# Internet Appendix for
# Can Rare Events Explain the Equity Premium Puzzle?

Christian Julliard*
*London School of Economics*

Anisha Ghosh†
*Carnegie Mellon University*

March 6, 2012

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

†Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, U.S.A., anishagh@andrew.cmu.edu.

**Abstract**

The Appendix contains additional methodological details that are not included in the paper for reasons of brevity, as well as tables and figures with additional robustness checks.

# 1 Blockwise Estimation

The optimizing behavior of the representative agent, in the time-additive power utility model, entails that $\left\{(C_t/C_{t-1})^{-\gamma}\,\mathbf{R}_t^e\right\}_{t=1}^{\infty}$ is a martingale difference sequence, i.e. it is not autocorrelated. Therefore, the distribution theory for the EL estimator outlined in the paper remains valid even if the stochastic process generating $\{C_t/C_{t-1};R_t^e\}_{t=1}^{\infty}$ is weakly dependent. However, serially correlated measurement error in consumption (see Wilcox (1992)) or model misspecification (see e.g. Parker and Julliard (2005)) could make the martingale property of the conditional Euler equation fail in the data, therefore jeopardizing the asymptotic justification of the standard errors and test statistics. To address this issue, we also consider the *blockwise* EL (Kitamura (1997)) estimator. The idea is to use blocks of consecutive observations to retrieve, non-parametrically, information about dependence in the data (this is closely related to the blockwise bootstrap, see e.g. Hall (1985)). More precisely, the observation $(C_t/C_{t-1})^{-\gamma}\,\mathbf{R}_t^e$ is replaced by

$$\widehat{f}(t,\gamma)=\frac{1}{2M+1}\sum_{s=-M}^{M}(C_{t+s}/C_{t-1+s})^{-\gamma}\,\mathbf{R}_{t+s}^e$$

where $M^2/T\to 0$, and $M\to\infty$ as $T\to\infty$. Then, the EL estimation is performed with $(C_t/C_{t-1})^{-\gamma}\,\mathbf{R}_t^e$ replaced by $\widehat{f}(t,\gamma)$, $t=M+1,M+2,...,T-M$. Kitamura (1997) shows that the asymptotic distribution of this estimator is

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

where $V$ and $D$ are defined as in the standard case but

$$S=\sum_{j=-\infty}^{\infty}E^F\left[(C_t/C_{t-1})^{-\gamma_0}\,\mathbf{R}_t^e\mathbf{R}_t^{e'}\,(C_t/C_{t-1})^{-\gamma_0}\right],$$

and $S$ can be estimated using a Newey and West (1987) HAC approach.

# 2 Time Aggregation and Multi-year Euler Equation Estimation

We repeated our estimation and testing results at multi-year frequencies taking time aggregation effects into account. In particular, we performed the estimation and testing of the consumption Euler equation at multiperiod frequencies when the first multiperiod observation is replaced by the minimum excess return on the market portfolio and the minimum consumption growth observed during the 4 years corresponding to the Great Depression. The results are presented in Table $B1$ below. Panels $A$ and $B$ report results for the 3-year and 5-year frequencies, respectively. The results are very similar to those presented in Table 2, Panel $C$ of the paper (that provides results at the five year frequency without taking time aggregation effects into account). This suggests that our results are robust to time aggregation effects.

**Table B1: Time-Aggregation Effects**

| | *EL* | *BETEL* |
|---|---|---|
| *Panel A: Three Years Euler Equation* | | |
| $\hat{\gamma}$ | 26.5 (7.31) | 26.5 [19.2, 42.1] |
| $\chi^2_{(1)}$ | 8.65 (0.003) | |
| $\Pr(\gamma \leq 10 \mid \text{data})$ | 0.00% | 0.00% |
| *Panel B: Five Years Euler Equation* | | |
| $\hat{\gamma}$ | 39.5 (9.74) | 39.5 [29.4, 76.0] |
| $\chi^2_{(2)}$ | 20.1 (.000) | |
| $\Pr(\gamma \leq 10 \mid \text{data})$ | 0.0% | 0.0% |

Note: EL and BETEL estimation results for the Euler Equation at multi-year frequencies taking time-aggregation effects into account, i.e. replacing the first multiperiod observation by the minimum excess return on the market portfolio and the minimum consumption growth observed during the 4 years corresponding to the Great Depression. The $\hat{\gamma}$ row reports the EL point estimate (with s.e. underneath), and the BETEL posterior mode (with 95% confidence regions underneath), of the relative risk aversion coefficient. 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. The last row of each panel reports the BEL and BETEL posterior probabilities of $\gamma$ being smaller than, or equal to, 10.

# 3 The World's Largest 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 Ursua (2008b), we have 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. 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.

In particular, we ask *a*) how large should have been the consumption drop during the U.S. Great Depression in order to obtain small estimates of the coefficient of RRA and not reject the model, and *b*) whether a *consumption* disaster of this size, or larger, has ever occurred during the last two centuries in the 40 countries studied by Barro and Ursua (2008b): a sample of 95 identified consumption disasters.

To answer these questions, we modify our baseline annual data sample (1929-2009) by replacing the four data points corresponding to the Great Depression period with calibrated disaster observations. Note that we choose to use the sample that delivers the weakest rejection of the rare events hypothesis (see Tables 1 and 2).[1] To make our calibration comparable with the Barro and Ursua (2008b) consumption disasters data, we calibrate disasters of both different *sizes* and *lengths* – e.g. for a *two* year disaster with a cumulated 40% drop in consumption, we add to our sample *two* data points characterized by a cumulated consumption drop of 40%. Since in Barro and Ursua (2008b) we have no estimates of the stock market drop during the disaster, the market return during the disaster is calibrated to match the annualized excess return during the U.S. Great Depression.[2]

In the modified baseline samples we perform estimation of the risk aversion coefficient using, as before, a set of estimators that allow the probabilities attached to different states of the economy to differ from their sample frequencies, and for which one disaster is in principle

[1] Very similar results are obtained using the 1890-2009 total consumption sample, and are available from the authors upon request.

[2] In Section 3.2.2 of the paper, we relax this assumption by using an alternative data set (from Barro and Ursúa (2009)) that contains fewer countries but has the advantage of providing us with stock market returns during the disasters.

enough to approximate the tail behavior of the consumption distribution. Moreover, note that the sample frequency of disaster periods in the calibrated samples is already quite high: ranging from 1.3% for disasters that last only one year, to 9.5% for disasters that last 8 years.

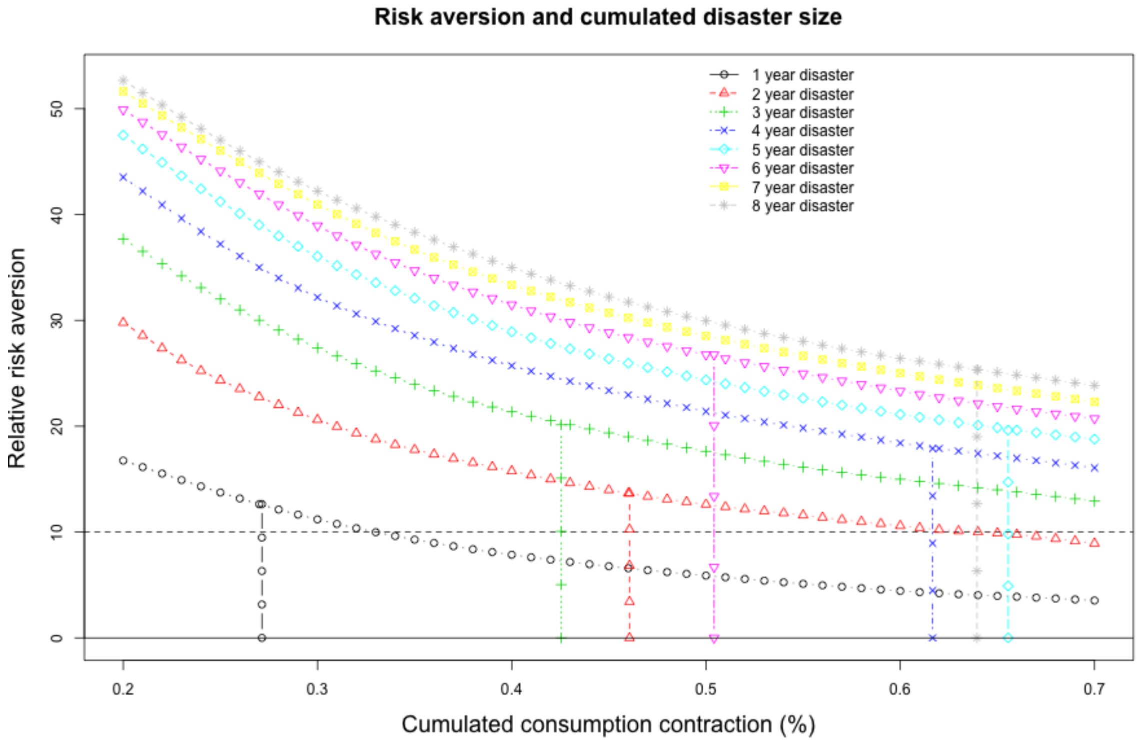


Figure 1: The curves report relative risk aversion estimates using the EL estimator as a function of different cumulated consumption contraction sizes and disaster lengths. The vertical lines report the maximum consumption contraction ever recorded in the data for an economic disaster of a given length. The horizontal black dashed line corresponds to a relative risk aversion of 10.

The results of this exercise are reported in Figure 1 for the $EL$ estimator. Each of the different curves correspond to a different disaster length (from 1 to 8 years), and each point on a curve gives the RRA coefficient estimate (measured on the vertical axis) corresponding to a disaster with a given cumulated consumption contraction (measured on the horizontal axis). Each of the vertical segments correspond to the *largest* consumption disaster of a given length *recorded* in the historical data. For instance, the figure shows that, to have a RRA estimate below 10, the U.S. should have experienced a one year consumption contraction of at least 35% (given by the intersection between the dashed horizontal line and the curve with black circles), while instead the largest one year consumption disaster ever recorded in the data corresponds to a contraction of only 27% (represented by the black vertical line with circles). That is, the one year contraction needed to rationalize the EPP is more than one-fourth larger than the largest one year consumption disaster ever recorded in the data. Similarly, focussing on two year disasters (denoted by the curve with red triangles) we see that we would have needed a two year consumption contraction of at least 68% to explain the equity premium with low risk aversion, while the largest two year consumption disaster ever recorded is only 43%. That is, to rationalize the EPP the U.S. should have experienced a two year consumption disaster that is 58% larger than the largest two year consumption disaster in the data. Focusing on the other disaster lengths we have also similar findings: *the consumption disasters that we should have observed in the U.S. to rationalize the EPP are much larger than the largest consumption disasters ever recorded in a sample (containing a total of 95 disaster events) of 4162 annual consumption observations for 40 different countries during the last two centuries.*

We repeat the exercise using the $BETEL$ estimation approach. Results are presented in Figure 2, and they are in line with the ones obtained using the $EL$ estimation approach: the magnitude of the consumption disasters that we should have observed in the U.S. during the Great Depression in order to rationalize the EPP with a low level of risk aversion, are larger than the largest consumption disasters ever recorded in a sample of more than 4162 annual consumption observations for 40 different countries.

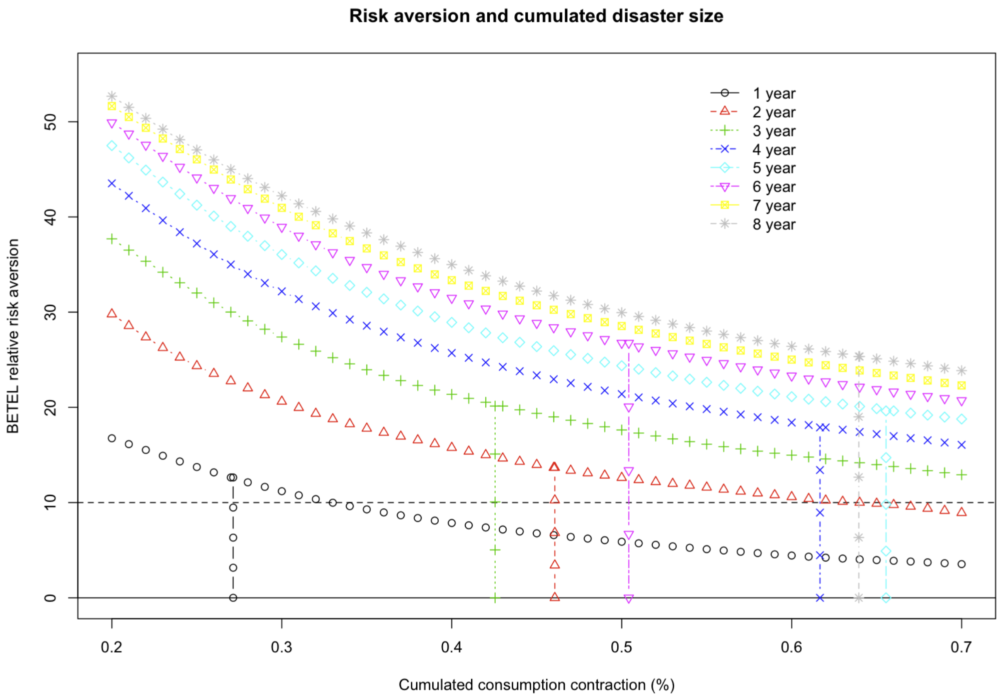


Figure 2: The curves report relative risk aversion estimates using the BETEL estimator, as a function of different cumulated consumption contraction size and disaster length. The vertical lines report the maximum consumption contraction ever recorded in the data for an economic disaster of a given length. The horizontal black dashed line correspond to a risk aversion of 10.

# 4 The World's Largest Stock Market *and* Consumption Disasters

In this section we repeat the two exercises presented in Section 3.2.2 of the paper using the $BETEL$ estimation approach.

In Figure 3 we report the estimates of the relative risk aversion coefficient obtained by replacing the U.S. Great Depression observations with the world disasters identified in Barro and Ursúa (2009). The results obtained using the $BETEL$ estimator are very similar to the ones reported in the paper for the $EL$ approach: in *none* of the 58 samples is the estimated risk aversion coefficient smaller than 10; the median estimate across samples, being about 41, is even higher than the estimate in the true U.S. data sample; the centered

95% area of the distribution of RRA estimates ranges from 16 to 72; and none of the point estimates lies below the 95% posterior probability area obtained in the true U.S. sample.

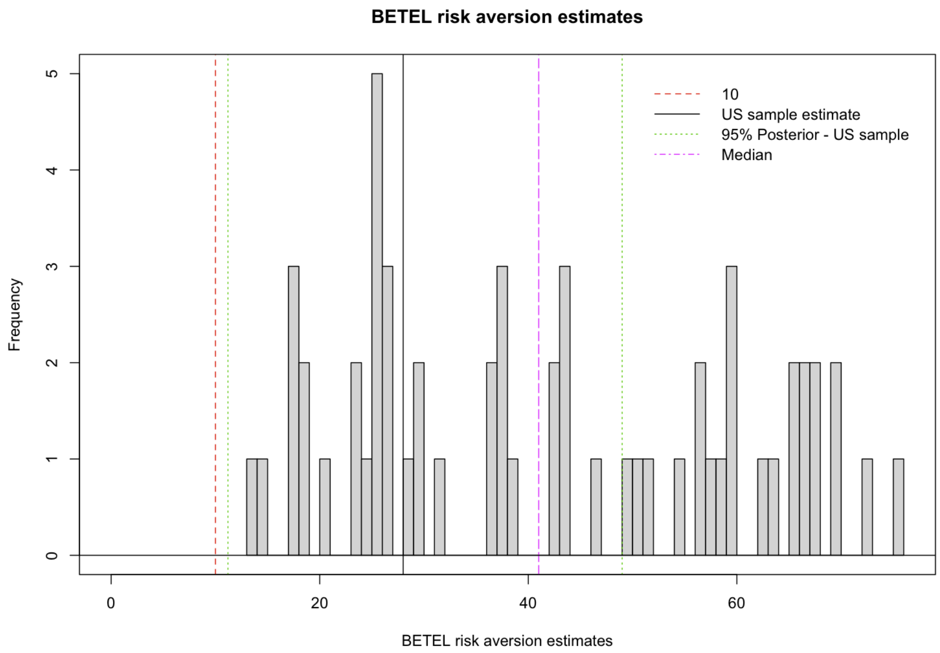


Figure 3: distribution of BETEL estimates of the RRA coefficient with disaster observations drawn from the empirical distribution of world disasters.

Figure 4 repeats the second exercise presented in Section 3.2.2 of the paper using the BETEL estimator. The figure shows that taking into account the multi year nature of the world economic disasters, a sample frequency of disasters of about 9.2% (a disaster every 10.9 years) is needed to explain the the equity premium with a risk aversion of 10, and a probability of disaster of about 14.5% (a disaster every 6.9 years) would be needed in order to explain the EPP (as the literature based on the "Standard Calibration" approach of disaster models pioneered by Barro (2006) does) with a risk aversion as small as 4. Once again, looking at the blue dashed line – based on having erroneously assumed that all the disasters had their full impact on the economy in only one year – we find a RRA estimate of about 4 with a sample frequency of disasters as small as 1.7%.

# 5 The Standard Calibration Approach

The results presented in the paper and in the previous sections of this Internet Appendix suggest that the rare events hypothesis is an unlikely explanation of the EPP. This conclusion is in contrast with the one of the literature that has followed the now "Standard Calibration" (SC) approach of disaster models pioneered by Barro (2006) (e.g. Gabaix (2007), Wachter (2011), Barro and Ursua (2008a) and many others) and that finds strong support for the rare events explanation of the EPP. In this section we investigate the reasons behind this discrepancy, and we show that the difference in results is driven by the fact that the SC approach calibrates *one-year consumption* drops using *multi-year GDP* contractions data, and that this mechanically overstates the degree of consumption risk of the stock market.

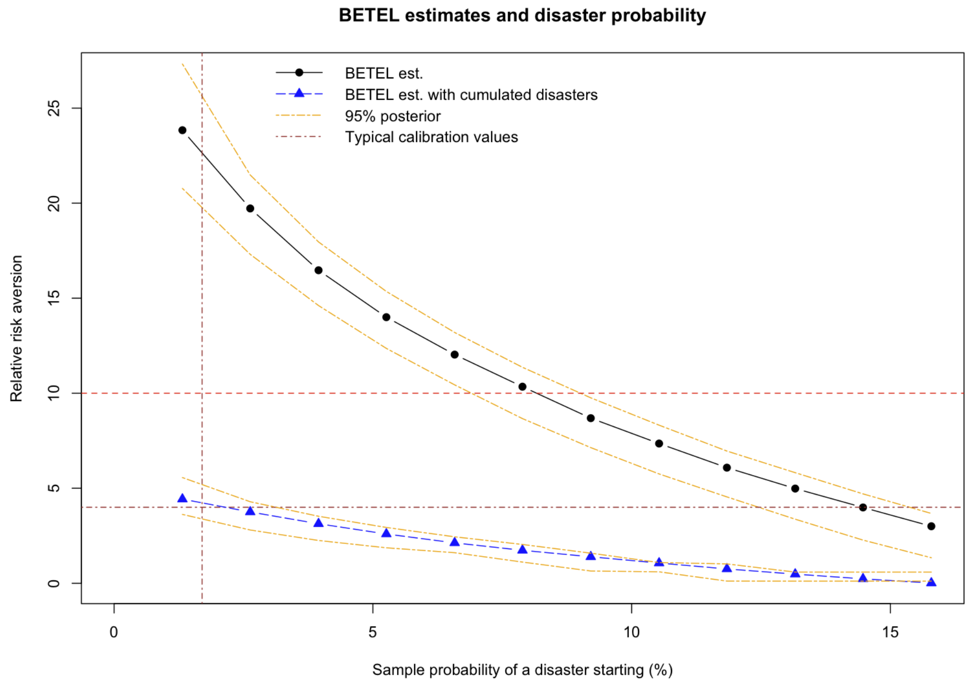


Figure 4: BETEL estimates of the RRA coefficient, with disasters drawn from their empirical world distribution, with (black solid line with circles) and without (blue dashed line with triangles) taking into account that disasters might last more than one year, as a function of the sample probability of disasters.

The key element in the SC approach is the calibration of the consumption contraction during a rare economic disaster. In order to identify a reasonable value for this quantity, Barro (2006) performs an extensive study of the major economic disasters of the twentieth century in a cross section of 35 countries (the data are taken from Maddison (2003)). The criterion used to identify an economic disaster is a cumulated multi-year drop in GDP per capita of more than 15%. Based on these identified disasters, the SC approach calibrates the distribution of the *one-year consumption drop* during a disaster as being equal to the empirical distribution of *multi-year GDP drop* in the cross-country sample of disasters.

This calibration choice raises two concerns. First, if the identified disasters tend to last for more than one year, this approach overstates the degree of consumption risk, since a risk averse agent fears one large contraction in consumption more than having the same shock spread over several years. Second, if agents are able to at least partially smooth income shocks over time, we would expect consumption to drop by less than GDP during disasters.

Indeed, both concerns seem to be supported by the data. Figure 5 reports the size and length, and length distribution histogram, of the 64 disasters identified in Barro (2006) using Maddison (2003) data and employed in the SC approach.[3] All the disasters are multi-year contractions: the average length of the disaster is about 4 years while the mode of the distribution is 3 years. Also, there is a positive, and statistically significant, relation between disaster length and size as indicated by the regression of disaster size on length reported in

[3]Note that we have dropped the Austrian GDP data point for the $1944-1945$ period since: *a*) Good and Ma (2006) question the accuracy of the Austrian GDP data for the first half of the Twentieth Century reported in Maddison (2003), and argue that they are constructed using "*crude 'back-of-the-envelope' calculations*;" *b*) the sources for Austrian data cited by Maddison (2003) *do not* report a GDP value for the $1944-1945$ period. Probably for the same reasons, Barro has also dropped this data point in his subsequent works.

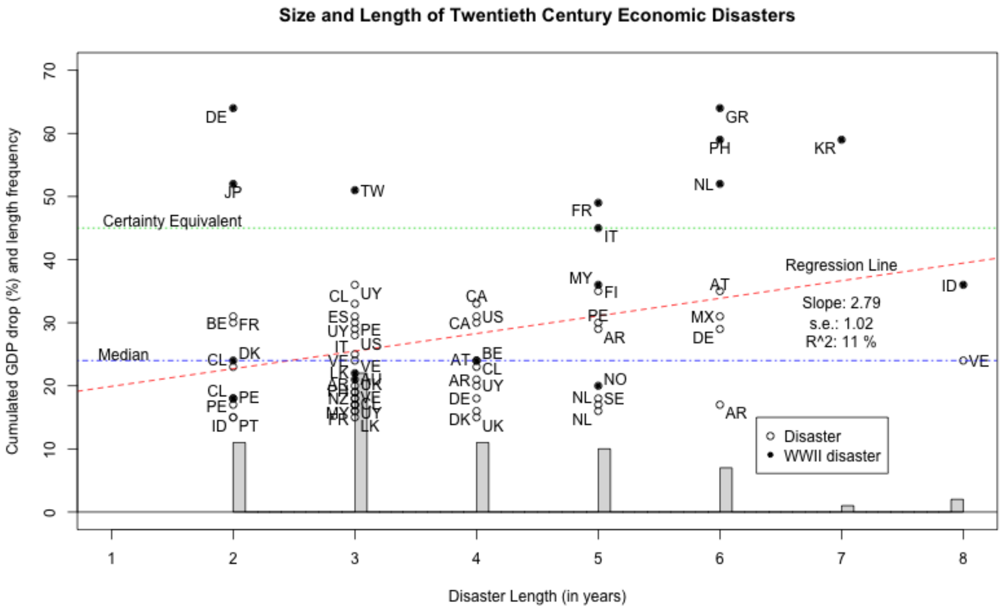


Figure 5: Cumulated size, length, and length histogram of GDP contractions during the 64 major economic disasters of the Twentieth Century. Reported in the figure are also: the median disaster size (blue dash dotted line); the certainty equivalent of the disasters distribution in terms of one-year consumption drop (green dotted line); the linear regression of disaster size on disaster length (red dashed line) as well as the regression measure of fit, slope coefficient estimate, and its standard error. Source: Barro (2006), Table I, and authors' calculations.

the figure (red dashed line). The figure also reports the certainty equivalent of the disasters distribution in terms of one-year consumption drop (green dotted line): about 88% of the multi-year contractions appear to be smaller than the one-year certainty equivalent.[4] That is, the largest 9 contractions have an overwhelming weight in determining the overall consumption risk of a disaster – and the median length of these largest contractions is 5 years.

Figure 6 plots the histogram of *annualized* GDP contractions in the 64 rare disasters considered in the SC and presented in Figure 5, as well as the certainty equivalent consumption drop of the SC approach (red dashed line), and the cumulated GDP (green dotted line) and cumulated consumption (blue dashed dotted line) drops during the U.S. Great Depression period. The first things to notice are that *a*) *all* the disasters are characterized by an annualized contraction smaller than the SC certainty equivalent (45%, dashed red line), *b*) the median annualized contraction (black continuous line) is less than a sixth of that value (being about 7%). That is, the SC certainty equivalent is extreme – and rare – even among rare events, since it lies on the far right tail of the distribution of annualized disasters and it is 14% higher than the largest annualized disaster ever recorded. These findings are not surprising since, as shown in Figure 5, all the twentieth century disaster episodes lasted several years.

Figure 6 also singles out the *cumulated* contractions in GDP (green dotted line) and consumption (blue dash-dotted line) during the U.S. Great Depression. What is remarkable is that the contraction in consumption (17%) is close to half the GDP drop (31%). Moreover, the U.S. Great Depression cumulated contraction in consumption is about 28% lower than

[4]The certainty equivalent is computed using the Barro (2006) calibrated preference parameters.

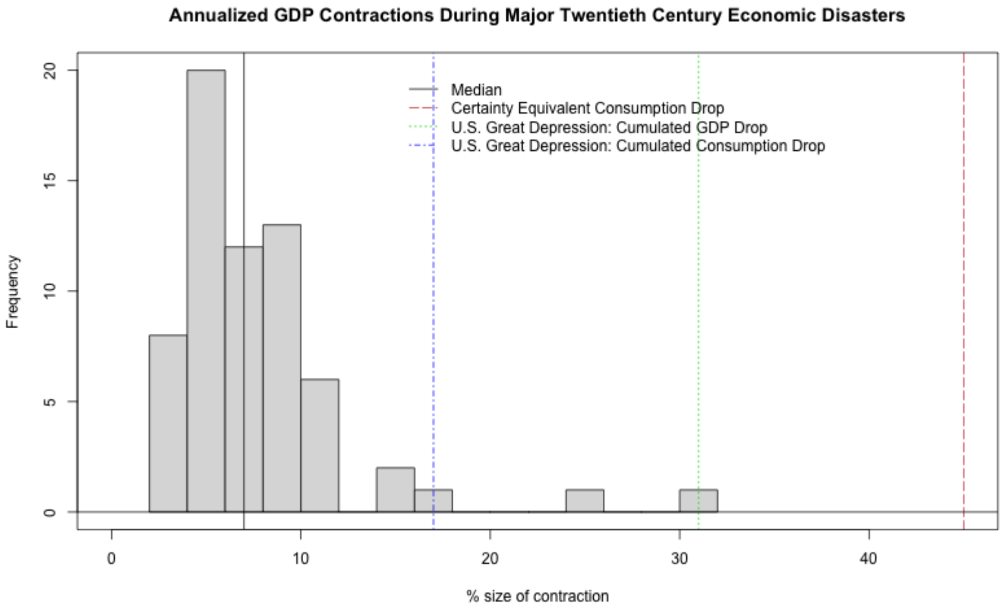


Figure 6: Histogram of annualized GDP contractions during the 64 major economic disasters of the Twentieth Century. Source: Barro (2006), Table I, and authors' calculations.

the SC certainty equivalent – even though the U.S. Great Depression cumulated contraction in GDP (31%) is above the average cumulated GDP contraction during disasters (this being 29%). That is, even in a country characterized by a larger than average economic disaster, the cumulated contraction in consumption is much smaller than the SC certainty equivalent value.

The above observations do not per se rationalize the discrepancy between the results presented in the previous sections and those obtained in the literature that employs the SC approach. We now address this issue formally. In order to do so, we perform the following counterfactual exercise.

First, we modify our baseline annual data sample (1929-2009) by replacing the *four* data points corresponding to the Great Depression period with *one* calibrated disaster observation. In order to assess the effect of SC approach, we calibrate the rare disaster observation in two ways: first as the *cumulated consumption drop* during the U.S. Great Depression, and second as the *cumulated GDP drop* during the same period. If these two assumptions were the reasons behind the discrepancy between our results and the SC ones, we would expect that $a$) using the cumulated consumption drop the estimated risk aversion parameter should be smaller than the one in Panel $A$ of Table 1 (namely, 28.5), and that $b$) using the cumulated contraction in GDP, we should find an even lower risk aversion coefficient and a high likelihood of having the Euler equation satisfied with a RRA smaller than 10. In both modified samples, the market return during the disaster is calibrated to match the annualized excess return during the U.S. Great Depression.

Second, we use these modified samples to estimate and test the consumption Euler Equation. The estimation is performed, as in the paper, using a set of estimators that are robust to rare events problems in the data, since they allow the probabilities attached to different states of the economy to differ from their sample frequencies. Moreover, using the calibration and simulation procedures described in Sections 1.2 and 4.2 of the paper, we elicit the likelihood of observing an EPP if the data were generated by the rare events probability distribution needed to rationalize the puzzle with a low level of risk aversion.

We undertake this counterfactual exercise – that uses a mix of real and fictitious data – in order to perform a controlled experiment, since we know from Table 1 and Table 4 that in the *true* 1929-2009 sample *i*) the C-CAPM is rejected and requires a high level of RRA (28.5) to rationalize the stock market risk premium, and that *ii*) 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 a sample of the same size as the historical one. This counterfactual exercise is presented in Table $B2$.

Panel $A$ of Table $B2$ calibrates the economic disaster to match the cumulative consumption drop during the U.S. Great Depression, while Panel $B$ uses the cumulative GDP contraction during the same period. The first row of each panel reports the $EL$ point estimate (with standard error in parentheses), as well as the BETEL posterior mode (with 95% confidence region in brackets), of the RRA coefficient $\gamma$. The second row of each panel reports the Empirical Likelihood Ratio test (with $p$-value in parentheses) for the joint hypothesis of a $\gamma$ as small as 10 and for the identifying restriction given by the consumption Euler Equation. The third row of each panel reports the BEL and BETEL posterior probabilities of $\gamma$ being smaller than, or equal to, 10. The fourth row of each panel reports the probability of observing an EPP as large as the historical one if rare events were the true explanation of the puzzle.

**Table B2: Estimation and Counterfactual EPP with Calibrated Disaster**

| | $EL$ | $BETEL$ |
|---|---|---|
| *Panel A. U.S. Great Depression Cumulated Consumption Drop.* | | |
| $\hat{\gamma}$ | 19.9<br>(5.16) | 19.9<br>[12.4, 37.5] |
| $\chi^2_{(1)}$ | 5.74<br>(.017) | |
| $\Pr(\gamma \leq 10 \mid \text{data})$ | 0.76% | 0.61% |
| $\Pr\left(epp_i^T(\gamma) \geq epp^T(\gamma)\right)$ | 1.69% | 0.83% |
| *Panel B. U.S. Great Depression Cumulated GDP Drop.* | | |
| $\hat{\gamma}$ | 10.7<br>(2.70) | 10.7<br>[6.80, 20.0] |
| $\chi^2_{(1)}$ | 0.07<br>(.798) | |
| $\Pr(\gamma \leq 10 \mid \text{data})$ | 29.7% | 29.3% |
| $\Pr\left(epp_i^T(\gamma) \geq epp^T(\gamma)\right)$ | 45.0% | 44.8% |

Note: EL and BETEL estimation results for the Euler Equation, and counterfactual probabilities of observing an equity premium puzzle as large as the historical one, when the annual data sample (1929-2009) is modified by replacing the Great Depression consumption observations (1929-1933) with one calibrated economic disaster observation. Panel A uses the cumulated U.S. consumption reduction during the Great Depression. Panel B uses the cumulated U.S. GDP reduction during the Great Depression. The first row of each panel 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 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. The third row of each panel reports the posterior probabilities of $\gamma$ being smaller than, or equal to, 10. The fourth row of each panel reports the counterfactual probability of observing a realized equity premium puzzle as large as the historical one for $\gamma = 10$.

The point estimates in Panel $A$ are almost *a third smaller* than the estimated value in the true 1929-2009 sample (Table 1, Panel $A$). That is, replacing the four Great Depression consumption observations with the cumulated consumption drop during the same period, the level of risk aversion needed to rationalize the equity premium is reduced by almost a third. Moreover, the Bayesian confidence interval covers values that are close to – but still larger than – 10. Nevertheless, as shown in the last three rows of Panel $A$, the rare events hypothesis is unlikely to rationalize the EPP with a risk aversion as low as 10 in a sample of the same size as the historical one.

In Panel $B$ we calibrate the consumption drop during the disaster to the same value as the cumulated GDP drop during the U.S. Great Depression. The point estimate of $\gamma$ is now 10.7 for both the estimators – *almost two thirds smaller* than in the true 1929-2009 sample – and values well below 10 cannot be rejected. Moreover, the posterior probability of a RRA smaller than 10 becomes about 29%. Most importantly, the fourth row stresses that in this case the rare events hypothesis becomes a very likely explanation of the EPP since, under this hypothesis and with a RRA of 10, the likelihood of observing an EPP in a sample of the same size as the historical one ranges from 44.8% to 45.0%.

Overall, the results in Table $B2$ show that if, as in the SC approach, we were to $a$) calibrate an annual model with a cumulative multi-year contraction during disasters, and $b$) overstate the cumulative consumption drop by replacing it with the GDP drop, we would reach the same conclusions as in the previous literature on rare events.

# 6 Probability Weights with Alternative Values of $\gamma$

Figure 7 plots the time series of the $EL$ and $BETEL$ probability weights constructed by setting $\gamma = 4$. Comparing Figure 7 with the corresponding figure in the paper (where $\gamma = 10$), it appears clearly that, changing the value of $\gamma$, the sets of events that need to receive higher probability weights in order to rationalize the EPP with a low level of RRA, stays unchanged. The only difference is that, in order to rationalize the puzzle with a lower level of risk aversion, the probabilities assigned to a few economy wide extremely bad states, such as market crashes concomitant with deep recessions, have to be marginally increased. In fact, the correlation between the $EL$ probability weights for $\gamma = 4$ and $\gamma = 10$ is 0.96 and that between the $BETEL$ weights is 0.98. This suggests robustness of the approach with respect to the choice of the relative risk aversion parameter.

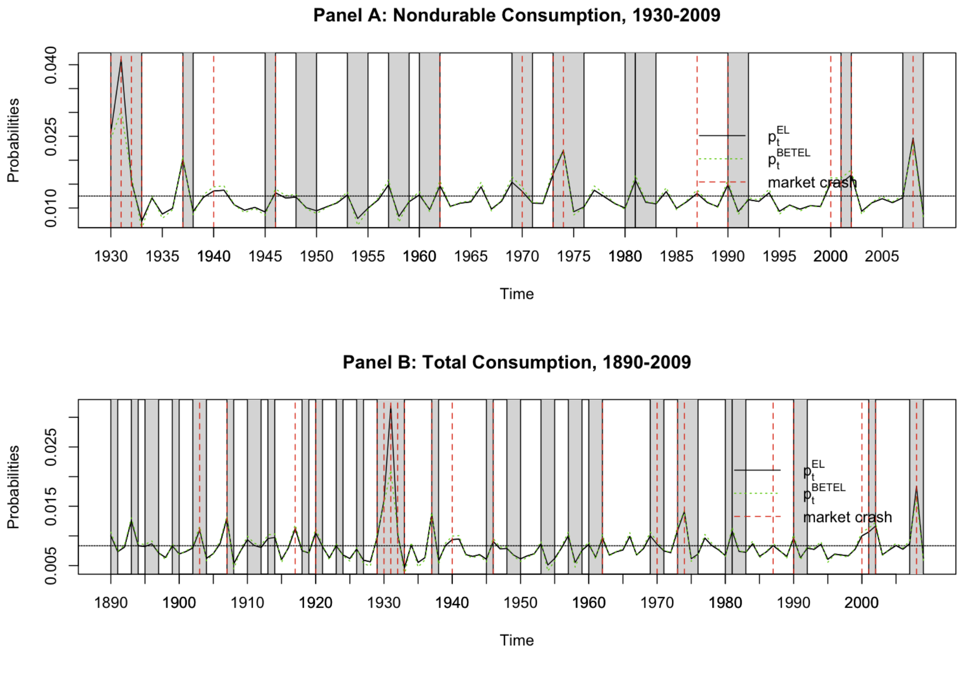


Figure 7: EL and BETEL estimated probabilities for $\gamma = 4$. Shaded areas are NBER recession periods. Vertical dashed lines are the stock market crashed identified by Mishkin and White (2002).

# 7 The Likelihood of the Equity Premium Puzzle with Blockwise Sampling

In this section we repeat the counterfactual exercise presented in Section 4.2 of the paper, but instead of drawing individual couples of consumption growth and excess returns, we draw consecutive *blocks* of data in order to preserve, as in a blockwise bootstrap procedure, the autocorrelation properties of the data. The $EL$ and $BETEL$ probabilities for the blocks are constructed using the blockwise approach described in Section 1 of this Appendix, and a window of five years in the two annual samples.

Results obtained using this approach are reported in Table $B3$, and the findings are largely in line with the ones in Table 4 of the paper: the median realized EPP in the counterfactual samples is very close to zero and its 95% confidence region does not include the historically observed value. Moreover, the probability of observing an EPP in the counterfactual samples is very small, and never larger than 2.5%. That is, as Table 4, Table $B3$ suggests that if one believes that the rare events hypothesis is the explanation of the EPP, one should also believe that the historically observed EPP is itself a rare event.

**Table B3: Counterfactual EPP with Blockwise Sampling**

| | $epp^T$ | $epp_i^T$ | $\Pr\left(epp_i^T \geq epp^T\right)$ |
|---|---|---|---|
| *Panel A. Annual Data: 1929-2009* | | | |
| $\hat{P}^{EL}\,(\gamma=4)$ | 6.9% | −0.1% [−4.3%, 4.7%] | 0.16% |
| $\hat{P}^{EL}\,(\gamma=10)$ | 5.9% | −0.0% [−5.3%, 5.9%] | 2.49% |
| $\hat{P}^{BETEL}\,(\gamma=4)$ | 6.9% | −0.1% [−3.4%, 3.6%] | 0.01% |
| $\hat{P}^{BETEL}\,(\gamma=10)$ | 5.9% | −0.0% [−4.5%, 4.9%] | 0.78% |
| *Panel B. Annual Data: 1890-2009* | | | |
| $\hat{P}^{EL}\,(\gamma=4)$ | 6.1% | −0.0% [−4.1%, 4.2%] | 0.22% |
| $\hat{P}^{EL}\,(\gamma=10)$ | 5.8% | 0.1% [−4.9%, 5.2%] | 1.41% |
| $\hat{P}^{BETEL}\,(\gamma=4)$ | 6.1% | −0.0% [−3.3%, 3.3%] | 0.02% |
| $\hat{P}^{BETEL}\,(\gamma=10)$ | 5.8% | 0.0% [−4.2%, 4.5%] | 0.61% |

Note: the $epp^T$ column reports the realized equity premium puzzle in the historical sample corresponding to the given level of $\gamma$; the $epp_i^T$ column reports the median realized equity premium puzzle (and its 95% confidence band underneath) in the counterfactual samples for to the given level of $\gamma$ and probability distribution $\hat{P}^j(\gamma)$, $j \in \{EL, ET\}$, used to generate the data. The $\Pr\left(epp_i^T \geq epp^T\right)$ column reports the probability of observing a realized equity premium puzzle as large as the historical one in the $epp^T(\gamma)$ column.

# 8 Additional Cross-Sectional Evidence

In this section we present results from Fama and MacBeth (1973) cross-sectional regressions when the 6 Fama-French portfolios are used to construct the $EL$ and $BETEL$ probability weights and the 25 Fama-French portfolios are used to perform the cross-sectional estimation.

The results are reported in Table $B4$. The first row shows that, using the sample moments of excess returns and pricing kernel, the estimated $\hat{\alpha}$ is large and statistically significant. This is the cross-sectional equivalent of the EPP, since it implies an average underpricing of the cross-section of the 25 portfolio returns of about 10% on an annual basis. The point estimate of $\hat{\lambda}$ is not statistically different from its theoretical value of unity, but this is due to the large standard error that makes it also not statistically different from zero. The joint test of the hypothesis $\alpha = 0,\ \lambda = 1$ is strongly rejected with p-value 0.0%. Overall, consistent with what has been widely documented in the literature, the performance of the C-CAPM is poor since the model is able to explain only 0.2% of the cross-sectional variance of risk premia of the Fama and French (1992) 25 portfolios.

The second row shows that the use of moments constructed under the probability measure $\hat{P}^{EL}(\gamma)$ substantially reduces the estimated $\alpha$ coefficient, implying a much smaller average mispricing (about 4.0% on a yearly basis), and the coefficient is not statistically significant – consistent with the implications of the consumption Euler equation. Even though the cross-sectional $R^2$ is higher than that obtained using the sample moments in Row 1, the estimated $\hat{\lambda}$ coefficient has the opposite sign to and is statistically different from what theory would predict. The Wald test strongly rejects the joint hypothesis $\alpha = 0,\ \lambda = 1$ with p-value 1.2%.

The results in Row 3 using the $\hat{P}^{BETEL}(\gamma)$ probability weights are very similar to those obtained using the $\hat{P}^{EL}(\gamma)$ weights: the model fits better the average risk premium but the estimated slope coefficient is statistically significantly negative and different from the theoretical value of unity, and the joint hypothesis $\alpha = 0,\ \lambda = 1$ is strongly rejected with p-value 0.0%.

**Table B4: Counterfactual Cross-Sectional Regressions**

| Moments: | $R^2$ | $\hat{\alpha}$ | $\hat{\lambda}$ | Wald test: $\alpha = 0,\ \lambda = 1$ | $\Delta \frac{Var(\beta_m)}{Var[E(R_m^e)]}$ |
|---|---|---|---|---|---|
| | | *Fama-French 25 Portfolios* | | | |
| Sample | 0.2% | 0.10 (0.026) | 0.38 (0.801) | 15.8 (0.00) | |
| $\hat{P}^{EL}(\gamma)$ | 9.8% | 0.04 (0.032) | −1.24 (0.796) | 8.8 (0.01) | 289.4% |
| $\hat{P}^{BETEL}(\gamma)$ | 16.8% | 0.06 (0.030) | −1.72 (0.790) | 14.0 (0.00) | 250.7% |

Note: Fama and MacBeth (1973) cross-sectional regression results for the 25 Fama-French portfolios. Row 1 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 $\gamma = 10$.

# References


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 $20^{th}$ 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.

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

GOOD, D. F., AND T. MA (2006): "The economic growth of Central and Eastern Europe in comparative perspective, 1870-1989," *European Review of Economic History*, 3(02), 103–137.

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

KITAMURA, Y. (1997): "Empirical Likelihood Methods with Weakly Dependent Processes," *Annals of Statistics*, 25, 2084–2102.

MADDISON, A. (2003): *The World Economy: Historical Statistics*. OECD, Paris.

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.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–08.

PARKER, J. A., AND C. JULLIARD (2005): "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy*, 113(1).

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.

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