Individual Discount Rates: A Meta-Analysis of Experimental Evidence*

Jindrich Matousek[a], Tomas Havranek[a,b], and Zuzana Irsova[a]

[a]Charles University, Prague
[b]Czech National Bank

April 2, 2019

Abstract

A key parameter estimated by lab and field experiments in economics is the individual discount rate—and the results vary widely. We examine the extent to which this variance can be attributed to observable differences in methods, subject pools, and potential publication bias. To address the model uncertainty inherent to such an exercise, we employ Bayesian model averaging. We find occasional but widespread publication bias against unintuitive results: in consequence, the mean reported discount rate is inflated twofold. Our results suggest that estimates decrease with the time horizon, a finding consistent with hyperbolic discounting. Discount rates are similar for money and health questions, but people tend to be less patient in exotic contexts (e.g., when offered a kiss from a movie star). Africans are less patient than people from other continents. Finally, the results of lab and field experiments differ systematically, and it also matters whether the experiment relies on students or uses broader samples of the population.

Keywords: Discount rate, experiment, publication bias, meta-analysis, Bayesian model averaging

JEL Codes: D01, C83, C90

*Corresponding author: Zuzana Irsova, `zuzana.irsova@ies-prague.org`. An online appendix with data and code is available at `meta-analysis.cz/discrate`. Matousek acknowledges support from the Grant Agency of Charles University (grant #848217) and SVV project #260463, Havranek acknowledges support from the Czech Science Foundation (grant #18-02513S), and Irsova acknowledges support from Charles University (project Primus/17/HUM/16). The views expressed here are ours and not necessarily those of the Czech National Bank.

1 Introduction

Intertemporal trade-offs are key to a host of decision problems at both the private and public levels. For some of these decisions, it is appropriate to employ the market discount rate, which is easily detectable from financial time series. For others, however, we must try to recover the underlying discount rates of individuals—rates that also reflect the underlying transaction costs of borrowing money that households face (Kovacs & Larson, 2008). Policies addressing climate change, particularly those underpinned by the literature on the social cost of carbon, constitute a typical example of choices for which individual discounting of future costs and benefits plays a crucial role (Tol, 1999; Goulder & Stavins, 2002; Fujii & Karp, 2008; Anthoff *et al.*, 2009).

Individual discount rates can be either observed from existing data (such as in Lawrance, 1991; Dreyfus & Viscusi, 1995; Warner & Pleeter, 2001) or measured experimentally (Benzion *et al.*, 1989; Chapman & Elstein, 1995; Coller & Williams, 1999; Harrison *et al.*, 2010, among others). This paper focuses on the latter group: experiments. Controlled experiments provide a natural framework for exploring time discounting in both laboratory and field conditions by enabling researchers to vary the parameters in order to infer the subject's preferences. However, despite decades of work and dozens of experiments devoted to eliciting time preferences, no consensus on how to best measure discounting has emerged (Andreoni *et al.*, 2015). It is safe to say that the discount rate differs across individuals and its estimates vary a great deal throughout the literature, sometimes by orders of magnitude (Coller & Williams, 1999; Frederick *et al.*, 2002).

In this paper, we take stock of the evidence and aim to trace the differences in the reported discount rates to the design of experiments while accounting for model uncertainty. To this end, we review the literature quantitatively using meta-analysis methods. We also control for the effects of selective reporting, a phenomenon found to be widespread in economics and other fields (Doucouliagos & Stanley, 2013; Ioannidis *et al.*, 2017). Focusing on aspects related to study design, methodology, and subject pool characteristics, we collect a set of 22 explanatory variables and employ Bayesian model averaging (BMA; Raftery *et al.*, 1997) to examine which ones matter the most for the differences among the reported estimates.

Our results suggest that selective reporting (which causes publication bias) represents an important factor in the literature. Insignificant and negative estimates are discriminated against. A zero or negative discount rate, of course, makes little sense in most contexts. However, given sufficient noise in the experimental setup, we should sometimes observe insignificant or negative estimates and sometimes observe very large positive estimates. If negative estimates (which are unintuitive) are discarded but large positive estimates (for which it is difficult to determine whether they are intuitive or not) are kept, harmful publication bias arises. This outcome is paradoxical because selective reporting can be beneficial at the micro level: for an individual study, it is most likely a wise choice not to build the story around negative estimates of the discount rate. However, at the macro level, the discarding rule is asymmetrical since large estimates are typically not omitted. We find that such publication bias exaggerates the mean reported discount rate twofold, from 0.34 to 0.68.

Aside from publication bias, which manifests as a correlation of the point estimates of the discount rate with their standard errors, the differences in the reported estimates seem to be caused primarily by the experimental design of discounting tasks. We find that the delay embedded in the task presented to subjects has a strong impact on the resulting estimates, which contradicts the property of a constant discount rate in the discounted utility theory (Frederick *et al.*, 2002). This finding is consistent with hyperbolic discounting—with longer time horizons under evaluation, people seem more patient, and vice versa. Next, we partially confirm domain independence in intertemporal choice (Loewenstein *et al.*, 2003): it matters what the experimental subjects should be patient or impatient about. Participants seem to be equally patient concerning money and health choices but much less patient when it comes to more exotic contexts (such as vacations, certificates, or kisses from movie stars). We also find that Africans tend to be more impatient than people from other continents.

Our results offer three implications for economics experiments in general. First, it matters whether the experiment is conducted in the lab or in the field. Lab experiments yield systematically larger discount rates, indicating greater impatience. Second, the composition of the sample of experimental subjects (the subject pool) has a systematic impact on the results. Experiments working exclusively with students show more evidence for patience than experiments using mixed population samples. Taken together, these two results might question the external validity of some experiments. However, we also have good news for the experimental economics community: Third, it does not matter systematically for the results whether experiments use real or hypothetical rewards.

The remainder of the paper is structured as follows. Section 2 reviews the basic concepts of discounted utility models and discusses the methods of discount rate elicitation. Section 3 describes our approach to data collection and presents an overview of our dataset. Section 4 examines the extent of publication bias using meta-regression and other meta-analysis techniques. Section 5 investigates the sources of heterogeneity in the estimated discount rates using Bayesian model averaging. Section 6 concludes the paper. Supplementary statistics and diagnostics for the BMA and robustness checks to all analyses presented in the main body are available in the appendices at the end of the manuscript and online at `meta-analysis.cz/discrate`.

2 Estimating the Discount Rate

In this section, we do not attempt to provide a comprehensive review of the literature on discounting but briefly describe the basic concepts that are necessary for the understanding of our meta-analysis. For a more detailed treatment, we refer the reader to the authoritative works by Frederick *et al.* (2002), Andersen *et al.* (2014), and Cheung (2016).

The theory of intertemporal choice and discounting dates back to Irving Fisher's *Theory of Interest* (Fisher, 1930) and Paul Samuelson's *Note on Measurement of Utility*, in which he postulated the discounted utility model (Samuelson, 1937). His model was widely accepted together with its central idea of concentrating various decisions about intertemporal choice into a single parameter—the discount rate. Several modifications to the original discount function have been introduced to capture various features, such as hyperbolic (Ainslie, 1975; Mazur, 1984) or quasi-hyperbolic (Phelps & Pollak, 1968; Laibson, 1997) discounting functions.

The discounted utility model captures the time preferences of an individual—more specifically, an individual's preference for immediate utility over delayed utility, represented by her intertemporal utility function $U^t(c_t, ..., c_T)$, which can be described by the functional form presented in equation 1:

$$U^t(c_t, ..., c_T) = \sum_{k=0}^{T-t} D(k) \cdot u(c_{t+k}), \tag{1}$$

where $D(k)$ is the discount function and $u(c_{t+k})$ is a instantaneous utility function that can be interpreted as an individual's well-being in period $t + k$. The discount function $D(k)$ represents the relative weight that the individual places in period t on her well-being in period $t + k$ and encompasses parameter δ, which represents the individual's discount rate. This discount function can have different functional forms.

The standard exponential model, a well-known functional form used in the majority of practical applications, follows:

$$D^E(k) = \frac{1}{(1+\delta)^k}, \qquad k \geq 0 \tag{2}$$

where the discount rate d is simply $d^E(k) = \delta$. The key feature of this model is that the discount rate $d^E(k)$ is constant over time, i.e., the rate at which an individual discounts future well-being between today and tomorrow is identical to the rate at which she discounts future well-being between today and tomorrow in one month. In contrast, a widely documented situation in which an individual has a declining rate of time preference is described as hyperbolic discounting, which generally means that the implicit discount rate over longer time horizons is lower than the implicit discount rate over shorter time horizons. A typical case from the family of hyperbolic discounting functions proposed by Mazur (1984) is described in equation 3:

$$D^H(k) = \frac{1}{1+\delta k}, \tag{3}$$

where the hyperbolic discount rate $d^H(k) = (1+\delta k)^{\frac{1}{k}} - 1$ (Andersen *et al.*, 2014).[1] Phelps & Pollak (1968) further introduced a quasi-hyperbolic specification of the discount function for use in a social planner problem:

$$D^{QH}(k) = \begin{cases} 1, & \text{if} \quad k = 0 \\ \frac{\beta}{(1+\delta)^k}, & \text{if} \quad k \geq 0 \end{cases} \tag{4}$$

where $\beta \leq 1$ and the quasi-hyperbolic discount rate $d^{QH}(k) = \left(\frac{\beta}{(1+\delta)^k}\right)^{-\frac{1}{k}} - 1$.[2] A characteristic feature of the quasi-hyperbolic specification is the discontinuity at time $t = 0$. This specification was applied by Laibson (1997) to model individual agent behavior.

Several experimental methods are available to elicit time preferences in both laboratory and field settings, such as lotteries, choice lists, and bidding; however, there is no consensus how to best measure discounting (Andreoni *et al.*, 2015). The basic method for eliciting individual discount rates is conceptually simple—asking subjects questions about whether they prefer an amount of money today (option A) or the same amount + \$$X$ tomorrow (option B), where $X > 0$. By changing X, a researcher can infer bounds for the subject's individual discount rate.[3] Experiments therefore involve a series of questions aligned in lists, such as in the classical choice list design of Coller & Williams (1999) or Harrison *et al.* (2002). Modifications to this basic method are further used to elicit preferences more precisely, such as variations in the delay between options A and B, the domain in which preferences are revealed (money, health, etc.),

[1] In a hyperbolic specification, the discount rate is the value of $d^H(k)$ that solves $D^H(k) = 1/(1+d^H)^k$, i.e., the equation $1/1+\delta k = 1/(1+d^H)^k$.

[2] Again, in the quasi-hyperbolic specification, the discount rate is the value of $d^{QH}(k)$ that solves $D^{QH}(k) = 1/(1+d^{QH})^k$, i.e., the equation $\beta/(1+\delta)^k = 1/(1+d^{HQ})^k$.

[3] The point of the first switch to option B gives a measure of the upper bound of her discount rate. We assume linear utility here for simplicity and discuss a relaxing of this assumption later.

and the magnitude or the nature of the reward (hypothetical or real).

Several types of elicitation methods are routinely used in the experimental literature (Frederick *et al.*, 2002): i) choice, ii) matching, iii) rating, and iv) pricing. The most common type of elicitation is the choice method, where subjects are presented alternative options and are asked to simply choose between them. This method provides discount rate intervals pre-generated by the experimenter rather than precise estimates of the discount rate for specific individuals. The matching method, in contrast, provides an exact inference of the individual's discount rate since she reveals her true indifference point by filling the blank field to equate two intertemporal options. In rating tasks, subjects evaluate individual options by rating their attractiveness on a predefined scale, while in pricing tasks, subjects specify their willingness to pay for individual options in which they either obtain or avoid a particular outcome. In contrast to choice and matching tasks, rating and pricing tasks allow the researcher to manipulate the time variable between subjects since immediate and delayed options are evaluated separately.

Each method described briefly above has its strengths and limitations. When subjects are asked to evaluate multiple options at once in a standard choice list, the earlier choices inevitably influence the choices made later. This procedural limitation—the anchoring effect—can be partially addressed by employing titration procedures and exposing subjects to a sequence of different opposing anchors (Frederick *et al.*, 2002). The timing of an outcome was found to have a much lower effect when evaluating a single option compared to a situation when two options occurring in different times are evaluated against each other at once (Loewenstein, 1987). The timing of two evaluating options is further argued to cause the more general problem of an additional risk or transaction costs imposed on the future option. The recent literature, represented by Harrison *et al.* (2005), Andersen *et al.* (2014), and others, deal with this risk by employing a front end delay, thereby shifting the immediate option to the nearer future and imposing transaction costs on the instant payoff.

Harrison *et al.* (2005) argues that standard choice tasks often executed through multiple price lists (MPL) have three possible disadvantages: i) they elicit only interval responses; ii) they allow subjects to switch back and forth while moving down the list; and iii) they can be subject to framing effects. Harrison *et al.* (2005) therefore introduces an *iterative Multiple Price List* (iMPL) that allows the subjects to iteratively specify their choices through refined options within an interval chosen in the last option.

The inference of discount rates from the experimental task depends on the utility function presented in the discounted utility model 1. This function, however, is unobserved and therefore usually assumed to be linear, generating biased estimates for individuals with non-linear utility functions (Cheung, 2016). Recent papers by Andersen *et al.* (2008, 2014) use the *joint elicitation strategy* to measure time preferences by controlling for non-linear utility. Using the equivalence of utility for risk and time, these authors use a series of binary choices to infer the discount function conditional on the utility function elicited through Holt & Laury (2002)'s risk preference task. Further modifications of the design to measure time preferences by controlling for non-linear utility include, among others, the work of Laury *et al.* (2012), who interact risk with time using a lottery to be paid out with probability p_t in time t and with probability p_{t+k} in time $t+k$, where $p_t \leq p_{t+k}$ and p_{t+k} vary through the choice list. Further experiments measuring time preferences while controlling for non-linear utility are conducted by Takeuchi (2011), who employs separate choices under risk and over time using matched pairs of payoffs; Andreoni & Sprenger (2012b), Andreoni & Sprenger (2012a), and Andreoni *et al.* (2015), who examine risk and time preferences through individual elicitation methods—convex time budgets and double multiple price list tasks—and Attema *et al.* (2016), who introduce a *direct method* to measure discounting that is not dependent on the knowledge or measurement of utility.

An alternative method for inferring discount rates was devised by Chabris *et al.* (2008a), who not only derive intertemporal preferences from standard choice tasks but also adopt an approach of using response times from these choices, i.e., how long it actually takes the subjects to choose between option A and option B. The authors assume that *"subjects should take longest to decide when the two options are most similar in their discounted values"* and therefore argue that the inference from response times should, in principle, work (Chabris *et al.*, 2008b, p. 7). The results of Chabris *et al.* (2008b) suggest that choice-based and response-time-based estimates are nearly identical in their setting.

3 The Dataset

The first step of a meta-analysis is the collection of primary studies. To this end, we search Google Scholar for the literature on discounting and then examine the references of the retrieved studies to search for other usable studies (this method is called "snowballing" in the meta-analysis context). We apply four inclusion criteria. Each study included in our dataset must be an experiment, either lab or field, and must report an estimate of the discount rate (or the discount factor in a way that allows recomputation to the discount rate). We also exclude estimates of the discount rate derived from very short delays (several hours)—these are extreme cases for which it is often difficult to find use in practice. A typical case for which we apply the third selection criterion is Loewenstein (1987), who derive discount factors using delays of 3 hours, 24 hours, and 3 days. The minimum delay for the discounting tasks we include in our dataset is therefore 1 week. Finally, we include only studies published in peer-reviewed journals. The major reason for this inclusion criterion is feasibility, but we also hope that peer review sets a lower bar for quality. Moreover, journal articles generally contain fewer typos and other mistakes in the presentation of results compared to unpublished manuscripts, which is important in meta-analysis.

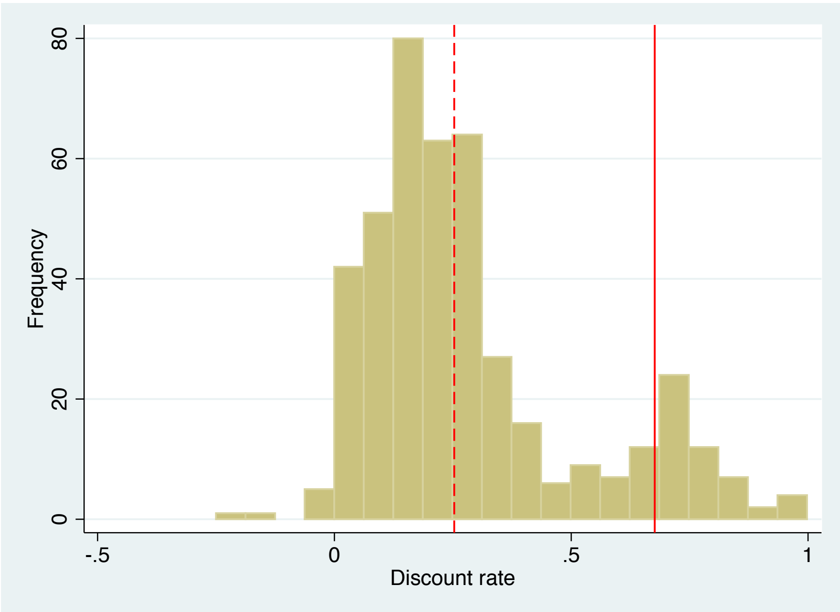
Table 1: Studies included in the meta-analysis

1	Andersen *et al.* (2006)	18	Coller & Williams (1999)
2	Andersen *et al.* (2008)	19	Deck & Jahedi (2015a)
3	Andersen *et al.* (2014)	20	Deck & Jahedi (2015b)
4	Andreoni & Sprenger (2012a)	21	Dolan & Gudex (1995)
5	Andreoni *et al.* (2015)	22	Duquette *et al.* (2012)
6	Attema *et al.* (2016)	23	Hardisty *et al.* (2013)
7	Bauer & Chytilová (2010)	24	Harrison *et al.* (2002)
8	Bauer & Chytilová (2013)	25	Harrison *et al.* (2010)
9	Benzion *et al.* (1989)	26	Loewenstein (1987)
10	Cairns & der Pol (1997)	27	McClure *et al.* (2007)
11	Cassar *et al.* (2017)	28	Meier & Sprenger (2010)
12	Chabris *et al.* (2008b)	29	Meier & Sprenger (2013)
13	Chapman & Elstein (1995)	30	Newell & Siikamäki (2015)
14	Chapman (1996)	31	Read & Read (2004)
15	Chapman & Winquist (1998)	32	Sutter *et al.* (2013)
16	Chapman *et al.* (1999)	33	Tanaka *et al.* (2010)
17	Chesson & Viscusi (2000)	34	Zauberman *et al.* (2009)

We added the last study on January 22, 2019, and terminated the data collection. Our final dataset covers 34 studies comprising 501 estimates of the discount rate. Of these, 447 were reported explicitly as discount rates, and the remaining 54 estimates were reported as discount factors that we recomputed to rates according to the corresponding discounting formulas. The oldest study in our sample was published in 1987, and our meta-analysis thus spans three decades of research in the area. An overview of primary studies included in the meta-analysis is presented in Table 1; the full dataset (together with estimation codes for R and Stata) is available in the online appendix at `meta-analysis.cz/discrate`.

Apart from the key variables for our analysis—the estimated discount rate and its standard error—we codify additional explanatory variables to control for the sources of variation in our data sample. We control for the type of the discounting estimate; that is, whether the estimate was originally reported as the discount rate or the discount factor. We include the length of the time horizon presented to the subjects, i.e., the delay of the experimental task. Some experiments vary the time horizon by holding the payoff constant. Such studies do not separate different time horizons into different treatments and hence do not report the exact time point at which the subjects made the switch. We therefore code the maximum horizon presented in the task and include a dummy variable to control for this method choice.

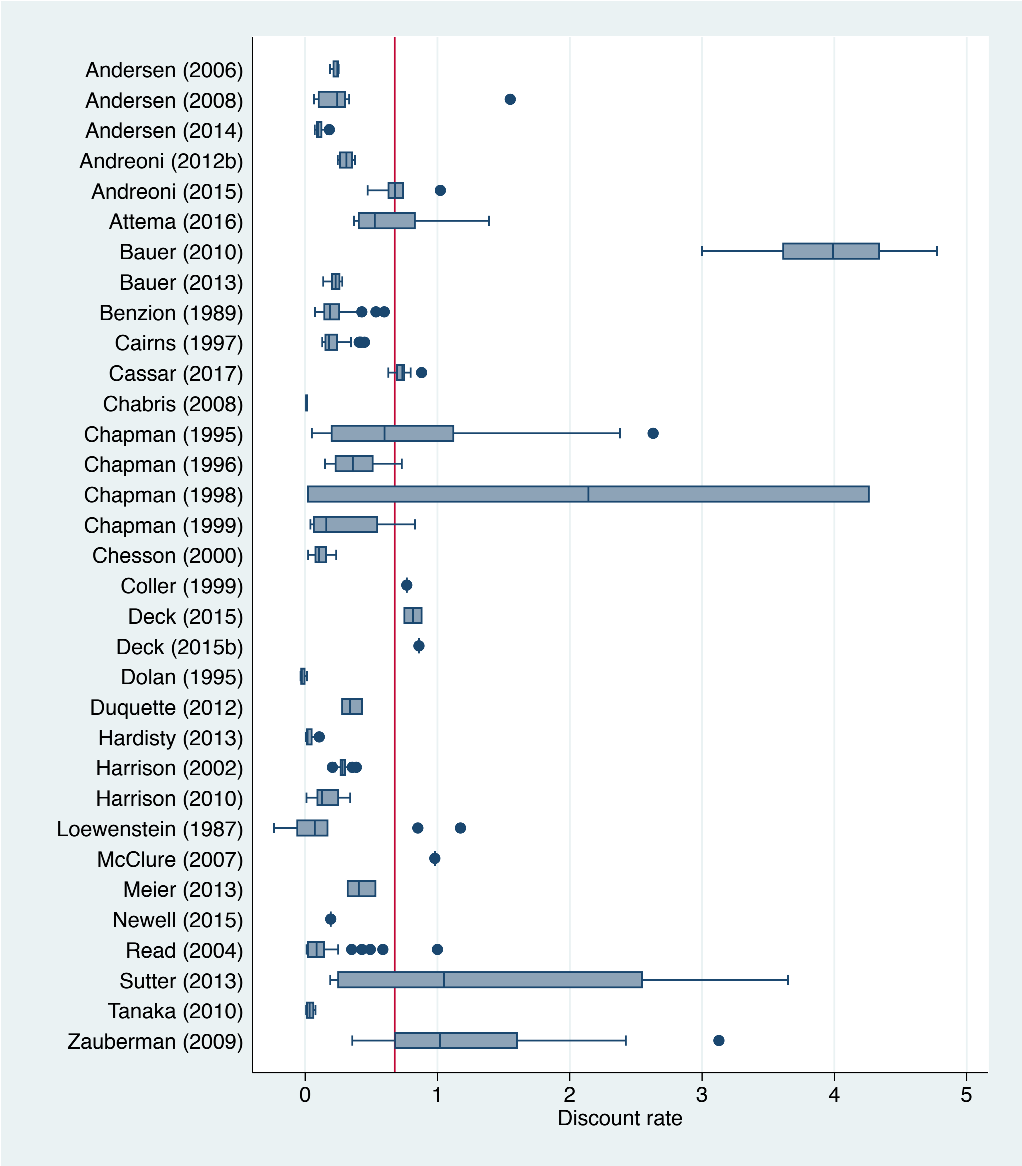
Figure 1: Histogram of discount rate estimates



Notes: The figure depicts a histogram of the discount rate estimates reported by individual studies. Extreme values are omitted from the graph but are accounted for in all regressions. The solid line denotes the sample mean; the dashed line denotes the sample median.

Moreover, we include a dummy variable describing whether the reported estimate relates to hyperbolic or exponential discounting. We further control for whether the study is performed in the lab or in the field; if payoffs used in the study are hypothetical or real, i.e., paid out at the end of the experiment; which elicitation method (choice, matching, and rating) and domain (money, health, etc.) is used to identify the estimate; and whether the framing of the task is positive (giving), negative (loosing) or neutral. We also control for the characteristics of the subject pool: whether it contains students or a more general sample of the population; the gender of the subjects it includes (exclusively males, females, or both); and the continent from which the subject pool was drawn. Additionally, we control for study age and the number of Google Scholar citations weighted by the number of years since the first version of the study

Figure 2: Study-level variation of discount rate estimates



Notes: The figure shows a box plot of discount rate estimates reported in individual primary studies (observations above the 99th percentile are omitted from the figure but are accounted for in all regressions).

Table 2: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Discount rate	501	0.676	1.255	−0.237	0.148	0.658	11.537
SE	314	0.053	0.088	0.0004	0.009	0.044	0.731
SE combined	501	0.083	0.145	0.009	0.009	0.087	1.505

Notes: SE combined = standard errors including those re-sampled from bootstrapping.

appeared in Google Scholar. We describe these variables in more detail in Section 5, which also includes the corresponding Bayesian model averaging analysis.

The estimated discount rates in our dataset have a mean of 0.68 and a standard deviation of 1.26. A histogram of the estimates is presented in Figure 1: the distribution is apparently skewed, with a median value of 0.25. Negative values of the discount rate estimates are rare, though present, and are almost exclusively the matter of the health domain (choosing to experience pain or illness now rather than later). The distribution thus offers several outliers on both sides. We address the potential influence of these outliers on our analysis by winsorizing at the 5% level (the results are robust to changes in the winsorization level).

To be able to employ modern meta-analysis methods, we need measures of precision for individual estimates. However, the standard errors of the discount rate estimates are reported only for 314 of the 501 estimates in our dataset. Researchers in the field sometimes mention that the discount rates they report are large and robust to various changes in the specifications, which constitutes the implicit apology for not reporting precision. As a robustness check (which delivers results close to our baseline and which is presented in the Appendix), we exclude these studies from the dataset and focus only on those for which standard errors can be obtained directly. However, doing so reduces the power of our estimations. Therefore, in the baseline case, we also use studies that do not report precision explicitly. To approximate precision at least at the study level, we apply the bootstrap re-sampling technique. We then combine the explicitly reported standard errors with the standard errors obtained by bootstrapping at the study level. The summary statistics for the key variables are presented in Table 2. The substantial within- and between-study heterogeneity of discount rate estimates, the rationale for a meta-regression analysis, is apparent from Figure 2.

4 Publication Bias

The selective reporting of some estimates (typically those that are intuitive and statistically significant) has been identified as a serious threat to the credibility of empirical economics research (Ioannidis *et al.*, 2017). When estimation noise is large, and therefore standard errors are large, researchers have incentives to preferentially report large point estimates that become statistically significant. McCloskey & Ziliak (2019) liken selective reporting to the Lombard effect, in which speakers increase their vocal effort in the presence of noise. Selective reporting (which is also called publication bias but is not confined to published papers) thus manifests as a correlation between point estimates and their standard errors, which is otherwise difficult to explain.

The general prior among economists and psychologists is that the discount rate is positive. People are impatient; they value the present more than the future. In contrast, a negative estimate of the discount rate means that an individual is willing to accept an offer in the future with a lower value than what is available now, indicating an extraordinary preference for such a state of the world. Negative estimates are rare in our sample but do occur, which suggests that any potential publication bias in the literature is occasional and not universal. These estimates typically concern the health domain in which the experimental subject is offered, for example, to choose between experiencing pain now or a year from now. We do not claim that the average discount rate should be zero or even negative. However, the crux of the publication bias problem is the following: with sufficient imprecision, we always obtain insignificant or negative estimates from time to time. For the same reason, we also obtain large positive estimates. If negative and zero findings are often discarded (they are obviously far from the true value), while large positive estimates are often retained (it is less obvious whether they are far from the true value), the literature as a whole presents distorted results. The typical reported estimate is biased upwards.

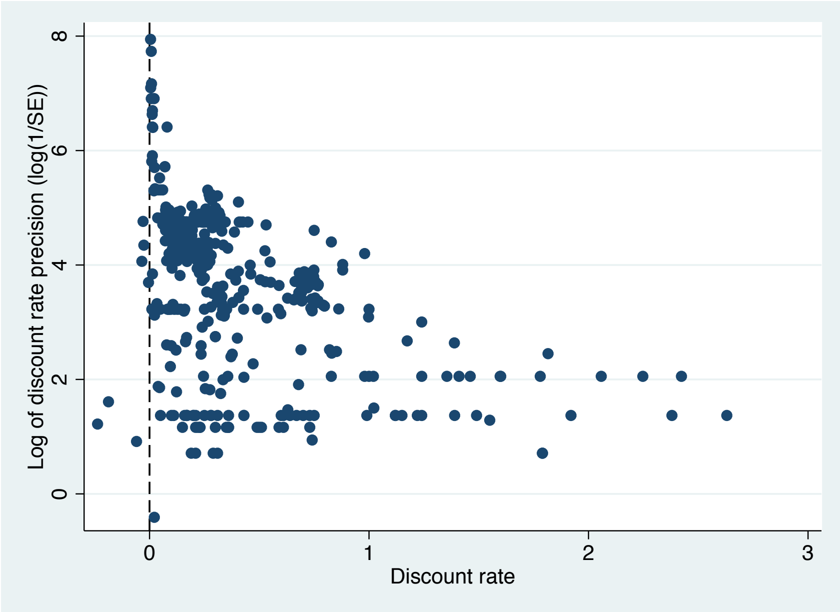
The idea of publication bias is illustrated by Figure 3, the so-called funnel plot (Egger *et al.*, 1997). The horizontal axis depicts the magnitude of the estimate, while the vertical axis depicts the estimate's precision. With no publication bias, the most precise estimates should be close to the underlying average effect. With decreasing precision, we obtain increasing dispersion, which creates the shape of an inverted funnel. However, there is no reason for asymmetry in the funnel. If, in contrast, imprecise negative estimates are discarded but imprecise large positive estimates are reported, we obtain asymmetry—which is precisely what we see from the figure. The funnel plot can thus serve as a visual check of publication bias (Stanley & Doucouliagos, 2010).

Next, we examine the correlation between the discount rate estimates and their standard errors quantitatively to test for the presence of publication bias. Following Stanley & Doucouliagos (2014), we apply a combined conditional estimator, the so-called *PET-PEESE*. First, we test $H_0 : \hat{\delta}_1 = 0$ with a simple meta-regression model used by Egger *et al.* (1997)—the *precision-effect* test in equation 5:

$$\hat{\delta}_{ij} = \delta_1 + \gamma_1 \cdot SE(\hat{\delta}_{ij}) + u_{ij}. \tag{5}$$

Here, the $\hat{\delta}_{ij}$ is the i-th estimate of the discount rate, and $SE(\hat{\delta}_{ij})$ is the corresponding standard error; u_{ij} is a disturbance term. The precision-effect test (PET) *"provides a valid basis for determining whether there is a genuine empirical effect beyond publication selection bias"* (Stanley & Doucouliagos, 2014, p. 64) but is known to give biased estimates of the underlying true effect when the effect is not zero (Stanley, 2008). The first

Figure 3: Funnel plot



Notes: The figure depicts the funnel plot of the discount rate estimates. Extreme values are omitted from the graph but are accounted for in all regressions.

part of Table 3 shows the results of the PET test for various model specifications; we cluster standard errors at the study level in all of them. We obtain positive and statistically significant estimates of γ_1—the coefficient that represents publication bias—in all specifications. Our results also suggest that the underlying mean discount rate is indeed positive. However, to be able to estimate its magnitude, we need to use the PEESE model, reported in the second panel of the table, to which we will turn soon.

In line with Stanley & Doucouliagos (2014), we use weighted specifications to estimate the meta-regression model since equation 5 contains heteroskedasticity by definition.[4] For the weight, we use the *precision* estimate, 1/SE, which gives more weight to more precise results, but a caveat regarding this weighting scheme is worth mentioning. The derivation of standard errors in economics is often an important feature of the model, and weighting by precision can create bias by itself if some studies underestimate the standard error. As an alternative weight, we use the inverse of the number of estimates reported per study. This approach equalizes the impact that each study has on the results. Both weighting schemes yield positive and significant results for publication bias, as well as the corrected mean discount rate.

[4]Heteroskedasticity arises because the standard error of the discount rate estimate—the independent variable—is a measure of the dispersion of the magnitude of the estimate of the discount rate—the dependent variable.

Table 3: Funnel asymmetry tests

PET tests	Precision	FE	FE+IV	OLS	Study
SE (pub. bias)	6.129***	12.01***	15.93***	4.737**	3.884***
	(2.305)	(1.916)	(0.663)	(2.217)	(1.275)
Constant	0.156***	0.0646**	0.267*	0.259***	0.374***
	(0.0378)	(0.0299)	(0.139)	(0.0717)	(0.105)
Observations	501	501	501	501	501
PEESE tests	Precision	FE	FE+IV	OLS	Study
SE^2	18.76**	27.48**	72.05	13.73**	10.83***
	(8.426)	(11.06)	(118.0)	(6.850)	(3.844)
Constant (corr. effect)	0.230***	0.188***	0.349	0.398***	0.492***
	(0.0322)	(0.0306)	(0.307)	(0.0960)	(0.126)
Observations	501	501	501	501	501

Notes: The table reports the results of 5 (PET test) and 6 (PEESE test). Standard errors of the regression parameters are clustered at the study level and shown in parentheses. SE = standard error. Precision = weighted by the inverse of the standard error; FE = study-level fixed effects; IV = Instrumental variables; OLS = ordinary least squares; Study = weighted by the inverse of the number of estimates reported per study. * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.

Since the underlying (unobserved) model of selective reporting is a complex function of the standard error, Stanley & Doucouliagos (2014) examine a quadratic approximation that yields satisfactory results. In line with Stanley & Doucouliagos (2014), we apply this quadratic estimator of the precision-effect estimate with standard errors (PEESE), described in equation 6, where δ_1 from equation 5 is constrained to zero:

$$\hat{\delta}_{ij} = \delta_2 + \gamma_2 \cdot SE^2(\hat{\delta}_{ij}) + u_{ij}. \tag{6}$$

The second part of Table 3 reports the results of equation 6 for different specifications. We again cluster standard errors at the study level since the estimates reported in the same study are unlikely to be independent. Almost all model specifications yield positive and significant results of publication bias as well as the underlying corrected effect. The fixed effect regression clustered at the study level and instrumented by the inverse square root of the sample size does not indicate a statistically significant mean discount rate beyond publication bias, but the power of this estimator is low. The point estimate, moreover, is close to that delivered by other techniques. We find it important to conduct the PEESE estimation with an instrument for the standard error because some omitted study characteristics may influence both the point estimate and the standard error, which would render our results for publication bias spurious. The number of observations in the primary study is an appealing instrument because it is unlikely to be strongly correlated with method choices. In the next section, we pursue this caveat further and control for many additional study characteristics. Our results concerning publication bias hold there as well.

Our results are consistent with the mean discount rate corrected for publication bias between 0.18 and 0.49. The average and median result is 0.34, compared to the uncorrected mean of the reported estimates, 0.68. We therefore conclude that the discounting literature overestimates the mean impatience approximately twofold. It is worth noting that the twofold inflation attributable to publication bias is precisely what Ioannidis *et al.* (2017) find for the economics literature as a whole. Our results also prove to be robust to ignoring observations with re-sampled standard errors. Table 6 in the appendix provides the results of meta-regressions performed only using discount rates with explicitly reported uncertainty measures.

We further check the robustness of our results using alternative methods of correcting for publication bias. We employ the "*Top10*" method presented by Stanley *et al.* (2010), who suggest that "*discarding 90% of the [most imprecise] published findings greatly reduces publication selection bias and is often more efficient than conventional summary statistics.*" (Stanley *et al.*, 2010, p. 70). The average discount rate of the 10% most precise estimates in our sample is 0.17, which would imply publication bias that is twice as strong as what our baseline meta-regression techniques suggest. Furthermore, employing the recent method of the weighted average of the adequately powered estimates (WAAP) presented by Ioannidis *et al.* (2017) produces a corrected mean for the discount rate of 0.2, which is very close to that presented by the Top10 method. Another recent (non-parametric) technique that can be used to correct publication bias in meta-analysis is the ingenious stem-based method by Furukawa (2019). The "stem" in the title of the technique refers to the stem of the funnel plot, and it is a clever generalization of the Top10 method. The result for our dataset is 0.22, which is, again, very close to Top10 and WAAP. Finally, the new method proposed by Andrews & Kasy (2019) offers an estimate for the corrected mean of 0.3. We conclude that our baseline result of 0.34 is conservative; the bias is likely stronger and the corrected mean discount rate smaller.

5 Heterogeneity

The substantial differences in the estimates of the discount rate reported in the experimental literature have already been stressed by several previous studies (Frederick *et al.*, 2002; Percoco & Nijkamp, 2009; Andersen *et al.*, 2014; Cheung, 2016). As Frederick *et al.* (2002, p. 352) puts it: "*While the discounted utility model assumes that people are characterized by a single discount rate, this literature reveals spectacular variation across (and* *even within) studies."* Figure 2 shows strong differences in the results at the study level. In this section, we try to explain the differences by regressing the estimated discount rates on their standard errors together with 21 additional explanatory variables that reflect observable variation in the context in which researchers obtain the estimates.

The first option for estimating such a model is simply running a regression with all the variables. The problem is that not all the variables are equally important; some are probably redundant, and including all variables would diminish the precision of our point estimates for the effects of the important variables. However, we do not know ex ante which variables are redundant. A common approach would be to eliminate potential redundant variables in a step-wise fashion; but in doing so, we can never be sure that we have arrived at the best underlying model. Furthermore, the theory can help us stress some particular variables, but we still do not want to completely ignore the remaining ones. In other words, we face extensive model uncertainty, which is a typical feature of meta-regression analysis. The formal response to model uncertainty is Bayesian model averaging (Raftery *et al.*, 1997), which is our method of choice.

Bayesian model averaging (BMA) tackles the problem of uncertainty by estimating models with all possible combinations of explanatory variables in the dataset[5] and constructing a weighted average over the estimated coefficients across all these models. The weights used for averaging stem from posterior model probabilities derived from from Bayes' theorem and are analogous to information criteria in frequentist econometrics. Posterior model probabilities (PMPs) measure how well the particular model fits the data, conditional on model size. BMA produces posterior inclusion probability (PIP) for each variable, which is the sum of the posterior model probabilities for the models in which the variable is included. Recent applications of Bayesian model averaging in meta-analyses in economics and finance include, for example, Havranek *et al.* (2017); Havranek & Irsova (2017); Havranek *et al.* (2018). More details on BMA can be found in Raftery *et al.* (1997) or Eicher *et al.* (2011).

The application of BMA, however, is not straightforward since estimating the millions of possible model combinations is infeasible. A solution is to approximate the whole model space by applying the Markov chain Monte Carlo algorithm that walks only through the models with high posterior model probabilities (Madigan *et al.*, 1995). For approximation we use the BMS package for R developed by Zeugner & Feldkircher (2015). Bayesian model averaging is sensitive to the estimation framework, particularly to the use of *priors* representing the researcher's prior beliefs on the probability of each model (the model prior) and regression coefficients (Zellner's g-prior).

[5]If the matrix of explanatory variables X contains K potential variables, this means estimating 2^K variable combinations, i.e., 2^K models. This estimation results in $2^{22} = 4,194,304$ models in our case.

In the baseline specification, we follow the agnostic priors supported by Eicher *et al.* (2011), who show that these intuitive priors yield the best predictive performance: the unit information prior ("UIP") for Zellner's g-prior, which assigns the prior the same weight as one observation of data, and the uniform model prior, which gives each model the same prior probability.[6]

5.1 Variables

The explanatory variables we have collected are listed in Table 4; we include the description of each variable, its mean, its standard deviation, and the mean weighted by the inverse of the number of estimates reported per study, which effectively levels the impact each study has on the results. We divide the explanatory variables into 4 categories: estimation characteristics, experimental characteristics, subject pool characteristics, and publication characteristics.

Estimation Characteristics

The variation among the reported discount rate estimates can stem from the theoretical assumptions of the intertemporal choice model used in the experimental task presented to subjects, that is, mainly from the type of the discounting model and the time horizon that subjects face in their decision. The studies included in our dataset use the hyperbolic discounting model most frequently (281 observations; 56% of the data), followed by the exponential discounting model (97; 19%). Special cases of discounting models such as exponential mixture share, quasi-hyperbolic discounting, or mixed general model occur rarely in our dataset (5; 1% of cases in total). Due to a lack of information reported in primary studies, we cannot identify the type of the discounting model in 24% of the cases and use this "unidentified" group as a reference category.

The time horizon of the decisions presented to the subjects spans from one week to 50 years, while the mean value is 4.54 years. Some studies, however, vary the time horizon while holding the payoff constant and therefore do not separate different time horizons into different treatments. Since such studies do not report the exact time point at which the subjects made the switch, we cannot always distinguish the exact time horizon corresponding to the estimate. We thus codify the maximum time horizon presented to the subjects in the task and include the dummy variable *Max delay* to

[6] A robustness check using the benchmark g-prior suggested by Fernández *et al.* (2001) and the beta-binomial model prior according to Ley & Steel (2009) can be found in appendix C; our main results would not change if we opted for this alternative set of priors.

Table 4: Summary of explanatory variables

Variable	Description	Mean	SD	WM
Discount rate	The estimated discount rate.	0.676	1.255	0.826
SE	The standard error of the estimate.	0.083	0.145	0.382
Estimation characteristics				
Discount factor	= 1 if the originally reported variable is the discount factor instead of the rate (we recompute all results into discount rates).	0.108	0.310	0.147
Hyperbolic disc.	= 1 if the discounting type is hyperbolic.	0.567	0.496	0.392
Exponential disc.	= 1 if the discounting type is exponential.	0.196	0.397	0.217
Delay	The logarithm of the time horizon of the task.	0.590	1.522	-0.330
Max delay	= 1 if only the maximum time horizon of the task can be codified.	0.315	0.465	0.408
Lab experiment	= 1 if a controlled laboratory experiment is used instead of a field experiment.	0.677	0.468	0.585
Experimental characteristics				
Real reward	= 1 if the reward subjects received is real instead of hypothetical.	0.467	0.499	0.741
Matching task	= 1 if matching is used for elicitation.	0.429	0.495	0.186
Health domain	= 1 if the experiment concerns health questions.	0.096	0.295	0.091
Other domain	= 1 if the experiment concerns questions other than health or money (vacation, certificates, or a kiss from a movie star).	0.084	0.277	0.071
Negative framing	= 1 if the framing of the experimental task is presented as negative, i.e., "loosing."	0.048	0.214	0.082
Neutral framing	= 1 if the framing of the experimental task is presented as neutral.	0.024	0.153	0.022
Subject pool characteristics				
Sample size	The logarithm of the sample size used for the experiment.	4.818	0.793	4.894
Students	= 1 if the subject pool consists of students only.	0.391	0.489	0.452
Males only	= 1 if the subject pool contains males only.	0.040	0.196	0.026
Females only	= 1 if the subject pool contains females only.	0.044	0.205	0.033
North America	= 1 if the experiment is conducted in North America.	0.315	0.465	0.558
Asia	= 1 if the experiment is conducted in Asia.	0.096	0.295	0.088
Africa	= 1 if the experiment is conducted in Africa.	0.054	0.226	0.029
Publication characteristics				
Citations	The logarithm of the number of citations the study received in Google Scholar normalized by the number of years since the first draft of the study appeared in Google Scholar.	2.762	1.292	2.702
Publication year	The standardized publication year of the study.	0.093	1.033	0.537

Notes: SD = standard deviation, WM = mean weighted by the inverse of the number of estimates reported per study.

control for this feature of experiments. Last but not least, we control the general estimation setup—that is, whether the study employs a controlled laboratory experiment or a field experiment.

Experimental Characteristics

Any experiment can be affected by procedural subtleties. The second set of explanatory variables therefore comprises experimental and behavioral characteristics of the task presented to the subject pool. Psychological research suggests that there should be no systematic difference observed between real and hypothetical payoffs in discounting experiments (Johnson & Bickel, 2002; Kühberger *et al.*, 2002; Locey *et al.*, 2011). The recent literature, however, provides more ambivalent results stating that hypothetical conditions yield patterns of discounting that mirror those for real effort tasks, but these may change with repeated exposure to the decisions. The nature of the payoffs provided with the repetition of those tasks therefore needs to be taken into account when designing discounting studies (Malesza, 2019). We therefore control for this payoff effect by extracting the information on the nature of the reward from primary studies; 53% of the discount rates are computed for hypothetical payoffs. Furthermore, this information can serve as a proxy for the size of the payoffs presented to subjects since large payoffs are often associated with hypothetical decisions and, comparatively, small payoffs with real decisions (Kühberger *et al.*, 2002).[7]

Following the reasoning of Frederick *et al.* (2002) and others, we control for the variation in the estimates caused by the elicitation method used in the experiment. We include a dummy variable for matching tasks, taking choice tasks as the reference category present in 57% of cases. An important behavioral aspect of the corresponding task is represented by the domain over which the intertemporal decision is made. The majority of observations utilize monetary payoffs (82%); we therefore use them as the natural reference category in this regard. We codify the remaining domains by using dummy variables, distinguishing between the health domain (9.6%) and other domains—typically, more exotic ones (vacation, certificate, or a kiss from a movie star—8.4%).

The design of any experiment is seldom immune to the issues of framing effects that refer to the finding that subjects often respond differently to different descriptions of the same problem (Tversky & Kahneman, 1981). The majority of discounting tasks are presented (framed) as positive monetary decisions, i.e., choices between a sure amount of money today and a greater amount tomorrow (92.8%). There are,

[7]Due to the lack of information reported in many primary studies, we could not codify the magnitude of the payoffs directly.

however, also negative framings of the tasks present in our dataset (4.8%). For example, Chapman & Winquist (1998) and Hardisty *et al.* (2013) use monetary losses in their experiments. Other studies with negative framing operate with the health domain (Dolan & Gudex, 1995; Read & Read, 2004). Neutral framing applies for only 2.4% of the observations.

Subject Pool Characteristics

We describe the subject pool characteristics of an individual study by several variables. First, we control for the size of the subject pool by coding the number of subjects used for deriving the estimate; the mean is 169.7. Second, we control for the composition of the subject pool by incorporating dummy variables reflecting whether the pool consists exclusively of male or female subjects. The majority of studies, however, use non-exclusive subject pools consisting of both males and females (91.6%).

A general concern of any experimental study is its external validity, i.e., the extent to which its results can be generalized to other situations. Economic experiments are often criticized for using university students (typically economics majors) as experimental subjects—a pool of people with specific characteristics not always generalizable to the whole population (Marwell & Ames, 1981; Carter & Irons, 1991; Frank *et al.*, 1993). The behavior of decision makers recruited from natural markets has been examined in a variety of contexts, and it has typically not differed from that exhibited by more standard (and far less costly) student subject pools (Davis & Holt, 1993, p. 17).[8] We control for the potential effect of a subject pool composed exclusively of student subjects. Finally, the heterogeneity in the reported discount rates may stem from different cultural characteristics of populations. The primary studies do not give us much information to build on, but at least we can control for continents out of which the subject pool was recruited. The majority of studies recruit subjects from European countries (58.3% obs.) and North America (31.5%). We also experimented with including dummy variables for each individual region, but doing so creates collinearity problems.

Publication Characteristics

We do not exclude any journal articles based on their supposedly poor quality, but we try to control for it—even poor-quality studies can bring useful information, especially if their results differ from those of high-quality studies. Some of the aspects related to

[8] See Davis & Holt (1993) for examples of this evidence. More recent evidence on differences between student and non-student samples is provided by Depositario *et al.* (2009).

quality are captured by the data and method characteristics described above. However, other quality aspects are surely more difficult to observe. Therefore we use two rough proxies: the age of the study and the number of citations. These are no perfect controls for quality, but other things being equal, newer and highly cited studies tend to be more reliable. For computing the age of the study we do not use the year of journal publication; due to different publication lags in different economics and psychology journals, such a measure would be useless. Therefore, we use the date of the first appearance of a draft of the paper in Google Scholar. For citations, we also use Google Scholar and compute the number of per-year citations that the primary study has obtained since the first draft appeared.

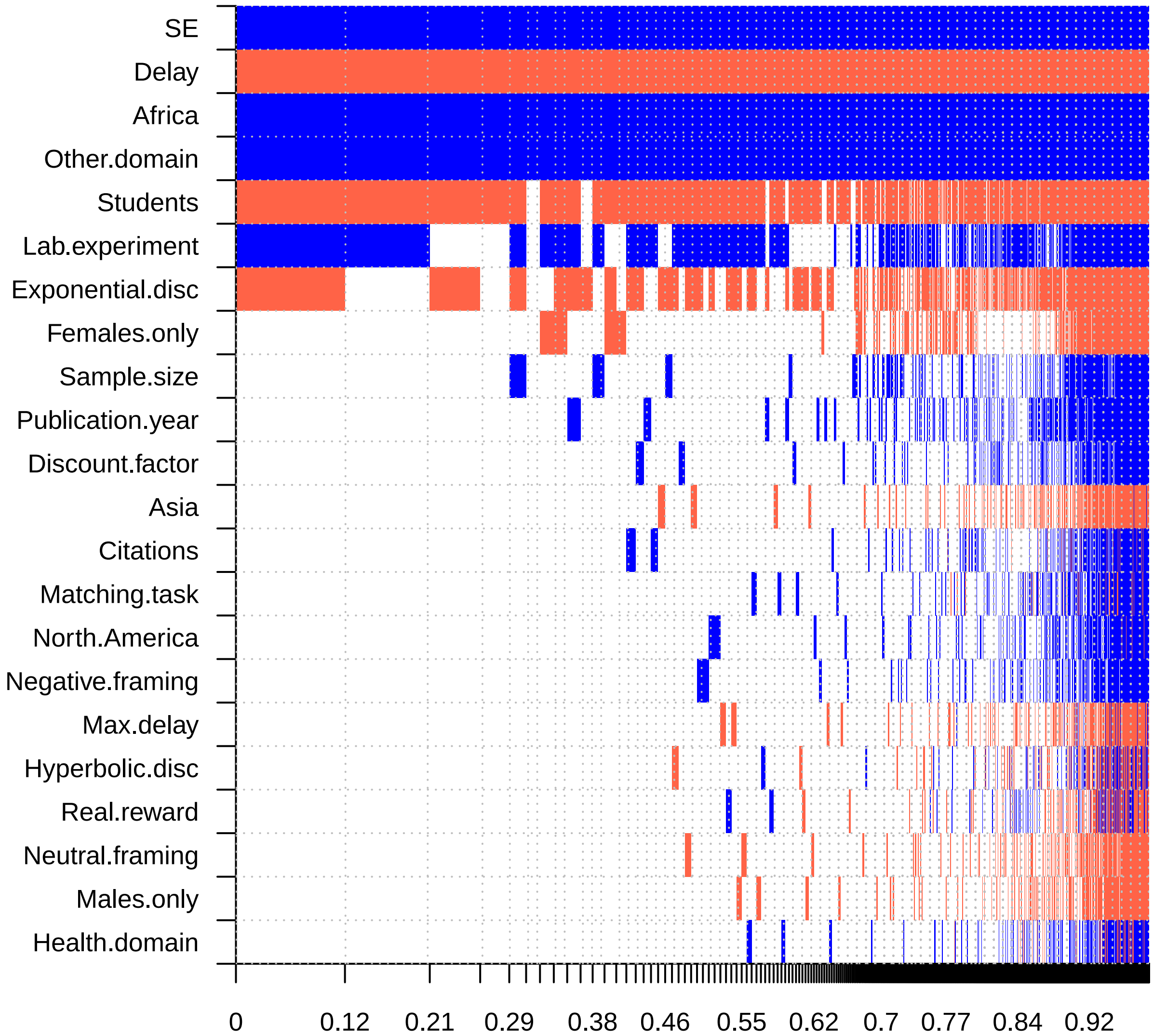
5.2 Results

The results of the BMA estimation are visualized in Figure 4. The variables are displayed on the vertical axis and sorted by posterior inclusion probability. PIP can be thought of as a Bayesian analogy of statistical significance—we therefore see the most "significant" variables at the top of the figure. The horizontal axis denotes individual regression models sorted according to the PMP, from left to right. The PMP represents how well the model fits the data relative to its size; the width of the columns is proportional to the PMP. The colors of individual cells denote the sign of the corresponding regression coefficients. Blue (darker in greyscale) depicts a positive sign, while red (lighter in greyscale) depicts a negative sign. Blank cells denote the exclusion of the variable from the given model.

The numerical results of BMA are reported in the left-hand panel of Table 5, which shows the posterior mean and standard deviation for each variable together with the posterior inclusion probability. Not counting the intercept, which is included by default in all models, seven variables have PIPs above 50%: the standard error, the dummy variable for exponential discounting, the delay in the discounting task, the dummy for lab experiments, the dummy for other (exotic) domains, the dummy for students in the subject pool, and the dummy for subjects drawn from Africa. In the remainder of this subsection, we will go through these results in more detail.

The first important result of the BMA analysis concerns publication bias. Standard errors are robustly correlated with the point estimates of the discount rate even when we control for 21 additional aspects of studies and estimates. The result corroborates our previous findings (especially those using an instrument for the standard error) that the correlation is not spurious and most likely does not result from an omission of factors that influence both the standard error and the point estimate. Moreover, both

Figure 4: Model inclusion in BMA (based on "UIP" g-prior)



Notes: The response variable is the estimate of the discount rate reported in a primary study. SE = standard error. 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 "UIP" Zellner's *g-prior* and "uniform" model probability prior recommended by Eicher *et al.* (2011). Blue (darker in greyscale) depicts included variables with a positive estimated sign. Red (lighter in greyscale) depicts included variables with a negative estimated sign. Variables with no color are not included in the given model. The numerical results of the BMA exercise are reported in Table 5.

Table 5: Explaining the heterogeneity in discount rate estimates

	Bayesian Model Averaging			Frequentist Check (OLS)		
Variable	Post. Mean	Post. SD	PIP	Coeff.	SE	p-val.
Intercept	0.314	NA	1.000	0.329	0.095	0.002
SE	3.010	0.480	1.000	3.406	0.717	0.000
Estimation characteristics						
Discount factor	0.004	0.025	0.067			
Hyperbolic disc.	-0.001	0.016	0.051			
Exponential disc.	-0.109	0.111	0.568	-0.200	0.118	0.099
Delay	-0.185	0.022	1.000	-0.190	0.047	0.000
Max delay	-0.001	0.014	0.052			
Lab experiment	0.107	0.097	0.627	0.146	0.077	0.068
Experimental characteristics						
Real reward	-0.000	0.017	0.051			
Matching task	0.003	0.024	0.059			
Health domain	0.001	0.017	0.043			
Other domain	0.371	0.085	0.998	0.415	0.142	0.006
Negative framing	0.004	0.030	0.054			
Neutral framing	-0.005	0.042	0.050			
Subject pool characteristics						
Sample size	0.005	0.017	0.121			
Students	-0.182	0.091	0.880	-0.229	0.118	0.060
Males only	-0.003	0.031	0.050			
Females only	-0.025	0.076	0.141			
North America	0.002	0.017	0.054			
Asia	-0.005	0.033	0.067			
Africa	2.576	0.148	1.000	2.501	0.231	0.000
Publication characteristics						
Citations	0.001	0.007	0.065			
Publication year	0.005	0.020	0.118			
Observations	501			501		

Notes: The frequentist check (OLS) includes variables recognized by the BMA as having a posterior inclusion probability above 50% (they also comprise the single best model identified by the BMA). Standard errors in the frequentist check are clustered at the study level. SD = standard deviation, PIP = Posterior inclusion probability, SE = standard error. Detailed descriptions of the variables are available in Table 4.

the posterior mean and the coefficient in the frequentist check suggest a strong correlation. We also provide a robustness check by estimating the BMA using a dataset with standard errors that are not approximated by bootstrapping, i.e., only with observations coded from the studies that explicitly provide a measure of statistical precision. The result of this analysis confirms the strong correlation between the point estimates of discount rates and their standard errors, which is strong evidence for publication bias in the literature.[9]

Results for Estimation Characteristics

The second important factor influencing the heterogeneity in discount rate estimates turns out to be the length of the delay over which the decision is made. This factor is inherently embedded as the parameter k in the discounted utility model presented in equation 1. According to the discounted utility theory, the values of all future outcomes should be discounted at a constant rate (Frederick *et al.*, 2002). This statement has two parts that are important for our analysis: the first part highlights that *all* outcomes, without difference, should be discounted, and the second notes that the discounting should be done at a *constant rate*. Regarding the second part, our results indicate that the discount rate is indeed not constant. There is a strong negative effect of the delay on the discount rate. This finding corroborates non-constant hyperbolic discounting, in which the discount rate over longer horizons is lower than that over short horizons. This finding is in line with the results of, among others, Mazur (1984), who presents evidence for hyperbolic discounting, or, more recently Tsukayama & Duckworth (2010), who find that subjects discount rewards more steeply when they find the discounting domain particularly tempting.

A related effect is the importance of the dummy for exponential discounting, of which the constant discount rate is an important property. Our analysis suggests that tasks with exponential setups, i.e., with a constant discount rate between decisions with different delays, decrease the individual elicited estimates. The estimates in our sample do not seem to be significantly different when hyperbolic discounting is applied, but we stress that these estimates are compared only to the baseline of a non-identified discounting type; moreover, a part of the effect related to this variable is already captured by the *Delay* variable described in the previous paragraph.

Our result regarding the length of the delay might be influenced by our inability in some cases to codify the exact time horizon for which the particular discount rate was estimated. Nevertheless, we control for the cases in which we were unable to

[9]The results of this robustness check are presented in the appendix C.

codify the time horizon precisely, and the dummy variable reflecting that the study reports only the maximum time horizon of the task does not prove important in the BMA exercise. So there is no significant difference between observations from different time ranges of the discounting tasks or, more specifically, there is no significant difference between experimental treatments that vary payoffs and those that vary the time horizon instead.

Two additional results related to estimation characteristics are important. The first result is the low posterior inclusion probability and therefore the absence of the variable *Discount factor* in most BMA models, which suggests that, for the magnitude of the discount rate, it does not matter whether the researcher reports the result as a rate or a factor. A second important result is the difference between field and laboratory experiments. This finding suggests that a controlled laboratory environment produces more evidence for impatience than field study environment.

Results for Experimental Characteristics

Several studies find that individual discount rates are not very correlated across different domains such as money and health, and this diversity is called domain independence. Cairns (1992), for example, estimates discount rates that are significantly lower for future health as compared to future wealth states; Chapman & Elstein (1995) demonstrate in two experiments that decision makers use different discount rates for health-related decisions and money-related decisions, with less patience for the health domain. See Loewenstein *et al.* (2003) for more examples of domain independence.

Our results suggest that people tend to be more impatient when making their decisions in more exotic domains than money and health: holiday preferences, gift certificates, kisses from movie stars (here the reward is obviously hypothetical, but we admit we would be impatient too). On the other hand, in contrast to much of the previous literature, there seems to be no statistical difference between monetary and health decisions. We therefore confirm domain independence only partially.

In Section 5.1, we referred to the literature suggesting there should be no difference whether real or hypothetical payoffs are used in discounting experiments. Our results confirm that it indeed does not matter whether the decision is made with fictive payoffs only. Real rewards do not systematically affect the estimates of the discount rate. Researchers can thus use hypothetical questions that have advantages in the elicitation of time preferences since hypothetical setting allows us to ask questions involving long time horizons and large payoffs (Wang *et al.*, 2016).

We find no substantial effect for the remaining experimental characteristics. Different experimental tasks do not bring substantially different results: matching does not seem to differ significantly from choice tasks, which suggests that the inference of an individual's discount rate by the matching method does not systematically outperform the interval elicitation provided by choice tasks. Framing also seems to little affect the resulting discount rate.

Results for Subject Pool Characteristics

The long-term debate over the external validity of the experiments performed on student samples is reflected in our analysis with the variable *Students*. Our results suggest that students make more patient choices in discounting tasks than the general population, which can be caused by several factors, out of which the standard argument would point to the self-selection of students into subject pools. The vast majority of experiments are conducted with university students majoring in economics, who have been shown, for example, to be more selfish than the general population (Marwell & Ames, 1981). Two types of hypotheses explain why this may be the case: 1) the selection hypothesis, according to which individuals concerned with economic incentives opt for economic studies, and 2) the learning hypothesis, which states that individuals studying economics learn behavioral patterns out of the theories and models they pursue (Carter & Irons, 1991). It might be true that not only more "selfish" individuals self-select into study fields such as economics but also that more patient students self-select into the roles of experimental subjects.

Our results provide strong evidence that discount rates elicited from subject pools in Africa significantly differ from those obtained in other parts of the world. The African population is, according to our analysis, much more impatient than the population of other continents. This result is in line with the results of the large cross-country study on time preferences by Wang *et al.* (2016, p. 17), who observe that *"Africa has the lowest percentage of participants choosing to wait* (33%)*."* The benchmark demographic area—Europe—seems to follow similar patterns of discounting as North America and Asia. We find no evidence of any impact of the sample size on the discount rate estimates; large and small studies seem to produce relatively similar results—on average, at least. Furthermore, neither exclusively male nor female subject pools report significantly different results of discount rates in our sample compared to the baseline (mixed) subject pools.

Finally, publication characteristics are not correlated with the reported discount rates. Quality, or at least the aspects of quality that are captured by our rough proxies (citations and age), does not systematically affect the outcomes. There are certainly quality aspects that we do not control for, and an obvious solution would be the addition of study-level fixed effects. We opt for this estimator in the previous section that focuses on publication bias, but here, it is not feasible: for many variables in which we are interested, the within-study variation is very small.

5.3 Robustness Checks

We perform several different sensitivity checks in order to confirm whether our baseline BMA results presented earlier in this section are reasonably robust. First, we combine the reduction in model uncertainty resulting from BMA estimation with traditional frequentist estimation. The best model identified by the BMA exercise includes seven explanatory variables (plus the intercept). These variables also have a posterior inclusion probability above 0.5 and therefore should, according to the classification by Kass & Raftery (1995), have a non-negligible impact on our response variable. We re-estimate this best BMA model using the standard OLS technique, clustering standard errors at the study level and weighting the regression by the inverse number of estimates per study to reduce the impact of large studies. The results of this estimation are provided in the right-hand panel of Table 5.

Next, we perform a robustness check using an alternative set of BMA priors, employing the benchmark g-prior suggested by Fernández *et al.* (2001) together with the beta-binomial model prior, which gives each model size equal prior probability (Ley & Steel, 2009). We label this estimation according to the g-prior parameter as "BRIC." The results of this robustness check are reported in Table 8 in the appendix and are similar to the baseline specification.

Finally, we also perform a robustness check by estimating the BMA on data using only non-combined standard errors (that is, if an observation reported in the study is not precise, we drop the observation). The results are again consistent with our baseline specification, confirming the impact of the seven variables we stress in this paper. Numerical as well as graphical results of this robustness check can be found in appendix C.

6 Concluding Remarks

We provide a quantitative synthesis of the literature that uses experiments to identify individual discount rates. By employing meta-regression methods, we detect selective reporting against null and negative results. The mean reported discount rate is 0.68. Using conservative techniques, we find that the mean drops to 0.34 after we correct for publication bias—that is, people are more patient on average than what is indicated by a naive summary of the conclusions of the experiments. Several complementary methods based on recent advances in meta-analysis point to an even stronger publication bias. This result is in contrast to Imai *et al.* (2019), who report no strong evidence of selective reporting in the literature estimating the present bias parameter.

The estimates of the discount rate vary a great deal. We explain this heterogeneity by using Bayesian model averaging, a method accounting for model uncertainty inherent in meta-analysis. We corroborate the presence of selective reporting in the literature by showing that the standard error is an important factor in the heterogeneity of discount rate estimates. Next, we find that a key feature influencing the reported discount rate is the delay that individual subjects face during the experiment. We also partially confirm domain independence stressed by the previous literature (Cairns, 1992; Chapman & Elstein, 1995; Loewenstein *et al.*, 2003) since discount rates for different questions (for example, health on one hand and a kiss from a movie star on the other) differ systematically. Other important results include the systematic difference between lab and field experiments and the importance of the composition of the subject pool.

The results of our study can be used in various settings. The discount rate has implications for decisions regarding savings, education, smoking, exercise, and other contexts of day-to-day behavior (e.g., Chabris *et al.*, 2008b; Meier & Sprenger, 2010). Accurate measures of discounting parameters can provide helpful guidance in welfare analyses on the potential impacts of policies and provide useful diagnostics for effective policy targeting (Andreoni *et al.*, 2015); moreover, they can be applicable to modeling political campaigns, advertisement, and R&D investment (Deck & Jahedi, 2015b). Other examples of applications are discussed by Deck & Jahedi (2015a), who examine discounting in strategic settings, such as auctions or experimental contests, in which it is often critical to accurately predict the behavior of counterparts.

Climate change policies, in which the individual pure rate of time preference or the social discount rate is needed to evaluate the long-term effects, can serve as an example of a welfare analysis application of our results. The pure rate of time preference together with the growth rate of per capita consumption and the elasticity of marginal utility of consumption create the basis for the calculation of the Ramsay discount rate consisting of *time* and *growth* discounting elements (Fearnside, 2002; Anthoff *et al.*, 2009; Foley *et al.*, 2013). Our discount rate synthesis together with the results of Havranek *et al.* (2015), who provide a meta-analysis of the elasticity of marginal utility of consumption, can be employed to calculate the pure rate of time preference from the Ramsay discount rate.

Three caveats of our results are in order. First, we are unlikely to cover all experiments ever conducted on the discount rate. In particular, the psychological literature on discounting is vast, and we search for primary studies using a Google Scholar query that follows the standards of reporting in economics. Nevertheless, a meta-analysis does not have to cover the entire universe of available studies; it is important only to avoid selecting studies based on their results. Second, only approximately two-thirds of the collected estimates are reported together with a measure of uncertainty from which we can directly compute standard errors. We address this problem partially by re-sampling standard errors at the study level for observations with missing data. (Limiting our attention to the studies that report precision would not change our main results.)

Third, although we control for the differences in many features of study design, experiments involve many unique methodological as well as procedural details that are difficult to codify but that can cause differences in the results of individual studies. Some of these unobserved features might be correlated not only with the reported discount rate but also with the reported standard error, which might make our results concerning publication bias spurious. We partially address this problem by using study fixed effects and by employing the number of observations in primary studies as an instrument for the standard error.

References

AINSLIE, G. (1975): "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin* **82(4)**: pp. 463–496.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, & E. E. RUTSTRÖM (2006): "Elicitation Using Multiple Price List Formats." *Experimental Economics* **9(4)**: pp. 383–405.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, & E. E. RUTSTRÖM (2008): "Eliciting Risk and Time Preferences." *Econometrica* **76(3)**: pp. 583–618.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, & E. E. RUTSTRÖM (2014): "Discounting Behavior: A Reconsideration." *European Economic Review* **71**: pp. 15–33.

ANDREONI, J., M. A. KUHN, & C. SPRENGER (2015): "Measuring Time Preferences: A Comparison of Experimental Methods." *Journal of Economic Behavior & Organization* **116**: pp. 451–464.

ANDREONI, J. & C. SPRENGER (2012a): "Estimating Time Preferences from Convex Budgets." *American Economic Review* **102(7)**: pp. 3333–3356.

ANDREONI, J. & C. SPRENGER (2012b): "Risk Preferences Are Not Time Preferences." *American Economic Review* **102(7)**: pp. 3357–3376.

ANDREWS, I. & M. KASY (2019): "Identification of and Correction for Publication Bias." *American Economic Review* **(forthcoming)**.

ANTHOFF, D., R. S. J. TOL, & G. W. YOHE (2009): "Risk Aversion, Time Preference, and the Social Cost of Carbon." *Environmental Research Letters* **4(2)**: pp. 240–242.

ATTEMA, A. E., H. BLEICHRODT, Y. GAO, Z. HUANG, & P. P. WAKKER (2016): "Measuring Discounting without Measuring Utility." *American Economic Review* **106(6)**: pp. 1476–1494.

BAUER, M. & J. CHYTILOVÁ (2010): "The Impact of Education on Subjective Discount Rate in Ugandan Villages." *Economic Development and Cultural Change* **58(4)**: pp. 643–669.

BAUER, M. & J. CHYTILOVÁ (2013): "Women, Children and Patience: Experimental Evidence from Indian Villages." *Review of Development Economics* **17(4)**: pp. 662–675.

BENZION, U., A. RAPOPORT, & J. YAGIL (1989): "Discount Rates Inferred from Decisions: An Experimental Study." *Management Science* **35(3)**: pp. 270–284.

CAIRNS, J. & M. DER POL (1997): "Constant and Decreasing Timing Aversion for Saving Lives." *Social Science & Medicine* **45(11)**: pp. 1653–1659.

CAIRNS, J. A. (1992): "Health, Wealth and Time Preference." *Project Appraisal* **7(1)**: pp. 31–40.

CARTER, J. R. & M. D. IRONS (1991): "Are Economists Different, and If So, Why?" *Journal of Economic Perspectives* **5(2)**: pp. 171–177.

CASSAR, A., A. HEALY, & C. VON KESSLER (2017): "Trust, Risk, and Time Preferences after a Natural Disaster: Experimental Evidence from Thailand." *World Development* **94**: pp. 90–105.

CHABRIS, C., D. LAIBSON, C. MORRIS, J. SCHULDT, & D. TAUBINSKY (2008a): "Measuring Intertemporal Preferences Using Response Times." *Working paper 14353*, National Bureau of Economic Research, Cambridge, MA.

CHABRIS, C. F., D. LAIBSON, C. L. MORRIS, J. P. SCHULDT, & D. TAUBINSKY (2008b): "Individual Laboratory-Measured Discount Rates Predict Field Behavior." *Journal of Risk and Uncertainty* **37(2-3)**: pp. 237–269.

CHAPMAN, G. B. (1996): "Temporal Discounting and Utility for Health and Money." *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22(3)**: pp. 771–791.

CHAPMAN, G. B. & A. S. ELSTEIN (1995): "Valuing the Future: Temporal Discounting of Health and Money." *Medical decision making* **15(4)**: pp. 373–386.

CHAPMAN, G. B., R. NELSON, & D. B. HIER (1999): "Familiarity and Time Preferences: Decision Making about Treatments for Migraine Headaches and Crohn's Disease." *Journal of Experimental Psychology: Applied* **5(1)**: pp. 17–34.

CHAPMAN, G. B. & J. R. WINQUIST (1998): "The Magnitude Effect: Temporal Discount Rates and Restaurant Tips." *Psychonomic Bulletin & Review* **5(1)**: pp. 119–123.

CHESSON, H. & W. K. VISCUSI (2000): "The Hheterogeneity of Time-Risk Tradeoffs." *Journal of Behavioral Decision Making* **13(2)**: pp. 251–258.

CHEUNG, S. L. (2016): "Recent Developments in the Experimental Elicitation of Time Preference." *Journal of Behavioral and Experimental Finance* **11**: pp. 1–8.

COLLER, M. & M. B. WILLIAMS (1999): "Eliciting Individual Discount Rates." *Experimental Economics* **2(2)**: pp. 107–127.

DAVIS, D. D. & C. A. HOLT (1993): *Experimental Economics*. Princeton: Princeton University Press.

DECK, C. & S. JAHEDI (2015a): "An Experimental Investigation of Time Discounting in Strategic Settings." *Journal of Behavioral and Experimental Economics* **54**: pp. 95–104.

DECK, C. & S. JAHEDI (2015b): "Time Discounting in Strategic Contests." *Journal of Economics & Management Strategy* **24(1)**: pp. 151–164.

DEPOSITARIO, D. P. T., R. M. NAYGA, X. WU, & T. P. LAUDE (2009): "Should Students Be Used as Subjects in Experimental Auctions?" *Economics Letters* **102(2)**: pp. 122–124.

DOLAN, P. & C. GUDEX (1995): "Time Preference, Duration and Health State Valuations." *Health Economics* **4(4)**: pp. 289–299.

DOUCOULIAGOS, C. & T. D. STANLEY (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

DREYFUS, M. K. & W. K. VISCUSI (1995): "Rates of Time Preference and Consumer Valuations of Automobile Safety and Fuel Efficiency." *The Journal of Law and Economics* **38(1)**: pp. 79–105.

DUQUETTE, E., N. HIGGINS, & J. HOROWITZ (2012): "Farmer Discount Rates: Experimental Evidence." *American Journal of Agricultural Economics* **94(2)**: pp. 451–456.

EGGER, M., G. DAVEY SMITH, M. SCHNEIDER, & C. MINDER (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* **315(7109)**: pp. 629–34.

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.

FEARNSIDE, P. M. (2002): "Time Preference in Global Warming Calculations: A Proposal for a Unified Index." *Ecological Economics* **41(1)**: pp. 21–31.

FERNÁNDEZ, C., E. LEY, & M. F. STEEL (2001): "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

FISHER, I. (1930): *The Theory of Interest*. New York: Macmillan.

FOLEY, D. K., A. REZAI, & L. TAYLOR (2013): "The Social Cost of Carbon Emissions: Seven Propositions." *Economics Letters* **121(1)**: pp. 90–97.

FRANK, R. H., T. GILOVICH, & D. T. REGAN (1993): "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives* **7(2)**: pp. 159–171.

FREDERICK, S., G. LOEWENSTEIN, & T. O'DONOGHUE (2002): "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* **40(2)**: pp. 351–401.

FUJII, T. & L. KARP (2008): "Numerical Analysis of Non-Constant Pure Rate of Time Preference: A model of Climate Policy." *Journal of Environmental Economics and Management* **56(1)**: pp. 83–101.

FURUKAWA, C. (2019): "Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method." *Working paper*, MIT.

GOULDER, L. H. & R. N. STAVINS (2002): "Discounting: An Eye on the Future." *Nature* **419(6908)**: pp. 673–674.

HARDISTY, D. J., K. F. THOMPSON, D. H. KRANTZ, & E. U. WEBER (2013): "How to Measure Time Preferences: An Experimental Comparison of Three Methods." *Judgment and Decision Making* **8(3)**: pp. 236–249.

HARRISON, G. W., M. IGEL LAU, E. E. RUTSTRÖM, & M. B. SULLIVAN (2005): "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues." In "Field experiments in economics," pp. 125–218. Emerald Group Publishing Limited.

HARRISON, G. W., M. I. LAU, & E. E. RUTSTRÖM (2010): "Individual Discount Rates and Smoking: Evidence from a Field Experiment in Denmark." *Journal of Health Economics* **29(5)**: pp. 708–717.

HARRISON, G. W., M. I. LAU, & M. B. WILLIAMS (2002): "Estimating Individual Discount Rates in Denmark: A Field Experiment." *American Economic Review* **92(5)**: pp. 1606–1617.

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 (2017): "Do Borders Really Slash Trade? A Meta-Analysis." *IMF Economic Review* **65(2)**: pp. 365–396.

Havranek, T., Z. Irsova, & O. Zeynalova (2018): "Tuition Fees and University Enrolment: A Meta-Regression Analysis." *Oxford Bulletin of Economics and Statistics* **80(6)**: pp. 1145–1184.

Havranek, T., M. Rusnak, & A. Sokolova (2017): "Habit Formation in Consumption: A Meta-Analysis." *European Economic Review* **95**: pp. 142–167.

Holt, C. A. & S. K. Laury (2002): "Risk Aversion and Incentive Effects." *American Economic Review* **92(5)**: pp. 1644–1655.

Imai, T., T. A. Rutter, & C. F. Camerer (2019): "Meta-Analysis of Present-Bias Estimation using Convex Time Budgets." *Working paper*, Stanford University.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The Power of Bias in Economics Research." *Economic Journal* **127(605)**: pp. F236–F265.

Johnson, M. W. & W. K. Bickel (2002): "Within-Subject Comparison of Real and Hypothetical Money Rewards in Delay Discounting." *Journal of the Experimental Analysis of Behavior* **77(2)**: pp. 129–146.

Kass, R. E. & A. E. Raftery (1995): "Bayes Factors." *Journal of the American Statistical Association* **90(430)**: pp. 773–795.

Kovacs, K. F. & D. M. Larson (2008): "Identifying Individual Discount Rates and Valuing Public Open Space with Stated-Preference Models." *Land Economics* **84(2)**: pp. 209–224.

Kühberger, A., M. Schulte-Mecklenbeck, & J. Perner (2002): "Framing Decisions: Hypothetical and Real." *Organizational Behavior and Human Decision Processes* **89(2)**: pp. 1162–1175.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics* **112(2)**: pp. 443–478.

Laury, S. K., M. M. McInnes, & J. Todd Swarthout (2012): "Avoiding the Curves: Direct Elicitation of Time Preferences." *Journal of Risk and Uncertainty* **44(3)**: pp. 181–217.

Lawrance, E. C. (1991): "Poverty and the Rate of Time Preference: Evidence from Panel Data." *Journal of Political Economy* **99(1)**: pp. 54–77.

Ley, E. & M. F. Steel (2009): "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to

Growth Regression." *Journal of Applied Econometrics* **24(4)**: pp. 651–674.

LOCEY, M. L., B. A. JONES, & H. RACHLIN (2011): "Real and Hypothetical Rewards." *Judgment and Decision Making* **6(6)**: pp. 552–564.

LOEWENSTEIN, G. (1987): "Anticipation and the Valuation of Delayed Consumption." *The Economic Journal* **97(387)**: pp. 666–684.

LOEWENSTEIN, G., D. READ, & R. F. BAUMEISTER (2003): *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation.

MADIGAN, D., J. YORK, & D. ALLARD (1995): "Bayesian Graphical Models for Discrete Data." *International Statistical Review* **63(2)**: pp. 215–232.

MALESZA, M. (2019): "The Effects of Potentially Real and Hypothetical Rewards on Effort Discounting in a Student Sample." *Personality and Individual Differences* **(forthcoming)**.

MARWELL, G. & R. E. AMES (1981): "Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods." *Journal of Public Economics* **15(3)**: pp. 295–310.

MAZUR, J. E. (1984): "Tests of an Equivalence Rule for Fixed and Variable Reinforcer Delays." *Journal of Experimental Psychology: Animal Behavior Processes* **10(4)**: pp. 426–436.

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?" *Journal of Economic Eduation* **(forthcoming)**.

MCCLURE, S. M., K. M. ERICSON, D. I. LAIBSON, G. LOEWENSTEIN, & J. D. COHEN (2007): "Time Discounting for Primary Rewards." *Journal of Neuroscience* **27(21)**: pp. 5796–5804.

MEIER, S. & C. SPRENGER (2010): "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* **2(1)**: pp. 193–210.

MEIER, S. & C. D. SPRENGER (2013): "Discounting Financial Literacy: Time Preferences and Participation in Financial Education Programs." *Journal of Economic Behavior & Organization* **95**: pp. 159–174.

NEWELL, R. G. & J. SIIKAMÄKI (2015): "Individual Time Preferences and Energy Efficiency." *American Economic Review* **105(5)**: pp. 196–200.

PERCOCO, M. & P. NIJKAMP (2009): "Estimating Individual Rates of Discount: A Meta-Analysis." *Applied Economics Letters* **16(12)**: pp. 1235–1239.

PHELPS, E. S. & R. A. POLLAK (1968): "On Second-Best National Saving and Game-Equilibrium Growth." *The Review of Economic Studies* **35(2)**: p. 185.

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.

READ, D. & N. L. READ (2004): "Time Discounting over the Lifespan." *Organizational Behavior and Human Decision Processes* **94(1)**: pp. 22–32.

SAMUELSON, P. (1937): "Note on Measurement of Utility." *Review of Economic Studies* **4(2)**: pp. 155–161.

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., S. B. JARRELL, & H. DOUCOULIAGOS (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox." *The American Statistician* **64**: pp. 70–77.

SUTTER, M., M. G. KOCHER, G. R. DANIELA, & S. T. TRAUTMANN (2013): "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* **103(1)**: pp. 510–531.

TAKEUCHI, K. (2011): "Non-Parametric Test of Time Consistency: Present Bias and Future Bias." *Games and Economic Behavior* **71(2)**: pp. 456–478.

TANAKA, T., C. F. CAMERER, & Q. NGUYEN (2010): "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review* **100(1)**: pp. 557–571.

TOL, R. S. J. (1999): "Time Discounting and Optimal Emission Reduction: An Application of FUND." *Climatic Change* **41(3-4)**: pp. 351–362.

TSUKAYAMA, E. & A. L. DUCKWORTH (2010): "Domain-Specific Temporal Discounting and Temptation." *Judgment and Decision Making* **5(2)**: pp. 72–82.

TVERSKY, A. & D. KAHNEMAN (1981): "The Framing of Decisions and the Psychology of Choice." *Science* **211(4481)**: pp. 453–458.

WANG, M., M. O. RIEGER, & T. HENS (2016): "How Time Preferences Differ: Evidence from 53 Countries." *Journal of Economic Psychology* **52**: pp. 115–135.

WARNER, J. T. & S. PLEETER (2001): "The Personal Discount Rate: Evidence from

Military Downsizing Programs." *American Economic Review* **91(1)**: pp. 33–53.

ZAUBERMAN, G., B. K. KIM, S. A. MALKOC, & J. R. BETTMAN (2009): "Discounting Time and Time Discounting: Subjective Time Perception and Intertemporal Preferences." *Journal of Marketing Research* **46(4)**: pp. 543–556.

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.

Appendices

A Robustness Checks of PET-PEESE (for Online Publication)

Table 6: Funnel asymmetry tests

PET tests	Precision	FE	FE+IV	OLS	Study
SE (pub. bias)	12.24***	13.75***	16.91***	11.50***	7.468**
	(3.041)	(1.502)	(0.714)	(3.619)	(3.590)
Constant	0.0731	0.0525**	0.221	0.108	0.248**
	(0.0473)	(0.0200)	(0.142)	(0.0856)	(0.112)
Observations	314	314	314	314	314
PEESE tests	Precision	FE	FE+IV	OLS	Study
SE^2	55.62***	79.87***	134.8	47.94***	29.73*
	(13.57)	(16.66)	(197.1)	(13.96)	(15.89)
Constant (corr. effect)	0.201***	0.129***	0.304	0.339***	0.417***
	(0.0430)	(0.0379)	(0.310)	(0.0815)	(0.134)
Observations	314	314	314	314	314

Notes: The table reports results of regressions 5 (PET test) and 6 (PEESE test) for observations with reported SE (standard error). The SE of regression parameters are clustered at the study level and shown in parentheses. Precision = weighted by the inverse of the standard error; FE = study-level fixed effects; IV = Instrumental variables; OLS = ordinary least squares; Study = weighted by the inverse of the number of estimates reported per study. * $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$.

B Diagnostics of BMA (for Online Publication)

Table 7: Summary of BMA estimation

Mean no. regressors 7.1669	Draws 2e+06	Burn-ins 1e+06	Time 4.401 minutes
No. models visited 520 596	Modelspace 4 194 304	Visited 12%	Topmodels 98%
Corr PMP 0.9999	No. Obs 501	Model Prior uniform	g-Prior UIP
Shrinkage-Stats Av=0.998			

Notes: In this specification, we employ the priors suggested by Eicher *et al.* (2011) based on the predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of the data).

Figure 5: Correlation matrix of BMA variables

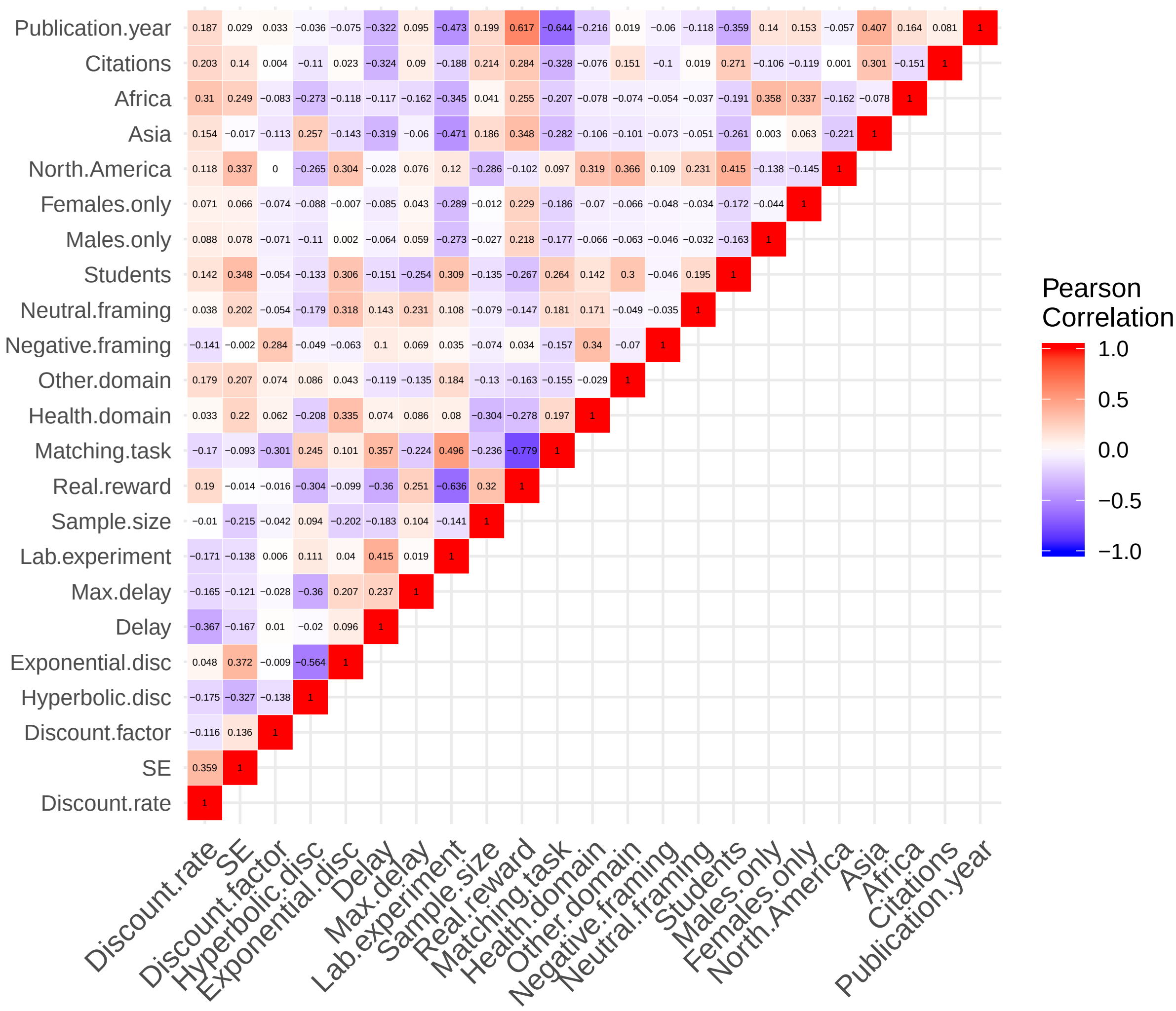
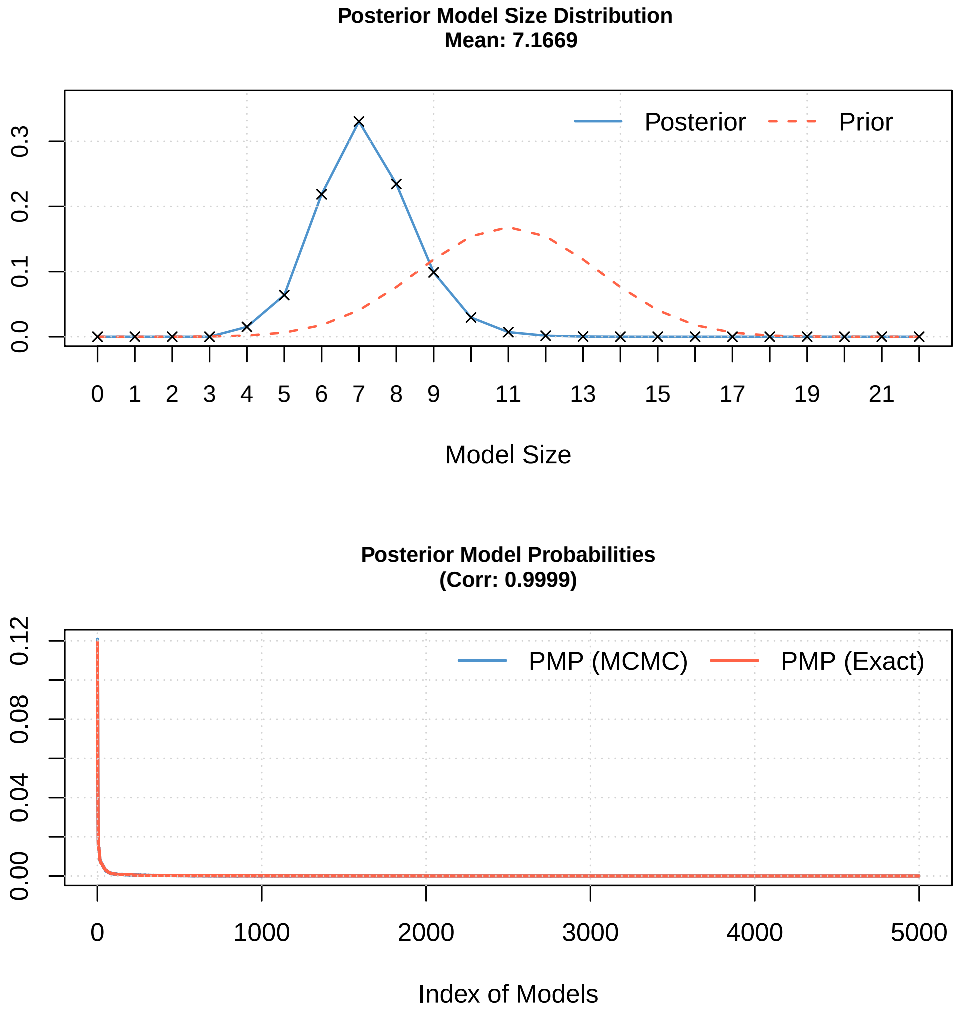
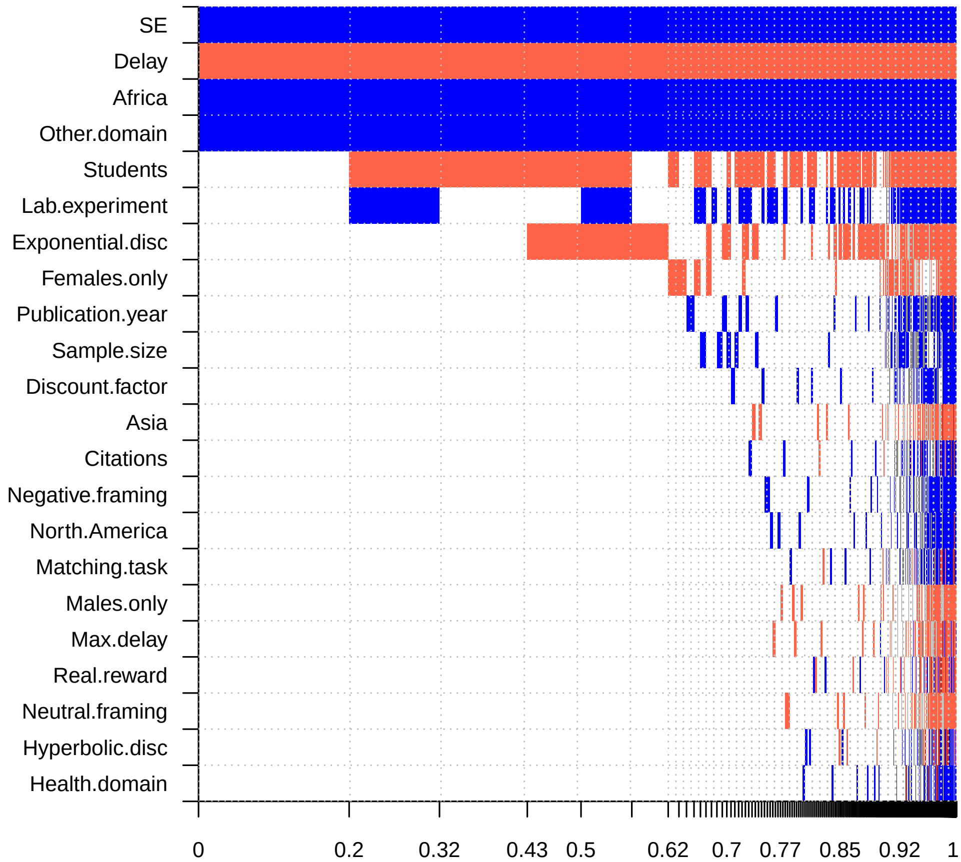


Figure 6: Model size and convergence for the UIP prior



C Robustness Checks of BMA (for Online Publication)

Figure 7: Model inclusion in BMA (based on the "BRIC" g-prior)



Notes: The response variable is the estimate of the discount rate. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes cumulative posterior model probabilities. Estimation based on "BRIC" Zellner's *g-prior* according to Fernández *et al.* (2001) and "random" model probability prior suggested by Ley & Steel (2009). Blue (darker in greyscale) depicts included variables with a positive estimated sign. Red (lighter in greyscale) depicts included variables with a negative estimated sign. Variables with no color are not included in the model. Numerical results of the BMA exercise are reported in Table 8.

Table 8: Heterogeneity in discount rate estimates

Variable	BRIC g-prior Post. Mean	BRIC g-prior Post. SD	BRIC g-prior PIP	Without re-sampling Post. Mean	Without re-sampling Post. SD	Without re-sampling PIP
Intercept	0.346	NA	1.000	-0.193	NA	1.000
SE combined	2.701	0.501	1.000	NA	NA	NA
SE	NA	NA	NA	2.877	0.608	0.999
Estimation characteristics						
Discount factor	0.002	0.017	0.031	0.820	0.096	1.000
Hyperbolic disc.	-0.000	0.009	0.022	0.008	0.033	0.113
Exponential disc.	-0.062	0.099	0.326	-0.000	0.022	0.059
Delay	-0.176	0.021	1.000	-0.117	0.019	1.000
Max delay	-0.000	0.009	0.022	-0.003	0.026	0.072
Lab experiment	0.057	0.088	0.339	0.190	0.087	0.907
Experimental characteristics						
Real reward	-0.000	0.011	0.022	-0.518	0.102	0.999
Matching task	0.000	0.013	0.024	-0.025	0.076	0.152
Health domain	0.000	0.011	0.019	-0.084	0.144	0.316
Other domain	0.378	0.084	0.998	0.032	0.087	0.168
Negative framing	0.002	0.020	0.025	0.004	0.038	0.067
Neutral framing	-0.002	0.027	0.022	NA	NA	NA
Subject pool characteristics						
Sample size	0.002	0.012	0.053	0.074	0.050	0.770
Students	-0.120	0.107	0.622	-0.301	0.080	0.995
Males only	-0.001	0.022	0.023	0.014	0.053	0.112
Females only	-0.014	0.058	0.073	0.005	0.036	0.072
North America	0.001	0.011	0.024	0.012	0.041	0.134
Asia	-0.002	0.021	0.028	-0.113	0.133	0.495
Africa	2.629	0.147	1.000	2.896	0.144	1.000
Publication characteristics						
Citations	0.000	0.004	0.025	0.118	0.027	0.999
Publication year	0.002	0.013	0.059	0.209	0.040	0.999
Observations	501			314		

Notes: We do not include the variable *Neutral framing* into this BMA analysis without re-sampling of missing standard errors since this variable is not present in the reduced dataset. SD = standard deviation, PIP = Posterior inclusion probability, SE = standard error.

Figure 8: Model size and convergence for the BRIC prior

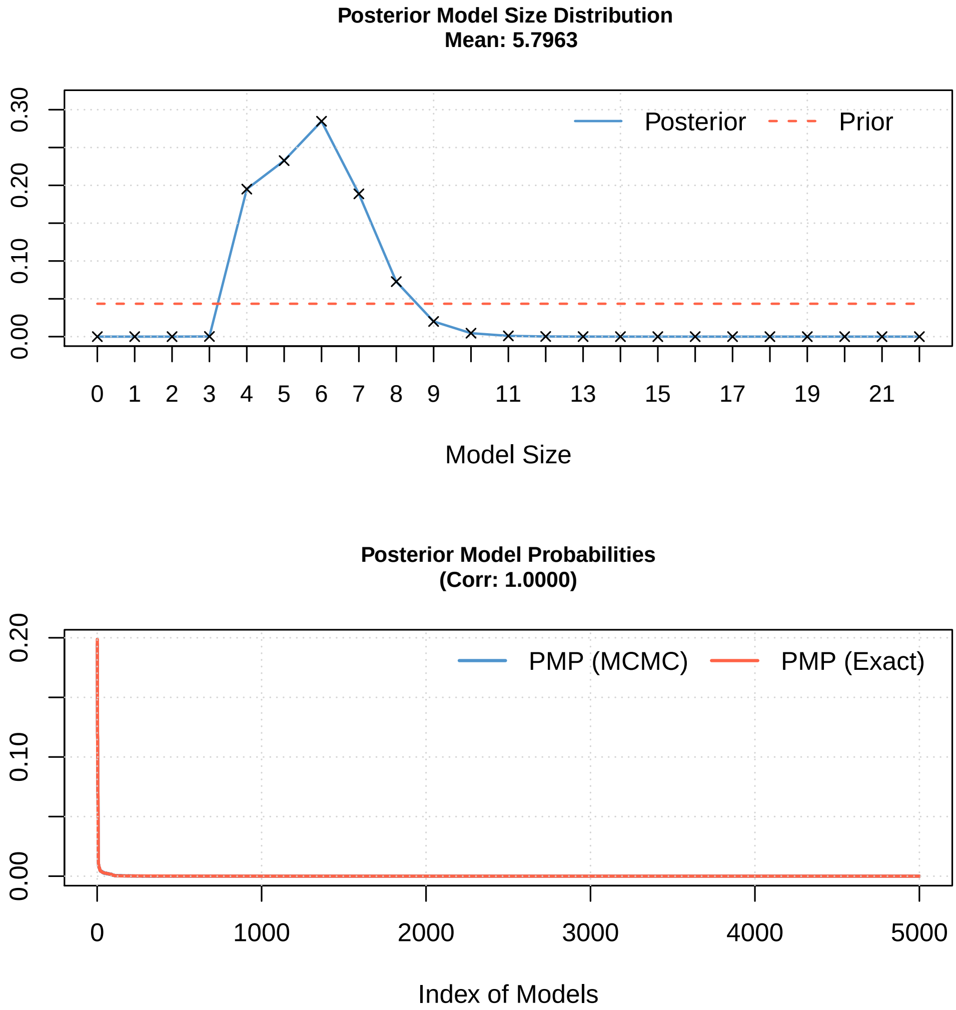
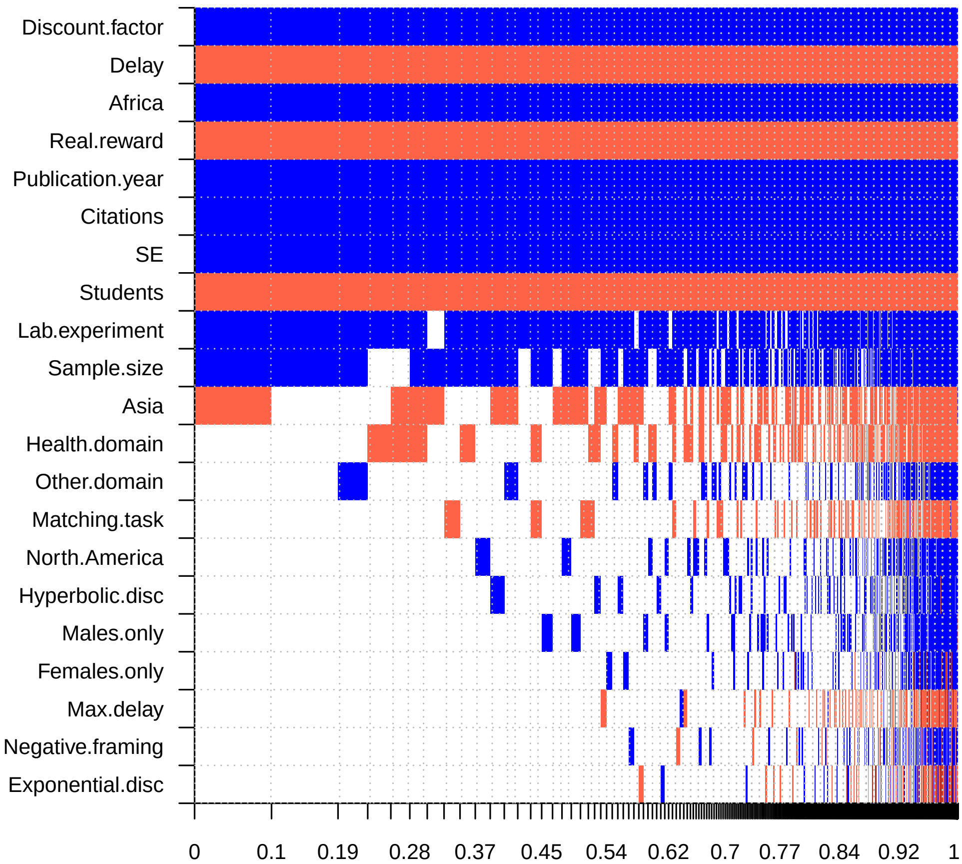
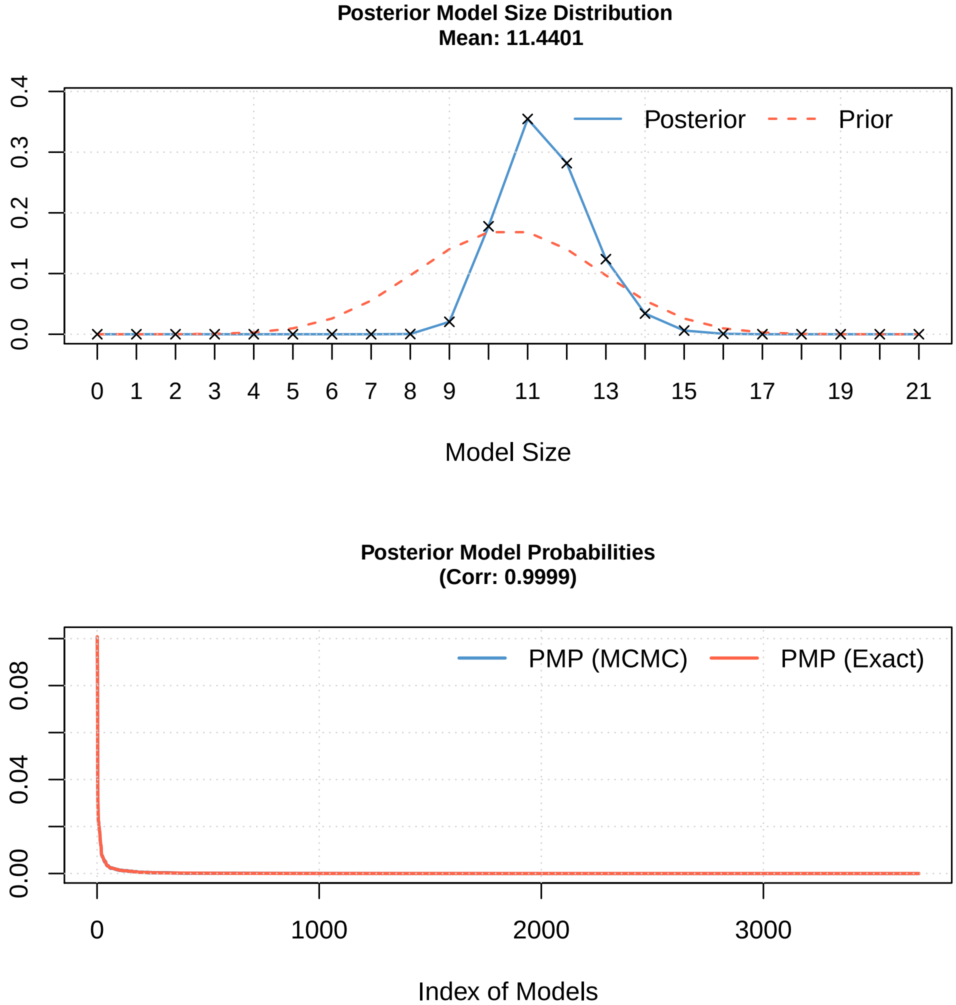


Figure 9: Model inclusion in BMA with non-combined SE (based on the "UIP" g-prior)



Notes: SE = standard error. The robustness check includes observations with explicitly reported measures of precision. The response variable is the estimate of the discount rate. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes cumulative posterior model probabilities. Estimation based on "UIP" Zellner's *g-prior* according to Fernández *et al.* (2001) and "random" model probability prior suggested by Ley & Steel (2009). Blue (darker in greyscale) depicts included variables with a positive estimated sign. Red (lighter in greyscale) depicts included variables with a negative estimated sign. Variables with no color are not included in the model. Numerical results of the BMA exercise are reported in Table 8.

Figure 10: Model size and convergence for non-combined SE (based on the "UIP" g-prior)



Notes: The robustness check includes only observations with directly reported measures of precision.