People Are Less Risk-Averse than Economists Think*

Ali Elminejad[a], Tomas Havranek[a,b], Zuzana Irsova[a]

[a]Charles University, Prague

[b]CEPR

June 27, 2022

Abstract

We collect 1,021 estimates from 92 studies that use the consumption Euler equation to measure relative risk aversion and that disentangle it from intertemporal substitution. We show that calibrations of risk aversion are typically larger than estimates thereof. Moreover, reported estimates are typically larger than the underlying risk aversion because of publication bias. After correction for the bias, the literature suggests a mean risk aversion of 1 in economics and 2–7 in finance contexts. The reported estimates are systematically driven by the characteristics of data (frequency, dimension, country, stockholding) and utility (functional form, treatment of durables). To obtain these results we use nonlinear techniques to correct for publication bias and Bayesian model averaging techniques to account for model uncertainty.

Keywords: Euler equation, risk aversion, Epstein-Zin preferences, meta-analysis, publication bias, Bayesian model averaging

JEL Codes: C83, D81, D90

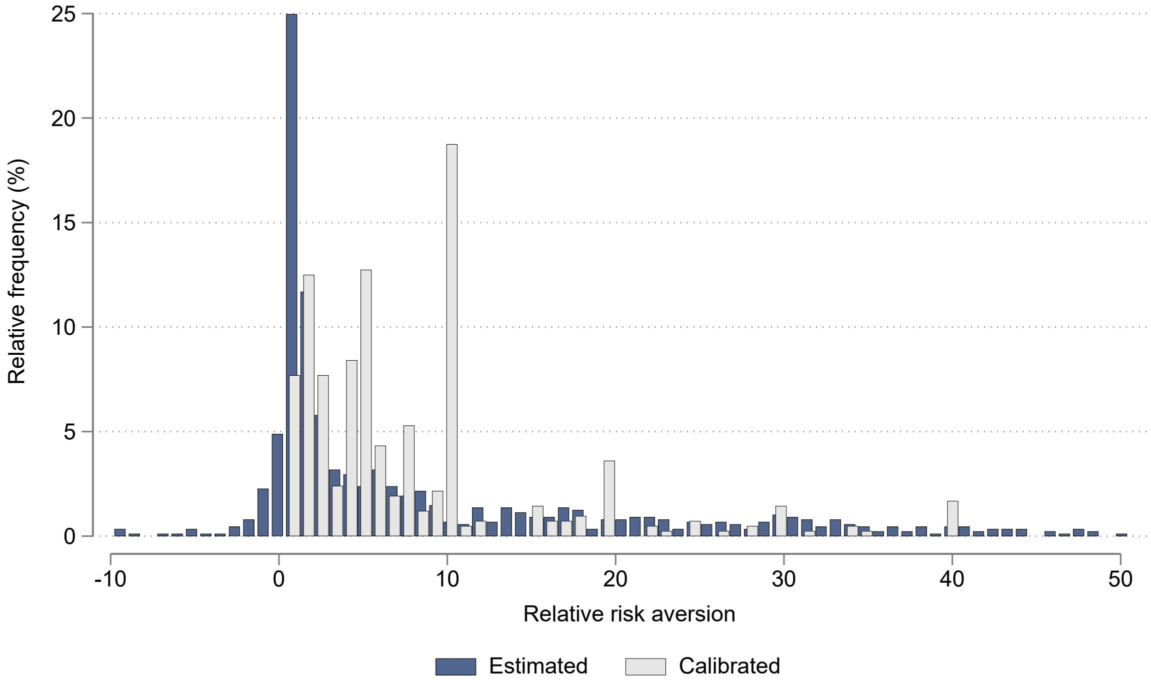
1 Introduction

Risk aversion is a key concept in economics and finance. Almost every structural model requires assumptions concerning relative risk aversion, and dozens of studies have estimated the corresponding coefficient using the consumption Euler equation. Yet no consensus on the appropriate

*An online appendix with data and code is available at `meta-analysis.cz/risk`. Corresponding author: Zuzana Irsova, `zuzana.irsova@ies-prague.org`.

calibration values has emerged, as Figure 1 demonstrates: common values are 2.5, 5, and 10, but 1 and 20 also appear often. Remarkably, the distribution of calibrations does not match the distribution of estimates. The most common estimated value is 1, while the most common calibration is 10. But the figure also shows that almost every calibrated value up to at least 50 can be justified by some empirical estimates. There are few guidelines on the calibrations of relative risk aversion, and no quantitative synthesis (or meta-analysis) has attempted to shed light on the issue. That is what we attempt to deliver in this paper.

Figure 1: Calibrations of risk aversion overtop most estimates thereof



Notes: The figure shows histograms of i) 1,021 estimates or relative risk aversion collected from 92 studies and ii) 446 calibrations of relative risk aversion collected from 200 studies. In both cases we only consider studies that separate risk aversion from intertemporal substitution. For ease of exposition, values below -10 and above 50 are excluded from the figure but included in all statistical tests. Summary statistics are available in Table 1. Separate figures for economics and finance literatures are available in Figure B1 and Figure B2, respectively.

The absence of a meta-analysis on the topic can perhaps be explained by the sheer size of the literature on risk aversion. Risk aversion can be estimated using lab experiments, surveys, labor-supply behavior, auction behavior, choices in insurance contracts, option prices, and game show contestant behavior (see, for example, Zhang *et al.*, 2014). We focus on the consumption Euler equation approach, which constitutes the benchmark framework employed in economics and finance. The problem is that most studies in this literature assume power utility, which means that relative risk aversion equals the reciprocal of the elasticity of intertemporal substitution, and hence the interpretation of the estimated parameter is unclear. We thus concentrate on the relatively small subset of the literature that separates risk aversion from intertemporal substitution. The separation is typically done by employing Epstein-Zin preferences (Epstein & Zin, 1989, 1991), but can also be achieved using habits in consumption, expected utility with a reference level of consumption, ambiguity aversion, or disappointment aversion. Even this small subset of the Euler equation literature yields 1,021 estimates from 92 studies. To construct Figure 1 we also collect 446 calibrations from 200 studies, once again only those that break the link between risk aversion and intertemporal substitution.

Four previous studies are intimately related to the analysis we present. Havranek (2015) conducts a meta-analysis of the elasticity of intertemporal substitution in consumption. After correcting the literature for various biases, he argues that the best guess concerning the mean elasticity of substitution is 1/3. Because almost all studies in his sample use power utility, the finding translates to the relative risk aversion of 3—if we accept the argument by Kocherlakota (1990), contrary to Hall (1988), that the parameter derived from the corresponding Euler equation is more informative about risk aversion than intertemporal substitution. Ascari *et al.* (2021) present a recent and meticulous estimation, robust to weak instruments, of all parameters that can be derived from the consumption Euler equation. They find that the potential range for relative risk aversion is wide. Brown *et al.* (2022) conduct a meta-analysis of loss aversion, a concept related to but distinct from relative risk aversion as commonly used in economics, and find that the mean loss aversion is around 2 after correction for several biases. Imai *et al.* (2021) present a meta-analysis of the present bias, which some argue (prominently, Dean & Ortoleva, 2019) is strongly related to risk preferences. The corrected mean present bias recovered by Imai *et al.* (2021) is between 0.95 and 0.97.

Key issues for meta-analysis are the twin problems of publication bias and p-hacking. Publication bias describes a situation in which authors, referees, or editors, intentionally or not, refuse to publish estimates that are statistically insignificant or inconsistent with the theory (for example, have the wrong sign). P-hacking is the effort by authors, again intentional or not, to produce publishable results: for example, by trying different subsamples or control variables until the estimate reaches statistical significance. McCloskey & Ziliak (2019) invoke a nice analogy to the Lombard effect in psychoacoustics: speakers involuntarily increase their vocal effort in the presence of noise. In a similar way can researchers respond to noise in their data or techniques and try harder till they obtain a point estimate large enough to compensate for the large standard error. Note that publication bias and p-hacking are observationally equivalent, so for parsimony we will use the term publication bias to describe both, as is common in the meta-analysis literature. Many studies have recently discussed how publication bias can exaggerate empirical estimates in economics (Brodeur *et al.*, 2016; Bruns & Ioannidis, 2016; Card *et al.*, 2018; Christensen & Miguel, 2018; DellaVigna *et al.*, 2019; Blanco-Perez & Brodeur, 2020; Brodeur *et al.*, 2020; Ugur *et al.*, 2020; Xue *et al.*, 2020; Neisser, 2021; Stanley *et al.*, 2021; DellaVigna & Linos, 2022; Stanley *et al.*, 2022), and the exaggeration can be twofold or more (Ioannidis *et al.*, 2017). Publication bias is natural, common in economics, and does not imply cheating or any ulterior motives on the part of the researchers. But it is a serious problem for the interpretation of the results in the literature, a problem meta-analysis can tackle.

Most meta-analysis techniques used for publication bias correction in economics and finance rely on the Lombard effect and regress estimates on their standard errors (meta-regression). Evidence of a nonzero slope is commonly taken as evidence for publication bias, and the constant in the regression measures the mean estimate conditional on maximum precision, often interpreted as the mean corrected for the bias. There are two problems with such a strategy. First, as shown by Andrews & Kasy (2019) and Stanley & Doucouliagos (2014), publication bias can be a nonlinear function of the standard error. Second, as discussed by Havranek *et al.* (2022), the assumption of no correlation between estimates and standard errors in the absence of publication bias can be problematic because of unobserved heterogeneity that affects both estimates and standard errors. To address these two problems, we employ recently developed nonlinear tests for publication bias: the selection model by Andrews & Kasy (2019), the weighted average of adequately powered estimates (Ioannidis *et al.*, 2017), the stem-based technique (Furukawa, 2021), the endogenous kink model (Bom & Rachinger, 2019), and the p-uniform* technique (van Aert & van Assen, 2021).

In the second part of the analysis we investigate the heterogeneity in the reported estimates of relative risk aversion. We identify 30 characteristics of data, specification, estimation, and publication that reflect the context in which the estimates are obtained and that may affect the estimates. The characteristics are so numerous because of the many choices researchers have to make when specifying their models. In consequence, substantial model uncertainty arises in meta-analysis when we want to relate estimates of risk aversion to estimation context. As a solution we use Bayesian model averaging (see, e.g., Zeugner & Feldkircher, 2015; Steel, 2020), which is the natural response to model uncertainty in a Bayesian setting; moreover, it is computationally less cumbersome than frequentist alternatives. Bayesian model averaging also allows us to partially address collinearity by employing the dilution prior (George, 2010), which penalizes models with a small determinant of the correlation matrix.

We find substantial publication bias in the empirical literature on relative risk aversion. The mean amount of exaggeration due to the bias is striking: about seven-fold in both economics and finance. The corrected mean relative risk aversion is 1 in the economics literature and 2–7 in the finance literature (where different correction techniques give quantitatively different results, but all agree that publication bias is strong). The correction for publication bias further widens the gap between typical estimates and typical calibrations presented earlier in Figure 1. In particular, the value of 10 most frequently used for calibration is inconsistent with the bulk of empirical estimates. In contrast, the second most common calibration, 5, is well within the plausible range of estimates suggested by the literature in finance (but not economics) contexts. Note also that the mean estimate of 1 obtained for economics does not lend itself to the recommendation of the logarithmic utility function in that field. The reason is, as we have mentioned earlier, that the elasticity of intertemporal substitution is typically not 1 but around 1/3 (Havranek, 2015). In finance contexts, power utility with relative risk aversion set at 3 thus seems relatively consistent with empirical evidence.

When we allow for heterogeneity by employing Bayesian model averaging, we confirm the finding of strong publication bias and a substantial difference in estimated risk aversion between economics and finance contexts—even after other aspects of data and methods are controlled for. In addition, studies that focus on stockholders tend to find substantially smaller values of risk aversion, which is consistent with both intuition and previous results (such as Mankiw & Zeldes, 1991). Finally, reported estimates of relative risk aversion are systematically related to data characteristics (frequency, dimension, and country coverage) and the definition of the utility function (the assumption of separability between durables and nondurables and the use of Epstein-Zin preferences in contrast to other methods for separating risk aversion from intertemporal substitution). The results are reasonably robust to alternative priors for Bayesian model averaging.

2 Data

Details on the estimation of relative risk aversion in the context of the consumption Euler equation are available in Appendix B; the estimation approaches followed by most studies are also clearly described by Epstein & Zin (1991) and Vissing-Jørgensen & Attanasio (2003). A more general overview of modeling risk aversion is presented by O'Donoghue & Somerville (2018). Appendix A provides details on the way we search the literature for estimates of relative risk aversion. We start with a search query in Google Scholar, which we prefer over alternative databases because of its universal coverage and full-text capabilities. The search query yields more than 3,500 studies. For feasibility, we only inspect the first 1,500 studies returned by the search. We read the abstracts of these studies and download those that indicate any chance of containing empirical estimates of risk aversion (about a half of the examined studies).

We read the downloaded studies and include those that conform to the following three criteria. First, the study must use the consumption Euler equation to obtain an empirical estimate of the coefficient of relative risk aversion. Second, the estimate must be reported together with the corresponding standard error or any statistics from which the standard error can be computed. Third, the study must separate risk aversion from intertemporal substitution. We collect both published and unpublished papers, and terminate the search on May 16, 2022. The search yields 92 papers (called "primary studies" in the meta-analysis terminology and listed in Table B1), which together provide 1,021 estimates of relative risk aversion. The sample of calibration studies is assembled using a similar search strategy with the following differences: in the search query we replace the word "estimate" with "calibration", restrict our attention to published papers, and stop once we collect 200 usable studies (ranked by the order they appear in the Google Scholar reply to our query). This approach yields 446 individual calibrated values of relative risk aversion.

In addition to calibrations, estimates, and the estimates' standard errors, we also collect 30 variables, described in Section 4, that reflect the context in which the estimates are obtained in primary studies: the characteristics of data, specification, estimation, and publication. This means we collect manually more than 30,000 data points. To reduce the danger of mistakes and typos, two of the co-authors collect the data independently, and the third co-author resolves inconsistencies between these two datasets. The resulting clean dataset is available in the online appendix at `meta-analysis.cz/risk` together with the code used in this analysis and the list of 200 calibration studies.

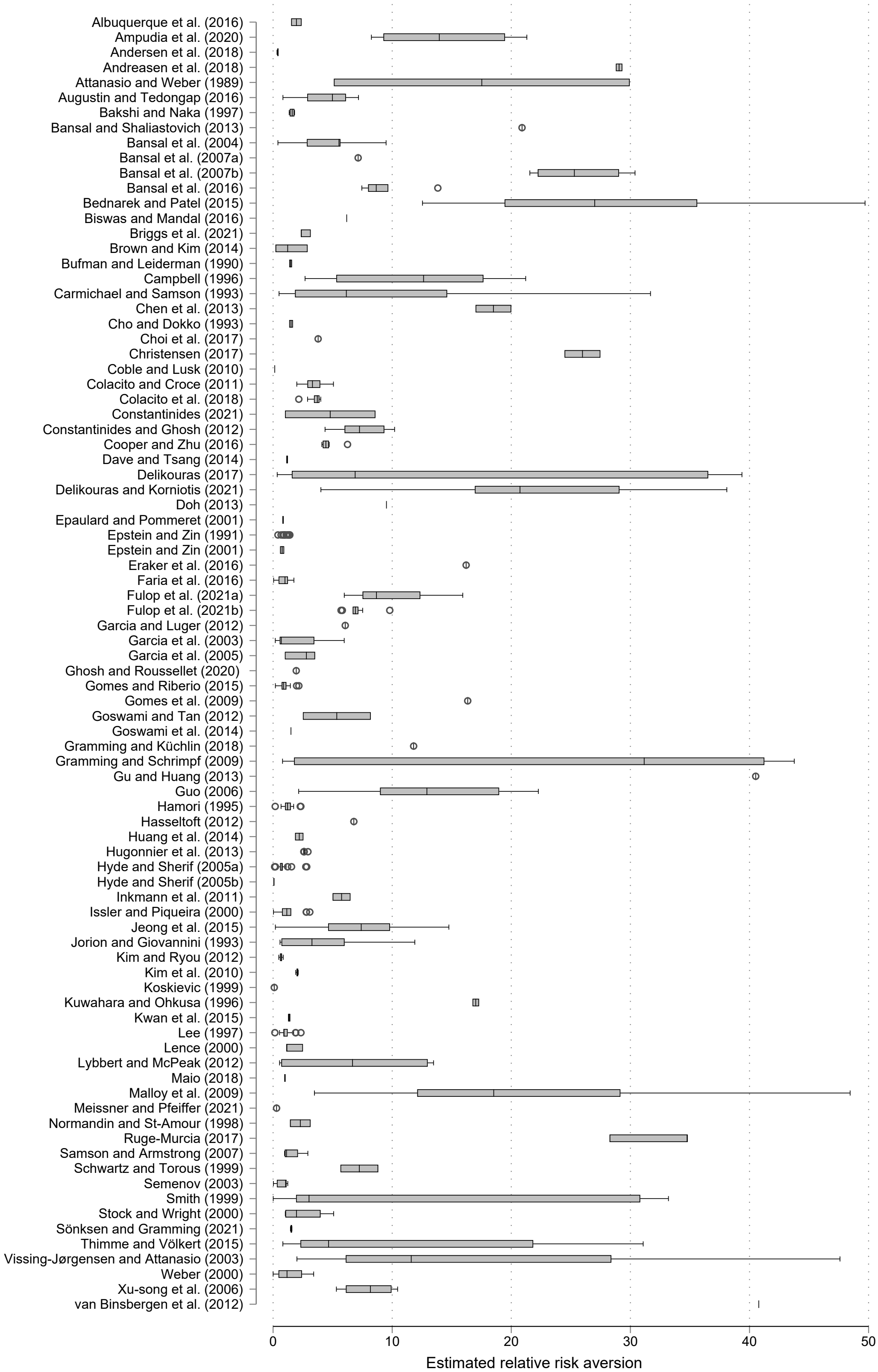
Table 1: Summary statistics of estimated and calibrated relative risk aversion

Panel A: Estimates				
	Observations	Mean	Median	Standard deviation
All 92 studies	1,021	23.36	3.77	98.58
Economics (58 studies)	590	7.50	1.26	30.74
Finance (34 studies)	431	45.05	17.82	144.71
Panel B: Calibrations				
	Observations	Mean	Median	Standard deviation
All 200 studies	446	14.33	6.00	30.10
Economics (115 studies)	237	17.14	6.00	35.74
Finance (85 studies)	209	11.12	6.00	21.66

Notes: We only consider studies that separate risk aversion from intertemporal substitution. Studies are classified into economics and finance categories based on the journals they were published in and using the journal classification of the Web of Science. If in the Web of Science the journal is included in both categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking. If a study is unpublished (15 studies in total), we classify it based on the prevailing publications of the corresponding author. In the meta-analysis we winsorize estimates at the 5% level. Summary statistics for benchmark calibrations from each study in Panel B are reported in Table B2.

Throughout the paper we distinguish between estimates obtained in economics and finance contexts. The precise boundary is hard to draw: estimates in economics are often, but not always, derived from approaches that focus on the entire economy, while finance estimates tend to focus almost exclusively on asset prices (see Appendix B for details). We choose a classification based on the journal in which the primary or calibration study is published and follow the categories defined by the Web of Science. If in the Web of Science the journal is included in both categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking. If a study is unpublished (15 primary studies in total), we classify it based on the prevailing publications of the corresponding author. In such a way each study can be unambiguously classified into either economics or finance.

Figure 2: Estimates of risk aversion vary both across and within studies



Notes: The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. For ease of exposition, outliers are excluded from the figure but included in all statistical tests.

Table 1 presents the summary statistics of estimates and calibrations, and Figure 2 shows that the estimates vary widely both within and across studies. As we have noted in the discussion of Figure 1, calibrations of risk aversion in the literature tend to be larger than most empirical estimates. But Table 1 also shows that the story differs between economics and finance. In economics, calibrations are indeed much larger than estimates, both in terms of mean and median values; the corresponding histogram is available in Figure B1 in Appendix B. In finance, the opposite is the case: estimates overtop calibrations (Figure B2). Calibrations in both fields are very similar to each other, with a median of 6 and mean around 15 (The pattern holds for the set of benchmark calibrations from each study; see Table B2.) Figure B2 shows that while even in finance the estimates of risk aversion between 1 and 10 are the most common, values around 20 and larger are also routinely reported.

Curiously, therefore, calibrations of relative risk aversion in both fields seem to have little basis in the distribution of the empirical estimates of the parameter in a given field. Instead, many calibrations simply quote Mehra & Prescott (1985), who argue that 10 is a reasonable upper bound for the coefficient of relative risk aversion. Because large risk aversion is often sought for calibration (for example, to help explain the equity premium puzzle), it follows that 10 is the most frequently used calibration value by a large margin. Values of 2.5, 5, and 20 present the most common robustness checks to the baseline calibration. Our goal in this paper is to help reconnect calibrations of risk aversion to empirical estimates thereof, and the first necessary step is the correction of the estimates for publication selection bias.

3 Publication Bias

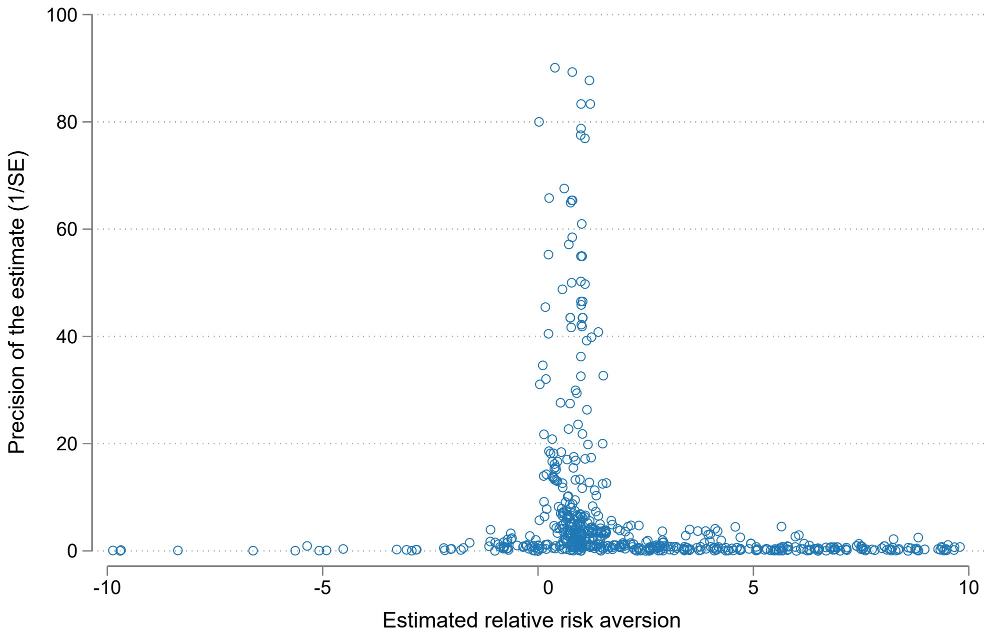
Economists expect that most people are risk-averse, and hence that the mean coefficient of relative risk aversion in any group is positive. This belief is reflected by the 446 calibrations shown earlier in Figure 1: all of them are positive. Negative or zero risk aversion bodes well with few economics and finance models. Of course, the underlying mean coefficient of relative risk aversion is most likely substantially positive. But unless it is huge, researchers will sometimes run into estimation contexts in which the estimate of the coefficient turns out to be insignificantly different from zero or even negative. Noise in the data or methods will produce such counter-intuitive results from time to time. In a similar way, noise will also produce estimates that are too large and away from the true mean. The problem is that while it is difficult to identify the implausibly large estimates (no upper threshold exists for risk aversion), researchers immediately spot and investigate those that are negative or statistically insignificant. Given such unintuitive results, researchers may choose not to report them, or try a different specification in the hope of obtaining results that are consistent with their priors. Such a censoring drives the mean reported risk aversion upwards from the true value, and this is what meta-analysts call publication bias (Card & Krueger, 1995; Stanley, 2001).

The process leading to publication bias is not necessarily detrimental to science, and certainly it does not need to involve any ulterior motives on the part of the researchers. In most cases it will improve the inference of any study if it does not focus on negative or insignificant estimates of relative risk aversion. After all, these "nonsensical" estimates are likely to be caused by some problems in data or methods. But researchers do not winsorize: they typically treat small estimates with suspicion, but not those that are large. Thus, at the level of the entire literature, a bias arises that exaggerates the true mean effect. The tension between these two aspects of publication bias is nicely illustrated by the following quote due to Uhlig (2012, p. 38) about empirical evidence on monetary policy transmission:

> At a Carnegie-Rochester conference a few years back, Ben Bernanke presented an empirical paper, in which the conclusions nicely lined up with a priori reasoning about monetary policy. Christopher Sims then asked him, whether he would have presented the results, had they turned out to be at odds instead. His half-joking reply was, that he presumably would not have been invited if that had been so. There indeed is the danger (or is it a valuable principle?) that a priori economic theoretical biases filter the empirical evidence that can be brought to the table in the first place.

How to test and correct for publication bias? The histogram of the estimates shown in Figure 1 does not really help, though it suggests that the bias is not universal: some negative estimates of risk aversion do appear in the literature. A neat way to measure publication bias is to compare the results of original studies and pre-registered replications (Kvarven *et al.*, 2020), the latter being unlikely to suffer from much bias. But there are no pre-registered replications of studies estimating relative risk aversion using the Euler equation; in general, pre-registration is most efficient in the experimental literature where researchers cannot inspect their data prior to pre-registration (Olken, 2015). To correct for the bias, we thus rely on

Figure 3: The funnel plot suggests publication bias



Notes: In the absence of publication bias (and any small-sample and heterogeneity-related biases), the plot should form a a symmetrical inverted funnel. Outliers are excluded from the figure for ease of exposition but included in all tests.

techniques traditionally used by medical researchers and new methods recently developed by econometricians and psychologists.

The starting point is a visual examination of the so-called funnel plot, often used in medical research (Egger *et al.*, 1997; Stanley & Doucouliagos, 2010). The funnel plot, Figure 3, is a scatter plot of point estimates on the horizontal axis and the estimates' precision (reciprocal of the standard error) on the vertical axis. In the absence of systematic heterogeneity, which will be examined in the next section, the most precise estimates should be close to the underlying mean coefficient of relative risk aversion. As precision decreases, the estimates should be more widely dispersed around the true mean value. Because in the absence of publication bias all estimates have the same chance of being reported, the funnel will be symmetrical: all imprecise estimates are published, both those that are negative and those that are huge and positive. Figure 3 shows that, first, the funnel is asymmetrical, which indicates publication bias against small estimates of risk aversion. Second, the most precise estimates are concentrated around 1.

Table 2 shows the results of more formal tests of funnel asymmetry and the underlying risk aversion beyond publication bias. The tests are regressions of estimates on standard errors and can also be interpreted as tests of the Lombard effect discussed in the Introduction (researchers

Table 2: Funnel asymmetry tests indicate modest risk aversion beyond publication bias

Panel A: All studies	WLS	FE	BE	Study
Standard error	1.865***	2.287***	2.837	3.062***
(*publication bias*)	(0.362)	(0.713)	(1.760)	(0.893)
	[0.956, 2.577]			[1.251, 4.900]
Constant	1.199***	1.084***	1.590***	1.533***
(*mean corrected RRA*)	(0.257)	(0.194)	(0.235)	(0.412)
	[0.725, 2.130]			[0.673, 2.476]
Observations	1,021	1,021	1,021	1,021
Studies	92	92	92	92

Panel B: Economics	WLS	FE	BE	Study
Standard error	1.392***	1.411	4.119***	3.604***
(*publication bias*)	(0.540)	(1.146)	(1.361)	(0.827)
	[0.383, 2.506]			[2.007, 5.293]
Constant	1.085***	1.082***	0.714***	0.822***
(*mean corrected RRA*)	(0.261)	(0.211)	(0.178)	(0.243)
	[0.654, 2.059]			[0.351, 1.464]
Observations	590	590	590	590
Studies	58	58	58	58

Panel C: Finance	WLS	FE	BE	Study
Standard error	1.859***	3.476***	0.817	2.168
(*publication bias*)	(0.449)	(0.169)	(3.061)	(1.654)
	[0.050, 2.895]			[-1.197, 5.548]
Constant	2.390***	1.107***	3.223***	2.888***
(*mean corrected RRA*)	(0.675)	(0.134)	(0.423)	(0.732)
	[0.812, 4.006]			[1.062, 4.89]
Observations	431	431	431	431
Studies	34	34	34	34

Notes: We regress estimates of relative risk aversion on their standard errors (weighted by inverse variance). Standard errors, clustered at the study level, are reported in parentheses. RRA = relative risk aversion. WLS = standard weighted least squares. FE = study fixed effects. BE = study between effects. Study = the inverse of the number of estimates reported per study is used as an additional weight. In square brackets we show the 95% confidence interval from wild bootstrap (Roodman *et al.*, 2019). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

increase their specification search effort in response to noise in their data or methods). The estimated slope in the regression measures the extent of publication bias. The intercept can be interpreted as the mean coefficient of relative risk aversion corrected for publication bias: if we assume that publication bias is indeed a linear function of the standard error. (This is a strong assumption that we will later relax.) We account for the obvious heteroskedasticity by weighting the regressions by inverse variance (Stanley & Doucouliagos, 2014, 2015). We employ four specifications: standard weighted least squares, study-level fixed effects, study-level between effects, and a specification that additionally weights estimates by the inverse of the number of estimates reported by each study, thus giving each study the same weight. All specifications except between effects report standard errors clustered at the study level; for the first and last specification we also report confidence intervals based on wild bootstrap.

In all cases we obtain estimated coefficients for publication bias that are positive and large, in line with the funnel plot. Most of them are also statistically significant at the 5% level. Given that this test for publication bias is known to have relatively low power (Stanley, 2008), the results are consistent with substantial bias. The corrected mean coefficient of relative risk aversion is around 1 for economics and 1–3 for finance, compared with uncorrected means of 7.5 and 45, respectively. The estimated exaggeration due to publication bias is striking and much larger than what is typical in economics: Ioannidis *et al.* (2017) report that the mean exaggeration due to publication bias is twofold. Next, we relax the assumption that publication bias is a linear function of the standard error, which has been criticized by Andrews & Kasy (2019) and Stanley & Doucouliagos (2014). In doing so, we rely on recently developed nonlinear models of publication bias.

We use five nonlinear techniques for publication bias correction. First, the weighted average of adequately powered estimates by Ioannidis *et al.* (2017). The technique estimates retrospective power for all estimates and yields a result that is the average of the estimates with power above 80% (weighted by inverse variance). Second, the selection model by Andrews & Kasy (2019). This rigorously founded technique estimates the probability that negative and insignificant estimates are not reported; the probability is then used to upweight these estimates. Third, the stem-based technique by Furukawa (2021). The technique exploits the trade-off between bias and variance: when more imprecise studies are added, publication bias increases, but variance

Table 3: Nonlinear corrections for publication bias

Panel A: All studies	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Furukawa (2021)	Bom & Rachinger (2019)	van Aert & van Assen (2021)
Mean corrected RRA	1.318***	0.960***	1.467***	1.199***	0.367***
	(0.250)	(0.035)	(0.951)	(0.046)	[0.002]
Observations	1,021	1,021	1,021	1,021	1,021
Studies	92	92	92	92	92
Panel B: Economics	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Furukawa (2021)	Bom & Rachinger (2019)	van Aert & van Assen (2021)
Mean corrected RRA	1.172***	0.910***	0.474***	1.085***	0.366***
	(0.250)	(0.030)	(0.390)	(0.052)	[0.002]
Observations	590	590	590	590	590
Studies	58	58	58	58	58
Panel C: Finance	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Furukawa (2021)	Bom & Rachinger (2019)	van Aert & van Assen (2021)
Mean corrected RRA	2.535***	11.196***	6.100***	2.390***	0.625***
	(0.662)	(1.212)	(0.885)	(0.112)	[0.008]
Observations	431	431	431	431	431
Studies	34	34	34	34	34

Notes: RRA = relative risk aversion. Standard errors are reported in parentheses; the p-uniform* technique due to van Aert & van Assen (2021) only yields p-values, which we report in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

decreases because more estimates are available. Furukawa (2021) minimizes the corresponding mean squared error that is the sum of bias and variance. Fourth, the endogeneous kink model by Bom & Rachinger (2019). The technique assumes that the relationship between estimates and standard errors is linear when precision is low but that no relationship exists when precision is sufficiently high. For example, if the p-value is 0.001, publication probability is not affected by small changes in the standard error. Fifth, the p-uniform* model by van Aert & van Assen (2021). The technique, developed in psychology, works with the distribution of p-values and uses the statistical principle that the distribution should be uniform at the true mean value of the coefficient of relative risk aversion.

The first four techniques introduced above assume that there is no correlation between estimates and standard errors in the absence of publication bias. This is a common meta-analysis assumption that traces its roots back to medical research, where meta-analysis was first developed and applied. But in economics the assumption is problematic. Most of the research here is observational, which means much more heterogeneity and choices that researchers have to make: and not all of the choices are (or can be) reported. It is entirely plausible that certain aspects of data or methods affect both estimates and standard errors, thereby creating a correlation even in the absence of publication bias. For example, studies using instrumental variables may correct for an endogeneity bias, but the resulting estimates also tend to be less precise. Indeed, the estimates are correlated with standard errors even if we employ a subsample of estimates that are likely to be published in any case because they are highly statistically significant (with p-values < 0.005); see Table C3 in Appendix C. In Table C2 we perform a test, due to Kranz & Putz (2022), of the Andrews & Kasy (2019) model. The lack of correlation between estimates and standard errors in the absence of publication bias is the key assumption of the model, but the Kranz & Putz (2022) test concerns the Andrews & Kasy (2019) model as a whole, including the assumption of constant publication probabilities for estimates with the same classification of statistical significance (for example, p-values between 0.05 and 0.1). The test rejects the validity of the model in the case of relative risk aversion. Only the p-uniform* does not rely on the uncorrelation assumption because here the identification is based on p-values, not on estimates and their standard errors.

The results of the nonlinear tests are shown in Table 3. All tests corroborate strong publication bias: the corrected mean coefficients of relative risk aversion are always much smaller than uncorrected means shown earlier in Table 1. But the individual results vary. The p-uniform* technique, which can be considered conceptually superior to other models in the context of risk aversion since it does not rely on the uncorrelation assumption, yields values of risk aversion below 1 for both economics and finance. The selection model yields a large estimate for finance, 11, but we have seen that in our dataset the model probably does not work well. The remaining results are more consistent and suggest relative risk aversion around 1 in economics and 2–6 in finance: with the qualification that our interpretation of the strength of publication bias is conservative because p-uniform* suggests an even stronger exaggeration. Finally, we also apply two new tests of p-hacking by Elliott *et al.* (2022) in Table C1, Appendix C. These tests also do not rely on the uncorrelation assumption, but need a huge sample and only test p-hacking without estimating the corrected risk aversion. Using these tests we reject the hypothesis of no bias in the entire sample but not in the individual subsamples of economics and finance studies.

4 Heterogeneity

Another way to relax the uncorrelation assumption is to explicitly allow for heterogeneity among the estimates of relative risk aversion. To this end we collect 30 aspects of the context in which the estimates are obtained. Using these additional variables we seek answers to three questions: Are our findings regarding publication bias robust to heterogeneity? Do some aspects of data or methods affect the reported estimates systematically? What is the literature's best guess regarding relative risk aversion in various contexts after correction for publication bias?

The variables are summarized in Table 4 and discussed in detail in the Appendix, Subsection D.1. For ease of exposition we divide them into four groups: data characteristics, specification characteristics, estimation techniques, and publication characteristics. The list of variables that control for the context in risk aversion estimation is potentially unlimited, but we do our best to account for differences that are most commonly discussed in the literature. Figure D1 shows that even with so many variables, collinearity is likely not a major issue for our analysis. Even so, we employ techniques that take collinearity into account.

Because we have so many variables, we need to use methods that account for model uncertainty. While all of the variables we collect have been implicated in the literature to potentially affect the reported risk aversion, it is unclear whether all variables indeed belong to the best model. If not, then the effects of important variables will be imprecisely estimated, perhaps drastically so. A natural solution to model uncertainty arises in the Bayesian framework as Bayesian model averaging (see Steel, 2020, for a great overview). Bayesian model averaging estimates many models that include various combinations of the explanatory variables we have collected and weights individual models by goodness of fit and parsimony. Because in our case there are too many possible models, we simplify this computationally demanding task by employing the Metropolis-Hastings algorithm of the `bms` package for R by Zeugner & Feldkircher (2015), which walks only through the most likely models. We also employ the dilution prior (George, 2010), which accounts for collinearity by adding a weight that is proportional to the determinant of the correlation matrix of the variables included in the individual model. Unfortunately Bayesian model averaging can only be combined with the linear test of publication bias, but we have shown in the previous section that the results of the linear tests are broadly consistent with more advanced nonlinear techniques.

Table 4: Definition and summary statistics of explanatory variables

Variable	Description	Mean	SD
Standard error	The standard error of the estimated coefficient of relative risk aversion.	76.65	730.63
Data characteristics			
Time span	The logarithm of the time span of the data used to estimate RRA.	3.45	0.92
Midpoint	The logarithm of the median year of the data used minus the earliest median year observed in primary studies.	3.82	0.63
Panel	= 1 if panel data are used (reference category: time series).	0.04	0.19
Cross-section	= 1 if cross-sectional data are used (reference category: time series).	0.20	0.40
Monthly	= 1 if data frequency is monthly or higher (reference category: annual).	0.25	0.43
Quarterly	= 1 if data frequency is quarterly (reference category: annual).	0.50	0.50
US	= 1 if the estimate relates to the United States (reference category: other countries).	0.74	0.44
EU	= 1 if the estimate relates to European countries (reference category: other countries).	0.11	0.31
Asia	= 1 if the estimate relates to developed Asian countries (reference category: other countries).	0.03	0.18
Developing	= 1 if the estimate relates to developing countries, including China (reference category: other countries).	0.06	0.24
Specification characteristics			
Epstein-Zin	= 1 if preferences are of the Epstein-Zin type (the remaining estimates are derived from specifications with internal habits, expected utility with a reference level of consumption, ambiguity aversion, or disappointment aversion).	0.90	0.30
Long-run risk	= 1 if estimation features long-run risks.	0.32	0.47
Fixed EIS	= 1 if the value of the elasticity of intertemporal substitution is fixed when estimating RRA.	0.25	0.43
Nonseparable durables	= 1 if the model allows for nonseparability between durable and nondurable consumption.	0.13	0.33
Total consumption	= 1 if total consumption is used instead of nondurable consumption.	0.10	0.30
Exact Euler	= 1 if the exact Euler equation is estimated instead of the log-linearized one.	0.37	0.48
Human capital	= 1 if human capital is accounted for in the estimation.	0.10	0.30
Stockholder	= 1 if the estimate relates to stockholders or wealthy households (reference category: mixed sample).	0.12	0.32
Nonstockholder	= 1 if the estimate relates to nonstockholders or poor households (reference category: mixed sample).	0.05	0.21
Estimation techniques			
Experimental	= 1 if the estimate is based on (quasi-)experimental data.	0.02	0.15
Implied	= 1 if the value of RRA is not reported explicitly but can be computed from other reported parameters.	0.12	0.32
GMM	= 1 if the generalized method of moments is used (reference category: OLS).	0.59	0.49
Simulations	= 1 if nonparametric simulation-based methods are used (reference category: OLS).	0.17	0.37
Second lag	= 1 if only second or higher lags are included among instruments.	0.16	0.36
Market return included	= 1 if market return is included among instruments.	0.32	0.47
Consumption included	= 1 if consumption is included among instruments.	0.35	0.48
Publication characteristics			
Publication year	The logarithm of the year when the study first appeared in Google Scholar minus the year when the earliest study in our dataset appeared in Google Scholar.	2.84	0.63
Top journal	= 1 if the estimate comes from a study published in the top five economics or top three finance journals.	0.30	0.46
Finance journal	= 1 if the estimate is reported in a finance journal.	0.42	0.49
Citations	The logarithm of the number of per-year citations of the study, according to Google Scholar.	1.72	1.40

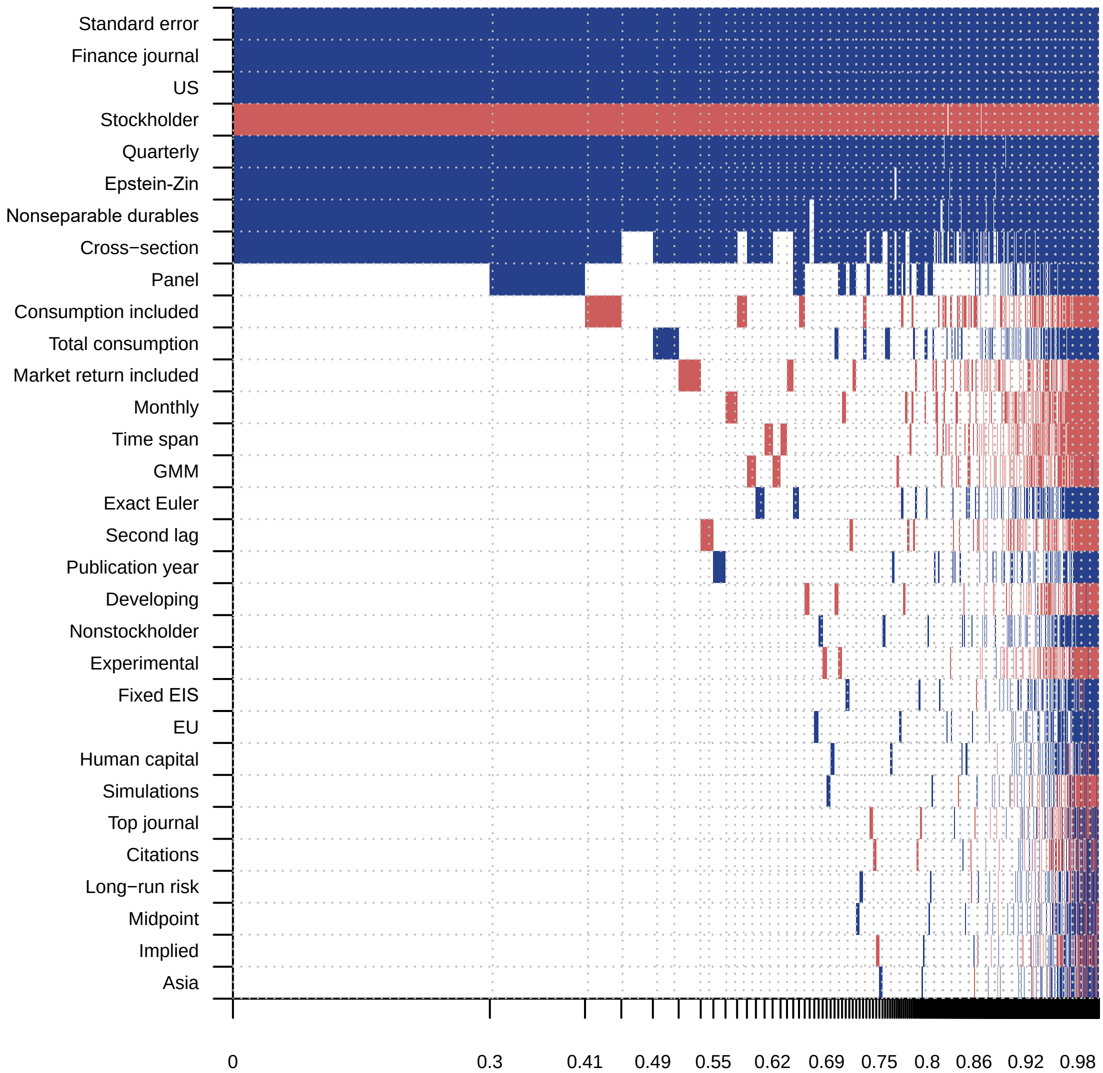
Notes: All estimates that we collect are derived from specifications that separate risk aversion from intertemporal substitution. RRA = relative risk aversion; EIS = elasticity of intertemporal substitution; GMM = general method of moments; SD = standard deviation. The table excludes the definition and summary statistics of reference categories, which are omitted from Bayesian model averaging. Regarding the variable *Finance journal*, we use the classification of the Web of Science. If in the Web of Science the journal is included in both economics and finance categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking.

The results of Bayesian model averaging are summarized graphically in Figure 4; more details are available in Table D1 and Figure D2 in Subsection D.2. The horizontal axis denotes cumulative posterior model probabilities: the weights received by each model. The most informative individual models, denoted by columns, therefore, are depicted on the left. Variables are sorted by posterior inclusion probability (the sum of posterior model probabilities of all models in which the variable is included) in descending order. This ordering means that the variables most useful in explaining the variation in estimated risk aversion are depicted at the top of the figure. The single most important variable is the standard error, which corroborates our previous results concerning publication bias. In total, there are 8 variables with posterior inclusion probability above 0.5, which means that these variables are systematically related to the published coefficients of relative risk aversion. The results of Bayesian model averaging can be sensitive to the priors used, but Figure 5 and Table D2 show that posterior inclusion probabilities do not change much when we apply alternative priors sometimes used in the literature.

The numerical results of Bayesian model averaging are reported in the left-hand part of Table 5. The right-hand part shows a simple frequentist robustness check, in which we run ordinary least squares using only the variables with posterior inclusion probability above 0.5 in Bayesian model averaging. The robustness check is broadly consistent with the results of Bayesian model averaging, but finds borderline statistical significance for several of the variables. The point estimates, however, are similar and suggest large effects of these characteristics. We find that, even if we control for estimation context, finance journals tend to report coefficients of relative risk aversion substantially larger than economics journals: by about 6. Another intuitive result is that stockholders are less risk-averse than nonstockholders. Again the difference in relative risk aversion is about 6. Next, we find that the results are driven by data and estimation characteristics: data dimension (cross-section vs. time series vs. panel data), data frequency (monthly vs. quarterly vs. annual), regional coverage (US vs. other countries), the specification of the utility function (Epstein-Zin vs. other approaches), and treatment of durables (separability vs. nonseparability). The heterogeneity results are described in more detail in Subsection D.2.

Finally, we compute relative risk aversion implied by the literature for different settings after correction for publication bias and other potential biases. For this exercise we use the results of Bayesian model averaging and compute the corresponding fitted values. To do so, we need

Figure 4: Model inclusion in Bayesian model averaging



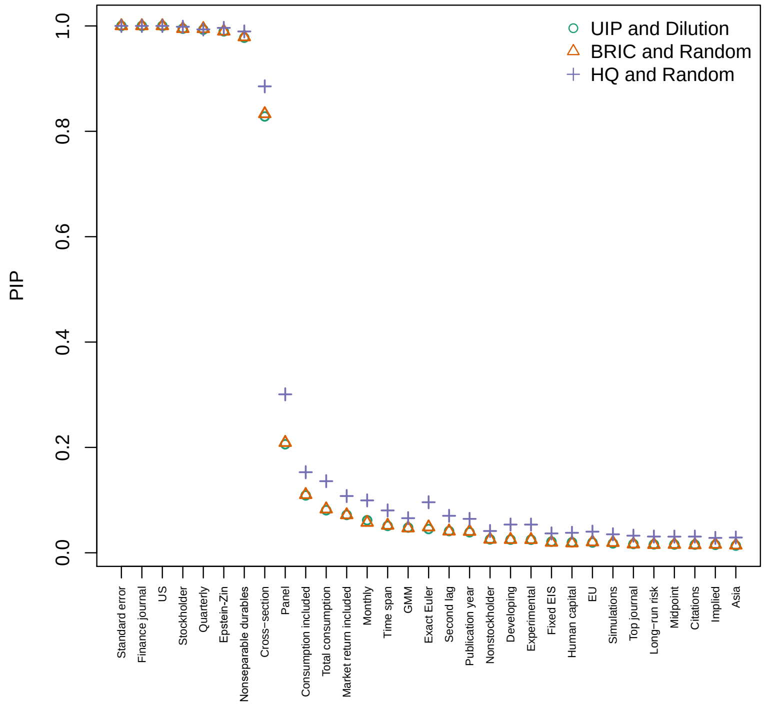
Notes: The response variable is the reported estimate of relative risk aversion; all estimates that we collect are derived from specifications that separate risk aversion from intertemporal substitution. The columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes cumulative posterior model probabilities. The estimation is based on the agnostic unit information prior recommended by Eicher *et al.* (2011) and the dilution prior suggested by George (2010), which takes collinearity into account. Blue color (darker in grayscale) = the variable has a positive estimated sign. Red color (lighter in grayscale) = the variable has a negative estimated sign. No color = the variable is excluded from the given model. Table 4 presents a detailed description of the variables. The numerical results are reported in Table 5.

Table 5: Why do estimates of risk aversion vary?

	Bayesian model averaging			Frequentist check (OLS)		
Variable	Post. Mean	Post. SD	PIP	Coeff.	SE	p-val.
Constant	-8.841	N.A.	1.000	-9.050	3.108	0.004
Standard error	0.980	0.035	1.000	0.980	0.070	0.000
Data characteristics						
Time span	-0.041	0.217	0.049			
Midpoint	0.003	0.084	0.014			
Panel	1.037	2.229	0.207			
Cross-section	3.424	1.866	0.833	4.098	1.841	0.026
Monthly	-0.117	0.569	0.057			
Quarterly	4.469	0.954	0.995	4.394	1.679	0.009
US	6.064	1.004	1.000	5.924	1.498	0.000
EU	0.024	0.270	0.019			
Asia	0.004	0.245	0.013			
Developing	-0.055	0.491	0.024			
Specification characteristics						
Epstein-Zin	5.488	1.370	0.991	5.592	3.390	0.099
Long-run risk	0.004	0.131	0.014			
Fixed EIS	0.024	0.276	0.020			
Nonseparable durables	4.834	1.372	0.979	5.008	3.354	0.135
Total consumption	0.207	0.801	0.080			
Exact Euler	0.063	0.345	0.045			
Human capital	0.018	0.239	0.017			
Stockholder	-5.768	1.341	0.995	-5.769	3.659	0.115
Nonstockholder	0.053	0.482	0.024			
Estimation techniques						
Experimental	-0.062	0.593	0.022			
Implied	-0.001	0.150	0.014			
GMM	-0.075	0.414	0.046			
Simulations	-0.005	0.231	0.017			
Second lag	-0.066	0.389	0.041			
Market return included	-0.116	0.486	0.070			
Consumption included	-0.195	0.628	0.108			
Publication characteristics						
Publication year	0.037	0.230	0.038			
Top journal	0.001	0.143	0.015			
Finance journal	6.358	0.949	1.000	6.297	1.565	0.000
Citations	-0.001	0.045	0.015			
Observations	1,021			1,021		
Studies	92			92		

Notes: The response variable is the reported estimate of relative risk aversion; all estimates that we collect are derived from specifications that separate risk aversion from intertemporal substitution. SD = standard deviation; PIP = posterior inclusion probability; SE = standard error. The left-hand panel applies BMA based on the unit information g-prior and the dilution model prior (Eicher *et al.* 2011; George 2010). See Zeugner & Feldkircher (2015) for a detailed description of the priors. The right-hand panel reports a frequentist check using ordinary least squares, which includes variables with PIPs above 0.5 in BMA. Standard errors in the frequentist check are clustered at the study level. Table 4 presents a detailed description of the variables.

Figure 5: Posterior inclusion probabilities across different prior settings



Notes: UIP = unit information prior; the prior has the same weight as one observation of data. Dilution model prior = the prior weight of each model is proportional to the determinant of the correlation matrix. BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior for the model space, which means that each model size has equal prior probability (Fernandez *et al.*, 2001). The HQ prior asymptotically mimics the Hannan-Quinn criterion. See Zeugner & Feldkircher (2015) for a detailed description of the priors.

to choose a specific value for each variable, which is inevitably subjective. We plug zero for the standard error to account for publication bias. To give more weight to studies with larger datasets and newer data, we plug in sample maxima for the time span and midpoint of data. We prefer if panel data, exact Euler equation, and Epstein-Zin preferences are used, first lags are not included among instruments (because of potential problems with time aggregation), the elasticity of intertemporal substitution is not fixed, and the estimate is not obtained via simulation. We also prefer if the study was published recently, in a top journal, and is frequently cited. All other variables are set to their sample means. Table 6 shows that such an exercise yields imprecise results, but the point estimate for economics is still around 1, consistent with our previous results. The implied estimate for finance is somewhat larger, around 7, but not far from the 2–6 range discussed in the previous section. The implied values of risk aversion for different contexts shown in Table 6 lie between 1 and 7.

Table 6: Implied risk aversion

	Mean	95% cred. int.
Overall best practice	3.73	[-7.36, 14.82]
Economics	1.24	[-10.25, 12.73]
Finance	7.16	[-3.85, 18.17]
US	5.81	[-5.64, 17.26]
EU	1.57	[-7.07, 10.22]
Stockholder	1.49	[-6.80, 9.79]
GMM	3.79	[-6.94, 14.52]
Quarterly data	6.33	[-4.61, 17.27]

Notes: The table uses benchmark BMA results to compute relative risk aversion conditional on selected aspects of data, methodology, and publication (see text for details). That is, the table attempts to answer the question what the mean risk aversion would look like if the literature was free of publication bias and all studies used the same strategy as the one we prefer. The 95% credible intervals are reported in parentheses.

5 Conclusion

We provide the first meta-analysis of the literature estimating relative risk aversion. We focus on studies that use the consumption Euler equation and that break the link (present with power utility) between risk aversion and intertemporal substitution. This means that we mostly focus on estimates that employ Epstein-Zin preferences. The literature provides 1,021 estimates reported in 92 studies; we also collect 446 calibrations of relative risk aversion from 200 studies. Our results suggest a wedge between estimates and calibrations: calibrations are often larger than estimates, especially in the economics literature. The wedge increases substantially when we correct the estimates of risk aversion for publication selection bias: the corrected mean estimate is 1 for economics and 2–7 for finance, which are the values we recommend for calibration. The finding for economics is consistent with Chetty (2006), who argues that data on labor supply behavior impose an upper bound of 2 on relative risk aversion. Our results also suggest that the estimates are systematically correlated with the context in which they are obtained, such as data dimension (time-series vs. cross-section vs. panel data), data frequency (monthly vs. quarterly vs. annual), country coverage (US vs. Europe), general form of the utility function (Epstein-Zin vs. other approaches), treatment of durables (separability vs. nonseparability), and whether or not the researcher focuses on stockholders.

Three qualifications are in order. First, our classification of studies into economics and finance fields is crude and follows the classification of journals in which the studies are published.

Two studies may use a similar strategy to identify relative risk aversion, but one can be published in an economics journal, the other in a finance journal. The advantage of the journal-based classification is its clarity and parsimony; a rule based on methodology or data would also inevitably be more subjective. The sharp difference between the distribution of estimates in economics and finance according to our definition suggests that the classification we use is informative. Second, most meta-analysis methods that we use invoke the classical assumption that in the absence of publication bias there is no correlation between estimates and standard errors. The assumption does not have to hold in the risk aversion literature, because estimation approaches vary widely and some may influence both estimates and standard errors. As a partial solution we employ the p-uniform* technique, which does not need this strong assumption. The technique suggests even stronger publication bias for both economics and finance. Third, we use more than one estimate from primary studies, which violates the standard meta-analysis assumption that all estimates are independent. We partially address this problem by clustering standard errors at the study level and using wild bootstrap.

References

van Aert, R. C. & M. van Assen (2021): "Correcting for publication bias in a meta-analysis with the p-uniform* method." *Working paper*, Tilburg University & Utrecht University.

Albuquerque, R., M. Eichenbaum, V. X. Luo, & S. Rebelo (2016): "Valuation risk and asset pricing." *The Journal of Finance* **71(6)**: pp. 2861–2904.

Ampudia, M., R. Cooper, J. Le Blanc, & G. Zhu (2018): "MPC Heterogeneity in Europe: Sources and Policy Implications." *Working Paper 25082*, National Bureau of Economic Research.

Andersen, S., G. W. Harrison, M. I. Lau, & E. E. Rutström (2018): "Multiattribute utility theory, intertemporal utility, and correlation aversion." *International Economic Review* **59(2)**: pp. 537–555.

Andreasen, M. M. (2012): "An estimated DSGE model: Explaining variation in nominal term premia, real term premia, and inflation risk premia." *European Economic Review* **56(8)**: pp. 1656–1674.

Andreasen, M. M., J. Fernández-Villaverde, & J. F. Rubio-Ramírez (2018): "The pruned state-space system for non-linear DSGE models: Theory and empirical applications." *The Review of Economic Studies* **85(1)**: pp. 1–49.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–94.

Ascari, G., L. M. Magnusson, & S. Mavroeidis (2021): "Empirical evidence on the Euler equation for consumption in the US." *Journal of Monetary Economics* **117(C)**: pp. 129–152.

Attanasio, O. P. & G. Weber (1989): "Intertemporal substitution, risk aversion and the Euler equation for consumption." *The Economic Journal* **99(395)**: pp. 59–73.

Augustin, P. & R. Tédongap (2016): "Real economic shocks and sovereign credit risk." *Journal of Financial and Quantitative Analysis* **51(2)**: pp. 541–587.

Bakshi, G. S. & A. Naka (1997): "An empirical investigation of asset pricing models using Japanese stock market data." *Journal of International Money and Finance* **16(1)**: pp. 81–112.

Bansal, R., A. R. Gallant, & G. Tauchen (2007a): "Rational pessimism, rational exuberance, and asset pricing models." *The Review of Economic Studies* **74(4)**: pp. 1005–1033.

Bansal, R., D. Kiku, & A. Yaron (2007b): "Risks for the Long Run: Estimation and Inference." *Working paper*, Duke University.

Bansal, R., D. Kiku, & A. Yaron (2016): "Risks for the long run: Estimation with time aggregation." *Journal of Monetary Economics* **82(C)**: pp. 52–69.

Bansal, R. & I. Shaliastovich (2013): "A long-run risks explanation of predictability puzzles in bond and currency markets." *The Review of Financial Studies* **26(1)**: pp. 1–33.

Bansal, R., T. D. Tallarini, & A. Yaron (2008): "The return to wealth, asset pricing and the intertemporal elasticity of substitution." *Meeting papers 918*, Society for Economic Dynamics.

Bansal, R. & A. Yaron (2004): "Risks for the long run: A potential resolution of asset pricing puzzles." *The journal of Finance* **59(4)**: pp. 1481–1509.

Bednarek, Z. & P. Patel (2015): "Long-run Risk, Durable Consumption Growth and Estimation of Risk Aversion." *Working paper*, California Polytechnic State University.

Bekhtiar, K., P. Fessler, & P. Lindner (2020): "Risky assets in Europe and the US: risk vulnerability, risk aversion and economic environment." *Working paper 2270*, European Central Bank.

Biswas, A. & B. Mandal (2016): "Estimating Preference Parameters from Stock Returns Using Simulated Method of Moments." *Annals of Financial Economics* **11(01)**: p. 1650005.

Blanco-Perez, C. & A. Brodeur (2020): "Publication Bias and Editorial Statement on Negative Findings." *The Economic Journal* **130(629)**: pp. 1226–1247.

Bom, P. R. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research Synthesis Methods* **10(4)**: pp. 497–514.

Bretscher, L., A. Hsu, & A. Tamoni (2020): "Fiscal policy driven bond risk premia." *Journal of Financial Economics* **138(1)**: pp. 53–73.

Briggs, J., D. Cesarini, E. Lindqvist, & R. Östling (2021): "Windfall gains and stock market participation." *Journal of Financial Economics* **139(1)**: pp. 57–83.

Brodeur, A., N. Cook, & A. Heyes (2020): "Methods Matter: P-Hacking and Causal Inference in Economics." *American Economic Review* **110(11)**: pp. 3634–3660.

Brodeur, A., M. Le, M. Sangnier, & Y. Zylberberg (2016): "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

Brown, A. L., T. Imai, F. Vieider, & C. Camerer (2022): "Meta-Analysis of Empirical Estimates of Loss-Aversion." *Journal of Economic Literature* **(forthcoming)**.

Brown, A. L. & H. Kim (2014): "Do individuals have preferences used in macro-finance models? An experimental investigation." *Management Science* **60(4)**: pp. 939–958.

Bruns, S. B. & J. P. A. Ioannidis (2016): "p-Curve and p-Hacking in Observational Research." *PloS ONE* **11(2)**: p. e0149144.

Bufman, G. & L. Leiderman (1990): "Consumption and asset returns under non-expected utility: Some new evidence." *Economics Letters* **34(3)**: pp. 231–235.

Campbell, J. Y. (1996): "Understanding risk and return." *Journal of Political Economy* **104(2)**: pp. 298–345.

Card, D., J. Kluve, & A. Weber (2018): "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association* **16(3)**: pp. 894–931.

Card, D. & A. B. Krueger (1995): "Time-series minimum-wage studies: A meta-analysis." *American Economic Review* **85(2)**: pp. 238–243.

Carmichael, B. & L. Samson (1993): "Excess returns determination: Empirical evidence from Canada." *Journal of Economics and Business* **45(1)**: pp. 35–48.

Chen, X., J. Favilukis, & S. C. Ludvigson (2013): "An estimation of economic models with recursive preferences." *Quantitative Economics* **4(1)**: pp. 39–83.

Chetty, R. (2006): "A New Method of Estimating Risk Aversion." *American Economic Review* **96(5)**:

pp. 1821–1834.

CHIAPPORI, P. & M. PAIELLA (2011): "Relative Risk Aversion Is Constant: Evidence From Panel Data." *Journal of the European Economic Association* **9(6)**: pp. 1021–1052.

CHO, J. & Y. DOKKO (1993): "Risk aversion in the expected and the nonexpected utility functions." *Review of Quantitative Finance and Accounting* **3(4)**: pp. 421–427.

CHOI, H., S. LUGAUER, & N. C. MARK (2017): "Precautionary saving of Chinese and US households." *Journal of Money, Credit and Banking* **49(4)**: pp. 635–661.

CHRISTENSEN, G. & E. MIGUEL (2018): "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature* **56(3)**: pp. 920–980.

CHRISTENSEN, T. M. (2017): "Nonparametric stochastic discount factor decomposition." *Econometrica* **85(5)**: pp. 1501–1536.

COBLE, K. H. & J. L. LUSK (2010): "At the nexus of risk and time preferences: An experimental investigation." *Journal of Risk and Uncertainty* **41(1)**: pp. 67–79.

COLACITO, R. & M. M. CROCE (2011): "Risks for the long run and the real exchange rate." *Journal of Political Economy* **119(1)**: pp. 153–181.

COLACITO, R., M. M. CROCE, F. GAVAZZONI, & R. READY (2018): "Currency risk factors in a recursive multicountry economy." *The Journal of Finance* **73(6)**: pp. 2719–2756.

CONSTANTINIDES, G. M. (2021): "Welfare costs of idiosyncratic and aggregate consumption shocks." *Working paper 29009*, National Bureau of Economic Research.

CONSTANTINIDES, G. M. & A. GHOSH (2011): "Asset pricing tests with long-run risks in consumption growth." *The Review of Asset Pricing Studies* **1(1)**: pp. 96–136.

COOPER, R. & G. ZHU (2016): "Household finance over the life-cycle: What does education contribute?" *Review of Economic Dynamics* **20(C)**: pp. 63–89.

COX, G. & X. SHI (2022): "Simple Adaptive Size-Exact Testing for Full-vector and Subvector Inference in Moment Inequality Models." *Review of Economic Studies* **(forthcoming)**.

DAVE, C. & K. P. TSANG (2014): "Recursive preferences, learning and large deviations." *Economics Letters* **124(3)**: pp. 329–334.

DEAN, M. & P. ORTOLEVA (2019): "The empirical relationship between nonstandard economic behaviors." *Proceedings of the National Academy of Sciences* **116(33)**: pp. 16262–16267.

DELIKOURAS, S. (2017): "Where's the kink? Disappointment events in consumption growth and equilibrium asset prices." *The Review of Financial Studies* **30(8)**: pp. 2851–2889.

DELIKOURAS, S. & G. M. KORNIOTIS (2021): "Asset pricing with and without garbage: Resurrecting aggregate consumption." *Working paper*, University of Miami Business School.

DELLAVIGNA, S. & E. LINOS (2022): "RCTs to Scale: Comprehensive Evidence From Two Nudge Units." *Econometrica* **90(1)**: pp. 81–116.

DELLAVIGNA, S., D. POPE, & E. VIVALT (2019): "Predict Science to Improve Science." *Science* **366(6464)**: pp. 428–429.

DOH, T. (2013): "Long-Run Risks in the Term Structure of Interest Rates: Estimation." *Journal of Applied Econometrics* **28(3)**: pp. 478–497.

EGGER, M., G. D. SMITH, M. SCHNEIDER, & C. MINDER (1997): "Bias in meta-analysis detected by a simple, graphical test." *BMJ* **315(7109)**: pp. 629–634.

EICHER, T. S., C. PAPAGEORGIOU, & A. E. RAFTERY (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

ELLIOTT, G., N. KUDRIN, & K. WÜTHRICH (2022): "Detecting p-hacking." *Econometrica* **90(2)**: pp. 887–906.

EPSTEIN, L. G. & S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* **57(4)**: pp. 937–969.

EPSTEIN, L. G. & S. E. ZIN (1991): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis." *Journal of Political Economy* **99(2)**: pp. 263–286.

EPSTEIN, L. G. & S. E. ZIN (2001): "The independence axiom and asset returns." *Journal of Empirical Finance* **8(5)**: pp. 537–572.

Eraker, B., I. Shaliastovich, & W. Wang (2016): "Durable goods, inflation risk, and equilibrium asset prices." *The Review of Financial Studies* **29(1)**: pp. 193–231.

Faria, A., R. Ornelas, & C. Almeida (2016): "Empirical selection of optimal portfolios and its influence in the estimation of Kreps-Porteus utility function parameters." *Brazilian Review of Econometrics* **36(1)**: pp. 43–62.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Model uncertainty in cross-country growth regressions." *Journal of Applied Econometrics* **16(5)**: pp. 563–576.

Fulop, A., J. Heng, J. Li, & H. Liu (2022): "Bayesian estimation of long-run risk models using sequential Monte Carlo." *Journal of Econometrics* **228(1)**: pp. 62–84.

Fulop, A., J. Li, H. Liu, & C. Yan (2021): "Estimating and Testing Long-Run Risk Models: International Evidence." *Working paper*, University of Manchester.

Furukawa, C. (2021): "Publication bias under aggregation frictions: From communication model to new correction method." *Working paper*, MIT.

Gandelman, N. & R. Hernández-Murillo (2015): "Risk aversion at the country level." *Federal Reserve Bank of St. Louis Review* **97(1)**: pp. 53–66.

Garcia, R. & R. Luger (2012): "Risk aversion, intertemporal substitution, and the term structure of interest rates." *Journal of Applied Econometrics* **27(6)**: pp. 1013–1036.

Garcia, R., R. Luger, & E. Renault (2003): "Empirical assessment of an intertemporal option pricing model with latent variables." *Journal of Econometrics* **116(1-2)**: pp. 49–83.

Garcia, R., E. Renault, & A. Semenov (2006): "Disentangling risk aversion and intertemporal substitution through a reference level." *Finance Research Letters* **3(3)**: pp. 181–193.

Garcia, R., E. Renault, & A. Semenov (2015): "A Consumption CAPM with a Reference Level." *Working paper*, University of Montreal.

George, E. I. (2010): "Dilution priors: Compensating for model space redundancy." In "Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown," pp. 158–165. Institute of Mathematical Statistics.

Ghosh, A. & G. Roussellet (2020): "Identifying beliefs from asset prices." In "Proceedings of Paris December 2019 Finance Meeting EUROFIDAI-ESSEC," McGill University.

Gomes, F. A. R. & P. F. Ribeiro (2015): "Estimating the elasticity of intertemporal substitution taking into account the precautionary savings motive." *Journal of Macroeconomics* **45(C)**: pp. 108–123.

Gomes, J. F., L. Kogan, & M. Yogo (2009): "Durability of output and expected stock returns." *Journal of Political Economy* **117(5)**: pp. 941–986.

Goswami, G. & S. Tan (2012): "Pricing the US residential asset through the rent flow: A cross-sectional study." *Journal of Banking & Finance* **36(10)**: pp. 2742–2756.

Goswami, G., S. Tan, & M. Waisman (2014): "Understanding the cross-section of the US housing bubble: The roles of lending, transaction costs, and rent growth." *Journal of Financial Stability* **15(C)**: pp. 76–90.

Grammig, J. & E.-M. Küchlin (2018): "A two-step indirect inference approach to estimate the long-run risk asset pricing model." *Journal of Econometrics* **205(1)**: pp. 6–33.

Grammig, J. & A. Schrimpf (2009): "Asset pricing with a reference level of consumption: New evidence from the cross-section of stock returns." *Review of Financial Economics* **18(3)**: pp. 113–123.

Gu, L. & D. Huang (2013): "Consumption, money, intratemporal substitution, and cross-sectional asset returns." *Journal of Financial Research* **36(1)**: pp. 115–146.

Guo, H. (2006): "Time-varying risk premia and the cross section of stock returns." *Journal of Banking & Finance* **30(7)**: pp. 2087–2107.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption." *Journal of Political Economy* **96(2)**: pp. 339–357.

Hamori, S. (1995): "On the test of the globalization of the Japanese equity market under the Kreps-Porteus preference." *Financial Engineering and the Japanese Markets* **2(2)**: pp. 123–137.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators." *Econometrica* **50(4)**: pp. 1029–1054.

Hansen, L. P., J. C. Heaton, & N. Li (2008): "Consumption strikes back? Measuring long-run risk." *Journal of Political Economy* **116(2)**: pp. 260–302.

Hansen, L. P. & K. J. Singleton (1982): "Generalized instrumental variables estimation of nonlinear rational expectations models." *Econometrica* **50(5)**: pp. 1269–1286.

Hardouvelis, G. A., D. Kim, & T. A. Wizman (1996): "Asset pricing models with and without consumption data: An empirical evaluation." *Journal of Empirical Finance* **3(3)**: pp. 267–301.

Hasseltoft, H. (2012): "Stocks, bonds, and long-run consumption risks." *Journal of Financial and Quantitative Analysis* **47(2)**: pp. 309–332.

Havranek, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

Havranek, T., Z. Irsova, L. Laslopova, & O. Zeynalova (2022): "Publication and Attenuation Biases in Measuring Skill Substitution." *Review of Economics and Statistics* **(forthcoming)**.

Havranek, T., T. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, & R. Van Aert (2020): "Reporting guidelines for meta-analysis in economics." *Journal of Economic Surveys* **34(3)**: pp. 469–475.

Horvath, R., L. Kaszab, & A. Marsal (2021): "Equity premium and monetary policy in a model with limited asset market participation." *Economic Modelling* **95**: pp. 430–440.

Huang, L., J. Wu, & R. Zhang (2014): "Exchange risk and asset returns: A theoretical and empirical study of an open economy asset pricing model." *Emerging Markets Review* **21(C)**: pp. 96–116.

Hugonnier, J., F. Pelgrin, & P. St-Amour (2013): "Health and (other) asset holdings." *Review of Economic Studies* **80(2)**: pp. 663–710.

Hyde, S. & M. Sherif (2005a): "Consumption asset pricing models: Evidence from the UK." *The Manchester School* **73(3)**: pp. 343–363.

Hyde, S. & M. Sherif (2005b): "Don't break the habit: structural stability tests of consumption asset pricing models in the UK." *Applied Economics Letters* **12(5)**: pp. 289–296.

Imai, T., T. A. Rutter, & C. F. Camerer (2021): "Meta-Analysis of Present-Bias Estimation Using Convex Time Budgets." *The Economic Journal* **131(636)**: pp. 1788–1814.

Inkmann, J., P. Lopes, & A. Michaelides (2011): "How deep is the annuity market participation puzzle?" *The Review of Financial Studies* **24(1)**: pp. 279–319.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research." *The Economic Journal* **127(605)**: pp. F236–F265.

Issler, J. V. & N. S. Piqueira (2000): "Estimating relative risk aversion, the discount rate, and the intertemporal elasticity of substitution in consumption for Brazil using three types of utility function." *Brazilian Review of Econometrics* **20(2)**: pp. 201–239.

Jeong, D., H. Kim, & J. Y. Park (2015): "Does ambiguity matter? Estimating asset pricing models with a multiple-priors recursive utility." *Journal of Financial Economics* **115(2)**: pp. 361–382.

Jorion, P. & A. Giovannini (1993): "Time-series tests of a non-expected-utility model of asset pricing." *European Economic Review* **37(5)**: pp. 1083–1100.

Kim, D. & J. Ryou (2012): "Time preference and saving rate: Implications for global imbalances." *Journal of Money and Finance* **26(3)**: pp. 61–91.

Kim, H., H. I. Lee, J. Y. Park, & H. Yeo (2010): "Macroeconomic uncertainty and asset prices: A stochastic volatility model." In "AFA 2010 Atlanta Meetings Paper," American Finance Association.

Kocherlakota, N. R. (1990): "Disentangling the Coefficient of Relative Risk Aversion from the Elasticity of Intertemporal Substitution: An Irrelevance Result." *Journal of Finance* **45(1)**: pp. 175–190.

Kogan, L., D. Papanikolaou, & N. Stoffman (2020): "Left behind: Creative destruction, inequality, and the stock market." *Journal of Political Economy* **128(3)**: pp. 855–906.

Korniotis, G. M. (2010): "Estimating panel models with internal and external habit formation." *Journal of Business & Economic Statistics* **28(1)**: pp. 145–158.

Koskievic, J.-M. (1999): "An intertemporal consumption–leisure model with non-expected utility." *Economics Letters* **64(3)**: pp. 285–289.

Kranz, S. & P. Putz (2022): "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics: Comment." *American Economic Review* **(forthcoming)**.

Kuwahara, Y. & Y. Ohkusa (1996): "An alternative estimation method for the OCE model." *Applied Economics Letters* **3(8)**: pp. 501–503.

Kvarven, A., E. Stromland, & M. Johannesson (2020): "Comparing meta-analyses and preregistered multiple-laboratory replication projects." *Nature Human Behaviour* **4(C)**: pp. 423–434.

Kwan, Y. K., C. K. Y. Leung, & J. Dong (2015): "Comparing consumption-based asset pricing models: The case of an Asian city." *Journal of Housing Economics* **28(C)**: pp. 18–41.

Lee, W. (1997): "Covariance risk, consumption risk, and international stock market returns." *The Quarterly Review of Economics and Finance* **37(2)**: pp. 491–510.

Lence, S. H. (2000): "Using consumption and asset return data to estimate farmers' time preferences and risk attitudes." *American Journal of Agricultural Economics* **82(4)**: pp. 934–947.

Lybbert, T. J. & J. McPeak (2012): "Risk and intertemporal substitution: Livestock portfolios and off-take among Kenyan pastoralists." *Journal of Development Economics* **97(2)**: pp. 415–426.

Maio, P. F. (2018): "Does Inflation Explain Equity Risk Premia?" *Working paper*, Hanken School of Economics.

Malloy, C. J., T. J. Moskowitz, & A. Vissing-Jørgensen (2009): "Long-run stockholder consumption risk and asset returns." *The Journal of Finance* **64(6)**: pp. 2427–2479.

Mankiw, N. G. & S. P. Zeldes (1991): "The consumption of stockholders and nonstockholders." *Journal of Financial Economics* **29(1)**: pp. 97–112.

McCloskey, D. N. & S. T. Ziliak (2019): "What Quantitative Methods Should We Teach to Graduate Students? A Comment on Swann's Is Precise Econometrics an Illusion?" *The Journal of Economic Education* **50(4)**: pp. 356–361.

Mehra, R. & E. C. Prescott (1985): "The equity premium: A puzzle." *Journal of Monetary Economics* **15(2)**: pp. 145–161.

Meissner, T. & P. Pfeiffer (2022): "Measuring preferences over the temporal resolution of consumption uncertainty." *Journal of Economic Theory* **200(C)**: p. 105379.

Neisser, C. (2021): "The Elasticity of Taxable Income: A Meta-Regression Analysis." *Economic Journal* **131(640)**: pp. 3365–3391.

Normandin, M. & P. St-Amour (1998): "Substitution, risk aversion, taste shocks and equity premia." *Journal of Applied Econometrics* **13(3)**: pp. 265–281.

O'Donoghue, T. & J. Somerville (2018): "Modeling Risk Aversion in Economics." *Journal of Economic Perspectives* **32(2)**: pp. 91–114.

Olken, B. A. (2015): "Promises and Perils of Pre-analysis Plans." *Journal of Economic Perspectives* **29(3)**: pp. 61–80.

Pommeret, A. & A. Epaulard (2001): "Agents' Preferences, the Equity Premium, and the Consumption-Saving Trade-Off: An Application to French Data." *IMF Working Papers 2001/117*, International Monetary Fund.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* **92(437)**: pp. 179–191.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, & M. D. Webb (2019): "Fast and Wild: Bootstrap Inference in Stata Using Boottest." *The Stata Journal* **19(1)**: pp. 4–60.

Ruge-Murcia, F. (2017): "Skewness risk and bond prices." *Journal of Applied Econometrics* **32(2)**: pp. 379–400.

Samson, L. & M. Armstrong (2007): "Preferences and observed risk premia: an empirical analysis." *Applied Economics Letters* **14(6)**: pp. 435–439.

Schildberg-Hörisch, H. (2018): "Are Risk Preferences Stable?" *Journal of Economic Perspectives* **32(2)**: pp. 135–154.

Schwartz, E. & W. N. Torous (1999): "Can we disentangle risk aversion from intertemporal substitution in consumption?" *Finance working paper 25-99*, UCLA.

Semenov, A. (2003): "An Empirical Assessment of a Consumption CAPM with a Reference Level under

Incomplete Consumption Insurance." *Working paper 2003-5*, York University.

Smith, D. C. (1999): "Finite sample properties of tests of the Epstein–Zin asset pricing model." *Journal of Econometrics* **93(1)**: pp. 113–148.

Sönksen, J. & J. Grammig (2021): "Empirical asset pricing with multi-period disaster risk: A simulation-based approach." *Journal of Econometrics* **222(1)**: pp. 805–832.

Stanley, T. D. (2001): "Wheat from chaff: Meta-analysis as quantitative literature review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Stanley, T. D. & H. Doucouliagos (2010): "Picture this: A simple graph that reveals much ado about research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

Stanley, T. D. & H. Doucouliagos (2014): "Meta-regression approximations to reduce publication selection bias." *Research Synthesis Methods* **5(1)**: pp. 60–78.

Stanley, T. D. & H. Doucouliagos (2015): "Neither fixed nor random: weighted least squares meta-analysis." *Statistics in Medicine* **34(13)**: pp. 2116–2127.

Stanley, T. D., H. Doucouliagos, & J. P. A. Ioannidis (2022): "Retrospective median power, false positive meta-analysis and large-scale replication." *Research Synthesis Methods* **13(1)**: pp. 88–108.

Stanley, T. D., H. Doucouliagos, J. P. A. Ioannidis, & E. C. Carter (2021): "Detecting publication selection bias through excess statistical significance." *Research Synthesis Methods* **12(6)**: pp. 776–795.

Steel, M. F. J. (2020): "Model Averaging and its Use in Economics." *Journal of Economic Literature* **58(3)**: pp. 644–719.

Stock, J. H. & J. H. Wright (2000): "GMM with weak identification." *Econometrica* **68(5)**: pp. 1055–1096.

Thimme, J. & C. Völkert (2015): "Ambiguity in the cross-section of expected returns: An empirical assessment." *Journal of Business & Economic Statistics* **33(3)**: pp. 418–429.

Ugur, M., S. A. Churchill, & H. M. Luong (2020): "What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power." *Research Policy* **49(1)**: p. 103866.

Uhlig, H. (2012): "Economics and reality." *Journal of Macroeconomics* **34(1)**: pp. 29–41.

Van Binsbergen, J. H., J. Fernández-Villaverde, R. S. Koijen, & J. Rubio-Ramírez (2012): "The term structure of interest rates in a DSGE model with recursive preferences." *Journal of Monetary Economics* **59(7)**: pp. 634–648.

Vissing-Jørgensen, A. & O. P. Attanasio (2003): "Stock-market participation, intertemporal substitution, and risk-aversion." *American Economic Review* **93(2)**: pp. 383–391.

Weber, C. E. (2000): "Rule-of-thumb consumption, intertemporal substitution, and risk aversion." *Journal of Business & Economic Statistics* **18(4)**: pp. 497–502.

Weil, P. (1989): "The equity premium puzzle and the risk-free rate puzzle." *Journal of Monetary Economics* **24(3)**: pp. 401–421.

Xu-Song, X., M. Li-li, & W. Ming (2006): "Estimation of behavior parameters based on recursive utility in asset pricing theory." In "2006 International Conference on Management Science and Engineering," pp. 282–286.

Xue, X., W. R. Reed, & A. Menclova (2020): "Social capital and health: A meta-analysis." *Journal of Health Economics* **72(C)**: p. 102317.

Yang, W. (2011): "Long-run risk in durable consumption." *Journal of Financial Economics* **102(1)**: pp. 45–61.

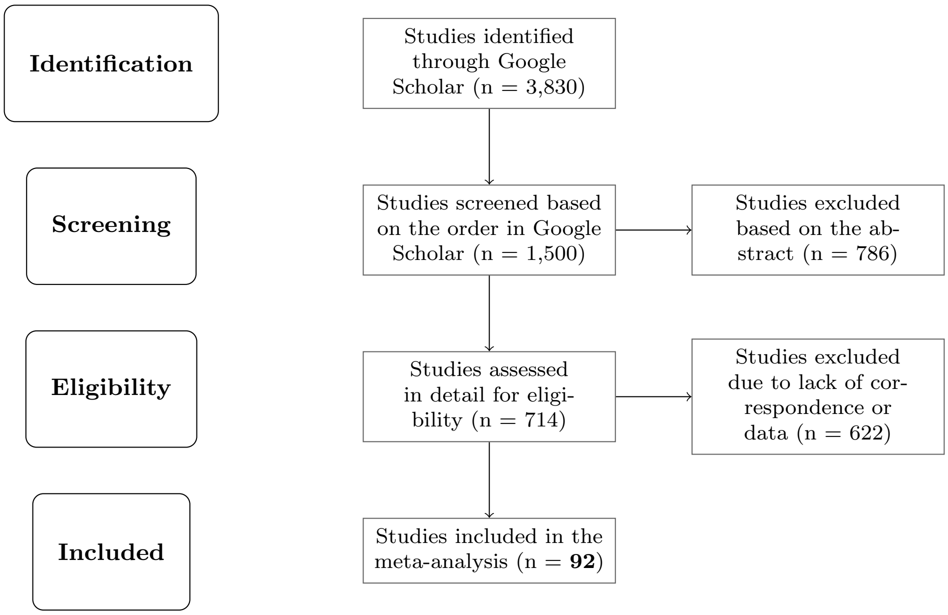
Yogo, M. (2006): "A consumption-based explanation of expected stock returns." *The Journal of Finance* **61(2)**: pp. 539–580.

Zeugner, S. & M. Feldkircher (2015): "Bayesian model averaging employing fixed and flexible priors: The BMS package for R." *Journal of Statistical Software* **68(4)**: pp. 1–37.

Zhang, R., T. J. Brennan, & A. W. Loc (2014): "The origin of risk aversion." *Proceedings of the National Academy of Sciences* **111(50)**: pp. 17777–17782.

A Details of Literature Search

Figure A1: PRISMA flow diagram



Notes: We use the following query in Google Scholar: `‘‘relative risk aversion’’ AND estimate AND (‘‘recursive utility’’ OR Epstein-Zin)`. Note that Google Scholar provides fulltext search, not only the search of the title, abstract and keywords; consequently, our query is very general. For the dataset of calibrations we use the same query but replace `estimate` with `calibration`; here we inspect the studies by the order in which they are returned by Google Scholar and stop once we reach 200 usable calibration studies. The search for both estimates and calibrations was terminated on May 16, 2022. The list of the 92 studies included in the meta-analysis is available in Table B1; the list of calibration studies is available in the online appendix at `meta-analysis.cz/risk`. All estimates and calibrations in our sample separate risk aversion from intertemporal substitution. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses. More details on PRISMA and reporting standards of meta-analysis in general are provided by Havranek *et al.* (2020).

B Estimation of Relative Risk Aversion and Additional Summary Statistics (for Online Publication)

As we have noted, there are several ways how to estimate relative risk aversion, and a useful overview is available in Zhang *et al.* (2014). Potential frameworks include human subject experiments and surveys, labor-supply behavior, deductible choices in insurance contracts, auction behavior, option prices, and contestant behavior on game shows. In this paper we focus on the consumption Euler equation, which constitutes by far the most common framework used in the fields of economics and finance.

Underlying the framework is the concept of expected utility (even though, in order to separate risk aversion from intertemporal substitution, the exact form of recursive preferences used in most studies in our sample generally does not imply expected utility). The expected utility hypothesis assumes that agents in the economy are risk-averse, meaning that their preferences are concave and exhibit a diminishing marginal return utility. Hence, the degree of risk aversion is related to the curvature of the utility function. Given a form of utility function $u(c)$ where c denotes consumption, the coefficient of relative risk aversion (RRA) is defined as

$$RRA = -\frac{u''(c)}{u'(c)}c. \tag{B1}$$

The degree of relative risk aversion can be increasing, decreasing, or constant. In economics and finance, the largest strand of the literature employs preferences with constant relative risk aversion (CRRA), i.e., isoelastic utility (power utility function), to study agents' behavior within the economy. Measuring the structural parameters associated with household preferences, such as the coefficient of relative risk aversion and the elasticity of intertemporal substitution (EIS), is important since they affect decisions on savings/investing and, consequently, asset prices in the economy. For instance, the degree of risk aversion plays a crucial role in the capital asset pricing model (CAPM) or consumption capital asset pricing model (CCAPM) since it heavily affects the investor's consumption and wealth portfolio, which ultimately alter asset prices.

Within the expected theory framework, a standard isoelastic utility function does not disentangle the attitude towards risk from intertemporal substitution as they are reciprocals of each other. The nonseparability of RRA and EIS ranks among the main critiques of the standard power utility function. The property means that when one of the parameters is large, the other has to be low, which is not necessarily realistic and consistent with empirical findings and commons sense. Hence, other forms of nonexpected utility must be considered to measure the degree of relative risk aversion isolated from the EIS. The most common solutions are recursive preferences of the type developed by Epstein & Zin (1989, 1991) and Weil (1989) (EZW hereinafter). This form of preferences constitutes a generalization of the standard power utility function in which the parameters governing EIS and RRA are separated. The separability of attitudes toward risk and intertemporal substitution makes the EZW recursive utility a suitable choice to estimate the degree of relative risk aversion. The EZW recursive utility function is a constant elasticity of substitution (CES) aggregator over the current and discounted future utility of consumption, taking the following form:

$$U_t = \left[(1-\beta)c_t^{1-\frac{1}{\psi}} + \beta\mu_t\left(U_{t+1}\right)^{1-\frac{1}{\psi}}\right]^{\frac{\psi}{\psi-1}}, \tag{B2}$$

where $0 < \beta < 1$ is the discount factor and $\psi \geq 0$ is the EIS. Households' private consumption in period t is denoted by c_t and the risk-adjusted expectation operator is given by

$$\mu_t\left(U_{t+1}\right) = \left(\mathbb{E}_t U_{t+1}^{1-\gamma}\right)^{\frac{1}{1-\gamma}}. \tag{B3}$$

Employing (B1) with some modifications, it is straightforward to show that $\gamma \geq 0$ is the coefficient of relative risk aversion for EZW preferences. The recursive utility preferences collapse to the familiar standard CRRA utility function if $\gamma = \frac{1}{\psi}$. Additionally, when $\gamma > \frac{1}{\psi}$, the EZW preferences imply that the household prefers an early resolution of uncertainty, and a late resolution of uncertainty if $\gamma < \frac{1}{\psi}$. Assuming a representative agent model with one type of consumption goods, maximizing the intertemporal utility of the household in (B2) subject to an intertemporal budget constraint results in two types of Euler equations:

$$\mathbb{E}_t\left[\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\psi}}\right)^{\eta}(R_{t+1}^M)^{\eta-1}R_{t+1}^i\right] = 1, \tag{B4}$$

and

$$\mathbb{E}_t\left[\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\psi}}\right)^{\eta}(R_{t+1}^M)^{\eta}\right] = 1, \tag{B5}$$

where $\eta = \frac{1-\gamma}{1-\frac{1}{\psi}}$, R^M_{t+1} is the gross return on the optimal portfolio, and R^i_{t+1} is the gross return on asset i between t and $t+1$. To test the separability hypothesis, it is necessary to include the following equation

$$E_t \left[\frac{\left(\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} R^M_{t+1} \right)^{\eta} - 1}{\eta} \right] = 0. \tag{B6}$$

Table B1: Studies included in the meta-analysis

Albuquerque *et al.* (2016)	Dave & Tsang (2014)	Inkmann *et al.* (2011)
Ampudia *et al.* (2018)	Delikouras (2017)	Issler & Piqueira (2000)
Andersen *et al.* (2018)	Delikouras & Korniotis (2021)	Jeong *et al.* (2015)
Andreasen (2012)	Doh (2013)	Jorion & Giovannini (1993)
Andreasen *et al.* (2018)	Pommeret & Epaulard (2001)	Kim & Ryou (2012)
Attanasio & Weber (1989)	Epstein & Zin (1991)	Kim *et al.* (2010)
Augustin & Tédongap (2016)	Epstein & Zin (2001)	Kogan *et al.* (2020)
Bakshi & Naka (1997)	Eraker *et al.* (2016)	Koskievic (1999)
Bansal & Shaliastovich (2013)	Faria *et al.* (2016)	Kuwahara & Ohkusa (1996)
Bansal *et al.* (2008)	Fulop *et al.* (2022)	Kwan *et al.* (2015)
Bansal *et al.* (2007a)	Fulop *et al.* (2021)	Lee (1997)
Bansal *et al.* (2007b)	Garcia & Luger (2012)	Lence (2000)
Bansal *et al.* (2016)	Garcia *et al.* (2003)	Lybbert & McPeak (2012)
Bednarek & Patel (2015)	Garcia *et al.* (2015)	Maio (2018)
Biswas & Mandal (2016)	Ghosh & Roussellet (2020)	Malloy *et al.* (2009)
Bretscher *et al.* (2020)	Gomes & Ribeiro (2015)	Meissner & Pfeiffer (2022)
Briggs *et al.* (2021)	Gomes *et al.* (2009)	Normandin & St-Amour (1998)
Brown & Kim (2014)	Goswami & Tan (2012)	Ruge-Murcia (2017)
Bufman & Leiderman (1990)	Goswami *et al.* (2014)	Samson & Armstrong (2007)
Campbell (1996)	Grammig & Küchlin (2018)	Schwartz & Torous (1999)
Carmichael & Samson (1993)	Grammig & Schrimpf (2009)	Semenov (2003)
Chen *et al.* (2013)	Gu & Huang (2013)	Smith (1999)
Cho & Dokko (1993)	Guo (2006)	Sönksen & Grammig (2021)
Choi *et al.* (2017)	Hamori (1995)	Stock & Wright (2000)
Christensen (2017)	Hardouvelis *et al.* (1996)	Thimme & Völkert (2015)
Coble & Lusk (2010)	Hasseltoft (2012)	Van Binsbergen *et al.* (2012)
Colacito & Croce (2011)	Horvath *et al.* (2021)	Vissing-Jørgensen & Attanasio (2003)
Colacito *et al.* (2018)	Huang *et al.* (2014)	Weber (2000)
Constantinides (2021)	Hugonnier *et al.* (2013)	Xu-Song *et al.* (2006)
Constantinides & Ghosh (2011)	Hyde & Sherif (2005a)	Yogo (2006)
Cooper & Zhu (2016)	Hyde & Sherif (2005b)	

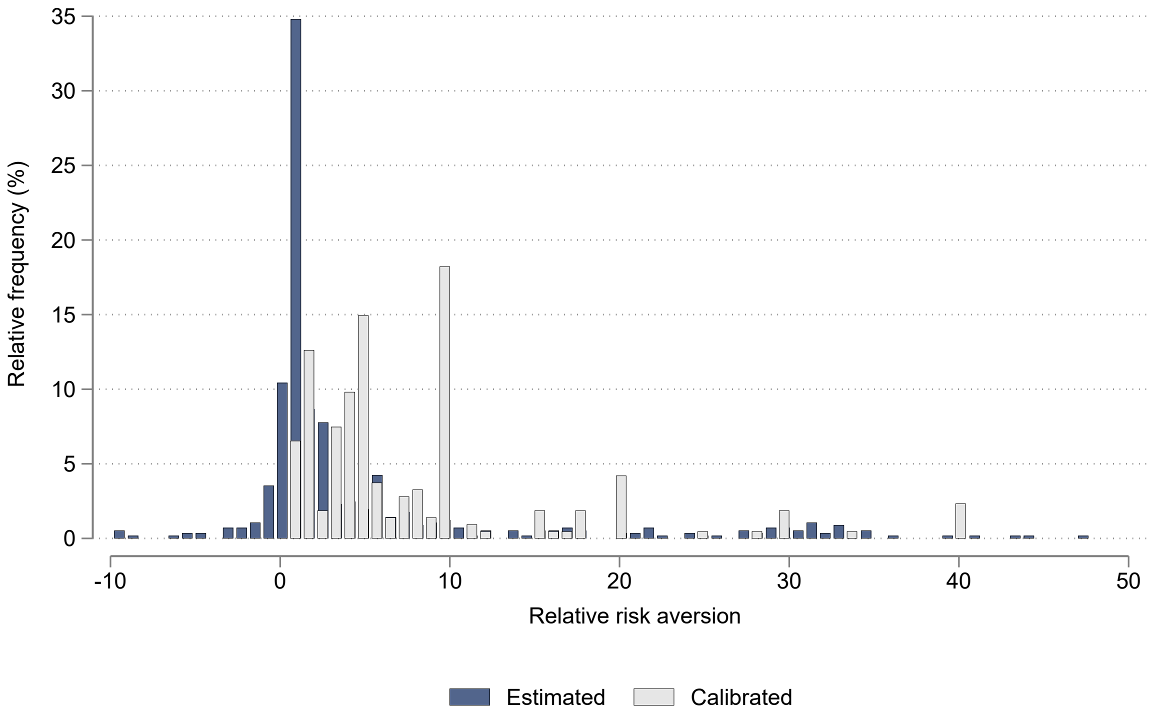
Moreover, assuming that consumption growth and asset returns are jointly log-normally distributed, (B5) takes the form of an equivalent log-linearized version. In the log-linearized version of the equation, the riskiness of an asset depends on the conditional variance of the asset's real return, the conditional covariance of the asset's real return with both consumption growth and the portfolio's real return. If the preferences reduce to the standard power utility function, i.e., $\eta = 1$, covariance risk becomes irrelevant, while in the case of EZW preferences,

Table B2: Summary statistics of benchmark calibrations

	Observations	Mean	Median	Standard deviation
All studies	200	13.13	5.93	28.62
Economics	115	16.58	5.20	36.61
Finance	85	8.47	6.00	9.14

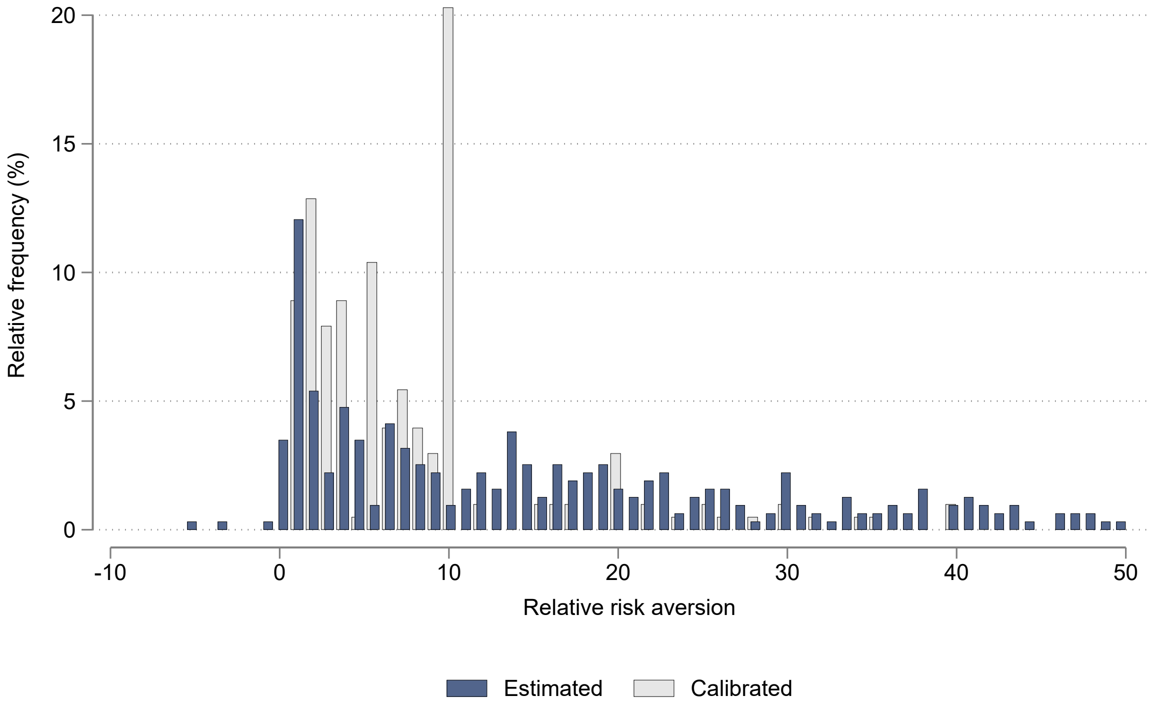
Notes: The table only considers one benchmark calibration per study (the calibration most stressed by the authors) and only includes published studies that separate risk aversion from intertemporal substitution. Studies are classified into economics and finance categories based on the journals they were published in and using the journal classification of the Web of Science. If in the Web of Science the journal is included in both categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking. Summary statistics for all calibrations from each study are reported in Table 1.

Figure B1: Estimated and calibrated relative risk aversion in economics



Notes: The figure shows histograms of i) 590 estimates or relative risk aversion collected from 58 economics studies and ii) 237 calibrations of relative risk aversion collected from 115 economics studies. In both cases we only consider studies that separate risk aversion from intertemporal substitution. Studies are classified into economics and finance categories based on the journals they were published in and using the journal classification of the Web of Science. If in the Web of Science the journal is included in both categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking. For ease of exposition, values below −10 and above 50 are excluded from the figure but included in all statistical tests. Summary statistics are available in Table 1.

Figure B2: Estimated and calibrated relative risk aversion in finance



Notes: The figure shows histograms of i) 431 estimates or relative risk aversion collected from 34 finance studies and ii) 209 calibrations of relative risk aversion collected from 85 finance studies. In both cases we only consider studies that separate risk aversion from intertemporal substitution. Studies are classified into economics and finance categories based on the journals they were published in and using the journal classification of the Web of Science. If in the Web of Science the journal is included in both categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking. For ease of exposition, values below -10 and above 50 are excluded from the figure but included in all statistical tests. Summary statistics are available in Table 1.

both covariance risk and consumption risk effectively explain assets' riskiness. Regarding the theoretical and empirical implications, Epstein & Zin (1991), Campbell (1996), and Vissing-Jørgensen & Attanasio (2003) provide more details on the log-linearized Euler equation.

The most frequently employed econometric approach to estimate the structural parameters of (B4) and (B6) or the log-linearized versions of the equations is the generalized method of moments (GMM) proposed by Hansen (1982) and Hansen & Singleton (1982). Unlike other methods in the literature, the assumptions regarding the absence of heteroskedasticity and autocorrelation of residuals do not need to hold. Moreover, the GMM estimates are consistent and asymptotically efficient, unlike ordinary least squares (OLS). To implement the technique, it is necessary to identify a set of instruments that are correlated with the included endogenous variables. Market returns, stock returns, disposable income, human capital, consumption growth, and their lagged values (one-period or more) are some of the most common instruments used in the literature (see e.g., Chen *et al.*, 2013; Faria *et al.*, 2016; Jeong *et al.*, 2015; Yogo, 2006).

Besides OLS and GMM methods, maximum likelihood estimation (MLE) is another econometric technique used to estimate the relative risk aversion parameter (e.g., Hugonnier *et al.*, 2013; Normandin & St-Amour, 1998). Conditional on distributional assumptions, this method can provide estimates with higher statistical power than those of GMM. In the case of equilibrium models, such as dynamic stochastic general equilibrium (DSGE) models, MLE-based estimations are widely used. For instance, using an MLE procedure, Van Binsbergen *et al.* (2012) estimate RRA in a DSGE model with recursive preferences. The Bayesian method of estimation is another approach widely used in the literature and, in particular, DSGE models. Among others, Bretscher *et al.* (2020) follow a Bayesian approach to estimate the relative risk aversion parameter of EZW preferences in a New-Keynesian DSGE model. The economics literature often relies on the latter two methods to deal with investors' behavior and asset returns along with the equilibrium of the whole economy at the aggregate level. On the other hand, finance literature mainly focuses on a narrower part of the economy, i.e., the behavior of investors within the asset markets, and uses extensive data on stock market returns. Hence, the finance literature mainly employs CAPM or CCAPM models (or their extensions and alternatives) that traditionally require GMM or OLS techniques to estimate the coefficient of RRA.

Additionally, one strand of literature uses simulation-based methods to estimate the degree of risk aversion along with other structural parameters. For example, the simulated method of moments that can be considered a particular case of GMM is a widely used simulation-based technique to estimate the coefficient of relative risk aversion in the Euler equation derived from recursive preferences as it tackles the problem of aggregating consumption over time (see e.g., Albuquerque *et al.*, 2016). Moerover, the presence of internal habit formation in households' preferences can lead to a wedge between the RRA and the EIS as they are not the inverse of each other. Similar to models with recursive preferences, habit formation models employ estimation techniques such as GMM and OLS to estimate the coefficient of risk aversion. In this regard, Korniotis (2010) provides a detailed discussion on the estimation procedure regarding risk aversion in internal and external habit formation models. Other alternative models include expected utility with a reference level of consumption (Garcia *et al.*, 2006), multiple-priors recursive utility with ambiguity aversion (Jeong *et al.*, 2015), recursive preferences with smooth ambiguity aversion (Thimme & Völkert, 2015), and recursive preferences with disappointment aversion (Delikouras, 2017). Finally, a relatively limited literature estimates the RRA by combining the nonexpected utility model and (quasi) experimental methods. See Brown & Kim (2014) and Briggs *et al.* (2021) for a detailed procedure of quasi-experimental estimation of relative risk aversion in the presence of recursive preferences.

C Extensions and Tests of Publication Bias Models (for Online Publication)

Table C1: Tests of p-hacking due to Elliott *et al.* (2022)

	All studies	Economics	Finance
Test for non-increasingness	0.004	0.104	1.000
Test for monotonicity and bounds	0.001	0.142	0.577
Observations ($p <= 0.15$)	755	409	346
Total observations	1,021	590	431

Notes: The table shows p-values for each test; the null hypothesis is no p-hacking. The techniques rely on the conditional chi-squared test of Cox & Shi (2022). The first technique is a histogram-based test for non-increasingness of the p-curve, the second technique is a histogram-based test for 2-monotonicity and bounds on the p-curve and the first two derivatives.

Table C2: Specification test for the Andrews & Kasy (2019) model

	All studies	Economics	Finance
Correlation	0.606	0.517	0.530
	[0.552, 0.656]	[0.434, 0.593]	[0.413, 0.643]

Notes: Following Kranz & Putz (2022), the table shows the correlation coefficient between the logarithm of the absolute value of the estimated inverse elasticity and the logarithm of the corresponding standard error, weighted by the inverse publication probability estimated by the Andrews & Kasy (2019) model. If the assumptions of the model hold, the correlation is zero. Bootstrapped 95% confidence interval in parentheses.

Table C3: Regressing estimates on standard errors when $p < 0.005$

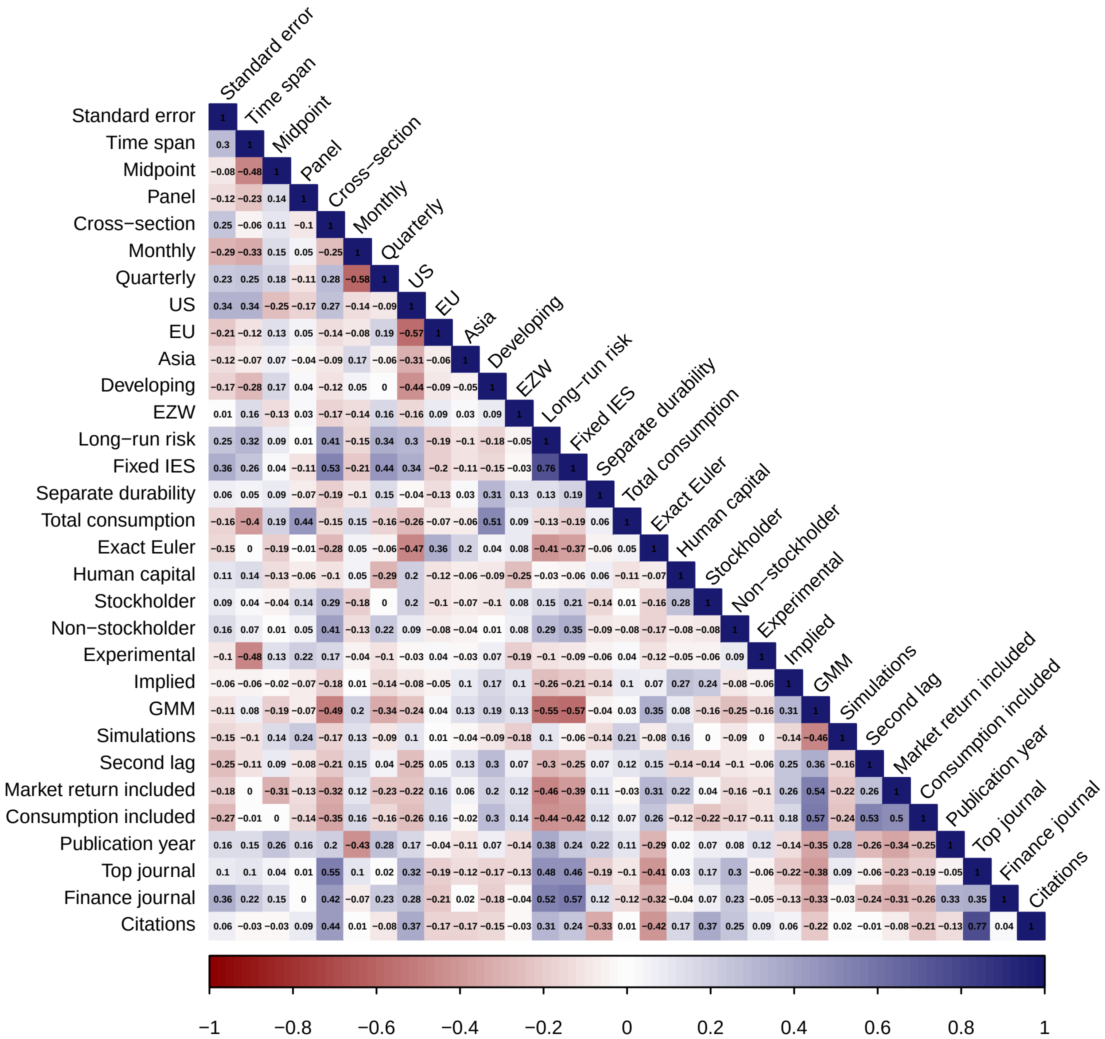
	All studies	Economics literature	Finance literature	Top journals	Implied estimate	Experimental study
Standard error	3.646***	4.171***	3.383***	3.376***	2.871***	4.261***
	(0.181)	(0.314)	(0.171)	(0.146)	(0.0134)	(0.136)
	[3.324, 4.131]	[3.016, 5.150]	[3.098, 3.866]	[2.094, 5.116]	[0.583, 4.067]	[-7.625, 16.230]
Observations	479	300	179	156	33	18
	United States	Developing country	OLS method	GMM method	Quarterly data	Annual data
Standard error	3.570***	3.722***	3.546***	3.592***	3.544***	4.033***
	(0.178)	(0.323)	(0.235)	(0.308)	(0.192)	(0.437)
	[3.259, 4.045]	[-8.862, 4.209]	[3.121, 4.325]	[2.974, 4.548]	[3.202, 4.046]	[2.848, 5.116]
Observations	327	39	155	232	247	82

Notes: The constant is included in the all regressions but not reported in the table. Standard errors, clustered at the study level, are shown in parentheses. 95% confidence intervals from the wild bootstrap are in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Summary Statistics, Extensions, and Additional Discussion of Heterogeneity Models (for Online Publication)

D.1 Variables

Figure D1: Correlation matrix of BMA variables



Notes: The figure shows Pearson correlation coefficients for the variables described in Table 4.

Data characteristics All the variables are defined and summarized in Table 4 in the main body of the paper. The first category that we consider is a set of variables concerning different characteristics of the samples used in the primary studies. We introduce eight dummy variables accounting for differences in the data. Two variables account for the difference in the data dimension: panel and cross-sectional data. Most of the reported estimates (about 76%) in primary studies are obtained using time series data, which we use as the reference category. Moreover, we codify two variables capturing data frequency. Datasets with monthly data or higher frequencies (i.e., weekly, daily) are used for 25% of the estimates, while 50% are obtained from more conventional datasets with quarterly data. Four other dummy variables denote the geographical coverage of the reported estimates. The largest group is based on the US data, accounting for 74% of estimates. The mean estimate from the US data is 31, which is substantially higher than the mean estimate of non-US data, which equals 3. This is consistent with the stream of the literature estimating higher relative risk aversion for American households compared to other countries (Gandelman & Hernández-Murillo, 2015).

On the other hand, the second largest group of estimates, using European data, exhibits the opposite pattern. The European sample, comprising around 11% of the collected estimates, yields a mean around 3, while the mean estimate of non-EU datasets is 26. Two other dummy variables denote Asian and developing countries consisting of 3% and 6% of the estimates, respectively. In addition to the dummy variables, we define two variables capturing the time properties of the datasets. The first variable, time span, captures the period of data (in terms of years) used to estimate risk aversion. To control for a potential time trend reflecting structural changes in preferences (Chiappori & Paiella, 2011; Schildberg-Hörisch, 2018), we include the midpoint of the data as an additional explanatory variable. The earliest median year of data is 1930 in Campbell (1996), which we subtract from other studies' median years to derive a relative midpoint for each study.

Specification characteristics We codify nine dummy variables to capture different aspects of the specifications for estimating relative risk aversion. The first dummy variable denotes estimates based on the EZW recursive preferences, which are used for 90% of the estimates in our sample. The remaining 10% of the estimates are derived from other techniques that allow researchers to distinguish between risk aversion and intertemporal substitution: models with habits (Korniotis, 2010), expected utility with a reference level of consumption (Garcia *et al.*, 2006), multiple-priors recursive utility with ambiguity aversion (Jeong *et al.*, 2015), recursive preferences with smooth ambiguity aversion (Thimme & Völkert, 2015), and recursive preferences with disappointment aversion (Delikouras, 2017). Next, we define a dummy variable regarding the long-run risk (LLR) model proposed by Bansal & Yaron (2004). The LLR framework contains a representative agent consumer with recursive preferences allowing for distinguishing between the RRA and EIS. The framework's other main feature is the expected consumption growth containing a small but highly persistent long-run consumption risk.

Furthermore, the LLR framework also allows for a time-varying risk premium on assets and nonindependent and identically distributed consumption growth. Using the LLR model, Hansen *et al.* (2008) show that the long-run risk channel can explain several problematic stylized facts in asset markets. Almost one-third (32%) of the estimates in primary studies are obtained within the LLR framework. The next variable accounts for the case when the estimated coefficients of relative risk aversion are obtained when the elasticity of intertemporal substitution is fixed in the estimation process. Around 25% of coefficients in the sample are estimated in the presence of fixed EIS. Several studies document that the estimation of EIS within a model with recursive preferences is not only empirically tricky but also irrelevant to the estimated risk aversion (e.g., Constantinides & Ghosh, 2011; Malloy *et al.*, 2009). However, there is no consensus in the literature about the exact value of the EIS, as documented by Havranek (2015) and Havranek *et al.* (2015).

Around 13% of the estimates are obtained in a framework where the utility function allows for nonseparability between durables and nondurables. An extensive asset pricing literature estimates the risk aversion coefficient when only nondurable goods and services are considered for consumption. There are studies, however, documenting the importance of durable goods and two-good models in estimating risk aversion (e.g., Bednarek & Patel, 2015; Yang, 2011). Similarly, we codify a dummy variable corresponding to the use of total consumption. Furthermore, more than one-third of the reported coefficients of RRA in our sample are estimated using a nonlinear (exact) Euler equation. The log-linearization of the Euler equation requires parametric restrictions on the structural parameters and the consumption growth and asset return, resulting in different estimates from the nonlinear case. Hence, we consider the effect of linearization of the Euler equation on the estimated risk aversion by defining a dummy variable accounting for the reported estimates obtained from the exact Euler equation.

Additionally, we add a variable to control for the role of human capital in estimating the coefficient of relative risk aversion. Since the return on human capital is not observable, it is common to use returns on equity or labor income as a proxy in the literature (Campbell, 1996). Among others, Grammig & Schrimpf (2009) argue that asset pricing models augmented by human capital provide more reliable results. Slightly more than ten percent of the reported estimates are obtained using models that include human capital. Finally, two additional variables control for estimates computed exclusively for stockholders (or rich households) and nonstockholders (or poor households). Not surprisingly, as shown in Table 4, stockholders often show lower risk aversion than nonstockholders. The mean estimate of the coefficient of relative risk aversion for stockholders is almost 10, while the mean estimate for nonstockholders is more than five times larger, equal to 53. Only 5% of the estimates correspond to non-stockholders and 12% for stockholders.

Estimation techniques The next category of variables considers various methods and approaches used to estimate RRA in the literature. The first dummy variable captures (quasi) experimental approaches. The variable indicates both laboratory experiments (e.g., Meissner & Pfeiffer, 2022) and quasi-experimental (e.g., Lybbert & McPeak, 2012) studies. The mean of such estimates is about 2, significantly lower than the mean estimate of non-experimental studies (24): though there are few (quasi) experimental studies that rely on the Euler equation. Next, we define a variable corresponding to the cases where the RRA is not directly estimated but implied by estimating other parameters in the model. The implied RRA might differ from the estimated coefficients in terms of magnitude and precision. The variable thus can be a source of heterogeneity among the estimates in the literature. The implied estimates form 12% of the sample.

Regarding the econometric approach, we define two variables capturing the techniques used in the literature. First, the GMM variable denotes the estimated coefficients obtained within the GMM framework, accounting for 59% of estimates reported in the primary studies. The second variable captures simulation-based estimates. The LLR models often employ simulation-based methods such as the simulated method of moments to estimate parameters (Hasseltoft, 2012). Almost 17% of estimates in our collected sample are simulation-based. We employ the OLS estimates as the baseline category. Estimates obtained by the generalized least squares (GLS) method are also included in the baseline category. The relevance and exogeneity of instruments are essential factors affecting the reliability of estimates. We thus introduce three dummy variables to control for the instruments used in the estimation procedure. The first variable captures estimates if the second or higher lags are included among instruments, accounting for almost 16% of estimates. We also control for the fact whether market returns are included among instruments by adding a dummy variable capturing 32% of the estimates in our sample. Finally, we include a similar dummy variable regarding the presence of consumption growth among instruments (35% of the estimates).

Publication characteristics The last group of variables reflects publication differences and measures of quality not captured by the previous variables. First, since more recent studies are more likely to provide newer methods and innovations regarding both theory and data, we control for the publication year of the estimate. Second, we categorize the estimates into economics literature and finance literature. To this end, we codify a dummy variable indicating estimates from the finance literature, which comprise 42% of the collected dataset. Studies are classified into economics and finance categories based on the journals they were published in and using the journal classification of the Web of Science. If in the Web of Science the journal is included in both categories, we follow the classification of the "most similar" journal according to the Scientific Journal Ranking. If a study is unpublished (15 studies in total), we classify it based on the prevailing publications of the corresponding author. As shown in Table 1, the mean of finance estimates (45) is much higher than that of our reference category, economics literature (7.5). Finally, we control for publication in top-five economics or top-three finance journals. The estimates from top journals account for 30% of the estimates reported in the primary studies. We also consider the number of citations to be a proxy for the ex-post quality of a publication and introduce a variable reflecting the number of per-year citations of each study.

D.2 Results

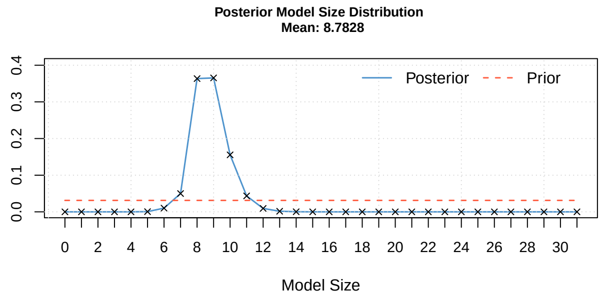
Figure 4 in the main body of the paper illustrates the results of Bayesian model averaging. The horizontal axis denotes the cumulative posterior model probabilities, and each column corresponds to one regression model. The explanatory variables are sorted by their posterior inclusion probabilities in descending order. The blue color (darker in grayscale) and red color (lighter in grayscale) denote the positive posterior mean and negative posterior mean, respectively. A

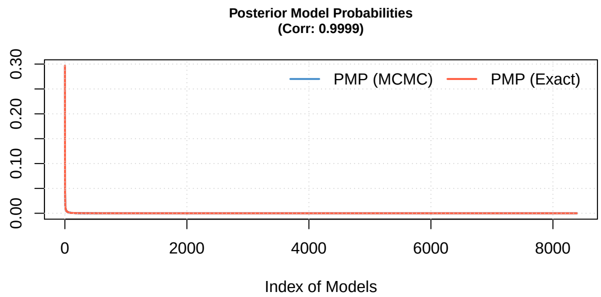
Table D1: Summary of the benchmark BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
8.794	$2 \cdot 10^6$	$1 \cdot 10^6$	2.654 mins	229,513
Modelspace	*Models visited*	*Topmodels*	*Corr PMP*	*No. Obs.*
$2.1 \cdot 10^9$	0.0011%	100	0.999	1,021
Model prior	*g-prior*	*Shrinkage-stats*		
Uniform/ 15.5	UIP	Av=0.999		

Notes: The results of this BMA specification are reported in Table 5. We account for collinearity among explanatory variables by employing the dilution prior suggested by George (2010); we also use the information prior recommended by Eicher *et al.* (2011). See Zeugner & Feldkircher (2015) for a detailed description of the priors.

Figure D2: Model size and convergence for the benchmark BMA model





Notes: The figure illustrates the posterior model size distribution and the posterior model probabilities of the BMA exercise reported in Table 5.

Table D2: Results for alternative BMA priors

	BRIC g-prior			HQ g-prior		
Variable	Post. Mean	Post. SD	PIP	Post. Mean	Post. SD	PIP
Constant	-8.837	N.A.	1.000	-8.844	N.A.	1.000
Standard error	0.980	0.035	1.000	0.977	0.036	1.000
Data characteristics						
Time span	-0.044	0.225	0.052	-0.066	0.275	0.080
Midpoint	0.004	0.091	0.016	0.006	0.127	0.031
Panel	1.044	2.234	0.209	1.484	2.534	0.301
Cross-section	3.422	1.867	0.833	3.641	1.727	0.885
Monthly	-0.121	0.581	0.057	-0.200	0.724	0.099
Quarterly	4.467	0.959	0.994	4.378	0.981	0.993
US	6.068	1.007	1.000	6.108	1.052	1.000
EU	0.026	0.283	0.020	0.053	0.404	0.040
Asia	0.005	0.259	0.014	0.012	0.377	0.029
Developing	-0.055	0.495	0.025	-0.129	0.754	0.054
Specification characteristics						
Epstein-Zin	5.484	1.383	0.990	5.573	1.313	0.996
Long-run risk	0.005	0.139	0.015	0.006	0.198	0.031
Fixed EIS	0.025	0.290	0.020	0.038	0.350	0.037
Nonseparable durables	4.834	1.376	0.979	4.952	1.297	0.990
Total consumption	0.214	0.813	0.083	0.352	1.035	0.136
Exact Euler	0.068	0.360	0.049	0.142	0.518	0.096
Human capital	0.019	0.247	0.018	0.045	0.368	0.038
Stockholder	-5.767	1.350	0.995	-5.924	1.307	0.998
Nonstockholder	0.056	0.497	0.025	0.073	0.561	0.041
Estimation techniques						
Experimental	-0.071	0.640	0.025	-0.178	1.017	0.054
Implied	-0.001	0.162	0.016	0.002	0.218	0.028
GMM	-0.075	0.415	0.047	-0.093	0.461	0.066
Simulations	-0.007	0.249	0.019	-0.017	0.345	0.035
Second lag	-0.066	0.389	0.041	-0.107	0.488	0.070
Market return included	-0.119	0.492	0.072	-0.169	0.572	0.108
Consumption included	-0.199	0.633	0.110	-0.265	0.712	0.153
Publication characteristics						
Publication year	0.039	0.236	0.040	0.059	0.288	0.064
Top journal	0.001	0.151	0.016	0.001	0.217	0.033
Finance journal	6.358	0.949	1.000	6.251	0.938	1.000
Citations	-0.001	0.047	0.015	-0.002	0.068	0.031
Observations	1,021			1,021		
Studies	92			92		

Notes: The response variable is estimated relative risk aversion. SD = standard deviation, PIP = Posterior inclusion probability. The left-hand panel applies BMA based on the BRIC g-prior (the benchmark g-prior for parameters with the beta-binomial model prior). The right-hand panel reports the results of BMA based on HQ g-prior, which asymptotically mimics the Hannan-Quinn criterion. See Zeugner & Feldkircher (2015) for a detailed description of the priors. Table 4 presents a detailed description of all the variables.

blank cell means that the variable is not included in the model. The results indicate that there are eight explanatory variables with the highest values of PIP that are likely systematically effective in explaining the size of the estimated coefficient of relative risk aversion reported in primary studies.

Table 5 in the main body of the paper presents the corresponding numerical results. The left panel presents BMA results for each explanatory variable by reporting posterior mean, posterior inclusion probability, and posterior standard deviation. Apart from the intercept, there are three *decisive* (according to the Raftery *et al.*, 1997, classification) variables with PIP equal to 1 (standard error, US data, and finance journal). Four other variables have PIPs between 0.95 and 0.99 (quarterly data, stockholder, EZW preferences, and separate durability). We label these coefficients as variables with a *strong* impact. Finally, one *substantial* explanatory variable has a PIP between 0.75 and 0.95 (cross-sectional data). Additionally, Table 5 reports the results of the frequentist check (OLS) in the right-hand panel, including the explanatory variables with PIP larger than 0.5. The results reported in both panels are consistent since the estimated coefficients exhibit similar signs and magnitude. However, two variables estimated by OLS are marginally statistically insignificant.

Data characteristics Our findings indicate the importance of three decisive variables among data characteristics affecting the size of the estimates. First, studies based on US data tend to report higher estimates than those of other countries. The empirical literature shows contradicting results regarding cross-country heterogeneity in risk aversion. Our BMA results are consistent with the stream of the literature indicating a higher risk aversion for the United States. Gandelman & Hernández-Murillo (2015) show that the United States has a relatively high degree of risk aversion among developed countries. On the other hand, a fraction of studies find the share of American households holding risky assets is higher than their counterparts in other countries, and this implies a lower degree of risk aversion in the United States (Bekhtiar *et al.* 2020).

Second, our results suggest that estimates based on cross-sectional data tend to be typically larger than the estimates obtained from time series or longitudinal data. This result is consistent with the strand of the literature concerning the cross-section of stock returns that requires a higher degree of risk aversion to reconcile aggregate consumption and market returns (see e.g., Grammig & Schrimpf, 2009; Malloy *et al.*, 2009). Significant cross-sectional variations in excess returns conflate the relationship of assets and consumption risk, which results in larger estimates of structural parameters such as the coefficient of RRA. Third, BMA results indicate that studies employing quarterly data tend to report larger estimates of relative risk aversion. On the other hand, the variable denoting frequencies higher than quarterly data, i.e., monthly frequency data, is not an insignificant explanatory variable in all BMA settings. In addition, our results suggest that the other data characteristics are not systematically correlated with the magnitude of the coefficient of relative risk aversion.

Specification characteristics Our results suggest that differences in assumed preferences may have a systematic effect on the size of the estimate. Studies that employ Epstein-Zin-Weil preferences report a higher degree of risk aversion on average than those with other types of preferences, e.g., internal habit formation model. Furthermore, we find that allowing for nonseparability of durables in the utility function is associated positively with larger reported estimates. A linear combination of the discounted future nondurable and durable consumption growth determines these models' expected asset log returns. For instance, Yogo (2006) and Bednarek & Patel (2015) show that durable consumption growth plays a significant role in the pricing of stock returns, and a higher share of durable consumption in the total expenditure will result in larger estimates of relative risk aversion. Similarly, Yang (2011) finds that since both equity premium and the stock return volatility change linearly with the share of durable goods, an increase in the risk aversion coefficient can explain the increase in the premium due to the presence of durable goods in the model.

In addition, we find that stockholders are systematically less risk-averse compared to the general population. This finding aligns with the economic theory intuition that participating in stock markets indicates a lower risk aversion, while non-stockholders show a higher level of risk aversion that prevents them from holding risky assets. There is an extensive literature documenting results similar to our BMA results. Using the 17 years of data from PSID, Mankiw & Zeldes (1991) document that the implied coefficient of relative risk aversion based on stockholder consumption is one-third of those of all families in the US. Similarly, using the EZW preferences, Malloy *et al.* (2009) find that the risk aversion coefficient is, in general, lower for the stockholders and decreases with the level of wealth of stockholders. Their structural estimates for the stockholders and the wealthiest third of stockholders are 15 and 10, respectively. We do not find evidence that the estimates obtained within the LLR model or a nonlinear Euler equation are systematically different from the rest of the estimates. Similarly, BMA results do not show that a fixed EIS and total consumption or human capital in the estimated model systematically affect the size of reported estimates.

Estimation techniques All variables related to the estimation approaches are negatively associated with the magnitude of reported estimates. However, the posterior mean for most of them is barely different from zero. More importantly, BMA results show that none of them is systematically important in determining the size of the coefficient of relative risk aversion. Among the variables in this category, only the variable reflecting instrumented consumption growth exhibits a PIP larger than 0.10, while the rest have PIPs between 0.01 and 0.07. These results remain the same also when we employ alternative BMA priors (Table D2).

Publication characteristics Regarding the variables controlling for the quality of publications, we do not find evidence that publication year, publication in a top-five and top-three journal, or the number of citations are systematically effective in explaining the size of the reported estimates. In contrast, we confirm our previous observation that the finance literature tends to report higher estimates of RRA compared to the economics literature. BMA results indicate that finance estimates are larger than those reported in the economics literature by 6.4 on average. One explanation might be the impact of the influential studies in the finance literature. There are high-quality publications widely cited within the finance literature reporting huge estimates (e.g., Yogo, 2006; Malloy *et al.*, 2009). Such studies become benchmark studies that other researchers follow, resulting in larger estimates of the coefficient of RRA in the field.