Can Rare Events Explain the Equity Premium Puzzle?[*]

Christian Julliard[†]
Department of Economics
London School of Economics

Anisha Ghosh[‡]
Department of Economics
London School of Economics

First version: October 2007. This version: November 6, 2007.

Preliminary – Comments welcome

Abstract

Probably not. First, allowing the probabilities attached to the states of the economy to differ from their sample frequencies, the Consumption-CAPM is still rejected by the data and requires a very high level of Relative Risk Aversion (RRA) in order to rationalize the stock market risk premium. Second, we elicit the likelihood of observing an Equity Premium Puzzle (EPP) if the data were generated by the rare events probability distribution needed to rationalize the puzzle with a low level of RRA. We find that the historically observed EPP would be very unlikely to arise. Third, we find that the rare events explanation of the EPP significantly *worsens* the ability of the Consumption-CAPM to explain the cross-section of asset returns. This is due to the fact that by assigning higher probability to bad – economy wide – states in which consumption growth is low and all the assets in the cross-section tend to yield low returns, it reduces the cross-sectional dispersion of consumption risk relative to the cross-sectional variation in average returns.

Keywords: Rare Events, Equity Premium Puzzle, Generalized Empirical Likelihood, Semi-parametric Bayesian Inference, Calibration, Cross-Section of Asset Returns, Peso Phenomenon.

JEL classification: C11, C14, E17, G12.

1   Introduction

The average excess return on the U.S. stock market relative to the one-month Treasury Bill – the so called equity risk premium – has been about 7% per year over the last century. Nevertheless, the representative agent model with time separable CRRA utility, calibrated to match micro evidence on households' attitude toward risk and the time series properties of consumption and asset returns, generates a risk premium of less than 1%. This quantitative discrepancy was originally dubbed by Mehra and Prescott (1985) as the Equity Premium Puzzle (EPP). Given the dramatic long-term investment implications of this differential rate of return, over the last two decades the equity premium puzzle has been the focus of a substantial research effort in Economics and Finance.[1]

In this paper we study the ability of the *rare events hypothesis*, pioneered by Rietz (1988) and recently rediscovered by a growing literature (e.g. Veronesi (2004), Barro (2006), Gabaix (2007a)), to rationalize the equity premium puzzle. This hypothesis is conceptually simple. Suppose that in every period there is an *ex ante* small probability of an extreme stock market crash and economic downturn (that is, a Great Depression-like state of the economy). Risk averse equity owners will demand a high equity premium to compensate for the extreme losses they may incur during these unlikely – but exceptionally harmful – states of the world. In a finite sample, if such states happen to occur with a frequency lower than their true probability, *ex post* realized risk premia will be high even though *ex ante* expected returns are low – that is, in such a scenario equity owners are compensated for crashes and economic contractions that happen not to occur. Moreover, to an outside observer investors will appear irrational in the sample, and economists will tend to overestimate their risk aversion and underestimate the consumption risk of the stock market.

Our contribution to the analysis of the rare events hypothesis is three-fold. First, adopting an information-theoretic alternative to the Generalized Method of Moments[2] (see Owen (1991, 2001), Kitamura and Stutzer (1997), Kitamura (2006)), we estimate the consumption Euler equation allowing explicitly the probabilities attached to different states of the economy to differ from their sample frequencies. The methodologies we use belong to the Generalized Empirical Likelihood family, and *i*) are by construction

---

[1]Attempts to rationalize the puzzle include alternative assumptions on preferences (e.g. Abel (1990), Constantinides (1990), Epstein and Zin (1991), Campbell and Cochrane (1999), Bansal and Yaron (2004)), incomplete markets (e.g. Constantinides and Duffie (1996), Brav, Constantinides, and Geczy (2002)), market imperfections (e.g. Heaton and Lucas (1996), Constantinides, Donaldson, and Mehra (2002)), limited stock market participation (e.g. Mankiw and Zeldes (1991), Attanasio, Bank, and Tanner (2002), Brav, Constantinides, and Geczy (2002), Vissing-Jorgensen (2002)), problems of temporal aggregation (e.g. Heaton (1993), Gabaix and Laibson (2001)), and behavioral explanations (e.g. Barberis, Huang, and Santos (2001)). However, according to Mehra and Prescott (2003), none of the proposed explanations has been so far fully satisfactory (see also Campbell (1999, 2003)).

[2]Hansen (1982) and Hansen and Singleton (1982).

more robust to a rare events problem in the data,[3] ii) tend to have better small sample and asymptotic properties than the standard GMM approach (see e.g. Kunitomo and Matsushita (2003), Newey and Smith (2004) and Kitamura (2006)), iii) allow us to perform Bayesian posterior inference (Lazar (2003), Schennach (2005)) that does not rely on asymptotic properties that are less likely to be met, in finite sample, in the presence of rare events.

We find that the Consumption Capital Asset Pricing Model (C-CAPM) is still rejected by the data, and requires a very high level of relative risk aversion in order to rationalize the stock market risk premium.

Second, we show that our information-theoretic estimation approaches can also be used to identify, nonparametrically, the rare events distribution needed to rationalize the equity premium puzzle with a low level of risk aversion. In contrast with the *ad hoc* distributional assumptions and calibrations used in the previous literature on the rare events hypothesis, our methodology identifies the closest distribution, in the Kullback-Leibler Information sense, to the true unknown distribution of the data. That is, it provides the most likely rare events explanation of the equity premium puzzle. We show that our identified rare events distributions are qualitatively and quantitatively in line with the ones advocated by Rietz (1988) and Barro (2006) – that is, our data-driven procedure finds that only modest increases in the likelihood of observing extremely bad states such as the Great Depression are needed to rationalize the equity premium puzzle.

With these estimated rare events distributions at hand, we generate counterfactual histories of data of the same length as the historical time series. This allows us to elicit the probability of observing an equity premium puzzle in samples of the same size as the historical ones. We find that if the data were generated by the rare events distribution needed to rationalize the equity premium puzzle with a low level of risk aversion, the puzzle itself would be very unlikely to arise. We interpret this finding as suggesting that, if one is willing to believe that the rare events hypothesis is the explanation of the equity premium puzzle, one should also believe that the puzzle itself is a rare event.

Third, we study whether rare events can rationalize the poor performance of the Consumption-CAPM in pricing the cross-section of asset returns. We find that imposing on the data the rare events explanation of the equity premium puzzle *worsens* the ability of the Consumption-CAPM to explain the cross-section of asset returns. This is due to the fact that, in order to rationalize the equity premium puzzle with a low level of risk aversion, we need to assign higher probability to bad – economy wide – states such as deep recessions and market crashes. Since during market crashes and deep recessions consumption growth tends to be low and *all* the assets in the cross-section

---

[3]This is due both to the Large Deviations properties of our estimation and testing approach (see e.g. Kitamura (2006)), and to the "weak law of large numbers for rare events" rationale for estimators based on relative entropy minimization (Brown and Smith (1990)).

tends to yield low returns, this reduces the cross-sectional dispersion of consumption risk across assets, making it harder for the model to explain the cross-section of risk premia.

This finding also suggests that explanations of the equity premium puzzle based on agents' expectations of an economy wide disaster (e.g. a financial market meltdown) that has *not* materialized in the sample[4] – a so called *peso* phenomenon – would also reduce the ability of the Consumption-CAPM to price the cross-section of asset returns, since such an expectation would reduce the cross-sectional dispersion of consumption risk across assets.

The remainder of the paper is organized as follows. Section 2 reviews the related literature on the rare events hypothesis. Section 3 presents the theoretical underpinnings of the estimation and testing approaches considered. A data description is provided in Section 4. Estimation and testing results are presented in Sections 5. Section 6.1 presents the rare events distribution of the data needed to rationalize the equity premium puzzle with a low level of risk aversion. In Section 6.2, we ask what would be the likelihood of observing an equity premium puzzle, in a sample of the same size as the historical ones, if the rare events hypothesis were the true explanation of the puzzle. In Section 6.3, we analyze the implications of the rare events hypothesis for the ability of the C-CAPM to price the cross-section of asset returns. Section 7 concludes. Additional robustness check, and methodological details, are provided in the Appendix

2 Rare Events – Related Literature

"Un coup de dés jamais n'abolira le hasard."[5] Mallarmé (1897).

In this section we sketch the links to the existing literature on the rare events hypothesis.

Already in the sixties Mandelbrot (1963, 1967) pointed out that extreme financial asset price swings are more likely than what is often assumed in our models, and that returns on risky assets tend to have much thicker tails than a Gaussian distribution. The need of dealing with these empirical regularities is at the origin of: the attempts to apply *extreme value theory* (a branch of statistics that focuses on extreme deviations from the median) to quantitative financial analysis (see e.g. Beirlant, Schoutens, and Segers (2004)); the popularity, within the risk management industry, of the value-at-risk approach for assets valuation (the VaR measures the worst anticipated loss over a period); the inclusion of jump and Lévy processes into derivatives pricing models, and the development of tail-related financial risk measures.

The first to suggest that tail events in the distribution of asset returns and consumption might be the reason behind the equity premium puzzle originally documented

---

[4]See e.g. Danthine and Donaldson (1998).

[5]"A throw of dice will never abolish chance."

by Mehra and Prescott (1985), is Rietz (1988). As Mehra and Prescott, Rietz considers a (Markovian) finite-state version of the Lucas (1978) exchange economy. In this setting (as in the Consumption-CAPM of Rubinstein (1976) and Breeden (1979)), the optimizing behavior of the agent leads to the consumption Euler equation

$$E\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e\right] = 0, \tag{1}$$

where E is the unconditional expectation operator, C_t denotes the time t consumption flow, γ is the relative risk aversion coefficient, and R_t^e is the return on the stock market in excess of the risk free rate.

The only difference between the Mehra-Prescott and the Reitz frameworks is that in the former the state of the economy is good or bad with equal probabilities, while the latter adds a low probability depression-like state to capture rare, but severe, economic downturns and market crashes. Calibrating the depression state to match the economic contraction registered during the Great Depression, Rietz finds that a less than 2% probability for such a state is enough to make the model match the historically observed equity premium. Moreover, Danthine and Donaldson (1998) show that such a result also holds in a production economy setting.

More recently, Barro (2006) constructs a model of the equity-premium that extends Rietz (1988), and calibrates disaster probabilities from the twentieth century global history – especially the sharp contractions associated with the Great Depression and the World Wars. He argues that the potential for rare economic disasters explains a lot of asset pricing puzzles including the high equity premium, the low risk-free rate, and volatile stock returns. Gabaix (2007a), combining the Reitz's hypothesis with an ad hoc assumption about the data generating processes (the "linearity-generating processes" of Gabaix (2007b)), argues that rare events can potentially rationalize not only the equity premium, but also nine more puzzles in macro-finance. Similarly, but adding agents' learning to the framework, Veronesi (2004) concludes that the peso problem hypothesis implies most of the stylized facts about stock returns, including time-varying volatility, asymmetric volatility reaction to good and bad news, and excess sensitivity of price reaction to dividend changes.[6]

Beside the equity premium, the concept of rare events has also been applied to the study of a wide array of subjects such as: the term structure of interest rates,[7] exchange rates fluctuations and the forward-premium puzzle,[8] households' investment in annuities,[9] the "smirk" patterns documented in the index options market.[10]

---

[6]See also Sandroni (1998) on the interaction between learning and rare events.

[7]See e.g. Lewis (1990) and Bekaert, Hodrick, and Marshall (2001).

[8]Gourinchas and Tornell (2004) and Farhi and Gabaix (2007).

[9]Lopes and Michaelides (2007).

[10]Liu, Pan, and Wang (2005).

The common characteristic of all the studies that have focused on the rare events hypothesis and the equity premium puzzle, is the use of a calibration approach.[11] Since the equity premium is a first moment, and first moments are extremely sensitive to outliers, results in the rare events setting tend to be very sensitive to the calibration choice. For example, Copeland and Zhu (2006) extend the Barro (2006) closed-economy model to a two country setting, and they show that – using Barro's own calibration – the model implies levels of the equity risk premium far lower than those typically observed in the data e.g. they reverse Barro's original finding.

An obvious alternative to calibration is to estimate directly the consumption Euler equation (1), and this has been done extensively in the literature. The standard approach is to use consumption and stock market data to estimate the parameter γ in equation (1) as

$$\hat{\gamma} := \arg\min g\left(E^T\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e\right], E^T\left[R_t^e\right], E^T\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\right]\right) \qquad (2)$$

for some function $g\left(.\right)$, where $E^T\left[x_t\right] = \frac{1}{T}\sum_{t=1}^{T} x_t$ i.e. the distributions of the pricing kernel and returns is proxied with an empirical distribution that assigns probability $1/T$ to each realized state (observation) in the sample, and then judge whether $\hat{\gamma}$ (or some function of $\hat{\gamma}$, like the implied expected returns or a test statistic) is "reasonable." The GMM inference, for example, belongs to this class (see Hansen (1982), Hansen and Singleton (1982)). In some cases, distributional assumptions are directly made, and the inference is conditional on this (e.g log-normal returns and consumption growth). The replacement of the unconditional moments with sample moments is justified by a weak law of large numbers and central limit arguments. Nevertheless, in a finite sample, the presence of extreme rare events, that happen to occur with a lower (or higher) frequency than their true probability, can have dramatic effects on the sample first moments. Therefore, inference based on equation (2) might be unreliable. Saikkonnen and Ripatti (2000) illustrate this point with a Monte Carlo exercise, and they document an extremely poor performance of the GMM estimator of the Euler equation in the presence of rare events – even in relatively large samples.

Two conclusions can be drawn from the above review of the literature. First, in calibrating a rare events model, we would ideally remove any degree of freedom regarding the modeling of the underlying true distribution of the data. That is, it would be desirable to a) avoid parametric assumptions about the distribution of the

---

[11]Some indirect empirical evidence supporting the rare events hypothesis can be found in Brown, Goetzmann, and Ross (1995) and Goetzmann and Jorion (1999a) (see also Goetzmann and Jorion (1999b)). The first of these papers shows that the time series of equity returns on the U.S. stock market might be affected by survival bias. The second considers a large international cross-section of market returns, and finds that the average high return on the U.S. stock market belongs to the right tail of the cross-sectional distribution.

data, and *b*) use an approach that makes the calibrated distribution as close as possible to the true, unknown, distribution of the data.

Second, in estimating and testing the consumption Euler equation in the presence of a potential rare events phenomenon, it would be desirable to use an approach that allows the probabilities attached to different states of the economy to differ from their sample frequencies – that is, an approach that explicitly allows to take rare events into account.

Our paper does both of these things, and additionally derives the implications of imposing the rare events explanation of the equity premium puzzle on the cross-section of asset returns – a feature that, to the best of our knowledge, we are the first to analyze.

3   Econometric Methodology

In this section we present and discuss the econometric methodology used in our study. The methods presented belong to the Generalized Empirical Likelihood family, and readers familiar with this literature can skip this section without loss of continuity.

Consider a model characterized by the following moment condition

$$E^{\mu}\left[f(\mathbf{z}_t; \theta_0)\right] = \int f(\mathbf{z}_t; \theta_0) d\mu = \mathbf{0}, \quad \theta \in \Theta \subset \mathbb{R}^s, \tag{3}$$

where f is a known $\mathbb{R}^q$-valued function, $q \geqslant s$, and $\mathbf{0}$ denotes a vector of zeros of size q. The econometrician observes draws of an $\mathbb{R}^k$-valued random variable, $\{\mathbf{z}_t\}_{t=1}^T$, where each $\mathbf{z}_t$ is distributed according to an unknown probability measure μ, and θ_0 denotes the true – unknown – value of θ. The approaches used in this paper can be applied to both *i.i.d.* or (weakly) dependent data.[12] In the case of *i.i.d.* observations the nonparametric log likelihood at $(p_1, p_2, ..., p_T)$ is

$$\ell_{NP}(p_1, p_2, ..., p_T) = \sum_{t=1}^T \log(p_t), \ (p_1, ..., p_T) \in \Delta$$

where Δ denotes the simplex $\left\{(p_1, ..., p_T): \ \sum_{t=1}^T p_t = 1, \ p_t \geq 0, \ t = 1, ..., T\right\}$. This last expression can be interpreted as the log likelihood for a multinomial model, where the support of the multinomial distribution is given by the empirical observations, $\{\mathbf{z}_t\}_{t=1}^T$, even though the distribution μ of $\mathbf{z}_t$ is *not* assumed to be multinomial – it is indeed left unspecified. The Empirical Likelihood (EL) estimator of Owen (1988, 1990, 1991) parametrizes the moment condition (3) with $(\theta, p_1, p_2, ..., p_T) \in \Theta \times \Delta$, and is

---

[12]See e.g. Kitamura (1997) for a definition of weak definition of weak dependence.

given by

$$\left(\widehat{\theta}_{EL}, \widehat{p}_1^{EL}, ..., \widehat{p}_T^{EL}\right) = \underset{\{\theta, p_1, ..., p_T\} \in \Theta \times \Delta}{\arg\max} \ell_{NP} = \sum_{t=1}^T \log(p_t) \quad \text{subject to} \sum_{t=1}^T f(\mathbf{z}_t; \theta) p_t = \mathbf{0}.$$
(4)

Thus, the EL estimation defines a function that appears analogous to a parametric likelihood function and yet enables inference that does not require distributional assumptions. Moreover, the nonparametric maximum likelihood estimator (NPMLE) of the unknown probability measure μ is given by $\hat{\mu}_{EL} = \sum_{t=1}^T \widehat{p}_t^{EL} \delta_{\mathbf{z}_t}$, where $\delta_{\mathbf{z}}$ denotes a unit mass at $\mathbf{z}$. This is an efficient estimator for μ, i.e. for a function $a(\mathbf{z}; \theta_0)$ of $\mathbf{z}$, $\sum_{t=1}^T a(\mathbf{z}_t; \widehat{\theta}_{EL}) \widehat{p}_t^{EL}$ is a more efficient estimator of $E[a(\mathbf{z}; \theta_0)]$ than the naive sample mean $\frac{1}{T} \sum_{t=1}^T a(\mathbf{z}_t; \widehat{\theta}_{EL})$, and can be shown to be semiparametrically efficient (see Kitamura (2006) and Brown and Newey (1998)).

The EL estimator, for both $i.i.d.$ and weakly dependent data, also has an important information-theoretic interpretation (see e.g. Kitamura and Stutzer (1997)). To see this let M be the set of all probability measures on $\mathbb{R}^k$, and for each parameter vector $\theta \in \Theta$, define the following set of probability measures

$$P(\theta) \equiv \{p \in M : E^p[f(\mathbf{z}_t; \theta)] = \mathbf{0}\}$$

which are also absolutely continuous with respect to the measure μ in equation (3). Therefore, $\mathcal{P} \equiv \cup_{\theta \in \Theta} P(\theta)$ is the set of all the probability measure that are consistent with the model characterized by the moment condition in equation (3). The EL estimator can be shown to solve the following optimization problem

$$\inf_{\theta \in \Theta} \inf_{p \in P(\theta)} K(\mu, p) = \inf_{\theta \in \Theta} \inf_{p \in P(\theta)} \int \log (d\mu/dp) \, d\mu,$$
(5)

$$\text{subject to } E^p[f(\mathbf{z}_t; \theta)] = \mathbf{0},$$

where $K(\mu, p)$ is the Kullback-Leibler Information Criterion (KLIC) divergence from μ to p (White (1982)). Therefore $K(\mu, p) \geqslant 0$, and it will hold with equality if and only if $\mu = p$. If the model is correctly specified, i.e. if there exists a θ_0 satisfying (3), we have that $\mu \in P(\theta_0)$, and μ solves (5) delivering a KLIC value of 0. On the other hand, if the model is misspecified, μ is not an element of $\mathcal{P}$ and for each θ there is a positive KLIC distance $K(\mu, p) > 0$ attained by the solution $p(\theta)$. Thus, the EL approach searches for a $\hat{\mu}_{EL}(\theta)$ that makes the estimated distribution as close as possible – in the information sense – to the true unknown one.

Since the KLIC is not symmetric, the closely related Exponential Tilting (ET) estimator of Kitamura and Stutzer (1997) can be obtained by inverting the roles of μ and p in (5). That is, the ET estimator solves

$$\widehat{\theta}_{ET} = \inf_{\theta \in \Theta} \inf_{p \in P(\theta)} K(p, \mu) = \inf_{\theta \in \Theta} \inf_{p \in P(\theta)} \int \log (dp/d\mu) \, dp,$$
(6)

$$\text{subject to } E^p \left[f(\mathbf{z}_t; \theta) \right] = \mathbf{0}.$$

To first order, the ET and EL estimators are asymptotically equivalent to the optimal GMM estimator (Kitamura and Stutzer (1997), Qin and Lawless (1994)), i.e. have an asymptotic normal distribution given by

$$\sqrt{T} \left(\widehat{\theta}_j - \theta_0 \right) \xrightarrow{d} N(0, V), \quad j \in \{EL, ET\}$$

$$V = (D'S^{-1}D)^{-1}, \quad D = E^\mu \left[\partial f(\mathbf{z}, \theta_0) / \partial \theta' \right], \quad S = E^\mu \left[f(\mathbf{z}, \theta_0) f(\mathbf{z}, \theta_0)' \right]$$

However, Newey and Smith (2004) show that these estimators have smaller second-order bias than the GMM estimator. They also show that the bias-corrected EL estimator is third-order efficient. Moreover, Kunitomo and Matsushita (2003) provide a detailed numerical study of EL and GMM, and find that the distribution of the EL estimator tends to be more centered and concentrated around the true parameter value. They also report that the asymptotic normal approximation appears to be more appropriate for EL than for GMM.

Beside the desirable *local* asymptotic efficiency mentioned above, the empirical likelihood approach also has – unlike the GMM estimator – desirable *global* properties. The conventional asymptotic efficiency considerations focus on the behavior of the estimator in a *shrinking* close neighborhood of the true value of the parameters of interest. Efficiency theory based on the *large deviation* (LD) principle instead, focuses on the behavior of the estimator in a *fixed* neighborhood of the truth. Kitamura (2001) shows that testing based on the empirical likelihood ratio (ELR, described below) is asymptotically optimal in the large deviation sense, that is ELR is *uniformly most powerful*.[13] This is important for our empirical investigation, since large deviation efficiency is particularly appealing when estimating and testing in a setting in which the unknown distribution of the data might be characterized by rare events that can take on extreme values (since, in finite sample, the estimator is likely not to lie in a close neighborhood of the truth).

The reason behind the good asymptotic and finite sample properties of the empirical likelihood approach is that the KLIC, as pointed out by Robinson (1991), is extremely sensitive to any deviation of one probability measure from another. Moreover, since both EL and ET are based on the minimization of the relative entropy between the estimated and the unknown probability measure (captured by the inner minimizations in equations (5) and (6)), they endogenously re-weight rare events to have the model in equation (3) fit the data.[14]

Since both the EL and ET estimators are the solutions to convex optimization problems, the Fenchel duality applies (see Borwein and Lewis (1991)b and Kitamura (2006)),

---

[13] This property is sometimes referred to as generalized Neyman-Pearson optimality.

[14] See the weak law of large numbers for rare events of Brown and Smith (1990).

therefore reducing dramatically the dimensionality of the optimization problem. In particular, the solution to the inner minimization problem in equation (5) is a multinomial distribution with support given by the empirical observations $\mathbf{z}_t$, $t = 1, ..., T$ (Csiszar (1975)), and the probability weight assigned the t-th observation is

$$p_t^{EL}(\theta) = \frac{1}{T(1 + \lambda(\theta)'f(\mathbf{z}_t, \theta))}, \quad t = 1, ..., T$$

where $\lambda(\theta) \in \mathbb{R}^q$ is the solution to the following unconstrained convex problem

$$\lambda(\theta) = \arg\min_{\lambda} - \sum_{t=1}^{T} \log(1 + \lambda'f(\mathbf{z}_t, \theta)).$$

Similarly, the solution to the inner minimization problem in equation (6) is also a multinomial distribution with probability weight on the t-th observation given by

$$p_t^{ET}(\theta) = \frac{e^{\lambda(\theta)'f(\mathbf{z}_t, \theta)}}{\sum_{t=1}^{T} e^{\lambda(\theta)'f(\mathbf{z}_t, \theta)}}, \quad t = 1, ..., T$$

where $\lambda(\theta) \in \mathbb{R}^q$ is the solution to

$$\lambda(\theta) = \arg\min_{\lambda} \frac{1}{T} \sum_{t=1}^{T} e^{\lambda'f(\mathbf{z}_t, \theta)}$$

Likelihood based testing is also possible within the EL and ET frameworks. In the EL setting, Owen (1991, 2001) shows that a joint test of the overidentifying restrictions in equation (3), and the parameter restrictions $\theta_0 = \overline{\theta}$, may be performed by forming the nonparametric analog of the parametric likelihood ratio statistic, and this ELR test statistic has an asymptotic χ^2 distribution. Similarly, theorem 4 of Kitamura and Stutzer (1997) shows that an analogous likelihood ratio test can be constructed using the ET estimator.

The EL and ET estimated probability weights ($\{\hat{p}_t^j(\theta)\}$, $j \in \{EL, ET\}$) can also be used for Bayesian inference.

First, even though the empirical likelihood defines a *profile likelihood,* Lazar (2003) provides simulation evidence showing that, if the profile EL function is used as the likelihood part of the Bayes theorem, accurate posterior inference can be performed.[15]

---

[15]Based on the Monahan and Boos (1992) Kolmogorov-Smirnov criterion as a way of deciding that a likelihood alternative is valid for posterior inference, and an examination of the frequentist properties of the Bayesian intervals, Lazar (2003) concludes that it is reasonable to use EL within the Bayesian paradigm.

That is, given a prior $\pi(\theta)$, a Bayesian (empirical likelihood, BEL) posterior can be formed as

$$p(\theta|\{\mathbf{z}_t\}) \propto \pi(\theta) \times \Pi_{t=1}^T \hat{p}_t^{EL}(\theta). \tag{7}$$

Second, Schennach (2005) provides a well-defined probabilistic interpretation of the ET function that justifies its use in Bayesian inference. She shows that this likelihood function naturally arises as the nonparametric limit of a Bayesian procedure that places a type of noninformative prior on the space of distributions,[16] and that a posterior distribution can be obtained, as in equation (7), using the $\left\{p_t^{ET}(\theta)\right\}_{t=1}^T$ probability weights (this is the Bayesian Exponentially Tilted Empirical Likelihood, BETEL).

The application of the (generalized) empirical likelihood approaches just outlined to the estimation of consumption Euler equation (1), relies on the fact that the optimizing behavior of the representative agent, in the time-additive power utility model, leads to the conditional Euler equation

$$E_{t-1}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e\right] = 0, \tag{8}$$

where $E_{t-1}[.]$ denotes the expectation operator conditional on time $t-1$ information set. The above expression entails that $\left\{(C_t/C_{t-1})^{-\gamma} R_t^e\right\}_{t=1}^{\infty}$ is a martingale difference sequence, i.e. it is not autocorrelated. Therefore, the distribution theory for the EL and ET estimators outlined above remains valid even if the stochastic processes generating $\{C_t/C_{t-1}; R_t^e\}_{t=1}^{\infty}$ are weakly dependent. Nevertheless, serially correlated measurement error in consumption (see Wilcox (1992)) could make the martingale property of the conditional Euler equation (8) fail in the data. In Appendix A.1.1 we show how to deal with this issue, and results robust to violations of the martingale property are also provided.

4 Data Description

Ideally, the empirical analysis of the rare events hypothesis should be based on the longest possible sample. As a consequence, due to the different starting periods of available annual and quarterly consumption series, we focus on two samples of data: an annual data sample starting at the onset of the Great Depression (1929-2006), and a quarterly data sample starting in the post World War II period (1947:Q1-2003:Q3). Our proxy for the market return is the Center for Research in Security Prices (CRSP) value-weighted index of all stocks on the NYSE, AMEX, and NASDAQ. The

---

[16]The prior on the space of distributions gives preference to distributions having small support and, among the ones with the same support, it favors the entropy-maximizing ones. Moreover, it becomes uniform as $T \to \infty$.

proxy for the risk-free rate is the one-month Treasury Bill rate. Quarterly (annual) returns for the above assets are computed by compounding monthly returns within each quarter (year), and converted to real using the personal consumption deflator. For consumption, we use per capita real personal consumption expenditures on nondurable goods from the National Income and Product Accounts (NIPA). We make the standard "end-of-period" timing assumption that consumption during quarter t takes place at the end of the quarter. We make this choice so that the entire period that C_t covers is contained in the information set of the agent before the time $t + 1$ return is realized.[17]

For the cross-sectional analysis, we use the quarterly returns on the 25 Fama and French (1992) portfolios, and construct excess returns as these returns less the return on the 3-month Treasury Bill rate. We focus in this case on quarterly data only, since the cross-sectional estimation approach will be asymptotically justified by the assumption that the time dimension is large relative to the cross-sectional one. We concentrate on the Fama-French portfolios because they have a large dispersion in average returns that is relatively stable across subsamples, and because they have been used extensively to evaluate asset pricing models. The 25 Fama-French portfolios are the intersections of five portfolios formed on size (market equity) and five portfolios formed on the ratio of book equity to market equity. We denote a portfolio by the rank of its market equity and then the rank of its book-to-market ratio, so that portfolio 51 is the largest quintile of stocks by market equity and the smallest quintile of stocks by book-to-market. These portfolios are designed to focus on two key features of average returns: the size effect – firms with small market value have, on average, higher returns – and the value premium – firms with high book values relative to market equity have, on average, higher returns.

A relevant question for the robustness of our empirical approach is whether infrequent, economy wide, negative events are observed in our sample. We have good reasons to believe this is the case.

First, in our annual (quarterly) sample we observe 11 (7) out of the 15 major stock market crashes of the twentieth century identified by Mishkin and White (2002) plus the 2002 market crash.[18] Moreover, these include the largest one-day decline in stock market values in U.S. history – October 19th 1987, aka " Black Monday" – and 10 (5 in the quarterly sample) out of the 10 largest contractions of the Dow Jones Industrial Average index during the twentieth century.[19]

Second, our annual (quarterly) sample covers 13 (10) out of the 22 NBER recessions registered since 1900, and about 52.3% (32.4%) of the months of economic contraction

---

[17]The alternative timing convention, used by Campbell (1999) for example, is that consumption occurs at the beginning of the period.

[18]Mishkin and White (2002) identify a stock market crash as a period in which either the Dow Jones Industrials, the S&P500 or the NASDAQ index drops by at least 20 percent in a time window of either one day, five days, one month, three months or one year.

[19]Source: Dow Jones.

recorded over the same period.[20]

Third, in our annual (quarterly) sample we observe 4 (3) out of the 10 major U.S. wars since the 1775-1783 Revolution to the end of the twentieth century. Moreover, the 4 (3) wars in our annual (quarterly) sample amounts for 91.7% (22.3%) of the total cost of wars (in real terms) and for about 46.1% (34.3%) of the total months of wars in U.S. history, and about 88% (45%) of the enrolled forces in conflict during the twentieth century.[21]

Fourth, our annual (quarterly) sample covers about 85% (59%) of the Major Hurricanes, responsible for about 39% (18%) of the hurricane-related deaths, and 87% (68%) of the Deadly Earthquakes, responsible for about 20% (12%) of the earthquake-related deaths, recorded in the U.S. since 1900.[22,23]

5   Estimation Results

"Really, the most natural thing to do with the consumption-based model
is to estimate it and test it, as one would do for any economic model."
Cochrane (2005).

In this section we present estimation and testing results for the consumption Euler equation (1) using the Empirical Likelihood (EL), Exponential Tilting (ET), Bayesian EL (BEL) and Bayesian Exponentially Tilted Empirical Likelihood (BETEL) methods described in Section 3. We focus on these methods since they endogenously allow the probabilities attached to different states of the economy to depart from their sample frequencies – as the rare events hypothesis implies. Moreover, their asymptotic and small sample properties should deliver more robust and sharper inference in the presence of a rare events problem in the data.

Table 1 shows the estimation results for quarterly (Panel A) and annual (Panel B) data. The first row of both panels reports the point estimates of the relative risk aversion coefficient γ. The EL and ET frequentist estimates are, respectively, 102 and 146 in the quarterly sample, and 32 (for both estimators) in the annual sample.

---

[20]Source: National Bureau of Economic Reseacrh.

[21]Source: The United States Civil War Center. The wars considered in the calculations reported are: the Revolution (1775-1783), the War of 1812 (1812-1815), the Mexican War (1846-1848), the Civil War (1861-1865), the Spanish American War (1898), World War I (1917-1918), World War II (1941-1945), the Korean War (1950-1953), the Vietnam War (1964-1972), and the Gulf War (1990-1991). The Iraq War is excluded from the sample since complete statistics are currently unavailable.

[22]Source: U.S. Geological Survey Earthquake Hazards Program.

[23]Natural disasters have a long history of bearing relevant consequences for both stock markets and real economic activity. For example, the credit crisis known as the Panic of 1907 was originated by the losses stemming from the San Francisco earthquake the year before, that had hammered British insurers and generated rumors of insolvency for the biggest north American banks, and ultimately lead to the creation of the Federal Reserve.

Moreover, even though the standard errors of estimates (reported in brackets below the estimated coefficients) are large, all the point estimates are statistically larger than 10 (the upper bound of the "reasonable" range for the relative risk aversion coefficient) at standard confidence levels. The BEL and BETEL posterior distributions of this parameter (computed under a uniform prior on $\gamma \in \mathbb{R}^+$) also peak at very high values: 102 and 90, for BEL and BETEL respectively, in the quarterly sample, and at 32 in the annual sample. Moreover, the posterior 95% confidence intervals (in square brackets) never include values of γ smaller that 13.4.

Table 1: Euler Equation Estimation

	EL	ET	BEL	BETEL
	Panel A: Quarterly Data			
$\hat{\gamma}$	102 (48.0)	146 (32.3)	102 [24.8, 263.1]	90 [19.5, 164.9]
$\chi^2_{(1)}$	9.87 (.002)	10.65 (.001)		
$\Pr\left(\gamma \leq 10 \vert \text{data}\right)$			.64%	.92%
	Panel B: Annual Data			
$\hat{\gamma}$	32 (10.5)	32 (10.5)	32 [13.4, 64.1]	32 [13.8, 57.1]
$\chi^2_{(1)}$	5.26 (.022)	5.93 (.015)		
$\Pr\left(\gamma \leq 10 \vert \text{data}\right)$			1.00%	.84%

The second row of both panels reports tests for the joint hypothesis of a γ as small as 10 and for the identifying restriction given by the consumption Euler equation (1). These tests are the Empirical Likelihood Ratio (ELR) test of Owen (1991, 2001), (in the first column) and the likelihood ratio test proposed in theorem 4 of Kitamura and Stutzer (1997). Under the null hypothesis, both statistics follow asymptotically a χ^2 distribution with one degree of freedom. As revealed by the *p*-values reported below the test statistics, both tests reject the hypothesis of the Euler equation being satisfied by a γ as small as 10. Finally, the third row of each panel reports the posterior probabilities of γ being smaller than, or equal to, 10 given the observed data. This probability is small, and never larger than 1%, for both samples and both BEL and BETEL posteriors.

Overall, the results in Table 1 indicate that even adopting an estimation procedure that allows the probabilities attached to different states of the economy to differ from their sample frequencies, and is therefore robust to rare events problems in the data, the Consumption-CAPM is still rejected and requires a very high level of relative risk aversion to rationalize the stock market risk premium. Moreover, as a robustness check, Table A1 in the Appendix reports estimation and testing results that are robust to violations of the martingale difference property of the conditional Euler equation (that might, for example, be generated by serially correlated measurement error in

14

consumption as discussed in Wilcox (1992)). This robustness check confirms the results in Table 1.

6   Counterfactual Analysis

In this section, instead of jointly estimating the coefficient of relative risk aversion and the probabilities associated with different states of the economy, we fix the γ parameter to a "reasonable" value, and ask the ET and EL estimation procedures to identify the distribution of the data that would solve the Equity Premium Puzzle in the historical samples. This procedure can be interpreted as calibrating a rare events model (that solves the EPP) in a formal – data driven – fashion that minimizes the distance (in the information sense) between the model distribution and the true unknown distribution of the data.

With this estimated distribution at hand, we can ask the following relevant counterfactual questions.

First, suppose that the data were generated by the rare events distribution needed to explain the equity premium puzzle with a low level of risk aversion. Under this distribution, what would be the probability of observing an equity premium puzzle in samples of the same size as the historical ones? That is, if rare events that did not happen frequently enough in the historical sample were the true reason behind the equity premium puzzle, what would be the likelihood of observing such a puzzle?

Second, suppose rare events were the cause of the equity premium puzzle. Would taking these events into account also explain why the C-CAPM performs poorly in pricing the cross-section of asset returns? Or would it worsen the cross-sectional failure of the model?

In Section 6.1 we present the constructed rare events distribution of the data, while its implications for the likelihood of observing an equity premium puzzle and for the cross-section of asset returns are discussed, respectively, in Section 6.2 and Section 6.3.

6.1   A World without the Equity Premium Puzzle

To see how the EL and ET estimation procedure can be used to estimate a distribution of the data that rationalizes the equity premium puzzle, note that the consumption Euler equation (1) implies the following identity

$$\frac{E^F\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}R_t^e\right]}{E^F\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\right]} \equiv \underbrace{E^F\left[R_t^e\right] + \frac{Cov^F\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma};R_t^e\right]}{E^F\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\right]}}_{=:epp^F(\gamma)},$$

where F is the true, unknown, probability distribution of the data, and the right hand side is a measure of the equity premium puzzle under F, since it is given by the difference between the expected risk premium on the market and the risk premium implied by the Consumption CAPM. If the C-CAPM were the true model of the economy, we would have $epp^F(\gamma) = 0$ at the true γ. Note also that the EL and ET procedures estimate nonparametrically, for any value of γ, the unknown true probability distribution F with the probability weights $\{\hat{p}_t^j(\gamma)\}_{t=1}^T$ (where $j \in \{EL, ET\}$) such that

$$\sum_{t=1}^T \left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e \hat{p}_t^j(\gamma) = 0 \ \forall \gamma.$$

Therefore, indicating with $\hat{P}^j(\gamma)$ the probability measures defined by $\{\hat{p}_t^j(\gamma)\}_{t=1}^T$, $j \in \{EL, ET\}$, we have that

$$E^{\hat{P}^j(\gamma)}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e\right] = 0 \rightarrow epp^j(\gamma) = 0,$$

as long as $\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}$ and R_t^e have finite first and second moments under $\hat{P}^j(\gamma)$.

That is, fixing the relative risk aversion coefficient γ, we can use the EL and ET procedures to construct the probability distribution needed to solve the equity premium puzzle. As discussed in Section 3, these procedures will be consistent. Moreover, this approach will minimizes the Kullback-Leibler divergence between the estimated distribution and the unknown data generating process. That is, in the same fashion as a Maximum Likelihood estimator, this approach will minimize the distance (in the information sense) between the model and the true data generating process. Therefore, this procedure can be interpreted as calibrating a rare events model, that solves the equity premium puzzle, in a rigorous data-driven fashion, since the estimated $\hat{P}^j(\gamma)$ will be the closest distribution, among all the distributions that could rationalize the puzzle, to the true unknown data generating process. This implies that, if rare events are the true explanation of the equity premium puzzle, the estimated $\hat{P}^j(\gamma)$, $j \in \{EL, ET\}$, should identify their distribution.

In what follows we discuss the properties and implications of the estimated $\hat{P}^j(\gamma)$ assuming $\gamma = 10$, that is a level of relative risk aversion at the upper bound of what is commonly considered the "reasonable" range for this parameter (e.g. Gollier (2002)). In the Appendix we also report results for $\gamma = 5$.

With the $\hat{P}^j(\gamma)$ estimates at hand, the first question to ask is whether the implied state probabilities make economic sense. A priori, we would expect that the rare events distribution needed to rationalize the equity premium puzzle assigns a relatively higher weight to a few particularly bad states of the economy. Figure 1 suggests that this is exactly what the estimated $\hat{P}^j(\gamma)$ do.

16

Figure 1 reports EL and ET probability estimates, NBER recession periods (shaded areas), and the major stock market crashes identified by Mishkin and White (2002) plus the 2002 market crash (vertical dashed lines). Panel A reports estimated probability weights for quarterly data over the sample 1947:Q2-2003:Q3, while Panel B focuses on annual data over the sample 1930-2006. In the annual sample, we classify a given year as a recession if a NBER recession was registered in at least one of the quarters.
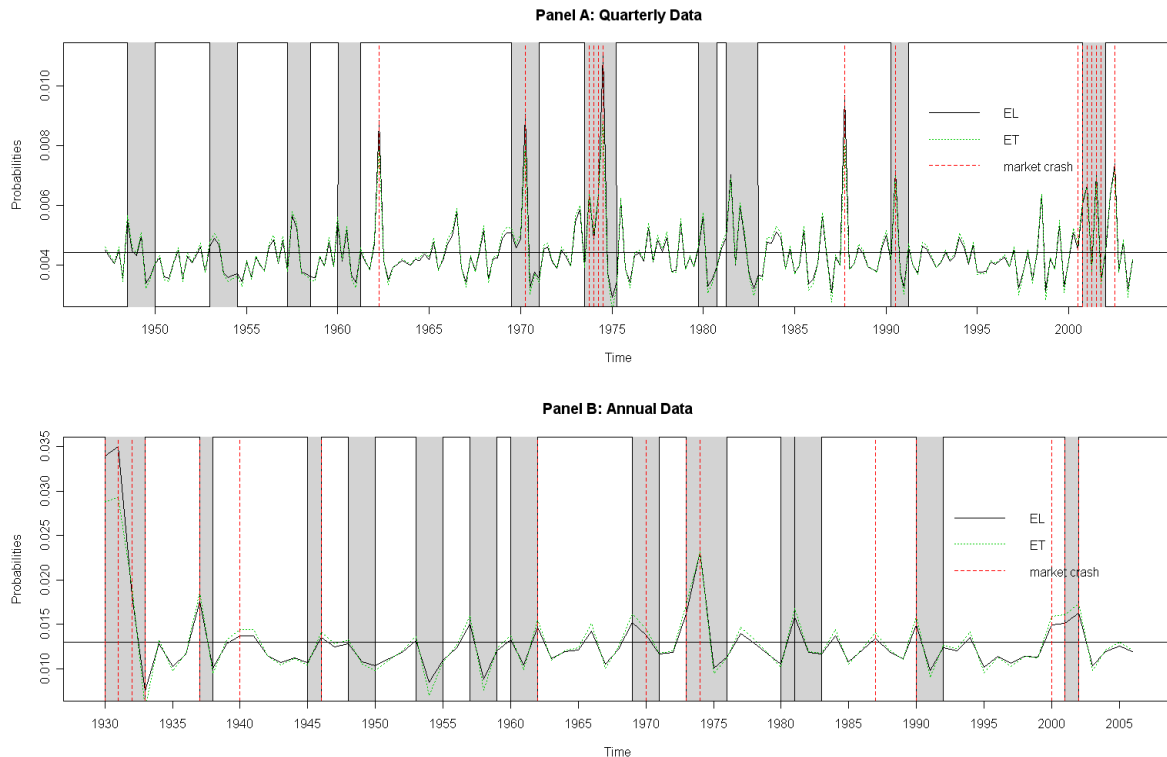

Figure 1: EL and ET estimated probabilities for $\gamma = 10$. Shaded areas are NBER recession periods. Vertical dashed lines are the stock market crashes identified by Mishkin and White (2002).

Several things are evident in Figure 1. First, the EL and ET estimated weights are extremely similar – the correlation between the two estimates is about .97 for both quarterly and annual data – suggesting robustness of these approaches. Second, both estimates tend to assign a relatively higher probability weight to recession periods: the frequency of recession in the quarterly (annual) sample is 19.9% (35.1%) while the EL and ET estimated probability of being in a recession period are, respectively, 21.3% and 20.9% (39.5% and 38.6%). Third, the increases in the probability of observing a recession are largely driven by assigning a higher probability to few recession periods

17

that are concomitant with market crash episodes. Fourth, the EL and ET estimated distributions assign higher probability to most of the identified periods of stock market crash. The sampling frequency of stock market crashes in the quarterly (annual)[24] data is 6.6% (20.8%) while the EL and ET estimated probabilities of a stock market crash are, respectively, 10.2% and 9.6% (28.2% and 27.9%). Fifth, the estimated probabilities tend to put the highest weights on few periods characterized by both a stock market crash and a recession – that is, states in which the consumption risk of the stock market is particularly high, like during the Great Depression period and the 1973-1975 recession. Nevertheless, even the probabilities attached to these states are still fairly small compared to the sampling frequency of the observations: for quarterly (annual) data the sampling frequency is .4% (1.3%), while the highest EL and ET probability weights are, respectively, 1.1% and .9% (4.7% and 3.4%).

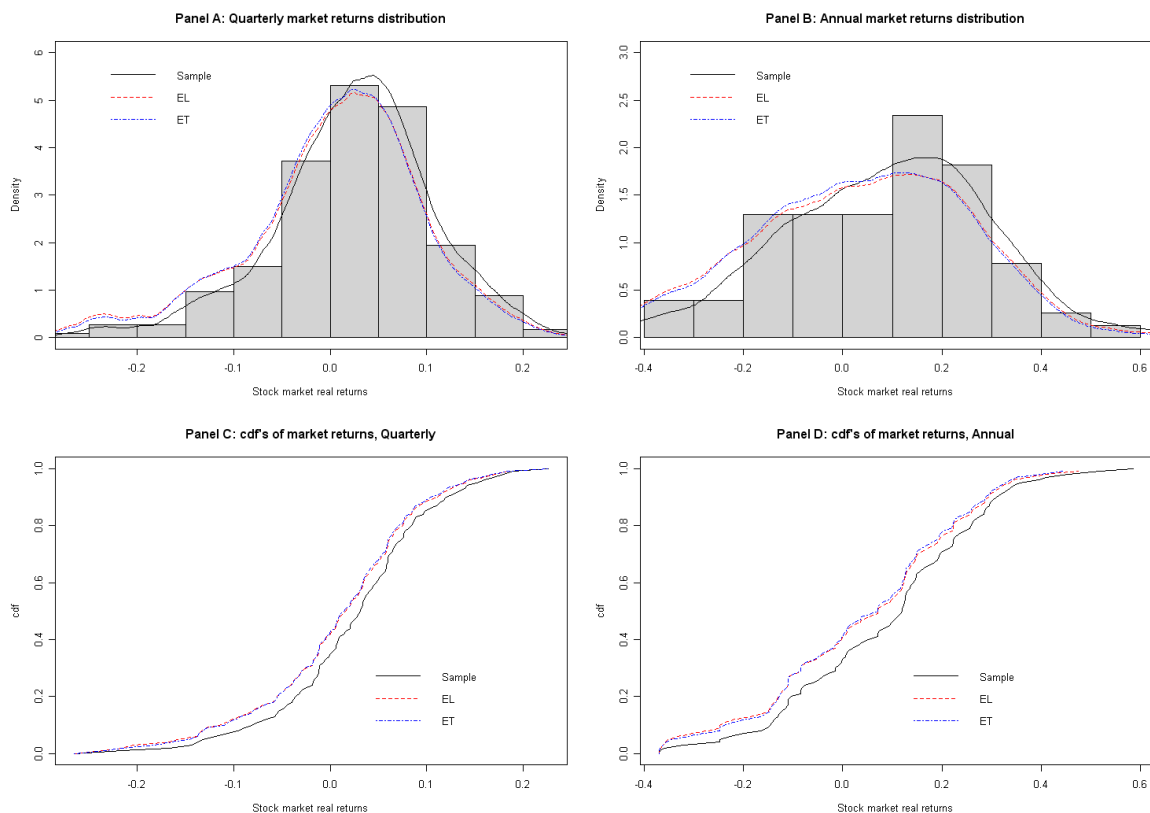Figure 2: Sample, EL and ET market returns distributions. EL and ET distributions are computed setting $\gamma = 10$.

---

[24]We classify a given year as a stock market crash year if at least one of the Mishkin and White (2002) crash episodes was recorded.

The implications of the estimated probability weights for the distribution of stock market real returns are summarized in Figure 2. The upper two panels report the histograms of quarterly (Panel A) and annual (Panel B) stock market real returns, (Epanechnikov) kernel estimate of the sample distribution, and weighted (Epanechnikov) kernel estimates where the weights are given by the estimated $\hat{P}^j(\gamma)$ probabilities. Panel C (for quarterly data) and Panel D (for annual data) report the empirical cumulative distribution function in the sample, and using the EL and ET probability weights.

The first thing to notice is that the rare event distribution needed to rationalize the equity premium puzzle implies thicker negative tails, and a more left skewed distribution than what has been historically observed. Moreover, the EL and ET probability weights generate a leftward shift in the distribution of returns when compared with the historical data. This leftward shift implies a reduction of both median and mean stock market return: the implied annual median (mean) risky rate is about 4.9%-6.4% (2.1%-5.0%). These numbers are in line with the rare events calibrated model of Barro (2006) that finds an expected risky rate in the 3.7%-8.4% range. One last point worth stressing is that, as shown in Panels C and D of Figure 2, the implied empirical distributions of market returns under $\hat{P}^{EL}(\gamma)$ and $\hat{P}^{ET}(\gamma)$ are extremely similar, once again suggesting robustness of the approach proposed.

Rare events models stress that the equity premium puzzle can be rationalized by assigning a higher probability to particularly bad states of the economy in which *both* market returns and consumption growth are low, since these are the states in which the consumption risk of the stock market is the highest. Figure 3 shows that this is indeed an implication of the EL and ET estimated probability weights.

Each panel of Figure 3 reports the scatter plot of stock market excess returns (horizontal axis) and consumption growth (vertical axis), also singling out observations that correspond to NBER recessions and to the stock market crash periods identified by Mishkin and White (2002). Each Panel also reports the level curves of Epanechnikov kernel estimates of the joint distribution of excess returns and consumption growth. The upper three panels focus on the quarterly sample, while the lower three panels report results for the annual sample. Panels A and D focus on the sample distributions, while Panels B and E and Panels C and F report, respectively, the EL and ET implied joint distributions (obtained by performing weighted kernel estimates with the weights given by the EL and ET estimated probabilities). The lower left portion of each panel represents states of the world in which the consumption risk of the stock market is highest i.e. observations that are characterized by both low excess returns and low consumption growth. Not surprisingly, this is also the area were recessions and stock market crashes tend to appear more often. Comparing the level curves in Panels A and D with the ones in Panel B, C, E and F, it appears clearly that the ET and ET probability weights skew the joint distribution of consumption growth and market returns toward the lower left portion of the graphs, therefore increasing the likelihood of
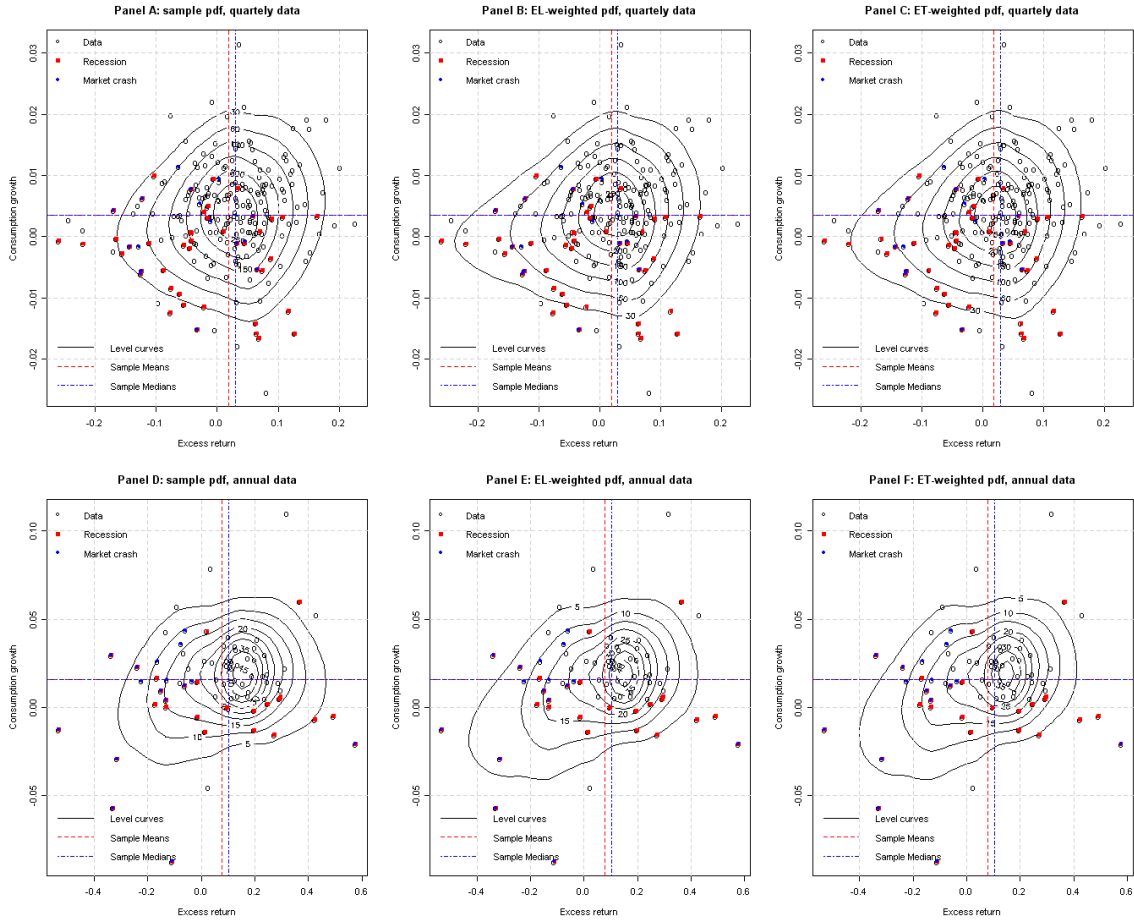
Figure 3: Level curves of the joint distribution of consumption growth and stock market excess returns.

high stock market consumption risk states.[25] Moreover, most of the shift in probability mass happens on the lowest level curve i.e. on the tail on the joint distribution, as the rare events explanation of the equity premium puzzle would imply.

Overall, the results of this section suggest that using the EL and ET approaches to construct distributions of the data that rationalize the equity premium puzzle with a low level of risk aversion, and at the same time are as close as possible to the true unknown distribution of the data, deliver results that are: *a*) robust, since both approaches have extremely similar implications, and *b*) in line with what the rare events

---

[25]In the quarterly (annual) data, the probability of jointly observing excess returns and consumption growth that are below their historical medians becomes about 30.2% (29.3%) for the EL estimates and 30.3% (29.2%) for the ET ones, therefore increasing the likelihood of such events by about 5%.

hypothesis predicts should be the mechanisms needed to rationalize the equity premium puzzle.

In the next two sections we ask whether a rare event model characterized by the $\hat{P}^j(\gamma)$ probability weights discussed above a) would be likely to deliver an equity premium puzzle in a sample of the same size as the historical one, and b) can help explain the inability of the standard Consumption CAPM to price the cross-section of asset returns.

6.2   How likely is the Equity Premium Puzzle?

The $\hat{P}^j(\gamma)$, $j \in \{EL, ET\}$, measures just discussed provide the most probable (in the likelihood sense) rare events explanation of the equity premium puzzle. But under these measures, what is the likelihood of observing an equity premium puzzle in a sample of the same size as the historical one?

To answer this question we perform the following counterfactual exercise. First, we use the estimated $\hat{P}^j(\gamma)$, $j \in \{EL, ET\}$, distributions to generate counterfactual samples of data of the same size as the historically ones.

That is, we use the $\left\{ \hat{p}_t^j(\gamma) \right\}_{t=1}^T$, $j \in \{EL, ET\}$, probabilities to draw with replacement from the observed data $\left\{ \frac{C_t}{C_{t-1}} ; R_t^e \right\}_{t=1}^T$, and use these draws to form samples of size T. We generate a total of $100,000$ counterfactual samples in this fashion (for both quarterly and annual data).

Second, in each sample i we compute the *realized* equity premium puzzle, $epp_i^T(\gamma)$, as

$$epp_i^T(\gamma) = E^T\left[R_{i,t}^e\right] + \frac{Cov^T\left[\left(\frac{C_{i,t}}{C_{i,t-1}}\right)^{-\gamma} ; R_{i,t}^e\right]}{E^T\left[\left(\frac{C_{i,t}}{C_{i,t-1}}\right)^{-\gamma}\right]} \tag{9}$$

where $E^T[.]$ and $Cov^T[.]$ denote the sample moments operators, and γ is fixed to the same low level used to construct $\hat{P}^j(\gamma)$. Moreover, in each generated sample we perform a GMM estimation of the coefficient of relative risk aversion. The results of this counterfactual exercise are summarized in Table 2. The first column reports the equity premium puzzle, as a function of γ, in the historical samples. The second column reports the median and, in squared brackets, the 95% confidence interval of the equity premium puzzle in the counterfactual samples. The third column reports the probability of observing, in the counterfactual samples, a realized equity premium puzzle as large as – or larger than – the historical one. The last column reports the median and, in squared brackets, the 95% confidence interval of the estimated γ coefficient in the counterfactual samples. Panel A focuses on quarterly data while Panel B hinges upon annual observations. Quarterly rates in Panel A are annualized for the sake of comparison with the annual ones in Panel B.

Table 2: Counterfactual Equity Premium Puzzle

	$epp^T(\gamma)$	$epp_i^T(\gamma)$	$\Pr\left(epp_i^T(\gamma) \geq epp^T(\gamma)\right)$	$\hat{\gamma}_{i,GMM}$
		Panel A: Quarterly Data		
$\hat{P}^{EL}(\gamma = 5)$	7.4%	0.0% [−4.6%, 4.7%]	0.10%	5 [−41, 67]
$\hat{P}^{EL}(\gamma = 10)$	7.3%	0.0% [−4.7%, 4.7%]	0.12%	10 [−36, 69]
$\hat{P}^{ET}(\gamma = 5)$	7.4%	0.0% [−4.6%, 4.5%]	0.10%	5 [−43, 66]
$\hat{P}^{ET}(\gamma = 10)$	7.3%	0.0% [−4.6%, 4.5%]	0.13%	10 [−40, 70]
		Panel B: Annual Data		
$\hat{P}^{EL}(\gamma = 5)$	7.2%	0.0% [−5.4%, 5.3%]	0.37%	5 [−21, 29]
$\hat{P}^{EL}(\gamma = 10)$	6.5%	0.0% [−5.7%, 5.7%]	1.22%	10 [−12, 32]
$\hat{P}^{ET}(\gamma = 5)$	7.2%	0.0% [−5.1%, 5.1%]	0.33%	5 [−24, 29]
$\hat{P}^{ET}(\gamma = 10)$	6.5%	0.0% [−5.4%, 5.5%]	0.98%	10 [−13, 33]

The first row, first column, of Panel A shows that the assumption of a relative risk aversion coefficient of 5 implies, in the 1947:Q2-2003:Q3 sample, an equity premium puzzle of about 7.4% per year. The second column shows instead that the median equity premium puzzle in the counterfactual sample generated by the EL probabilities with $\gamma = 5$ is about 0%, and that the upper bound of its 95% confidence band is only 4.7% – that is, the confidence interval does *not* include the historically observed equity premium puzzle.[26] Moreover, in the counterfactual samples, a negative equity premium puzzle seems almost as likely as a positive one. This is due to the fact that increasing the probability attached to extremely bad states of the economy, we make it more likely to observe too many of these events in a finite sample, therefore increasing the likelihood of observing a negative equity premium puzzle in the counterfactual samples.

The third column shows that, for a risk aversion of 5, the likelihood of observing an equity premium puzzle as large as the historical one would be extremely low – about 0.10%. The last column reports the median estimate and the 95% confidence bands, in square brackets, of the estimated relative risk aversion coefficient.[27] The estimates are centered around the true value used to generate the samples, but the 95% interval is very large, ranging from −41 to 67. This finding is in line with the evidence of a poor performance of the GMM estimator in the presence of rare events (Saikkonnen and Ripatti (2000)).

---

[26]Median and confidence bands are computed from the percentiles of $\left\{epp_i^T(\gamma)\right\}_{i=1}^{100,000}$.

[27]Median and confidence bands are computed from the percentiles of $\left\{\hat{\gamma}_{i,GMM}\right\}_{i=1}^{100,000}$.

The second row, first column, of Panel A shows that assuming a relative risk aversion of 10 reduces only marginally the equity premium puzzle in the 1947:Q2-2003:Q3 sample, to about 7.3% per year. The median epp_i^T in the counterfactual samples is still zero and once again the 95% confidence interval does not contain the historical value. Moreover, the likelihood of observing an equity premium puzzle as large as the historical one increases only slightly to about 0.12%. The last column shows that the estimates of γ are centered around the truth, but that in the presence of rare events this parameter can be dramatically misestimated.

The last two rows of Panel A of Table 2 use the ET probabilities instead of the EL ones. The results are largely in line with the ones in the first two rows: the median equity premium puzzle in the counterfactual sample is zero and its 95% confidence bands are too tight to include the historical values. Moreover, the equity premium puzzle is very unlikely to arise: its probability is less than 0.13%.

Panel B of Table 2 reports the same exercise as in Panel A, but it uses annual observations over the sample 1930-2006. The median epp_i^T is about zero for both EL and ET probabilities and, independently of the value of γ considered, the 95% confidence bands do not include the historical values of the puzzle. Most importantly, even in this sample the probability of observing an equity premium puzzle of the same magnitude as the historical one is extremely small, ranging from 0.33% to 1.22%. Once again, the GMM estimates of γ are centered around the truth but the confidence bands tend to be very large, suggesting that in the presence of rare events the risk aversion coefficient is likely to be misestimated.

As a robustness check, in the Appendix we present a similar counterfactual exercise that is robust to potential violations of the martingale difference property of the conditional consumption Euler equation. This approach, combined with a simple modification of the procedure for drawing counterfactual samples (presented in Section A.1.4), also allows us to preserve the autocorrelation properties of consumption growth and returns. As shown by Table A2 in the Appendix, this robustness check confirms the results in Table 2.

Over all, the results presented in this section imply that if the data were generated by the rare events distribution needed to rationalize the equity premium puzzle, the puzzle itself would be very unlikely to arise in samples of the same size as the historical ones. This suggests that *if one is willing to believe that the rare events hypothesis is the explanation of the equity premium puzzle, one should also believe that the puzzle itself is a rare event.*

6.3   Rare events and the cross-section of asset returns

Having identified the EL and ET probability weights needed to rationalize the equity premium puzzle with a low level of risk aversion, we can explore whether rare events

are a viable explanation of the poor performance of the C-CAPM in pricing the cross-section of asset returns.

The consumption Euler equation implies that the expected excess return on any asset should be fully explained by the asset's consumption risk, where the latter is measured by the covariance of the excess return with the ratio of marginal utilities between two consecutive periods. That is, for any asset i with associated excess return $R_{i,t}^e$, and denoting with F the true distribution of the data, we have that the following relation

$$E^F\left[R_{i,t}^e\right] = \alpha - \underbrace{\frac{Cov^F\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}; R_{i,t}^e\right]}{E^F\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\right]}}_{=:\beta_i}\lambda \tag{10}$$

should hold exactly (at the true γ) with $\alpha = 0$ and $\lambda = 1$. Similarly, linearizing the pricing kernel to get rid of γ (see e.g. Parker and Julliard (2003)), we have that the following relation

$$E^F\left[R_{i,t}^e\right] = \alpha + \underbrace{Cov^F\left(\ln\frac{C_t}{C_{t-1}}; R_{i,t}^e\right)}_{=:\beta_i}\lambda \tag{11}$$

should hold with $\alpha = 0$ and $\lambda > 0$. The β_i terms in equations (10) and (11) can be interpreted as a measure of the consumption risk that an agent undertakes investing in asset i.

It is a well documented empirical regularity that the above implications of the C-CAPM are rejected by the data when the moments in equations (10) and (11) are replaced with their sample analogs (see e.g. Mankiw and Shapiro (1986), Breeden, Gibbons, and Litzenberger (1989), Campbell (1996), Cochrane (1996), Lettau and Ludvigson (2001), Parker and Julliard (2005)). Nevertheless, if the empirical failures of the C-CAPM were the outcome of rare events that happened to occur with too low frequency in the historical sample, we would expect equations (10) and (11) to hold if the moments were constructed taking explicitly these rare events into account. That is, using the probability weights $\hat{P}^j(\gamma)$, $j \in \{EL, ET\}$, needed to rationalize the equity premium puzzle under the rare events hypothesis (and presented in Section 6.1), we would expect

$$E^{\hat{P}^j(\gamma)}\left[R_{i,t}^e\right] = \alpha - \frac{Cov^{\hat{P}^j(\gamma)}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}; R_{i,t}^e\right]}{E^{\hat{P}^j(\gamma)}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\right]}\lambda, \tag{12}$$

to hold with $\alpha = 0$ and $\lambda = 1$, and

$$E^{\hat{P}^j(\gamma)}\left[R_{i,t}^e\right] = \alpha + Cov^{\hat{P}^j(\gamma)}\left(\ln\frac{C_t}{C_{t-1}}; R_{i,t}^e\right)\lambda \tag{13}$$

24

to hold with $\alpha = 0$ and $\lambda > 0$.

To test this hypothesis we use the quarterly cross-section of Fama-French 25 portfolios and construct excess returns as these returns less the return on the 3-month Treasury Bill rate. To obtain empirical estimates of α and λ, we use the two-step Fama and MacBeth (1973) cross-sectional regression procedure, adapted to take into account that the moments in equations (12) and (13) should be constructed under the $\hat{P}^j(\gamma)$ probability measures rather than as sample analogs. The EL and ET probability weights $(\hat{P}^j(\gamma))$ are the ones described in Section 6.1 under the assumption that $\gamma = 10$, and we therefore estimate equation (12) setting $\gamma = 10$ in the pricing kernel. The point estimate of α measures the extent by which the model fails to price the average equity premium in the cross-section of the Fama and French (1992) 25 portfolios. The estimation procedure is described in detail in Appendix A.2. Cross-sectional estimation results are reported in Table 3.

The first column of the table reports the cross-sectional R^2 for all the models considered.[28] The second and third columns report, respectively, the point estimates of α and λ and their standard errors (in brackets). In order to disentangle the channels through which the rare events distributions $\hat{P}^j(\gamma)$ affect the cross-sectional performance of the C-CAPM, Table 3 also reports two additional statistics. The fourth column reports the percentage change in the ratio of the cross-sectional variance of consumption risk measures to the cross-sectional variance of average excess returns (that is $Var(\beta_i)/Var(E[R_{i,t+1}^e])$) caused by using the $\hat{P}^j(\gamma)$ probability weights, instead of sample averages, in computing the moments in equations (12) and (13). The fifth column reports the change in the cross-sectional variance of the correlations between the pricing kernel and excess returns in equations (12) and (13), that is the change in $Var(\rho_i)$, where ρ_i is defined as

$$\rho_i := corr\left(\left(\frac{C_t}{C_{t-1}} \right)^{-\gamma}; R_{i,t}^e \right)$$

for the non-linear representation in equation (10), and as

$$\rho_i := corr\left(\ln \frac{C_t}{C_{t-1}}; R_{i,t}^e \right)$$

for the linearized case in equation (11). Panel A of Table 3 focuses on the estimation of equations (10) and (12), while Panel B reports the estimation results of equations (11) and (13).

---

[28]Details on how to construct this statistic under the $\hat{P}^j(\gamma)$ probability measures are reported in Appendix A.2.

Table 3: Counterfactual Cross-Sectional Regressions

Moments:	R^2	$\hat{\alpha}$	$\hat{\lambda}$	$\Delta\frac{Var(\beta_i)}{Var\left(E\left[R^e_{i,t+1}\right]\right)}$	$\Delta Var\left(\rho_i\right)$
			Panel A: C-CAPM, $\gamma = 10$		
Sample	0.11	0.017 (0.005)	6.28 (5.04)		
$\hat{P}^{EL}\left(\gamma\right)$	0.00	0.007 (0.006)	−1.15 (5.09)	−35.4%	−18.4%
$\hat{P}^{ET}\left(\gamma\right)$	0.00	0.006 (006)	−0.78 (5.09)	−38.2%	−12.9%
			Panel B: linearized C-CAPM		
Sample	0.12	0.017 (0.005)	63.35 (49.89)		
$\hat{P}^{EL}\left(\gamma\right)$	0.00	0.007 (0.006)	−12.18 (50.31)	−34.9%	−19.4%
$\hat{P}^{ET}\left(\gamma\right)$	0.00	0.006 (0.006)	−8.49 (50.37)	−37.8%	−13.7%

Note: Fama and MacBeth (1973) standard errors in parenthesis under the estimated coefficients.

The first row of Panel A reports the results of estimating equation (10) using the sample moments of excess returns and pricing kernel. The large and statistically significant estimated $\hat{\alpha}$ is the cross-sectional equivalent of the equity premium puzzle, since it implies an average underpricing of the cross-section of returns of about 7% on an annual basis. Moreover, the point estimates of $\hat{\lambda}$ is more than six times larger than what is implied by the C-CAPM, and this is due to the fact that the consumption risk measures of the asset returns (the β_i's) computed from the sample averages are far too low to be consistent with the sample averages of risk premia. Nevertheless, the point estimate of $\hat{\lambda}$ is not statistically different from its theoretical value, but this is due to the large standard error that makes it not statistically different from zero. Overall, the performance of the C-CAPM is poor since the model is able to explain only 11% of the cross-sectional variance of risk premia of the Fama and French (1992) 25 portfolios.

The second row of Panel A of Table 3 focuses on the estimation of equation (12) using the EL probability weights, $\hat{P}^{EL}\left(\gamma\right)$, needed to rationalize the equity premium puzzle with a level of relative risk aversion of 10. The use of moments constructed under the probability measure $\hat{P}^{EL}\left(\gamma\right)$ substantially reduces the estimated α coefficient, implying a much smaller average mispricing (about 2.8% on a yearly basis), and the coefficient is not statistically significant – as implied by the consumption Euler equation. The $\hat{\lambda}$ coefficient instead, even though not statistically different from 1, has the opposite sign of what theory would predict. Moreover, the cross-sectional measure of fit is about 0.

Comparing the results in the first and second rows of Panel A, it is clear that, with the exception of the reduction in the average mispricing $\hat{\alpha}$, the C-CAPM performs even worse under the $\hat{P}^{EL}\left(\gamma\right)$ probability measure. What drives this result? In order to increase the ability of the C-CAPM to price the cross-section of returns, the $\hat{P}^{EL}\left(\gamma\right)$ empirical measure should in principle increase the cross-sectional variance of

consumption risk, $Var(\beta_i)$, relative to the cross-sectional variance of average risk premia, $Var\left(E\left[R_{i,t}^e\right]\right)$. But the entry on column four, second row, of Panel A shows that the exact opposite happens: moving from sample moments to the $\hat{P}^{EL}(\gamma)$ weighted moments, $Var(\beta_i)/Var\left(E\left[R_{i,t}^e\right]\right)$ is reduced by about 35.4%. And, as shown by the fifth column of the same row, this is due to the fact that the cross-sectional variation of the correlation between the pricing kernel and excess returns, $Var(\rho_i)$, is reduced by about 18.4% using the $\hat{P}^{EL}(\gamma)$ weights. This last finding is a direct consequence of the rare events explanation of the equity premium puzzle. In order to rationalize the equity premium puzzle with a low level of risk aversion, we need to assign higher probability to bad – economy wide – states such as deep recessions and market crashes. But in a market crash or a deep recession *all* the assets in the cross-section tends to yield low returns and consumption growth tend to be lower. Therefore, increasing the probability of these type of states has two effects. On one hand, it can rationalize the average risk premium on the market since, at the same time, it increase the consumption risk of investing in financial assets and reduces the expected returns. On the other hand, it makes it harder to explain the cross-section of risk premia, since it reduces the cross-sectional variability of consumption risk across assets.

The third row of Panel A of Table 3 use the $\hat{P}^{ET}(\gamma)$ probability weights for the estimation of equation (12). The results are very similar to the ones obtained using the $\hat{P}^{EL}(\gamma)$ weights: the model fits better the average risk premium but its ability to explain the cross-section of returns is reduced due to the substantial reduction in both $Var(\beta_i)/Var\left(E\left[R_{i,t+1}^e\right]\right)$ and in the cross-sectional variation of the correlation between the pricing kernel and excess returns, $Var(\rho_i)$.

Panel B of Table 3 focuses on the cross-sectional estimation of the C-CAPM in its linearized form reported in equations (11) and (13). The advantage of using the linearized approach is that the results do not depend on the choice of a pre-specified level for the relative risk aversion parameter γ. The first row of Panel B uses the sample moments of returns and consumption growth and reproduces the standard poor performance of the C-CAPM in explaining the cross-section of asset returns (see e.g. Parker and Julliard (2003)): the model is able to explain only 12% of the cross-sectional variation of average excess returns; the large $\hat{\alpha}$ estimate is statistically significant and implies an average underpricing close to 7% on an annual basis; $\hat{\lambda}$ has the right sign but is not statistically different from zero. The second and third rows focus, respectively, on the implications of using the $\hat{P}^{EL}(\gamma)$ and $\hat{P}^{ET}(\gamma)$ weights in estimating the moments in equation (13). The results are substantially in line with the ones reported in Panel A. Using the $\hat{P}^j(\gamma)$ probability weights helps the C-CAPM fit better the average risk premium on the market but it worsens the ability of the model to explain the cross-sectional differences in average returns. This is due to the reduction in the cross-sectional variation of the consumption risk measures caused by attaching higher probability to infrequent, economy wide, bad states during which consumption growth is low and all the assets tend to perform poorly.

Overall, the estimates in Table 3 suggest that the rare events explanation of the equity premium puzzle worsens the poor performance of the C-CAPM in pricing the cross-section of asset returns. Moreover, the above results hold qualitatively for any $\gamma \in \left]0, 10\right]$.

In order to show that the above findings are not driven by a few outlier portfolios, Figure 4 reports average returns and consumption risk estimates using both sample moments and $\hat{P}^j(\gamma)$ weighted moments. In Figure 4, the upper three panels report the scatter plots of average excess returns and model implied risk premia from equations (10) and (12), and the fitted lines correspond to the cross-sectional estimates in Panel A of Table 3. The lower three panels report the scatter plots of average excess returns and the consumption risk of the Fama-French 25 portfolios implied by the linearized pricing relations in equations (10) and (12). The fitted lines in the lower panels correspond to the cross-sectional estimates in Panel B of Table 3. We denote a portfolio by the rank of its market equity and then the rank of its book-to-market ratio so that portfolio 51 is the largest quintile of stocks by market equity and the smallest quintile of stocks by book-to-market.

Panel A of Figure 4 reports average and implied excess returns using the sample moments. If the C-CAPM were to price correctly the cross-section, all the Fama-French 25 portfolios would lie on the 45 degree line (the black line in the figure). All the Portfolio in Panel A lie far above and to the left of the 45 degree line, since sample averages of returns are all much higher than the model implied ones. Panel B reports average and implied risk premia constructed using the $\hat{P}^{EL}(\gamma)$ weights. By thickening the left tail of the returns distributions, the average returns are lowered, and by increasing the probability of states characterized by both low returns and low consumption growth, the model implied risk premia on the Fama-French 25 portfolios increases. As a consequence, the scatter plot in Panel B is shifted down and to the right compared with the one in Panel A. This shift makes the portfolios lie both above and below the 45 degree line, therefore delivering a better fit of the average risk premium on the portfolios, as shown by the lower $\hat{\alpha}$ in Table 3. Nevertheless, the $\hat{P}^{EL}(\gamma)$ weights fail to line up the portfolios on the 45 degree line due to the decrease in the cross-sectional dispersion of implied risk premia relative to the cross-sectional dispersion of average returns, therefore failing to improve the pricing of the cross-section of average returns. Panel C uses the $\hat{P}^{ET}(\gamma)$ weights in constructing average and implied risk premia, and delivers the same qualitative results as the ones obtained in Panel B using the $\hat{P}^{EL}(\gamma)$ weights.

The lower three Panels of Figure 4 reports average excess returns and the covariances between consumption growth and excess returns using sample moments (Panel D), $\hat{P}^{EL}(\gamma)$ weighted moments (Panel E), and $\hat{P}^{ET}(\gamma)$ weighted moments (Panel F). Comparing Panels E and F with Panel D, we observed a shift of the scatter plots toward the lower right corner, that increases the ability of the C-CAPM to fit the average risk premium on the 25 portfolios under the $\hat{P}^j(\gamma)$ probability measures. Nevertheless, we

do not observe the increase in the cross-sectional variation of consumption risk that would be necessary to make the model fit the cross-sectional differences in average risk premia.
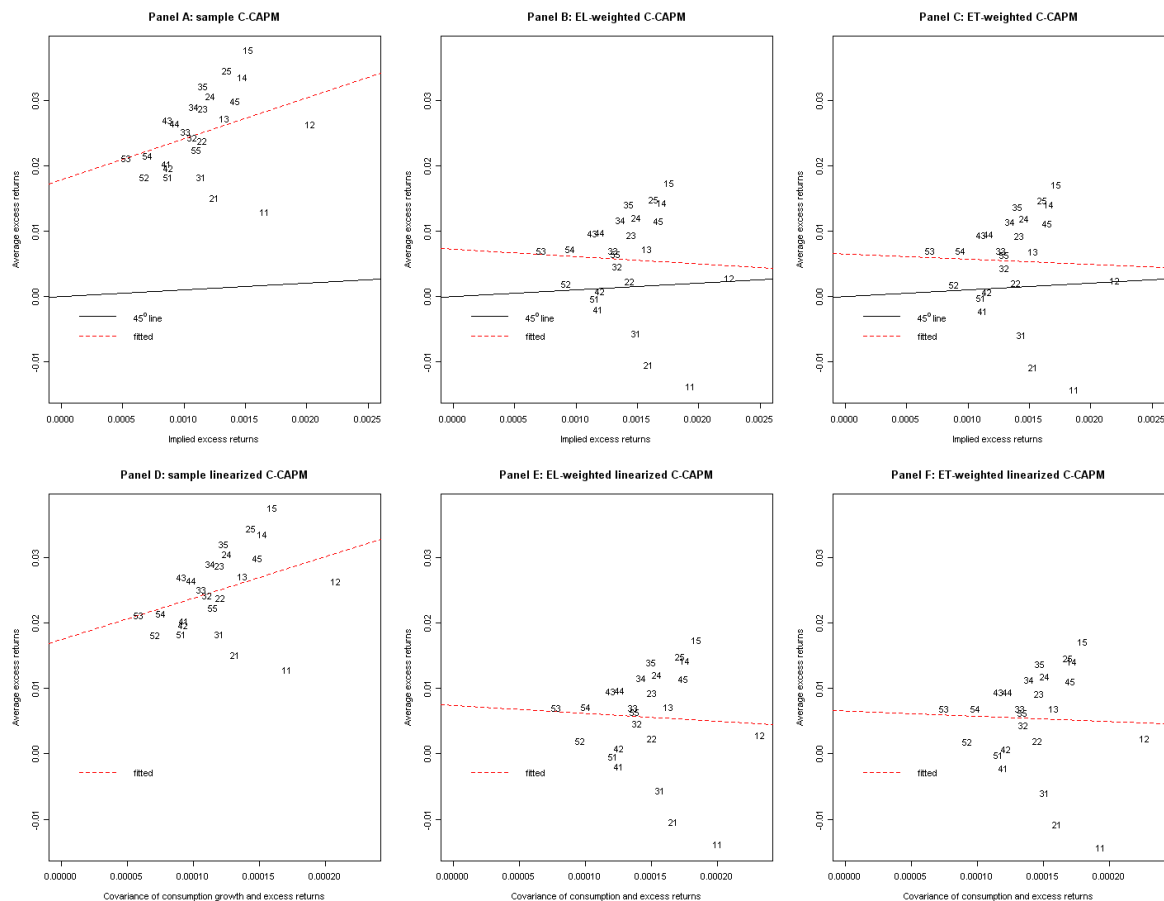

Figure 4: Rare events and the cross-section of asset returns.

Overall, the results of this section suggest that forcing on the data the rare events rationalization of the equity premium puzzle would worsen the already poor performance of the C-CAPM in pricing the cross-section of asset returns. This finding is driven by the fact that, in order to rationalize the equity premium puzzle with a low level of risk aversion, we need to assign higher probability to bad – economy wide – states such as recessions and market crashes. Since during market crashes and deep recessions *all* the assets in the cross-section tend to yield low returns and consumption growth tends to be low, this reduces the cross-sectional variability of consumption risk across assets, making it harder for the model to explain the cross-section of risk

premia. This finding also suggests that explanations of the equity premium puzzle based on agents' expectations of an economy wide disaster (e.g. a financial market meltdown) that has *not* materialized in the sample – a so called *peso* phenomenon – would also reduce the ability of the Consumption-CAPM to price the cross-section of asset returns, since such an expectation would reduce the cross-sectional dispersion of consumption risk across assets.

7   Conclusion

In this paper we study the ability of the rare events hypothesis to rationalize the equity premium puzzle. Performing econometric inference with an approach that endogenously allows the probabilities attached to the states of the economy to differ from their sample frequencies, we find that the consumption Euler equation with time-additive CRRA preferences is still rejected by the data, and that a very high level of relative risk aversion is needed in order to rationalize the stock market risk premium.

We identify the most likely rare events distribution of the data needed to rationalize the puzzle, and show that the constructed distribution is in line with the predicaments of the rare events hypothesis. Nevertheless, we find that, if the data where generated by such a distribution, an equity premium puzzle of the same magnitude as the historical one would be very unlikely to arise. We interpret this finding as suggesting that, if one is willing to believe in the rare events explanation of the equity premium puzzle, one should also believe that the puzzle itself is a rare event.

We also show that imposing on the data the rare events explanation of equity premium puzzle, substantially worsens the ability of the Consumption-CAPM to price the cross-section of asset returns. This is due to the fact that, in order to rationalize the EPP through a rare events (or peso) explanation, we need to assign higher probability to extremely bad, economy wide, states. Since in such states consumption growth is low and all the assets in the cross-section tend to perform poorly, the cross-sectional dispersion of consumption risk is reduced relative to the cross-sectional dispersion of asset returns, therefore reducing the ability of the Consumption-CAPM to explain the cross-section of returns.

The analytical approach undertaken in this paper can be extend to the study of other empirical regularities that, researchers have suggested, could be explained by the rare events hypothesis, e.g. exchange rates fluctuations and the forward-premium puzzle, the term structure of interest rates, and the "smirk" patterns documented in the index options market. Moreover, the information theoretic approach we suggest can be applied to the calibration of the nuisance distributions of *any* economic model that delivers well defined Euler equations. Furthermore, with minor methodological modifications, this approach can also be used to study the dynamic properties of the calibrated models.

References

ABEL, A. B. (1990): "Asset Prices Under Habit Formation and Catching Up with the Joneses," *American Economic Review*, 80, 38–42.

ATTANASIO, O., J. BANK, AND S. TANNER (2002): "Asset Holding and Consumption Volatility," *Journal of Political Economy*, 110(4), 771–92.

BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*.

BARBERIS, N., M. HUANG, AND T. SANTOS (2001): "Prospect Theory and Asset Prices," *Quarterly Journal of Economics*, 2001.

BARRO, R. J. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121(3), 823–866.

BEIRLANT, J., W. SCHOUTENS, AND J. SEGERS (2004): "Mandelbrot's Extremism," CentER discussion paper 2004-125.

BEKAERT, G., R. J. HODRICK, AND D. A. MARSHALL (2001): "Peso problem explanations for term structure anomalies," *Journal of Monetary Economics*, 48(2), 241–270.

BORWEIN, J. M., AND A. LEWIS (1991): "Duality relationships for entropy-type minimization problems," *SIAM Journal of Control and Optimization*, 29, 325–338.

BRAV, A., G. M. CONSTANTINIDES, AND C. C. GECZY (2002): "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence," *Journal of Political Economy*, 110(4), 793–824.

BREEDEN, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–96.

BREEDEN, D. T., M. R. GIBBONS, AND R. H. LITZENBERGER (1989): "Empirical Test of the Consumption-Oriented CAPM," *The Journal of Finance*, 44(2), 231–262.

BROWN, B. W., AND W. K. NEWEY (1998): "Efficient semiparametric estimation of expectations," *Econometrica*, 66(2), 453–464.

BROWN, D. E., AND R. L. SMITH (1990): "A Correspondance Principle for Relative Entropy Minimization," *Naval Research Logistics*, 37(2).

BROWN, S. J., W. N. GOETZMANN, AND S. A. ROSS (1995): "Survival," *Journal of Finance*, 50(3).

CAMPBELL, J. Y. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–345.

——— (1999): "Asset Prices, Consumption, and The Business Cycle," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1C, chap. 19, pp. 1231–1303. Elsevier Press.

——— (2003): "Consumption-Based Asset Pricing," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, chap. 13. North-Holland, Amsterdam.

CAMPBELL, J. Y., AND J. H. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–51.

COCHRANE, J. H. (1996): "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104(3), 572–621.

——— (2005): *Financial Markets and the Real Economy*. Now Publishers Inc.

CONSTANTINIDES, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98(2), 519–43.

CONSTANTINIDES, G. M., J. B. DONALDSON, AND R. MEHRA (2002): "Junior Can'T Borrow: A New Perspective On The Equity Premium Puzzle," *The Quarterly Journal of Economics*, 117(1), 269–296.

CONSTANTINIDES, G. M., AND D. DUFFIE (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104(2), 219–40.

COPELAND, L., AND Y. ZHU (2006): "Rare Disasters and the Equity Premium in a Two-Country World," working paper.

CSISZAR, I. (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *Annals of Probability*, 3, 146–158.

DANTHINE, J., AND J. DONALDSON (1998): "Non-Falsified Expectations and General Equilibrium Asset Pricing: The Power of the Peso," PenneWebber Working paper, Columbia University.

EPSTEIN, L. G., AND S. E. ZIN (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99(2), 263–286.

FAMA, E. F., AND K. R. FRENCH (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

FAMA, E. F., AND J. MACBETH (1973): "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

FARHI, E., AND X. GABAIX (2007): "Rare Disasters and Exchange Rates: a Theory of the Forward Premium Puzzle," mimeo.

GABAIX, X. (2007a): "Linearity-Generating Processes: A Modelling Tool Yielding Closed Forms for Asset Prices," working paper, M.I.T.

——— (2007b): "A Simple, Unified, Exactly Solved Framework for Ten Puzzles in Macro Finance," working paper, M.I.T.

GABAIX, X., AND D. LAIBSON (2001): "The 6D bias and the equity premium puzzle," in *N.B.E.R. Macroeconomics Annual 2001*, ed. by B. Bernanke, and K. Rogoff, pp. 257–311. Cambridge: MIT Press.

GOETZMANN, W., AND P. JORION (1999a): "Global Stock Markets in the Twentieth Century," *Journal of Finance*.

——— (1999b): "Re-emerging Markets," *Journal of Financial and Quantitative Analysis*, 34(1).

GOLLIER, C. (2002): "What does the classical theory have to say about household portfolios?," in *Household Portfolios*, ed. by T. J. Luigi Guiso, and M. Haliassos, chap. 1. MIT Press, Cambridge.

GOURINCHAS, P.-O., AND A. TORNELL (2004): "Exchange rate puzzles and distorted beliefs," *Journal of International Economics*, 64(2), 303–333.

HALL, P. (1985): "Resampling and coverage processes," *Stochastic Processes and Their Applications*, 19, 259–269.

HANSEN, L. P. (1982): "Large Sample Properties of Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86.

HEATON, J. (1993): "The Interaction Between Time-nonseparable Preferences and Time Aggregation," *Econometrica*, 61(2), 353–85.

HEATON, J., AND D. LUCAS (1996): "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 104(3), 443–87.

KITAMURA, Y. (1997): "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics*, 25, 2084–2102.

——— (2001): "Asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica*, 69, 1661–1672.

——— (2006): "Empirical Likelihood Methods in Econometrics: Theory and Practice," Cowles Foundation Discussion Papers 1569, Cowles Foundation, Yale University.

KITAMURA, Y., AND M. STUTZER (1997): "An Information-Theoretic Alternative To Generalized Method Of Moments Estimation," *Econometrica*, 65(4), 861–874.

KUNITOMO, N., AND Y. MATSUSHITA (2003): "On Finite Sample Distributions of the Empirical Likelihood Estimator and the GMM estimator," Manuscript.

LAZAR, N. (2003): "Bayesian Empirical Likelihood," *Biometrika*, 90, 319–326.

LETTAU, M., AND S. LUDVIGSON (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1286.

LEWIS, K. K. (1990): "Was there a "peso problem" in the U.S. term structure of interest rates: 1979-1982?," NBER WP 3282.

LIU, J., J. PAN, AND T. WANG (2005): "An Equilibrium Model of Rare-Event Premia and its Implications for Option Smirks," *Review of Financial Studies*, 18, 131–164.

LOPES, P., AND A. MICHAELIDES (2007): "Rare events and annuity market participation," *Finance Research Letters*, 4(2), 82–91.

LUCAS, JR., R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1426–14.

MALLARMÉ, S. (1897): "Un Coup de dés jamais n'abolira le hasard," *Cosmopolis*, 17, 417–427.

MANDELBROT, B. B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36, 394–419.

——— (1967): "The variation of some other speculative prices," *Journal of Business*, 40, 393–413.

MANKIW, N. G., AND M. D. SHAPIRO (1986): "Risk and Return: Consumption Beta Versus Market Beta," *Review of Economics and Statistics*, 68, 452–59.

MANKIW, N. G., AND S. ZELDES (1991): "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, 29(1), 97–112.

MEHRA, R., AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145–61.

——— (2003): "The Equity Premium Puzzle in Retrospect," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, vol. 2. North-Holland, Amsterdam.

MISHKIN, F. S., AND E. N. WHITE (2002): "U.S. Stock Market Crashes and Their Aftermath: Implications for Monetary Policy," NBER Working Papers 8992, National Bureau of Economic Research, Inc.

MONAHAN, J., AND D. BOOS (1992): "Proper Likelihoods for Bayesian Analysis," *Biometrika*, 79(2), 271–278.

NEWEY, W., AND R. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

OWEN, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

——— (1990): "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18(1), 90–120.

——— (1991): "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19(4), 1725–1747.

——— (2001): *Empirical Likelihood.* Chapman and Hall.

PARKER, J. A., AND C. JULLIARD (2003): "Consumption Risk and Cross-Sectional Returns," NBER Working Papers 9538, National Bureau of Economic Research, Inc.

——— (2005): "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy*, 113(1).

QIN, J., AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.

RIETZ, T. A. (1988): "The equity risk premium a solution," *Journal of Monetary Economics*, 22(1), 117–131.

ROBINSON, P. M. (1991): "Consistent nonparametric entropy-based testing," *Review of Economic Studies*, 58(3), 437–453.

RUBINSTEIN, M. (1976): "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics*, 7, 407–25.

SAIKKONNEN, P., AND A. RIPATTI (2000): "On the Estimation of Euler Equations in the Presence of a Potential Regime Shift," *The Manchester School Supplement*, pp. 92–121.

SANDRONI, A. (1998): "Learning, Rare Events, and Recurrent Market Crashes in Frictionless Economies Without Intrinsic Uncertainty," *Journal of Economic Theory*.

SCHENNACH, S. M. (2005): "Bayesian exponentially tilted empirical likelihood," *Biometrika*, 92(1), 31–46.

VERONESI, P. (2004): "The Peso Problem Hypothesis and Stock Market Returns," *Journal of Economic Dynamics and Control*, 28(4).

VISSING-JORGENSEN, A. (2002): "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110(4), 825–53.

WHITE, H. (1982): "Maximum Likelihood of Misspecified Models," *Econometrica*, 50, 1–25.

WILCOX, D. W. (1992): "The Construction of U.S. Consumption Data: Some Facts and Their Implications For Empirical Work," *American Economic Review*, 82, 922–941.

A Appendix

A.1 Robustness

A.1.1 Blockwise Empirical Likelihood and Exponential Tilting

The empirical application of the econometric methodology presented in Section 3 relies on the fact that the optimizing behavior of the representative agent, in the time-additive power utility model, leads to the Euler equation

$$E_{t-1}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e\right] = 0$$

where $E_{t-1}[.]$ denotes the expectation operator conditional on time $t-1$ information set. The above expression implies that $\left\{\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e\right\}_{t=1}^{\infty}$ is a martingale difference sequence, i.e. it is not autocorrelated. This implies that the standard errors and test statistics in Table 1 remain valid even if the stochastic processes generating $\left\{\frac{C_t}{C_{t-1}}, R_t^e\right\}_{t=1}^{\infty}$ are (weakly) dependent.

Nevertheless, serially correlated measurement error in consumption (see Wilcox (1992)) could make the martingale property of the conditional Euler equation fail in the data, therefore jeopardizing the asymptotic justification of the tests presented in Table 1.

To address the above issue, we focus on the *blockwise* empirical likelihood (Kitamura (1997)) and exponential tilting (Kitamura and Stutzer (1997)) estimators. The idea is to use blocks of consecutive observations to retrieve – nonparametrically – information about dependence in the data (this is closely related to the blockwise bootstrap, see e.g Hall (1985)). More precisely, the observation $f(\mathbf{z}_t,\theta) = \left(\frac{C_t}{C_{t-1}}\right)^{-\gamma} R_t^e$ is replaced by

$$\widehat{f}(t,\theta) = \frac{1}{2M+1} \sum_{s=-M}^{M} f(\mathbf{z}_{t+s},\theta)$$

where $M^2/T \to 0$, and $M \to \infty$ as $T \to \infty$. Then, EL and ET methods, as outlined in Section 3, are applied with $f(x_t,\theta)$, $t = 1, 2, ..., T$, replaced with $\widehat{f}(t,\theta)$, $t = M + 1, M + 2, ..., T - M$. Kitamura (1997) shows that the asymptotic distribution of these estimators is

$$\sqrt{T}\left(\widehat{\theta}_j - \theta_0\right) \xrightarrow{d} N(0, V), \quad j \in \{EL, ET\},$$

where

$$V = (D'\Omega^{-1}D)^{-1}, \quad D = E^{\mu}\left[\partial f(x_t,\theta_0)/\partial\theta'\right], \quad \Omega = \sum_{j=-\infty}^{\infty} E^{\mu}\left[f(x_t,\theta_0)f(x_{t+j},\theta_0)'\right].$$

A.1.2 Blockwise Estimation Results

In this section we use the blockwise estimation methodology outlined in Section A.1.1 to estimate and test the consumption Euler equation (1). Results obtained using this approach, and a window of four years in the annual sample, and of 12 quarters in the quarterly sample, are reported in Table A1.

Table A1 : Euler Equation Estimation

	EL	ET	BEL	$BETEL$
	Panel A: Quarterly Data			
$\hat{\gamma}$	100 (16.6)	124 (12.2)	100 [72.3, 133.0]	86 [65.0, 107.2]
$\chi^2_{(1)}$	8.67 (.003)	12.2 (.001)		
$\Pr(\gamma \leq 10 \vert \text{data})$			0.00%	0.00%
	Panel B: Annual Data			
$\hat{\gamma}$	37 (5.35)	37 (5.35)	37 [22.1, 54.8]	37 [22.6, 53.9]
$\chi^2_{(1)}$	4.42 (.035)	7.19 (.007)		
$\Pr(\gamma \leq 10 \vert \text{data})$			0.02%	0.01%

Point estimates and testing results in Table A1 are in line with the ones in Table 1. First, frequentist estimates of the relative risk aversion coefficient are very large and statistically larger than 10 at any standard confidence level. Second, joint tests of the consumption Euler equation and the restriction of a γ as small as 10, reject the null hypothesis. Third, Bayesian posteriors peak at high levels of γ and the 95% confidence regions do not include values smaller that 22. Fourth, given the data, the posterior probability of a relative risk aversion of 10 (or smaller) is extremely small.

A.1.3 Probability Weights with Alternative Values of γ

Figure 5 plots the time series of the EL and ET probability weights constructed setting $\gamma = 5$. Comparing Figure 5 with 1, it appears clearly that, changing the value of γ, the sets of events that need to receive an higher probability weight in order to rationalize the EPP with a low level of RRA, stays unchanged. The only difference is that, in order to rationalize the puzzle with a lower level of risk aversion, the probability assigned to a few economy wide extremely bad states such, as market crashes concomitant with deep recessions, have to be marginally increased. This suggests robustness of the approach with respect to the choice of the relative risk aversion parameter.
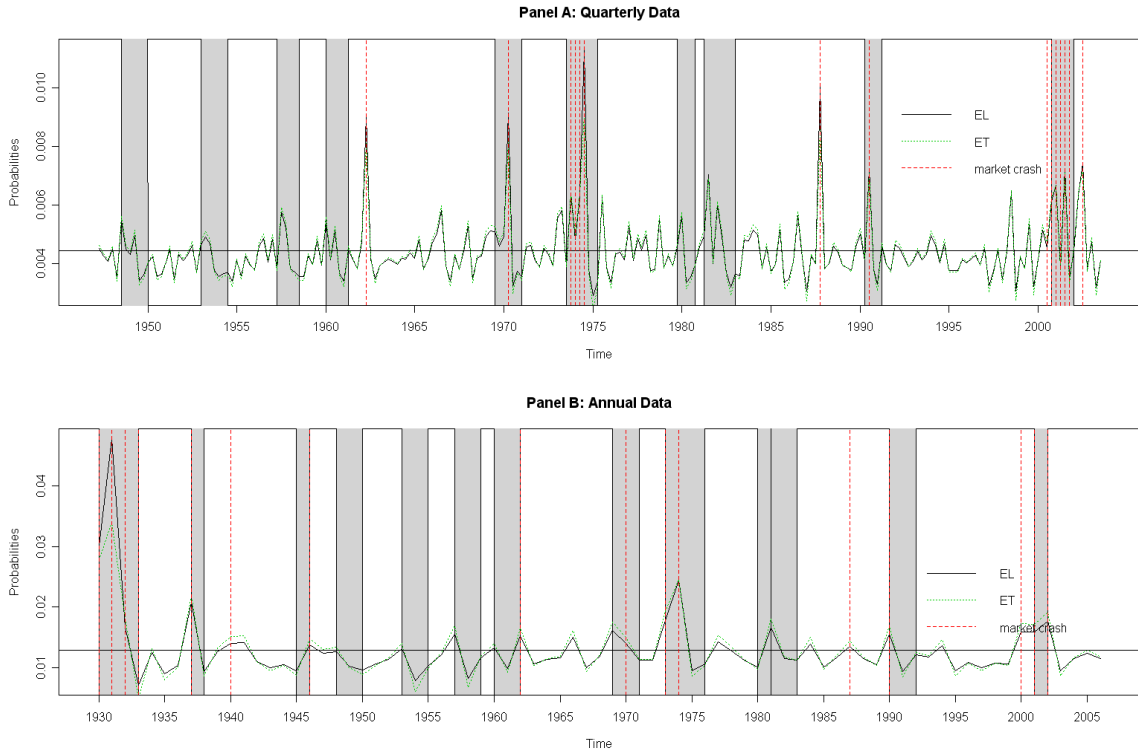
Figure 5: EL and ET estimated probabilities for $\gamma = 5$. Shaded areas are NBER recession periods. Vertical dashed lines are the stock market crashed identified by Mishkin and White (2002).

A.1.4 The Likelihood of the Equity Premium Puzzle with Blockwise Sampling

In this section we repeat the counterfactual exercise presented in Section 6.2, but instead of drawing individual couples of consumption growth and excess returns, we draw consecutive *blocks* of data in order to preserve, as in a blockwise bootstrap, the autocorrelation properties of the data. The EL and ET probabilities for the blocks are constructed using the blockwise approach described in Section A.1.1, and a window of four years in the annual sample, and of 12 quarters in the quarterly sample. Results obtained using this approach are reported in Table A2. The findings in Table A2 are largely in line with the ones in Table 2: the median realized equity premium puzzle in the counterfactual samples is very close to zero and its 95% confidence region does not include the historically observed value; the estimated coefficient of risk aversion in the counterfactual samples does not have any systematic bias but its 95% confidence interval is very large, suggesting that in the presence of rare events this parameter

39

might be dramatically misestimated. Moreover, the probability of observing an equity premium puzzle in the counterfactual samples is very small, and never larger than 1.02%. That is, as Table 2, Table A2 suggests that if one believes that the rare events hypothesis is the explanation of the EPP, one should also believe that the historically observed EPP is itself a rare event.

Table A2: Counterfactual EPP with Blockwise Sampling

	$epp^T(\gamma)$	$epp_i^T(\gamma)$	$\Pr\left(epp_i^T(\gamma) \geq epp^T(\gamma)\right)$	$\hat{\gamma}_{i,GMM}$
		Panel A: Quarterly Data		
$\hat{P}^{EL}(\gamma=5)$	7.4%	0.1% [−5.7%, 5.5%]	0.34%	5 [−56, 76]
$\hat{P}^{EL}(\gamma=10)$	7.3%	0.1% [−5.7%, 5.5%]	0.42%	10 [−56, 80]
$\hat{P}^{ET}(\gamma=5)$	7.4%	0.1% [−4.9%, 4.6%]	0.08%	5 [−49, 75]
$\hat{P}^{ET}(\gamma=10)$	7.3%	−0.2% [−5.2%, 4.3%]	0.08%	10 [−47, 80]
		Panel B: Annual Data		
$\hat{P}^{EL}(\gamma=5)$	7.2%	−0.1% [−4.6%, 4.7%]	0.08%	5 [−14, 20]
$\hat{P}^{EL}(\gamma=10)$	6.5%	−0.1% [−4.9%, 5.4%]	1.02%	10 [−5, 27]
$\hat{P}^{ET}(\gamma=5)$	7.2%	−0.2% [−3.7%, 3.7%]	0.00%	4 [−22, 28]
$\hat{P}^{ET}(\gamma=10)$	6.5%	−0.1% [−4.3%, 4.6%]	0.26%	10 [−9, 35]

A.2 Probability Weighted Fama-MacBeth Regressions

In order to estimate the parameters α and λ in equations (12) and (13) we follow a Fama and MacBeth (1973) two step procedure, adapted to take into account that the moments should be constructed under $\hat{P}^j(\gamma)$ probability measures instead that as sample analogs.

In a first step, we construct the consumption risk measure in equation (12) by computing

$$
\hat{\beta}_i^j \quad := \quad -\frac{Cov^{\hat{P}^j(\gamma)}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}; R_{i,t}^e\right]}{E^{\hat{P}^j(\gamma)}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\right]}
$$

$$
= \quad -\frac{\sum_{t=1}^{T}\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}R_{i,t}^e\hat{p}_t^j - \left[\sum_{t=1}^{T}\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\hat{p}_t^j\right]\left[\sum_{t=1}^{T}R_{i,t}^e\hat{p}_t^j\right]}{\left[\sum_{t=1}^{T}\left(\frac{C_t}{C_{t-1}}\right)^{-\gamma}\hat{p}_t^j\right]},
$$

where $j \in \{EL, ET\}$, and γ is fixed to the same value used to construct the weights $\{\hat{p}_t^j\}_{t=1}^T$. Similarly, for equation (12) we construct

$$
\begin{aligned}
\hat{\beta}_i^j \quad : \quad &= Cov^{\hat{P}^j(\gamma)}\left(\ln \frac{C_t}{C_{t-1}}; R_{i,t}^e\right) \\
&= \sum_{t=1}^T \ln\left(\frac{C_t}{C_{t-1}}\right) R_{i,t}^e \hat{p}_t^j - \left[\sum_{t=1}^T \ln\left(\frac{C_t}{C_{t-1}}\right) \hat{p}_t^j\right]\left[\sum_{t=1}^T R_{i,t}^e \hat{p}_t^j\right].
\end{aligned}
$$

In a second step, for each t we run the cross–sectional regression

$$
R_{i,t}^e = \alpha_t + \hat{\beta}_i^j \lambda_t + \varepsilon_{i,t}, \tag{14}
$$

where $\varepsilon_{i,t}$ is a mean zero cross-sectional error term, obtaining the sequence of estimates $\left\{\hat{\alpha}_t, \hat{\lambda}_t\right\}_{t=1}^T$. The point estimates for α and λ are then constructed as

$$
\hat{\alpha} := \sum_{t=1}^T \hat{\alpha}_t \hat{p}_t^j \text{ and } \hat{\lambda} := \sum_{t=1}^T \hat{\lambda}_t \hat{p}_t^j.
$$

Finally, we use the weighted sampling variation of $\left\{\hat{\alpha}_t, \hat{\lambda}_t\right\}_{t=1}^T$ to construct the standard deviations of the above estimators:

$$
\sigma^2(\hat{\alpha}) := \frac{1}{T}\sum_{t=1}^T (\hat{\alpha}_t - \hat{\alpha})^2 \hat{p}_t^j, \; \sigma^2\left(\hat{\lambda}\right) := \frac{1}{T}\sum_{t=1}^T \left(\hat{\lambda}_t - \hat{\lambda}\right)^2 \hat{p}_t^j.
$$

Note that the estimated $\hat{\alpha}$ and $\hat{\lambda}$ are numerically equivalent to the coefficients we would obtain by running the cross-sectional regression

$$
\left[\sum_{t=1}^T R_{i,t}^e \hat{p}_t^j\right] = \alpha + \hat{\beta}_i^j \lambda + \eta_i
$$

where η_i is a mean zero error term.

The cross-sectional R^2 corresponding to these Fama and MacBeth (1973) regressions is constructed as

$$
R^2 := 1 - \frac{Var\left(E^{\hat{P}^j(\gamma)}\left[R_{i,t}^e\right] - \hat{R}_{i,t}^e\right)}{Var\left(E^{\hat{P}^j(\gamma)}\left[R_{i,t}^e\right]\right)}
$$

where

$$
E^{\hat{P}^j(\gamma)}\left[R_{i,t}^e\right] := \sum_{t=1}^T R_{i,t}^e \hat{p}_t^j
$$

41

and

$$\hat{R}^e_{i,t} := \hat{\alpha} + \hat{\beta}^j_i \hat{\lambda}. \tag{15}$$

Note that the standard Fama and MacBeth (1973) cross-sectional regression approach (that does not use probability weights) can be recovered by setting $\hat{p}^j_t = 1/T$ $\forall t$.

The estimation algorithm just outlined is the one used to produce the results in Table 3 of Section 6.3.