Meta-Analysis of Present-Bias Estimation using Convex Time Budgets

Taisuke Imai Tom A. Rutter Colin F. Camerer *

July 8, 2019

Abstract

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 the estimates exhibit substantial heterogeneity across studies. There is evidence of modest selective reporting in the direction of overreporting present-bias. The primary source of this heterogeneity is the type of reward, either monetary or non-monetary reward, as discussed in the literature, but the effect is weakened after correcting for potential selective reporting. In the studies using the monetary reward, the delay until the issue of the reward associated with the "current" time period is shown to influence the estimates of present bias parameter.

JEL Classification codes: D90, C91

Keywords: present bias, structural behavioral economics, meta-analysis, selective reporting

*Imai: Department of Economics, LMU Munich, taisuke.imai@econ.lmu.de. Rutter: Stanford Law School, Stanford University, trutter@law.stanford.edu. Camerer: Division of the Humanities and Social Sciences, California Institute of Technology, camerer@hss.caltech.edu. This is a part of the project "A Large-Scale, Interdisciplinary Meta-Analysis on Behavioral Economics Parameters" supported by the Social Science Meta-Analysis and Research Transparency (SSMART) Grants from Berkeley Initiative for Transparency in the Social Sciences (BITSS). We thank Stefano DellaVigna, Peter Schwardmann, Charles Sprenger, and Tom Stanley for helpful comments. Imai acknowledges financial support by the Deutsche Forschungsgemeinschaft through CRC TRR 190. Rutter acknowledges the support of the 2016 SURF Fellowship from the California Institute of Technology.

1 Introduction

Most choices create benefits and costs that occur at different points in time. Domains of these intertemporal 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 quasi-hyperbolic discounted utility model (QHD; Laibson, 1997; Phelps and Pollak, 1968), also known as the present-biased preferences model, is an extension of the exponentially discounted utility model (EDU; Koopmans, 1960; Samuelson, 1937). It is designed to capture dynamically inconsistent choices while retaining some of the tractability of EDU. In QHD an agent values a consumption stream $(x_0, \ldots, x_T)$ according to

$$U(x_0, \ldots, x_T) = u(x_0) + \beta \sum_{t=1}^{T} \delta^t u(x_t), \tag{1}$$

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 that QHD includes EDU as a special case when $\beta = 1$ (there is no present-bias). 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 all empirical estimates of present-biased preferences measured with the experimental method called the Convex Time Budget (CTB; Andreoni and Sprenger, 2012) and meta-analyze those data. The meta-analysis gives tentative answers to four 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.? (iv) How much will more data change these answers?

Our meta-analysis collects 220 estimates of the present-bias parameter (β in equation (1); hereafter *PB*) in the QHD model from 31 studies reported in 28 articles included in the dataset. To

[1] See, for example, DellaVigna and Malmendier (2006), Gruber and Kőszegi (2001), Heidhues and Kőszegi (2010), and O'Donoghue and Rabin (1999, 2001) for applications of (naïve) present-biased preferences and O'Donoghue and Rabin (2015) for a short overview.

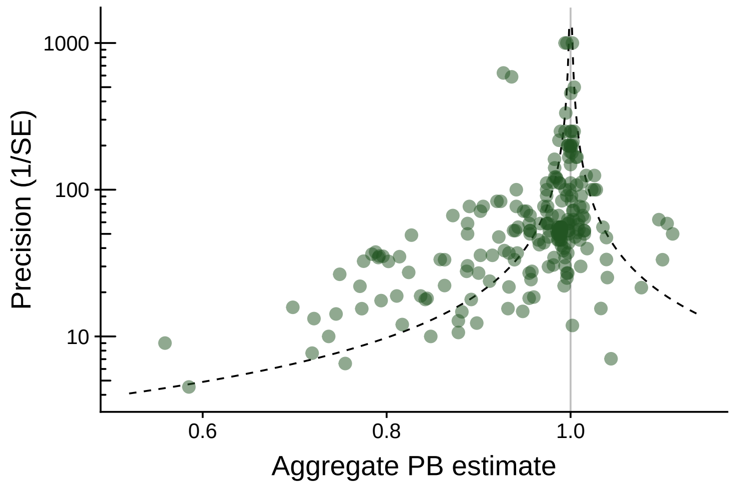


FIGURE 1: Funnel plot of estimates of present bias parameter (*PB*). The y-axis (precision; inverse standard error) is presented in the log-scale. The dotted curves indicate the boundaries for rejection of the null hypothesis of no present bias ($PB = 1$; vertical grey line) for a two-sided test at the 5% level.

give a quick preview, the distribution of estimates and the relation with their associated standard errors is presented in the "funnel plot" in Figure 1. A significant 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 show present bias rather than future bias.

While meta-analysis is indeed a method, the contribution of our paper is *not* primarily methodological. 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; Beshears et al., 2017; Meier and Sprenger, 2010), health decisions (Fang and Wang, 2015), labor contracts (Bisin and Hyndman, 2018; Kaur et al., 2010, 2015), demand for commitment devices (Ashraf et al., 2006; Beshears et al., 2015; John, forthcoming), and others.

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 re-analyze), establishing central tendency of numerical estimations, exploring cross-study moderators which affect estimates, and testing for various kinds of publication bias.

Meta-analysis is designed to accumulate scientific knowledge, and also detect nonrandom reporting or publication of estimates that deviate from the average. Since it was first introduced by Glass (1976), meta-analysis has been playing 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 intertemporal choices and *PB* include the classic piece by Frederick et al. (2002) and more recent coverage by Cohen et al. (forthcoming) and Ericson and Laibson (2019). These articles are narrative and do not provide systematic collection and analysis of empirical observations (they rather describe subsets of important contributions and themes which emerge across studies). [4]

The next section explains how we construct the dataset. Section 3 describes observable characteristics of the studies and variation in experimental design. Section 4 presents the results.

[2]See a list of relevant publications indexed on RePec at: `https://ideas.repec.org/k/metaana.html`.

[3]Echenique et al. (2018) also conduct a meta-analysis on choice data from CTB experiments, to measure the consistency of individual-level choices with several models of time preferences, including EDU, QHD, and general time separable utility (of which EDU and QHD are special cases). They use a nonparametric revealed-preference-based approach developed by Echenique et al. (2016). Time-separability is frequently violated, especially when subjects made fewer choices at the corners of the budget lines. This fact implies that the estimates of *PB* described in our CTB meta-analysis all come from a model which rests on empirically dubious assumptions. Nonetheless, we hope our meta-analysis is useful, just as the narrative reviews have been, since estimates of *PB* are of general interest. There is also no general procedure to recover an improved replacement for *PB* in the face of time-separability violations.

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

2 Data and Method

2.1 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 intertemporal 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. 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.

2.2 Identification and Selection of Relevant Studies

Every good meta-analysis starts by casting a wide net trying to identify *all* 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 sufficient sample size and to be able to check indicators of publication bias and selective reporting.

We searched articles which employed the CTB protocol using Google Scholar, first by querying papers that cited Andreoni and Sprenger (2012), Andreoni et al. (2015), and Augenblick et al. (2015). We also searched for papers with the keyword 'convex time budget'. These two sets of searches, done on November 28 and December 15, 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, which conducted experiments or surveys involving the CTB protocol. Note that this broad inclusion criterion keeps studies even if QHD parameters are not estimated. These studies do not contribute

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

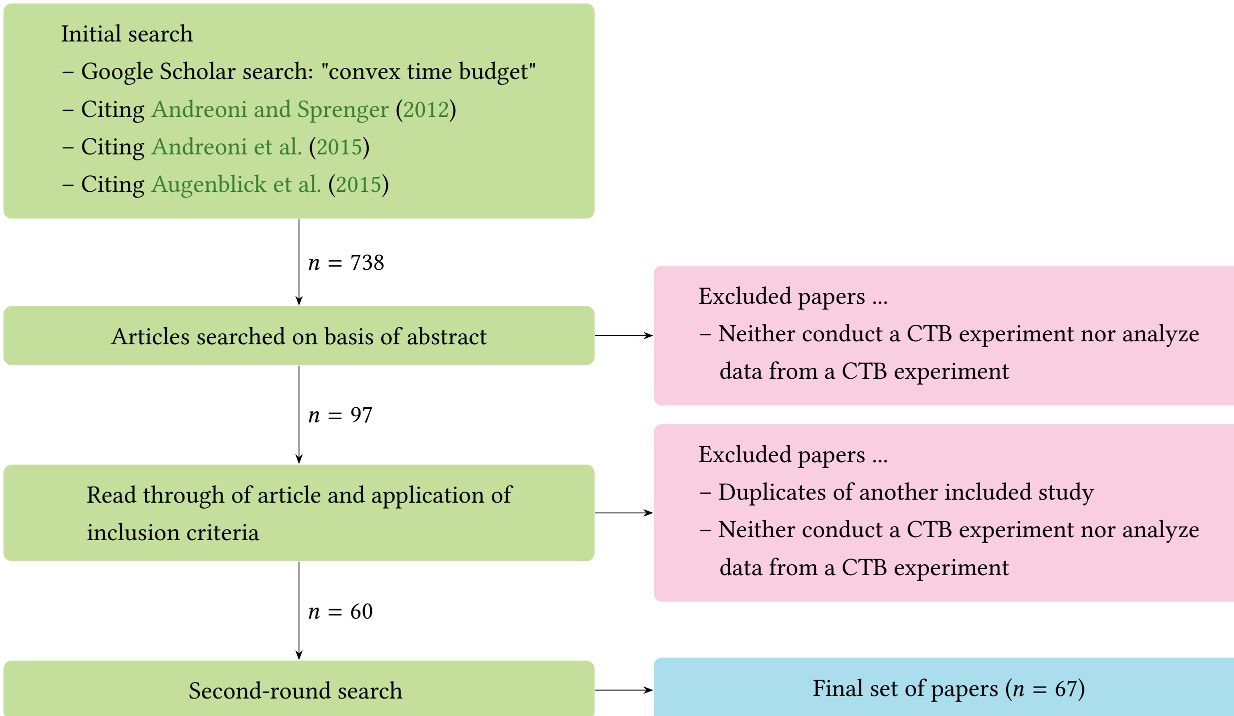


Figure 2: Paper search and data construction.

to our main mata-analysis but still provide 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, too.

We performed the second-round search (using the same query) and updated the database in the Fall of 2018. The final dataset includes 67 articles. [6] Figure 2 illustrates our selection procedure.

Note that in keeping with good meta-analysis practice, our inclusion criteria specifically exclude other studies which are 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 find a decay effect in which a few hours of delay reduces present bias substantially. There are many, many other papers in economics, psychology and cognitive neuroscience which are important but are not included because they did not use CTB. [7]

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

[7] We are currently conducting a larger-scale meta-analysis using papers which estimate discounting parameters using any method, extending the scope beyond CTB.

2.3 Data Construction

After identifying relevant articles, we assembled the dataset by coding estimation results and characteristics of the experimental design. We call a collection of estimates a “study” when they are from the same experimental design. These two units of observations, an article and a study, coincide in many cases, but it allows us to distinguish two conceptually different experiments reported in a single paper (e.g., monetary reward and effort-cost versions of CTB in Augenblick et al., 2015).

Our primary variable of interest is 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) as well. Studies report either aggregate-level parameter estimates (i.e., pool choice data from all subjects and estimate a set of parameters for the “representative subject”) or some summary statistics, such as the mean or median of individual-level estimates. We coded these two reporting types of estimate separately. [8] We also coded standard errors of parameter estimates from aggregate-level analysis in order to control simply 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, check, gift card); subject pool (e.g., children, college student, general population); and so on. Table B.4 in Online Appendix lists variables coded in the study. Some studies implemented the CTB protocol with some treatment variations, such as hunger, cognitive resource depletion, financial education intervention, time pressure, and so on (Table B.3 in 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 condition).

3 Features of Studies and Experimental Designs

We identified 67 articles that conducted experiments or surveys that used the CTB protocol or a modification, where 36 of them are published (or “in press”) including nine articles published in

[8] In our main meta-analysis 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 with the standard deviation of the distribution. See Figure B.1 in Online Appendix.

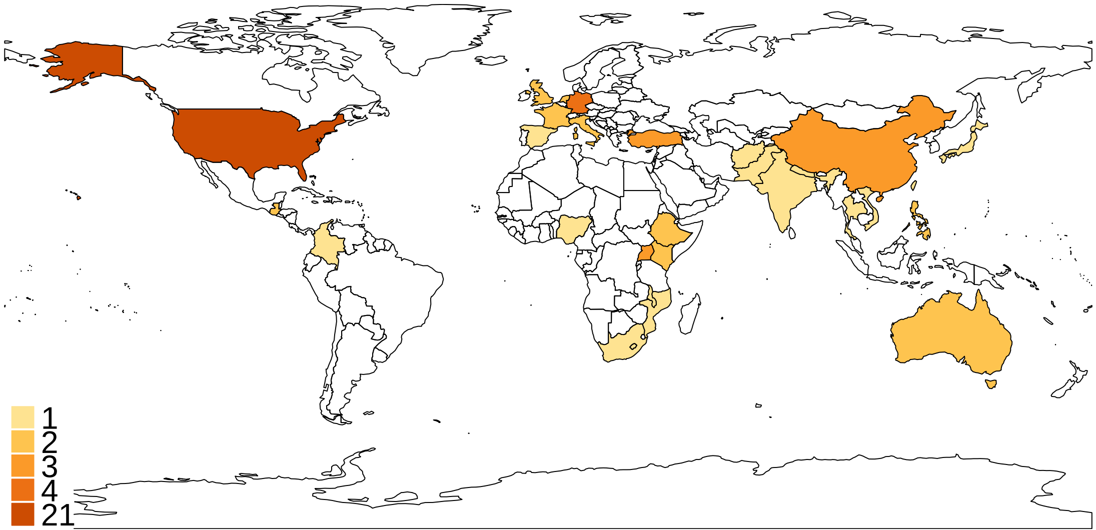


FIGURE 3: Number of studies by country.

one of the "Top 5" journals (as of December 31, 2018). There are 41 articles that report structurally estimated QHD parameters either at the aggregate level or at the individual level. The median number of estimates reported in an article is three. Seven studies reported more than 10 estimates, and two of them reported more than 30 (Table B.1 in Online Appendix).

Observable features of experimental design do not exhibit marked difference between studies with parameter estimates and those without (Tables 1 and 2; Figure C.5 in Online Appendix).

Roughly half of the studies report laboratory experiments. Online experiments constitute fewer than 20% of the studies in the dataset. Only one experiment which studied choices made by children in a classroom. As shown in Figure 3, studies were conducted in 29 different countries, although a third of studies analyzed data from the USA. [9]

Most of the studies recruited participants from the population of college/university students, or a 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) (and see Brown et al. (2009) for earlier results with liquid primary reinforcers, not using CTB). Studies which used monetary reward differed in how future payments were made: some used bank transfer or sent checks to the subjects, but in some other experiments subjects came back to the laboratory to

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

Table 1: Characteristics of CTB studies in the dataset (1).

	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 12/31/2018)				
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 standard errors			28	77.8
Individual-level estimates			10	27.8

Note: *"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).

pick up the payments.

These observable study characteristics exhibit some patterns of co-occurrence (Figures C.6-

Table 2: Characteristics of CTB studies in the dataset (2).

	All CTB studies		Studies with estimates	
	Frequency	Proportion (%)	Frequency	Proportion (%)
Total number of studies	67		36	
Subject population				
Kids and teens	7	10.4	1	2.8
Univ. students	28	41.8	15	41.7
General pop.	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				
Bank transfer	19	28.4	11	30.6
Pickup	5	7.5	3	8.3
Check	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 of future payment	46	68.7	23	63.9
Equalize transaction cost	52	77.6	28	77.8

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

C.8 in Online Appendix). For example, laboratory experiments tended to have student subjects while field studies are more likely to recruit from the general population.

Experimental elicitation of time preferences requires researchers to design experiments so

Table 3: Characteristics of budgets and time frames.

	All CTB studies (60)				Studies with estimates (38)			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Number of budget sets	17.69	14.50	1.00	55	21.88	20.00	4.00	55
Number of time frames	3.18	2.00	1.00	10	3.78	3.00	1.00	10
Minimum delay length (days)	34.89	28.00	1.00	365	40.88	30.00	1.00	365
Maximum delay length (days)	166.40	32.50	1.00	7,300	236.85	56.00	1.00	7,300
Mean delay length (days)	90.72	30.00	1.00	3,285	123.95	42.00	1.00	3,285

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., Cohen et al., forthcoming; Ericson and Laibson, 2019). [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 in transaction costs of 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 which 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 $1 + r$ over k periods; and so on. Table 3 summarizes the ranges and central tendencies of these design variables.

On average, researchers asked 22 questions to recover QHD parameters. 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 is a little over one month and the maximum waiting period is six to eight months.

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

Table 4: Characteristics of aggregate-level PB parameter 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 + NLS	62	27.3
Tobit	107	47.1
Multinomial logit + 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

Finally, we look at the assumptions and econometric approaches employed to structurally estimate QHD parameters (Table 4). There are 227 estimates in the dataset, and a significant majority assume a constant relative risk aversion (CRRA) specification for the instantaneous utility function u in the model (1). The typical specification for studies using real-effort tasks is a convex effort cost function. There are 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 nonlinear least squares (NLS) and Tobit approaches.

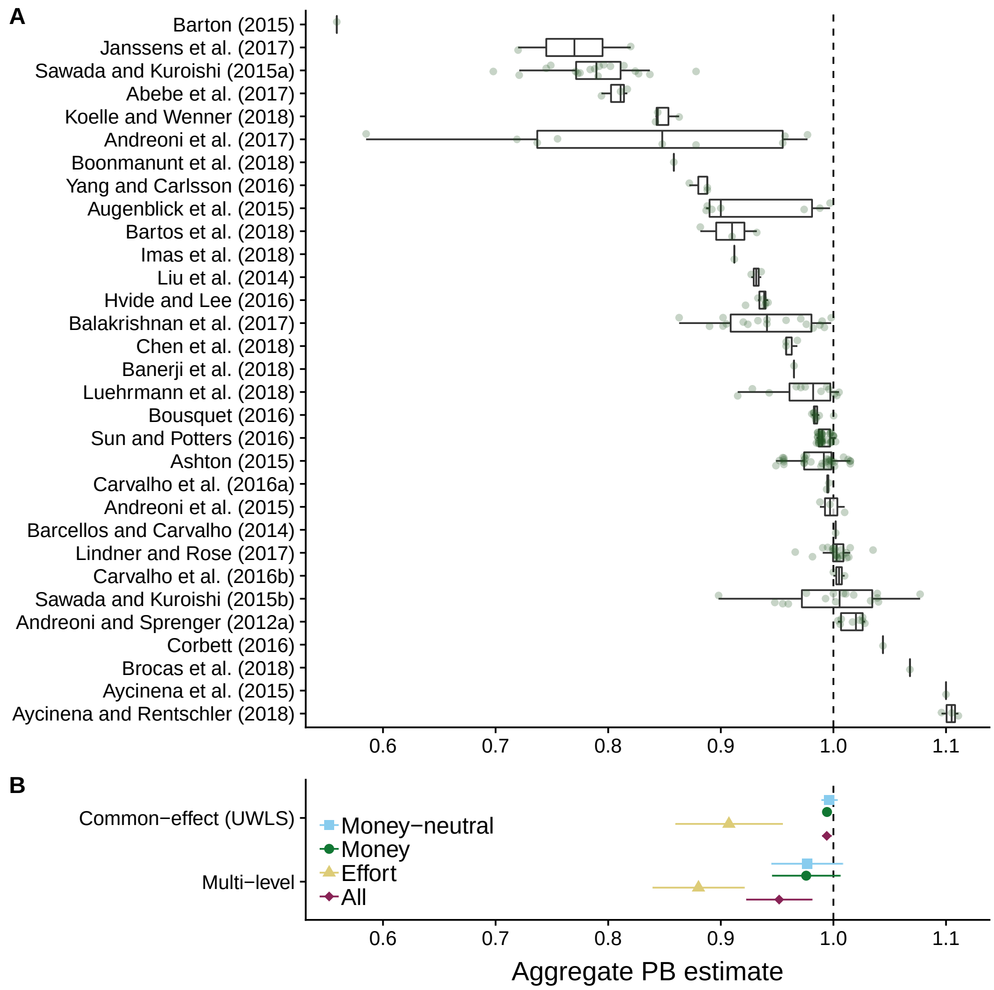


FIGURE 4: Present bias parameter estimates. The vertical dotted line indicates no present/future bias.

4 Results

Aggregate-level estimates of the present-bias parameter from each article in the dataset are shown in Figure 4A. About 77% of these estimates (170) 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 standard errors (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 dataset.

4.1 Meta-Analytic Synthesis of Present Bias Estimates

We start by providing a meta-analytic estimation of the "average" *PB* in the dataset. These analyses provide a tentative answer to the question: What is the average value of *PB*?

In a simple meta-analytic framework, the common-effect model is

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

where PB_j is the jth estimate of present-bias in the dataset ($j = 1, \ldots, m$), 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 $\varepsilon_j \sim \mathcal{N}(0, v_j^2)$ and the sampling variance v_j^2 is known. We can obtain the common-effect estimate of PB_0 as the weighted average of individual estimates:

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

where the weights are given by the inverse variance, $w_j = 1/v_j^2$. In this average, estimates with higher precision (smaller standard errors) are given larger weights. If we assume that the sampling variance is known only up to some unknown multiplicative constant (i.e., $v_j^2 = \phi \tilde{v}_j^2$ for some $\phi > 0$), equation (2) becomes the unrestricted weighted least squares model (UWLS; Stanley and Doucouliagos, 2015). [11]

In the random-effects meta-analysis (DerSimonian and Laird, 1986), we assume 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 (grand mean) and the sampling error ξ_j. It is further assumed that $\xi_j \sim \mathcal{N}(0, \tau^2)$, where τ^2 is the observation-specific heterogeneity that must be estimated. [12] Note that the random-effects model (3) reduces to the common-effect model (2) when $\tau^2 = 0$. Stanley (2008) shows, using simulations, that the common-effect approach is less biased in the presence of selective reporting. The random-effects estimates are presented in Online Appendix C.4.

Note that our dataset includes *statistically dependent* estimates of *PB* since many studies included in our meta-analysis report multiple estimates from the same experiment (e.g., using dif-

[11] The common-effect and the unrestricted weighted least squares models give the same weighted average $\overline{PB}_0$ but their associated variances are different. The unknown constant ϕ is given by the residual variance from the standard weighted least squares.

[12] For a recent discussion on the use of random-effects meta-analysis in medical decision making, see Manski (2019).

Table 5: Meta-analytic average of present bias parameter.

	All studies		Monetary (all)		Monetary ("neutral")		Effort cost	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{PB}_0$	0.9941	0.9518	0.9943	0.9758	0.9964	0.9766	0.9072	0.8802
	(0.0020)	(0.0149)	(0.0020)	(0.0154)	(0.0036)	(0.0161)	(0.0242)	(0.0208)
p-value	0.0069	0.0031	0.0107	0.1334	0.3317	0.1640	0.0050	0.0004
Model	UWLS	Multi-level	UWLS	Multi-level	UWLS	Multi-level	UWLS	Multi-level
Observations	217	217	193	193	140	140	24	24
Studies	29	29	20	20	19	19	9	9

Note: p-values are from the two-sided test of the null hypothesis $H_0 : PB = 1$. Standard errors in parentheses are cluster-robust (Hedges et al., 2010). Three observations with large influence measure ($|DFBETAS| > 1$) are excluded.

ferent 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 standard error from the regression without the target observation. Following Bollen and Jackman (1985), we identify any observations to be influential if $|DFBETAS| > 1$ (i.e., the observation shifts the coefficient at least one standard error). [13] 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. [14]

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

Table 5 reports the first set of results (odd-numbered columns; also presented in Figure 4B). All specifications show $\overline{PB}_0 < 1$, indicating present bias, and the null hypothesis of no present

[13] *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 standard error of $\widehat{\gamma}_k^{(m)}$.

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

bias (i.e., $H_0 : PB = 1$) is rejected at the conventional level of $p = 0.05$ in all but one specification. The overall $\overline{PB}_0$ is 0.99, which is significantly different from one at the 1% significance level. The only estimate which is not significantly different from one (at the 5% level) comes from the subset of observations from CTB studies using monetary reward without any treatment variations. We observe smaller average $\overline{PB}_0$ in the real-effort version of CTB studies compared to the CTB studies using monetary reward.

As an alternative approach to handle dependent PB estimates, we also apply multi-level meta-analysis (Konstantopoulos, 2011; Van den Noortgate et al., 2013). [15] Let PB_{ij} denote the jth estimate of 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 $\varepsilon_{ij} \sim \mathcal{N}(0, v_{ij}^2)$ 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 a lot 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 $\mathbf{E}[PB_{ij}] = PB_0$.

In this multi-level specification, we find $\overline{PB}_0$'s that are smaller than the corresponding estimates from UWLS approach (Table 5). The overall $\overline{PB}_0$ in the literature is about 0.95 (see column (2) of Table 5). The value 0.95 is therefore the tentative best guess of the overall value of PB_0. However, previewing results below, it also appears that $\overline{PB}_0$ is close to one for choices over money, and is smaller, around 0.88-0.91, $\overline{PB}_0$ for choices over effort (see columns (7) and (8) of Table 5).

4.2 Identifying and Correcting for Selective Reporting

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

[15] 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.

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 which derive an unpredicted or unconventional result may be underpublished. We will refer to this misproduction of results as "publication bias". There are several possible sources of publication bias. 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 sources 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 publication bias (except clumsy frauds). However, in a group of related studies there are ways to detect possible collective publication bias.

The QHD model emerged to explain observed patterns of present-biased choices, including procrastination and challenges self-control. Publication bias 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 than could generate the observed pattern of present-biased choices that the QHD model is trying to capture.

The funnel plot is a useful method for detecting selective reporting (and counterfactually 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 a slight asymmetry even though the magnitude may not be huge (see also Online Appendix Figures C.3 and C.4, which present funnel plots for monetary-CTB and effort-CTB separately).

A common procedure for detecting and correcting for publication selection bias is the FAT-PET-PEESE procedure (Stanley and Doucouliagos, 2012, 2014). [16] In the absence of selective reporting, the reported estimates of the present-bias parameter should be uncorrelated with their standard errors. In the presence of selective reporting, on the other hand, the reported estimates are correlated with their standard errors (more imprecise estimates in the unconventional direction will go unreported). This motivates a simple regression model for detection of selective

[16] This is an acronym for combination of *Funnel Asymmetry Test* (FAT), *Precision Effect Test* (PET), and *Precision Effect Estimates with Standard Errors* (PEESE).

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 standard errors 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 standard error 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). An equivalent expression of the model is given by dividing (4) by SE_{ij}:

$$t_{ij} = \alpha_1 + \alpha_0 \cdot (1/SE_{ij}) + \nu_{ij},$$

where t_{ij} is the t-statistic of each estimate and $\nu_{ij} = \varepsilon_{ij}/SE_{ij}$. 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) for discussion (especially on the limitations of these approaches).

Table 6 reports results from estimation of model (4) using the unrestricted weighted least squares. We again exclude three overly influential observations identified above. The estimated values of α_1 are negative, indicating that less precise (i.e., larger SE) studies do yield lower estimates of PB. However, we do not reject the null hypothesis that the coefficient on SE is zero.

We also apply the *latent studies* method for identification and correction for publication bias proposed by Andrews and Kasy (forthcoming), discussed in detail in Online Appendix C.7. These results are shown in Table 6 and Table C.8. None of the relative publication probabilities for estimates with different z-values are significantly different from one. Since there does not appear to be substantial publication selection, the adjusted study estimates from the latent studies model are very similar to the original study estimates (shown in Figure C.27 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 Figure C.29.

Taken together, we view our results as demonstrating that there is evidence of modest selective reporting in the direction of overreporting $PB < 1$, though it is marginally significant by

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		0.9266	
		(0.0029)		(0.0029)		(0.0036)		(0.0253)	

Note: Estimated by weighted least squares. Standard errors 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.

conventional tests. This bias also appears strongest for studies using effort. Correcting for selective reporting gives values of *PB* that are still close to one for money and lower, 0.90-0.93, for effort.

4.3 Explaining Heterogeneity

We have thus far assumed that the variability in reported estimates are mainly due to sampling errors, either at the observation level or study level, or both, and a modest amount of 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. Online Appendix Figures C.9-C.19 visualize the effects of some representative study characteristics on reported estimates, looking at each characteristic in isolation.

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

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \boldsymbol{\gamma} \boldsymbol{X}_{ij} + \varepsilon_{ij}, \tag{5}$$

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

Results from this meta-regression analysis report a tentative answer to the question: How does *PB* vary reliably with methods, subject population, and other study characteristics?

In the first set of meta-regressions presented in Table 7, we restrict samples to those using monetary reward. We consider eight 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*), subject population (omitted category is *Subject: Kids*), 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 *Computerized*). [17] We also include several additional variables which are specific to experiments involving monetary reward: method of reward delivery (omitted category is *Delivery: Check*) and treatment of confounding factors such as uncertainty regarding future reward and transaction costs (omitted category is *Ignored* in both variables). We estimate the model using the unrestricted weighted least squares (Stanley and Doucouliagos, 2017).

The effects of study characteristics on estimated *PB* parameter exhibit interesting patterns. For example, regression coefficients reported in Table 7 suggest that: university students and the general population are less present-biased than children; field experiments tend to find less present-biased preferences compared to lab studies; dealing with uncertainty about future reward makes estimated *PB* smaller; and dealing with transaction costs makes estimated *PB* larger. However, these effects are sensitive to which other characteristics are simultaneously controlled for. We do not observe the effects of reward delivery method, and whether or not to jointly estimate background consumption has little impact on the estimates of *PB*.

Note that the timing of "immediate" payment appears to matter as discussed in the literature. Compared to studies which guaranteed to deliver the "immediate" rewards within the day of the experiment, estimated *PB* is smaller (more present-biased) when these "immediate" rewards were delivered by the end of experiment.

[17] In Abebe et al. (2017), the immediate reward was delivered on the next day of the experimental session. In other words, their definition of $t = 0$ is extended to "today and tomorrow." Since our definition of "immediate" is limited up to the day of the experiment, estimates from this study (and only those estimates) are categorized into *"Immediate" pay: No immediate rewards.*

Table 7: Explaining the heterogeneity of reported estimates (monetary reward).

	(1)	(2)	(3)	(4)	(5)	(6)
SE of PB estimate	−0.915	−1.141*	−1.327*	−1.248**	−1.951**	−1.711*
	(0.471)	(0.526)	(0.525)	(0.454)	(0.636)	(0.668)
Non-neutral condition	−0.008	−0.008	−0.015	−0.006	−0.003	−0.006
	(0.006)	(0.006)	(0.009)	(0.005)	(0.005)	(0.007)
Subject: University students	−0.003	−0.006	0.023			
	(0.008)	(0.008)	(0.023)			
Subject: General population	−0.010	−0.019	0.015			
	(0.010)	(0.013)	(0.029)			
Location: Field				0.066**	0.071**	0.090***
				(0.025)	(0.022)	(0.025)
Location: Class				0.011	0.022	0.029*
				(0.013)	(0.013)	(0.014)
Location: Online				−0.010	−0.031*	−0.026
				(0.005)	(0.014)	(0.016)
“Immediate” pay: Within day	0.033	0.030*	0.030*	0.048**	0.050***	0.051***
	(0.018)	(0.015)	(0.012)	(0.017)	(0.012)	(0.014)
“Immediate” pay: Not reported	−0.015	−0.011	0.014	−0.066	−0.060	0.046
	(0.056)	(0.053)	(0.067)	(0.056)	(0.051)	(0.065)
Delivery: Cash	0.018	0.011	0.012	0.029	0.017	0.024
	(0.018)	(0.016)	(0.017)	(0.021)	(0.018)	(0.018)
Delivery: Bank	−0.006	−0.007	−0.009	−0.004*	−0.003	−0.006
	(0.004)	(0.006)	(0.008)	(0.002)	(0.004)	(0.005)
Delivery: Other	−0.008	−0.007	−0.010*	−0.008	−0.008*	−0.011**
	(0.006)	(0.007)	(0.005)	(0.004)	(0.004)	(0.003)
Estimation: Tobit		0.005	0.013*		0.018*	0.016
		(0.005)	(0.006)		(0.009)	(0.009)
Estimation: Other		0.004	0.006		−0.002	−0.001
		(0.007)	(0.006)		(0.006)	(0.006)
Estimation: B.g. consumption		−0.005	0.002		−0.001	−0.001
		(0.006)	(0.006)		(0.006)	(0.007)
Deal uncertainty			−0.015**			−0.005
			(0.006)			(0.004)
Deal transaction cost			0.053			0.111**
			(0.041)			(0.038)
Paper and pencil			0.021			−0.017
			(0.025)			(0.013)
Constant	0.981***	0.989***	0.916***	0.963***	0.963***	0.854***
	(0.020)	(0.019)	(0.075)	(0.017)	(0.014)	(0.052)
Observations	193	193	193	193	193	193
R^2	0.372	0.384	0.442	0.457	0.500	0.523
Adjusted R^2	0.341	0.343	0.394	0.427	0.464	0.480

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

Monetary vs. non-monetary rewards. Models of intertemporal choices are fundamentally about utility flow at each time period and not about the receipt of monetary payments. A large share of existing empirical studies has measured time preferences using time-dated monetary payments, but additional assumptions (such as monetary payments are "consumed" at the time of receipt) are necessary to infer individuals' discount functions from observed choices in this approach. More recent studies try to directly control the timing of utility flow using, for example, real-effort tasks (Augenblick et al., 2015; Augenblick and Rabin, 2019; Carvalho et al., 2016), and report evidence that non-monetary rewards provide estimates of present bias parameter that are smaller than those from the standard monetary reward studies.

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

$$PB_{ij} = \alpha_0 + \alpha_1 \cdot SE_{ij} + \boldsymbol{\gamma}\boldsymbol{X}_{ij} + \boldsymbol{\lambda}(SE_{ij} \cdot \boldsymbol{Z}_{ij}) + \varepsilon_{ij}, \tag{6}$$

which extends equation (5) to allow for any factors that can potentially influence selective reporting (captured by $SE_{ij} \cdot \boldsymbol{Z}_{ij}$). 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, neutral condition" in columns (1, 4); "monetary studies, neutral condition, lab, immediate rewards delivered within the day" in columns (2, 5); and "monetary studies, neutral condition, lab, immediate rewards delivered within the day, estimation with NLS" in columns (3, 6).

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)-(3)). The other three specifications (columns (4)-(6)) include *SE* as well as its interaction with *Reward: Money* to correct for potential selective reporting in each type of CTB experiments. The estimated coefficients on *Reward: Money* are positive but not statistically significant at the 5% level. It suggests that the difference between "bias-corrected" average *PB* from monetary and non-monetary studies is not (yet) strong.

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***	0.907***	0.993***	0.993***	0.993***
	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)
SE of PB estimates				−2.057***	−2.057***	−2.057***
				(0.414)	(0.414)	(0.414)
Reward: Money	0.089***	0.093***	0.093***	0.015	0.015	0.016
	(0.024)	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)
× Non-neutral condition	−0.003	−0.011	−0.012	−0.011**	−0.009	−0.006
	(0.004)	(0.007)	(0.007)	(0.004)	(0.006)	(0.006)
× Location: Field		0.055*	0.057**		0.065**	0.071***
		(0.023)	(0.021)		(0.024)	(0.018)
× Location: Class		0.026	0.026		0.035*	0.037***
		(0.018)	(0.017)		(0.017)	(0.011)
× Location: Online		0.006	0.004		−0.003	−0.026
		(0.006)	(0.009)		(0.005)	(0.014)
× "Immediate": By end of exp		−0.039*	−0.039*		−0.036*	−0.042***
		(0.018)	(0.017)		(0.017)	(0.011)
× "Immediate": Not reported		−0.125*	−0.127*		−0.115*	−0.112*
		(0.061)	(0.060)		(0.056)	(0.051)
× Estimation: Tobit			0.002			0.019*
			(0.006)			(0.009)
× Estimation: Other			−0.005			−0.002
			(0.005)			(0.004)
× *SE* of PB estimates				0.374	0.740	0.065
				(0.854)	(0.593)	(0.708)
Observations	217	217	217	217	217	217
R^2	0.054	0.370	0.375	0.249	0.456	0.504
Adjusted R^2	0.045	0.349	0.348	0.235	0.432	0.478
$H_0 : PB_{\text{effort}} = 1$		$p = 0.0004$			$p = 0.7747$	

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

Discussion. 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-occurence of study characteristics in the data (Figures C.7 and C.8 in Online Appendix), and thus made somewhat arbitrarily. Stanley and Doucouliagos (2012) recommend using a general-to-specific approach, also known as a backward stepwise model selection. It starts with including all explanatory variables, and the least statistically significant variable is removed from the model one at a time. This procedure continues until only statistically significant variables remain in the model.

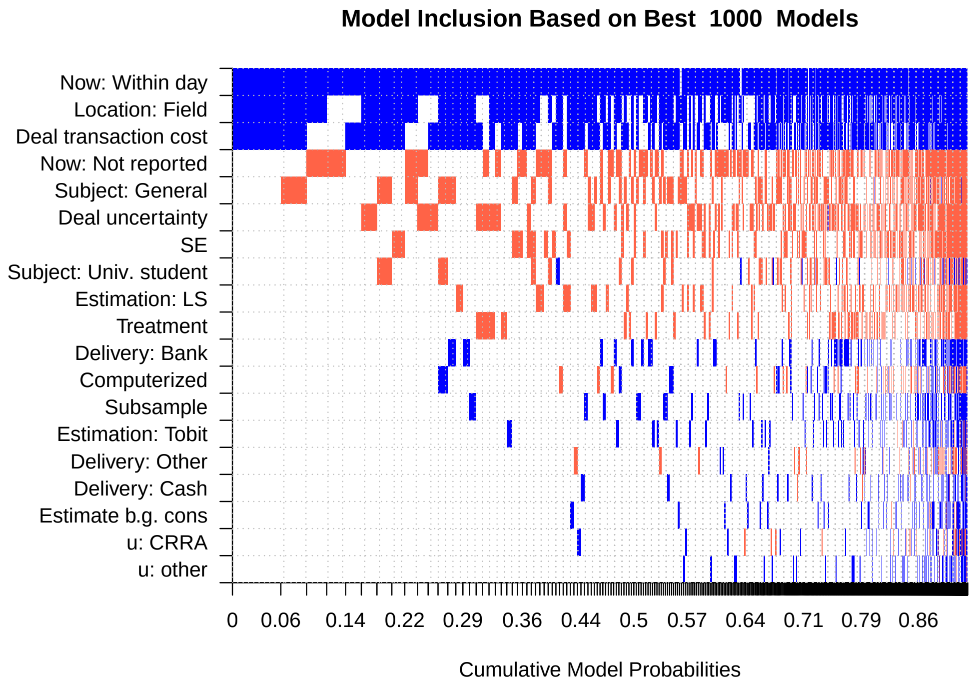
We augment our meta-regression analysis with the application of Bayesian model averaging (BMA) to tackle the model uncertainty resulting from the large number of explanatory variables we could have included in our meta-regression model (Hoeting et al., 1999; Moral-Benito, 2015; Steel, forthcoming). BMA 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.5. For applications of BMA in meta-analysis in economics, see Havránek et al. (2015), Havránek et al. (2017), and Iršová and Havránek (2013).

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.5 of the Online Appendix). In this figure, columns denote individual models where variables are sorted by posterior model probability in a descending order. Blue cells (darker cells in grayscale) indicate that the variable is included in the model and has a positive coefficient, while red cells (lighter cells in grayscale) indicate that the variable has a negative coefficient. White cells indicate that the variable is not included in the model.

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

5 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 collect 220 estimates from 28 articles and find that the meta-analytic average of the present-bias parameter is around 0.95, which is significantly smaller than one, after taking the multi-level nature of the data into consideration. The values for monetary-reward studies are close to one, however, and effort-based studies have lower values, around 0.9-0.93. We also find that estimates vary greatly across studies, primarily due to their different study characteristics. Our meta-regression analysis suggests that CTB experiments with non-monetary rewards indeed found estimates that are "more present biased" than those from CTB with typical monetary rewards, but the effect is weakened after correcting for potential



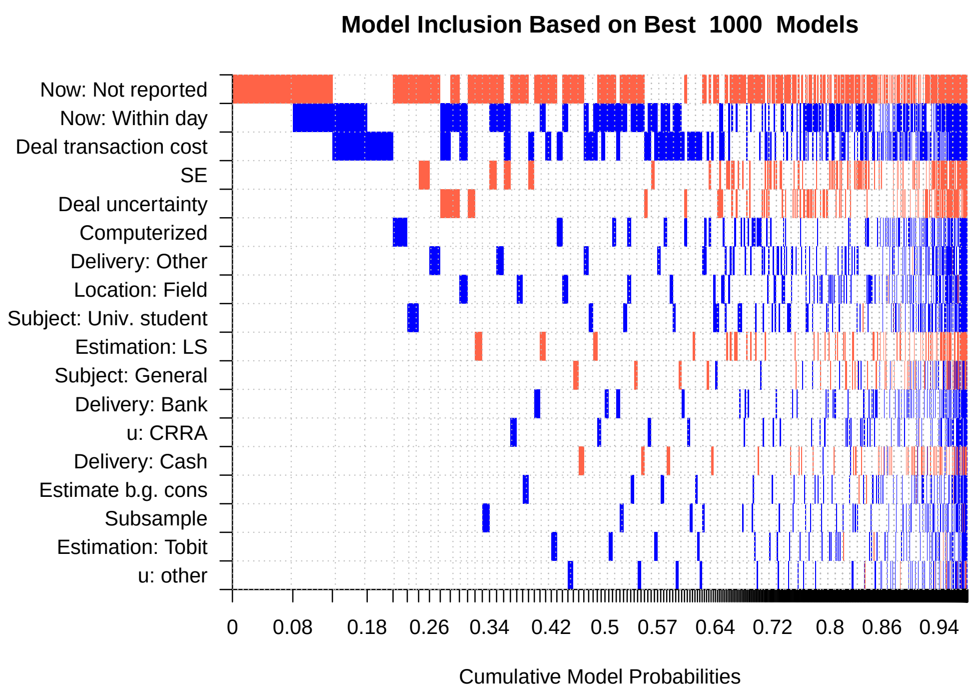


FIGURE 5: Model inclusion. Observations from monetary-CTB studies only. The top panel uses observations from both neutral and non-neutral conditions, while the bottom panel discards data from non-neutral conditions.

selective reporting. Furthermore, we found evidence to confirm the suggestion by Ericson and Laibson (2019) regarding the importance of the delay until the issue of the reward associated with the "current" time period; across a range of specifications in both our meta-regression and Bayesian model averaging approach, studies that delivered rewards associated with the "current" period by the end of the experiment, as 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 behavior of subjects.

In addition, we found suggestive evidence concerning the importance of a factor on estimates of present bias 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 follow Andreoni and Sprenger's (2012) original econometric strategy and report estimates using both NLS and Tobit (or estimates with and without background consumption). These methods ignited significant debate in the literature (see, for example, the discussion in Andreoni et al., 2015). However, our meta-analysis showed that the econometric strategy makes little difference.

Indeed, some design characteristics that have consumed a lot of professional attention do not appear to have effects on *PB* that are robust across meta-regression specifications. These (tentative) non-effects suggest that it is not a good idea to constrain experimental practices to some kind of "ideal design"; instead, variations in design will enable updating of the meta-analytic database so we can learn more rapidly.

References

Abebe, G., S. Caria, and E. Ortiz-Ospina (2017): "The Selection of Talent: Experimental and Structural Evidence from Ethiopia," Unpublished manuscript.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015): "Measuring Time Preferences: A Comparison of Experimental Methods," *Journal of Economic Behavior & Organization*, 116, 451–464.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

Andrews, I. and M. Kasy (forthcoming): "Identification of and Correction for Publication Bias," *American Economic Review*.

Angeletos, G.-M., D. Laibson, A. Repetto, J. Tobacman, and S. Weinberg (2001): "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 15, 47–68.

Ashraf, N., D. Karlan, and W. Yin (2006): "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121, 635–672.

Augenblick, N. (2018): "Short-Term Time Discounting of Unpleasant Tasks," Unpublished manuscript.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130, 1067–1115.

Augenblick, N. and M. Rabin (2019): "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," *Review of Economic Studies*, 86, 941–975.

Barcellos, S. H. and L. Carvalho (2014): "Information about Self-Control and Intertemporal Choices," Unpublished manuscript.

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

Beshears, J., J. J. Choi, C. Clayton, C. Harris, D. Laibson, and B. C. Madrian (2017): "Optimal Illiquidity," Unpublished manuscript.

Beshears, J., J. J. Choi, C. Harris, D. Laibson, B. C. Madrian, and J. Sakong (2015): "Self Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?" NBER Working Paper No. 21474.

Bisin, A. and K. Hyndman (2018): "Present-Bias, Procrastination and Deadlines in a Field Experiment," NBER Working Paper No. 19874.

Bollen, K. A. and R. W. Jackman (1985): "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases," *Sociological Methods & Research*, 13, 510–542.

Brown, A. L., Z. E. Chua, and C. F. Camerer (2009): "Learning and Visceral Temptation in Dynamic Saving Experiments," *Quarterly Journal of Economics*, 124, 197–231.

Carvalho, L. S., S. Meier, and S. W. Wang (2016): "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday," *American economic review*, 106, 260–284.

Cohen, J., K. Ericson, D. Laibson, and J. White (forthcoming): "Measuring Time Preferences," *Journal of Economic Literature*.

Cubitt, R. P. and D. Read (2007): "Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption?" *Experimental Economics*, 10, 369–389.

DellaVigna, S. and U. Malmendier (2006): "Paying Not to Go to the Gym," *American Economic Review*, 96, 694–719.

DerSimonian, R. and N. Laird (1986): "Meta-Analysis in Clinical Trials," *Controlled Clinical Trials*, 7, 177–188.

Echenique, F., T. Imai, and K. Saito (2016): "Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations," Caltech HSS Working Paper 1388.

——— (2018): "Revealed Preference Tests of Behavior in Convex Time Budget Experiments," Unpublished manuscript.

Egger, M., G. D. Smith, M. Schneider, and C. Minder (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test," *BMJ*, 315, 629–634.

Ericson, K. M. and D. Laibson (2019): "Intertemporal Choice," in *Handbook of Behavioral Economics - Foundations and Applications*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Amsterdam: Elsevier/North-Holland, vol. 2, 1–67.

Fang, H. and Y. Wang (2015): "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions," *International Economic Review*, 56, 565–596.

Frederick, S., G. Loewenstein, and T. O'donoghue (2002): "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40, 351–401.

Furukawa, C. (2019): "Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method," Unpublished manuscript.

Glass, G. V. (1976): "Primary, Secondary, and Meta-Analysis of Research," *Educational Researcher*, 5, 3–8.

Gruber, J. and B. Kőszegi (2001): "Is addiction "rational"? Theory and evidence," *Quarterly Journal of Economics*, 116, 1261–1303.

Gurevitch, J., J. Koricheva, S. Nakagawa, and G. Stewart (2018): "Meta-Analysis and the Science of Research Synthesis," *Nature*, 555, 175–182.

Halevy, Y. (2014): "Some Comments on the Use of Monetary Rewards in the Measurement of Time Preferences," Unpublished manuscript.

Havránek, T., R. Horváth, Z. Iršová, and M. Rusnak (2015): "Cross-Country Heterogeneity in Intertemporal Substitution," *Journal of International Economics*, 96, 100–118.

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

Hedges, L. V., E. Tipton, and M. C. Johnson (2010): "Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates," *Research Synthesis Methods*, 1, 39–65.

Heidhues, P. and B. Kőszegi (2010): "Exploiting Naïvete about Self-Control in the Credit Market," *American Economic Review*, 100, 2279–2303.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–401.

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

Iršová, Z. and T. Havránek (2013): "Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis," *World Development*, 42, 1–15.

John, A. (forthcoming): "When Commitment Fails – Evidence from a Field Experiment," *Management Science*.

Kaur, S., M. Kremer, and S. Mullainathan (2010): "Self-Control and the Development of Work Arrangements," *American Economic Review: Papers & Proceedings*, 100, 624–28.

——— (2015): "Self-Control at Work," *Journal of Political Economy*, 123, 1227–1277.

Konstantopoulos, S. (2011): "Fixed Effects and Variance Components Estimation in Three-level Meta-Analysis," *Research Synthesis Methods*, 2, 61–76.

Koopmans, T. C. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, 287–309.

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

Liu, E. M., J. Meng, and J. T.-y. Wang (2014): "Confucianism and Preferences: Evidence from Lab Experiments in Taiwan and China," *Journal of Economic Behavior & Organization*, 104, 106–122.

Manski, C. F. (2019): "Meta-Analysis for Medical Decisions," NBER Working Paper No. 25504.

Meier, S. and C. Sprenger (2010): "Present-Biased Preferences and Credit Card Borrowing," *American Economic Journal: Applied Economics*, 2, 193–210.

Moral-Benito, E. (2015): "Model Averaging in Economics: An Overview," *Journal of Economic Surveys*, 29, 46–75.

O'Donoghue, T. and M. Rabin (1999): "Doing It Now or Later," *American Economic Review*, 89, 103–124.

——— (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, 116, 121–160.

——— (2015): "Present Bias: Lessons Learned and to be Learned," *American Economic Review*, 105, 273–279.

Phelps, E. S. and R. A. Pollak (1968): "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–199.

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

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review," *Journal of Economic Perspectives*, 15, 131–150.

——— (2005): "Beyond Publication Bias," *Journal of Economic Surveys*, 19, 309–345.

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

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

——— (2014): "Meta-Regression Approximations to Reduce Publication Selection Bias," *Research Synthesis Methods*, 5, 60–78.

——— (2015): "Neither Fixed nor Random: Weighted Least Squares Meta-Analysis," *Statistics in Medicine*, 34, 2116–2127.

——— (2017): "Neither Fixed nor Random: Weighted Least Squares Meta-Regression," *Research Synthesis Methods*, 8, 19–42.

Stanley, T. D., S. B. Jarrell, and H. Doucouliagos (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox," *The American Statistician*, 64, 70–77.

Steel, M. F. J. (forthcoming): "Model Averaging and its Use in Economics," *Journal of Economic Literature*.

Van den Noortgate, W., J. A. López-López, F. Marín-Martínez, and J. Sánchez-Meca (2013): "Three-Level Meta-Analysis of Dependent Effect Sizes," *Behavior Research Methods*, 45, 576–594.