META-ANALYSIS OF PRESENT-BIAS ESTIMATION USING CONVEX TIME BUDGETS*

Taisuke Imai, Tom A. Rutter and Colin F. Camerer

We examine 220 estimates of the present-bias parameter from 28 articles using the Convex Time Budget protocol. The literature shows that people are on average present-biased, but estimates exhibit substantial heterogeneity across studies. There is evidence of modest selective reporting in the direction of over-reporting present bias. The primary source of heterogeneity is the type of reward, either monetary or non-monetary, but this effect is weakened after correcting for selective reporting. In studies using monetary rewards, the delay until the issue of the reward associated with the 'current' time period influences estimates of the present-bias parameter.

Most choices create benefits and costs that occur at different points in time. Domains of these inter-temporal choices include health (e.g., eating and exercise), financial decision-making (e.g., saving for retirement), pursuit of education, household decisions, and more. In many of these domains, introspection and experimental evidence suggest that people often exhibit *present bias*: people prefer a smaller immediate reward to a larger delayed reward in the present, but they reverse their preferences when these two alternatives are shifted to the future by the same amount of time. Understanding how and why people make such present-biased choices in many domains informs design of government policy, corporate practices and clinical practices.

The exponentially discounted utility model (EDU; Samuelson, 1937; Koopmans, 1960) is the standard model of inter-temporal choice in economics. The model assumes that an individual's inter-temporal preferences are governed by a parameter δ, called the *discount factor*, and that she attaches the relative weight δ^t to the utility from consumption she receives t periods in the future. The quasi-hyperbolic discounted utility model (QHD; Phelps and Pollak, 1968; Laibson, 1997), also known as the present-biased preferences model, is a one-parameter extension of EDU. It is designed to capture dynamically inconsistent choices while retaining the tractability of EDU. In QHD an agent (at period 0) values a consumption stream $(x_0, \ldots, x_T)$ according to

$$U(x_0, \ldots, x_T) = u(x_0) + \beta \sum_{t=1}^{T} \delta^t u(x_t), \tag{1}$$
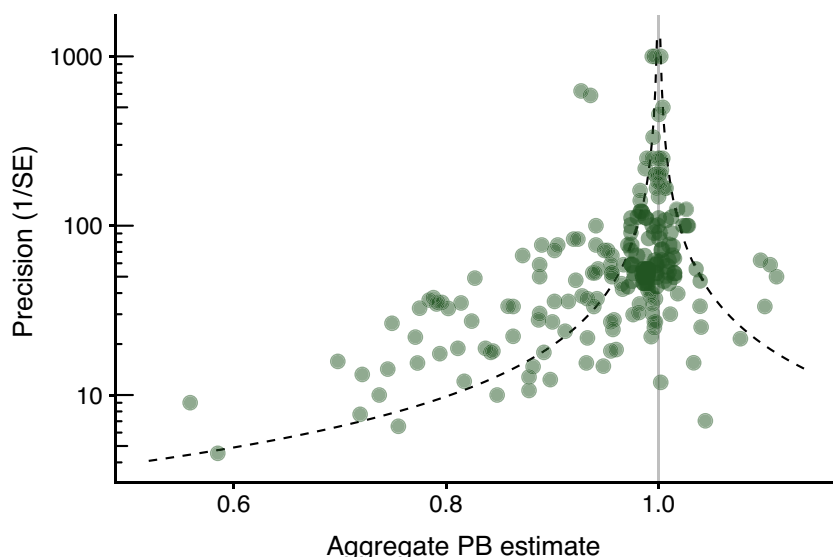
Fig. 1. *Funnel Plot of Estimates of Present-Bias Parameter (PB).*
Notes: The *y*-axis (precision; inverse SE) is presented in the log scale. The dotted curves indicate the boundaries for rejection of the null hypothesis of no present bias ($PB = 1$; vertical grey line) for a two-sided test at the 5% significance level.

where $\delta > 0$ is a traditional discount factor and $\beta > 0$ captures present bias. Note that the utilities from 'future' periods ($t \geq 1$) are exponentially weighted as in the standard EDU, while this stream of future utilities is also discounted by β. Note also that QHD includes EDU as a special case when $\beta = 1$ (there is no present bias; time-consistency). QHD is the most widely used representation of present-biased preferences, although other functional forms (particularly variants of hyperbolic discounting) will exhibit present bias too.[1]

In this paper we assemble a data set of empirical estimates of present-biased preferences measured with the experimental method called the Convex Time Budget (CTB; Andreoni and Sprenger, 2012) and meta-analyse those data. The meta-analysis gives tentative answers to three questions:

(*i*) What is an average value of β?
(*ii*) Is there selective reporting or publication bias?
(*iii*) How does β vary reliably with types of rewards, subject population, estimation methods, etc.?

Our meta-analysis collects 220 estimates of the present-bias parameter in the QHD model (β in (1); hereafter *PB*) from 31 studies reported in 28 articles. The distribution of estimates and the relation with their associated SEs is presented in the 'funnel plot' in Figure 1. A significant

---

[1] See, for example, O'Donoghue and Rabin (1999; 2001), Gruber and Kőszegi (2001), DellaVigna and Malmendier (2006), and Heidhues and Kőszegi (2010) for applications of (naïve) present-biased preferences, and O'Donoghue and Rabin (2015) for a short overview. See Ericson and Laibson (2019) for a broad coverage of models of what they term 'present-focused' preferences, including, but not restricted to, QHD.

proportion of estimated *PB*s are smaller than one, indicating present bias rather than future bias. The dotted curves indicate the boundaries for rejection of the null hypothesis of no present bias ($PB = 1$) for a two-sided test at the 5% significance level; estimates outside the boundaries are rejections. The figure shows that many studies *did not* find strong evidence to reject the null of $PB = 1$, but those that do reject the null hypothesis tend to show present bias ($PB < 1$) rather than future bias ($PB > 1$).

We now provide a preview of our results. We find statistically significant evidence of present bias overall; our meta-analytic average *PB* is between 0.95 and 0.97. However, the reported estimates differ systematically by the type of reward: the values for monetary-reward studies are close to one, indicating the absence of present bias, while studies with non-monetary rewards report a lower average *PB* of 0.88. We also find evidence suggesting selective reporting, in the direction of over-reporting $PB < 1$ in studies using a non-monetary reward. Within the studies using monetary rewards, the delay until the issue of the 'current period' ($t = 0$) reward is shown to robustly influence estimated *PB*.

Our contribution is substantive because it presents the best available estimates of *PB* and how much they vary. This evidence should be useful to many empirical economists for whom a *PB* has been applied, including in household finance (e.g., Angeletos *et al.*, 2001; Meier and Sprenger, 2010; Beshears *et al.*, 2020), health decisions (Fang and Wang, 2015), labour contracts (Kaur *et al.*, 2010; 2015; Bisin and Hyndman, 2020), demand for commitment devices (Ashraf *et al.*, 2006; Beshears *et al.*, 2015; John, 2020) and others. It should also be useful for experimentalists who want to understand which aspects of the design might influence their estimates of *PB*.

Meta-analysis presumes that, along with conventional 'narrative' reviews, it is useful to compile studies using specific inclusion criteria and compare numbers measured in different studies. It hardly bears mentioning that even in the presence of quantitative meta-analyses, narrative reviews will always be useful. They allow insightful commentary on which studies authors believe are particularly interesting, diagnostic, or deserving of replication and extension, in a way that meta-analysis does not easily permit.

At the same time, narrative reviews do not typically specify inclusion criteria and usually do not compare study results on one or more quantitative metrics. As a result, until a meta-analysis such as ours, it is fair to say that even the most expert scholars are not fully aware of what all existing studies have to say about the numerical size and variation in *PB*. Meta-analysis goes further by compiling accessible cross-study data (which others can reanalyse), establishing the central tendency of numerical estimations, exploring cross-study moderators that affect estimates and testing for various kinds of selective reporting.

Meta-analysis is designed to accumulate scientific knowledge, and also detect non-random reporting or publication of estimates that deviate from the average. Since it was first introduced by Glass (1976), meta-analysis has played an important role in evidence-based practices in medicine and policy (Gurevitch *et al.*, 2018). However, meta-analysis has been less common in economics until recently (Stanley, 2001).[2] The current study is the first systematic meta-analysis on the structural estimation of present bias in QHD, focusing specifically on empirical approaches based on the CTB protocol.[3] Prominent reviews of evidence about inter-temporal choices and

---

[2] See a list of relevant publications indexed on RePec at: https://ideas.repec.org/k/metaana.html (last accessed: 4 November 2020).

[3] In a companion paper Imai *et al.* (2018) conduct a large-scale meta-analysis of empirical estimates of discount rates. The data set covers estimates from both experimental and non-experimental studies in economics, psychology, neuroscience, medicine and other fields. In a contemporaneous and independent work, Matoušek *et al.* (2020) conduct

PB include the classic piece by Frederick *et al.* (2002) and more recent coverage by Ericson and Laibson (2019) and Cohen *et al.* (2020). These articles are narrative and do not provide systematic collection and analysis of empirical observations (they rather describe subsets of important contributions and themes that emerge across studies).[4]

The next section explains how we construct the data set. Section 2 describes observable characteristics of the studies and variation in experimental design. Section 3 presents the results.

1. Data and Method

1.1. *The Convex Time Budget Protocol*

There is a large body of evidence on estimation of time preferences, including present-biased preferences. Many experimental methods have been proposed in the literature, but here we focus on the method called the Convex Time Budget (CTB), introduced by Andreoni and Sprenger (2012).[5]

The main goal of this method is to elicit all the parameters of the QHD model—the discount factor δ, present bias β and instantaneous utility function u—in a single experimental instrument. Subjects in a CTB experiment are asked to choose a 'bundle' of rewards (x_t, x_{t+k}) delivered at two points in time $(t, t + k)$ under an inter-temporal budget constraint with a k-period gross interest rate of $1 + r$. By asking a series of allocation questions with varying $(t, t + k)$ and $1 + r$, one can identify parameters of the QHD model.[6] See more details in Online Appendix A.

The CTB protocol instantly became popular. The protocol has been applied not only in laboratory experiments but also in field experiments in developing countries. As we describe below, we have variation in several aspects of CTB design, which we exploit in meta-regression analysis.

1.2. *Identification and Selection of Relevant Studies*

Every good meta-analysis starts by casting a wide net trying to identify relevant studies. In order to deliver an unbiased meta-analysis, it is important to make sure that identification and selection of papers are guided by unambiguously defined inclusion criteria. In our case the main criterion is to 'include all articles that conducted experiments or surveys with the CTB protocol'. We searched for both published and unpublished papers to have a sufficient sample size and to be able to check indicators of publication bias and selective reporting.

We searched for articles that employed the CTB protocol using Google Scholar, first by querying papers that cited Andreoni and Sprenger (2012), Andreoni *et al.* (2015) or Augenblick *et al.* (2015). We also searched for papers with the keyword 'convex time budget'. These two

a similar meta-analysis of experimentally measured discount rates, not *PB*, examining 927 estimates reported in 56 published articles. They find evidence of selective reporting against zero or negative discount rates. They also find that the differences in reported estimates are systematically associated with the domain (money and health), location (lab and field), subject pool and gain–loss framing.

[4] Cohen *et al.* (2020) document the design characteristics of 222 empirical studies identified using Google Scholar, but they do not analyse parameter estimates reported in these studies.

[5] An experimental design concept that is similar to CTB is discussed in Cubitt and Read (2007).

[6] Roughly speaking, variation in gross interest rates $1 + r$ identifies the curvature of the instantaneous utility function u, variation in the delay length k identifies the discount factor δ, and whether the sooner payment date is today ($t = 0$) or not identifies present bias β. Since the key driver of the identification of β is the change in allocations between time points $(0, k)$ and $(t, t + k)$, the CTB protocol is able to recover not only present bias but also future bias. Online Appendix A illustrates optimal allocation decisions in the CTB protocol for a present-biased as well as a future-biased agent against the benchmark of the time-consistent agent (Figure A.2).
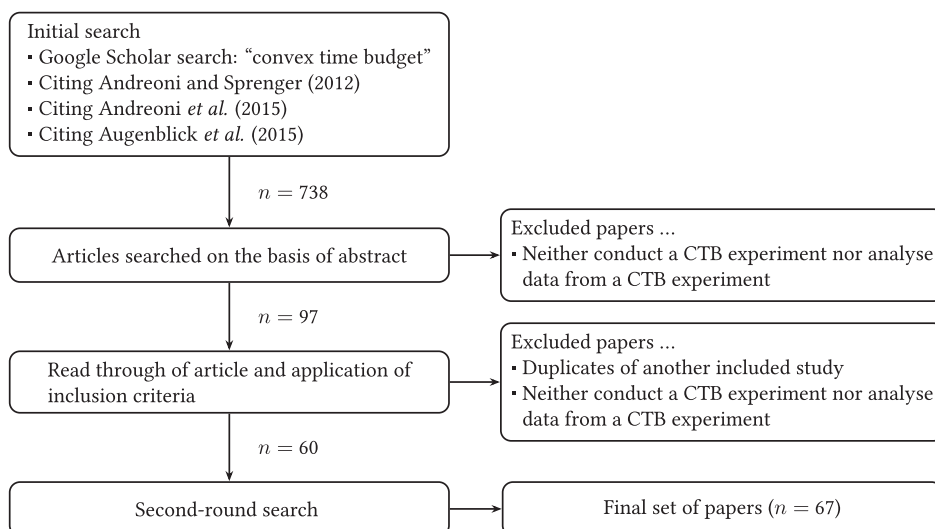
Fig. 2. *Paper Search and Data Construction.*

sets of searches, done on 28 November and 15 December 2017, returned a total of 738 results (including overlaps), which we further narrowed down by examining the titles and the abstracts.

As mentioned above, we searched for any articles, both published and unpublished, conducting experiments or surveys involving the CTB protocol. Note that this broad inclusion criterion kept studies even if QHD parameters were not estimated. These studies did not contribute to our main meta-analysis but still provided some additional information regarding how the CTB protocol has been used in the literature. For this reason we kept track of these studies without estimates as well.

We performed the second-round search (using the same query) and updated the database in the fall of 2018. The final data set includes 67 articles.[7] Figure 2 illustrates our selection procedure.

Note that our inclusion criteria specifically excluded other studies that were informative about present bias. Narrative reviews are better equipped to weave discoveries from such papers into a coherent conclusion. For example, Augenblick (2018) varies time of delivery of initial payments, and finds a decay effect in which a few hours' delay reduces present bias substantially. There are many, many other papers in economics, psychology and cognitive neuroscience that are important but were not included because they did not use CTB.

1.3. *Data Construction*

After identifying relevant articles, we assembled the data set by coding estimation results and characteristics of the experimental design. We called a collection of estimates a *study* when they were from the same experimental design. These two units of observation—an article and a study—coincided in many cases, but allowed us to distinguish two conceptually different

---

[7] Tables B.1 and B.2 in the Online Appendix list all studies (and their basic design characteristics) in the data set, split by the existence of parameter estimates. Online Appendix D presents the full list of references.

experiments reported in a single article. For example, monetary reward and effort-cost versions of CTB in Augenblick *et al.* (2015) were two separate 'studies'.

Our primary variable of interest was the estimate of present-biasedness, but we also coded other parameters in the QHD model (such as discount factor, utility curvature and parameter for stochastic choice, if available). Studies reported either aggregate-level parameter estimates (i.e., pool choice data from all subjects and estimate a set of parameters for the 'representative subject') or certain summary statistics, such as the mean or median of individual-level estimates. We coded these two types of estimate separately.[8] We also coded SEs of parameter estimates from aggregate-level analysis in order to control for the quality of the study in the meta-analysis reported below.

We also coded variables describing characteristics of experimental design and econometric strategies. These variables include, among others: location of the experiments (e.g., laboratory, field, online), types of reward (e.g., real or hypothetical, money, effort), delivery method (e.g., cash, cheque, gift card) and subject pool (e.g., children, college students, general population). Table B.4 in the Online Appendix lists the variables coded in the study. Some studies implemented the CTB protocol with a number of treatment variations, such as hunger, cognitive resource depletion, financial education intervention, time pressure and so on (Table B.3 in the Online Appendix). We coded a dummy variable for treatment. We call a study *neutral* if there is no treatment variation (there is a single data set of experimental conditions).

2. Features of Studies and Experimental Designs

We identified 67 articles that conducted experiments or surveys that used the CTB protocol, of which 36 were published (or 'in press'), including 9 articles published in one of the 'top 5' journals (as of 31 December 2018). There were 36 articles that reported structurally estimated QHD parameters either at the aggregate level or at the individual level. The median number of estimates reported in an article was three. Ten studies reported more than 10 estimates, and two of these reported more than 30.

Observable features of experimental design did not exhibit a marked difference between studies with parameter estimates and those without (Tables 1 and 2; Figure C.3 in the Online Appendix).

Roughly half of the studies reported laboratory experiments. Online experiments constituted less than 20% of the studies in the data set. Only one experiment studied choices made by children in a classroom. Studies were conducted in 29 different countries, as shown in Figure C.2, although a third of studies analysed data from the USA.[9]

Most of the studies recruited participants from the population of college/university students, or the general population (including retirees). It is important to note that several studies in our sample estimated QHD parameters using non-monetary rewards (more precisely, using the cost of working on tedious real-effort tasks), following Augenblick *et al.* (2015). Studies that used monetary rewards differed in how future payments were made: some used bank transfer or sent

---

[8] In our main meta-analysis below, we focus only on the aggregate-level estimates, since there are not many individual-level estimates and the reporting format is not common across these studies. More precisely, we identified only 44 individual-level estimates from 10 studies. Six of these estimates are the mean of the distribution and the other 38 are the median. The former six estimates are accompanied by the SD of the distribution. See Figure B.1 in the Online Appendix.

[9] These 29 countries/regions are Afghanistan, Australia, China, Colombia, Ethiopia, France, Germany, Guatemala, India, Italy, Japan, Kenya, Malawi, Mozambique, Nepal, Netherlands, Nigeria, Pakistan, the Philippines, Singapore, South Africa, Spain, Taiwan, Thailand, Turkey, Uganda, UK, USA and Vietnam.

Table 1. *Characteristics of CTB Studies in the Data Set (I).*

	All CTB studies		Studies with estimates	
	Frequency	Proportion (%)	Frequency	Proportion (%)
Total number of studies	67	100.0	36	100.0
Content of study				
Report *PB* parameter estimates	36	53.7		
Publication status (as of 31 December 2018)				
Published	36	53.7	17	47.2
Published in 'top 5' journal	9	13.4	3	8.3
Type of study				
Lab experiment	29	43.3	15	41.7
Field experiment	27	40.3	14	38.9
Online experiment	10	14.9	6	16.7
Classroom	1	1.5	1	2.8
Geographic location				
Continent: North America	22	32.8	13	36.1
Continent: Europe	13	19.4	8	22.2
Continent: Asia	17	25.4	9	25.0
Continent: Africa	11	16.4	5	13.9
Continent: Oceania	2	3.0	0	0.0
Continent: South America	2	3.0	1	2.8
Reporting of PB parameter estimates				
Aggregate-level estimates			31	86.1
with SEs			28	77.8
Individual-level estimates			10	27.8

Notes: Published in 'top 5' journal indicates that the paper is published (or 'in press') in one of the following journals: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*. *Reporting of parameter estimates*: a paper is counted as reporting a particular type of estimate if it reports *at least one* specification reporting the given type of estimate. Five additional studies reported estimates of EDU parameters, not QHD (i.e., no *PB* parameter in the model).

cheques to the subjects, but in some other experiments subjects came back to the laboratory to pick up the payments.

These observable study characteristics exhibited some patterns of co-occurrence (Figures C.4–C.6 in the Online Appendix). For example, laboratory experiments tended to have student subjects, while field studies were more likely to recruit from the general population.

Experimental elicitation of time preferences requires researchers to design experiments so that the effects of potential confounding factors are minimized. As discussed in the literature, two notable examples of potential confounding factors are the uncertainty or distrust of future payment and the differences in transaction costs between receiving outcomes at earlier and later dates (e.g., Ericson and Laibson, 2019; Cohen *et al.*, 2020).[10] Andreoni and Sprenger (2012) dealt with these issues using the following strategies: (*i*) they gave the experimental participants the business cards of the researcher (and told them to reach out if they did not receive the payment) to increase trust; and (*ii*) they split the participation fee into two parts, one delivered together with the 'sooner payment' and the other delivered with the 'later payment', to reduce the difference

[10] Our view is that both uncertainty about payment and transaction costs are minor factors that many previous experiments have controlled effectively, in the sense that they do not change estimates of *PB* by numerical amounts that would give one pause in deciding whether *PB* should be investigated in applications. See Halevy (2014) for similar scepticism.

Table 2. *Characteristics of CTB Studies in the Data Set (II).*

	All CTB studies		Studies with estimates	
	Frequency	Proportion (%)	Frequency	Proportion (%)
Total number of studies	67	100.0	36	100.0
Subject population				
Children and teens	7	10.4	1	2.8
University students	28	41.8	15	41.7
General population	32	47.8	20	55.6
Reward type				
Real incentive	65	97.0	34	94.4
Certain	63	94.0	36	100.0
Gains	59	88.1	29	80.6
Money	53	79.1	29	80.6
Effort	9	13.4	8	22.2
Reward delivery method				
Bank transfer	19	28.4	11	30.6
Pickup	5	7.5	3	8.3
Cheque	10	14.9	6	16.7
Cash	8	11.9	7	19.4
PayPal	2	3.0	2	5.6
CTB implementation				
Corner allowed	58	86.6	30	83.3
Computer	28	41.8	19	52.8
Deal with confounding factors				
Uncertainty about future payments	46	68.7	23	63.9
Equalize transaction cost	52	77.6	28	77.8

Notes: A study is counted as offering a certain type of reward if it offers the reward to *at least one* of the samples the study analyses.

Table 3. *Characteristics of Budgets and Time Frames.*

	All CTB studies (65)				Studies with estimates (38)			
	Mean	Median	Min.	Max.	Mean	Median	Min.	Max.
Number of budget sets	17.57	14	1	55	21.63	20	4	55
Number of time frames	3.14	2	1	10	3.89	3	2	10
Minimum delay length (days)	34.44	28	1	365	24.56	29	1	42
Maximum delay length (days)	163.98	30	1	7,300	52.78	49	1	150
Mean delay length (days)	89.43	30	1	3,285	36.89	34	1	84

Notes: Two studies using short time horizons (Barton, 2015; Imas *et al.*, 2018) are not included in the bottom three rows.

in transaction costs caused by receiving rewards at two different points in time. Many of the later studies in our sample also followed these strategies.

Let us now turn to the detail of the CTB protocol. There are several variables researchers can specify: number of budgets (i.e., questions), set of time frames (pairs (t, k) of 'sooner' payment date t and delay length k), gross interest rates over k periods, and so on. Table 3 summarizes the ranges and central tendencies of these design variables.

Table 4. *Characteristics of Aggregate-Level PB Estimates.*

	Frequency	Proportion (%)
Number of estimates	227	
SE reported	220	96.9
Instantaneous utility function u		
Estimated	222	97.8
Imputed	2	0.9
Fixed	3	1.3
Specification of u		
Constant relative risk aversion (CRRA)	183	80.6
Constant absolute risk aversion (CARA)	15	6.6
Other	6	2.6
Convex effort cost	22	9.7
Estimation method		
OLS or NLS	62	27.3
Tobit	107	47.1
Multinomial logit or maximum likelihood	25	11.0
Background consumption		
Fixed at zero	134	59.0
Fixed at non-zero value	70	30.8
Estimated	23	10.1

On average, researchers asked 22 questions to recover QHD parameters. In all protocols the questions were asked close together in time.[11] Subjects made allocation decisions on four different (t, k) pairs on average, implying that each time frame was associated on average with five levels of gross interest rates over k periods. The length of delay between the 'sooner' payment and the 'later' payment varied substantially across studies. On average, the minimum waiting period was a little over one month and the maximum waiting period was six to eight months.

Finally, we look at the assumptions and econometric approaches employed to structurally estimate QHD parameters (Table 4). There were 227 estimates in the data set, and a significant majority assumed a constant relative risk aversion (CRRA) specification for the instantaneous utility function u in model (1). The typical specification for studies using real-effort tasks was a convex effort cost function. There were five observations where the utility curvature was either fixed at some exogenous value or imputed from an additional elicitation task such as a multiple price list (Holt and Laury, 2002).

The popular econometric approach is (two-limit) Tobit regression, since researchers need to handle censoring due to corner choices. See Andreoni and Sprenger (2012) and Augenblick *et al.* (2015) for a detailed explanation of identification and estimation using non-linear least squares (NLS) and Tobit approaches.

---

[11] It is conceivable that people are 'artificially' consistent, giving the same early–late allocations under time frames $(0, k)$ and (t, k). Such a desire to appear consistent will lead to estimates of *PB* biased towards one. If this is the case, the most we can say is that *PB* estimates represent a bound (an upper bound if *PB* is less than one and a lower bound if *PB* is greater than one). A different procedure that increases elapsed time between responses might produce *PB* values closer to one. Note, however, that Imai and Camerer (2018) use an adaptively optimal experimental design procedure that selects individually tailored budget lines and time frames based on each subject's responses to the previous questions. In that design the questions subjects face vary substantially from trial to trial, and $(0, k)$ and (t, k) budget lines with the same level are rarely presented together. In that design it is more difficult to select allocations in an artificially consistent manner, yet the estimated *PB* values are similar to those in the standard non-adaptive design (with monetary reward) covered here. While this is just one study, it suggests that a procedural change that happened to reduce between-trial consistency did not change the value of estimated *PB* significantly.
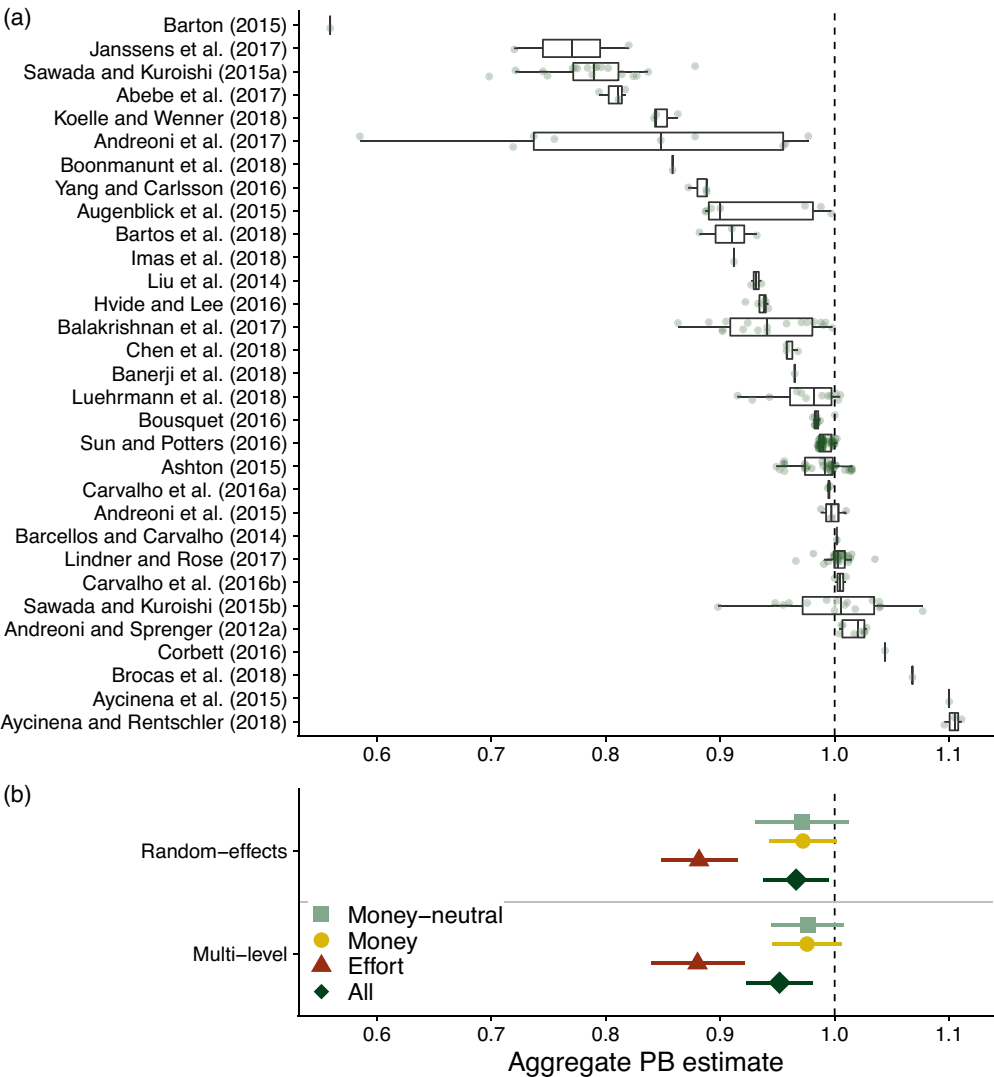
Fig. 3. *Present-Bias Parameter Estimates. Panel (a): Boxplots of present-bias estimates reported in each paper. Panel (b): Random-effects and multi-level estimation of average present bias.*
Notes: The vertical dotted line indicates no present/future bias.

3. Results

Aggregate-level estimates of the present-bias parameter from each article in the data set are shown in Figure 3(a). About 77% of these estimates are below one, indicating present bias. It is clear from the figure that these estimates vary not only between studies but also within each study. We have 220 aggregate-level estimates with SEs (Table 4). In this section we first calculate the 'average' present-bias parameter using the standard meta-analytic technique. We next investigate

the existence or absence of selective reporting. Finally, we investigate the heterogeneity of observed estimates using the moderator variables coded in our data set.

3.1. *Meta-Analytic Synthesis of Present-Bias Estimates*

We start by providing a meta-analytic estimation of the 'average' *PB* in the data set. The analysis below provides a tentative answer to the question *What is the average value of PB measured by the CTB protocol?*

We begin by setting up the simplest meta-analytic framework, the common-effect model. This is

$$PB_j = PB_0 + \varepsilon_j, \tag{2}$$

where PB_j is the jth estimate of present-bias parameter in the data set ($j = 1, \ldots, m$), PB_0 is the 'true' present-bias parameter that is assumed to be common to all observations in the data and ε_j is the sampling error. It is assumed that $\varepsilon_j \sim \mathcal{N}(0, \nu_j^2)$ and the sampling variance ν_j^2 is known. We can obtain the common-effect estimate of PB_0 as the weighted average of individual estimates:

$$\overline{PB}_0^{\mathrm{CE}} = \frac{\sum_{j=1}^m w_j^{\mathrm{CE}} PB_j}{\sum_{j=1}^m w_j^{\mathrm{CE}}},$$

where the weights are given by the inverse variance, $w_j^{\mathrm{CE}} = 1/\nu_j^2$. In this average, estimates with higher precision (smaller SEs) are given larger weights.

The random-effects meta-analysis (RE; DerSimonian and Laird, 1986) assumes that

$$PB_j = \mu_j + \varepsilon_j = PB_0 + \xi_j + \varepsilon_j, \tag{3}$$

where ε_j is a sampling error of PB_j as an estimate of μ_j, and the estimate-specific 'true' effect μ_j is decomposed into PB_0 (the grand mean) and the sampling error ξ_j. It is further assumed that $\xi_j \sim \mathcal{N}(0, \tau^2)$, where τ^2 captures the between-observation heterogeneity, beyond the mere sampling variance, that must be estimated. Note that the random-effects model (3) reduces to the common-effect model (2) when $\tau^2 = 0$. The random-effects estimate $\overline{PB}_0^{\mathrm{RE}}$ is again the weighted average of the individual PB_j, but now the weights are given by $w_j^{\mathrm{RE}} = 1/(\nu_j^2 + \hat{\tau}^2)$, where $\hat{\tau}^2$ is the estimate of τ^2.

Our data set includes *statistically dependent* estimates of *PB*, since many studies included in our meta-analysis report multiple estimates from the same experiment (e.g., using different econometric approaches or using different subsamples). In order to account for the dependency, we use cluster-robust variance estimation to account for correlation of estimates among each study (Hedges *et al.*, 2010).

We also address the issue of 'overly influential' observations (i.e., leverage points) by calculating *DFBETAS* (Belsley *et al.*, 1980), which measures how much the regression coefficient changes if one observation is removed, standardized by the coefficient SE from the regression without the target observation. Following Bollen and Jackman (1985), we identify any observations as influential if $|DFBETAS| > 1$ (i.e., the observation shifts the coefficient at least one SE).[12]

---

[12] *DFBETAS* is intended to measure the impact of removing observation m on the kth coefficient. Let $\widehat{\gamma}_k$ and $\widehat{\gamma}_k^{(m)}$ be the estimated kth coefficient with and without observation m, respectively. Then, the impact of observation m is given by $DFBETAS_m = (\widehat{\gamma}_k - \widehat{\gamma}_k^{(m)})/SE(\widehat{\gamma}_k^{(m)})$, where $SE(\widehat{\gamma}_k^{(m)})$ is the SE of $\widehat{\gamma}_k^{(m)}$.

Table 5. *Meta-Analytic Average of Present-Bias Parameter.*

	All studies		Monetary (all)		Monetary ('neutral')		Effort cost	
	(1) RE	(2) ML	(3) RE	(4) ML	(5) RE	(6) ML	(7) RE	(8) ML
$\overline{PB}_0$	0.9663	0.9518	0.9723	0.9758	0.9716	0.9766	0.8815	0.8802
	(0.0147)	(0.0149)	(0.0150)	(0.0154)	(0.0209)	(0.0161)	(0.0171)	(0.0208)
p-value	0.0297	0.0031	0.0805	0.1334	0.1898	0.1640	0.0001	0.0004
$\hat{\tau}^2$	0.0031		0.0029		0.0037		0.0021	
I^2	98.0824		98.1257		97.5754		45.9587	
I^2_{within}		0.7528		0.9336		0.4389		9.2236
I^2_{between}		98.1997		97.9088		97.6832		39.5324
Observations	217	217	193	193	140	140	24	24
Studies	29	29	20	20	19	19	9	9

Notes: p-values are from the two-sided test of the null hypothesis H_0: $PB = 1$. SEs in parentheses are cluster-robust (Hedges *et al.*, 2010). τ^2 in the random-effects model is estimated using the restricted maximum likelihood method. Three observations with a large influence measure ($|DFBETAS| > 1$) are excluded.

This procedure identifies three influential observations in our data: one estimate from Barcellos and Carvalho (2014) and two estimates from Liu *et al.* (2014). We remove these three estimates from our simple meta-analysis presented in this subsection.[13]

We estimate the meta-analytic averages for four different subsets of the data: (*i*) all estimates; (*ii*) observations from studies using monetary rewards; (*iii*) observations from 'neutral' studies using monetary rewards; and (*iv*) observations from studies using the real-effort version of CTB.

Table 5 reports the results from the random-effects specification (odd-numbered columns), which are also presented in Figure 3(b).[14] All specifications show $\overline{PB}_0^{\text{RE}} < 1$, indicating present bias. The overall $\overline{PB}_0^{\text{RE}}$ is 0.97, which is statistically significantly different from one at the 5% significance level. Two estimates from CTB studies using a monetary reward are also smaller than one, but we cannot reject the null hypothesis of no present bias. We observe a smaller $\overline{PB}_0^{\text{RE}}$ of 0.88 in the real-effort version of CTB studies compared to those using monetary rewards. We explore and discuss this difference below in Subsection 3.3.

From the I^2 statistic (Higgins and Thompson, 2002), we observe that 98% of the total variability in estimates from monetary CTB and 46% of the total variability in estimates from real-effort CTB are due to between-observation heterogeneity rather than sampling variance.[15] Note that estimates from real-effort CTB are less precisely estimated (i.e., they are associated with larger SEs) compared to those from monetary CTB.

[13] Online Appendix C.4 presents results with these three estimates included.

[14] Anticipating the amount of between-study heterogeneity in estimated *PB*, we directly jump to the random-effects model. Results from the common-effect specification are reported in Table C.1 in the Online Appendix.

[15] The I^2 statistic gives the amount of heterogeneity relative to the total amount of variance in the observed effects. Formally, the I^2 statistic is computed by

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + s^2} \times 100\%,$$

where $\hat{\tau}^2$ is the estimated value of τ^2 and

$$s^2 = \frac{(m-1)\sum w_j}{(\sum w_j)^2 + \sum w_j^2},$$

is the 'typical' sampling variance of the observed effect sizes, where $w_j = 1/\nu_j^2$.
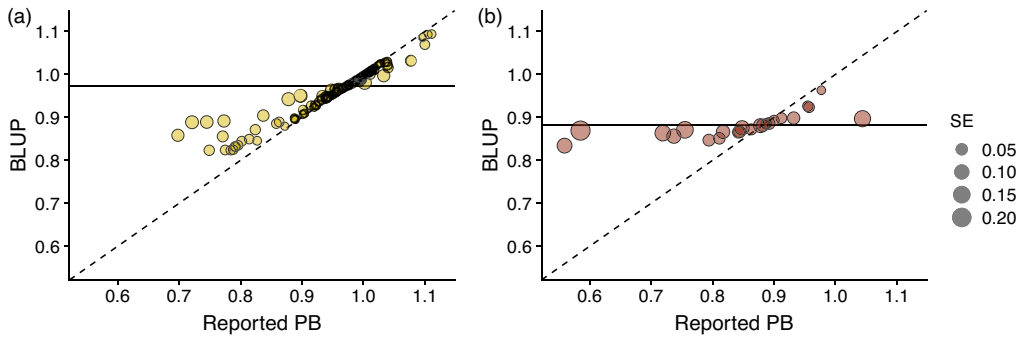
© 2021 Royal Economic Society.

Fig. 4. *Reported PB and the Corresponding Best Linear Unbiased Predictors. Panel (a): Studies with monetary rewards. Panel (b): Studies with real effort.*
Notes: Reported estimates PB_j are drawn proportionally to the associated SEs. Solid horizontal lines correspond to random-effects estimates (0.9723 and 0.8815).

Given the random-effects estimates $\overline{PB}_0^{\text{RE}}$ and the estimated degree of heterogeneity $\hat{\tau}^2$, we can construct the best linear unbiased predictors (BLUPs), also known as the empirical Bayes estimates:

$$\widehat{PB}_j = \omega_j \overline{PB}_0^{\text{RE}} + (1 - \omega_j)PB_j = \frac{\nu_j^2}{\nu_j^2 + \hat{\tau}^2}\overline{PB}_0^{\text{RE}} + \frac{\hat{\tau}^2}{\nu_j^2 + \hat{\tau}^2}PB_j,$$

where the weight ω_j captures the degree to which the estimates are 'pooled' together.[16] BLUPs lie between PB_j and $\overline{PB}_0^{\text{RE}}$, and the relative position depends on the size of sampling variability ν_j^2 and the degree of heterogeneity $\hat{\tau}^2$. Figure 4 demonstrates how 'shrinkage' of BLUPs works, especially for the less precise estimates.

As an alternative approach for handling statistically dependent PB estimates within each paper, we also apply a multi-level random-effects model (ML; Konstantopoulos, 2011; Van den Noortgate *et al.*, 2013).[17] Let PB_{ij} denote the jth estimate of the PB parameter from study i. The first level is $PB_{ij} = \mu_{ij} + \varepsilon_{ij}$, where μ_{ij} is the 'true' present-bias parameter and $\varepsilon_{ij} \sim \mathcal{N}(0, \nu_{ij}^2)$ for the jth estimate in study i. The second level is $\mu_{ij} = \lambda_i + \xi_{ij}^{(2)}$, where λ_i is the average present-biasedness in study i and $\xi_{ij}^{(2)} \sim \mathcal{N}(0, \tau_{(2)}^2)$. Finally, the third level is $\lambda_i = PB_0 + \xi_i^{(3)}$, where PB_0 is the population average of PB and $\xi_i^{(3)} \sim \mathcal{N}(0, \tau_{(3)}^2)$. These equations are combined into a single model:

$$PB_{ij} = PB_0 + \xi_{ij}^{(2)} + \xi_i^{(3)} + \varepsilon_{ij}.$$

A small value of $\tau_{(2)}^2$ indicates that the estimates are similar at the study level (i.e., there is little within-study variation of different estimates). A large $\tau_{(3)}^2$ suggests that the 'true' present-bias parameter varies considerably across studies. Under the typical assumption of $\text{Cov}(\tau_{(2)}^2, \tau_{(3)}^2) = \text{Cov}(\tau_{(2)}^2, \varepsilon_{ij}) = \text{Cov}(\tau_{(3)}^2, \varepsilon_{ij}) = 0$, we have $\mathbf{E}[PB_{ij}] = PB_0$.

---

[16] This is called the 'pooling factor' in Bayesian hierarchical modelling (Gelman and Hill, 2007; Meager, 2019).

[17] More precisely, we assume a 'three-level' model structure. The common-effect model (2) and the random-effects specification (3) described above can be seen as 'two-level' models where the first level is $PB_j = \mu_j + \varepsilon_j$ and the second levels are $\mu_j = PB_0$ for the common-effect model and $\mu_j = PB_0 + \xi_j$ for the random-effects model.

In this multi-level specification, estimates (presented in the even-numbered columns in Table 5) are close to the results from the random-effects approach discussed above. The overall average of *PB* is 0.95, which is statistically significantly different from one. While average *PB* from monetary studies is around 0.98, we cannot reject the null hypothesis of no present bias. Finally, effort CTB produces a smaller average *PB* of 0.88. The heterogeneity measures I^2 adjusted to the multi-level specification indicate that, in studies with monetary rewards, 98% of total variance is due to between-study heterogeneity. In the real-effort version of CTB studies, between-study heterogeneity is estimated to be less than 40%. In both cases, within-study heterogeneity is small.

Taken together, we find that the average value of *PB* measured by the CTB protocol is between 0.95 and 0.97. We do not observe statistically significant present bias, on average, in studies using monetary rewards, but those with real effort produce a smaller average *PB* of 0.88. Note that there is a genuine heterogeneity in estimates from monetary studies. Below we further explore this heterogeneity using meta-regression models with plausible moderator variables (Subsection 3.3).

3.2. *Identifying and Correcting for Selective Reporting*

This section provides a tentative answer to our second question: *Is there selective reporting or publication bias?*

Scientific cumulation of knowledge is thrown off track and slowed down by selective reporting or publication of results. The typical concern is when the sign or magnitude of a statistical relationship is strongly predicted by theory, or becomes conventionally believed after preliminary studies. Then new studies that derive an unpredicted or unconventional result may be under-reported or under-published. We will refer to this misproduction of results as *selective reporting* or *publication bias*.

There are several possible sources of selective reporting. One is conscious fraud. Another is '*p*-hacking', in which multiple analyses are run to get the expected effect (without accounting for multiple comparisons during the specification search). A third source is that scientists who discover a genuine contradictory effect (and do not *p*-hack their way out of it) may simply not report results in any form, such as a conference presentation or preprint; the contradictory effect ends up in a 'file drawer'. A fourth source is that even if scientists attempt to publish contradictory effects, journals may implicitly screen them out or encourage, in the review process, *p*-hacking.

For a single study it is very difficult to detect any of these kinds of selective reporting (except clumsy frauds). However, in a group of related studies there are ways to detect possible collective selective reporting.

The QHD model emerged to explain observed patterns of present-biased choices, including procrastination and challenges to self-control. Selective reporting would therefore seem most likely to exaggerate the number of studies estimating the present-bias parameter to be significantly below one, since an estimate of the present-bias parameter below one is consistent with preferences that could generate the observed pattern of present-biased choices that the QHD model is trying to capture.

The funnel plot provides a useful first step for detecting selective reporting (and counter-factually correcting for it). Selective reporting will lead to 'missing studies', which create an asymmetry in the funnel plot. Figure 1 presents suggestive evidence of selective reporting—there is an asymmetry, even though the magnitude may not be huge (see also Online Appendix Figure C.1, which presents funnel plots for monetary CTB and effort CTB separately).

Given the relatively large SEs of some of the studies in our sample, it is noticeable that we do not see as many studies as we might expect (in the aggregate) with an imprecise estimate of the present-bias parameter consistent with *future bias* ($PB > 1$). Since future bias is viewed as an 'unreasonable' finding (in light of voluminous evidence documenting $PB < 1$), the lack of such findings apparent from the funnel plot provides initial evidence that selective reporting may be an important factor in this literature.

A common procedure for detecting and correcting for publication selection bias is the FAT-PET procedure (Stanley and Doucouliagos, 2012; 2014).[18] In the absence of selective reporting, the reported estimates of the present-bias parameter should be uncorrelated with their SEs. In the presence of selective reporting, on the other hand, the reported estimates are correlated with their SEs (more imprecise estimates in the unconventional direction will go unreported). This motivates a simple regression model for detection of selective reporting:

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \varepsilon_{ij}, \tag{4}$$

where PB_{ij} and SE_{ij} are again the jth estimates of the present-bias parameter and their associated SEs reported in the ith study. In this model $\alpha_1 \neq 0$ captures the degree of selective reporting bias. The estimate of α_0 naturally serves as an estimate of the selection-corrected effect size (since it corresponds to an extrapolated effect size with zero SE and hence perfect precision). Note that the variance of ε_{ij} in this regression will vary across estimates. Therefore, it is often suggested to use weighted least squares (WLS) with the inverse of the variance of the study's estimate ($1/SE_{ij}^2$) as the weight (Stanley and Doucouliagos, 2012). This model allows us to test the asymmetry of the funnel plot (FAT; Egger *et al.*, 1997; Stanley, 2005; 2008) as well as whether there is a genuine effect beyond publication selection (PET). See Stanley and Doucouliagos (2012) and Stanley (2017) for in-depth discussion (especially on the limitations of these approaches).

Table 6 reports results from estimation of model (4) using the unrestricted WLS (Stanley and Doucouliagos, 2015). The estimated values of α_1 are negative, indicating that less precise (i.e., larger SE) studies are associated with lower estimates of PB (i.e., more present-biased). We do not reject the null hypothesis that the coefficient on SE is zero in studies with monetary rewards—columns (3)–(6)—while the relationship is statistically significant for studies using effort cost: columns (7)–(8). The intercept α_0 represents an estimate of 'true' underlying PB that has been corrected for selective reporting. The results indicate that the 'bias-corrected' estimate of PB is statistically indistinguishable from one, due to the strong relationship between reported PB estimates and their SEs.

It has been argued that the performance of commonly used bias-correction methods such as the FAT-PET procedure depends on the nature of the data, and no single method dominates the other in all circumstances (Alinaghi and Reed, 2018; Carter *et al.*, 2019; Hong and Reed, 2019). Therefore, we also report results from other bias-correction methods recently introduced in the literature.

We first apply the *latent studies* method for identification of and correction for selective reporting proposed by Andrews and Kasy (2019), which models the conditional probability of publication as a function of a study's results (discussed in detail in Online Appendix C.7).[19] The results are shown in the bottom panel of Table 6 (and Table C.6 in Online Appendix).

---

[18] This is an acronym for a combination of *funnel asymmetry test* (FAT) and *precision effect test* (PET).

[19] While Andrews and Kasy (2019) model conditional *publication* probabilities, our application of the method is intended to capture conditional *reporting* probabilities.

Table 6. *Funnel Plot Asymmetry and Precision Effect Testing.*

		All studies		Monetary (all)		Monetary ('neutral')		Effort cost	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SE of PB estimate	α_1	−1.4498	−0.3679	−1.3185	−0.2480	−1.6776	−0.1872	−2.0571	−1.8720
		(0.6187)	(0.3329)	(0.7260)	(0.3410)	(1.0459)	(0.3917)	(0.4412)	(0.1093)
Constant	α_0	1.0002		0.9998		1.0077		0.9931	
		(0.0032)		(0.0032)		(0.0056)		(0.0255)	
FAT ($H_0: \alpha_1 = 0$)	*p*-value	0.0265	0.2785	0.0852	0.4759	0.1261	0.6385	0.0016	0.0000
PET ($H_0: \alpha_0 = 1$)	*p*-value	0.9475		0.9393		0.1831		0.7931	
Study fixed effect		No	Yes	No	Yes	No	Yes	No	Yes
Observations		217	217	193	193	140	140	24	24
Number of studies		29	29	20	20	19	19	9	9
R^2		0.1823	0.8429	0.1400	0.8377	0.1777	0.9055	0.5100	0.9503
Adjusted R^2		0.1785	0.8186	0.1355	0.8189	0.1717	0.8906	0.4877	0.9183
Other bias-correction methods									
Latent-studies method	$\overline{PB}_0$	0.974		0.987		0.939		0.904	
		(0.040)		(0.051)		(0.064)		(0.016)	
Stem-based method	$\overline{PB}_0$	0.9910		0.9910		0.9992		0.9287	
		(0.0029)		(0.0029)		(0.0036)		(0.0274)	

Notes: Estimated by WLS. SEs are clustered at the study level. Three observations with large influence measure ($|DFBETAS| > 1$) are excluded. In the specification with study fixed effects, the constant term is dropped and all the dummy variables for the studies are included. Details of the latent-studies method and the stem-based method are presented in Online Appendices C.7 and C.8, respectively.

Although none of the relative reporting probabilities for estimates with different intervals of Z-values ($Z < -1.96$, $-1.96 \leq Z < 0$, $0 \leq Z < 1.96$) is individually significantly different from one, joint tests show evidence that, for monetary studies, selective reporting acts to 'squeeze' *PB* estimates towards one from both sides (so in this case, selective reporting acts to hide, instead of exaggerate, statistically significant findings). For effort studies, selective reporting does in fact cause over-representation of statistically significant estimates of present bias, consistent with our FAT-PET results. That being said, the degree of selective reporting is not drastic—the adjusted study estimates from the latent studies model are very similar to the original study estimates (shown in Figure C.26 of the Online Appendix).

Finally, we apply the *stem-based bias-correction method* developed by Furukawa (2019); (adapting Stanley *et al.*, 2010), which is discussed in more detail in Online Appendix C.8. Intuitively, this method provides a weighted average of the estimates from an optimally chosen subset of the most precise studies. The results show insignificant aggregate evidence for present bias across the most precise studies. However, when only studies in which subjects make decisions over allocations of effort are included, we find significant levels of present bias, as shown in the bottom panel of Table 6 (and Table C.7 in the Online Appendix).

Taken together, we view our results as demonstrating that there is evidence suggesting the existence of modest selective reporting in the direction of over-reporting *PB* < 1 in studies using a real-effort task. Correcting for potential selective reporting pushes values of average *PB* upward towards one. They are still close to one for monetary studies. For effort studies, values are still lower than those for monetary studies, but the estimated degree of present-biasedness depends on the method used for bias correction.

3.3. *Explaining Heterogeneity*

We have thus far assumed that the variability in reported estimates is due mainly to sampling errors—at either the observation level or the study level, or both—and potential selective reporting. However, these estimates come from studies that use a variety of experimental designs, participants and econometric approaches, which may result in systematic variation in reported estimates.[20] This section provides a tentative answer to our third question: *How does reported PB vary reliably with observable study characteristics?*

In order to explain heterogeneity, we now add a set of moderator variables to model (4):

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \boldsymbol{\gamma} \cdot \boldsymbol{x}_{ij} + \varepsilon_{ij}, \qquad (5)$$

where $\boldsymbol{x}_{ij}$ is a vector of observable characteristics of the jth estimate from study i and $\boldsymbol{\gamma}$ is a coefficient vector.

In the first set of meta-regressions, presented in Table 7, we restrict samples to those using monetary rewards. We consider six basic sets of moderators as $\boldsymbol{x}_{ij}$. These variables are categorized into: treatment dummy (omitted category is *Neutral condition*), location of the experiment (omitted category is *Location: lab*), timing of immediate reward payment (omitted category is *by the end of the experiment*), estimation method (omitted category is *Estimation: least squares*), treatment of background (b.g.) consumption (omitted category is *Estimation: no b.g. consumption*) and interface (omitted category is *Computerised*). We also

---

[20] Online Appendix Figures C.7–C.17 visualize the effects of some representative study characteristics on reported estimates, looking at each characteristic in isolation.

Table 7. *Explaining the Heterogeneity of Reported Estimates (Studies with Monetary Rewards).*

	(1)	(2)	(3)	(4)	(5)	(6)
SE of *PB* estimate	−1.248**	−1.951**	−1.711*	−0.969*	−1.104*	−0.600
	(0.454)	(0.636)	(0.668)	(0.461)	(0.550)	(0.512)
Non-neutral condition	−0.006	−0.003	−0.006	−0.002	−0.004	−0.001
	(0.005)	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)
Location: field	0.066**	0.071**	0.090***	0.221***	0.184***	0.107**
	(0.025)	(0.022)	(0.025)	(0.048)	(0.052)	(0.038)
Location: class	0.011	0.022	0.029*	0.098**	0.049*	−0.007
	(0.013)	(0.013)	(0.014)	(0.035)	(0.022)	(0.028)
Location: online	−0.010	−0.031*	−0.026	−0.019	−0.016	−0.016
	(0.005)	(0.014)	(0.016)	(0.012)	(0.013)	(0.014)
'Immediate' pay: within day	0.048**	0.050***	0.051***	0.030**	0.019	0.027
	(0.017)	(0.012)	(0.014)	(0.009)	(0.014)	(0.019)
'Immediate' pay: not reported	−0.066	−0.060	0.046	−0.080	−0.148	−0.038
	(0.056)	(0.051)	(0.065)	(0.045)	(0.085)	(0.070)
Delivery: cash	0.029	0.017	0.024	0.009	−0.002	0.009
	(0.021)	(0.018)	(0.018)	(0.013)	(0.018)	(0.019)
Delivery: bank	−0.004*	−0.003	−0.006	0.045**	0.004	−0.030
	(0.002)	(0.004)	(0.005)	(0.016)	(0.008)	(0.017)
Delivery: other	−0.008	−0.008*	−0.011**	0.013	−0.001	−0.012*
	(0.004)	(0.004)	(0.003)	(0.008)	(0.005)	(0.005)
Estimation: Tobit		0.018*	0.016	0.005	0.002	0.010
		(0.009)	(0.009)	(0.006)	(0.009)	(0.008)
Estimation: other		−0.002	−0.001	−0.005	−0.014	−0.002
		(0.006)	(0.006)	(0.008)	(0.010)	(0.008)
Estimation: b.g. consumption		−0.001	−0.001	−0.000	−0.003	0.007
		(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Deal uncertainty			−0.005	0.025**	0.008	0.004
			(0.004)	(0.009)	(0.007)	(0.008)
Deal transaction cost			0.111**	0.084**	0.011	0.077
			(0.038)	(0.031)	(0.052)	(0.054)
Paper and pencil			−0.017	−0.044	−0.031	−0.025
			(0.013)	(0.031)	(0.019)	(0.017)
Credit card				0.183***		
				(0.051)		
Withdrawal					0.308*	
					(0.131)	
Emergency funds impossible						−0.216
						(0.123)
Constant	0.963***	0.963***	0.854***	0.764***	0.690***	0.966***
	(0.017)	(0.014)	(0.052)	(0.048)	(0.089)	(0.099)
Observations	193	193	193	193	193	193
R^2	0.457	0.500	0.523	0.718	0.657	0.597
Adjusted R^2	0.427	0.464	0.480	0.691	0.623	0.557

Notes: Observations with large influence measure ($|DFBETAS| > 1$) are excluded. Study fixed effects are not included in the model. SEs are clustered at the study level. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

include several additional variables that are specific to experiments involving monetary rewards: method of reward delivery (omitted category is *Delivery: cheque*); treatment of confounding factors, such as uncertainty regarding future reward and transaction costs (omitted category is *Ignored* in both variables); and proxies for ease of access to financial markets at the country level. We estimate the model using unrestricted WLS (Stanley and Doucouliagos, 2017).

The effects of study characteristics on the estimated *PB* parameter exhibit notable patterns. Regression coefficients reported in Table 7 (focusing on the first three columns for

now) suggest that: (*i*) field experiments tend to find lower levels of present-biased prefer-
ences compared to lab studies; (*ii*) dealing with transaction costs makes the estimated *PB*
larger; (*iii*) we do not observe systematic effects of the reward-delivery method; (*iv*) we
do not observe systematic effects of econometric approaches (e.g., Tobit or NLS); and (*v*)
whether or not to jointly estimate background consumption has little impact on the estimates
of *PB*.

Compared to studies that guaranteed to deliver the 'immediate' rewards within the day of the
experiment, the estimated *PB* is smaller (more present-biased) when these 'immediate' rewards
were delivered by the end of the experiment. It is possible to reason that, if the 'immediate'
reward is paid at the end of the experiment, that design increases uncertainty about future
payments, which in turn exaggerates the behavioural *PB* measure (in the direction of the larger
present-biasedness). To examine this potential confounding factor, we include the dummy *Deal
uncertainty* in a regression specification in column (3) aiming to control for the confidence of
future payment delivery. We find that this dummy itself has a statistically insignificant coefficient,
while the dummy *'Immediate' payment: within day* continues to have a statistically significant
effect, with its magnitude virtually unchanged compared to column (2). These observations
suggest that the timing of 'immediate' (i.e., $t = 0$) payment appears to matter, as documented in
Balakrishnan *et al.* (2017).

3.3.1. *Comparing monetary and non-monetary rewards*

Underlying models of inter-temporal choices are fundamentally about utility flows at each time
period, and not about the receipt of monetary payments. A large share of existing empirical
studies have measured time preferences using time-dated monetary payments, but additional
assumptions (such as monetary payments being 'consumed' at the time of receipt) are necessary
to infer individuals' discount functions from observed choices in this approach (Mulligan, 1996;
Chabris *et al.*, 2008; Cohen *et al.*, 2020). More recent studies try to directly control the timing
of utility flow using, for example, real-effort tasks (e.g., Augenblick *et al.*, 2015; Carvalho *et al.*,
2016; Augenblick, 2018; Fedyk, 2018; Augenblick and Rabin, 2019), and report evidence that
non-monetary rewards provide estimates of the present-bias parameter that are smaller (in the
sense of conveying greater levels of present bias) than those from the standard monetary reward
studies.

Building on this discussion, our next set of meta-regressions directly compares *PB* estimates
from studies with monetary and non-monetary rewards, correcting for selective reporting and
several study characteristics, to see whether the apparent difference in present bias is evident
from CTB alone. We set up a general regression model:

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \alpha_2 \cdot SE_{ij}^2 + \boldsymbol{\gamma} \cdot \boldsymbol{x}_{ij} + \lambda_1 \cdot (SE_{ij} \cdot z_{ij}) + \lambda_2 \cdot (SE_{ij}^2 \cdot z_{ij}) + \varepsilon_{ij},$$

which extends (5) to allow for any factors that can potentially influence selective reporting
(captured by $SE_{ij} \cdot z_{ij}$ and $SE_{ij}^2 \cdot z_{ij}$). We include a dummy for monetary studies and its interaction
with several study characteristics, so that the constant term (α_0) captures the average *PB* estimate
from non-monetary studies.

Table 8 reports the results. The main variable of interest is the coefficient on the dummy *Reward:
money*, which captures the difference between the average *PB* from non-monetary studies and that
from the 'baseline' monetary studies. The definition of 'baseline' studies is: 'monetary studies,

Table 8. *Explaining the Heterogeneity of Reported Estimates (Monetary vs. Non-Monetary Rewards).*

	(1)	(2)	(3)	(4)	(5)	(6)
Constant (*PB* from effort CTB)	0.907***	0.907***	0.993***	0.993***	0.932***	0.932***
	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)
SE of *PB* estimates			−2.057***	−2.057***		
			(0.414)	(0.414)		
SE^2 of *PB* estimates					−10.918***	−10.918***
					(2.829)	(2.829)
Reward: money	0.089***	0.093***	0.015	0.016	0.068**	0.069**
	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)
× Non-neutral condition	−0.003	−0.012	−0.011**	−0.006	−0.007*	−0.010
	(0.004)	(0.007)	(0.004)	(0.006)	(0.003)	(0.006)
× Location: field		0.057**		0.071***		0.064**
		(0.021)		(0.018)		(0.020)
× Location: class		0.026		0.037***		0.031*
		(0.017)		(0.011)		(0.015)
× Location: online		0.004		−0.026		−0.006
		(0.009)		(0.014)		(0.010)
× 'Immediate': by end of experiment		−0.039*		−0.042***		−0.041**
		(0.017)		(0.011)		(0.015)
× 'Immediate': not reported		−0.127*		−0.112*		−0.113*
		(0.060)		(0.051)		(0.052)
× Estimation: Tobit		0.002		0.019*		0.009
		(0.006)		(0.009)		(0.007)
× Estimation: other		−0.005		−0.002		−0.004
		(0.005)		(0.004)		(0.004)
× *SE* of *PB* estimates			0.374	0.065		
			(0.854)	(0.708)		
× SE^2 of *PB* estimates					−36.379	−26.427*
					(22.497)	(13.157)
Observations	217	217	217	217	217	217
R^2	0.054	0.375	0.249	0.504	0.222	0.456
Adjusted R^2	0.045	0.348	0.235	0.478	0.207	0.427
H_0: $PB_{effort} = 1$	$p = 0.0004$		$p = 0.7747$		$p = 0.0078$	

Notes: Observations with large influence measure ($|DFBETAS| > 1$) are excluded. Study fixed effects are not included in the model. SEs are clustered at the study level. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

neutral condition' in the odd columns, and 'monetary studies, neutral condition, lab, immediate rewards delivered within the day, estimation with NLS' in the even columns.[21]

As discussed in the literature, studies using non-monetary rewards estimate present-bias parameters that are generally smaller than those from the standard monetary reward studies, regardless of the definition of the baseline in monetary studies: columns (1)–(2). The other specifications include either *SE* or SE^2, as well as its interaction with *Reward: money*. The estimated coefficients on *Reward: money* are not statistically significant when *SE* is included, but are significantly positive when SE^2 is used. These results suggest that the difference between average *PB* from monetary and non-monetary studies shrinks when potential selective reporting is corrected for. However, the size of this difference $PB_{money} - PB_{effort}$ depends on the assumption imposed on the relationship between reported *PB* and *SE*.

---

[21] In the meta-regression models presented in Table 8, we do not include dummy variables for design characteristics in non-monetary studies. This is due solely to a power issue—there are only 24 estimates from nine effort-CTB studies in our data set. It is therefore important to revisit these meta-regression analyses after the literature has accumulated more estimates from CTB studies using non-monetary rewards.

3.3.2. *Access to financial markets*

The literature has discussed several drawbacks to the use of dated monetary payments (Coller and Williams, 1999; Cubitt and Read, 2007; Cohen *et al.*, 2020). For example, in the presence of complete financial markets (and an associated lack of borrowing constraints), an optimizing individual's rate of time preference over monetary rewards should be equal to the market interest rate at which the individual can save and borrow.

To explore whether access to credit markets and liquid savings may drive a wedge between utility parameters estimated in money-earlier-or-later studies and the 'true' parameters associated with the individual's discounting of utility flows, we explore the extent to which lack of access to financial markets and methods of saving/borrowing is associated with the display of greater levels of present bias in studies with monetary reward. (Andreoni *et al.*, 2018, refer to this potential effect as the arbitrage channel.)

The data we use are from the World Bank's Global Findex database (Demirgüç-Kunt *et al.*, 2018).[22] The data set consists of nationally representative samples of adults from 140 countries and focuses on the ability of adults in different countries to access financial services. We focus on three variables for our analysis to proxy for the ease of access to financial markets for experimental participants (see Figures C.18–C.20 in the Online Appendix):

(*i*) *Proportion of the adult population with a credit card.* Credit cards provide a relatively cheap way for consumers to borrow against future income, generally up to 30 days in the future. We expect that having access to a relatively easy source of borrowed funds will cause individuals to display less present-biased behaviour in money-later-or-earlier tasks.

(*ii*) *Proportion of the adult population who have made a withdrawal from a financial institution account in the last year.*[23] Since this proxies having access to liquid savings in the current period, we expect that individuals who have made withdrawals within the last year are less likely to display present-biasedness in money-earlier-or-later tasks, since if they desired to increase their current-period consumption, they could withdraw from their savings accounts instead of relying on the current-period reward from the experiment.

(*iii*) *Proportion of the adult population who would not be able to come up with emergency funds within the next month.*[24] We expect this to have a positive relationship with observed present-biasedness in money-earlier-or-later studies, since individuals who are unable to come up with emergency funds within the next month are more likely to have consumption closely following income in each period, and so monetary flows may more accurately proxy true utility flows for these individuals (provided other sources of income remain constant over time).

Columns (4)–(6) in Table 7 show the results. The coefficients have signs in the expected direction, and two of the variables—*Credit card* and *Withdrawal*—have statistically significant positive effects: studies conducted in countries/regions where more individuals have easier access to financial markets (through credit cards or withdrawals from liquid savings accounts) tend to exhibit less present-biasedness. These results indicate that some part of the observed heterogeneity in estimated *PB* can be attributed to the degree to which individuals have access to financial

---

[22] These data can be accessed at: https://globalfindex.worldbank.org/ (last accessed: 5 November 2020).

[23] A financial institution is defined by Demirgüç-Kunt *et al.* (2018) as a bank or another type of financial institution, such as a credit union, a microfinance institution, a cooperative, or the post office, or having a debit card.

[24] 'Emergency funds' are defined as 5% of gross national income per capita in the local currency.

markets, and that the arbitrage channel discussed by Andreoni *et al.* (2018) has some effect on the estimated *PB* in money-earlier-or-later studies.

3.3.3. *Model uncertainty*

The selection of variables and the order of inclusion in the first meta-regression analysis presented in Table 7 are based on prior discussion in the literature as well as co-occurrence of study characteristics in the data (Figures C.5 and C.6 in the Online Appendix), and thus made somewhat arbitrarily.

We augment our meta-regression analysis with the application of Bayesian model averaging (BMA) to tackle the model uncertainty resulting from the large number of explanatory variables we could have included in our meta-regression model (Hoeting *et al.*, 1999; Moral-Benito, 2015; Steel, 2020). BMA runs multiple regressions with different subsets of the explanatory variables (models) and marginalizes over models to obtain the posterior density of the parameters. We provide a more detailed explanation in Online Appendix C.6.[25]

The results of our application of BMA are in line with those reported in Table 7. Figure 5 is representative of our results (the full set of results is provided in Section C.6 of the Online Appendix).

4. Conclusion

We present a quantitative meta-analysis of estimates of the present-bias parameter in the QHD model using choice data from CTB experiments. We collected 220 estimates from 28 articles and found that the meta-analytic average of the present-bias parameter was between 0.95 and 0.97, which is statistically significantly smaller than one. The values for monetary-reward studies were close to one, indicating absence of present bias, on average. On the contrary, effort-based studies reported a lower meta-analytic average of 0.88, a statistically significant present bias. There is evidence suggesting selective reporting in studies using a real-effort CTB, and bias-corrected estimates of average *PB* vary from 0.90–0.93 to 0.99, depending on the method used for correction.

We also found that estimates varied greatly across studies, due primarily to their different study characteristics. Our meta-regression analysis suggests that CTB experiments with non-monetary rewards indeed found estimates that were 'more present-biased' than those from CTB studies with typical monetary rewards. One reason for this difference suggested in the literature relates to decision-makers arbitraging monetary payments using market interest rates, so that allocations in monetary CTB studies were in fact not representative of underlying parameters in the decision-maker's utility function (Cohen *et al.*, 2020). We found evidence that access to financial markets is in fact associated with estimates of *PB* close to one (as opposed to estimates showing present-biased behaviour), suggesting that this arbitrage channel does play at least some role in explaining results from monetary CTB designs.

Furthermore, we found evidence to confirm the importance of the delay until the issue of the reward associated with the 'current' period (e.g., Balakrishnan *et al.*, 2017; Ericson and Laibson, 2019); across a range of specifications in both our meta-regression and BMA approaches, studies that delivered rewards associated with the 'current' period by the end of the experiment, as

---

[25] For applications of BMA in meta-analysis in economics, see Iršová and Havránek (2013) and Havránek *et al.* (2015; 2017).
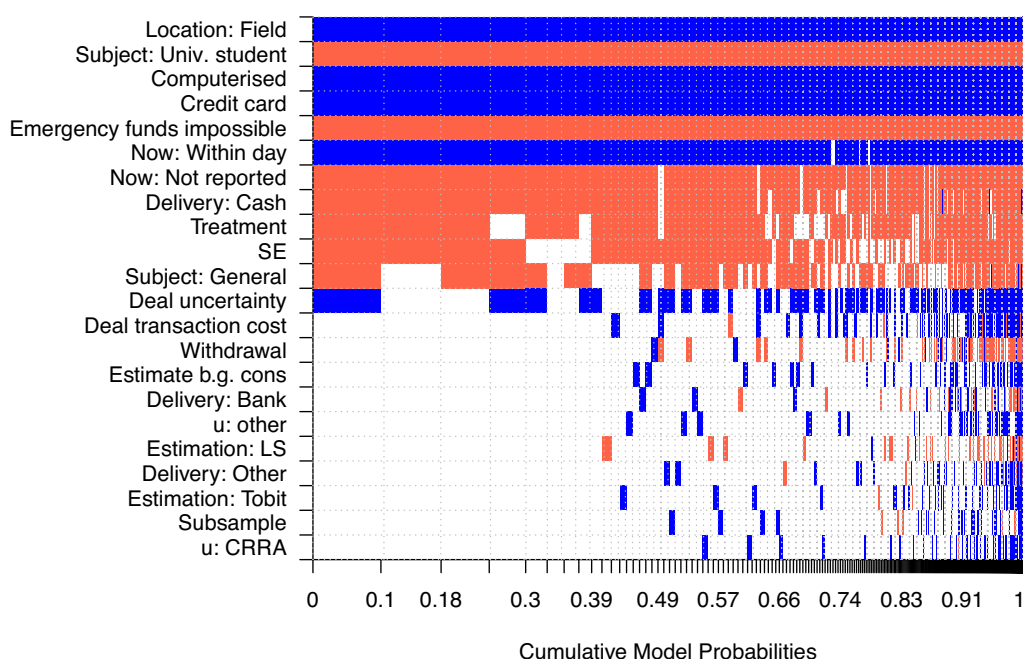
Model Inclusion Based on Best 616 Models

Fig. 5. *Model Inclusion Results from Bayesian Model Averaging.*
Notes: In this figure, observations are from monetary CTB studies only. Columns denote individual models in which variables are sorted by posterior model probability in descending order. Blue cells (darker cells in greyscale) indicate that the variable is included in the model and has a positive coefficient, while red cells (lighter cells in greyscale) indicate that the variable has a negative coefficient. White cells indicate that the variable is not included in the model.

opposed to only by the end of the day, tended to yield lower estimates of the present-bias parameter, indicating greater levels of present bias in the behaviour of subjects.

In addition, we found suggestive evidence concerning the importance for estimates of present bias of a factor that has so far not been widely discussed: the location of the study—whether it takes place in a laboratory or in the field. Both meta-regression and BMA suggest that subjects in laboratory experiments show larger present bias than subjects in field experiments. Many studies followed Andreoni and Sprenger's (2012) original econometric strategy and reported estimates using both NLS and Tobit (or estimates with and without background consumption). These methods have ignited significant debate in the literature (see, for example, the discussion in Andreoni *et al.*, 2015). However, our meta-analysis shows that the econometric strategy makes little difference.

Our findings naturally lead to two follow-on questions. First, as well as being statistically different from one, are deviations of estimated *PB* from one significantly different from one in a *practical* sense? Second, given that many study characteristics have a systematic influence on the estimated degree of present bias, is there is a preferred method for eliciting present bias?

Regarding the first question, at least in the setting of effort, where we estimate *PB* to be roughly 0.90–0.93, present bias seems to be a first-order modelling concern. With a per-period *discount*

© 2021 Royal Economic Society.

rate of 4%, and a present bias of 0.9, the effective first-period discount rate is roughly 15%. Such a discrepancy between the effective first-period discount rate and the discount rate for subsequent periods is substantial enough that it should merit consideration when analysing, for example, individual behaviour in the workplace under different contracts (Kaur *et al.*, 2015) or common self-control problems.

Regarding the second question—an extremely challenging and important one—we do not think our meta-analysis is capable of identifying a preferred method for eliciting present bias. The meta-regression effects can only tell us which methods produce estimates that are reliably different from those produced by others.

Approaching this question forces one to take a stand on the conceptual status of present bias. Is it a stable trait, and, if so, is there an ideal method that would come as close to the true value as possible? Or does expressed present bias change according to elicitation method *and*, very likely, in different natural-choice domains? The latter view is expressed by Frederick and Loewenstein (2008, p. 233), who write:

> *Like others . . . our findings suggest that respondents possess a variety of cognitive schemas, each of which can be evoked or suppressed by subtle contextual features. Thus, we believe that the major challenge for decision researchers lies not in honing parametric specifications, but in acquiring a broader understanding of the varied constituents of preferences and the problem representations that bring them to the fore.*

We suggest a middle path between the stable-trait view and the contextualist view. In psychometric language, measuring trait-like quantities well aspires to achieve two goals: reliability and validity.

Reliability means that there is low measurement error. For example, test-retest reliability means that a person answers the same question the same way, if it is asked twice.[26] 'Construct' validity is how well a measure is associated to a general construct.

We think that economic validity is best operationalized as good generalizability from one type of *PB* estimate to a different behaviour that is thought (theoretically) to be correlated with *PB*. Ideal examples of this are studies in which laboratory or survey estimates are associated with natural behaviours at the individual level. For example, Meier and Sprenger (2010) measured *PB* using CTB and found it to be correlated with credit card use and debt level.[27]

In sum, we think the criteria for a preferred method are high reliability and good economic validity. Unfortunately, our current meta-analysis cannot measure either of these criteria well, but more ambitious studies linking estimation and natural-data observation could do so, and certainly should.

LMU Munich, Germany
Opportunity Insights, USA
California Institute of Technology, USA

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

---

[26] Preferably, the retest is created with no ability to clearly recall the first answer, as in the footnote 12 discussion about artificial consistency.

[27] In a less related example, Chapman *et al.* (2018) compared a price-list method and an adaptive method for measuring risk attitudes. They found that the adaptive method appeared to have lower measurement error and, as a result, high correlations with plausible demographic variables.

References

Alinaghi, N. and Reed, W.R. (2018). 'Meta-analysis and publication bias: how well does the FAT-PET-PEESE procedure work?', *Research Synthesis Methods*, vol. 9(2), pp. 285–311.

Andreoni, J., Gravert, C., Kuhn, M.A., Saccardo, S. and Yang, Y. (2018). 'Arbitrage or narrow bracketing? On using money to measure intertemporal preferences', Working Paper No. 25232, National Bureau of Economic Research.

Andreoni, J., Kuhn, M.A. and Sprenger, C. (2015). 'Measuring time preferences: a comparison of experimental methods', *Journal of Economic Behavior & Organization*, vol. 116, pp. 451–64.

Andreoni, J. and Sprenger, C. (2012). 'Estimating time preferences from convex budgets', *American Economic Review*, vol. 102(7), pp. 3333–56.

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

Angeletos, G.M., Laibson, D., Repetto, A., Tobacman, J. and Weinberg, S. (2001). 'The hyperbolic consumption model: calibration, simulation, and empirical evaluation', *Journal of Economic Perspectives*, vol. 15(3), pp. 47–68.

Ashraf, N., Karlan, D. and Yin, W. (2006). 'Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines', *Quarterly Journal of Economics*, vol. 121(2), pp. 635–72.

Augenblick, N. (2018). 'Short-term time discounting of unpleasant tasks', Unpublished manuscript.

Augenblick, N., Niederle, M. and Sprenger, C. (2015). 'Working over time: dynamic inconsistency in real effort tasks', *Quarterly Journal of Economics*, vol. 130(3), pp. 1067–115.

Augenblick, N. and Rabin, M. (2019). 'An experiment on time preference and misprediction in unpleasant tasks', *Review of Economic Studies*, vol. 86(3), pp. 941–75.

Balakrishnan, U., Haushofer, J. and Jakiela, P. (2017). 'How soon is now? Evidence of present bias from convex time budget experiments', Working Paper No. 23558, National Bureau of Economic Research.

Barcellos, S.H. and Carvalho, L. (2014). 'Information about self-control and intertemporal choices', Unpublished manuscript.

Barton, B. (2015). 'Interpersonal time inconsistency and commitment', Unpublished manuscript.

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Hoboken, NJ: Wiley.

Beshears, J., Choi, J.J., Clayton, C., Harris, C., Laibson, D. and Madrian, B.C. (2020). 'Optimal illiquidity', Working Paper No. 27459, National Bureau of Economic Research.

Beshears, J., Choi, J.J., Harris, C., Laibson, D., Madrian, B.C. and Sakong, J. (2015). 'Self control and commitment: can decreasing the liquidity of a savings account increase deposits?', Working Paper No. 21474, National Bureau of Economic Research.

Bisin, A. and Hyndman, K. (2020). 'Present-bias, procrastination and deadlines in a field experiment', *Games and Economic Behavior*, vol. 119, pp. 339–57.

Bollen, K.A. and Jackman, R.W. (1985). 'Regression diagnostics: an expository treatment of outliers and influential cases', *Sociological Methods & Research*, vol. 13(4), pp. 510–42.

Carter, E.C., Schönbrodt, F.D., Gervais, W.M. and Hilgard, J. (2019). 'Correcting for bias in psychology: a comparison of meta-analytic methods', *Advances in Methods and Practices in Psychological Science*, vol. 2(2), pp. 115–44.

Carvalho, L.S., Meier, S. and Wang, S.W. (2016). 'Poverty and economic decision-making: evidence from changes in financial resources at payday', *American Economic Review*, vol. 106(2), pp. 260–84.

Chabris, C.F., Laibson, D.I. and Schuldt, J.P. (2008). 'Intertemporal choice', in (S.N. Durlauf and L.E. Blume, eds.), *The New Palgrave Dictionary of Economics*, pp. 168–77, London: Palgrave Macmillan.

Chapman, J., Snowberg, E., Wang, S. and Camerer, C. (2018). 'Loss attitudes in the U.S. population: evidence from dynamically optimized sequential experimentation (DOSE)', Working Paper No. 25072, National Bureau of Economic Research.

Cohen, J., Ericson, K., Laibson, D. and White, J. (2020). 'Measuring time preferences', *Journal of Economic Literature*, vol. 58(2), pp. 299–347.

Coller, M. and Williams, M.B. (1999). 'Eliciting individual discount rates', *Experimental Economics*, vol. 2(2), pp. 107–27.

Cubitt, R.P. and Read, D. (2007). 'Can intertemporal choice experiments elicit time preferences for consumption?', *Experimental Economics*, vol. 10(4), pp. 369–89.

DellaVigna, S. and Malmendier, U. (2006). 'Paying not to go to the gym', *American Economic Review*, vol. 96(3), pp. 694–719.

Demirgüç-Kunt, A., Klapper, L., Singer, D., Ansar, S. and Hess, J. (2018). *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*, Washington, DC: World Bank Group.

DerSimonian, R. and Laird, N. (1986). 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, vol. 7(3), pp. 177–88.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. (1997). 'Bias in meta-analysis detected by a simple, graphical test', *BMJ*, vol. 315(7109), pp. 629–34.

Ericson, K.M. and Laibson, D. (2019). 'Intertemporal choice', in (B.D. Bernheim, S. DellaVigna and D. Laibson, eds.), *Handbook of Behavioral Economics: Foundations and Applications*, pp. 1–67, Amsterdam: Elsevier/North-Holland.

Fang, H. and Wang, Y. (2015). 'Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions', *International Economic Review*, vol. 56(2), pp. 565–96.

Fedyk, A. (2018). 'Asymmetric naïveté: beliefs about self-control', Unpublished manuscript.

Frederick, S. and Loewenstein, G. (2008). 'Conflicting motives in evaluations of sequences', *Journal of Risk and Uncertainty*, vol. 37(2), pp. 221–35.

Frederick, S., Loewenstein, G. and O'Donoghue, T. (2002). 'Time discounting and time preference: a critical review', *Journal of Economic Literature*, vol. 40(2), pp. 351–401.

Furukawa, C. (2019). 'Publication bias under aggregation frictions: theory, evidence, and a new correction method', Unpublished manuscript.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge: Cambridge University Press.

Glass, G.V. (1976). 'Primary, secondary, and meta-analysis of research', *Educational Researcher*, vol. 5(10), pp. 3–8.

Gruber, J. and Kőszegi, B. (2001). 'Is addiction "rational"? Theory and evidence', *Quarterly Journal of Economics*, vol. 116(4), pp. 1261–303.

Gurevitch, J., Koricheva, J., Nakagawa, S. and Stewart, G. (2018). 'Meta-analysis and the science of research synthesis', *Nature*, vol. 555(7695), pp. 175–82.

Halevy, Y. (2014). 'Some comments on the use of monetary rewards in the measurement of time preferences', Unpublished manuscript.

Havránek, T., Horváth, R., Iršová, Z. and Rusnak, M. (2015). 'Cross-country heterogeneity in intertemporal substitution', *Journal of International Economics*, vol. 96(1), pp. 100–18.

Havránek, T., Rusnak, M. and Sokolova, A. (2017). 'Habit formation in consumption: a meta-analysis', *European Economic Review*, vol. 95, pp. 142–67.

Hedges, L.V., Tipton, E. and Johnson, M.C. (2010). 'Robust variance estimation in meta-regression with dependent effect size estimates', *Research Synthesis Methods*, vol. 1(1), pp. 39–65.

Heidhues, P. and Kőszegi, B. (2010). 'Exploiting naïvete about self-control in the credit market', *American Economic Review*, vol. 100(5), pp. 2279–303.

Higgins, J.P.T. and Thompson, S.G. (2002). 'Quantifying heterogeneity in a meta-analysis', *Statistics in Medicine*, vol. 21(11), pp. 1539–58.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). 'Bayesian model averaging: a tutorial', *Statistical Science*, vol. 14(4), pp. 382–401.

Holt, C.A. and Laury, S.K. (2002). 'Risk aversion and incentive effects', *American Economic Review*, vol. 92(5), pp. 1644–55.

Hong, S. and Reed, W.R. (2019). 'A performance analysis of some new meta-analysis estimators designed to correct publication bias', Unpublished manuscript.

Imai, T. and Camerer, C.F. (2018). 'Estimating time preferences from budget set choices using optimal adaptive design', Unpublished manuscript.

Imai, T., Rutter, T.A. and Camerer, C.F. (2018). 'Meta-analysis of estimation of time discounting of rewards', Unpublished manuscript.

Imas, A., Kuhn, M. and Mironova, V. (2018). 'Waiting to choose', Unpublished manuscript.

Iršová, Z. and Havránek, T. (2013). 'Determinants of horizontal spillovers from FDI: evidence from a large meta-analysis', *World Development*, vol. 42, pp. 1–15.

John, A. (2020). 'When commitment fails: evidence from a field experiment', *Management Science*, vol. 66(2), pp. 503–29.

Kaur, S., Kremer, M. and Mullainathan, S. (2010). 'Self-control and the development of work arrangements', *American Economic Review: Papers & Proceedings*, vol. 100(2), pp. 624–28.

Kaur, S., Kremer, M. and Mullainathan, S. (2015). 'Self-control at work', *Journal of Political Economy*, vol. 123(6), pp. 1227–77.

Konstantopoulos, S. (2011). 'Fixed effects and variance components estimation in three-level meta-analysis', *Research Synthesis Methods*, vol. 2(1), pp. 61–76.

Koopmans, T.C. (1960). 'Stationary ordinal utility and impatience', *Econometrica*, vol. 28(2), pp. 287–309.

Laibson, D. (1997). 'Golden eggs and hyperbolic discounting', *Quarterly Journal of Economics*, vol. 112(2), pp. 443–78.

Liu, E.M., Meng, J. and Wang, J.T.Y. (2014). 'Confucianism and preferences: evidence from lab experiments in Taiwan and China', *Journal of Economic Behavior & Organization*, vol. 104, pp. 106–22.

Matoušek, J., Havránek, T. and Iršová, Z. (2020). 'Individual discount rates: a meta-analysis of experimental evidence', Unpublished manuscript.

Meager, R. (2019). 'Understanding the average impact of microcredit expansions: a Bayesian hierarchical analysis of seven randomized experiments', *American Economic Journal: Applied Economics*, vol. 11(1), pp. 57–91.

Meier, S. and Sprenger, C. (2010). 'Present-biased preferences and credit card borrowing', *American Economic Journal: Applied Economics*, vol. 2(1), pp. 193–210.

Moral-Benito, E. (2015). 'Model averaging in economics: an overview', *Journal of Economic Surveys*, vol. 29(1), pp. 46–75.

Mulligan, C. (1996). 'A logical economist's argument against hyperbolic discounting', Unpublished manuscript.

O'Donoghue, T. and Rabin, M. (1999). 'Doing it now or later', *American Economic Review*, vol. 89(1), pp. 103–24.

O'Donoghue, T. and Rabin, M. (2001). 'Choice and procrastination', *Quarterly Journal of Economics*, vol. 116(1), pp. 121–60.

O'Donoghue, T. and Rabin, M. (2015). 'Present bias: lessons learned and to be learned', *American Economic Review*, vol. 105(5), pp. 273–79.

Phelps, E.S. and Pollak, R.A. (1968). 'On second-best national saving and game-equilibrium growth', *Review of Economic Studies*, vol. 35(2), pp. 185–99.

Samuelson, P.A. (1937). 'A note on measurement of utility', *Review of Economic Studies*, vol. 4(2), pp. 155–61.

Stanley, T.D. (2001). 'Wheat from chaff: meta-analysis as quantitative literature review', *Journal of Economic Perspectives*, vol. 15(3), pp. 131–50.

Stanley, T.D. (2005). 'Beyond publication bias', *Journal of Economic Surveys*, vol. 19(3), pp. 309–45.

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*, vol. 70(1), pp. 103–27.

Stanley, T.D. (2017). 'Limitations of PET-PEESE and other meta-analysis methods', *Social Psychological and Personality Science*, vol. 8(5), pp. 581–91.

Stanley, T.D. and Doucouliagos, H. (2012). *Meta-Regression Analysis in Economics and Business*, London: Routledge.

Stanley, T.D. and Doucouliagos, H. (2014). 'Meta-regression approximations to reduce publication selection bias', *Research Synthesis Methods*, vol. 5(1), pp. 60–78.

Stanley, T.D. and Doucouliagos, H. (2015). 'Neither fixed nor random: weighted least squares meta-analysis', *Statistics in Medicine*, vol. 34(13), pp. 2116–27.

Stanley, T.D. and Doucouliagos, H. (2017). 'Neither fixed nor random: weighted least squares meta-regression', *Research Synthesis Methods*, vol. 8(1), pp. 19–42.

Stanley, T.D., Jarrell, S.B. and Doucouliagos, H. (2010). 'Could it be better to discard 90% of the data? A statistical paradox', *The American Statistician*, vol. 64(1), pp. 70–7.

Steel, M.F.J. (2020). 'Model averaging and its use in economics', *Journal of Economic Literature*, vol. 58(3), pp. 644–719.

Van den Noortgate, W., López-López, J.A., Marín-Martínez, F. and Sánchez-Meca, J. (2013). 'Three-level meta-analysis of dependent effect sizes', *Behavior Research Methods*, vol. 45(2), pp. 576–94.