

# Can Rare Events Explain the Equity Premium Puzzle?\*

Christian Julliard<sup>†</sup>                      Anisha Ghosh<sup>‡</sup>  
*London School of Economics*      *Carnegie Mellon University*

March 13, 2012

## Abstract

Probably not. First, allowing the probabilities of the states of the economy to differ from their sample frequencies, the Consumption-CAPM is still rejected in both U.S. and international data. Second, the recorded world disasters are too small to rationalize the puzzle unless one assumes that disasters occur every 6-10 years. Third, if the data were generated by the rare events distribution needed to rationalize the equity premium puzzle, the puzzle itself would be unlikely to arise. Fourth, the rare events hypothesis, by reducing the cross-sectional dispersion of consumption risk, *worsens* the ability of the Consumption-CAPM to explain the cross-section of returns.

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

*JEL classification:* C11, C14, E17, G12.

---

\*We benefited from helpful comments from Ravi Bansal, Robert Barro, Geert Bekaert, Markus Brunnermeier, Mike Chernov, George Constantinides, Max Croce, Jean-Pierre Danthine, Bernard Dumas, Xavier Gabaix, Rick Green, Raj Mehra, Alex Michaelides, Jonathan Parker, Jessica Wachter, two anonymous referees, and seminar participants at 2008 CEPR ESSMF Conference, 2008 NBER Summer Institute, 2008 SED Conference, 2010 AEA Meetings, Cambridge University, Carnegie Mellon University, CEMFI, Duke University, Ente Einaudi, HEC Paris, London School of Economics, University of Manchester, New York Fed, Oxford University, Paris School of Economics, Royal Holloway, Sorbonne University, UC Irvine, University of North Carolina, University of Leicester, University of Minnesota, University of Naples, University of Venezia, University of Warwick.

<sup>†</sup>Department of Finance and FMG, London School of Economics, Houghton Street, London WC2A 2AE, U.K., and CEPR, c.julliard@lse.ac.uk.

<sup>‡</sup>Corresponding author, Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, U.S.A., Ph: (+1) 412-268-4063, anishagh@andrew.cmu.edu. Requests for reprints to be sent to this author.

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) and, given its dramatic long-term investment implications, has been the focus of a substantial research effort in Economics and Finance over the last two decades.<sup>1</sup>

In this paper we analyze the ability of the *rare events hypothesis*, pioneered by Rietz (1988) and recently revived by a growing literature (e.g. Barro (2006), Gabaix (2012)), to rationalize the EPP. In particular, we study whether U.S. and international data on the history of economic disasters offer, as argued in the previous literature, support for this theory.

The rare events 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 four-fold. First, adopting an information-theoretic alternative to the Generalized Method of Moments, we estimate the consumption Euler equation for the equity risk premium allowing explicitly the probabilities attached to different states of the economy to differ from their sample frequencies. 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 (RRA) in order to rationalize the stock market risk premium. Moreover, this result holds for a variety of data sources and samples, including ones that start as far back as 1890, and not only for the U.S., but also for eight other OECD countries.

Second, since for many of the countries that experienced disasters we do not have the time series data needed to estimate the consumption Euler equation, but we have data on the sizes of the economic contractions and stock market returns during disasters, we ask whether the rejection of the rare events hypothesis could be due to the U.S. being relatively “lucky” in not experiencing a much larger disaster in its history. We find that the disasters present in the world data do not offer support for the rare events explanation of the EPP, unless we are willing to believe that disasters should

---

<sup>1</sup>However, according to Mehra and Prescott (2003), none of the proposed explanations so far has been fully satisfactory (see also Campbell (1999, 2003) and Mehra (2008)).

be happening every 6-10 years – that is, with about one order of magnitude higher frequency than in the typical calibration approach (a disaster every 59 years) used to provide support for the rare events explanation of the EPP.<sup>2</sup> These results are in line with the ones in Backus, Chernov, and Martin (2009), who find that the potential consumption disasters implied by index options data are much smaller than what is typically used in the standard calibration approach of rare events models.

The econometric methodologies we use belong to the Generalized Empirical Likelihood family and *i*) are by construction more robust to a rare events problem in the data, *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)), and *iii*) allow us to perform Bayesian posterior inference (see Schemmach (2005)) that does not rely on asymptotic properties that are less likely to be met, in finite sample, in the presence of rare events.

Third, using a novel – data driven – approach to calibration to construct, non-parametrically, the rare events distribution needed to rationalize the EPP with a low level of risk aversion, 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 EPP 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 EPP, 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 EPP, one should also believe that the puzzle itself is a rare event. In contrast with the *ad hoc* distributional assumptions and calibrations used in the previous literature on rare events, our methodology identifies the closest distribution, in the Kullback-Leibler Information sense, to the true unknown distribution of the data. That is, our calibration provides the *most likely* rare events explanation of the EPP.

Fourth, we study whether rare events can rationalize the poor performance of the C-CAPM in pricing the cross-section of asset returns. We find that imposing on the data the rare events explanation of the EPP *worsens* the ability of the C-CAPM to explain the cross-section of asset returns. This is due to the fact that, in order to rationalize the EPP with a low level of risk aversion, we need to assign higher probabilities 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 tend 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.

Overall, our findings suggest that the rare events hypothesis is an unlikely explanation of the EPP. Moreover, we find that, given the historical data, the most likely process for consumption and returns dynamics needed to rationalize the EPP is one

---

<sup>2</sup>We also show that the conclusion obtained in the literature, that finds support for the rare events hypothesis, is driven by the calibration of *one-year* contractions during disasters as being equal to the *cumulated multi-year* contractions recorded in the data. This finding is in line with Nakamura, Steinsson, Barro, and Ursúa (2010) that, studying an international cross-section of consumption time series and (differently from us) focusing on consumption data in isolation from stock market dynamics, find that modeling disasters as occurring only in one year is inappropriate.

that increases the likelihood of recessions and market crashes by about 4-6% compared to the historical frequency. This suggests that a more likely explanation of the puzzle should be searched in recession, rather than disaster, risk.

Note that our analysis is based on CRRA preferences since the previous literature has argued, and shown with calibration exercises, that there is no need for more complex preferences to explain the EPP once rare disasters are taken into account. For instance, Epstein and Zin (1989) and CRRA preferences yield very similar implications for the EPP when the disaster frequency and intensity are calibrated in the same fashion – this can be seen for example by comparing the calibration results of Wachter (2011) with the ones of Barro (2006) (nevertheless, the former paper shows that, with recursive preferences, the stock market excess volatility can also be explained). Moreover, Bansal, Kiku, and Yaron (2010) show that, in the Long-Run Risk framework with Epstein-Zin preferences, large cyclical disaster risk has a trivial effect on the risk premium.<sup>3</sup> However, both Bansal, Kiku, and Yaron (2010) and Drechsler and Yaron (2011) show that small (of the order of  $-2.5 \times 10^{-4}$ , i.e. not disaster-like) infrequent jumps in the persistent component of consumption and dividend growth can have large asset pricing implications.

The remainder of the paper is organized as follows. Section 1 presents the theoretical underpinnings of the estimation and testing approaches considered. A data description is provided in Section 2. Estimation and testing results are presented in Section 3. Section 4.1 presents the rare events distribution of the data needed to rationalize the EPP with a low level of risk aversion. In Section 4.2, we ask what would be the likelihood of observing an EPP, in samples of the same size as the historical ones, if the rare events hypothesis were the true explanation of the puzzle. In Section 4.3, 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 5 concludes. Additional robustness checks and methodological details are provided in the Appendix.

## 1 Methodology

Our analysis is based on the C-CAPM of Rubinstein (1976) with time-additive power utility, in which the optimizing behavior of the representative agent leads to the consumption Euler equation

$$E^F \left[ (C_t/C_{t-1})^{-\gamma_0} \mathbf{R}_t^e \right] = \mathbf{0}, \quad \gamma \in \Theta \subset \mathbb{R}^+, \quad (1)$$

where the unconditional expectation is taken with respect to the unknown (true) probability measure  $F$ ,  $C_t$  denotes the time  $t$  consumption flow,  $\mathbf{R}_t^e \in \mathbb{R}^k$  is a vector of asset returns in excess of the risk free rate, and  $\gamma_0$  is the true – unknown – value of the RRA coefficient.

---

<sup>3</sup>However, both Bansal, Kiku, and Yaron (2010) and Drechsler and Yaron (2011) show that small (of the order of  $-2.5 \times 10^{-4}$ ), i.e. not disaster-like, infrequent jumps in the persistent component of consumption and dividend growth can have large asset pricing implications.

The standard inference approach for this model is to use consumption and stock market data to estimate the risk aversion parameter in Equation (1) as

$$\hat{\gamma} := \arg \min_{\gamma \in \Theta} g \left( E^T \left[ (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e \right], E^T \left[ \mathbf{R}_t^e \right], E^T \left[ (C_t/C_{t-1})^{-\gamma} \right] \right), \quad t = 1, \dots, T, \quad (2)$$

for some function  $g(\cdot)$ , where  $T$  denotes the sample size and  $E^T[x_t] \equiv \frac{1}{T} \sum_{t=1}^T x_t$  is an estimate of the unconditional moment  $E^F[x_t]$ , i.e. the distribution of the pricing kernel and returns is proxied with an empirical distribution that assigns probability  $p_t = 1/T$  to each realized state (observation) in the sample. The GMM inference, for example, belongs to this framework (see e.g. Hansen (1982)). The replacement of the unconditional moments with their sample analogs is justified by weak law of large numbers and central limit theorem arguments. However, 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 moments. Consequently, inference based on Equation (2) might be unreliable.<sup>4</sup>

This issue is particularly worrisome for the estimation of the RRA coefficient: if in a given sample economic disasters occur with a frequency lower than their true probability, the estimators defined in Equation (2) will tend to rationalize the realized risk premium by postulating a higher risk aversion than its true value. Therefore, a rejection of the C-CAPM based on a very high estimate of the risk aversion coefficient might simply be the consequence of the usage of estimators that constrain the probabilities ( $p_t$ ) attached to different states of the economy to equal their sample frequencies ( $1/T$ ).

Inference in a framework that relaxes the  $p_t = 1/T$  constraint is exactly what the estimation approaches used in our paper provide. Moreover, as explained in Section 1.2, these estimation approaches can be modified to develop a novel information-theoretic calibration method. This calibration procedure avoids parametric assumptions about the distribution of the data and makes the calibrated distribution as close as possible to the true, unknown, distribution of the data.

## 1.1 Estimation Method

We use two estimation and testing methods: the Empirical Likelihood (EL) of Owen (1988, 1990, 1991) and the Bayesian Exponentially Tilted Empirical Likelihood (BETEL) of Schennach (2005).<sup>5</sup> These estimators, and related test statistics, tend to have better small sample and asymptotic properties than the standard GMM approach (see e.g. Kitamura (2006)) and also allow us to perform Bayesian posterior inference (see

---

<sup>4</sup>Saikkonen 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. Also, Kocherlakota (1997) shows that if the marginal distribution of the shocks is too heavy-tailed, standard GMM testing of the Euler equation is unreliable.

<sup>5</sup>These approaches can be applied to both *i.i.d.* or weakly dependent data (see e.g. Kitamura (1997) for a definition of weak dependence). Moreover, two other related approaches, namely the Exponential Tilting (*ET*) of Kitamura and Stutzer (1997) and the Bayesian Empirical Likelihood (*BEL*) of Lazar (2003), produce very similar results that are available upon request.

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.

Consider first the EL approach. In a parametric maximum likelihood (ML) procedure, the data are assumed to have been generated by a parametric probability distribution that is a function of a vector of unknown parameters. The econometrician estimates the values of the parameters from an observed data sample by maximizing the likelihood of the sample. The EL estimator has the interpretation of a *non-parametric* ML estimator. In this approach, instead of assuming a parametric distribution for the data, the value of the probability mass at each sample point, denoted by  $p_t$ , is treated as a parameter to be estimated so as to maximize the non-parametric log likelihood for a multinomial model subject to the moment restriction given by the consumption Euler equation (1). In particular, the EL estimator parametrizes the moment condition (1) with  $(\gamma, p_1, p_2, \dots, p_T) \in \Theta \times \Delta$ , and is given by

$$(\hat{\gamma}^{EL}, \hat{p}_1^{EL}, \dots, \hat{p}_T^{EL}) = \arg \max_{\{\gamma, p_1, \dots, p_T\} \in \Theta \times \Delta} \ell = \sum_{t=1}^T \log(p_t) \quad \text{s.t.} \quad \sum_{t=1}^T p_t (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e = \mathbf{0}, \quad (3)$$

where  $\Delta$  denotes the simplex  $\{(p_1, \dots, p_T) : \sum_{t=1}^T p_t = 1, p_t \geq 0, t = 1, \dots, T\}$ ,  $\ell$  is the non-parametric log-likelihood function, and the ML estimator of the unknown probability measure  $F$  is given by  $\hat{F}^{EL} \equiv \sum_{t=1}^T \hat{p}_t^{EL} \delta_{\mathbf{z}_t}$ , where  $\delta_{\mathbf{z}_t}$  denotes a unit mass at the empirical observation  $\mathbf{z}_t \equiv [C_t/C_{t-1}, \mathbf{R}_t^e]$ .

The EL estimator has also 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  $\mathbb{R}^+ \times \mathbb{R}^k$ . For each parameter value  $\gamma \in \Theta$ , define the set of (absolutely continuous with respect to  $F$ ) probability measures that satisfy the consumption Euler equation:  $P(\gamma) \equiv \{p \in M : E^p [(C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e] = \mathbf{0}\}$ . Therefore,  $\mathcal{P} \equiv \cup_{\gamma \in \Theta} P(\gamma)$  is the set of all the probability measures that are consistent with the model characterized by the moment conditions in Equation (1). The EL estimation can then be shown to select a  $\gamma$ , and a probability measure  $p$ , so as to

$$\inf_{\gamma \in \Theta} \inf_{p \in P(\gamma)} K(F, p) \equiv \inf_{\gamma \in \Theta} \inf_{p \in P(\gamma)} \int \log(dF/dp) dF, \quad \text{s.t.} \quad E^p [(C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e] = \mathbf{0}, \quad (4)$$

where  $K(F, p)$  is the Kullback-Leibler Information Criterion (KLIC) divergence from  $F$  to  $p$  (White (1982)). Therefore,  $K(F, p) \geq 0$  and it will hold with equality if and only if  $F = p$ . If the model is correctly specified, i.e. if there exists a  $\gamma_0$  satisfying the consumption Euler equation (1), we have that  $F \in P(\gamma_0)$ , and  $F$  solves (4) delivering a KLIC value of 0. On the other hand, if the model is misspecified,  $F$  is not an element of  $\mathcal{P}$  and for each  $\gamma$  there is a minimum KLIC distance  $K(F, p) > 0$  attained by the solution of the inner minimization in Equation (4). Thus, the EL approach searches for an estimate of  $F$  that makes the estimated distribution as close as possible – in the information sense – to the true unknown one.

Moreover, since the EL estimator is the solution to a convex optimization problem, Fenchel duality applies (see e.g. Borwein and Lewis (1991)), therefore reducing dramatically the dimensionality of the optimization problem (see Appendix A.1.2).

To first order, the EL estimator is asymptotically equivalent to the optimal GMM estimator (see e.g. Qin and Lawless (1994) and Appendix A.1.1). However, Newey and Smith (2004) show that this estimator has smaller second-order bias than GMM, and 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.

Besides the desirable *local* asymptotic efficiency mentioned above, the EL 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 deviations* 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, i.e. the *elr* test is *uniformly most powerful*.<sup>6</sup> 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 a finite sample, the estimator is likely not to lie in a close neighborhood of the truth).<sup>7</sup>

Likelihood based testing is also possible within the EL framework (Owen (1991, 2001)). In particular, a joint test of the over-identifying restrictions in Equation (1) and the parameter restriction  $\gamma_0 = \bar{\gamma}$  may be performed by using the empirical likelihood ratio statistic:

$$elr(\bar{\gamma}) = -2[\ell(\bar{\gamma}) - (-T \log(T))]$$

where  $\ell(\bar{\gamma})$  is the log-likelihood in Equation (3) evaluated at  $\gamma = \bar{\gamma}$ , and the second term is the value of the maximized log-likelihood function in the absence of any moment restriction (i.e. at  $p_t = 1/T \forall t$ ). The above likelihood ratio statistic has an asymptotic  $\chi^2_{(k)}$  distribution under the null.

Probability weights obtained via entropy minimization can also be used to perform Bayesian posterior inference. Schennach (2005) shows that, given a prior  $\pi(\gamma)$ , a Bayesian (Bayesian Exponentially Tilted Empirical Likelihood, BETEL) posterior can be formed as

$$p\left(\gamma \mid \{C_t/C_{t-1}, \mathbf{R}_t^e\}_{t=1}^T\right) \propto \pi(\gamma) \times \prod_{t=1}^T \hat{p}_t^{BETEL}(\gamma), \quad (5)$$

where the  $\hat{p}_t^{BETEL}(\gamma)$  are obtained as the solution to

$$\inf_{\gamma \in \Theta} \inf_{p \in P(\gamma)} K(p, F) = \inf_{\gamma \in \Theta} \inf_{p \in P(\gamma)} \int \log(dp/dF) dp, \quad \text{s.t. } E^p[(C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e] = \mathbf{0}. \quad (6)$$

The posterior distribution in Equation (5) naturally arises as the non-parametric limit of a Bayesian procedure that places a type of non-informative prior on the space of

<sup>6</sup>This property is sometimes referred to as the generalized Neyman-Pearson optimality.

<sup>7</sup>See also the weak law of large numbers for rare events of Brown and Smith (1990) as a rationale for relative entropy estimators.

distributions.<sup>8</sup> Also, since the probability weights in Equation (6) are the solution to a convex optimization problem, the solution to the inner minimization problem in Equation (6) can be easily obtained from the dual optimization (see Appendix A.1.2).

Note that in our estimations we focus on the unconditional version of the consumption Euler equation, i.e. Equation (1). That is, we focus on the unconditional probability distribution of disasters. This unconditional approach has the advantage that it does not rule out the possibility of a time-varying conditional distribution of disasters. That is, our approach would be consistent even in a setting with CRRA preferences and time varying probability of disasters as in Gabaix (2012). Also, focusing on the unconditional restrictions and using a non-parametric modeling of the unconditional distribution, our approach is robust to potential regime switching in the consumption and returns dynamics, as long as the regime switching process has a stationary unconditional distribution.

## 1.2 Calibration Method

The relative entropy minimizations in Equations (4) and (6) involve two nested minimizations. The inner minimizations identify, given the preference parameters, the probability measure that is the closest to the true, unknown data generating process. This implies that, by dropping the outer minimization and fixing, rather than estimating, the preference parameters, one can use the inner entropy minimization to provided a non-parametric calibration of the data generating process in the model. The calibrated probability measures identified in this fashion can then be used to simulate the model. To the best of our knowledge, we are the first to propose such an information-theoretic approach to calibration. This approach, that can be applied to *any* economic model that delivers well defined moment conditions, has the appealing feature of making the calibrated model as close as possible – in the information sense – to the true unknown one, and enables model evaluation that is free from ancillary distributional assumptions. In a nutshell, we provide a maximum likelihood calibration method in which only the deep parameters (e.g. preferences, elasticities, etc.) need to be fixed exogenously, while instead the data generating process is calibrated endogenously using the data in a fashion that maximizes the likelihood of the model to be the true model of the economy.

## 2 Data Description

### 2.1 U.S. Data

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 total and nondurable consumption series, we focus on two annual data samples: a baseline data sample starting at the onset of the Great Depression (1929-2009)

---

<sup>8</sup>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  $T \rightarrow \infty$ .



and a longer data set (1890-2009) obtained from Campbell (2003) and Robert Shiller’s website. We use the shorter sample as our baseline since only over this period total consumption can be disaggregated into its nondurable and durable components, and we use the longer sample as a robustness check.

For the 1929-2009 data sample, 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 proxy for the risk free rate is the one-month Treasury Bill rate. Annual returns for the above assets are computed by compounding monthly returns within each 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).

For the longer data set, the return on the S&P composite index is used as a proxy for the market return. Due to data availability issues, we use the prime commercial paper rate as a proxy for the risk free rate, therefore partially underestimating the magnitude of the EPP. Consumption refers to the real per capita *total* personal consumption expenditures. See Campbell (1999, 2003) for a detailed data description.

We make the standard “end-of-period” timing assumption that consumption during quarter  $t$  takes place at the end of the quarter.

For the cross-sectional analysis, we focus on a variety of portfolios that are widely used in financial economics and that are described in detail in Appendix A.2.

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, both of our data samples include two out of the 65 major rare economic disasters of the twentieth century identified in Barro (2006):<sup>9</sup> the Great Depression (1929-1933) and the World War II aftermath (1944-47). The economic contractions associated with these two episodes (respectively, a 31% and 28% drop in GDP per capita) are both much larger than the median contraction during disasters (this being a 24% drop in GDP per capita in Table I of Barro (2006)), that is they are among the *worst* disasters of the twentieth century. Moreover, the U.S. consumption contraction during the Great Depression is also above the median of the 84 major consumption disasters recorded since the early nineteenth century (Barro and Ursua (2008a)).<sup>10</sup> Moreover, in our longer sample, Barro and Ursua (2008b) (that extends the Barro (2006) original data set) identify *five* disasters in the U.S.

Second, in our baseline annual sample we observe 11 out of the 15 major stock market crashes of the twentieth century identified by Mishkin and White (2002) plus the 2002 and 2008 market crashes, while in the longer sample we have all the 15 market crashes.<sup>11</sup> Moreover, these include the largest one-day decline in stock market values in

---

<sup>9</sup>Barro (2006) studies a set of 35 countries with GDP data from Maddison (2003), and identifies a disaster as a peak-to-trough cumulated contraction in GDP of at least 15%.

<sup>10</sup>Barro and Ursua (2008a) study an unbalanced panel of 21 countries that provides a total of 2638 yearly observations, and identify a consumption disaster as a peak-to-trough cumulated reduction in consumption of at least 10%.

<sup>11</sup>Mishkin and White (2002) identify a stock market crash as a period in which either the Dow Jones Industrial, 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.

U.S. history, and all of the 10 largest contractions of the Dow Jones Industrial Average index during the twentieth century.<sup>12</sup>

## 2.2 International and Disasters Data

Even though our U.S. data include some of the largest disasters ever recorded in the world history, a few other countries have experienced larger disasters than the U.S. As a consequence we also employ a variety of international data sources.

First, we use the Campbell (2003) international data set of consumption and stock market data that gives us a long time series for the U.K. (that includes two of the largest disasters of the twentieth century identified in Barro (2006)), and shorter post war samples for seven additional OECD countries.

Second, Barro and Ursúa (2009, Table 2, Panel A) provides us data on both consumption and stock market disasters for a large cross section of countries. The authors, looking at a cross-section of 25 countries since the early 1870s, document 58 events of concurrent stock market and consumption disaster (defined as multi-year real returns of  $-25\%$  or less and multi-year macroeconomic declines of  $10\%$  or more).

## 3 Estimation Results

In this section we present estimation and testing results for the consumption Euler Equation (1) using the methodology described in Section 1. In Subsection 3.1 we present results based on U.S. data, while in Subsection 3.2 we focus on international data.

### 3.1 U.S. Evidence

Table 1 shows the estimation results for the consumption Euler Equation (1) for the baseline data sample (1929-2009). Panel A presents results when the set of assets consists of only the excess return on the market portfolio.

The first row reports the point estimates of the RRA coefficient  $\gamma$ . The EL point estimate is 28.5. The standard error of the estimate (in parentheses) shows that the point estimate is statistically larger than 10 (the upper bound of the “reasonable” range for the RRA coefficient) at standard confidence levels. The BETEL posterior distribution of this parameter (computed under an improper uniform prior on  $\gamma \in \mathbb{R}^+$ ) also peaks at a high value of 28.5. Moreover, the posterior 95% confidence interval (in square brackets) does not include values of  $\gamma$  smaller than 11.5.

The second row reports a test for the joint hypothesis of a  $\gamma$  as small as 10 and for the identifying restriction given by the consumption Euler Equation (1). This is the Empirical Likelihood Ratio (*elr*) test of Owen (1991, 2001). Under the null hypothesis, the test statistic follows asymptotically a  $\chi^2$  distribution with one degree of freedom. As revealed by the *p*-value reported in parentheses below the test statistic, the test

---

<sup>12</sup>Source: Dow Jones.

rejects the hypothesis of the Euler equation being satisfied by a  $\gamma$  as small as 10 with  $p$ -value smaller than 4%.

Finally, the third row reports the posterior probability of  $\gamma$  being smaller than, or equal to, 10 given the observed data. This probability is very small at 1.67% for the BETEL posterior. A very similar posterior probability of only 1.89% is obtained when the profile EL function is used as the likelihood part of the Bayes theorem along with an improper uniform prior on  $\gamma \in \mathbb{R}^+$  to obtain the posterior distribution of the RRA coefficient (hereafter referred to as the BEL posterior).

\*\*\* Table 1 about here \*\*\*

While the information-theoretic estimation approaches that we use are theoretically robust to a rare events problem in the data, a natural question is whether they retain their superior properties in small samples in the presence of rare events. We assess this issue in Section A.3.1 of the Appendix. In particular, we generate samples, of the same size as the historical one, using a calibrated rare events model. We calibrate the model in two ways. First, we generate samples from the model and distribution of disasters of Barro (2006). Second, we generate samples using the information-theoretic calibration approach described in Sections 1 and 4.2. In each generated sample, we perform the EL and BETEL estimation, and with this small sample distribution of the parameter estimates we are able to compute finite sample  $p$ -values for the point estimates in Panel A of Table 1. The small sample  $p$ -values of the above point estimates arising if the true  $\gamma$  were smaller than 10 range from 3.95% to 4.33% (see Table A1, Panel A, Columns 1 and 4). That is, if the data were generated by a rare events distribution, the EL and BETEL estimators would be very unlikely to deliver the large estimates of the RRA coefficient obtained in the historical data because of a small sample issue.

In Panel B, we estimate the model using as an additional moment restriction the Euler equation for the risk free rate. This challenges the model to explain both the EPP and the risk free rate puzzle as well as requires the estimation of the subjective discount factor parameter,  $\delta$ , in addition to the risk aversion coefficient. Following the work of Kocherlakota (1996), our prior for  $\delta$  is flat on  $\mathbb{R}^+$  i.e. we do not restrict ex ante the coefficient to be smaller than 1. The results are very similar to those in Panel A – the EL point estimate of the risk aversion coefficient is high at 28.5, the  $\chi^2$  test rejects the model with  $p$ -value 2.0%, the BETEL posterior distribution peaks at  $\gamma = 28.6$ , the posterior 95% confidence interval is asymmetric, attaching high posterior probabilities to high values of  $\gamma$  and low ones to values smaller than 10 and does not include values of  $\gamma$  smaller than 14.5, and the posterior probability of  $\gamma$  being smaller than, or equal to, 10 given the observed data is at most 0.54%. Therefore, the results in Panel A do not seem to be driven by the exclusion of the risk free rate.

Panel C reports estimation and testing results when the set of assets consists of the excess returns on the 6 Fama-French portfolios formed from the intersection of 2 size-sorted and 3 book-to-market-equity-sorted portfolios (see Fama and French (1992)). The rejection of the model is even stronger in this case. The EL point estimate of  $\gamma$  is even higher at 34.2 and statistically larger than 10 at conventional significance levels. The  $\chi^2$  test strongly rejects the model with  $p$ -value 0.0%. The mode of the BETEL posterior distribution is attained at  $\gamma = 30.3$  and the posterior 95% confidence interval

does not include values of  $\gamma$  smaller than 14.7. Finally, the posterior probability of  $\gamma$  being smaller than, or equal to, 10 given the observed data is at most 0.53%.

Note that, the rejection of the model in Table 1, differently from a rejection based on a GMM estimator or any other estimator that assigns equal weight to each data point, cannot be ascribed to a rare events problem in the data. If in a given sample economic disasters were to occur with a frequency lower than their true probability, an estimator (like GMM) that assigns a  $1/T$  weight to each data point will tend to rationalize the realized risk premium by postulating a higher risk aversion, and the rejection of the model could be ascribable to an undersampling of disasters. The estimates and tests in Table 1, and the resulting rejection of the model, are based instead on an endogenous (optimal) reweighting of each individual observation, and are therefore robust to a potential undersampling of disasters.<sup>13</sup>

\*\*\* Table 2 about here \*\*\*

The results in Table 1 are obtained using real per capita personal consumption expenditure on nondurable goods, under the assumptions that the decision interval of the representative consumer is a year and that consumption adjusts instantaneously to market return innovations. In Table 2, we investigate whether the rejection of the model in Table 1 is due to using an incorrect measure of consumption or a violation of either of the two assumptions.

Panel *A* presents results when consumption is measured as the real per capita *total* (durables, nondurables, and services) personal consumption expenditure, data for which is available for a longer sample period 1890-2009. The EL point estimate of  $\gamma$  is 49.3 – even higher than those obtained in Table 1. Even though the standard error of the estimate is too large to reject a  $\gamma$  smaller than 10 at standard confidence levels, this is due to the fact that the likelihood is quite flat for high values of  $\gamma$  – despite being very steep for low values – making the (symmetric) Gaussian asymptotics misleading. The asymmetry of the likelihood function is illustrated in Figure 1 that plots the profile empirical likelihood as a function of the risk aversion coefficient. Indeed, Row 2 shows that the  $\chi^2$  test strongly rejects the model with a  $p$ -value smaller than 1% – and this is due to the fact that the Empirical Likelihood ratio test, unlike the Gaussian asymptotic standard errors, by construction correctly takes into account the asymmetry of the likelihood function. The BETEL posterior distribution of this parameter also peaks at a very high value of 49.3. Moreover, the posterior 95% confidence interval does not include values of  $\gamma$  smaller than 23.3. Also note that the Bayesian confidence interval shows that the likelihood of the data is very asymmetric, attaching high posterior probabilities to high values of  $\gamma$  and very low ones to values smaller than 20, therefore generating too large frequentist standard errors. In fact, the third row of Panel *A* shows that the posterior probability of  $\gamma$  being smaller than, or equal to, 10 given the observed data is at most only 0.11%. As for the results in Table 1, we compute small samples  $p$ -values for the point estimates of  $\gamma$  being smaller than 10 (see Table *A1*,

---

<sup>13</sup>In Table *A2* of the Appendix, as a robustness check, we report GMM estimates corresponding to the ones in Table 1 and find similar results. Note, however, that the GMM procedure tends to yield large average pricing errors for the estimations in Panels *B* and *C*.

Panel B), and find values in the 4.45%-4.93% range. That is, the estimates in Table 2, Panel A, are very unlikely to be the consequence of a small sample problem.

Parker and Julliard (2005) show that testing the C-CAPM focusing on the short run correlation of consumption and returns might miss important dynamics due to slow consumption adjustment to market innovations. As a consequence, as an additional robustness check, we follow Parker and Julliard (2005) and measure consumption risk by the covariance of the market return and consumption growth cumulated over many periods following the return.<sup>14</sup> We report results for a horizon of eight years, because the maximum duration of a disaster in the historical world data is 8 years (Barro and Ursúa (2009, Table 2, Panel A)), but obtain very similar results for all intermediate horizons. Panel B shows that using this more robust approach the model is still strongly rejected: the EL point estimate of  $\gamma$  is very high at 39.3 (but with misleading large Gaussian standard errors due to the asymmetry of the likelihood) and the  $\chi^2$  test (correctly taking into account the asymmetry of the likelihood) rejects the model with  $p$ -value 1.9%, the BETEL posterior distribution peaks at  $\gamma = 39.3$ , the posterior 95% confidence interval is asymmetric and does not include values of  $\gamma$  smaller than 28.3, and the posterior probability of  $\gamma$  being smaller than, or equal to, 10 given the observed data is 0.0%.<sup>15</sup>

\*\*\* Figure 1 about here \*\*\*

Finally, Panel C reports results similar to Table 1, Panel A, but at the 5-year, instead of the annual, frequency using overlapping 5-year consumption growth and returns. This gives the C-CAPM a better chance to explain the EPP based on the finding in Brainard, Nelson, and Shapiro (1991) that the longer the horizon of the investor, the better the C-CAPM performs relative to the CAPM. Panel C shows that the EL point estimate is higher than that at the annual frequency (32.9 vs. 28.5) and statistically larger than 10 at standard confidence levels. The  $\chi^2$  test rejects the model with a  $p$ -value of 0.0%. The BETEL posterior distribution of  $\gamma$  also peaks at a high value of 32.9 and the posterior 95% confidence interval does not include values smaller than 24.6. Moreover, our results are robust to potential time-aggregation effects in that we obtain very similar results if the multiperiod observation corresponding to the Great Depression is replaced by the minimum excess return on the market and the minimum consumption growth observed during the period.<sup>16</sup>

Compared to the previous literature, we find that overall our estimates of the risk aversion parameter are larger – although not statistically different – than the ones of Kocherlakota (1996) (this difference is attributable to the different proxies for the risk free rate and our longer sample) but much smaller than the ones reported in Parker (2001) that focuses on the post WWII sample.

---

<sup>14</sup>That is, instead of the Euler Equation (1), we use the moment restriction  $E^F[R_{t,t+S}^f (C_{t+S}/C_{t-1})^{-\gamma_0} \mathbf{R}_t^e] = \mathbf{0}$ , where  $S > 0$  is the number of periods over which consumption risk is measured, and  $R_{t,t+S}^f$  is the risk free rate between time  $t$  and  $t + S$ .

<sup>15</sup>Note that, for this version of the C-CAPM, the conditional Euler equation error is no longer a martingale difference sequence. Therefore, we rely on the blockwise EL approach (see Kitamura (1997)), details of which are reported in an online Appendix.

<sup>16</sup>These results are available in an online Appendix.

Overall, the results in Tables 1 and 2 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 C-CAPM is still rejected and requires a very high level of RRA to rationalize the stock market risk premium.

## 3.2 International Evidence

In this section we check whether the rejection of the rare events explanation of the EPP is an artifact of having focused only on U.S. data. We do this in two ways.

First, we re-estimate the Euler Equation (1) for all the countries for which we have extended time series for consumption and stock market data. Second, since for many of the countries that experienced disasters we *do not* have the time series data needed to estimate the consumption Euler equation – but we have data on the sizes of the economic contractions during disasters – we ask whether the rejection of the rare events hypothesis could be due to the U.S. being relatively “lucky” in not experiencing a much larger disaster in its history.

### 3.2.1 Estimation Results with U.K. and other International Data

Here we report estimation results of the consumption Euler Equation (1) using U.K. annual data over the sample 1919-1994 and for seven more OECD countries using smaller samples. The data are the ones used in Campbell (2003).

\*\*\* Table 3 about here \*\*\*

The findings in Table 3, that focuses on U.K. data and covers a sample in which two of the largest world disasters of the twentieth century were recorded (see Barro and Ursua (2008b)), are in line with the ones in Tables 1 and 2: the estimate of the risk aversion coefficient is very high and statistically larger than 10, the C-CAPM is strongly rejected using the empirical likelihood ratio test statistic, and the posterior probability of the RRA coefficient being as small as 10, or smaller, is at most .12%.

In Table A3 of Appendix A.4 we re-estimate the consumption Euler Equation (1) using data on seven other OECD countries (Australia, Canada, France, Germany, Netherlands, Sweden and Japan) over the available samples. This robustness check largely confirms the above results: the RRA estimates range from 19.1 to 309.6 and are generally statistically larger than 10 (only in one estimate out of 14 the posterior probability of an RRA smaller than, or equal to, 10 is larger than 10%). However, note that the international data are often overlapping in time and, as a consequence, should not be considered independent across countries.

### 3.2.2 The World’s Largest Stock Market and Consumption Disasters

For many of the countries that, over the last two centuries, have experienced disaster events, we do not have the time series data needed to estimate the consumption Euler equation. Nevertheless, thanks to the work of Barro and Ursúa (2009), we have

estimates of the sizes of the economic contractions as well as stock market returns during disaster events for a large cross-section of 25 countries over samples that start as early as 1870. These authors document 58 events of concurrent stock market and consumption disaster (defined as multi-year real returns of  $-25\%$  or less and multi-year macroeconomic declines of  $10\%$  or more).<sup>17</sup> As a consequence, we can ask whether the rejection of the rare events hypothesis in our data samples could be due to the U.S. being relatively “lucky” in not experiencing much larger disasters in its history.

To answer this question, we modify our baseline annual data sample (1929-2009) by replacing the four data points corresponding to the U.S. Great Depression with calibrated disasters observations. Note that we choose to use the sample that delivers the weakest rejection of the rare events hypothesis (see Tables 1 and 2) and perform two exercises.<sup>18</sup>

In the first exercise, we bootstrap the possible disasters that could have occurred during the U.S. Great Depression using the international disasters data set. That is, we generate as many samples as the number of identified disasters (58). In each of these samples, the Great Depression contraction is calibrated to match the length, the consumption drop, and the stock market performance of one of the Barro and Ursúa (2009) identified disasters. For instance, for a 3 years disaster characterized by a cumulated consumption drop of  $30\%$  and a stock market return of  $-60\%$ , we add 3 equal data points that produce a cumulated consumption drop of  $30\%$  and a cumulated stock market return of  $-60\%$ . We then use the EL and BETEL estimators to perform inference in each of these samples. The rationale behind this exercise is that we want to assess whether the empirical rejection of the rare events explanation of the EPP presented in Section 3.1 was due to the fact that the U.S. has been relatively “lucky” not experiencing a much larger disaster during the Great Depression period. Note that we are not viewing the recorded disasters in different countries as independent events (and, indeed, they are not) but only as potential disasters that could have happened in the US. By performing the estimation in all these samples we generate a sampling variation for the RRA estimates that can be compared with the point estimates obtained in Section 3 using the true U.S. data.<sup>19</sup>

\*\*\* Figure 2 about here \*\*\*

The distribution of the EL estimates of the RRA coefficient is reported in Figure 2. Several things are worth noticing. First, in *none* of the 58 samples is the estimated risk

---

<sup>17</sup>Note that these matched international disasters are likely, by construction, to overestimate the consumption downturn during a disaster since: *a*) the macro contraction is measured using GDP data when consumption data are not available, and *b*) for wartime contractions (the largest disasters), “when C [consumption] and GDP data were available, the macro contraction was based on C when this decline was greater than that in GDP. When the GDP decline was larger, the macro contraction was computed as the average of the C and GDP contractions” (Barro and Ursúa (2009), page 31).

<sup>18</sup>Very similar results are obtained using the 1890-2009 total consumption sample, and are available from the authors upon request.

<sup>19</sup>Moreover, the procedure of adding calibrated disasters to the true sample can be viewed as a Bayesian dummy prior observation approach in which the calibrated disasters play the role of priors about how big the consumption and stock market economic downturn could be during an economic disaster.

aversion coefficient smaller than 10. That is, if the U.S. had experienced *any* of the joint consumption and stock market disasters recorded in the data, the equity premium would still be too large to be rationalized with a small RRA coefficient. Second, the median estimate across samples, being about 41, is even higher than the estimate in the true U.S. data sample. Moreover, the centered 95% area of the distribution of RRA estimates ranges from 16.0 to 72.7. Third, about 67% of the estimated RRA coefficients are higher than the point estimate in the true U.S. sample. Fourth, none of the point estimates lies below the 95% frequentist confidence band of the EL estimator in the true sample. Fifth, in each sample we compute the joint  $\chi^2$  test of the consumption model being true and the Euler equation being satisfied with a RRA smaller than 10, and we find that in only 10.3% of the cases we would not be able to reject the model at the 10% confidence level. Sixth, in each sample we compute the posterior probability of the Euler equation being satisfied with a RRA coefficient smaller or equal to 10, and we find that in only 6.9% of the samples this probability is higher than, or equal to, 5%. Overall, these results stress the soundness of the empirical rejection of the rare events explanation of the EPP obtained in Section 3 using U.S. data: even after bootstrapping the possible disasters from a large international cross section of data, we still reject the rare events explanation of the EPP at standard confidence levels.

In the second exercise, we construct a counterfactual sample that emulates the “standard calibration” (SC) approach pioneered by Barro (2006) (e.g. Gabaix (2012) and many others), and that finds strong support for the rare events explanation of the EPP. But, differently from this literature, we take formally into account that several of the disasters detected in the data, instead of occurring only in one period, lasted several years.

The SC approach calibrates the consumption process to *a*) mimic the U.S. data in the non-disaster states, while *b*) it assumes that the consumption drop during a disaster is uniformly drawn from the data set of economic disasters identified in a cross-section of countries. In order to follow the same approach, we construct a counterfactual sample that *a*) contains the observations of our baseline U.S. data sample (1929-2009) without the data points corresponding to the Great Depression period (in order to mimic the non-disaster states), and *b*) contains *all* the joint consumption and stock market disasters identified in Barro and Ursúa (2009). We then modify the relative sample frequency of disaster and non-disaster periods to make our results comparable with the SC approach (that generally allows to rationalize the EPP with a probability of disasters of about 1.7% and a RRA of about 4). These calibrated samples are then used to estimate the RRA coefficient using the EL and BETEL estimators.

The results of this exercise using the EL estimator are reported in Figure 3, where we measure the RRA estimates on the vertical axis and the sample frequency of disaster states on the horizontal axis. In the figure we report two sets of estimates. The first one, denoted by the solid line with circles, is obtained taking into account that the contractions during disasters are spread over several years (we do so by adding, for each disaster, as many observation as the disaster length in years, and spreading the contraction over these multiple periods). The second one instead, denoted by the dashed line with triangles, is obtained disregarding the fact that several disasters lasted



for multiple periods, i.e. we add one observation per disaster and calibrate the one year economic contraction as being equal to the multiple years contraction. For each of these sets of estimates we also report frequentist and Bayesian 95% confidence bands. We also single out the canonical calibration of the RRA (4) and disaster probability (1.7%) of the SC approach (denoted, respectively, by the horizontal and vertical dash dotted lines).

\*\*\* Figure 3 about here \*\*\*

Focusing on the estimates based on the calibration that correctly takes into account the multi year nature of economic disasters (continuous line with circles), we see that at the SC value of the disaster probability the estimated risk aversion is in the mid twenties and the confidence bands are way above 10. Moreover, the figure indicates that a sample frequency of a disaster starting of about 9.6% – that is, *a disaster starting every 10.4 years* – is needed to explain the equity premium with a risk aversion of 10, and an even more extreme value – a probability of disaster of about 15.1% i.e. *a disaster every 6.6 years* – is needed to be able to explain, as the literature based on the SC does, the equity premium with a RRA as small as 4. Interestingly, looking at the dashed line with triangles – based on having erroneously assumed, as in the SC approach, that all the disasters had their full impact on the economy in only one year – we obtain a RRA estimate of about 4 with a sample frequency of disaster as small as 1.7%.

These set of findings suggest that: *a) the economic disasters present in the data of a large cross-section of countries do not offer support for the rare events explanation of the equity premium puzzle unless we are willing to believe that economic disasters should be happening every 6-10 years; b) the opposite conclusion, drawn by the literature that has followed the “standard calibration” approach, is due to having calibrated one year contractions during disasters as being equal to the cumulated multi year contractions recorded in the data.*<sup>20</sup>

The second point above is crucial. Indeed, in an online Appendix, we show formally that the discrepancy between the results obtained in this paper and the ones obtained in the literature that employs a calibration approach à la Barro (2006) is entirely driven by the counterfactual assumption that disasters have their full impact on the economy in one period, while instead in the data – the same data that this literature uses to calibrate the magnitude of the disasters – disasters always last multiple periods.<sup>21</sup>

---

<sup>20</sup>Using the Barro and Ursua (2008b) data set, that provides estimates of the sizes of the economic contractions during disaster events for a large cross-section of 40 countries over samples that start as early as 1800, we also show that the consumption disasters that we should have observed in the U.S. to rationalize the EPP with a low level of risk aversion are much larger than the largest consumption disasters ever recorded in the world history. These results are available in an online Appendix.

<sup>21</sup>Note that the results presented in this subsection are based on the EL estimator. However, analogous results are obtained for the BETEL estimator, and are available in an online Appendix.

## 4 Counterfactual Analysis

In this section, instead of jointly estimating the coefficient of RRA,  $\gamma$ , and the probabilities associated with different states of the economy, we fix the  $\gamma$  parameter to a “reasonable” value, and ask the EL and BETEL estimation procedures to identify the distribution of the data that would solve the EPP in the historical sample. 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 EPP with a low level of risk aversion. Under this distribution, what would be the probability of observing an EPP in a sample of the same size as the historical one? That is, if rare events that did not happen frequently enough in the historical sample were the true reason behind the EPP, what would be the likelihood of observing such a puzzle? Second, suppose rare events were the cause of the EPP. 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 4.1, we present the constructed rare events distribution of the data, while its implications for the likelihood of observing an EPP and for the cross-section of asset returns are discussed, respectively, in Sections 4.2 and 4.3.

### 4.1 A World without the Equity Premium Puzzle

To see how the EL and BETEL estimation procedures can be used to calibrate a distribution of the data that rationalizes the EPP, note that the consumption Euler Equation (1) implies the following identity

$$\frac{E^F [(C_t/C_{t-1})^{-\gamma} R_t^e]}{E^F [(C_t/C_{t-1})^{-\gamma}]} \equiv \underbrace{E^F [R_t^e] + \frac{Cov^F [(C_t/C_{t-1})^{-\gamma}, R_t^e]}{E^F [(C_t/C_{t-1})^{-\gamma}]}}_{=: epp^F(\gamma)}, \quad (7)$$

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

$$\sum_{t=1}^T (C_t/C_{t-1})^{-\gamma} R_t^e \hat{p}_t^j(\gamma) = 0, \quad \forall \gamma. \quad (8)$$

Therefore, denoting by  $\hat{P}^j(\gamma)$  the probability measure defined by  $\{\hat{p}_t^j(\gamma)\}_{t=1}^T$ , we have that

$$E^{\hat{P}^j(\gamma)} [(C_t/C_{t-1})^{-\gamma} R_t^e] = 0 \Rightarrow epp^j(\gamma) = 0, \quad (9)$$

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

That is, fixing the RRA coefficient  $\gamma$ , we can use the EL and BETEL procedures to construct the probability distribution needed to solve the EPP. As discussed in Section 1, this procedure is consistent. Moreover, this calibration approach minimizes the Kullback-Leibler divergence between the calibrated distribution and the unknown data generating process. That is, in the same fashion as a Maximum Likelihood estimator, this approach minimizes 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 EPP, 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 EPP, the estimated  $\hat{P}^j(\gamma)$  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 RRA at the upper bound of what is commonly considered the “reasonable” range for this parameter (e.g. Gollier (2002), Mehra and Prescott (1985)).<sup>22</sup>

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 EPP assigns relatively higher weights to a few particularly bad states of the economy. Figure 4 suggests that this is exactly what the estimated  $\hat{P}^j(\gamma)$  do.

\*\*\* Figure 4 about here \*\*\*

Figure 4 reports EL and BETEL probability estimates, NBER recession periods (shaded areas), and the major stock market crashes identified by Mishkin and White (2002) plus the 2002 and 2008 market crashes (vertical dash dotted lines).<sup>23</sup> Panel *A* reports estimated probability weights for annual data over the sample 1929-2009, while Panel *B* focuses on annual data over the longer sample 1890-2009.

Several features are evident in Figure 4. First, the EL and BETEL estimated weights are extremely similar – the correlation between the two estimates is above .94 in both the data sets considered – 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 annual sample over 1929-2009 is 37.5% while the EL and BETEL estimated probabilities of being in a recession period are, respectively, 41.3% and 40.7%. Similarly, in the 1890-2009 sample, the EL and BETEL probabilities increase the likelihood of being in a recession year by, respectively, 3.4% and 3.0%.

<sup>22</sup>We obtain very similar results, reported in an online Appendix, for  $\gamma = 4$ .

<sup>23</sup>Note that we classify a given year as a recession period if a NBER recession was registered in at least one of its quarters. Similarly, 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 during the period.

Third, the increases in the probabilities of observing a recession are largely driven by assigning higher probabilities to few recession periods that are concomitant with market crash episodes. Fourth, the EL and BETEL estimated distributions assign higher probabilities to most of the identified periods of stock market crash. The sampling frequency of stock market crashes in the 1929-2009 data is 21.3% while the EL and BETEL estimated probabilities of a stock market crash are, respectively, 28.1% and 27.9%. Similarly, in the 1890-2009 sample, the EL and BETEL probabilities increase the likelihood of a stock market crash in a year by, respectively, 6.1% and 5.8%. 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, the 1973-1975 recession, and the recent credit crisis. Nevertheless, even the probabilities attached to these states are still fairly small compared to the sampling frequency of the observations: for the annual data over 1929-2009 the sampling frequency is 1.3%, while the highest EL and BETEL probability weights are, respectively, 3.0% and 2.6%. Similarly, the sampling frequency in the 1890-1995 sample is 0.8%, while the highest EL and BETEL probability weights are, respectively, 2.8% and 2.0%.

\*\*\* Figure 5 about here \*\*\*

The implications of the estimated probability weights for the distribution of stock market real returns are summarized in Figure 5. The figure reports the histograms of annual stock market real returns over the period 1929-2009 (Panel A) and over the period 1890-2009 (Panel B), (Epanechnikov) kernel estimates of the empirical distribution, and weighted (Epanechnikov) kernel estimates, where the weights are given by the estimated  $\hat{P}^j(\gamma)$  probabilities.

The figure reveals that the rare events distribution needed to rationalize the EPP implies thicker negative tails, and a more left skewed distribution than what is obtained using the empirical (sample) weights. Moreover, the EL and BETEL probability weights generate a leftward shift in the distribution of returns when compared with the empirical distribution. This leftward shift implies a reduction in both the median and the mean stock market return: the implied annual median (mean) return is about 3.9%-7.0% (3.4%-4.1%) compared to 8.0%-11.7% (7.7%-8.4%) obtained using the sample weights. These numbers are in line with the rare events calibrated model of Barro (2006) that finds an expected risky rate in the range 3.7%-8.4%.

Rare events models stress that the EPP can be rationalized by assigning higher probabilities 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 6 shows that this is indeed an implication of the EL and BETEL estimated probability weights.

Each panel of Figure 6 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 annual sample over the period 1929-2009 and the

lower three panels correspond to the annual sample over the period 1890-2009. Panels *A* and *D* focus on the sample distributions, while Panels *B* and *E*, and Panels *C* and *F*, report, respectively, the EL and BETEL implied joint distributions (obtained by performing weighted kernel estimation with the weights given by the EL and BETEL 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 where recessions and stock market crashes tend to appear more often. Comparing the level curves in Panels *A* and *D* with the ones in the other panels, it appears clearly that the EL and BETEL probability weights skew the joint distribution of consumption growth and market returns toward the lower left portion of the graphs, thereby increasing the likelihood of high stock market consumption risk states. Moreover, most of the shift in probability mass happens on the lowest level curve i.e., in the tail of the joint distribution, as the rare events explanation of the EPP would imply.

\*\*\* Figure 6 about here \*\*\*

Overall, the results of this section suggest that using the EL and BETEL approaches to calibrate distributions of the data that rationalize the EPP with a low level of risk aversion, and that are at the same time 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 hypothesis predicts should be the mechanisms needed to rationalize the EPP.

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

## 4.2 How likely is the Equity Premium Puzzle?

The  $\hat{P}^j(\gamma)$ ,  $j \in \{EL, BETEL\}$ , measures just discussed provide the most probable (in the likelihood sense) rare events explanation of the EPP. But, under these measures, what is the likelihood of observing an EPP 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  $\hat{P}^j(\gamma = 10)$ ,  $j \in \{EL, BETEL\}$ , distributions to generate counterfactual samples of data of the same size as the historical ones. That is, we use the  $\{\hat{p}_t^j(\gamma)\}_{t=1}^T$ ,  $j \in \{EL, BETEL\}$ , probabilities to draw with replacement from the observed data  $\{C_t/C_{t-1}; R_t^e\}_{t=1}^T$ , and use these draws to form samples of size  $T$ . We generate a total of 10,000 counterfactual samples in this fashion (for both annual data sets).

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

$$epp_i^T(\gamma) = E^T [R_{i,t}^e] + Cov^T [(C_{i,t}/C_{i,t-1})^{-\gamma}, R_{i,t}^e] / E^T [(C_{i,t}/C_{i,t-1})^{-\gamma}], \quad (10)$$

where  $E^T [\cdot]$  and  $Cov^T [\cdot]$  denote the sample moment operators, and  $\gamma$  is fixed to the same low level used to construct  $\hat{P}^j(\gamma)$ .

The results of this counterfactual exercise are summarized in Table 4. Panels *A* and *B* focus on annual observations over the periods 1929-2009 and 1890-2009, respectively. The first column reports the EPP, as a function of  $\gamma$ , in the historical samples. The second column reports the median and, in square brackets, the 95% confidence interval of the realized EPP in the counterfactual samples.<sup>24</sup> The third column reports the probability of observing, in the counterfactual samples, a realized EPP at least as large as the historical one. The last three columns report statistics for the realized equity premia in the counterfactual samples. In particular columns four, five and six report, respectively, the median equity premium (and 95% confidence band underneath), its probability of being negative, and the probability of it being at least as large as the historical value.

\*\*\* Table 4 about here \*\*\*

The first row, first column, of Panel *A* shows that the assumption of a RRA coefficient of 10 implies, in the 1929-2009 sample, an EPP of 5.9% per year. The second column shows instead that the median realized EPP in the counterfactual samples generated by the EL probabilities with  $\gamma = 10$  is 0%, and that the upper bound of its 95% confidence band is only 5.8% – that is, the confidence interval does *not* include the historically observed EPP. Moreover, in the counterfactual samples, a negative realized EPP seems almost as likely as a positive one. This is due to the fact that increasing the probabilities attached to extremely bad states of the economy makes it more likely to observe too many of these events in a finite sample, therefore increasing the likelihood of observing a negative EPP in the counterfactual samples. The third column shows that, the likelihood of observing an EPP at least as large as the historical one would be extremely low – about 2.18%. The fourth column, shows that, in the counterfactual samples, the median equity premium is 2.7% but its distribution is quite wide with the 95% high probability area ranging from  $-2.2\%$  to  $7.5\%$ . This implies that, consistent with the results for the EPP in the second column, negative equity premia are quite likely: their probability, reported in the fifth column, is about 13.6%. Also, the last column shows that the probability of observing an equity premium at least as large as the historical value is extremely small at 2.7% (once again, consistent with the results for the EPP in the third column).

The second row of Table 4, Panel *A* uses the BETEL probabilities instead of the EL ones. The results are largely in line with the ones in the first row: the median EPP in the counterfactual samples is zero and its 95% confidence bands are too tight to include the historical values. Moreover, the historical EPP and equity premium are very unlikely to arise (their probabilities being, respectively, only 1.84% and 2.25%), and both negative EPP and equity premium are very likely to occur in the counterfactual samples.

Panel *B* reports the results for the annual sample over the period 1890-2009. The results are largely similar to those in Panel *A*. The median  $epp_i^T$  across counterfactual

---

<sup>24</sup>Median and confidence bands are computed from the percentiles of  $\{epp_i^T(\gamma)\}_{i=1}^{10,000}$ .

samples is zero in all cases and the 95% confidence intervals do not include the historically observed EPP. Moreover, an EPP and an equity premium, of at least the same magnitudes as the historical ones, are very unlikely to arise in a sample of the same length as the historical one.<sup>25</sup>

Overall, the results presented in this section imply that if the data were generated by the rare events distribution needed to rationalize the EPP, 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 EPP, one should also believe that the puzzle itself is a rare event.*

### 4.3 Rare events and 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

$$E^F [R_{i,t}^e] = \alpha - \underbrace{\frac{Cov^F [(C_t/C_{t-1})^{-\gamma}, R_{i,t}^e]}{E^F [(C_t/C_{t-1})^{-\gamma}]}}_{\equiv \beta_i} \lambda \quad (11)$$

should hold exactly (at the true  $\gamma$ ) with  $\alpha = 0$  and  $\lambda = 1$ . The  $\beta_i$  term in Equation (11) can be interpreted as a measure of the consumption risk that an agent undertakes by 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 Equation (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 Equation (11) to hold if the moments were constructed taking explicitly these rare events into account. That is, using the probability weights  $\hat{P}^j(\gamma)$ ,  $j \in \{EL, BETEL\}$ , needed to rationalize the returns on a cross-section of assets for a given level of the risk aversion coefficient  $\gamma$ , we would expect

$$E^{\hat{P}^j(\gamma)} [R_{i,t}^e] = \alpha - \underbrace{\frac{Cov^{\hat{P}^j(\gamma)} [(C_t/C_{t-1})^{-\gamma}, R_{i,t}^e]}{E^{\hat{P}^j(\gamma)} [(C_t/C_{t-1})^{-\gamma}]}}_{\equiv \beta_i^j} \lambda, \quad (12)$$

---

<sup>25</sup>As a robustness check, in an online 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, also allows us to preserve the autocorrelation properties of consumption growth and returns. This robustness check confirms the results in Table 4.

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

To test this hypothesis we need to choose first a set of assets to construct the probability weights  $\hat{P}^j(\gamma)$ , and then a second set of assets to perform the cross-sectional estimation in Equation (12) under the  $\hat{P}^j(\gamma)$  measure, and the two sets of assets should be different otherwise we would, by construction, obtain a perfect fit in the cross-sectional estimation. To construct the EL and BETEL probability weights ( $\hat{P}^j(\gamma)$ , obtained as described in Section 4.1), we use the excess returns on the 6 Fama-French portfolios (formed from the intersection of 2 size-sorted and 3 book-to-market-equity sorted portfolios). We use the 6 portfolios, instead of simply the market return as in the previous counterfactual exercise, to give the model a better chance to price the cross-section of asset returns (nevertheless, using the market return to construct the probability weights we obtain results qualitatively very similar).<sup>26</sup> As test assets for the cross-sectional estimation, we use instead: (a) the 10 Momentum portfolios, and (b) the 10 Industry portfolios. We also experimented with other sets of assets (e.g. the Fama-French 25 portfolios) obtaining qualitatively similar results.<sup>27</sup>

To obtain empirical estimates of  $\alpha$  and  $\lambda$ , we use the two-step Fama and MacBeth (1973) cross-sectional regression procedure, adapted to take into account that the moments in Equation (12) should be constructed under the  $\hat{P}^j(\gamma)$  probability measures rather than as sample analogs. The weights are extracted under the assumption that  $\gamma = 10$  and we, therefore, estimate Equation (12) setting  $\gamma = 10$  in the pricing kernel. The point estimate of  $\alpha$  measures the extent by which the model fails to price the average equity premium in the cross-section of portfolios. The estimation procedure is described in detail in Appendix A.1.3.

Panels *A*, and *B*, of Table 5 report, respectively, the cross-sectional estimation results for the 10 Momentum and the 10 Industry portfolios. The first column of the table reports the cross-sectional  $R^2$  for all the models considered.<sup>28</sup> The second and third columns report, respectively, the point estimates of  $\alpha$  and  $\lambda$  and their standard errors (in parentheses). The fourth column presents a joint test of the hypothesis  $H_0 : \alpha = 0, \lambda = 1$  and its  $p$ -value (in parentheses).<sup>29</sup> 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, Column 5 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  $Var(\beta_m)/Var(E[R_{m,t+1}^e])$ ) caused by using the  $\hat{P}^j(\gamma)$  probability weights, instead of sample averages, in computing the moments in Equation (12).

\*\*\* Table 5 about here \*\*\*

Consider first Panel *A* that examines the ability of the rare events hypothesis to price the cross-section of 10 Momentum portfolios. The first row reports the results of estimating Equation (11) using the sample moments. The point estimate of  $\alpha$  is not

<sup>26</sup>We use 6 portfolios, rather than the Fama-French 25 portfolios, because of the small available time series at the annual frequency.

<sup>27</sup>These are reported in an online Appendix.

<sup>28</sup>How to construct this statistic under the  $\hat{P}^j(\gamma)$  measures is explained in Appendix A.1.3.

<sup>29</sup>This is a Wald Test with an asymptotic  $\chi^2$  distribution with 2 degrees of freedom under the null.



statistically different from its theoretical value of zero. The estimated slope coefficient  $\lambda$  has the right sign, albeit statistically larger than its model-implied value of unity. Column 1 shows that the model is able to explain 79.2% of the cross-sectional variance of risk premia of the 10 Momentum portfolios.

The second row of Panel *A* focuses on the estimation of Equation (12) using the *EL* probability weights,  $\hat{P}^{EL}(\gamma)$ . The cross-sectional  $R^2$  reduces to 0.1% – two orders of magnitude smaller than 79.2% obtained using the sample weights in Row 1. Neither the intercept nor the slope coefficient is statistically different from zero. Comparing the results in Row 1 to those in Row 2, it is clear that, the C-CAPM performs much worse under the  $\hat{P}^{EL}(\gamma)$  probability measure. What drives this result? In order to increase the ability of the C-CAPM to price the cross-section of returns, the  $\hat{P}^{EL}(\gamma)$  measure should in principle increase the cross-sectional dispersion of consumption risk relative to the cross-sectional dispersion of average risk premia. But the entry in column five, second row shows that the exact opposite happens: moving from sample moments to the  $\hat{P}^{EL}(\gamma)$  weighted moments, the ratio of the cross-sectional variance of consumption risk measures to the cross-sectional variance of average excess returns is reduced by 75.3%. This last finding is a direct consequence of the rare events explanation of the EPP. In order to rationalize the EPP 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 tend to yield low returns and consumption growth tends 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 increases 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 dispersion of consumption risk across assets. This last finding is in line with the large empirical literature that documents an increase in correlation across assets during market downturns (see e.g. Ang and Chen (2002), Longin and Solnik (2001), Erb, Harvey, and Viskanta (1994)).<sup>30</sup>

Panel *A*, Row 3 uses the  $\hat{P}^{BETEL}(\gamma)$  probability weights for the estimation of Equation (12). The results are very similar to those obtained using the  $\hat{P}^{EL}(\gamma)$  weights: neither the intercept nor the slope coefficient is statistically different from zero, the cross-sectional  $R^2$  is three orders of magnitude smaller than that obtained using the sample weights in Row 1, and the ratio of the cross-sectional variance of consumption risk measures to the cross-sectional variance of average excess returns is reduced by 79.3% by moving from sample moments to the  $\hat{P}^{BETEL}(\gamma)$  weighted moments.

Panel *B* reports results for the 10 Industry portfolios. Column 1 shows that moving from the sample moments to the  $\hat{P}^{EL}(\gamma)$  and  $\hat{P}^{BETEL}(\gamma)$  weighted moments does not improve the cross-sectional performance of the model. Column 5 shows that, as with the 10 Momentum portfolios, the use of  $\hat{P}^{EL}(\gamma)$  and  $\hat{P}^{BETEL}(\gamma)$  weighted moments

---

<sup>30</sup>Nevertheless, note that, from a theoretical standpoint, an increase in the correlation across asset returns does not necessarily imply a reduction in the cross-sectional dispersion of risk premia. However, in the U.S. data, during the Great Depression period and the recent financial crisis, the spike in the correlation between the “Growth” and “Value” and the “Small” and “Large” portfolios is associated with a dramatic reduction in the Value and Size premia.

reduces the ratio of the cross-sectional variance of consumption risk measures to the cross-sectional variance of average excess returns by 21.8% and 11.8%, respectively, compared to the sample moments.

The results in Table 5 hold qualitatively for any  $\gamma \in ]0, 10]$ , and also using different sets of assets to construct the probability measures  $\hat{P}^j(\gamma)$  and different testing assets for the cross-sectional tests. Moreover, we obtain very similar results using a linearized version of the pricing kernel. The above results suggest that forcing on the data the rare events rationalization of the EPP worsens 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 EPP 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 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 EPP based on agents’ expectations of an economy wide disaster (e.g. a financial market meltdown) that has *not* materialized in the sample would also reduce the ability of the C-CAPM to price the cross-section of asset returns, since such an expectation would also reduce the cross-sectional dispersion of consumption risk across assets.

## 5 Conclusion

In this paper we study the ability of the rare events hypothesis to rationalize the EPP. 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 RRA is needed in order to rationalize the stock market risk premium. Moreover we show that: *a)* this result holds not only for the U.S., but also for eight other OECD countries; *b)* the disasters present in the data of a large cross-section of countries do not offer support for the rare events explanation of the EPP unless we are willing to believe that disasters should be happening every 6-10 years; *c)* the support for the rare events hypothesis found by the literature that has followed the “standard calibration” approach à la Barro (2006), is due to having calibrated one year contractions during disasters as being equal to the cumulated multi year contractions recorded in the data.

We also 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 were generated by such a distribution, an EPP 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 EPP, one should also believe that the puzzle itself is a rare event.

Last but not least, we show that imposing on the data the rare events explanation

of EPP, substantially worsens the ability of the C-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 explanation, we need to assign higher probabilities 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 C-CAPM to explain the cross-section of returns.

The analytical approach undertaken in this paper can be extended 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 propose can be applied to the calibration of the underlying distribution of *any* economic model that delivers well defined moment conditions.

## References

- ANG, A., AND J. CHEN (2002): “Asymmetric correlations of equity portfolios,” *Journal of Financial Economics*, 63(3), 443–494.
- BACKUS, D., M. CHERNOV, AND I. MARTIN (2009): “Disasters implied by equity index options,” NBER Working Papers 15240, National Bureau of Economic Research, Inc.
- BANSAL, R., D. KIKU, AND A. YARON (2010): “Long-Run Risks, the Macroeconomy, and Asset Prices,” *American Economic Review (Papers and Proceedings)*, 100(2), 542–46.
- BARRO, R. J. (2006): “Rare Disasters and Asset Markets in the Twentieth Century,” *Quarterly Journal of Economics*, 121(3), 823–866.
- BARRO, R. J., AND J. URSUA (2008a): “Consumption Disasters in the 20<sup>th</sup> Century,” manuscript.
- (2008b): “Macroeconomic Crises since 1870,” *Brookings Papers on Economic Activity*, 1, 255–350.
- BARRO, R. J., AND J. F. URSÚA (2009): “Stock-Market Crashes and Depressions,” Working Paper 14760, National Bureau of Economic Research.
- BORWEIN, J. M., AND A. LEWIS (1991): “Duality Relationships for Entropy-type Minimization Problems,” *SIAM Journal of Control and Optimization*, 29, 325–338.
- BRAINARD, W. C., W. R. NELSON, AND M. D. SHAPIRO (1991): “The Consumption Beta Explains Expected Returns at Long Horizons,” Yale University Manuscript.
- 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, D. E., AND R. L. SMITH (1990): “A Correspondance Principle for Relative Entropy Minimization,” *Naval Research Logistics*, 37(2).
- 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.
- COCHRANE, J. H. (1996): “A Cross-Sectional Test of an Investment-Based Asset Pricing Model,” *Journal of Political Economy*, 104(3), 572–621.
- CSISZAR, I. (1975): “I-Divergence Geometry of Probability Distributions and Minimization Problems,” *Annals of Probability*, 3, 146–158.
- DRECHSLER, I., AND A. YARON (2011): “What’s Vol Got to Do with It,” *Review of Financial Studies*, 24(1), 1–45.

- EPSTEIN, L. G., AND S. E. ZIN (1989): “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework,” *Econometrica*, 57, 937–968.
- ERB, C. B., C. R. HARVEY, AND T. E. VISKANTA (1994): “Forecasting International Equity Correlations,” *Financial Analysts Journal*, 50(6), 32–45.
- 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.
- GABAIX, X. (forthcoming): “Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance,” *Quarterly Journal of Economics*.
- 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.
- HANSEN, L. P. (1982): “Large Sample Properties of Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.
- 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.
- KOCHERLAKOTA, N. R. (1996): “The Equity Premium: It’s Still a Puzzle,” *Journal of Economic Literature*, 34(1), 42–71.
- (1997): “Testing The Consumption Capm With Heavy-Tailed Pricing Errors,” *Macroeconomic Dynamics*, 1(03), 551–567.
- 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.
- LEWELLEN, J., S. NAGEL, AND J. SHANKEN (2010): “A skeptical appraisal of asset pricing tests,” *Journal of Financial Economics*, 96(2), 175–194.
- LONGIN, F., AND B. SOLNIK (2001): “Extreme Correlation of International Equity Markets,” *Journal of Finance*, 56(2), 649–676.

- MADDISON, A. (2003): *The World Economy: Historical Statistics*. OECD, Paris.
- MANKIW, N. G., AND M. D. SHAPIRO (1986): “Risk and Return: Consumption Beta Versus Market Beta,” *Review of Economics and Statistics*, 68, 452–59.
- MEHRA, R. (2008): *Handbook of the Equity Premium*. Elsevier, New York.
- 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.
- NAKAMURA, E., J. STEINSSON, R. BARRO, AND J. URSÚA (2010): “Crises and Recoveries in an Empirical Model of Consumption Disasters,” Working Paper 15920, National Bureau of Economic Research.
- NEWBY, 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.
- (1991): “Empirical Likelihood for Linear Models,” *The Annals of Statistics*, 19(4), 1725–1747.
- (2001): *Empirical Likelihood*. Chapman and Hall.
- PARKER, J. A. (2001): “The Consumption Risk of the Stock Market,” *Brookings Papers on Economic Activity*, 2, 279–348.
- PARKER, J. A., AND C. JULLIARD (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.
- RUBINSTEIN, M. (1976): “The Valuation of Uncertain Income Streams and the Pricing of Options,” *Bell Journal of Economics*, 7, 407–25.
- SAIKKONEN, 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.
- SCHENNACH, S. M. (2005): “Bayesian Exponentially Tilted Empirical Likelihood,” *Biometrika*, 92(1), 31–46.

- STUTZER, M. (1996): “A simple Nonparametric Approach to Derivative Security Valuation,” *Journal of Finance*, LI(5), 1633–1652.
- WACHTER, J. (2011): “Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?,” NBER Working Papers 14386, National Bureau of Economic Research, Inc.
- WHITE, H. (1982): “Maximum Likelihood of Misspecified Models,” *Econometrica*, 50, 1–25.

# A Appendix

## A.1 Additional Methodological Details

### A.1.1 Asymptotics

The asymptotic distribution of the EL estimator is given by

$$\sqrt{T} (\widehat{\gamma}^{EL} - \gamma_0) \xrightarrow{d} N(0, V),$$

$$V = (D'S^{-1}D)^{-1}, \quad D = E^F \left[ \frac{\partial \{(C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e\}}{\partial \gamma} \Bigg|_{\gamma=\gamma_0} \right], \quad (13)$$

$$S = E^F \left[ (C_t/C_{t-1})^{-\gamma_0} \mathbf{R}_t^e \mathbf{R}_t^{e'} (C_t/C_{t-1})^{-\gamma_0} \right]. \quad (14)$$

### A.1.2 Dual Solution

The dual solution of the inner minimization problem in Equation (4) is a multinomial distribution with support given by the empirical observations  $\{C_t/C_{t-1}, \mathbf{R}_t^e\}$ ,  $t = 1, \dots, T$  (Csiszar (1975)), and the probability weight assigned to the  $t$ -th observation is

$$\widehat{p}_t^{EL}(\gamma) = \frac{1}{T(1 + \boldsymbol{\lambda}(\gamma)' (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e)}, \quad t = 1, \dots, T, \quad (15)$$

where  $\boldsymbol{\lambda}(\gamma) \in \mathbb{R}^k$  is the solution to the following (dual) unconstrained convex problem

$$\boldsymbol{\lambda}(\gamma) = \arg \min_{\boldsymbol{\lambda}} - \sum_{t=1}^T \log(1 + \boldsymbol{\lambda}' (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e). \quad (16)$$

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

$$\widehat{p}_t^{ET}(\gamma) = \frac{e^{\boldsymbol{\lambda}(\gamma)' (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e}}{\sum_{t=1}^T e^{\boldsymbol{\lambda}(\gamma)' (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e}}, \quad t = 1, \dots, T, \quad (17)$$

where  $\boldsymbol{\lambda}(\gamma) \in \mathbb{R}^k$  is the solution to

$$\boldsymbol{\lambda}(\gamma) = \arg \min_{\boldsymbol{\lambda}} \frac{1}{T} \sum_{t=1}^T e^{\boldsymbol{\lambda}' (C_t/C_{t-1})^{-\gamma} \mathbf{R}_t^e}. \quad (18)$$

### A.1.3 Probability Weighted Fama-MacBeth Regressions

In order to estimate the parameters  $\alpha$  and  $\lambda$  in Equation (12), we follow a Fama and MacBeth (1973) two step procedure, adapted to take into account that the moments



should be constructed under the  $\hat{P}^j(\gamma)$  probability measures instead of as sample analogs.

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

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

where  $j \in \{EL, BETEL\}$ , and  $\gamma$  is fixed to the same value used to construct the weights  $\{\hat{p}_t^j\}_{t=1}^T$ .

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

$$R_{i,t}^e = \alpha_t + \hat{\beta}_i^j \lambda_t + \varepsilon_{i,t}, \quad (19)$$

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

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

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

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

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

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

where  $\eta_i$  is a mean zero error term.

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

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

$$E^{\hat{P}^j(\gamma)} [R_{i,t}^e] := \sum_{t=1}^T R_{i,t}^e \hat{p}_t^j, \quad \hat{R}_i^e := \hat{\alpha} + \hat{\beta}_i^j \hat{\lambda}. \quad (20)$$

Note that the standard Fama and MacBeth (1973) cross-sectional regression ap-

proach (that does not use probability weights) can be recovered by setting  $\hat{p}_t^j = 1/T$ .

## A.2 Cross-Sectional Data Description

For the cross-sectional analysis, we use the returns on the 6 and the 25 Fama and French (1992) portfolios, formed, respectively, from the intersection of two size-sorted and three book-to-market-equity-sorted and five size-sorted and five book-to-market-equity-sorted portfolios, as well as the returns on the 10 Momentum portfolios and 10 Industry portfolios. We use the momentum and industry portfolios, in addition to the size- and book-to-market-equity-sorted portfolios, because, as pointed out by Lewellen, Nagel, and Shanken (2010), the latter portfolios have a strong factor structure making it relatively easy for any proposed factor to produce a high cross-sectional  $R^2$ . Lewellen, Nagel, and Shanken (2010) propose that one simple way to improve asset-pricing tests to make them more convincing is to expand the set of test assets to include other portfolios obtained by sorting stocks on the basis of other characteristics. We construct annual excess returns on these portfolios as their deflated annual returns less the inflation-adjusted rolled over return on 1-month Treasury Bills.

## A.3 Robustness of Estimations

### A.3.1 Assessing the Power of the Estimation Approach

In this section, we use Monte Carlo simulation and resampling procedures to assess the small sample performance, in the presence of rare events, of the information-theoretic estimation approaches described in Section 1. In particular, we compute the small sample  $p$ -values for the *EL* and *BETEL* estimators, and for all the other test statistics reported in Tables 1 and 2, under the assumption that the data are generated by a calibrated rare events model. We do this in two ways.

First, we simulate data from the Barro (2006) model using the author's calibrated values of preference parameters, probability of disasters, as well as the empirical distribution of disaster sizes and duration – that is, we generate data from the Barro model but we explicitly take into account that disasters last multiple years.

Second, as described in Section 4.1, we use our information-theoretic estimation approaches to identify, non-parametrically, the most likely rare events distributions needed to rationalize the EPP with  $\gamma = 10$  (we obtain very similar results setting  $\gamma \in ]0, 10]$ ) and draw hypothetical samples from these distributions.

With the above calibration approaches we generate 10000 samples of exactly the same length as the ones used for the estimations reported in Tables 1 and 2. In each of these simulated samples we perform the same type of estimation and hypothesis testing reported in Section 3.1. The results of this small sample performance evaluation are reported in Table A1. Panel *A* reports results for samples analogous to the 1929-2009 sample, while Panel *B* uses samples of the same length as the 1890-2009 sample.

In each panel, Column 1 reports the probability of obtaining, in the counterfactual samples, *EL* estimates of the RRA as large as, or larger than, the one obtained in the historical sample. That is, we compute  $\Pr(\hat{\gamma}_i \geq \hat{\gamma})$ , where  $\hat{\gamma}_i$  is the point estimate

obtained in the  $i$ -th counterfactual sample and  $\hat{\gamma}$  is the estimate obtained in the historical sample. Column 2 reports the probability that the BEL posterior probability of  $\gamma$  being smaller than, or equal to, 10 is smaller than the value obtained in the historical data. Column 3 reports the probability of obtaining, in the counterfactual samples, an empirical likelihood ratio test statistic at least as large as the one obtained in the historical sample. Finally, Columns 4 and 5 report analogous results to Columns 1 and 2 for the BETEL estimator.

\*\*\* Table A1 about here \*\*\*

The first row in each panel refers to the results obtained when simulating data from the Barro model, while the second and third rows refer to data simulated from the calibrations outlined in Section 4.1.

Focusing on Panel A, we see that: the small sample  $p$ -values for the EL (Column 1) and BETEL (Column 4) estimates in Table 1, Panel A, are about 4%; the likelihood of the small posterior probability of  $\gamma \leq 10$  obtained in Table 1 ranges from .87% to 1.67% for the EL estimator (Column 2) and from 1.25% to 1.68% for the BETEL estimator (Column 5); and the finite sample  $p$ -value for the  $\chi^2$  statistic in Panel A of Table 1 ranges from 1.85% to 10.87% (Column 3).

Panel B computes small sample statistics for the point estimates and tests in Table 2, Panel A. The results are largely in line with the ones in the previous Panel: the finite sample  $p$ -values of the EL and BETEL estimates are all smaller than 5%, while the  $p$ -values of the  $\chi^2$  statistic range from 0 to 10%, and the likelihood of obtaining the small posterior probabilities of a  $\gamma \leq 10$  is smaller than 1.68%.

Comparing the above results with the asymptotic  $p$ -values, and posterior probabilities in Tables 1 and 2, we find that: the asymptotic standard errors tend to overstate the uncertainty about the estimated coefficient (this is largely due to the Gaussian approximation of the likelihoods since, as can be seen in Figure 1, the likelihood is highly asymmetric); the in sample posterior probabilities appear quite precise, and are overall more accurate than the frequentist confidence intervals; the small sample  $p$ -values for the  $\chi^2$  statistic are smaller, or larger, than the asymptotic ones depending on the calibrated data generating process.

Overall, the results presented in this section imply that *if the data were generated by the rare events distribution, the relative entropy based estimation methods presented in Section 1 would be very unlikely to deliver the large estimates of relative risk aversion that are obtained in the historical samples.*

### A.3.2 GMM Estimation

Table A2 reports GMM estimation results for the consumption Euler Equation (1) using US data when consumption is measured as the real per capita personal consumption expenditure on nondurables.

\*\*\* Table A2 about here \*\*\*

## A.4 Estimation with Additional International Data

Table A3 reports BEL and BETEL estimation results for the consumption Euler Equation (1) using data for seven OECD countries. Given the relatively short nature of the samples at hand, we focus on Bayesian estimators in order to avoid relying on asymptotic inference.

\*\*\* Table A3 about here \*\*\*

Despite the short samples, the posterior probabilities of a risk aversion parameter smaller than, or equal to, 10 are consistently very low, and in only one case out of 14 this probability is more than 10%.

**Table 1: Euler Equation Estimation, 1929-2009**

	<i>EL</i>	<i>BETEL</i>
Panel A. Market Return		
$\hat{\gamma}$	28.5 (9.23)	28.5 [11.5, 46.6]
$\chi^2_{(1)}$	4.31 (.038)	
$\Pr(\gamma \leq 10 \text{data})$	1.89%	1.67%
Panel B. Market Return and Risk Free Rate		
$\hat{\gamma}$	28.5 (9.23)	28.6 [14.5, 50.5]
$\hat{\delta}$	0.93 (0.17)	0.95 [0.81, 1.14]
$\chi^2_{(2)}$	7.86 (0.020)	
$\Pr(\gamma \leq 10 \text{data})$	0.54%	0.51%
Panel C. 6 Book-to-Market and Size Portfolios		
$\hat{\gamma}$	34.2 (7.87)	30.3 [14.7, 43.3]
$\chi^2_{(1)}$	29.8 (0.00)	
$\Pr(\gamma \leq 10 \text{data})$	0.53%	0.52%

Note: EL and BETEL estimation results for the Euler Equation (1). The  $\hat{\gamma}$  and  $\hat{\delta}$  rows report the EL point estimates (with s.e. underneath), and the BETEL posterior modes (with 95% confidence regions underneath), of the relative risk aversion coefficient and intertemporal discount rate, respectively. The  $\chi^2$  row of each panel reports the Empirical Likelihood Ratio test (with  $p$ -value underneath) for the joint hypothesis of a  $\gamma$  as small as 10 and for the identifying restriction given by the Euler Equation (1). The last row of each panel reports the BEL and BETEL posterior probabilities of  $\gamma$  being smaller than, or equal to, 10.

**Table 2: Robustness Tests**

	<i>EL</i>	<i>BETEL</i>
<i>Panel A. Total Consumption, 1890-2009</i>		
$\hat{\gamma}$	49.3 (39.4)	49.3 [23.3, 125.7]
$\chi^2_{(1)}$	6.75 (.009)	
$\Pr(\gamma \leq 10 \text{data})$	0.11%	0.10%
<i>Panel B. Parker and Julliard (2005)</i>		
$\hat{\gamma}$	39.3 (57.0)	39.3 [28.3, 58.8]
$\chi^2_{(1)}$	5.52 (0.019)	
$\Pr(\gamma \leq 10 \text{data})$	0.00%	0.00%
<i>Panel C. Five Years Euler Equation</i>		
$\hat{\gamma}$	32.9 (7.96)	32.9 [24.6, 63.4]
$\chi^2_{(1)}$	17.1 (.000)	
$\Pr(\gamma \leq 10 \text{data})$	0.0%	0.0%

Note: EL and BETEL estimation results for Equation (1). The first row of each panel reports the EL point estimates (with s.e. underneath), and the BETEL posterior modes (with 95% confidence regions underneath), of the relative risk aversion coefficient  $\gamma$ . The second row of each panel reports the Empirical Likelihood Ratio test (with  $p$ -value underneath) for the joint hypothesis of a  $\gamma$  as small as 10 and for the identifying restriction given by the Euler Equation (1). The third row of each panel reports the posterior probabilities of  $\gamma$  being smaller than, or equal to, 10.

**Table 3: U.K. Euler Equation Estimation**

	<i>EL</i>	<i>BETEL</i>
<i>Annual Data: 1919-1994</i>		
$\hat{\gamma}$	65.3 (24.4)	65.3 [35.8, 131.3]
$\chi^2_{(1)}$	8.76 (.003)	
$\Pr(\gamma \leq 10 \text{data})$	.12%	.05%

Note: EL and BETEL estimation results for the consumption Euler Equation (1) using U.K. data. The first row reports the EL point estimate (with s.e. underneath), and the BETEL posterior mode (with 95% confidence region underneath), of the relative risk aversion coefficient  $\gamma$ . The second row reports the Empirical Likelihood Ratio test statistic (with  $p$ -value underneath) for the joint hypothesis of a  $\gamma$  as small as 10 and for the identifying restriction given by the consumption Euler Equation (1). The third row reports the posterior probabilities of  $\gamma$  being smaller than, or equal to, 10.

**Table 4: Counterfactual Equity Premium Puzzle**

	$epp^T$	$epp_i^T$	$\Pr(epp_i^T \geq epp^T)$	$\overline{R_{i,t}^e}^T$	$\Pr(\overline{R_{i,t}^e}^T < 0)$	$\Pr(\overline{R_{i,t}^e}^T \geq \overline{R_t^e}^T)$
<i>Panel A. Nondurable Consumption, 1929-2009</i>						
$\hat{P}^{EL}(\gamma)$	5.9%	0.0% [-5.7%, 5.8%]	2.18%	2.70% [-2.2%, 7.5%]	13.59%	2.70%
$\hat{P}^{BETEL}(\gamma)$	5.9%	0.0% [-5.5%, 5.6%]	1.84%	2.53% [-2.3%, 7.3%]	14.66%	2.25%
<i>Panel B. Total Consumption, 1890-2009</i>						
$\hat{P}^{EL}(\gamma)$	5.8%	0.0% [-4.5%, 4.5%]	0.59%	1.95% [-1.7%, 5.6%]	14.80%	1.06%
$\hat{P}^{BETEL}(\gamma)$	5.8%	0.0% [-4.3%, 4.3%]	0.43%	1.76% [-1.8%, 5.3%]	15.96%	0.48%

Note: the first column reports the realized equity premium puzzle (defined in equation (10)) in the historical sample corresponding to  $\gamma = 10$ ; the second column reports the median realized equity premium puzzle (and its 95% confidence band underneath) in the counterfactual samples for  $\gamma = 10$  and probability distribution  $\hat{P}^j(\gamma)$ ,  $j \in \{EL, BETEL\}$ , used to generate the data; the third column reports the probability of observing a realized equity premium puzzle at least as large as the historical one; the fourth column reports the median (and its 95% confidence band underneath) equity premium in the counterfactual samples while the fifth and sixth columns report the probability that the counterfactual equity premium is, respectively, smaller than zero, and as large (or larger) than the historical value.



**Table 5: Counterfactual Cross-Sectional Regressions**

Moments:	$R^2$	$\hat{\alpha}$	$\hat{\lambda}$	Wald test: $\alpha = 0, \lambda = 1$	$\Delta \frac{Var(\beta_m)}{Var[E(R_m^e)]}$
<i>Panel A. 10 Momentum Portfolios</i>					
Sample	79.2%	0.00 (0.033)	5.31 (1.070)	29.6 (0.00)	
$\hat{P}^{EL}(\gamma)$	0.1%	-0.01 (0.032)	0.45 (0.635)	1.25 (0.530)	-75.3%
$\hat{P}^{BETEL}(\gamma)$	0.0%	0.01 (0.031)	-0.22 (0.662)	3.84 (0.15)	-79.3%
<i>Panel B. 10 Industry Portfolios</i>					
Sample	0.4%	0.08 (0.024)	0.21 (1.387)	12.7 (0.00)	
$\hat{P}^{EL}(\gamma)$	0.6%	0.02 (0.027)	0.29 (0.946)	0.63 (0.731)	-21.8%
$\hat{P}^{BETEL}(\gamma)$	0.1%	0.02 (0.027)	0.12 (0.921)	1.02 (0.602)	-11.8%

Note: Fama and MacBeth (1973) cross-sectional regression results for 10 Momentum (Panel A) and 10 Industry Portfolios (Panel B). Row 1 of each panel reports results for the sample moments while Rows 2 and 3 report the same for the EL and BETEL probability-weighted moments, respectively. The EL and BETEL probability weights are obtained using the 6 Fama-French portfolios and  $\gamma = 10$ .

**Table A1: Finite Sample Performance**

	<i>EL</i>	<i>BEL posterior</i>	$\chi^2_{(1)}$	<i>BETEL</i>	<i>BETEL posterior</i>
<i>Panel A. Annual Data: 1929-2009</i>					
Barro (2006)	4.33%	1.67%	10.87%	4.21%	1.68%
$\hat{P}^{EL}$	3.95%	0.91%	1.85%	3.95%	1.25%
$\hat{P}^{BETEL}$	4.07%	0.87%	1.92%	4.07%	1.29%
<i>Panel B. Annual Data: 1890-2009</i>					
Barro (2006)	4.85%	1.50%	10.08%	4.61%	1.68%
$\hat{P}^{EL}$	4.93%	0.07%	0.48%	4.90%	0.18%
$\hat{P}^{BETEL}$	4.47%	0.05%	0.00%	4.45%	0.10%

Note: the table reports the finite sample  $p$ -values of the EL (Column 1) and BETEL (Column 4) estimates, and the  $\chi^2$  test (Column 3), as well as the finite sample likelihood of the small posterior probabilities (Columns 2 and 5), reported in Panels A of Tables 1 and 2.

**Table A2: GMM Estimation, 1929-2009**

<i>Panel A. Market Return</i>	
$\hat{\gamma}$	28.5 (9.23)
Mean  Pricing Err.	0.11%
<i>Panel B. Market Return and Risk Free Rate</i>	
$\hat{\gamma}$	28.6 (9.22)
$\hat{\delta}$	0.93 (0.17)
Mean  Pricing Err.	7.30%
<i>Panel C. 6 FF Portfolios</i>	
$\hat{\gamma}$	32.4 (7.84)
$\chi^2_{(5)}$	15.3 (0.009)
Mean  Pricing Err.	2.27%

Note: GMM estimation results for the consumption Euler Equation (1). The  $\hat{\gamma}$  and  $\hat{\delta}$  rows report, respectively, the point estimates (with s.e. underneath) of the RRA coefficient and the intertemporal discount factor. The  $\chi^2$  row of Panel C reports the overidentifying restrictions test (with p-value underneath). The last row of each panel reports the mean absolute pricing error.

**Table A3: Euler Equation Estimation with International Data**

	<i>BEL</i>		<i>BETEL</i>	
	$\hat{\gamma}$	$\Pr(\gamma \leq 10 \text{data})$	$\hat{\gamma}$	$\Pr(\gamma \leq 10 \text{data})$
<i>Australia</i> <i>1970:Q1-1998:Q2</i>	53.6 [8.66, 164.2]	5.8%	53.8 [8.52, 158.3]	5.9%
<i>Canada</i> <i>1970:Q1-1998:Q2</i>	87.7 [33.3, 466.2]	1.2%	87.7 [29.1, 374.5]	1.5%
<i>France</i> <i>1970:Q1-1998:Q2</i>	45.6 [11.1, 117.9]	4.5%	49.1 [8.21, 89.9]	6.1%
<i>Germany</i> <i>1972:Q2-1998:Q2</i>	168.8 [47.5, 367.9]	1.3%	183.7 [29.1, 283.9]	2.0%
<i>Netherland</i> <i>1977:Q2-1998:Q3</i>	309.6 [129.2, 481.0]	0.05%	309.4 [128.0, 432.3]	0.04%
<i>Sweden</i> <i>1970:Q1-1999:Q3</i>	112.7 [47.3, 315.2]	0.43%	107.1 [34.7, 144.3]	0.86%
<i>Japan</i> <i>1970:Q1-1998:Q3</i>	19.5 [5.12, 196.2]	10.0%	19.1 [4.16, 148.9]	12.4%

Note: BEL and BETEL estimation results for the consumption Euler Equation (1). The first and third columns report the posterior modes (with 95% confidence regions underneath), of the relative risk aversion coefficient  $\gamma$ . The second and fourth columns report the probabilities of  $\gamma$  being smaller than, or equal to, 10.

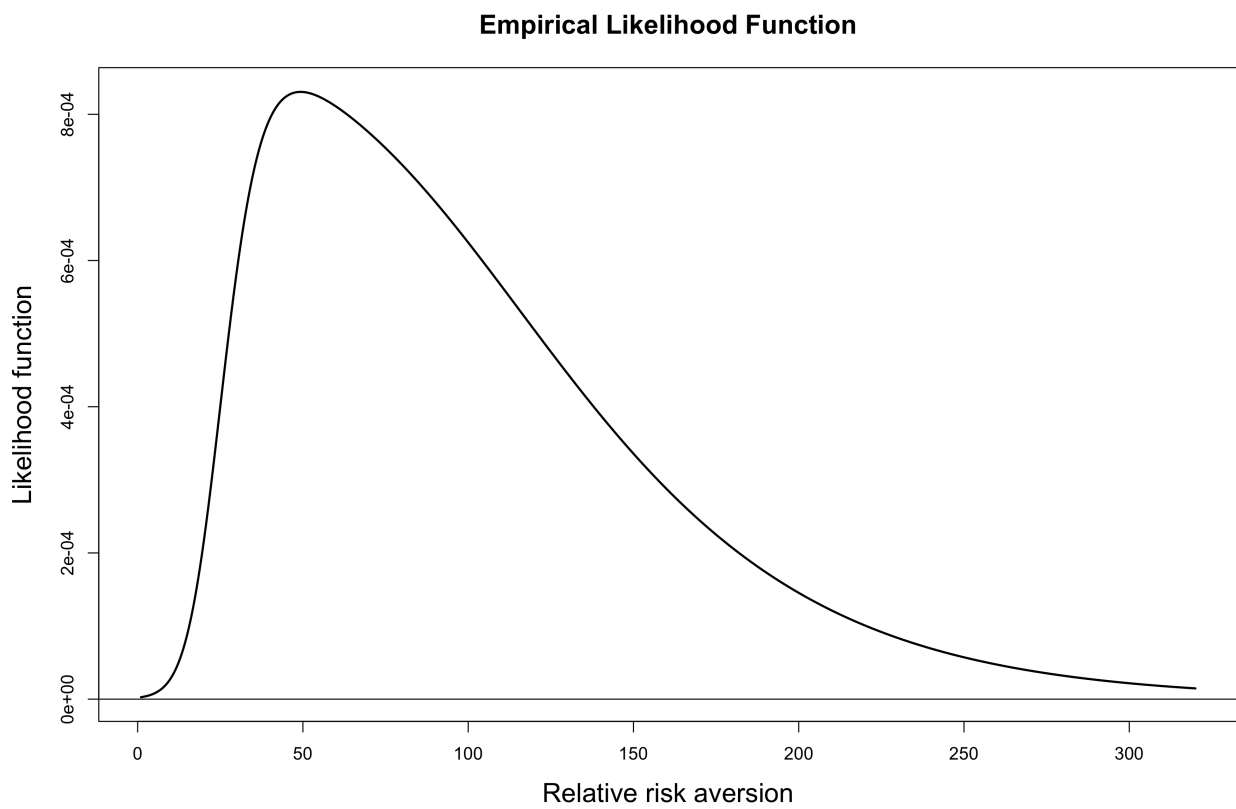


Figure 1: Empirical Likelihood function.

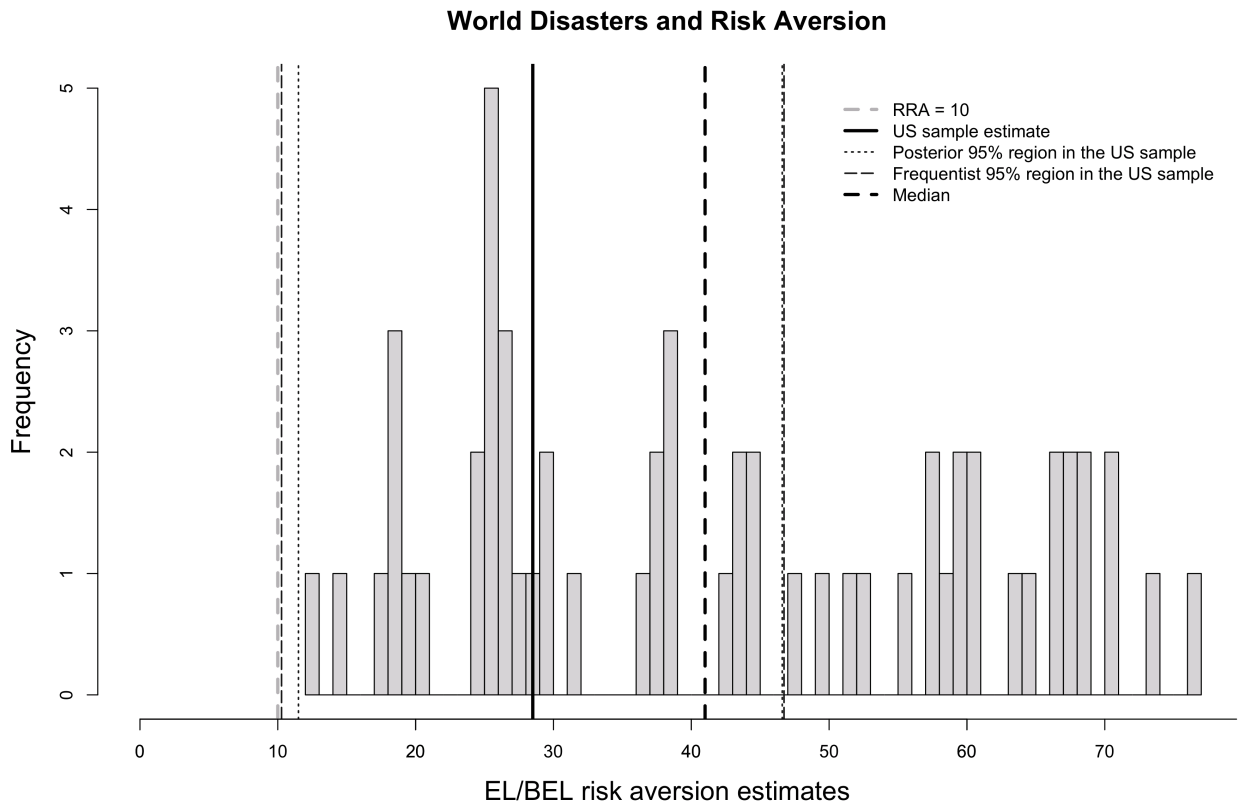


Figure 2: Distribution of the EL estimates of the RRA coefficient with disaster observations drawn from the empirical distribution of world disasters.

### Risk Aversion and World Disasters Probabilities

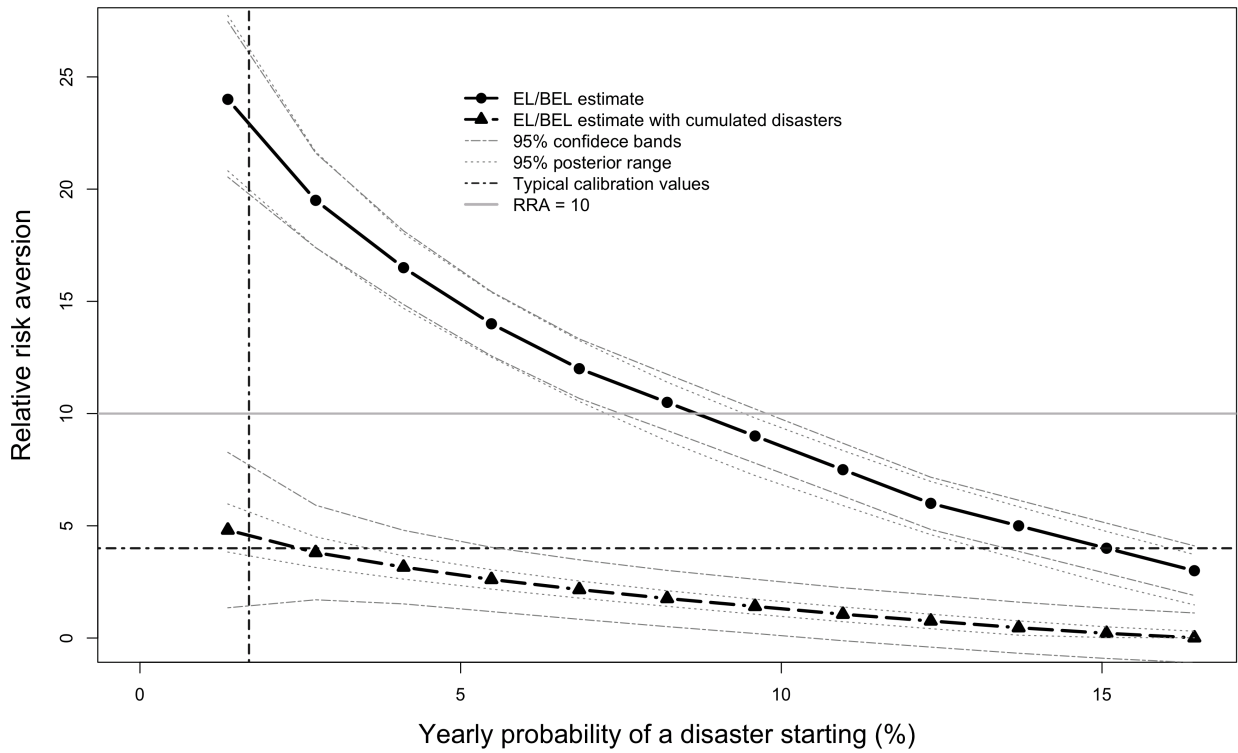


Figure 3: EL estimates of the RRA coefficient, with disasters drawn from their empirical distribution, with (black solid line with circles) and without (blue dashed line with triangles) taking into account that disasters might last more than one year, as a function of the sample probability of disasters.

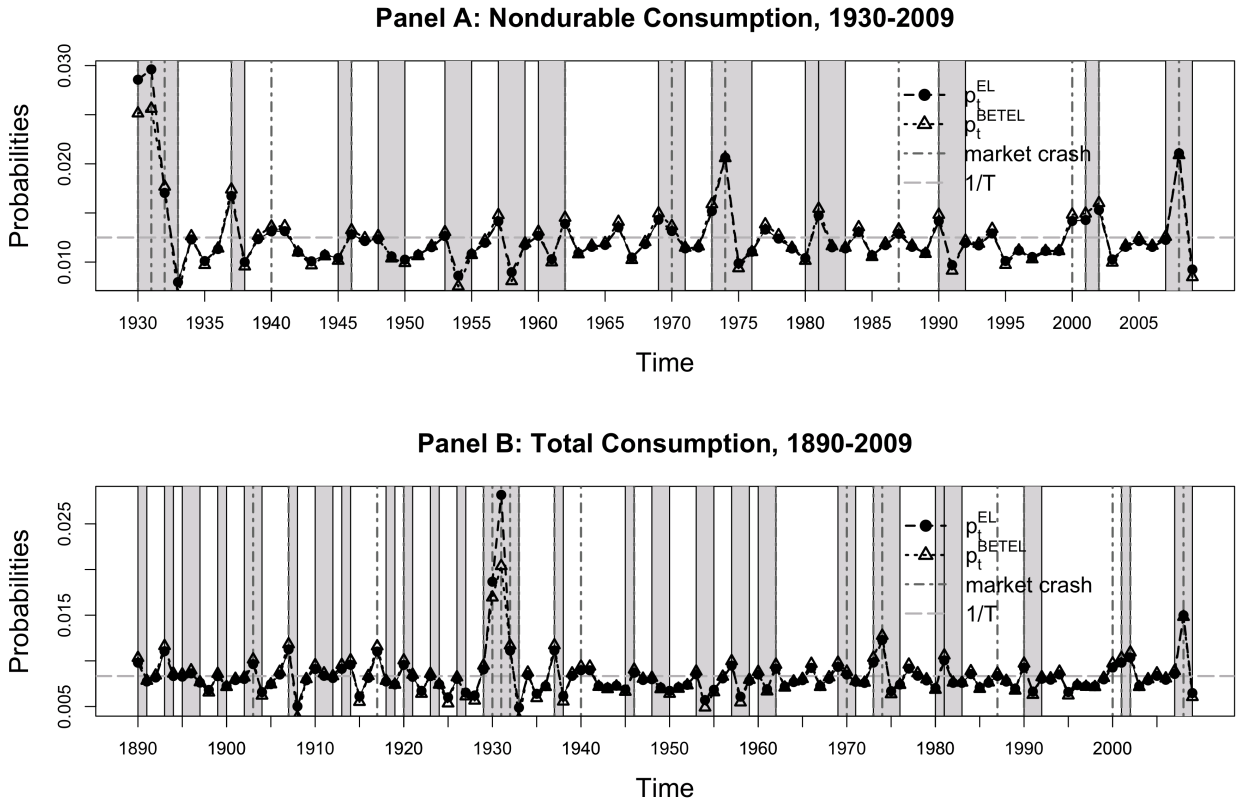


Figure 4: EL and BETEL estimated probabilities needed to solve the equity premium puzzle with  $\gamma = 10$ . Shaded areas are NBER recession periods. Vertical dashed lines are the stock market crashes identified by Mishkin and White (2002). The horizontal line in each panel indicates the sampling frequency ( $1/T$ ).



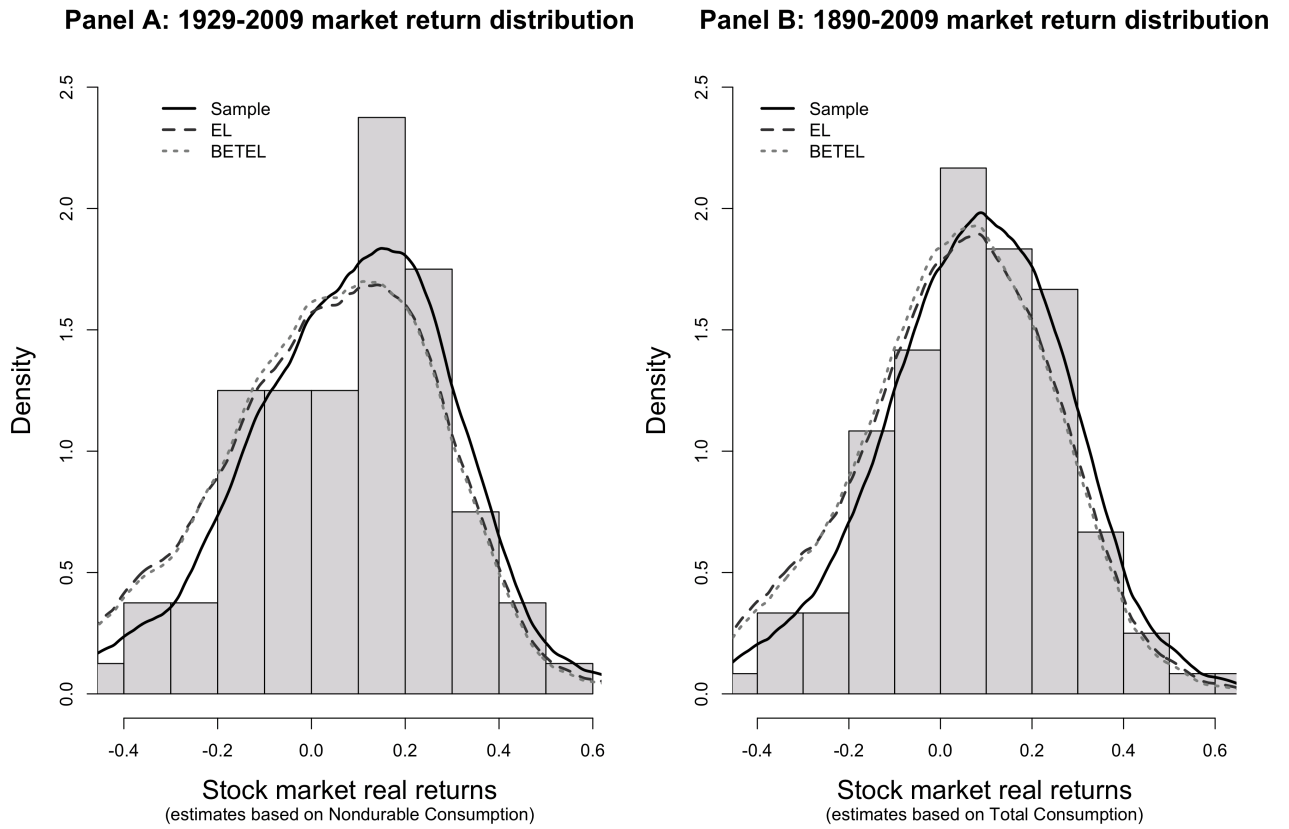


Figure 5: Sample, EL, and BETEL market returns distributions (computed setting  $\gamma = 10$ ).

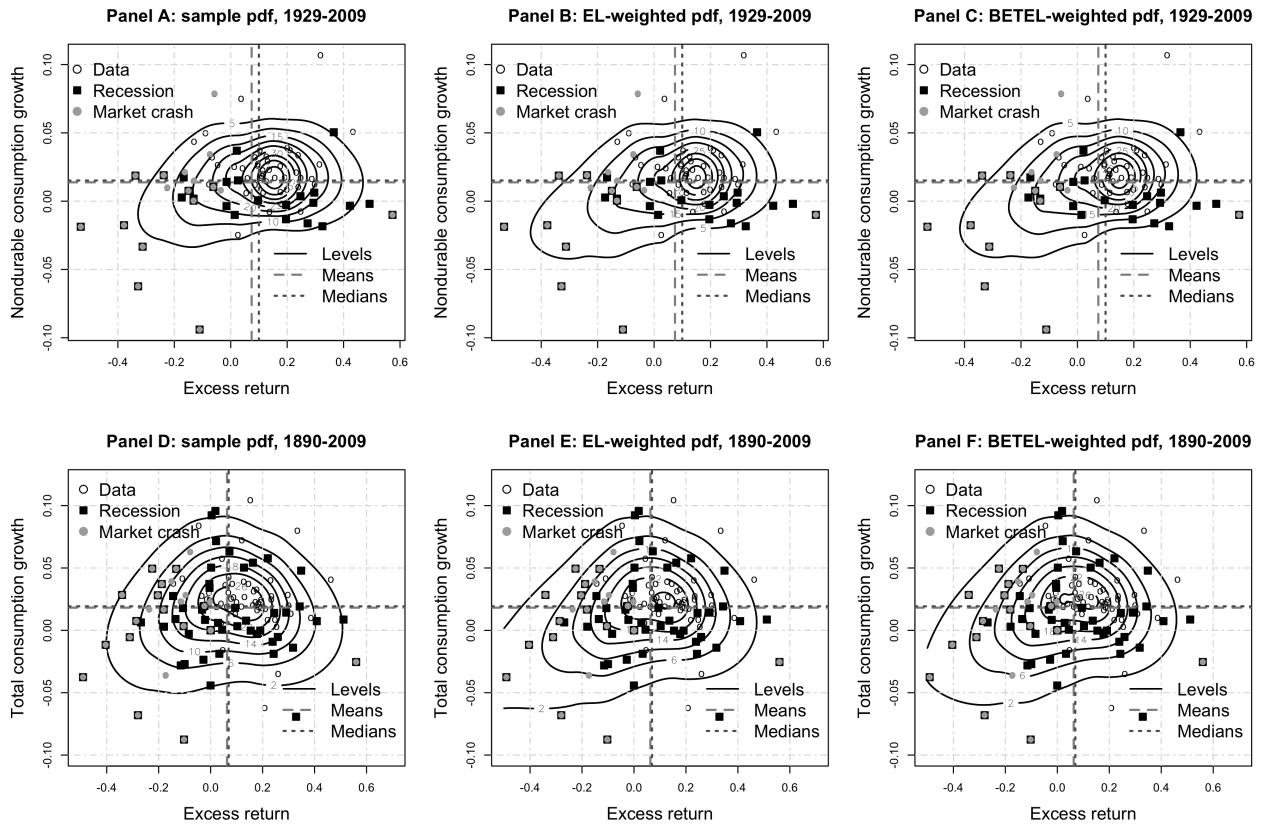


Figure 6: Level curves of the joint distribution of consumption growth and stock market excess returns.