

Non-parametric distributions, uncertainty, and asset prices

Viktor Tsyrennikov* Heidi Verheggen†

January 7, 2015

Abstract

We estimate the process of the U.S. consumption growth non-parametrically. Agents in the model behave as if they had pessimistically distorted beliefs. If we keep the strength of the distortion constant, then the estimated nonparametric distribution has a substantial edge over the normal distribution in matching the facts about asset returns. But the same distortion applied to the nonparametric distribution produces a less realistic belief, *i.e.* that is farther away from the original distribution. The advantage of the nonparametric distribution then disappears once we insist that agents' beliefs remain close to the original distribution.

Keywords: model uncertainty, robustness, asset pricing, non-normality

JEL classification: G11, G12

1 Introduction

The distribution of consumption growth is one of the key factors determining asset returns. The empirical distribution of the U.S. consumption growth, unlike the normal distribution that is widely used in theoretical models, is

*email: viktor.tsyrennikov@gmail.com. Corresponding author.

†heidiv@sas.upenn.edu. I would like to acknowledge the support of the National Science Foundation Graduate Research Fellowship Program, Grant No. DGE-1321851.

skewed and fat-tailed. How should the investor facing such a consumption process invest? Because the empirical distribution is skewed towards adverse low-growth outcomes the investor should buy risk-free bonds that guarantee a fixed return independently of the consumption growth rate and he should avoid risky stocks that provide low return precisely in the states when they are needed most, when the consumption growth is low. But could these features of the consumption growth process alone explain the two long-standing asset return puzzles described in Mehra and Prescott [1988]: the low risk free rate of return and the high risk premium?

Two deviations from normality were widely studied by the literature. First, Rietz [1988] and Barro [2006] assume that the consumption growth is subject to rare disaster shocks that are infrequent large decreases. Second, Bansal and Yaron [2004] assume that consumption growth is in part driven by a small but highly persistent component, often referred to as long-run risk. The above assumptions induce caution on the part of model agents, that is they invest more in safe bonds and less in risky equity than they otherwise would. This cautious behavior drives the risk-free rate of return down and the risk premium up. While being successful explanations, both of the distributional assumptions are difficult to verify. For example, Barro [2006] assumes that consumption could decline 29% in any given year with probability 1.7%. This assumption is an extrapolation from a heterogeneous sample of advanced economies, many of which suffered from wars and other periods of instability avoided by the U.S. Could such the disaster event occur in the U.S.? Possibly, but it is very unlikely. Hansen et al. [2008] find that the macroeconomic data is too short to identify the long-run risk component assumed in Bansal and Yaron [2004]. Others consider even more elaborate, thus harder to verify, models of the consumption growth process. Gourio [2012] and Wachter [2012] consider time-varying disaster probabilities. Gabaix [2012] assumes, in addition, that the consumption growth is a “linearity-generating” process: keeping normally distributed innovations he breaks the linearity of the growth rate process. Weitzman [2007] shows that Bayesian learning can lead to a thick-tailed posterior distribution. Cecchetti et al. [2000] assume that agents consider an oversimplified model of consumption growth and for this reason they form wrong, pessimistic given the

U.S. historical data, forecasts. Finally, Martin [2013] studies how different cumulants of the consumption growth process affect asset returns.

Unlike the above studies, we estimate the distribution of the consumption growth rate non-parametrically from over a century worth of the U.S. data. It is still assumed to be a linear auto-regressive process, but no assumptions are imposed on the distribution of the growth innovations. The estimated non-parametric distribution has a minor mode in the left tail. That is, there is evidence of large consumption contractions, as hypothesized by Rietz and Barro, but not as large as they assume. This does not mean, however, that agents did not expect larger decreases in consumption. But how far could an agent's expectations and the reality reasonably be? In the model that we describe agents' expectations differ consistently from the true distribution. The discrepancy between the two is endogenously restricted by an entropy constraint. The latter determines the set of plausible distributions, or beliefs, that an agent could assume. We then use the detection error probability (DEP) concept advocated by Hansen and Sargent [2007] to guide the choice of this set of reasonable beliefs. The resulting decision making model has a simple representation – the “multiplier” preferences. This preference specification is part of a larger class of Kreps and Porteus [1978] and Epstein and Zin [1989] utility functions. Because of this relation, the parameter that governs the tightness of the entropy constraint can be also interpreted as risk aversion to intratemporal fluctuations in wealth as in Tallarini [2000]. For this reason our paper is closest in spirit to Barillas et al. [2009] who exploit the same relation albeit assuming normally distributed consumption growth rate innovations.

We show that if the asset returns are computed as a function of the unobserved risk aversion parameter then deviating from the normality assumption improves the model predictions substantially. The reason is that with a left-skewed distribution of the consumption growth, agents behave in a more cautious fashion. Thus, they choose to hold safe bonds even if its return is low and demand a premium for holding risky equity. When we return to the original interpretation and compute returns for a given level of DEP, then the advantage of the non-parametric model disappears. This is because deviations from the non-parametric distribution are easier to detect

using statistical methods. That is, the entropy constraint becomes tighter and this limits the amount of pessimism that can be reasonably built into an agent’s behavior. And without pessimism asset return predictions are farther from the reality.

The paper also develops several useful analytical expressions that relate asset returns to the detection error probability. This can be done, though, only for the case in which the growth innovations are normally distributed. We describe how to perform the calculation for an arbitrary distribution numerically. We conclude with a speculation of what could occur in settings with time-varying pessimism.

2 Model

2.1 Static model

We first describe a static decision problem that motivates the objective function used in the dynamic model. Payoff $V(z|d)$ depends on a decision maker’s (DM) decision d and state $z \in \mathcal{Z}$ that is a realization of a random variable Z . Let $f(z)$ denote the DM’s subjective probability of outcome $z \in \mathcal{Z}$. This is his model of the world. However, the DM is concerned that f may not be the true distribution. These concerns may arise because the data is finite, inaccurate or incomplete.

The DM could test his distribution f . The best way to proceed is to use the likelihood-ratio (LLR) test, which is the most powerful test. Using the LLR test, the DM could build a set of distributions that are “reasonably” similar. This set can be conveniently summarized using the Kullback-Liebler divergence or entropy. *Entropy* of distribution f_1 relative to f_2 is defined as follows:

$$E(f_1, f_2) = \int \ln(f_1(z)/f_2(z))f_1(z)dz.$$

The higher the entropy is, the more different f_1 and f_2 are. It is non-negative, its minimal value is 0 and it is achieved when $f_1 = f_2$.

An agent who doubts his distribution f evaluates each action d according

to the following criterion:

$$\min_{\tilde{f}} \int V(z|d)\tilde{f}(z)dz \quad (1a)$$

where the minimization is subject to the entropy constraint:

$$\int \ln(\tilde{f}(z)/f(z))\tilde{f}(z)dz = E(\tilde{f}, f) \leq \eta. \quad (1b)$$

The entropy constraint determines the set of distributions that the agent is willing to consider “similar.” The agent then acts as if the worst possible distribution in this set were the true distribution.

To see where this takes us, we need to solve the above optimization problem. So, let θ^{-1} be the Lagrange multiplier on the entropy constraint. Then the first-order optimality condition for $\tilde{f}(z)$ is:

$$V(z|d) - \theta^{-1}(\ln(\tilde{f}(z)/f(z)) + 1) = 0.$$

The above and the fact that \tilde{f} must integrate to one imply:

$$\tilde{f}(z) = f(z) \frac{\exp(-\theta V(z|d))}{\int \exp(-\theta V(s|d))f(s)ds}.$$

The resulting objective is:

$$\int V(z|d)\tilde{f}(z)dz = \ln \int \exp(-\theta V(z|d))f(z)dz. \quad (2)$$

The parameter θ measures an agent’s concern for model mis-specification, *i.e.* the degree to which an agent doubts the distribution f . It can be obtained from the entropy constraint. But because there is a one-to-one (positive) relation between θ and η either of the two can be used.

The expression in (2) measures the DM’s welfare for any action d that he could choose. We say that the DM has the *multiplier* preferences. Had we left the objective as in (1) we would say that the DM has the *constraint* preferences. The two differ because of the following singularity: if the payoffs were constant, then the multiplier preferences would select $\tilde{f} = f$, while any \tilde{f} in the “admissible” set yields the same utility as would be recognized by the constraint preferences. Luckily, this never occurs in the setting that we are about to describe.

2.2 Dynamic model

We now describe the dynamic model environment. Time is indexed by $t = 0, 1, 2, \dots$. The state of the economy in period t is a first-order Markov process $g_t \in \mathbb{R}$. The initial state of the economy g_0 is known. We use g^t to denote the history of the state up to period t : g_0, g_1, \dots, g_t . The aggregate output in the economy is denoted by $Y(g^t)$. Its evolution is summarized by the following equation:

$$Y(g^{t+1})/Y(g^t) = e^{g_{t+1}}, \quad Y(g_0) \equiv 1. \quad (3)$$

We now make distributional assumptions about g_t :

$$\ln(g_t) - \mu = \rho(\ln(g_{t-1}) - \mu) + \varepsilon_t, \quad \text{where } \varepsilon_t \sim iid F.$$

Distribution F is known, and in our baseline specification we assume that it is a normal distribution.

The representative agent ranks different consumption plans $\{c_t(g^t)\}$ using the following recursive utility function:

$$V(g^t) = \ln(c(g^t)) - \frac{\beta}{\theta} \ln E [\exp(-\theta V(g^{t+1}))|g^t], \quad \beta \in (0, 1). \quad (4)$$

One should recognize the multiplier preferences from (2) embedded in the above. The parameter θ measures an agent's concern for model misspecification. As θ increases an agent doubts his distribution f more and contemplates a larger set of alternative models. When $\theta \rightarrow 0$ a standard time separable preference specification obtains: $V(g^t) = \ln(c(g^t)) + \beta E [V(g^{t+1})|g^t]$. That is, an agent's utility is a sum of the current period utility and the discounted expected future utility. As mentioned above, the parameter θ can be also interpreted as an agent's risk aversion. If dispersion of the agent's future consumption increases, his utility decreases in proportion to θ . An interested reader may consult Backus et al. [2004] for a detailed discussion of different preference specifications, interpretation of their parameters, and relation between them.

Financial markets are dynamically complete. That is after any history g^t the financial markets trade a continuum of securities that are contingent on possible realizations of the growth rate tomorrow g_{t+1} . Thus a security g'

pays one unit of consumption only if the growth rate tomorrow is $g_{t+1} = g'$. To illustrate, by purchasing all the securities $g' \leq E(g_{t+1})$ an agent insures himself against the event that the growth rate is at or below the expected value. Let the price of security g' after history g^t be denoted by $Q(g^t, g')$.

Let $a(g^t, g')$ be the quantity of security g' that an agent purchases after history g^t . Then an agent must satisfy the following budget constraint:

$$c(g^t) + \int_{g'} Q(g^t, g') a(g^t, g') dg' = y(g^t) + a(g^{t-1}, g_t). \quad (5)$$

It states that the value of consumption and security purchases must equal current income and the payoff from securities purchased in the previous period.

Notice that apart from the exogenous income and the endogenous choices of an agent the only variable that enters the budget constraint is his financial wealth $a(g^t, g')$. Thus, in what follows we replace state history with a pair $(a(g^t), g_t)$. These are the only two variables needed to make an informed decision. The financial wealth determines what purchases are feasible. The current state, not history because it is a first-order Markov process, is needed to compute expected values. For this reason we use $V(a(g^t), g_t)$ instead $V(g^t)$ to denote the life-time utility. We only use this notation temporarily to explain the optimization problem of an agent and to derive the equilibrium prices.

Consider now the optimization problem of an agent with financial wealth $a(g^t)$ when the current state of the economy is g_t . An agent, subject to the budget constraint (5), chooses current consumption and a portfolio of securities to maximize his life time utility starting from g^t :

$$V(a(g^t), g_t) = \max_{c(g^t), a(g^t, g'), \forall g'} \left\{ \ln(c(g^t)) - \frac{\beta}{\theta} \ln \left(\int_{g'} \exp(-\theta V(a(g^t, g'), g')) dF(g'|g_t) \right) \right\}.$$

This is the Bellman equation associated with an agent's optimization problem.

The first-order optimality conditions and the envelope condition for the above optimization problem are provided in the appendix B. Combining the

two we derive the equilibrium price $Q(g^t, g')$:

$$Q(g^t, g') = \frac{\beta}{g'} \tilde{f}(g'|g_t), \quad (6)$$

where

$$\tilde{f}(g'|g^t) \equiv \frac{\exp\{-\theta V(a(g^t, g'), g')\} f(g'|g^t) dg'}{\int_{s'} \exp\{-\theta V(a(g^t, s'), s')\} f(s'|g^t) ds'}. \quad (7)$$

Observe that $\tilde{f} \equiv d\tilde{F}$ is a distribution function because $\int_{g'} d\tilde{F}(g'|g_t) = 1$. It is referred to as distorted conditional distribution of the economy's state because it relocates probability mass ("distorts") towards adverse outcomes. The price $Q(g^t, g')$ still depends on the unknown value function.

To compute V , first, impose that the representative agent's consumption must equal the total output produced: $c(g^t) = Y(g^t), \forall g^t$. Lemma 2 in the appendix shows that function V can be factored into two components: $V(g^t) = W(g_t) + (1 - \beta)^{-1} \ln Y(g^{t-1})$. Then the distorted density can be computed using W only:

$$\tilde{F}(g'|g_t) \equiv \frac{\exp\{-\theta V(g^t, g')\} dF(g'|g^t)}{\int_{s'} \exp\{-\theta V(g^t, s')\} dF(s'|g^t)} = \frac{\exp\{-\theta W(g')\} dF(g'|g_t)}{\int_{s'} \exp\{-\theta W(s')\} dF(s'|g_t)}. \quad (8)$$

Notice that the distribution \tilde{F} is conditioned on g_t not g^t as the right-hand side depends only on g_t .

The value function W must satisfy the following recursive equation:

$$W(g_t) = \ln g_t - \frac{\beta}{\theta} \ln E [\exp(-\theta W(g_{t+1})) | g^t]. \quad (9)$$

That the above equation has a unique solution can be shown as follows. Let $T(W)(x) \equiv \ln(x) - (\beta/\theta) \ln E[\exp(-\theta V(x))], \forall x$ be the mapping defined by the recursive equation. It is easy to show that it satisfies Blackwell's conditions, see [Stokey et al., 1989, p.54]. So, it defines a contraction mapping on the space of continuous functions. If we restrict the support of the growth rate, as has to be done in the numerical analysis, then we can narrow our attention to the space of continuous and bounded functions. The latter is a complete metric space and according to the contraction mapping theorem, see [Stokey et al., 1989, p.50], there is a unique W satisfying the recursive equation.

According to lemma 3 in the appendix W must take the following form:

$$W(g_t) = A + \frac{\ln g_t}{(1 - \beta)(1 - \beta\rho)}. \quad (10)$$

where A is a positive constant. The value of A is irrelevant for computation of \tilde{F} . Combining (8) and (10), and assuming that f is a normal density gives:

$$\tilde{f}(g'|g_t) = f(g'|g_t) \cdot \exp \left\{ -\frac{(1 - \beta)(1 - \beta\rho)\theta\varepsilon + 0.5\theta^2\sigma^2}{[(1 - \beta)(1 - \beta\rho)]^2} \right\}, \quad (11)$$

where $\varepsilon = g' - (1 - \rho)\mu - \rho \ln g_t$. The representative agent believes that the conditional distribution of g' is:

$$\tilde{f}(g'|g_t) = N \left((1 - \rho) \left[\mu - \frac{\theta\sigma^2}{(1 - \beta)(1 - \beta\rho)} \right] + \rho \ln g_t, \sigma^2 \right). \quad (12)$$

Equivalently, the representative agent believes that growth innovations ε_{t+1} are distributed according to $N(-\frac{\theta\sigma^2}{(1-\beta)(1-\beta\rho)}, \sigma^2)$ while under the true distribution their mean is zero. In other words, he is *pessimistic* and makes his decisions in a way that would deliver a satisfactory utility level even if his pessimistic predictions turned out to be correct. This also means that he would achieve a higher level of life-time utility had he not doubted the consumption growth process. As was noted before, the degree of concern or precaution is governed by θ . The larger the θ the more pessimistic are the agent's implied beliefs. Notice also that uncertainty is disabled when there is no risk, i.e. when $\sigma^2 = 0$. The pessimism is also stronger when the persistence of the growth rate is higher, for then the long-run variance of the growth rate is higher. When the agent is more patient, that is β is higher, future utility becomes more important and this causes the agent to act more cautiously/pessimistically. Finally, we point out that the agent distorts only the mean, but not the variance. This is a special property of normally distributed risks. We will see later that allowing for non-normality results in a more complicated distortion pattern.

2.3 Asset returns

Our interest is in two assets. The first, that we refer to as equity, is a claim to the aggregate output in the economy. The second is the risk-free bond. The

payoffs of these two assets can be mimicked by appropriate combinations of the traded securities, and so their prices can be determined. The risk-free bond pays one unit in every state tomorrow and its payoff can be replicated by a portfolio consisting of one unit of each traded security. The bond price, denoted by $q^b(g^t)$, must equal the value of this portfolio:

$$q^b(g^t) = \int_{g'} Q(g^t, g') \cdot 1 \cdot dg' = \tilde{E}[\exp(-\ln g')]. \quad (13)$$

An equity claim pays $Y(g^{t+s})$ units of good in periods $t + s$, $\forall s \geq 1$ and its payoff can be replicated by a portfolio of $q^e(g^t, g') + Y(g^t, g')$ units of each security g' . The value of this portfolio would allow buying an equity claim and its dividends tomorrow. Its price is:

$$q^e(g^t) = \int_{g'} Q(g^t, g')(q^e(g^t, g') + Y(g^t, g'))dg'. \quad (14)$$

The value of the equity claim trends together with the value of the aggregate output. For this reason, we instead compute the price-dividend ratio $\omega(g_t) \equiv q^e(g^t)/Y(g^t)$ that is stationary:

$$w(g_t) = \int_{g'} Q(g^t, g')(w(g') + 1)g'dg'. w(g_t) = \int_{g'} Q(g^t, g')(w(g') + 1)g'dg'. \quad (15)$$

The price-dividend ratio depends only on the current state and not on the history because g_t is a first-order Markov process. For the current combination of the preference specification and the consumption growth process it is in fact constant. The following constant price-dividend function solves the above recursive equation:

$$\omega(g_t) = \bar{\omega} \equiv \beta/(1 - \beta).$$

That is, the price of an equity claim equals $\bar{\omega}Y(g^t)$ in any period t and any history g^t . Once the bond price and the equity price-dividend ratio are determined, the expected returns can be determined as follows:

$$E[R^b] = E \left[\frac{1}{q^b(g^t)} \right] = \frac{1}{\beta \tilde{E}[1/g']}, \quad (16a)$$

$$E[R^e] = \frac{\bar{\omega} + 1}{\bar{\omega}} E[g'] = \frac{E[g']}{\beta}. \quad (16b)$$

Notice that the true expected returns are computed, as they have to be compared with the averages observed in the data. The latter were generated from the true data-generating process f , not the pessimistic belief \tilde{f} of the agent. Also observe that the asset returns depend only on two expectations: $\tilde{E}[1/g']$ and $E[g']$.

The assumption that the growth innovations are normally distributed allows us to compute the returns analytically:

$$E[R^b] = \beta^{-1} \exp \left\{ \mu - \frac{\sigma^2 \theta}{(1-\beta)(1-\beta\rho)} + 0.5\sigma^2 \left(\frac{2\rho^2 - 1}{1-\rho^2} \right) \right\},$$

$$E[R^e] = \beta^{-1} \exp \left\{ \mu + \frac{0.5\sigma^2}{1-\rho^2} \right\}.$$

Using the first-order Taylor series approximation, $\exp(x) - 1 \approx x$ if $x \approx 0$, we get:

$$E[r^b] = E[R^b - 1] \approx -\ln(\beta) + \mu - \frac{\sigma^2 \theta}{(1-\beta)(1-\beta\rho)} + 0.5\sigma^2 \left(\frac{2\rho^2 - 1}{1-\rho^2} \right),$$

$$E[r^e - r^b] = E[R^e - R^b] \approx \sigma^2 + \frac{\sigma^2 \theta}{(1-\beta)(1-\beta\rho)}.$$

The (net) return on the risk-free bond, $E[r^b]$, is decreasing and the risk premium is increasing in θ because a more risk-averse agent demands more of the safe bond more and less of the risky equity claim.

2.4 Measuring θ : detection error probability

The asset return expressions involve θ , the parameter governing the agent's concern for uncertainty. Choice of θ can be guided by the statistical approach to model selection. A detection-error probability (henceforth DEP) is the probability that the likelihood ratio test fails to identify the true model. With two alternatives at hand, f and \tilde{f} , it is defined mathematically as follows:

$$\text{DEP} = 0.5 \left[\text{prob}(f \text{ is accepted} \mid \tilde{f} \text{ is true}) + \text{prob}(\tilde{f} \text{ is accepted} \mid f \text{ is true}) \right].$$

It is advantageous to cast the asset returns in terms of DEP because we have a better sense of "appropriate" values for DEP as opposed to θ ; see Barillas

et al. [2009]. In either case, the intuition is that a plausibly “rational” DM should not consider easily (statistically) distinguishable alternative models. If we use θ to measure the discrepancy between models, it means θ cannot be too large, but this is vague. Alternatively, if DEP is used, it means DEP cannot be too small. But we have a relatively better sense of appropriate values for DEP: the threshold that is often employed in empirical analysis is 10%; we use it too.

Computation of the DEP when the two alternatives are normal distributions with different means is simple. Assume, without loss of generality, that $\mu_1 > \mu_2$. Consider a sample $\{g_t\}_{t=1}^T$. Under model j the log-likelihood of the data is $l_j = \text{const} - 0.5\sigma^{-2} \sum_{t=1}^T (e_t - \mu_j)^2/T$ where $e_t = \ln g_t - \rho \ln g_{t-1}$. Thus, model 1 is accepted when in fact model 2 is correct if:

$$l_1 - l_2 > 0 \quad \leftrightarrow \quad \frac{1}{T} \sum_{t=1}^T e_t > 0.5(\mu_1 + \mu_2).$$

This result is very intuitive: model j is selected if the observed average is closer to mean μ_j . The average $\frac{1}{T} \sum_{t=2}^T e_t$ is normally distributed. So, the probability of the above event is:

$$\text{prob}(l_1 > l_2 | \text{model 2 is true}) = \Phi(0.5(\mu_1 - \mu_2)\sqrt{T}/\sigma).$$

Similarly:

$$\text{prob}(l_2 > l_1 | \text{model 1 is true}) = \Phi(0.5(\mu_2 - \mu_1)\sqrt{T}/\sigma).$$

Exploiting symmetry of a normal distribution we get:

$$\text{DEP} = \Phi(0.5|\mu_2 - \mu_1|\sqrt{T}/\sigma)$$

We state this result in the following lemma.

Lemma 1. *The DEP for $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ and sample length T is $\Phi(-0.5|\mu_1 - \mu_2|\sqrt{T}/\sigma)$.*

When f is a normal density, then \tilde{f} is also normal, and Lemma 1 can be applied to the two densities. The corresponding means are $\mu_1 = 0$ and

$\mu_2 = -\frac{\theta\sigma^2}{(1-\beta)(1-\beta\rho)}$. This yields the following expression for the DEP of the growth rate process:

$$\text{DEP} = \Phi(-0.5\sqrt{T}\theta\sigma/(1-\beta)(1-\beta\rho)). \quad (17)$$

Equivalently, $\theta\sigma/(1-\beta)(1-\beta\rho) = -2\Phi^{-1}(\text{DEP})/\sqrt{T}$: this relation allows us to express asset returns in terms of EDP rather than θ :

$$E[r^b] \approx -\ln(\beta) + \mu + \Phi^{-1}(\text{DEP})\frac{2\sigma}{\sqrt{T}} + 0.5\sigma^2\left(\frac{2\rho^2-1}{1-\rho^2}\right), \quad (18a)$$

$$E[r^e - r^b] \approx \sigma^2 - \Phi^{-1}(\text{DEP})\frac{2\sigma}{\sqrt{T}}. \quad (18b)$$

In what follows we use $\text{DEP} = 0.10$, that corresponds to an econometrician making a classification mistake every tenth sample, implying that the models are very similar. The discount factor β is set to 0.98. The results for other values of the parameter β are not reported because it has only a second order effect on the equity premium. We estimate the remaining parameters using the data on the U.S. consumption growth data from 1883 to 2013: $(\rho, \sigma) = (0.2654, 0.0378)$. The observed risk-free rate and the equity premium for this period are respectively 0.8% and 5.3%. The model, in turn, predicts that the above should be 2.95% and 1.15%, respectively. That is, the model over-estimates the risk-free rate of return and under-estimates the risk premium. In the model without model uncertainty, that is, with $\text{DEP} = 0.50$, these equal 3.95% and 0.15% respectively. We explore whether the gap between the model and the data can be closed by discarding the unrealistic normality assumption.

3 Non-parametric density

Rather than assuming that the distribution of ε is normal, we estimate it from the data on annual per capita consumption growth rates in the U.S. from 1889 to 2013. This non-parametrically estimated distribution is then used as a true distribution in the calculations of the asset returns. Observe that the “distorted” distribution also changes as it is a transformation of the true distribution.

Non-normality of the consumption growth innovations is apparent in the data, see figure 1. A histogram of the consumption growth rate innovations ε illustrates the much-discussed asymmetric pattern of left skewness and a thick left tail. According to the Kolmogorov-Smirnov test the null hypothesis that the innovations are generated from a normal distribution is strongly rejected. The test rejects the null hypothesis at the 10 percent level, with a p-value of 0.692. But is this departure from normality economically important for asset returns? Relaxing this distributional assumption necessitates a more complex computational procedure, as we can no longer obtain analytical results. The next subsection is devoted to describing this procedure in detail, and the subsection that follows discusses the economic implications.

3.1 Computational Procedure

The computational procedure has three steps. The first step is to compute the non-parametric density estimate and the implied cumulative density function (cdf). The second step is to compute the value function and the distorted distribution. The third step is to compute the relation between the DEP and the risk-aversion parameter θ . The equilibrium price equations (13) and (14) remain unchanged; what changes is how the distorted expectations are computed. The price dividend ratio remains constant and equal to $\beta/(1-\beta)$ even with a non-normal distribution.

3.1.1 Estimation of the distribution of ε

Our method of choice is non-parametric kernel estimation, in particular the Gaussian kernel estimation with the least-squares cross-validation to choose the optimal bandwidth. A kernel estimator of a density function f is:

$$\hat{f}(\varepsilon) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{e_i - \varepsilon}{h}\right),$$

where $k(\cdot)$ is a kernel function, a standard normal density in our implementation. The bandwidth parameter h controls the degree of smoothing: the larger h is, the smoother the estimate of f is. Using statistics terminology, it governs the tradeoff between bias and variance of an estimate. Using a

small value for a bandwidth parameter results in an estimate with small bias but large variance, and vice versa for a large value. The statistical procedure seeks to minimize the mean-squared error that is the sum of bias and variance. The optimal bandwidth parameter balances the two components. We use the least-squares cross-validation method for bandwidth selection as described in Li and Racine [2006], section 1.3. This technique chooses the bandwidth to minimize the integrated squared error of the estimate:

$$h^* = \arg \min_h \int [\hat{f}(\varepsilon|h) - f(\varepsilon)]^2 d\varepsilon = \arg \min_h \int \hat{f}(\varepsilon|h)^2 d\varepsilon - 2 \int \hat{f}(\varepsilon)f(\varepsilon|h)d\varepsilon.$$

In the above the objective can be replaced with the following sample analog:

$$h^* = \arg \min_h \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \bar{k} \left(\frac{e_i - e_j}{h} \right) - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j \neq i}^n k \left(\frac{e_j - e_i}{h} \right).$$

where $\bar{k}(v) = \int k(u)k(v-u)du$ is the twofold convolution kernel. In the case of the Gaussian kernel, as it is assumed here, we have $\bar{k}(v) = \exp(-v^2/4)/\sqrt{4\pi}$.

Applying this estimation method to the consumption growth data, we find that the kernel estimate captures some significant features of non-normality: thick tails and a mode in the left tail that can be seen in figure 1 panel A. The mode in the left tail demonstrates the phenomenon that had given rise to the disaster risk literature – a higher likelihood of lower tail realizations of consumption growth, that can have significant implications for asset prices. In contrast to the disaster risk literature however, we do not arbitrarily introduce consumption growth “disaster” shocks, but rather “extract” them directly from the data.

Henceforth we take f , the estimated density of the growth rate innovations ε , to be the true underlying data-generating process.

Computation of the cumulative and the stationary densities

It is also required to compute the associated cumulative distribution function for ε that is needed to generate a random sample from the distribution of ε . To compute the CDF, we construct a dense uniform grid $\{\bar{\varepsilon}_1, \bar{\varepsilon}_2, \dots, \bar{\varepsilon}_n\}$ on the support of ε with an even number of sub-intervals and use (third-order

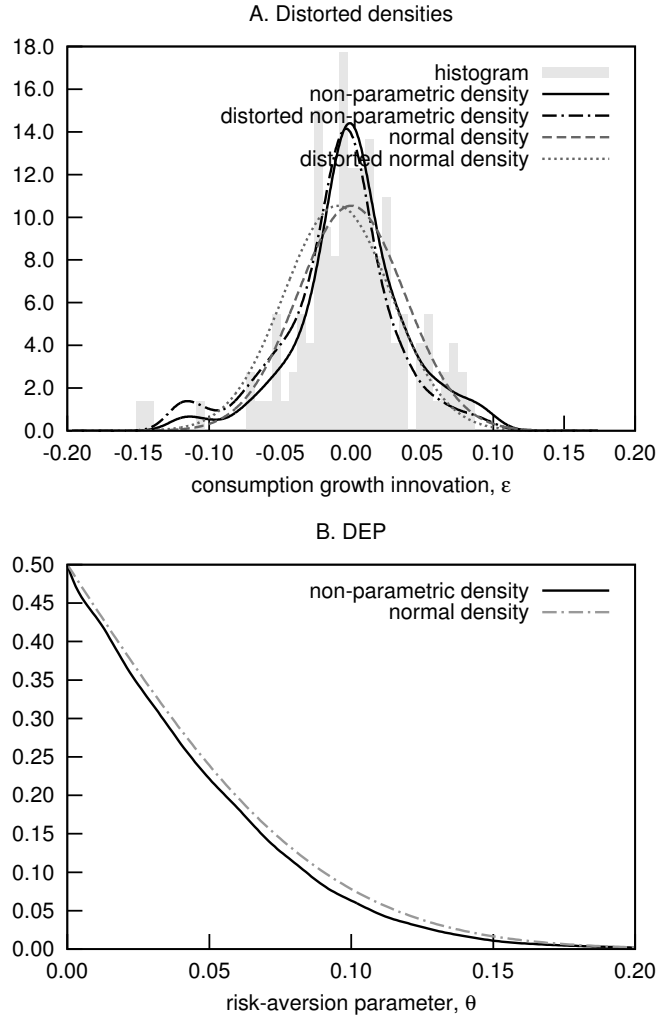


Figure 1: Histogram and estimates of the density of consumption growth innovations.

accurate) Simpson's rule for numerical integration:

$$F(\varepsilon_{2i}) = \int_{-\infty}^{\varepsilon_{2i}} f(v)dv \approx \frac{d}{3} \left[f(\bar{\varepsilon}_0) + 4 \sum_{j=1}^i f(\bar{\varepsilon}_{2j-1}) + 2 \sum_{j=1}^{i-1} f(\bar{\varepsilon}_{2j}) + f(\bar{\varepsilon}_i) \right],$$

where $d \equiv \bar{\varepsilon}_2 - \bar{\varepsilon}_1$. The data $\{\varepsilon_{2i}, F(\varepsilon_{2i})\}_{i=2}^{n/2}$ is then used to construct a spline approximation to F . Simpson's rule is also used to compute expectations in

all subsequent equations. In our computation we set $n = 1000$ and $\bar{\varepsilon}_1 = -5\sigma, \bar{\varepsilon}_n = 5\sigma$.

Unconditional asset returns must be computed using the stationary distribution of the consumption growth rate. In the case with normally distributed ε the stationary distribution is also normal with parameters that can be derived analytically. With the non-parametric f this calculation has to be performed numerically. The following recursive equation determines the relation between period- t and period- $t + 1$ distributions of the growth rate:

$$\varphi^{t+1}(y) = \int f(y - (1 - \rho)\mu - \rho x)\varphi^t(x)dx. \quad (19)$$

The stationary, or limiting, distribution is then computed iteratively starting with an initial guess φ^0 that is a spline approximation of $\sigma f((y - \mu)/\sigma)$ and updating our guess repeatedly using equation (19). The iteration is stopped when the sup norm between the two consecutive updates is less than 10^{-9} .

Computation of the value function and the distorted distribution

The normalized value function W must be computed numerically. As was noted above the Bellman equation (4) defines a contraction mapping on a complete metric space. According to the contraction mapping theorem, see [Stokey et al., 1989, p.50], W can be computed as the limit of the following recursive updating scheme:

$$W^{k+1}(g_t) = \ln(g_t) - \frac{\beta}{\theta} \ln [E (\exp \{-\theta W^k(g_{t+1})\} | g_t)].$$

Computation is started from $W^0(g_t) = 0$ and continues until $\|W^{k+1} - W^k\| < 10^{-6}$. For each k we use cubic splines with a 100 equally spaced knots to interpolate W^k on $[\min_t g_t - \sigma, \max_t g_t + \sigma]$.

Once W is known the implied distorted beliefs can be obtained using (7). To illustrate how the distorted distribution changes when the underlying DGP is non-normal, we plot the distorted density function for a parameter pair $(\beta, \theta) = (0.98, 0.10)$ in figure 1 panel A. The agent's pessimistic beliefs are no longer a simple shift in the mean of the underlying distribution. Instead, they tend to place higher probability on the tail realizations of consumption growth rates relative to the DGP, creating a larger mode in the

left tail of the agent’s distorted distribution and a smaller central mode that is now shifted leftward. Because the representative agent’s life-time utility is lower when consumption growth is low, the agent distorts the left tail of the distribution more than the right.

Computation of the DEP

To compute the DEP for the non-parametric distorted distribution, we turn to simulations. We simulate a large number N of samples of length T from distribution F and then compute the likelihood of each sample according to F and \tilde{F} . This can be done in two steps. First, generate a sample from a uniform distribution, $\{u_t\}_{t=1}^T$, using the Mersenne twister pseudo-random number generator. Second, transform each observation using the inverse of the cdf: $x_t = G^{-1}(u_t)$. Let the number of instances in which the likelihood computed using the distorted distribution is higher be $M_{\tilde{F}|F}$. We then reverse the roles of F and \tilde{F} and compute again the number of mistakenly classified samples, $M_{F|\tilde{F}}$. The DEP is approximated by:

$$\text{DEP} \approx 0.5(M_{F|\tilde{F}} + M_{\tilde{F}|F})/N.$$

This computation has to be repeated for each pair of (β, θ) as \tilde{F} depends on these two parameters.

Figure 1 panel B plots the DEP as a function of θ for the parametric and the non-parametric distribution estimates. The DEP is higher in the case with the normal rather than the non-parametric distribution. In other words, it is easier to distinguish the original and the distorted versions of the non-parametric distribution than those of the normal. It is so because in the case of the non-parametric distribution the distorted density shifts away from the middle of the distribution that is relatively likely, and hence “noticeable” to a statistician, under the true distribution. This means that while the non-parametric distribution may have better predictions for the asset returns, one has to consider lower values of the risk-aversion parameter θ .

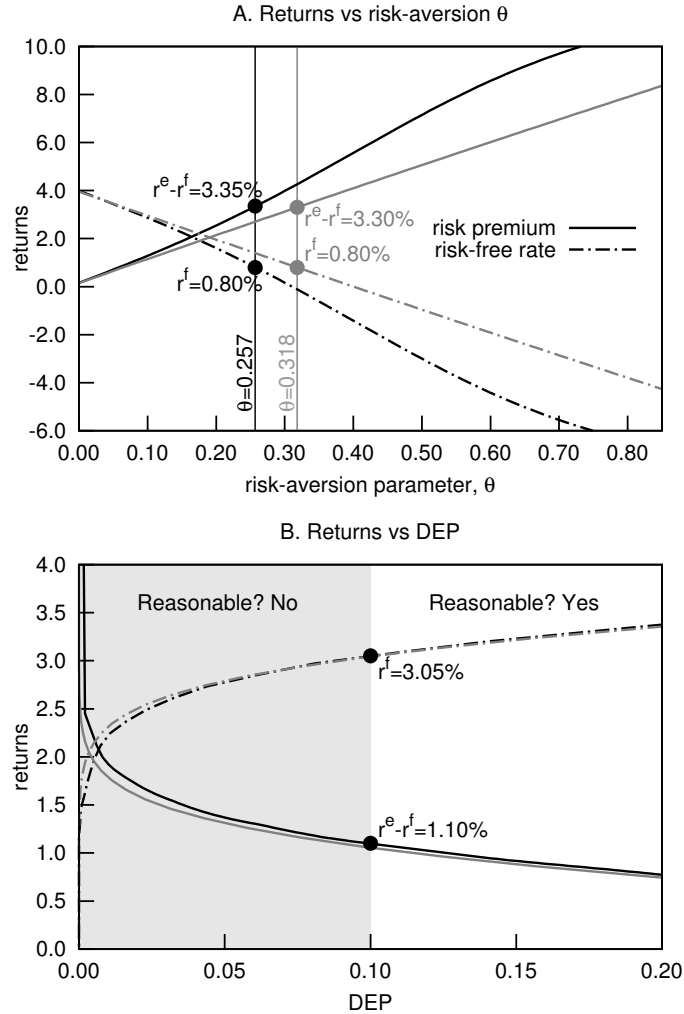


Figure 2: Asset returns as functions of the risk-aversion parameter θ . The black lines correspond to the non-parametric distribution of the growth innovations and the gray lines to the normal distribution. Vertical lines denote values of θ that match the observed risk-free rate of 0.8%. Parameters: $\beta = 0.98$.

3.2 Implications of non-normality for asset returns

The computed returns are plotted in figure 2. As θ increases, the risk free rate decreases and the risk-premium increases faster than in the case with the

normal growth innovations. If one commits to a given value of θ , then the asset return predictions are substantially better under the non-parametric distribution. But when the underlying distribution changes, so does the mapping from θ to DEP: see figure 1 panel B.

When we plot the returns as functions of the DEP, the advantage of the non-parametric distribution disappears. It is only with EDP less than 0.01 that the non-parametric distribution yields better predictions than the normal distribution. The key to understanding this is that at low values of θ , the asset returns depend very little on the shape of the distribution, as can be seen in figure 2. Yet the normal distribution corresponds to a higher DEP than the non-parametric distribution. Equivalently, if one keeps the DEP level fixed, then the model with the normal distribution admits higher values of the risk-aversion parameter θ : 0.318 vs 0.257. This overturns the advantage of the non-parametric distribution.

4 Conclusion

To match the risk-free rate of return that is observed in the data, one needs a value of the risk aversion parameter θ in $[0.25, 0.32]$. These parameter values are comfortably low. Admitting the non-parametric density improves the model's predictions significantly at all values of the risk-aversion in this range.

Unfortunately, the microeconomic evidence about θ is scarce and inconclusive. Consequently, we use the fact that the model solution can be interpreted as the one in which the representative agent has pessimistic beliefs. We then turn to statistical methods to shed light on the range of plausible values of θ . Surprisingly, in the case with the non-parametric distribution, the econometric tests would be more conclusive than with the normal distribution, yielding a smaller range of plausible beliefs and, therefore, θ s. When judged by the DEP, the model with the normal distribution performs marginally better.

To conclude, the estimated non-parametric density shows potential of improving asset-pricing predictions. But unfortunately the model is not rich enough to exploit it. The most immediate extension would be remove the

agent's confidence in the estimated value of the growth rate persistence ρ . When the agent has to distort both the distribution of the growth innovations and the persistence parameter, how would he change the latter? We conjecture that in this case the distorted distributions would become state-dependent. When the current growth rate is low, the agent would be inclined to think that this state may turn out to be more persistent than the evidence indicates. When the current growth state is high, he would guard against a possibility that the state is transitory. It would also be interesting to acknowledge that some parts of the distribution are estimated more precisely than others. That is to say one has to recognize that a model is estimated and the testing is performed on a particular data sample. For some samples, tail distortions may be easier to detect, but in other samples, these may be body distortions.

References

- David K. Backus, Bryan R. Routledge, and Stanley E. Zin. Exotic preferences for macroeconomists. *NBER Macroeconomics Annual*, 19:319–414, 2004.
- Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59:1481–1509, 2004.
- Francisco Barillas, Lars P. Hansen, and Thomas J. Sargent. Doubts or variability? *Journal of Economic Theory*, 144:2388–2418, 2009.
- Robert J. Barro. Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121(3):832–866, 2006.
- Stephen G. Cecchetti, Pok-sang Lam, and Nelson C. Mark. Asset pricing with distorted beliefs: Are equity returns too good to be true? *American Economic Review*, 90(4):787–805, 2000.
- Larry Epstein and Stanley Zin. Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57:937–969, 1989.

- Xavier Gabaix. An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, 127:645–700, 2012.
- Francois Gourio. Disaster risk and business cycles. *American Economic Review*, 102(6):2734–2766, 2012.
- Lars P. Hansen and Thomas J. Sargent. Robustness. *Princeton University Press*, 2007.
- Lars P. Hansen, John C. Heaton, and Nan Li. Consumption strikes back? measuring long-run risk. *Journal of Political Economy*, 116(2):260–302, 2008.
- David M. Kreps and Evan L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46:185–200, 1978.
- Qi Li and Jeffrey S. Racine. Nonparametric econometrics: Theory and practice. *Princeton University Press*, 2006.
- Ian W. Martin. Consumption-based asset pricing with higher cumulants. *Review of Economic Studies*, 80(2):745–773, 2013.
- Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, 1988.
- Thomas A. Rietz. The equity risk premium: A solution. *Journal of Monetary Economics*, 22(3):1117–1131, 1988.
- Nancy Stokey, Robert E. Lucas, and Edward C. Prescott. Recursive methods in economic dynamics. *Harvard University Press*, 1989.
- Thomas D. Tallarini. Risk-sensitive real business cycles. *Journal of Monetary Economics*, 45(3):507–532, 2000.
- Jessica Wachter. Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3):987–1035, 2012.
- Martin L. Weitzman. Subjective expectations and asset-return puzzles. *American Economic Review*, 97(4):1102–1130, 2007.

A Proofs

Lemma 2. *If g_t is a first-order Markov process then:*

$$V(g^t) = W(g_t) + \frac{y(g^{t-1})}{1 - \beta}, \quad \forall t, g^t.$$

Proof. Let $y(g^t) \equiv \ln Y(g^t)/(1 - \beta)$ and $W(g^t) \equiv V(g^t) - y(g^{t-1})/(1 - \beta)$. First, we derive the recursive equation for $W(g^t)$. Starting from the Bellman equation (4) we get:

$$\begin{aligned} W(g^t) &= \ln Y(g^t) - \frac{\beta}{\theta} \ln \left[E \left[\exp(-\theta(V(g^{t+1}) - y(g^t) + y(g^t)) | g^t) \right] \right] - y(g^{t-1}) \\ &= \ln Y(g^t) - \frac{\beta}{\theta} \ln \left[E \left[\exp(-\theta(V(g^{t+1}) - y(g^t)) | g^t) \right] \right] - y(g^{t-1}) \\ &= \ln g_t - \frac{\beta}{\theta} \ln \left[E \left[\exp(-\theta W(g^{t+1})) | g_t \right] \right]. \end{aligned}$$

The last line uses the fact that g_t is a first-order Markov process; so, it is sufficient to condition only on g_t . Finally, observe that the right hand side is a function of g_t not g^t . This implies that $W(g^t)$ must depend only on g_t as conjectured. \square

Lemma 3. *If g_t is a linear auto-regressive process then W is affine:*

$$\begin{aligned} W(g_t) &= A + B \ln g_t, \\ A &= (1 - \beta)^{-1} \beta [(1 - \rho)\mu - \ln[E[\exp(-\theta^2 B^2 \varepsilon)]]/\theta], \\ B &= 1/[(1 - \beta)(1 - \beta\rho)]. \end{aligned}$$

When the growth innovations are normally distributed then $A \equiv \beta[(1 - \rho)\mu - 0.5\sigma^2\theta/(1 - \beta)(1 - \beta\rho)]/(1 - \beta)^2/(1 - \beta\rho)$.

Proof. This can be verified using equation (4). \square

B Optimality conditions

Let $\lambda(g^t)$ be the Lagrange multiplier on the history- g^t budget constraint (5). The first-order optimality conditions for the representative agent's optimiza-

tion problem are:

$$\begin{aligned}c(g^t) : 0 &= 1/c(g^t) - \lambda(g^t), \\a(g^t, g') : 0 &= \beta V_a(a(g^t, g'), g') \frac{\exp(-\theta V(a(g^t, g'), g'))}{E[\exp(-\theta V(a(g^t, g'), g'))|g^t]} - \lambda(g^t) Q(g^t, g').\end{aligned}$$

The envelope condition is:

$$V_a(a(g^t), g_t) = \lambda(g^t).$$

The envelope condition implies that $V_a(a(g^t, g'), g') = \lambda(g^{t+1})$. Combining these relations with the first-order conditions yields (6).