as subjective posterior probability density functions, we are now ready to compute the Bayesian equity premium, the Bayesian risk-free interest rate, and Bayesian equity variability. 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, only the model never quite gets to a stationary distribution because some very small (but consequential for asset pricing) amount of learning takes place whenever $m < \infty$. When m is "very big" but less than infinity (one quadrillion is

a nice round number to keep in mind), the subjective Bayesian distribution of future growth rates is essentially unchanged by the arrival of a new datum point. Such a limiting situation where $10^{15} \leq m < \infty$ nullifies sampling error or learning effects and focuses the mind very 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 $\Pi(\delta)$ represent the value of $\ln E[R_e] - r_f$ as a function of δ that is obtained from formula (13) when the probability density function of g is $f_n(g \mid \delta, m)$ defined by equation (34). Plugging the subjective posterior distribution (34) into formula (13), we obtain

$$\Pi(\delta) \;\equiv\; \ln \frac{\int_{-\infty}^{\infty} e^{g}\, f_n(g \mid \delta, m)\, dg \;\int_{-\infty}^{\infty} e^{-\gamma g}\, f_n(g \mid \delta, m)\, dg}{\int_{-\infty}^{\infty} e^{(1-\gamma)g}\, f_n(g \mid \delta, m)\, dg}. \tag{37}$$

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*

$$\ln E[R_e] - r_f \;=\; \Pi(\delta_e). \tag{38}$$

Proof. Using (23) and the formula for the expectation of a lognormal random variable rewrite (37) (after cancelling terms in $\widehat{g}$) as

$$\Pi(\delta) \;=\; \ln \frac{\int_{\delta}^{\infty} e^{(n+1)/2nw}\, \psi_n(w \mid \delta, m)\, dw \;\int_{\delta}^{\infty} e^{(n+1)\gamma^2/2nw}\, \psi_n(w \mid \delta, m)\, dw}{\int_{\delta}^{\infty} e^{(n+1)(1-\gamma)^2/2nw}\, \psi_n(w \mid \delta, m)\, dw}. \tag{39}$$

It is readily apparent from examining (30) that as $\delta \to 0^+$, all three integrals in (39) approach $+\infty$. Therefore,

$$\lim_{\delta \to 0} \Pi(\delta) \;=\; \lim_{w \to 0} \ln \frac{\exp((n+1)/2nw)\;\exp((n+1)\gamma^2/2nw)}{\exp((n+1)(1-\gamma)^2/2nw)}. \tag{40}$$

Since

$$\ln \frac{\exp((n+1)/2nw)\ \exp((n+1)\gamma^2/2nw)}{\exp((n+1)(1-\gamma)^2/2nw)} \ = \ \frac{(n+1)\gamma}{nw}, \tag{41}$$

plugging (41) into (40) gives

$$\Pi(0) \ = \ \lim_{w\to 0}(n+1)\gamma/nw \ = \ +\infty. \tag{42}$$

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 (37) is continuous in δ. Since

$$\Pi(\infty) \ < \ \ln E[R_e] - r_f \ < \ \Pi(0), \tag{43}$$

condition (38) follows. ■

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 (34) is a t distribution of the form (36) – 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 (34) 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 to wreak 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 (34) is capable in principle of producing an explosion in formulas like (10) or (13), implying in the limit as $\delta \to 0^+$ an unboundedly large equity premium. In this normal-gamma Bayesian framework, the statistical fact that the moment generating function of a t-distribution is infinite has the important economic interpretation that, at least in principle, model-structure uncertainty has the potential to be a far more important determinant of asset prices 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 (whose structural parameters are *unknown* and must be estimated) than is this agent averse to the "weak force" of the pure risk per se of being exposed to the same underlying stochastic growth process (except with *known* structural parameters).

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 variability that could leave them holding the bag by wiping out their stock-market investments. This eerie sensation of low-δ background structural shadow-risk may not be simple to articulate, yet it frightens investors away from taking a more aggressive stance in equities and scares them into a position of wanting to hold instead 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 with standard parameter values. But *compared to what* is the observed value of $\ln E[R_e] - r_f$ "too big"? The answer given in the equity-premium literature is: "compared to 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 (13). 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 stationary 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 calibrated model, because such an event is not in the sample being used for the calibration.

Theorem 1 is trying to tell us that the statistical architecture of 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 *any* curved utility function $U(C)$ having the innocuous curvature property that $-CU''(C)/U'(C) > 0^+$ for all $C > 0$. Bayesian inferences from finite data fatten the posterior tails of probability density functions with dramatic consequences when expressed in subjective-expectation units of

future marginal utility – as the example of replacing the workhorse normal distribution by its t-like posterior distribution demonstrates. This "Bayesian-statistical peso problem" means that 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. The Bayesian peso problem is effectively saying 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 utility-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 extraordinarily 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 expectational uncertainty.

The important result in Schwarz (1999) can be interpreted as saying that, for essentially any reasonably-specified 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. It means that there is a generic sense in which 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 are they averse to the purely-stochastic growth-rate risk itself. When investors are modeled as perceiving and acting upon these inevitably-fat-tailed subjective posterior distributions, then a fully-rational general equilibrium interpretation integrates 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 (10) when the probability density function of g is $f_n(g \mid \delta, m)$ defined by equation (34). Plugging the subjective posterior distribution (34) into the right hand side of equation (10), the result is

$$\Phi(\delta) \;\equiv\; \rho - \ln \int_{-\infty}^{\infty} \exp(-\gamma g)\; f_n(g \mid \delta, m)\; dg. \tag{44}$$

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*

$$r_f = \Phi(\delta_f). \tag{45}$$

Proof. Use (23) and take the expectation of a lognormal distribution to rewrite (44) as

$$\Phi(\delta) \;=\; \rho + \gamma\widehat{g} - \ln \int_{\delta}^{\infty} \exp((n+1)\gamma^2/2nw)\; \psi_n(w \mid \delta, m)\; dw. \tag{46}$$

From examining (30), it is apparent that as δ is made to approach zero, the integral on the right hand side of (46) becomes unbounded. Therefore, $\Phi(0) = -\infty$. At the other extreme of δ is the deterministic Ramsey formula $\Phi(\infty) = \rho + \gamma\widehat{g}$. Thus,

$$\Phi(0) < r_f < \Phi(\infty), \tag{47}$$

and, since $\Phi(\delta)$ defined by (44) is continuous in δ, the conclusion (45) 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 an intense aversion to high-volatility low-precision situations. The bottom line once more is that a "Bayesian peso problem" causes false inferences about expected future utility, which are based upon mimicking 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 Normal-Equivalent Growth and Equity Variability

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 issues previously swept under the rug. We begin with the standard definition

$$r_e(g) \equiv \ln R_e(g), \tag{48}$$

and then apply (2) and (11) to obtain

$$r_e - E[r_e] = g - E[g]. \tag{49}$$

In this ultra-stark tree model with *i.i.d.* fruit and isoelastic utility, the only genuine "fundamental" of the system is the growth rate g. Stocks are being understood here as effectively representing one-period-ahead unit claims on future realizations of g in the form of the consumption payoff function $\exp(g)$. The idea that stock-market returns should vibrate consistently with the vibrations of their underlying "fundamentals" finds expression in this bare-bones macro-model by equation (49). A further decomposition of the one-period-ahead equity payoff-return into conventional dividends plus capital gains – with capital gains contributing most of the actual vibrating – can fill in some interesting and very useful supplementary details but, be that as it may, any dynamic generalization of this basic model (whether it be overlapping generations or infinitely-long-lived agents, stationary or non-stationary structure, with or without rational bubbles) should exhibit the same one-period-ahead properties as the simple reduced-form two-period specification of this paper.

According to (49), for an economy-wide comprehensive stock index, which embodies an implicit claim on the future aggregate output of the underlying real economy, the second order (and higher) central moments of g and r_e should coincide. Alas, the second moments of g and r_e are not even remotely equal in the time-series data. For the super-simple fruit-tree model of this paper, the "equity variability 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 can be interpreted as essentially posing sharply and answering precisely the basic question: between these two alternatives, *which one of the two empirically-observed variabilities* ($\widehat{\sigma}[r_e]$ standing in for the left hand side of equation (49), or $\widehat{\sigma}[g]$ standing in for the right hand side) *better represents the true welfare situation* (of the representative risk-averse agent, who is uncertain about what to expect from the unknown structure of future growth prospects)?

Because $\widehat{\sigma}[r_e] \gg \widehat{\sigma}[g]$, clearly expression (49) fails empirically to be a literal equality between random variables in an ex-post realized-frequency sense. The next question is: might

an equation *like* (49) be holding more figuratively here in some welfare-equivalent ex-ante sense of perceived subjective expectations? After all, 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 (and high-g) extreme outliers in the tails that have yet to be experienced. It follows that a subjective expected-utility-equivalent future growth rate specified to be normally distributed with the same mean $\widehat{g}$ as the posterior distribution $f_n(g \mid \delta, m)$ must have a significantly higher as-if-known variance than the realized sample variance from past data.

Suppose, for the sake of argument, that the representative agent feels more comfortable conceptualizing (or the modeler feels more comfortable telling a story about) the uncertainty of future growth prospects in terms of a familiar normal probability distribution. We now examine the connections between equity variability and a welfare-equivalent subjective belief in normally-distributed growth rates. Let

$$g_N \;\sim\; N(E[g_N], \sigma^2[g_N]) \tag{50}$$

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

$$H_N(g_N) \;\equiv\; \exp(g_N) \tag{51}$$

be the *subjective perception* of a stock-market payoff implicitly representing a unit claim on the lognormally distributed future aggregate output corresponding to (50). Such a unit claim gives rise to the *subjective perception* of a (geometrically measured) return on comprehensive economy-wide equity $r_N(g_N)$ satisfying the equation

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

which is exactly the normal counterpart here of (49).

The following representation theorem establishes the existence of a useful form of expected-utility indifference between g_N and g. This last proposition of the paper can be interpreted as providing a basic sense in which there might be *some* rationale for telling an as-if parable wherein the representative agent has a subjective normally-distributed welfare-equivalent

perception, which is consistent with (52), "as if" the future growth rate is g_N with known variability equal to the observed variability of returns on equity. In this as-if-normal interpretation, the true welfare situation is represented by the relatively high variability of returns on equity rather than by the relatively low variability of realized past growth rates. Because here $\sigma[r_N] = \sigma[g_N]$, at least from within the framework of this welfare-equivalent as-if-normal parable there is no longer a 'puzzling' mismatch of variabilities wanting to be explained.

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:*

$$E[U(C_0 \exp(g_N))] \;=\; E[U(C_0 \exp(g))], \tag{53}$$

$$E[g_N] \;=\; E[g] \;=\; \widehat{g}, \tag{54}$$

$$E[r_N] \;=\; E[r_e], \tag{55}$$

$$\sigma[r_N] \;=\; \sigma[g_N] \;=\; \widehat{\sigma}[r_e]. \tag{56}$$

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

$$\frac{1}{\sqrt{2\pi}S(\delta)} \int_{-\infty}^{\infty} \exp((1-\gamma)g_N - (g_N - \widehat{g})^2/2S(\delta))dg_N = \int_{-\infty}^{\infty} \exp((1-\gamma)g) f_n(g \mid \delta, m)dg \tag{57}$$

and note that for this definition (53) and (54) are satisfied by construction.

From plugging (11) into (6) into (2) and then taking logs,

$$r_e \;=\; g + \{\rho - \ln E[\exp((1-\gamma)g)]\}, \tag{58}$$

and, from plugging (51) into (6) into (2) and then taking logs,

$$r_N \;=\; g_N + \{\rho - \ln E[\exp((1-\gamma)g_N)]\}, \tag{59}$$

so that (55) then follows from (54), (57), (58), (59).

For $\delta = \infty$, the integral on the right hand side of equation (57) becomes $\exp((1-\gamma)\widehat{g})$, implying $S(\infty) = 0$. As $\delta \to 0$, the integral on the right hand side of equation (57) becomes unbounded, implying $S(0) = \infty$. Thus,

$$S(\infty) \;<\; \widehat{\sigma}[r_e] \;<\; S(0), \tag{60}$$

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

$$S(\delta_N) \;=\; \widehat{\sigma}[r_e], \tag{61}$$

which, when combined with (52), proves (56) 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 the model of this paper, the standard deviation of normally-distributed equity returns is presumed to be known by the agents, 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 generating a future time-series data sample from the prototype limiting case of the model where *m is extremely big* (but less than infinity, say one quadrillion), 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 (but not exactly) unchanged as new data arrive over time.

With the above near-stationary setup, the data generating process for future g is essentially a simulation of the t-like distribution (34), meaning that with very large m the generated growth rates are statistically indistinguishable from a simulation of *i.i.d.* draws from the distribution $g \sim N(\widehat{g}, \widehat{\sigma}^2[g])$. Simultaneously in this thought experiment, agents are maintaining a subjective perception of the future *as if* $g_N - \widehat{g}$, which is known to be normally distributed, is statistically indistinguishable from a linear amplification of $g - \widehat{g}$ by the magnifying factor $\widehat{\sigma}[r_e]/\widehat{\sigma}[g]$. Consistent with (52), the data generating process for future r_N is essentially a simulation of random draws from a known normal distribution, so that, except for sampling errors, $r_N \sim N(\widehat{r}_e, \widehat{\sigma}^2[r_e])$. Therefore, in this thought experiment the numerically-generated excess volatility of equity will simulate statistically $\widehat{\sigma}[r_e] - \widehat{\sigma}[g]$. Such a discrepancy in variabilities seems almost mind-bogglingly contradictory, however, because year after year of new realizations from the above data generating process are confirming that g appears *as if* it is normally distributed with standard deviation $\widehat{\sigma}[g]$, yet rational agent-investors in this economy nevertheless appear to an outside observer to be clinging stubbornly in their mind's eye to an unshakably-consistent irrational mental image *as if* their future

Quasi-Constant Parameter	**Value**
Mean arithmetic return on equity	$\ln E[R_e] \approx 7\%$
Geometric standard deviation of return on equity	$\sigma[r_e] \approx 17\%$
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 Economic "Stylized Facts"

welfare depends upon the realization of a much more variable normally-distributed growth rate, whose standard deviation is $\sigma[g_N]$ $(= \widehat{\sigma}[r_e] \gg \widehat{\sigma}[g])$.

7 Some Bayesian Calibration Exercises

Viewing the three partial-equilibrium theorems of the paper through the lens of the welfare-equivalent as-if-normal growth story of Theorem 3 gives an analytically-tractable relationship among Π, Φ, and S implied by the lognormality simplification. The following question then arises naturally: is there a reasonable sense in which approximately the same value of the hyperparameter δ comes close to matching simultaneously the actually-observed values of the equity premium, risk-free rate, and equity volatility? In other words, can the three degrees of freedom represented by $\Pi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ be essentially explained by the one degree of freedom represented parsimoniously by δ in this theory? The answer is basically "yes," which we will now proceed to show. The experimental outcome that all three empirically-observed values of the equity premium, risk-free rate, and equity variability come close to matching simultaneously their theoretically-predicted relationship conveys some overall sense of the degree to which the model is a relatively tight-fitting theoretical-empirical construct.

The proposed exercise will test whether the welfare-equivalent interpretation of Theorem 3 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 (52), an internally-consistent as-if story connecting together the actual stylized facts of our economic world. In Table 1, quasi-constant parameter settings have been selected that, I think, represent stylized-fact numbers well within the "comfort zone" for most economists. All rates are real and given by annual values. The data are intended to be an overall approximation of what has been observed for many countries over long time periods.

The model is explaining endogenously three quasi-constants $\Pi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ as

functions of the one free hyperparameter 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 $\Pi(\delta)$, $\Phi(\delta)$, and $S(\delta)$ equal to its observed value in Table 1 and then backing out the implied values of the other two remaining quasi-constants by inverting the two as-if-normal-growth reduced-form equations (17) and (18).

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 $\widehat{V} \equiv S^2(\delta_N)$,

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

to be compared with $\Pi(\delta_e) = 6\%$. From (18) with $\widehat{V} \equiv S^2(\delta_N)$,

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

to be compared with $\Phi(\delta_f) = 1\%$.

Defining δ_e to be the implicit solution of

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

we then have, from (18) and (17),

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

to be compared with $\Phi(\delta_f) = 1\%$. From (17) with $\widehat{V} \equiv \sigma^2[r_e]$,

$$\sigma[r_e] \;=\; \sqrt{\Pi(\delta_e)/\gamma} \;=\; 17\% \;=\; S \mid \delta_e,$$

to be compared with $S(\delta_N) = 17\%$.

Defining δ_f to be the implicit solution of

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

we then have, from (18) and (17),

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

to be compared with $\Pi(\delta_e) = 6\%$. From (18) with $\widehat{V} \equiv \sigma^2[r_e]$,

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

to be compared with $S(\delta_N) = 17\%$.

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

Continuing with the above as-if-normal-growth scenario, consider next 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 of certain growth is equivalent in expected utility to a change in the trend rate of

$$\Delta g = \frac{\gamma - 1}{2}\,\sigma^2[g]. \tag{62}$$

When $\gamma \approx 2$ and the historical value of $\sigma[g] \approx 2\%$ is used in (62), then $\Delta g \approx 0.02\%$, which is roughly the order of magnitude of numbers widely circulated as indicating that the cost of growth variability is so low that even a complete removal of macroeconomic uncertainty would 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 (62), 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 "equity variability puzzle," a "risk-free rate puzzle," and an "equity premium puzzle," whose magnitudes of discrepancy are 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 ob-

served in the data are telling an internally-consistent 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 variability of the subjective future growth rate *as if* it were equal to the observed variability 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] Lewellen, Jonathan, and Jay Shanken (2002), "Learning, Asset-Pricing Tests, and Market Efficiency." *Journal of Finance*, 57, 1113-1145.

[17] 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.

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

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

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

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

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

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