Shimodaira, Yuta; Shiozawa, Kohei; Inukai, Keigo

Working Paper

Investigation of the convex time budget experiment by parameter recovery simulation

ISER Discussion Paper, No. 1185

Provided in Cooperation with:
The Institute of Social and Economic Research (ISER), Osaka University

Suggested Citation: Shimodaira, Yuta; Shiozawa, Kohei; Inukai, Keigo (2022) : Investigation of the convex time budget experiment by parameter recovery simulation, ISER Discussion Paper, No. 1185, Osaka University, Institute of Social and Economic Research (ISER), Osaka

This Version is available at:
https://hdl.handle.net/10419/267797

Discussion Paper No. 1185

ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

INVESTIGATION OF THE CONVEX TIME BUDGET EXPERIMENT BY PARAMETER RECOVERY SIMULATION

Yuta Shimodaira
Kohei Shiozawa
Keigo Inukai

August 2022

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Investigation of the Convex Time Budget Experiment by Parameter Recovery Simulation

Yuta Shimodaira[1], **Kohei Shiozawa**[2], **and Keigo Inukai**[3,*]

[1]Graduate School of Economics, Osaka University, 1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan.
[2]Faculty of Economics, Takasaki City University of Economics, 1300 Kaminamie-machi, Takasaki, Gunma, 370-0801, Japan.
[3]Faculty of Economics, Meiji Gakuin University, 1-2-37 Shirokanedai, Minato, Tokyo, 108-8636, Japan.
[*]inukai@eco.meijigakuin.ac.jp

ABSTRACT

The convex time budget (CTB) method is a widely used experimental method for eliciting an individual's time preference. Researchers adopting the CTB experiment usually assume quasi-hyperbolic discounting utility as a behavioural model and estimate the parameters of the utility function. However, few studies using the CTB method have examined parameter recovery. We conduct simulations and find that the estimation error of the present bias parameter is so large that its effect is difficult to detect. The large error is due to the improper combination of the experimental method and the utility model, and it is not a problem we can deal with after the data collection. This paper suggests the importance of running parameter recovery simulations to audit estimation errors in the experimental design.

Introduction

The decision-making problem between different points in time is an extremely important issue for living organisms[1,2]. All life, including human life, is subject to the choice of taking immediate profit or giving up immediate profit for future profit. The problem of cross-selection has been studied in birds, mammals, and insects. These organisms have to make a choice between a small but immediate gain and a larger future gain. Most living things, including humans, are not good at sacrificing immediate gains for the sake of larger future gains. In the choice problem at different points in time, it is the discount rate that determines by how much future profits will be discounted over time.

Discount rates are measured using various methods, and some interesting results have emerged. The discount rate can change over time. Let us consider the following example. The choice is between eating one chocolate or delaying consumption by a week and eating two chocolates. Many people would probably give in to temptation and eat the one chocolate, rather than waiting a week. However, if the choice is now whether to eat one chocolate 1 week from now or two chocolates 2 weeks from now, people are more likely to wait 2 weeks. This is an anomaly named present bias[3], a contradiction of choice associated with the choice problem between different points in time. The existence of present bias suggests that the power of our will is weaker than we assume. The psychology of procrastinating on tasks we do not want to perform is one of the serious problems caused by these anomalies.

In addition, we cannot overlook the fact that there is an element of risk associated with the problem of choosing between different points in time. People may believe that short-term profits are certain, whereas future profits are uncertain. There is still much controversy about how risk influences the choice problem between different points in time. Therefore, experimental economists have proposed several ways to separate choice tasks and risk preferences between different points in time.

One method called the convex time budget (CTB) attempts to elicit simultaneously the effects of time discounting and risk attitude by directly estimating the curvature of the utility function using one single instrument[4]. In experimental economics, many experiments on intertemporal choice problems now adopt the CTB method in both laboratory and field settings[5–7]. When analysing preferences from behavioural data collected by the CTB method, researchers do not merely compare the intertemporal allocations across conditions, but also estimate the parameters of the quasi-hyperbolic discounting utility function[8,9]. A meta-analysis of studies using the CTB experiment showed that, on average, the experiment participants discounted the future payoff by 0.95–0.97 over the payoff available now[10].

Although researchers often examine the reliability of estimates a posteriori based on the magnitude of standard errors associated with the estimates, it is rare to examine the estimates' precision before undertaking an experiment. We do not understand how precisely we could estimate the utility function of an individual in a CTB experiment. For example, if an individual's present bias parameter estimate is 0.97, is this individual's behaviour truly biased? We can examine the errors of

the parameter estimates in the preliminary stages of implementation by conducting a simulation called "parameter recovery"[11]. Parameter recovery simulation is conducted using the following three steps. First, we generate artificial decision data using assumed parameter values. Second, we estimate the parameters from the artificial data using the software we will use after the real data collection. Third, we compare the estimated parameters to the true values to check whether they have been recovered precisely.

This paper audits parameter estimation error in typical CTB experiment designs. While studies using the CTB method are still be conducted, the interpretation of the confusion between time preference and risk preference is controversial[12–19]. This paper, however, does not use the CTB method to deal with time and risk preferences, but instead, discusses the use of parametric analysis within the CTB method. We show that the estimation precision of the present bias parameter is not good in the scope range of the parameter estimates reported so far. Researchers often focus on the existence of a present bias[10]. Our results suggest that it is inappropriate to adopt the CTB method to investigate a bias effect whose existence is doubtful and whose magnitude, if any, may be small.

It has been pointed out that the discerning rate and the ease of generating anomalies differ depending on the experimental method. In psychology, the reproducibility of experimental results is often a problem, and in experimental economics, reproducibility is an important issue to consider as well. Although it has been pointed out that the replication rate of experimental studies in experimental economics is better than that in psychology, it is also true that there is variability among experiments. This variability between experiments may be due to subject demographics and culture, but it may also depend on the measurement technique and parameter estimation method. We need to audit our experimental methods by conducting simulations at the experimental design stage to ensure the reproducibility of the experimental results.

In this paper, we perform parameter recovery simulations to 1) analyse whether discounting behaviours can be detected based on the standard errors associated with the estimates, and 2) evaluate the resolution of the parameter estimates from the distribution of the estimates. Then, we show that the combination of the CTB method and the quasi-hyperbolic discounting model cannot obtain estimates of the present bias parameter with small errors or correctly detect the bias if the actual effect size is small. Moreover, we discuss the reasons for the low resolution of the present bias parameter estimation.

Results

Setup

We now consider the decision-making problems associated with allocating the initial endowment $m = 20$ between the sooner and later periods. Let (c_t, c_{t+k}) denote an allocation bundle where c_t is the payoff for the sooner period t and c_{t+k} is for the $k = 70$ days later period. It only matters whether the sooner period t is 0 (i.e., present) or not, and for $t > 0$, the value of t does not matter, at least in our model. The exchange rate from tokens to material payoffs varies between the sooner and later periods, and we normalize the rate for the later period to be 1. We denote the exchange rate for the sooner payoff as $1 + r$, where $r \in [-0.4, 1]$ is an interest rate. We assume that income is exhausted, or the budget constraint binds the allocation bundle. Here, we can obtain the budget constraint for the decision problem as follows:

$$(1 + r)c_t + c_{t+k} = m. \tag{1}$$

To measure an individual's time preference, the experimenter asks the participants for their allocation (c_t, c_{t+k}) with changing t, k, and r.

We assume that each individual's time preference is represented by the following constant intertemporal elasticity of substitution and quasi-hyperbolic discounting (CES-QHD) utility function[8,9] for the intertemporal decision-making task described above:

$$U(c_t, c_{t+k}) = \frac{1}{\rho} c_t^{\rho} + \beta^{\mathbf{1}_{t=0}} \delta^k \frac{1}{\rho} c_{t+k}^{\rho}. \tag{2}$$

The variable $\mathbf{1}_{t=0}$ is an indicator for whether the sooner period is the present period. The parameter $\delta (> 0)$ is the 1-day discount factor, and the parameter $\beta (> 0)$ represents the present/future bias. The parameter ρ controls the curvature of the utility function and characterizes the intertemporal elasticity of substitution $\sigma = (1 - \rho)^{-1}$. The parameter ρ is sometimes interpreted as the degree of constant relative risk aversion in the risk context.

Our objective is to evaluate the errors in estimating the parameters δ and β from a dataset of (c_t, c_{t+k}) generated using a utility function for some ground-truth values. For the ground-truth values, we used 10 equally spaced values for δ and β from the range $0.9912 \leq \delta \leq 1.0025$ and $0.85 \leq \beta \leq 1.12$, respectively. Each interval from which we draw δ and β uses the ranges of the parameter distributions reported by AS. For the ground-truth curvature $\ln \sigma$, we used seven equally spaced values from the range $0.33 \leq \ln \sigma \leq 5.00$. We perturbed the generated data by adding a random number that follows a normal distribution with mean 0 and standard deviation $s \in \{0.01, 0.05, 0.10, 0.15, 0.20\}$. As the ratio of mean absolute deviation to standard

deviation is $\sqrt{2/\pi} \approx 0.8$, the generated data have, on average, a 0.8% error for the interval length allowed as a decision c_t for $s = 0.01$.

Detectability of Time Discounting

As a first measure to discuss the estimation error, we examined whether the estimated discount factor $\widehat{\delta}$ and the present/future bias parameter $\widehat{\beta}$ are distinguishable from 1, which indicates that the individual does not discount (or put a premium on) future payoffs. In previous studies, most attention has been paid to whether present-biased behaviour exists. We examine how far the true β is away from 1 to determine whether it can be distinguished from 1.

Fig. 1 shows the percentage of successfully rejected null hypotheses such that $\widehat{\delta} = 1$ and $\widehat{\beta} = 1$. For the discount factor parameter δ, when the ground-truth value is less than 0.9962, we can reject $\widehat{\delta} = 1$ in over 90% of cases regardless of the amount of added noise. For the case of $\delta > 1$, it may be more challenging to reject null hypotheses compared with the case of $\delta < 1$. To estimate δ accurately to place a premium on future payoffs, it is necessary to collect decision data at negative interest rates. In our simulation, however, we selected fewer interest rates in such a range, and therefore, the estimation accuracy was worse than that in the case of $\delta < 1$. For the present/future bias parameter β, in general, we may have more difficulty concluding that the estimates are not equal to 1 compared with the case of the discount factor parameter δ. Even when the true β is as small as 0.85, the success rate is below 90% for $s > 0.05$.

Error Size

To examine further the errors of the parameter estimates, instead of focusing on the estimated uncertainty of the parameter estimates for each individual, we analyse the actual variation of the estimates in a population with the same true parameter values. Here, we assume a population in which the three parameters—δ, β, and $\ln\sigma$—are distributed on a three-dimensional grid of ground-truth values that we set. Then, we check the distribution of estimates of each parameter in this population.

Fig. 2 shows the distribution of the estimated values of δ and β as a box plot (see Supplementary Analysis 1 for the $\ln\sigma$ estimates). If the box is above or below the red line, where the error in the estimate is 0, then the estimate is biased. In most cases, we find that the magnitude of the bias falls within the interquartile range of the estimates' distribution.

In addition to the existence of bias, we need to understand the resolution of the estimates. If the estimation is obtained using a higher resolution, we can precisely distinguish between any two individuals, even if the true parameter values are similar. By comparing the boxes' lengths (i.e., interquartile range), we consider the resolution: the minimum distance between the actual parameter values to be identifiable.

For the discount factor parameter δ, the whiskers of the estimates for any two adjacent ground-truths do not overlap and can be distinguished from each other for the smallest noise level $s = 0.01$. Even for the most extensive noise $s = 0.20$, the boxes do not overlap, whereas the whiskers do. We conclude that the experimental tasks considered in our simulations have enough resolution that as long as the distance between the true δ values of any two individuals is at least the ground-truth value spacing (1.3×10^{-3}), then we can distinguish between them, even assuming relatively large amounts of noise.

In contrast to the case of δ, the resolution of the present/future bias parameter β is generally not high. For $s = 0.01$, the whiskers for any two adjacent ground-truths do not overlap in most cases and can be only barely distinguished. However, whiskers and boxes often overlap when the noise is more prominent than for $s = 0.01$. For $s = 0.20$, the boxes overlap unless the true values of β are at least 0.1 away from each other. In the case of β, unlike the case of δ, we found that when comparing the magnitude of β for any two individuals using the experimental task we are addressing, the two individuals cannot be distinguished unless their true β values are farther apart than normally assumed.

Relative to the range of the prior distribution of β that we usually assume, the significant variance of the estimates suggests the possibility of errors. It has been argued that focusing only on statistically significant results using statistical tests with low power can lead to overestimation of effect sizes[20]. A meta-analysis of present bias parameter estimation indicated that the reported effect is strong such that it is suspected to be publication bias in the studies based on real effort tasks[10]. Our results raise further concerns regarding the overestimation of the present bias effect because greater noise in the estimation produces lower power regarding the statistical tests.

Why is the present bias estimation resolution low?

In the utility function, δ and β appeared as the term $D = \beta\delta^k$ (for $t = 0$). Fig. 3 shows a scatter plot of the estimated values of δ and β (for $\ln\sigma = 2.67$ and $s = 0.01$; see Supplementary Analysis 2 for the scatter plots including all $\ln\sigma$ and s) and a red line that satisfies $\beta\delta^{70} = 1$. Note that all points have been offset so that the ground-truth values coincide with $\delta = \beta = 1$ (indicated by the red cross). The points are distributed close to the red line. As the data used for parameter estimation include those at not only $t = 0$, but also $t > 0$, in principle, it is possible to distinguish between δ and β. In practice, however, even though the value of D itself can be estimated with reasonable precision, its components, δ and β, cannot be identified.

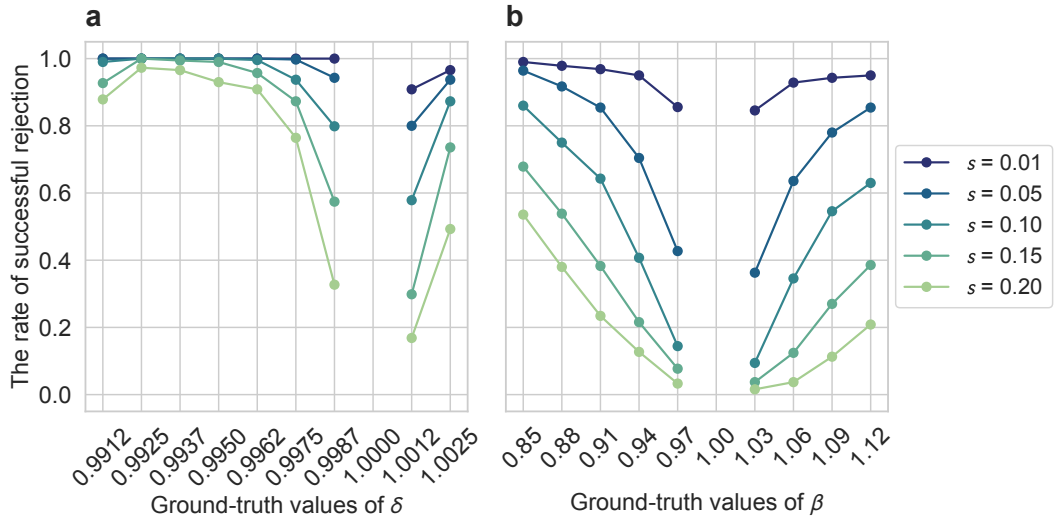

Figure 1. Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively. We conducted two-tailed Student's t-tests at the 5% significance level to examine whether the null hypothesis could be rejected for each simulation agent. We plotted each point focusing on a specific ground-truth δ (β) and a certain noise size s, after which, each point contained 10 replications of all combinations of ground-truth values of β (δ) and $\ln\sigma$, i.e., 700 simulation agents. We computed the test statistic using the standard error of the estimate, which is estimated by the jackknife method.

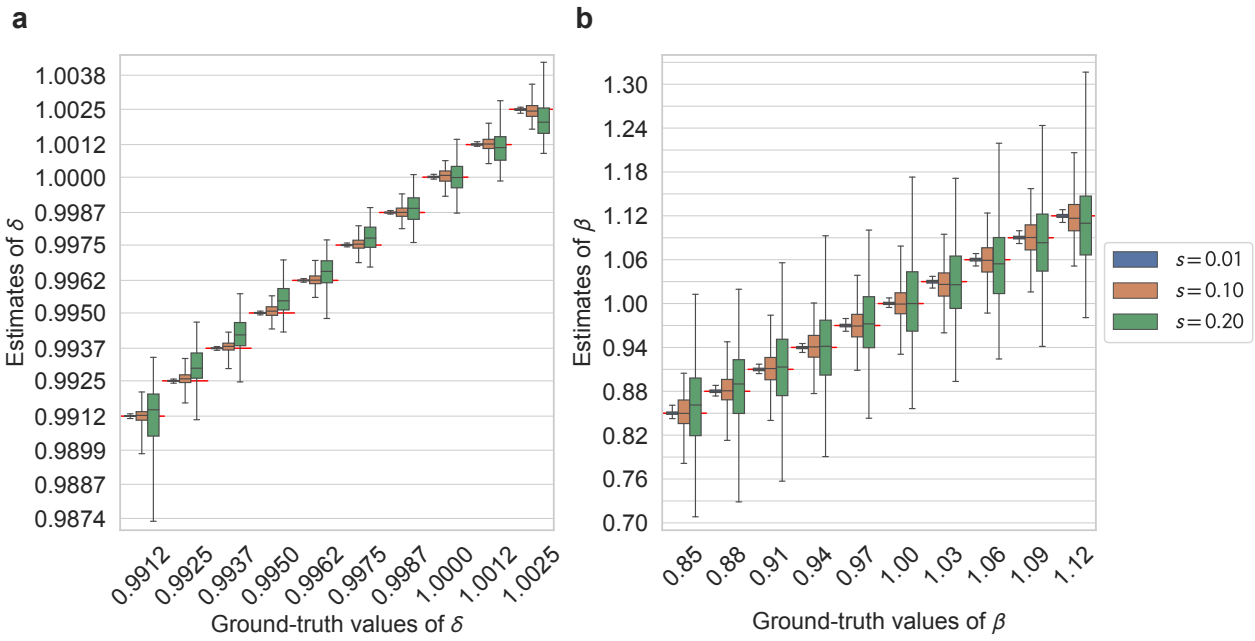

Figure 2. Box plot of the estimates of **a**) δ and **b**) β. Each box plot uses specific values of ground-truth δ (β) and certain noise size s, and each plot summarizes 10 replications of all combinations of ground-truth values of β (δ) and $\ln\sigma$, i.e., 700 simulation agents. The two ends of the box represent the first and third quartiles, respectively, and the two ends of the whiskers represent the 5th and 95th percentiles, respectively. On the red line, the error of the estimate is 0.

As $\mathrm{d}D/D = \mathrm{d}\beta/\beta + k\,\mathrm{d}\delta/\delta$, a 1% change in β results in a 1% change in D, but a 1% change in δ results in a k% change in D. As the difference between $\delta = 1$ and $\delta = 0.9987$ is 0.13%, the variation in D is 9.1% for $k = 70$. However, $\beta = 0.97$ is 3% smaller than $\beta = 1$ and yields a 3% variation for D, which is three times smaller than that for the δ case. The effect of the variation in D on decisions (or the demand function) depends on the curvature parameter $\ln\sigma$ and price $1 + r$. For example, in Fig. 4, we can compare the differences in decisions between individuals with $\ln\sigma = 2.67$ and different δ and β. We see that the discount behaviour of $\delta = 0.9987$ is larger than that of $\beta = 0.97$. Given the noise, it is more challenging to test whether the estimated β is less than 1 for an individual whose true β is 0.97 than whether the estimated δ is less than 1 for an individual whose true δ is 0.9987, because the difference in decisions is three times smaller. In the previous subsection, we observed that assuming significant noise $s = 0.20$, the resolution of δ is about 1.3×10^{-3}, which corresponds to the spacing of our ground-truth values, while the resolution of β is 0.1, which corresponds to about three times the spacing of our ground-truth values.

Eventually, low resolution of β estimation occurs because we try to identify values within a very narrow scope (or range of prior distribution) with a high precision. As is clear from the comparison of demand curves in Fig. 4, when the true difference in the values of β is less than 0.1, it is inherently difficult to identify individuals regardless of the econometric method used because the differences in behaviour are small (see Supplementary Analysis 3). In the expanded β scope, it is possible to distinguish between two individuals with given ground-truth values (see Supplementary Analysis 4). Although δ intuitively seems to require a very high resolution because it is a daily discount factor, it is actually possible to estimate it with sufficient resolution because its scope is broader than that of β. Because k depends on the scale of δ, we must expand the scope of δ if we make k a weekly discount factor. It should be noted that changing the value of k does not improve the resolution of β's estimation (see Supplementary Analysis 5).

Discussion

This paper evaluates the error of estimates for the time discounting parameters obtained using the CTB experiment[4] by conducting parameter recovery simulations[11]. Figs. 1 and 2 show that for the time discount factor δ, we have enough estimation resolution that $\delta = 0.9987$ is discriminable from $\delta = 1$. However, the precision of the estimate of β, which represents the present/future bias, is not good, and it is more difficult to conclude that the estimated $\beta = 0.97$ is smaller than 1, compared with the estimated $\delta = 0.9987$. Using a CTB experiment, we have tried to identify differences in β that are so small that discrimination is impossible. Considering the differences in behaviour corresponding to the differences in parameter values (see Fig. 4), we realize that the true β must be smaller than 0.9 to detect that the estimate is smaller than 1. In other words, when attempting to detect present bias using the problem set we used in our simulations, it can only be detected for individuals who discount future payoffs by more than 10%. Given the low resolution of the β estimation, it may be prone to overestimating or underestimating the effect of behavioural bias by chance, making publication bias more problematic. To be clear, we are not arguing that the CTB experiment itself should not be used, or that the CES-QHD model should not be used. This paper suggests the importance of understanding resolution for the design of experimental tasks.

Even if the differences in behaviour are small and drowned out by noise, adding more tasks can offset the effect of noise and improve the precision of the estimation (see Supplementary Analysis 6). However, it would be impractical to increase the number of tasks any further, given the workload on the participants in the experiment. In fact, the CTB experiments conducted following the original work[4] have tended to reduce the number of tasks[10].

It may be possible to improve the estimation precision by modifying how we generate the tasks instead of increasing the number of tasks (see Supplementary Analysis 6). The variables we can manipulate in generating the tasks are the sooner date t, the delay period k, and the price ratio $1 + r$. Primarily, there seems to be room for improvement in the generation of price variations. The problem set in our simulations includes prices smaller than 1 (i.e., the interest rate is negative) because we want to consider the parameters for individuals who place a premium, rather than a discount, on future payoffs. Suppose it is possible to ignore the existence of individuals with such unusual preferences and exclude them from the analysis. In that case, we can increase the number of tasks in positive interest rate domains by reducing the negative interest rate. In addition to modifying the price generation method, the variation of interval length k also matters in minimizing the estimation error, although the value itself is not an essential issue.

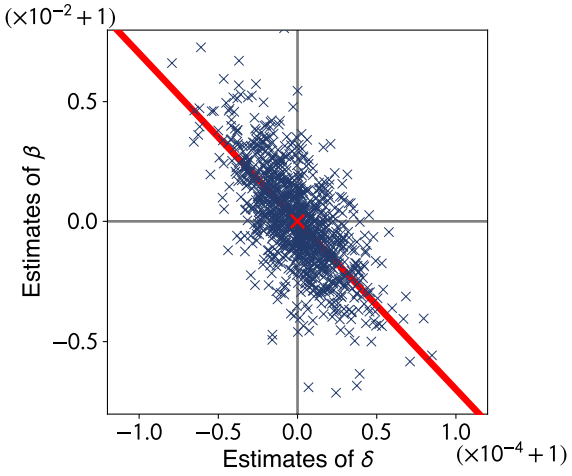

Figure 3. Scatter plot of each pair of estimated δ and β for $\ln\sigma = 2.67$ and $s = 0.01$. Each point is shifted so that the pair of corresponding ground-truth values coincides the red marked point $(\delta, \beta) = (1, 1)$. Both axes are on a logarithmic scale centred at 1. On the red line, $\beta\delta^{70} = 1$ is satisfied.

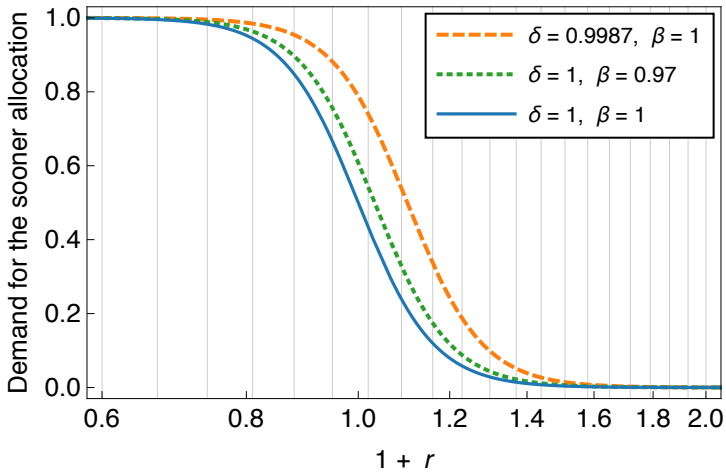

Figure 4. Demand curve for an individual whose curvature parameter is $\ln\sigma = 2.67$. The demand curve represents the relationship between the price $1 + r$ and the amount individuals are willing to allocate to the sooner period. The horizontal axis representing price $1 + r$ is on a logarithmic scale. Note that the individual faces the decision problem of allocating between now ($t = 0$) and $k = 70$ days later with the prices indicated by the vertical lines in the figure. The difference in behaviour when only δ decreases from 1 to 0.9987 is the difference between the blue and orange dashed curves. The difference when only β decreases from 1 to 0.97 is the difference between the blue and green dotted curves.

Methods

Behavioural Model

Here we discuss a theoretical model of participants' behaviour (c_t, c_{t+k}) for a given CTB experiment task $(t, k, 1+r, m)$. We suppose that an individual whose preferences are represented by the CES-QHD utility function $U(c_t, c_{t+k} \mid \delta, \beta, \ln\sigma)$ faces the utility maximization problem subject to the budget constraint $(1+r)c_t + c_{t+k} = m$. By solving, we obtain the following demand function:

$$g(t,k,1+r,m \mid \delta,\beta,\ln\sigma) = \begin{cases} \frac{1}{1+\left(\beta\delta^k\right)^{\sigma}\left(1+r\right)^{\sigma-1}} & \text{for} \quad t=0, \\ \frac{1}{1+\left(\delta^k\right)^{\sigma}\left(1+r\right)^{\sigma-1}} & \text{for} \quad t>0. \end{cases} \tag{3}$$

Note that the value of g corresponds to the sooner allocation c_t divided by its upper limit $m/(1+r)$, and thus, the function g maps into the interval $[0,1]$.

We perturbed the generated normalized sooner allocation $g(\bullet)$ by adding a random number ε, which follows a normal distribution with mean 0 and standard deviation $s \in \{0.01, 0.05, 0.10, 0.15, 0.20\}$. In the original experiment by Andreoni and Sprenger[4] (henceforth AS), participants were asked to select an integer in the interval from 0 to 100 as a normalized allocation, which corresponds to the value of g multiplied by 100. Given that forcing discrete choice causes rounding errors in decision-making, an error size of $s = 0.01$ is inevitable. We estimated the root mean squared error (RMSE) on the normalized sooner allocation from the AS experimental dataset. Regarding the distribution of the RMSE, the first quartile is 0.019, the median is 0.14, and the third quartile is 0.22. Comparing the distribution of the RMSE with the average size of the noise, $s = 0.20$ is not necessarily too large.

We truncated the noise-added value to the interval $[0,1]$. Thus, the noise-added data are always the interior points of the budget constraint line, and no end-points are chosen. It has been pointed out that if agents tend to choose extreme allocations, it is problematic to fit their data to the demand function using the least-squares method[13]. By truncating instead of censoring, we avoided this problem. However, the structure of noise associated with actual human decision-making needs to be discussed separately. While it is possible that our modelling for noise does not capture human behaviour well, it is at least consistent with the estimation method using nonlinear least squares as well as AS, as discussed below.

Experimental Tasks

We have two experimental situations (defined as a combination of an early period date t and a delay length k) for the experimental task: $t = 0$ (i.e., present) and $k = 70$ (days), and $t = 1$ (i.e., not present) and $k = 70$ (days). The delay length k is typically on the scale of weeks to months, and is rarely shorter than 1 week[10]. In each situation, we set 21 uniformly spaced prices chosen from $0.6 \leq 1+r \leq 2$. We fixed income m at 20 for simplicity. The number of tasks, i.e., the number of data points for each individual, is 42.

There are three critical differences between our problem set and the problem set used by AS. The first difference is that we chose the price $1+r$ from the range where the interest rate r is not only positive, but also negative. Few studies using CTB experiments, including AS, ask participants about negative interest rates. However, without asking about negative interest rates, it is impossible to estimate the discount factor for an individual who does not discount the future payoffs, but who does place a premium on them (for such an individual, the discount factor δ will be greater than 1). We also conducted simulations using the AS's original problem set and summarized the results in Supplementary Analysis 6.

Second, the delay k was set equal to 70 in this paper to simplify the discussion. Note that for AS's problem set, there were three conditions of k: 35, 70, and 98. We also investigate the effect of the number of conditions on k in Supplementary Analysis 6. Increasing or decreasing the variation in k affects the parameter estimation error. We confirm that increasing the variety of prices $1+r$ instead of increasing the variety of k leads to improved estimation precision.

Third, we limited the number of early period dates t to two, i.e., present or not present. Note that in AS's experiment, participants made decisions about the allocation between the future and a later future for $t = 7$ and $t = 35$ separately. Regarding the CES-QHD utility function model, there is no difference in decisions between $t = 7$ and $t = 35$, but it may affect real behaviour. For a real experimental design, it may be helpful to designate the variation in t to treat for bias in the parameter estimates, but we discarded that option.

Ground-truth Values

Table 1 shows the ground-truth parameter values that generate the decision data. We chose values for δ, β, and $\ln\sigma$; these values are evenly spaced as if from a uniform distribution. By combining the ground-truth values of the three parameters listed in Table 1, there are 700 synthetic individuals. As mentioned above, there are five noise levels s, and we generate 10 sets of data for each s, resulting in decision data for 35,000 agents.

Table 1. List of values used as ground-truth values for the three parameters characterizing the CES-QHD utility function and the standard deviation values for the added noise. We have also listed the corresponding ρ values for the curvature parameter $\ln\sigma$.

δ	0.9912	0.9925	0.9937	0.9950	0.9962	0.9975	0.9987	1.0000	1.0012	1.0025
β	0.85	0.88	0.91	0.94	0.97	1.00	1.03	1.06	1.09	1.12
$\ln\sigma$	0.33	1.11	1.89	2.67	3.44	4.22	5.00			
(ρ)	(0.283)	(0.671)	(0.849)	(0.931)	(0.968)	(0.985)	(0.993)			
s	0.01	0.05	0.10	0.15	0.20					

For the discounting parameters δ and β, we selected from a range covering the distribution of the estimates reported in AS's paper. Usually, the discount factor δ and the present bias β are assumed to be less than 1. However, because some studies report individuals with estimates greater than 1, we also included such values in our set of possible ground-truth values. We obtained estimates from the AS's experimental dataset. For the distribution of δ, the 5th percentile is 0.9917, the median is 0.9989, and the 95th percentile is 1.0018. For the distribution of β, the 5th percentile is 0.89, the median is 1.01, and the 95th percentile is 1.15.

We selected values of the curvature parameters $\ln\sigma$ from 0.33 ($\rho = 0.283$; nearly the Cobb–Douglas utility curvature) to 5 ($\rho = 0.993$; nearly linear curvature). We next describe in detail how we selected the curvature parameters. For mathematical clarity, we use $\ln\sigma = \ln(1/1-\rho)$ instead of the commonly used notation ρ to discuss the curvature parameter. The domain of $\ln\sigma$ is all real numbers, whereas the parameter ρ is defined on the domain $(-\infty, 1]$. Let us assume that $\ln\sigma = 0$ (or $\rho = 0$), which corresponds to the Cobb–Douglas utility function, is the centre of the curvature parameter space. For $\ln\sigma > 0$, the intertemporal allocations become substitutive, and complementary otherwise. As $\ln\sigma \longrightarrow -\infty$ (or $\rho \longrightarrow -\infty$), the utility function goes to a Leontief function: $U = \min\{c_t, c_{t+k}\}$, whose indifference curve is L-shaped, and which is known as the perfect complement utility function. As $\ln\sigma \longrightarrow +\infty$ (or $\rho \longrightarrow 1$), the utility function goes to a linear function: $U = c_t + \beta^{\mathbf{1}_{t=0}}\delta^k c_{t+k}$, which is the perfect substitution utility function.

Previous studies have reported that the curvature of participants' preferences in CTB experiments is generally linear. Regardless of the distribution of the actual parameter values, we should also check the estimation errors for individuals who would behave relatively complementarily because the utility function model does not explicitly exclude such individuals. However, when the curvature of the standard CES utility function $U(x,y) = (x^\rho + \phi y^\rho)^{1/\rho}$ is negative, the parameter ϕ, which corresponds to the discounting part $\beta^{\mathbf{1}_{t=0}}\delta^k$ in the CES-QHD utility, is no longer interpretable[21]. As the estimation errors of δ and β are inevitably large for $\ln\sigma < 0$, we excluded them from our analysis. Consequently, we chose ground-truth values for $\ln\sigma$ from 0.33 to 5. We also investigated the estimation errors in the same way for $\ln\sigma < 0$ and reported the results in Supplementary Analysis 7. Note that previous studies on the curvature of time preferences report that it is rare to observe individuals for whom $\ln\sigma$ is negative, regardless of whether the CTB method is used[4,18,22,23].

Estimation Methods

As we described above, for all individuals i characterized by $(\delta_i, \beta_i, \ln\sigma_i)$, and for all budget constraint lines $j \in \{1,\ldots,42\}$, we obtain the decision data $\widetilde{c}_i^j = g(t_j, k_j, 1+r_j, m_j \mid \delta_i, \beta_i, \ln\sigma_i) + \varepsilon$. Given the generated data, we estimated the three parameters using a nonlinear least squares method. Following AS, we used the "`nl`" command in Stata. Mathematically, we search for the values of $\widehat{\delta}$, $\widehat{\beta}$, and $\widehat{\ln\sigma}$ that minimize the sum of squared residuals:

$$\sum_{j=1}^{42}\left[\widetilde{c}_i^j - g\left(t_j, k_j, 1+r_j, m_j \mid \delta_i, \beta_i, \ln\sigma_i\right)\right]^2. \tag{4}$$

Although the curvature parameter $\ln\sigma$ is defined over all real numbers, outside of a specific range, the effect of $\ln\sigma$ variation on behaviour $g(\bullet)$ saturates: for an additional increase (decrease) in $\ln\sigma$, we cannot observe differences in the decision data and therefore we cannot observe an increase in substitutability (complementarity) as a behaviour (see Supplementary Analysis 8). We assume that $\ln\sigma$ less than -2.5 indicates perfect complements and $\ln\sigma$ greater than 5.5 indicates perfect substitutes. Therefore, we transformed the equation for $\ln\sigma$ to $\ln\sigma = f(\theta) = 4\tanh(\theta) + 1.5$ and estimated the latent variable θ. Through this manipulation, the estimated $\ln\sigma$ ranges between -2.5 and 5.5. Using the function $f(\theta)$, we obtained the $\ln\sigma$ estimates from the estimated θ.

For the parameter estimation, we set the convergence criterion as 10^{-5} and the maximum number of iterations as 200. By combining all parameters $(\delta_i, \beta_i, \ln\sigma_i)$ and s, there are 3,500 synthetic individuals, and we regenerate the decision data 10 times for each synthetic individual. Of the 3,500 synthetic individuals, 3,483 converged all 10 times, and the remaining 17 individuals had only one failure to converge.

References

1. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10**, 1625–1633 (2007).
2. McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 503–507 (2004).
3. O'Donoghue, T. & Rabin, M. Present Bias: Lessons Learned and To Be Learned. *Am. Econ. Rev.* **105**, 273–279 (2015).
4. Andreoni, J. & Sprenger, C. Estimating Time Preferences from Convex Budgets. *Am. Econ. Rev.* **102**, 3333–3356 (2012).
5. Augenblick, N., Niederle, M. & Sprenger, C. Working over Time: Dynamic Inconsistency in Real Effort Tasks. *The Q. J. Econ.* **130**, 1067–1115 (2015).
6. Carvalho, L. S., Meier, S. & Wang, S. W. Poverty and economic decision-making: Evidence from changes in financial resources at payday. *Am. Econ. Rev.* **106**, 260–84 (2016).
7. Blumenstock, J., Callen, M. & Ghani, T. Why Do Defaults Affect Behavior? Experimental Evidence from Afghanistan. *Am. Econ. Rev.* **108**, 2868–2901 (2018).
8. Laibson, D. Golden Eggs and Hyperbolic Discounting. *The Q. J. Econ.* **112**, 443–478 (1997).
9. O'Donoghue, T. & Rabin, M. Doing It Now or Later. *Am. Econ. Rev.* **89**, 103–124 (1999).
10. Imai, T., Rutter, T. A. & Camerer, C. F. Meta-analysis of present-bias estimation using convex time budgets. *The Econ. J.* **131**, 1788–1814 (2020).
11. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *Elife* **8**, e49547 (2019).
12. Andreoni, J. & Sprenger, C. Risk Preferences Are Not Time Preferences. *Am. Econ. Rev.* **102**, 3357–3376 (2012).
13. Harrison, G. W., Lau, M. I. & Rutström, E. E. Identifying time preferences with experiments: Comment (2013). https://cear.gsu.edu/wp-2013-09-identifying-time-preferences-with-experiments-comment/.
14. Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. Discounting behavior: A reconsideration. *Eur. Econ. Rev.* **71**, 15–33 (2014).
15. Cheung, S. L. Comment on "Risk Preferences Are Not Time Preferences": On the Elicitation of Time Preference under Conditions of Risk. *Am. Econ. Rev.* **105**, 2242–2260 (2015).
16. Epper, T. & Helga, F. D. Comment on "Risk Preferences Are Not Time Preferences": Balancing on a Budget Line. *Am. Econ. Rev.* **105**, 2261–2271 (2015).
17. Miao, B. & Zhong, S. Comment on "Risk Preferences Are Not Time Preferences": Separating Risk and Time Preference. *Am. Econ. Rev.* **105**, 2272–2286 (2015).
18. Andreoni, J., Kuhn, M. A. & Sprenger, C. Measuring time preferences: A comparison of experimental methods. *J. Econ. Behav. & Organ.* **116**, 451–464 (2015).
19. Andreoni, J. & Sprenger, C. Risk preferences are not time preferences: Reply. *Am. Econ. Rev.* **105**, 2287–2293 (2015).
20. van Zwet, E. W. & Cator, E. A. The significance filter, the winner's curse and the need to shrink. *Stat. Neerlandica* **75**, 437–452 (2021).
21. Thöni, C. A note on CES functions. *J. Behav. Exp. Econ.* **59**, 85–87 (2015).
22. Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. Eliciting risk and time preferences. *Econometrica* **76**, 583–618 (2008).
23. Cheung, S. L. Eliciting utility curvature in time preference. *Exp. Econ.* **23**, 493–525 (2020).

Acknowledgements

The authors thank Taisuke Imai and Nobuyuki Hanaki for helpful discussions. Y.S. acknowledges financial support provided through a Grant-in-Aid for JSPS Fellows (19J12097) from the Japan Society for the Promotion of Science. OnLine English (www.oleng.com.au) edited the manuscript to improve its readability.

Author contributions statement

K.S. and K.I. conceived the simulations, Y.S. and K.S. conducted the simulations, and Y.S. analysed the results and wrote the original manuscript. All authors reviewed the manuscript.

Data availability

The datasets generated during and/or analysed during this study are available from the corresponding author on reasonable request.

Additional information

Competing interests: The authors declare no competing interests.

Supplementary Analyses for "Investigation of the Convex Time Budget Experiment by Parameter Recovery Simulation"

Yuta Shimodaira*, Kohei Shiozawa†, and Keigo Inukai‡

August, 2022

SA.1 Box plots of the $\ln\sigma$ estimates

Fig. S1 shows box plots of the distribution of the estimated curvature parameter $\ln\sigma$, which was omitted from the main text. Each box focuses on a specific ground-truth $\ln\sigma$ and a certain noise size s, and then each plot summarizes 10 replications of all combinations of ground-truth values of δ and β, i.e., 1,000 simulation agents. The two ends of the box represent the first and third quartiles, respectively, and the two ends of the whiskers represent the 5th and 95th percentiles, respectively. For the red line, the error of the estimate is 0.

To ensure that the $\ln\sigma$ estimates converge, we transformed $\ln\sigma$ using the function $\ln\sigma = 4\tanh(\theta)+1.5$ and searched the latent variable θ. Through this manipulation, the estimated $\ln\sigma$ ranges between -2.5 and 5.5. The black horizontal dashed lines in the figure indicate the boundaries of $\ln\sigma$.

Contrary to the parameters δ and β, $\ln\sigma$ tends to have lower estimates, especially when the added noise is considerable. In the case of $s = 0.20$, $\ln\sigma$ appears to be saturated at about 2.2.

*Graduate School of Economics, Osaka University.
†Faculty of Economics, Takasaki City University of Economics.
‡Faculty of Economics, Meiji Gakuin University. E-mail: inukai@eco.meijigakuin.ac.jp

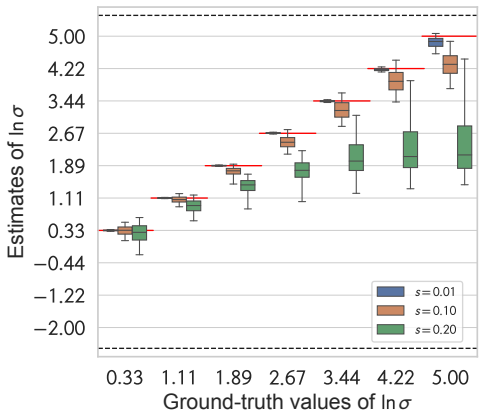

Figure S1: Box plots of the $\ln \sigma$ estimates.

SA.2 Scatter plots of estimated δ–β

Fig. S2 shows scatter plots of the estimated δ and β for each pair of specific ground-truth δ and β. Each δ–β plot includes estimates for all combinations of ground-truth $\ln\sigma$ and s. The darkness of the dots indicates the magnitude of the added noise; the darker the colour, the smaller the noise. The horizontal and vertical axes represent the δ and β estimates, respectively, and the grid shows the ground-truth values. Note that while the ground-truth values are linearly equally spaced, both axes use a logarithmic scale, so the grid lines are not exactly equally spaced. Fig. S2 is a matrix, with columns representing ground-truth δ and rows representing ground-truth β. A red cross indicates each pair of ground-truth values.

We have drawn a red curve (a straight line in log-log graphs) that satisfies $\beta\delta^{70} = \text{const.}$ and passes through the red crosses. The points in each scatter plot are distributed either near the red line or in a more vertical belt than the red line.

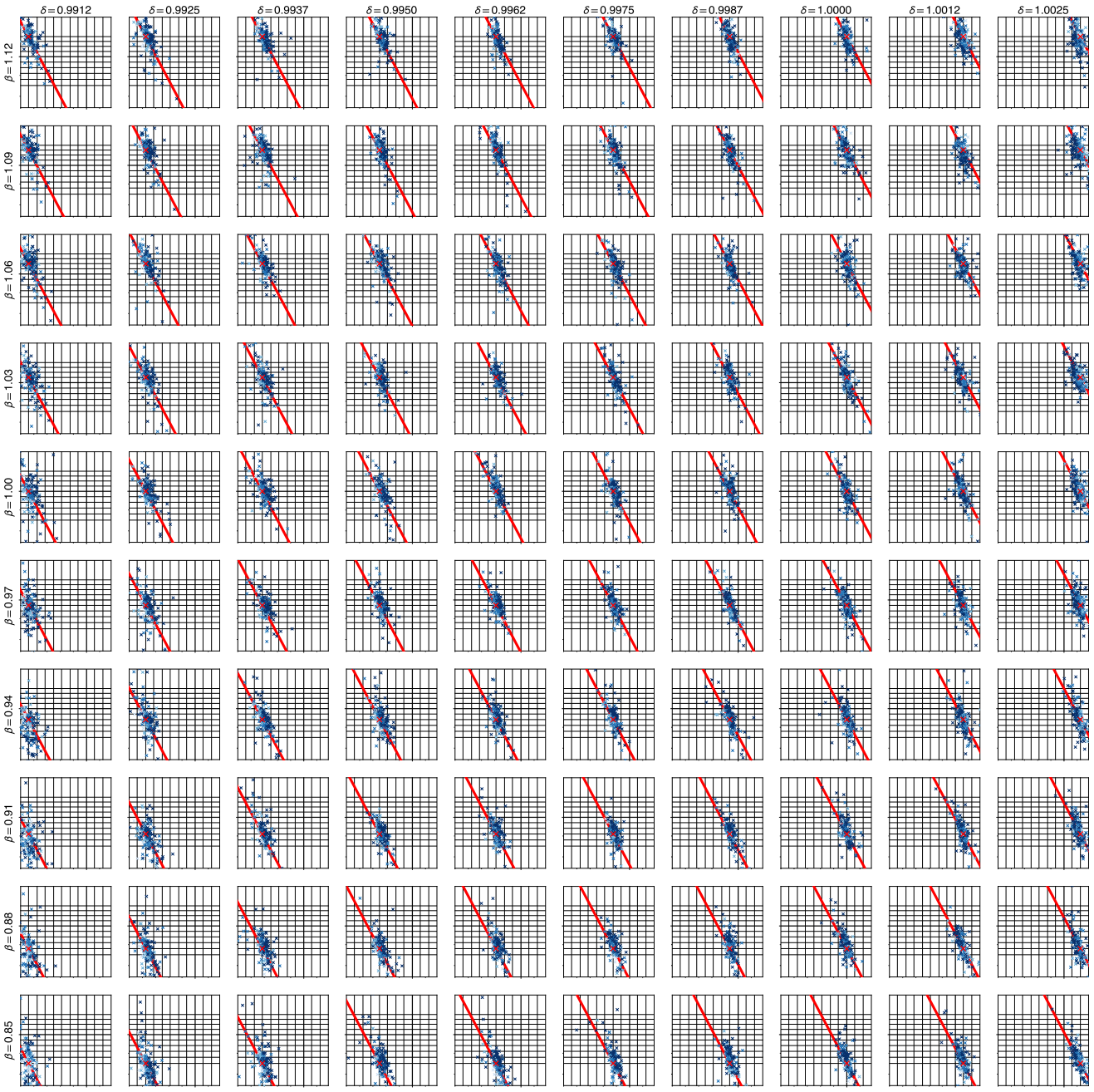

Figure S2: Scatter plots of estimated δ–β.

SA.3 Detect the present bias by using a paired two-sample t-test

We attempted to detect present bias by directly comparing the values of the decision data rather than by parameter β estimation. With two decisions required for the sooner periods, $t = 0$ and $t > 0$ for 21 different prices, we conducted a paired t-test for the two series. Fig. S3 shows the proportion of cases in which the two-tailed test rejected the null hypothesis that there was no difference in decisions at the 5% level. Compared with the success rates using the β estimates in the main analysis, it is more difficult to detect present bias when directly comparing decisions. Individuals whose utility function is almost linear will make drastic decisions, allocating everything to early periods on the one hand and late periods on the other, bounded by a specific price (the switching point). The difficulty of detection occurs because the effect we want to detect is drowned out by noise if we include data in the range where the decision does not change from that in the sooner period t.

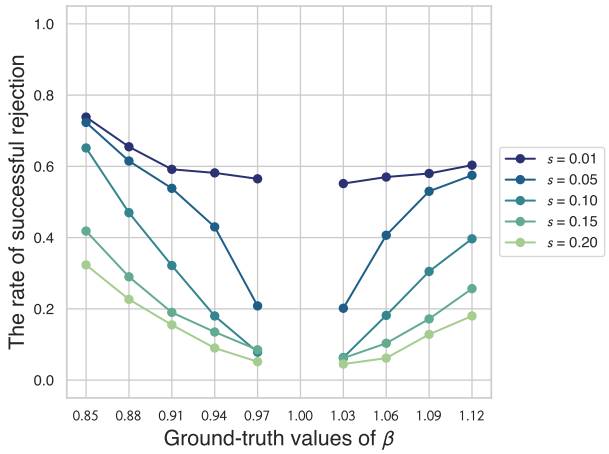

Figure S3: Rate of successful rejection of the null hypothesis that there was no difference in decisions for $t = 0$ and $t > 0$.

SA.4 Re-running the simulation with the expanded ground-truth β spacing

In the main analysis, we concluded that the poor accuracy of the β estimation reflected the required resolution, i.e., the spacing of the ground-truth values, being too narrow to identify in the first place. Here, we show the results of re-running the simulation with the spacing of the ground-truth β expanded by about three times. We used 10 equally spaced values from the range $0.50 \leq \beta \leq 1.40$ for the ground-truth β and the same ground-truth values for δ and $\ln \sigma$ as in the main analysis.

Fig. S4 shows box plots representing the distribution of the estimated δ, β, and $\ln \sigma$. The distributions of the δ and $\ln \sigma$ estimates do not change significantly from those in the main analysis, while the distribution of the β estimates shows that, based on the length of the box, the estimates are accurate enough to identify neighbours even when the noise size is $s = 0.20$.

Fig. S5 shows the rate of successful rejection of the null hypothesis that $\widehat{\delta} = 1$ and $\widehat{\beta} = 1$, respectively. From the graph of the success rates for β, we can see that if the true β is far enough from 1, we can correctly reject the absence of bias.

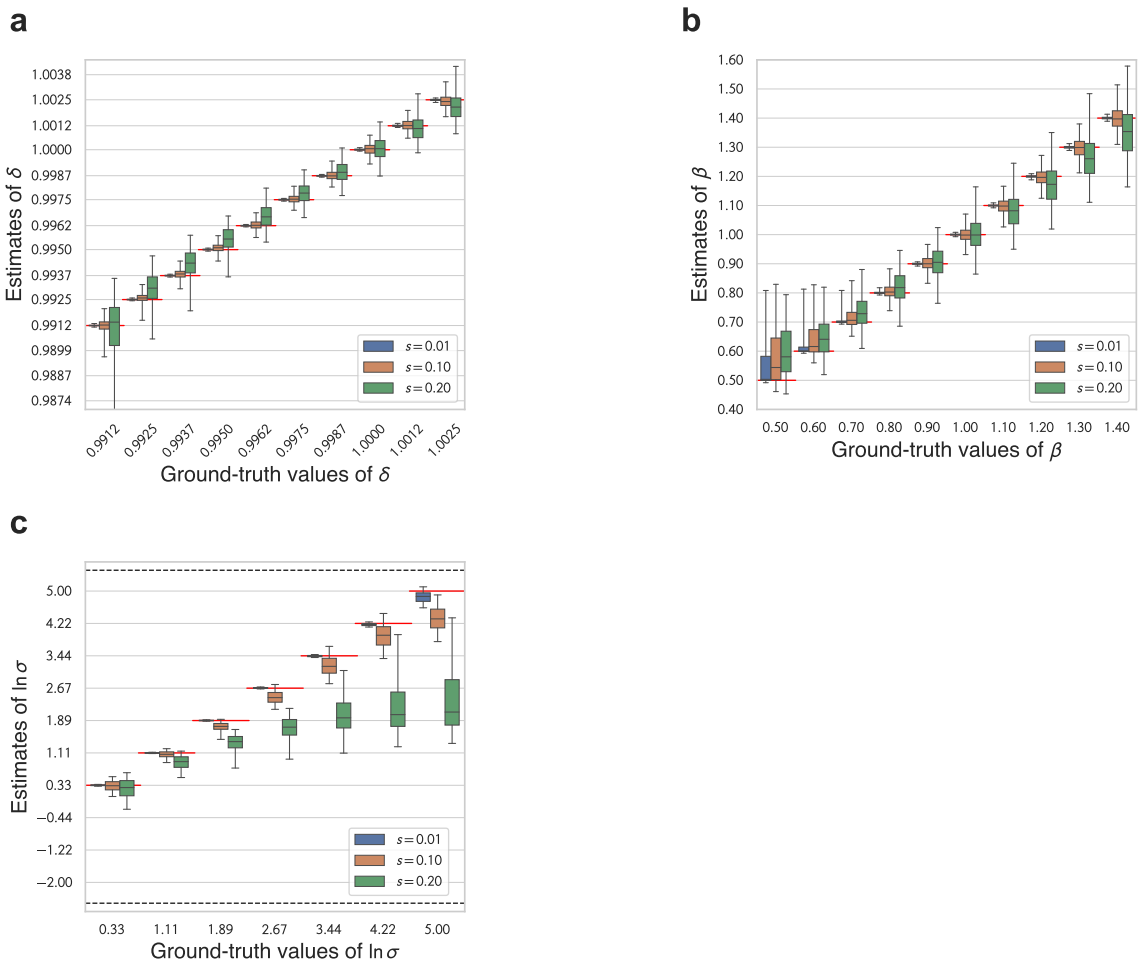

Figure S4: Box plots of **a**) δ, **b**) β, and **c**) $\ln \sigma$ estimates for an expanded ground-truth β range.

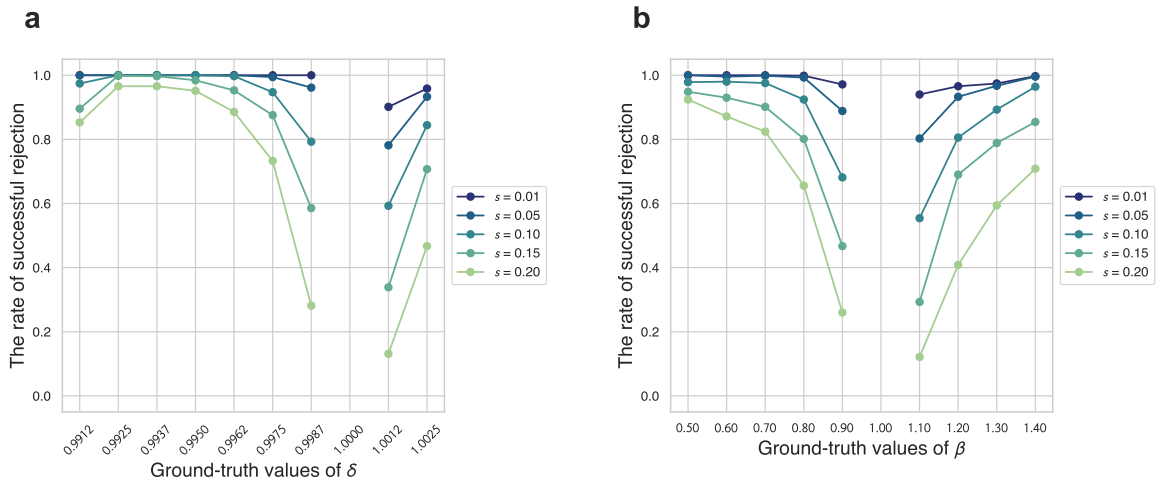

Figure S5: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for an expanded ground-truth β range.

SA.5 Re-running the simulation with $k = 1$

The value of k, which represents the length of the intertemporal period, depends on the scale of δ, and an enormous value of k has no intrinsic effect on the accuracy of parameter estimation. Let us assume that $k = 1$ "day". Figs. S6 and S7 show the results of the recovery simulation, where the data are generated with $k = 1$ for the same ground-truth values as in the main analysis. The resolution is very low for δ estimation. This suggests that, in an experimental setup where there is only 1 day between periods, it is challenging to estimate δ in the ground-truth range we have set. Compared with the figures for β shown in the main analysis, we can see that the estimation accuracy of β is almost the same for $k = 70$ as for $k = 1$.

Next, let us assume that $k = 1$ "period", where one period represents 70 days and δ is not a daily discount factor, but rather, a one "period" discount factor. In other words, set $k = 1$ and generate data using the ground-truth value of δ as the value in the main analysis to the power of 70. Figs. S8 and S9 show the results. Compared with the figures shown in the main analysis, we can see that in addition to the no change in resolution for β, there is little change in δ's resolution over the converted scope range.

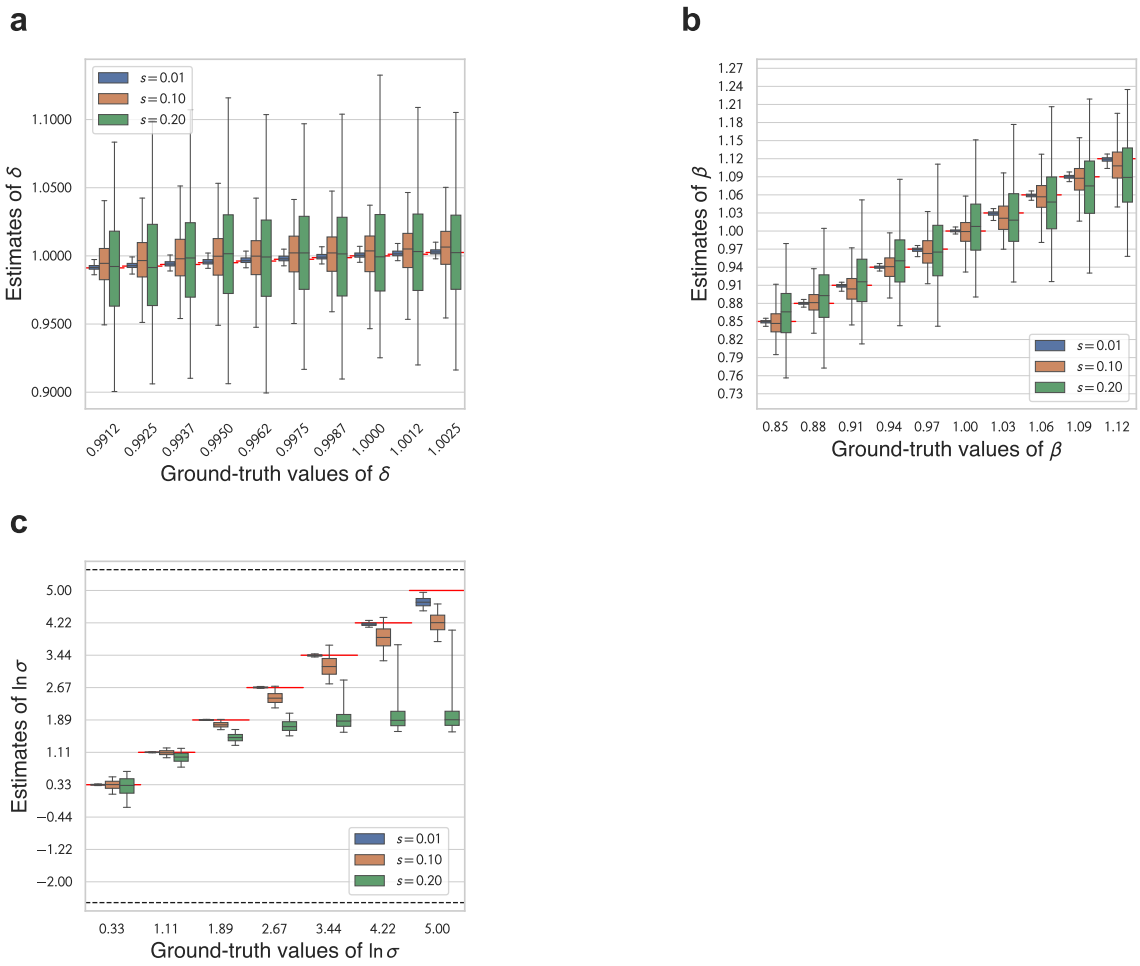

Figure S6: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for $k = 1$.

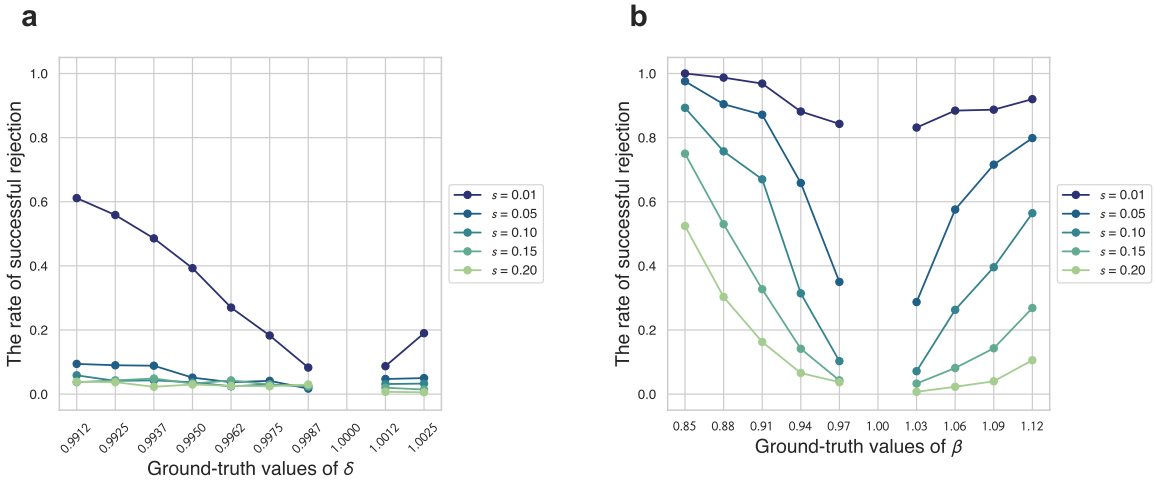

Figure S7: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for $k = 1$.

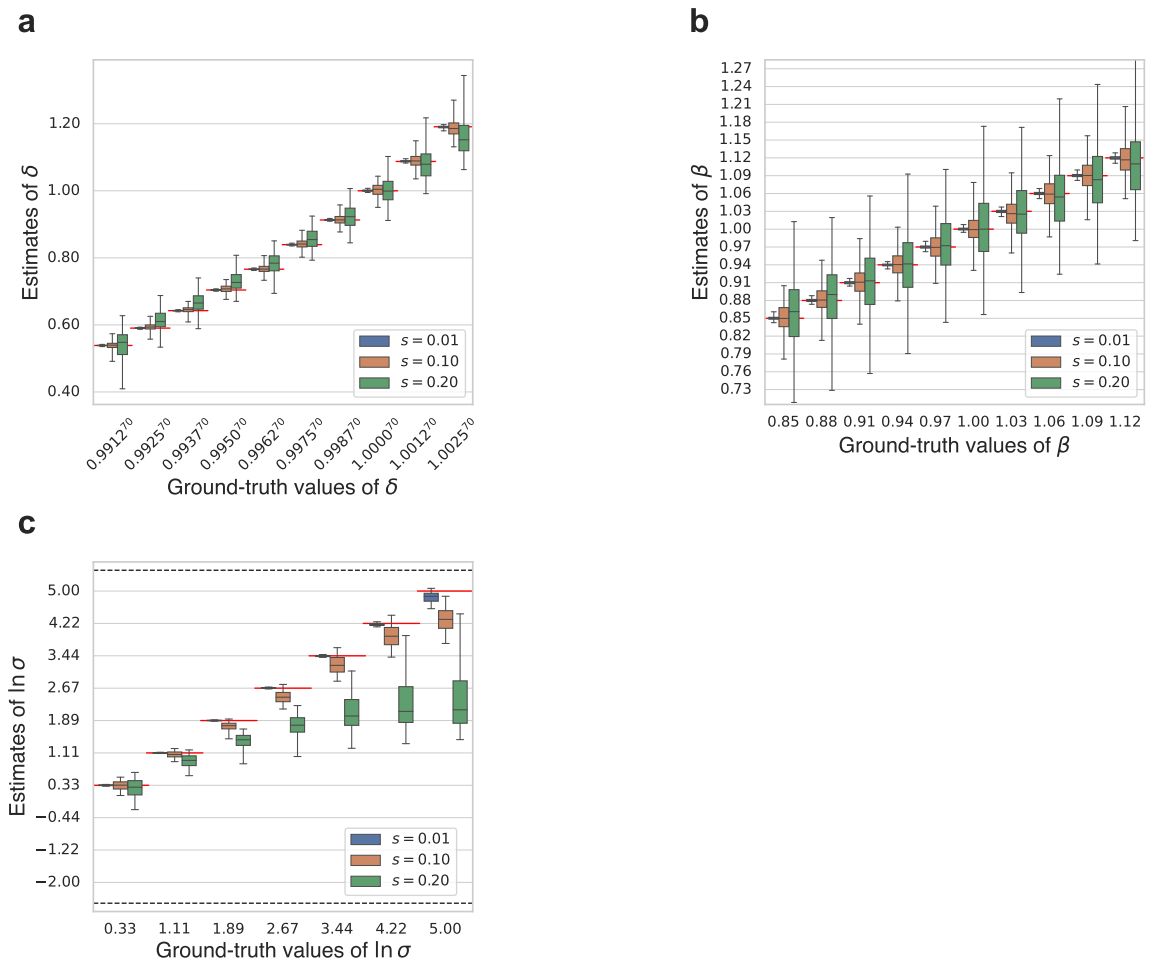

Figure S8: Box plots of **a**) δ, **b**) β, and **c**) $\ln \sigma$ estimates for $k = 1$ with an expanded ground-truth δ range.

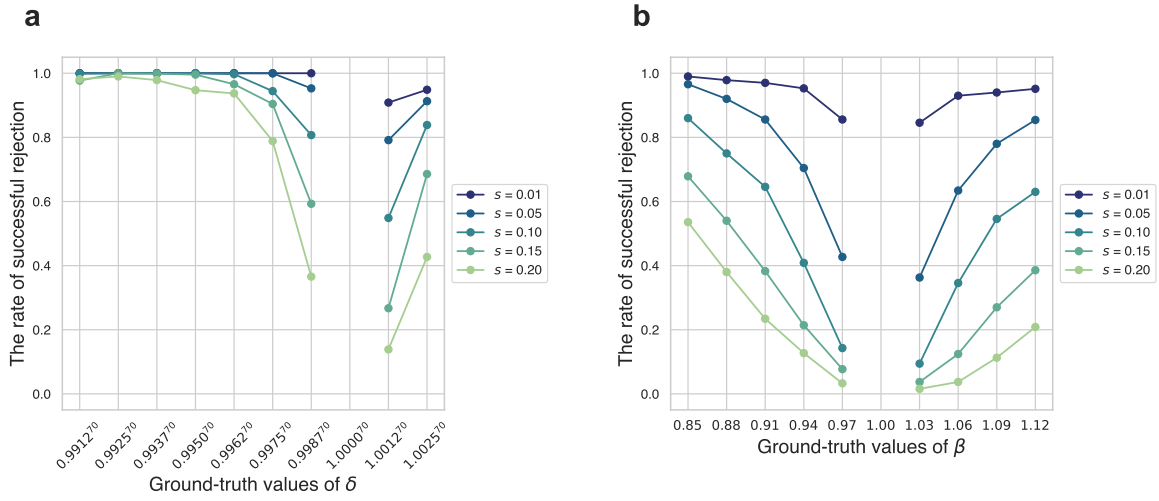

Figure S9: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for $k = 1$ with an expanded ground-truth δ range.

SA.6 Design of problem set

This section analyses how estimation accuracy varies depending on how the problem set is designed. The problem set used in the main analysis is referred to as PS0 in the following. For PS0, m is fixed to 20 and k is fixed to 70, and 21 uniformly spaced prices were drawn from the range $0.6 \leq 1+r \leq 2$ for $t=0$ and $t=1$, respectively, for a total of 42 problems in all.

First, we examine the effect of simply increasing the number of problems. Let PS1 be the problem set that withdraws 42 prices, or twice the number of PS0, from the same range of prices, and let PS2 be the problem set that withdraws 210 prices, or 10 times the number of PS0.

Next, we consider the strategy of changing the number of each variable types without increasing the number of problems. For problem set PS3, instead of increasing the number of k to three—35, 70, and 98—we set the number of prices $1+r$ to seven from the range $0.6 \leq 1+r \leq 2$ for each combination of t and k. For problem set PS4, instead of increasing the number of k to seven—14, 28, 42, 56, 70, 84, and 98—we set the number of prices $1+r$ to three—0.9, 1.2, and 1.5—for each combination of t and k. For problem set PS5, we fixed k to 70 and use 14 prices drawn from the range $0.6 \leq 1+r \leq 2$, once at $t=0$ and twice at $t=1$. This corresponds to setting two different t situations for $t>0$ (e.g., two decision makings for $k=70$: one between 7 and 77 days later and the other between 35 and 105 days later). For problem set PS6, we ignored the existence of individuals who place a premium on future payoffs, and did not consider negative interest rates; i.e., we fixed k to 70 and drew 21 prices from the range $1.05 \leq 1+r \leq 2$ for $t=0$ and $t=1$, respectively.

We conducted simulations using the problem sets PS1–PS6 as well as the original problem set used by Andreoni and Sprenger (2012) (henceforth AS). The problem set of AS is summarized in Table S1. For PS1 and PS2, we computed the standard error of the estimates using the inverse of the negative Hessian, whereas for PS3–PS6 and AS, we computed them using the jackknife method, as in the main analysis.

Figs. S10–S23 show the box plots of the estimates and the rate of successful rejection of the null hypothesis for the problem sets PS1–PS6 and AS. We also included the PS0 data as a dashed line in the success rate graph. Figs. S24 and S25 plot the absolute value of the error and the standard error of the estimate by problem set, respectively. We plotted the mean and median with bootstrap 95% confidence intervals. For the plot of the mean value, we used the logarithm scale because some values are enormous.

Comparing PS0 with PS1 and PS2, it is clear that estimation accuracy is improved for PS1 and PS2. Increasing the number of problems improves the accuracy. However, even with PS2, which has 420 tasks

in total, it is difficult to reject that the β estimate of an individual, whose actual value is 0.97, is not equal to 1 when the noise size is $s = 0.20$.

For PS3–PS5, it is difficult to conclude that any particular problem set is clearly superior to PS0. As PS6 and AS do not include prices with negative interest rates, the accuracy of δ estimation is inevitably very poor for individuals with $\delta > 1$, but improves in some cases for individuals with $\delta < 1$.

Table S1: Problem set of AS

t	k	$(1+r, m)$				
0	35	(1.05, 20)	(1.11, 20)	(1.25, 20)	(1.25, 25)	(1.43, 20)
0	70	(1.05, 20)	(1.11, 20)	(1.25, 20)	(1.25, 25)	(1.43, 20)
0	98	(1.05, 20)	(1.25, 20)	(1.25, 25)	(1.54, 20)	(2.00, 20)
7	35	(1.05, 20)	(1.11, 20)	(1.25, 20)	(1.25, 25)	(1.43, 20)
7	35	(1.05, 20)	(1.11, 20)	(1.25, 20)	(1.25, 25)	(1.43, 20)
7	70	(1.00, 20)	(1.05, 20)	(1.11, 20)	(1.25, 20)	(1.43, 20)
35	70	(1.05, 20)	(1.11, 20)	(1.25, 20)	(1.25, 25)	(1.43, 20)
35	98	(1.05, 20)	(1.25, 20)	(1.25, 25)	(1.54, 20)	(2.00, 20)
35	98	(1.05, 20)	(1.25, 20)	(1.25, 25)	(1.54, 20)	(2.00, 20)

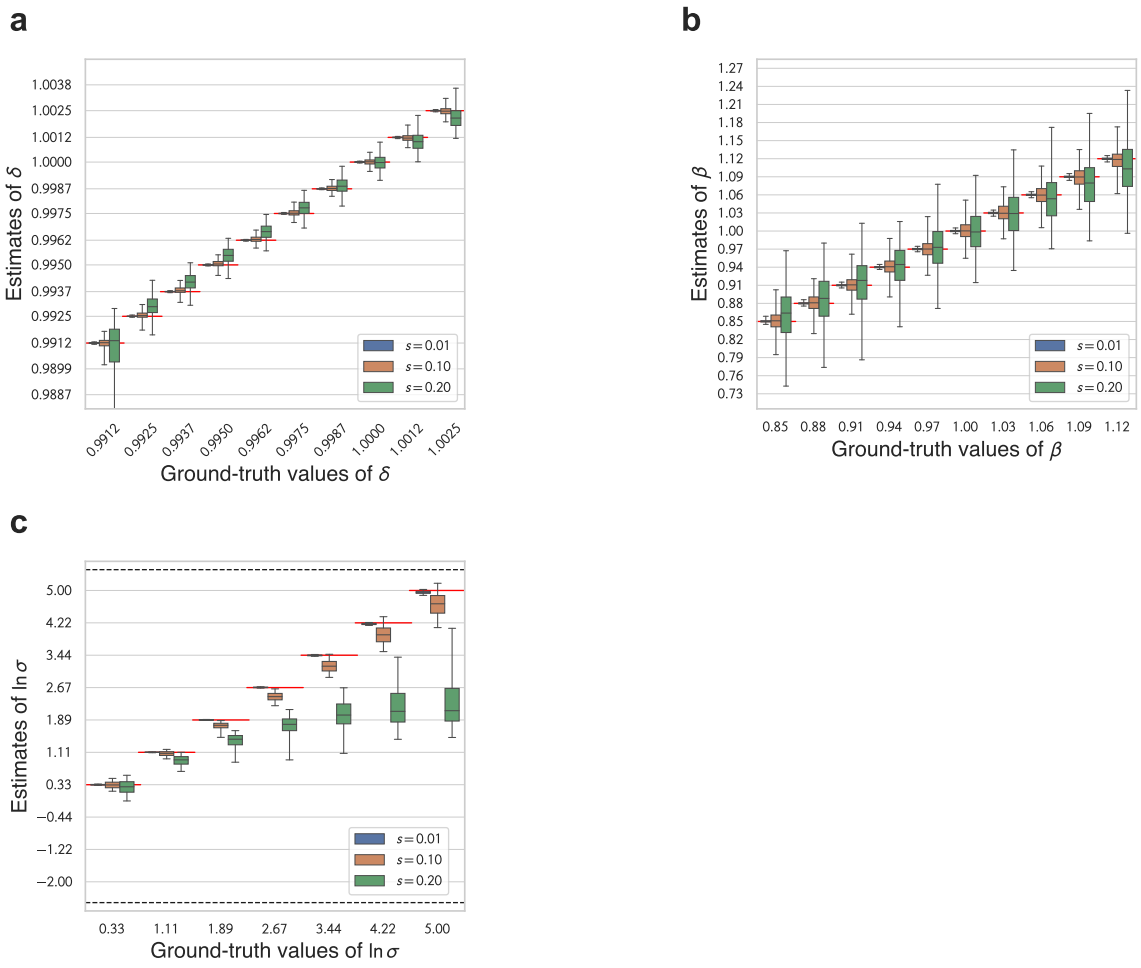

Figure S10: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for PS1.

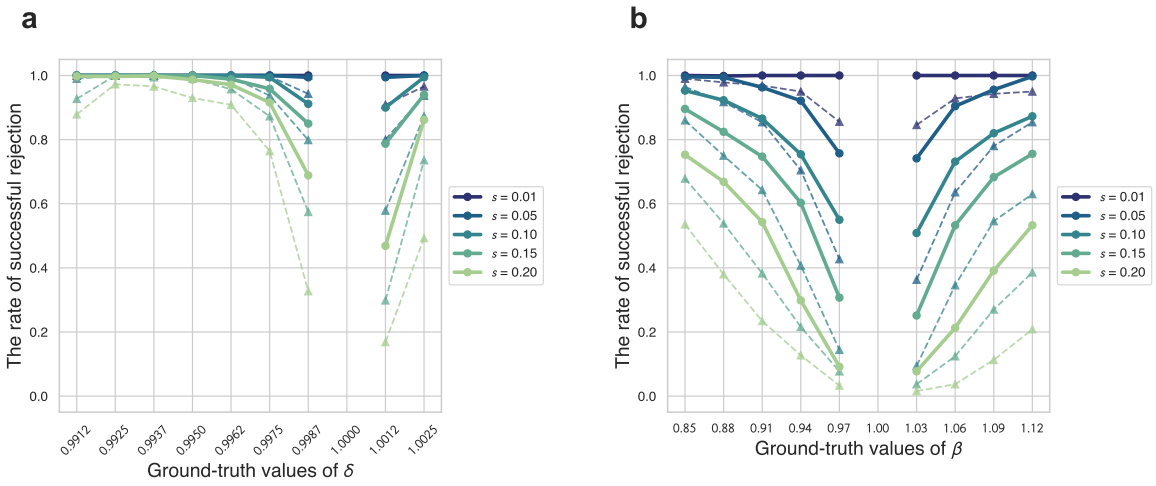

Figure S11: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for PS1.

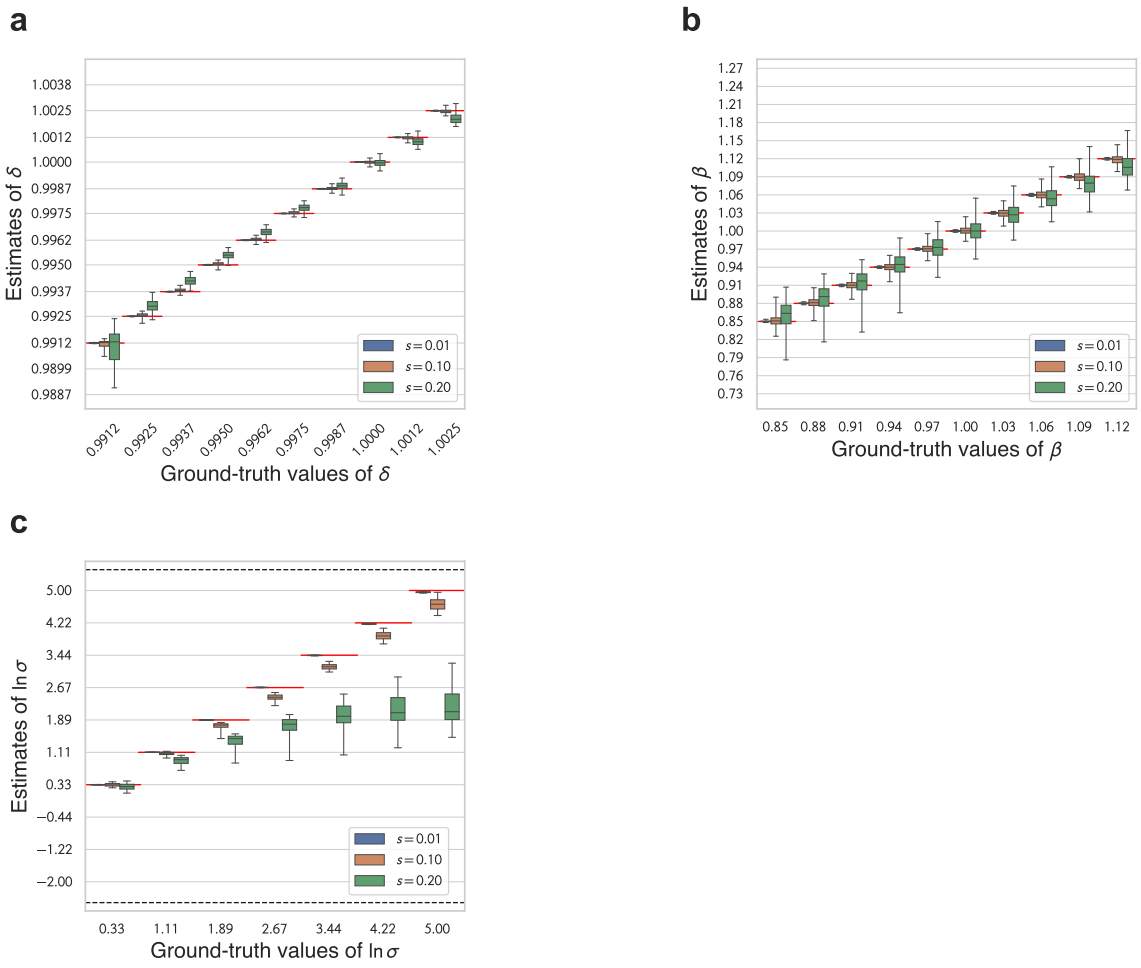

Figure S12: Box plots of **a**) δ, **b**) β, and **c**) $\ln \sigma$ estimates for PS2.

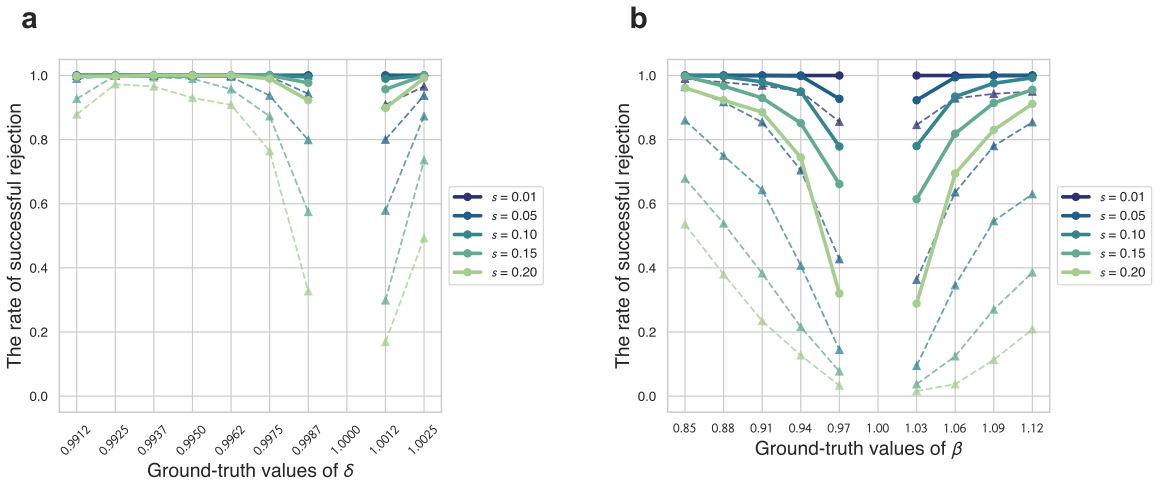

Figure S13: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for PS2.

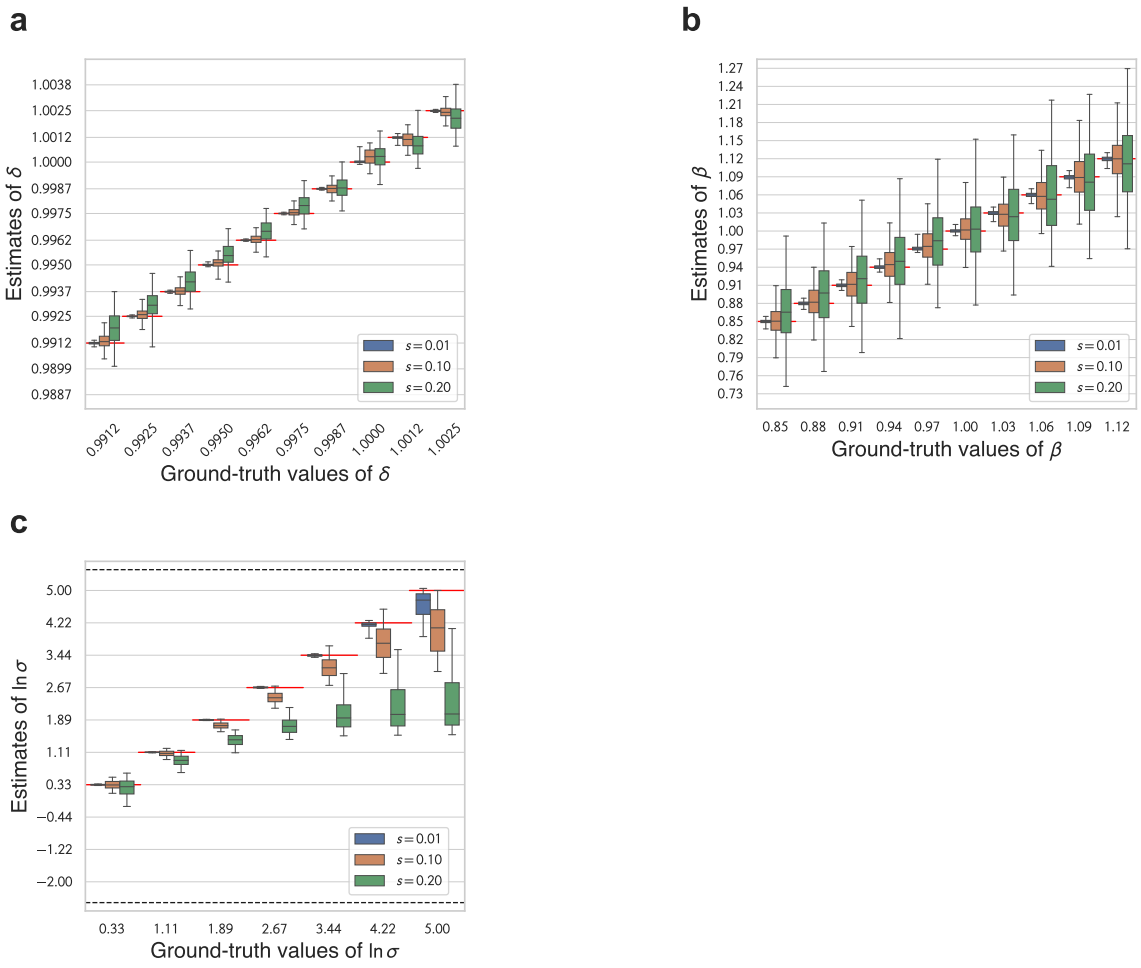

Figure S14: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for PS3.

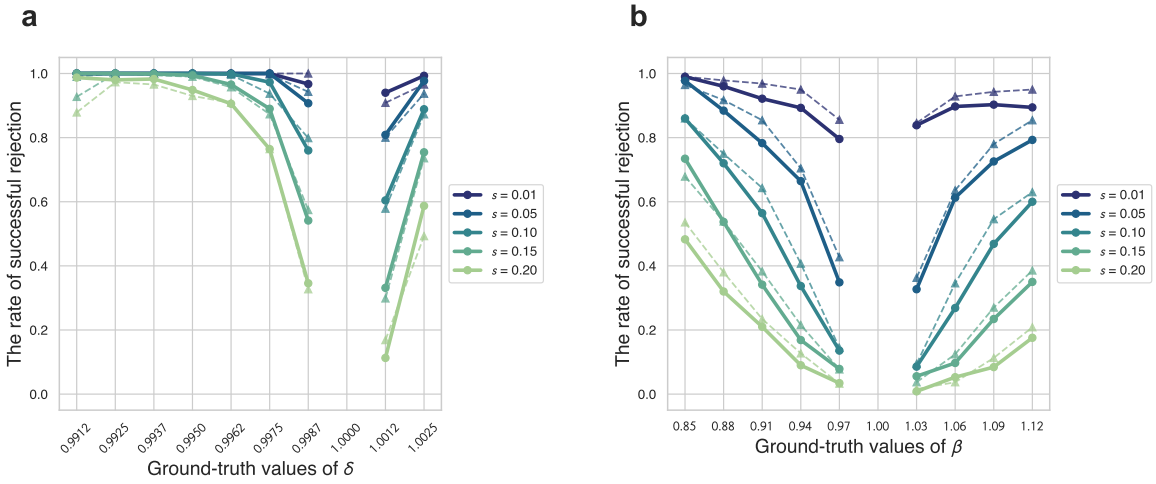

Figure S15: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for PS3.

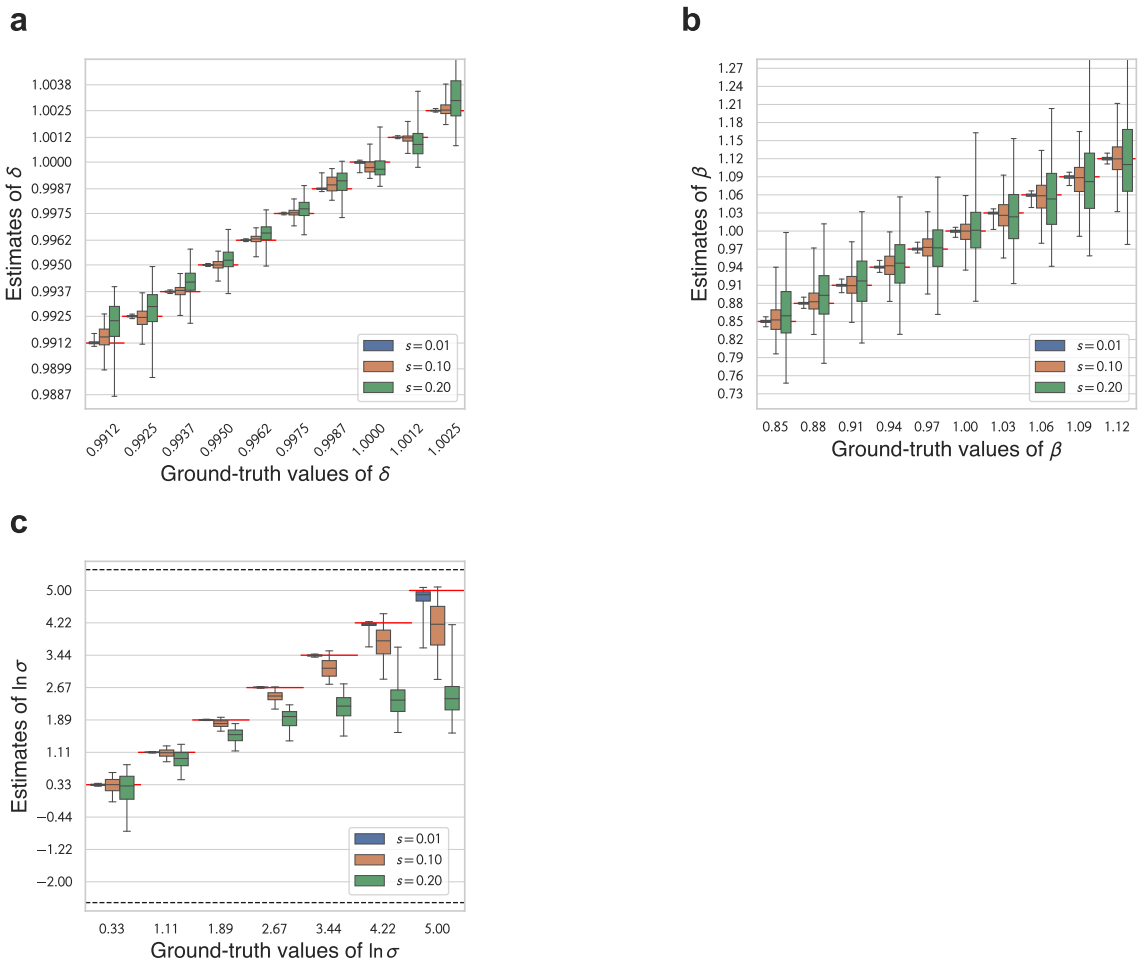

Figure S16: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for PS4.

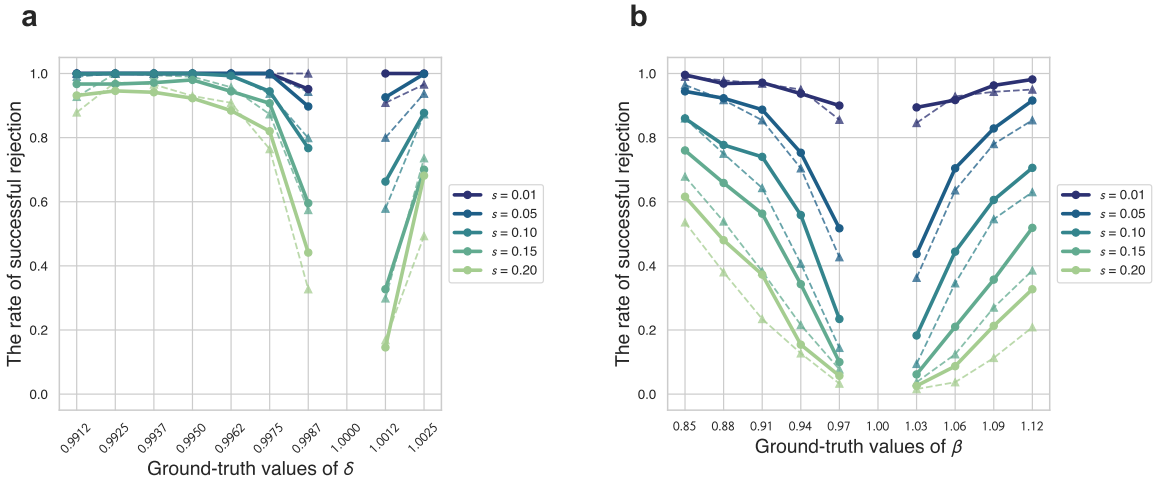

Figure S17: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for PS4.

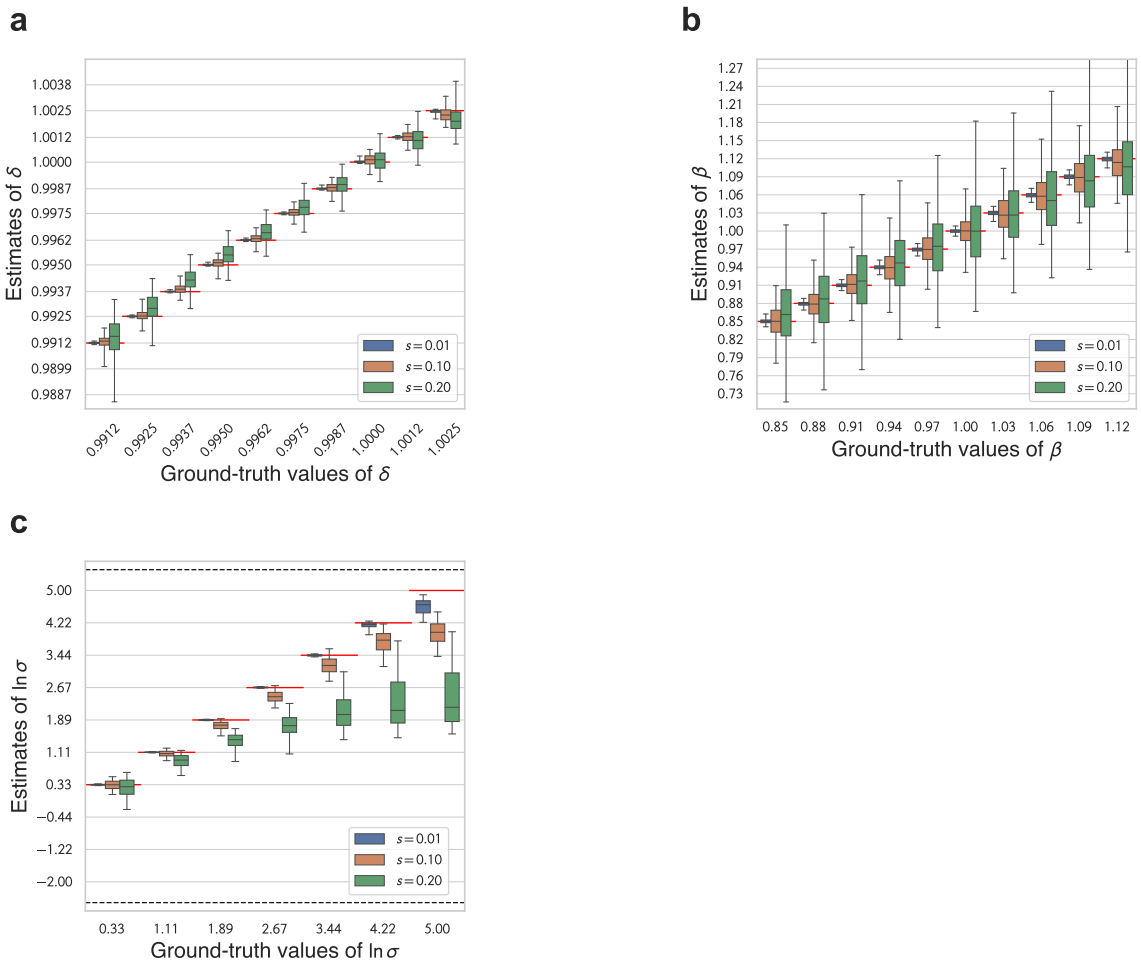

Figure S18: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for PS5.

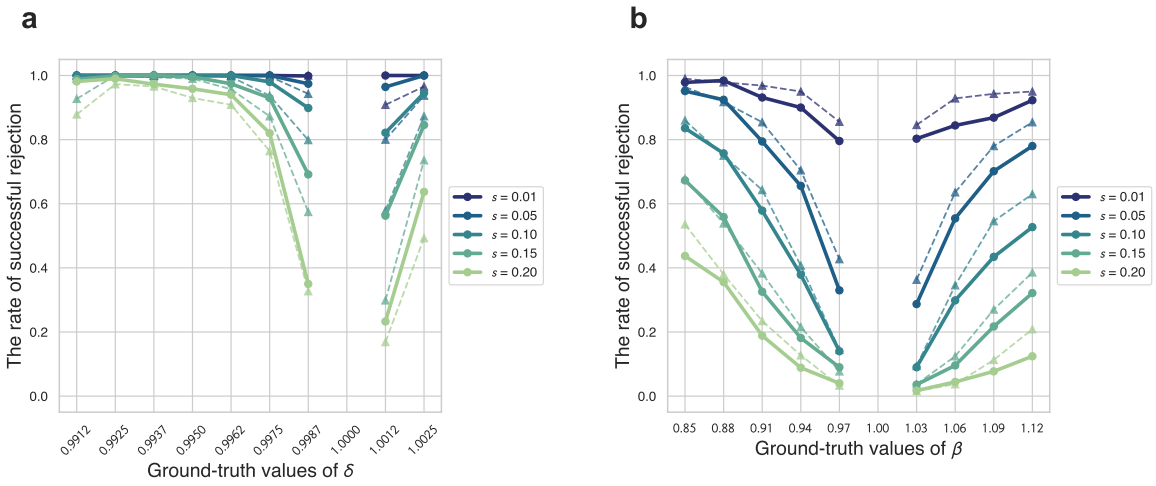

Figure S19: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for PS5.

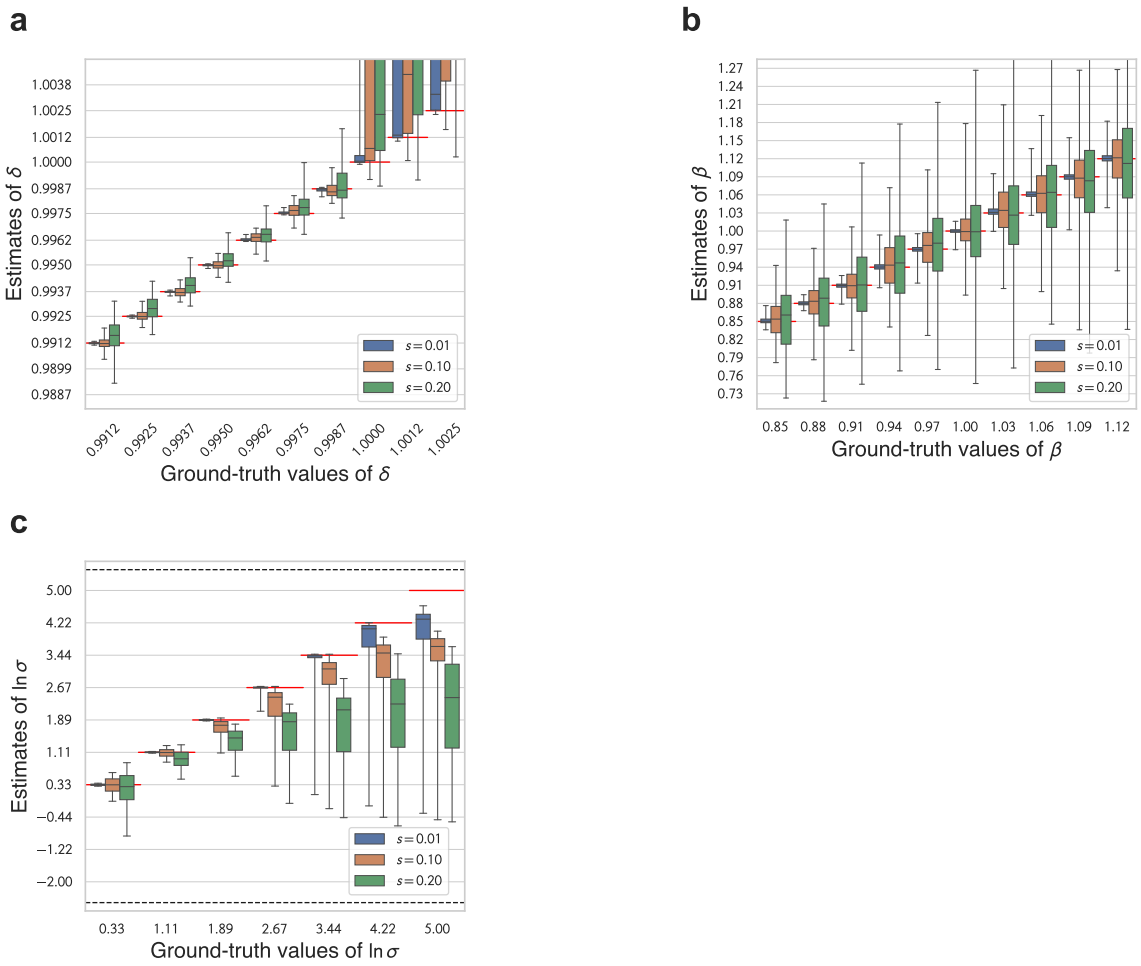

Figure S20: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for PS6.

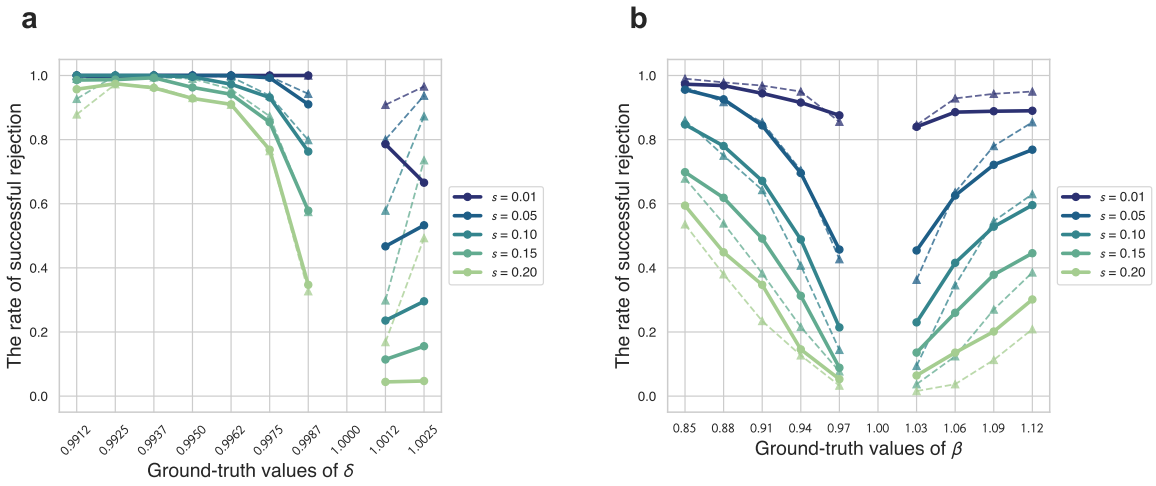

Figure S21: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for PS6.

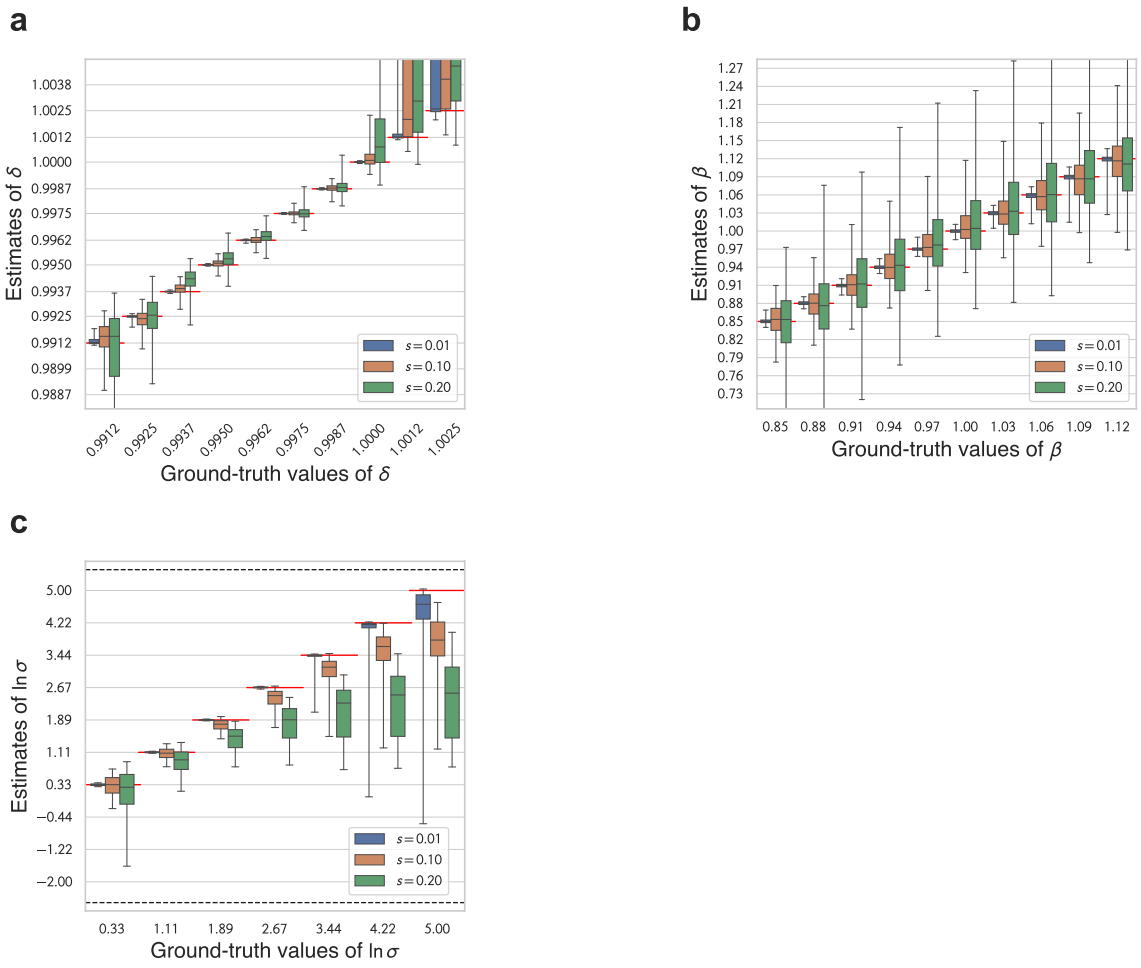

Figure S22: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates for AS.

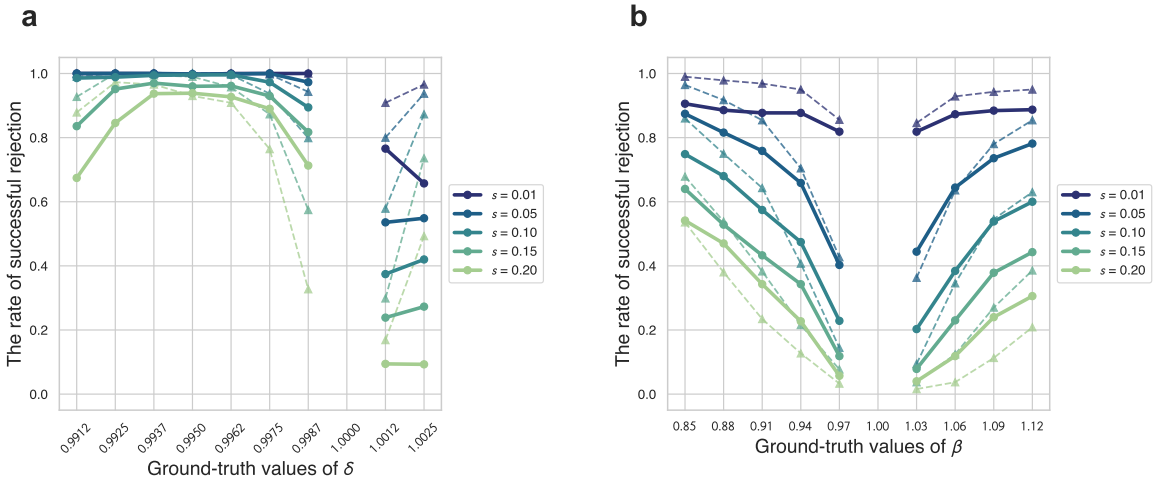

Figure S23: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, for AS.

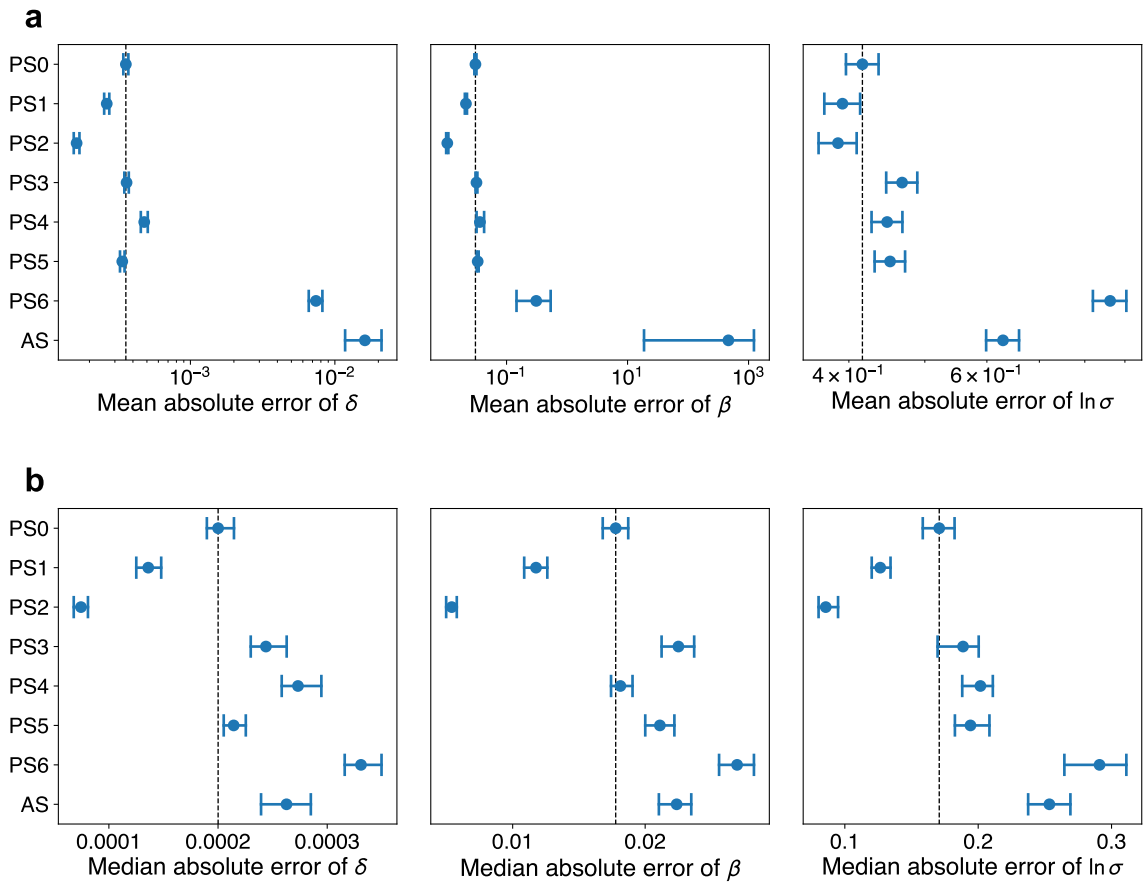

Figure S24: **a**) Mean and **b**) median of the absolute errors of the estimates for each problem set. Error bars represent the bootstrap 95% confidence intervals.

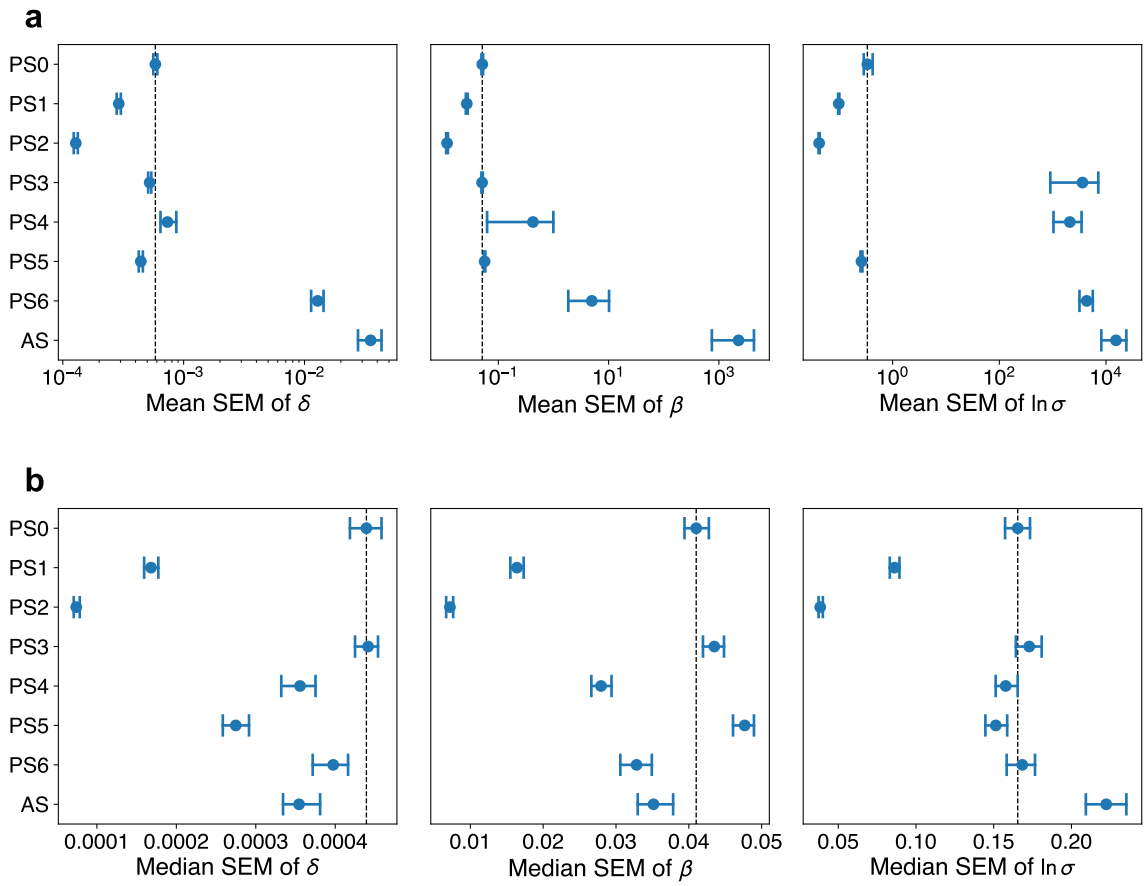

Figure S25: **a**) Mean and **b**) median of the standard errors of the estimates for each problem set. Error bars represent the bootstrap 95% confidence intervals.

SA.7 Case of negative ground-truth $\ln \sigma$

In the analysis in the main text, the ground-truth values of the curvature parameter $\ln \sigma$ were restricted to positive values, but here, we analyse the accuracy of parameter estimation for negative values of $\ln \sigma$ or the case of a lower elasticity. We used ground-truth values for $\ln \sigma$ of -2.00, -1.22, and -0.44, and the same ground-truth values for δ and β as in the main analysis.

Fig. S26 **a**) and **b**) show box plots representing the distribution of the estimated δ and β. Each plot summarizes 10 replications of all combinations of ground-truth values of β (δ) and $\ln \sigma$, i.e., 300 simulation agents. Although the median of the estimates is not significantly different from the case where $\ln \sigma$ is positive, the range of the distribution of the estimates (length of boxes and whiskers) is much larger than that in the case where $\ln \sigma$ is positive.

Fig. S26 **c**) shows box plots representing the distribution of the estimated $\ln \sigma$. Each plot summarizes 10 replications of all combinations of ground-truth values of δ and β, i.e., 1,000 simulation agents. For the case where ground-truth $\ln \sigma$ is positive (Fig. S1), even with a noise size of $s = 0.10$, the estimation was accurate enough not to overlap the whiskers of neighbouring plots, whereas the negative case is poor.

Fig. S27 shows the rate of successful rejection of the null hypothesis that $\widehat{\delta} = 1$ and $\widehat{\beta} = 1$, respectively. Each point contained 10 replications of all combinations of ground-truth values of β (δ) and $\ln \sigma$, i.e., 300 simulation agents. Compared with the case where $\ln \sigma$ is positive, the success rate is generally lower.

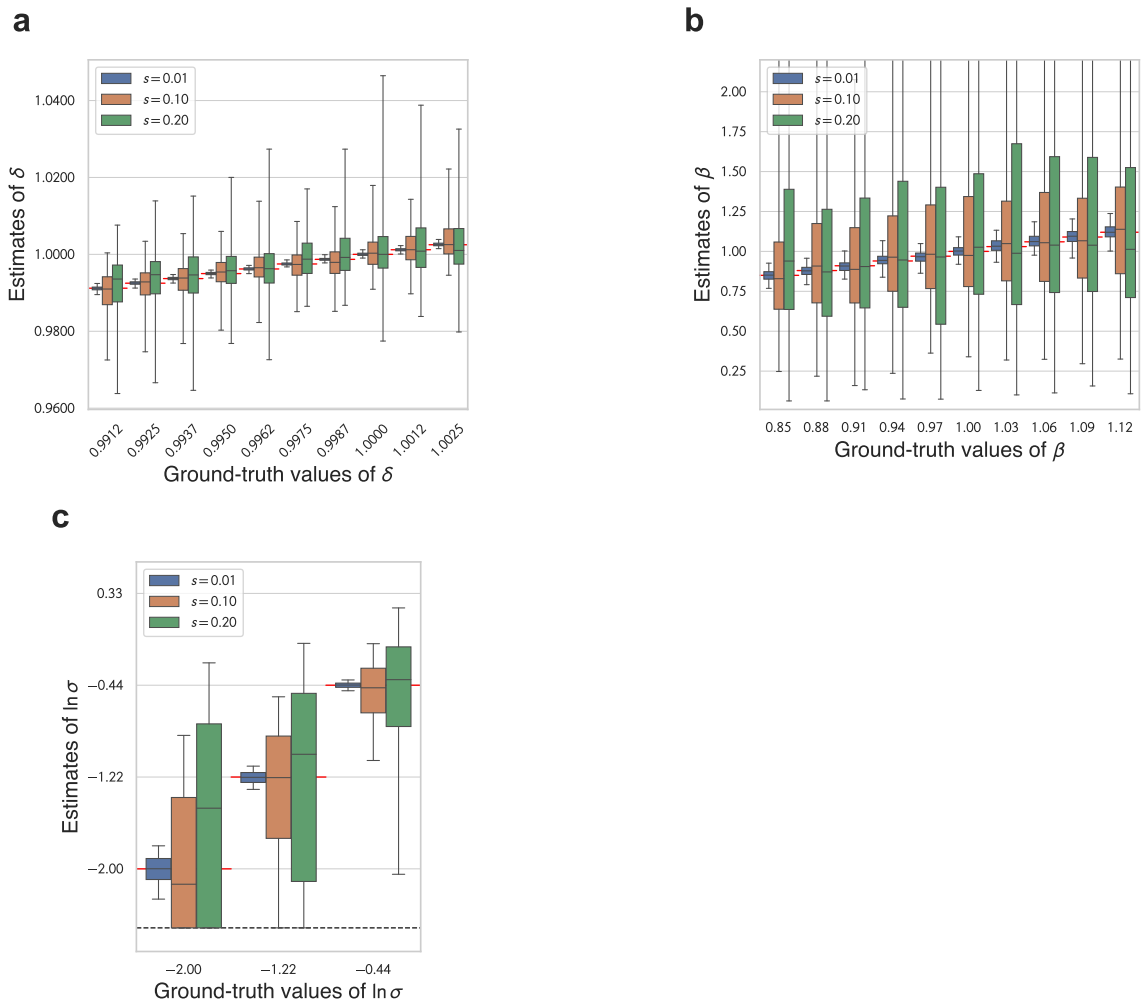

Figure S26: Box plots of **a**) δ, **b**) β, and **c**) $\ln\sigma$ estimates in the case where ground-truth $\ln\sigma$ is negative.

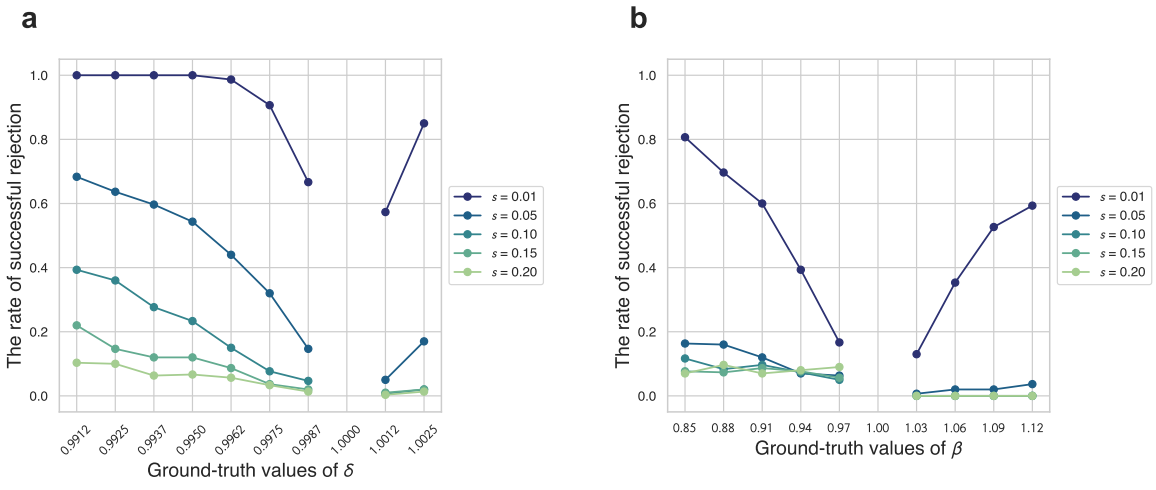

Figure S27: Rate of successful rejection of the null hypothesis that **a**) $\widehat{\delta} = 1$ and **b**) $\widehat{\beta} = 1$, respectively, in the case where $\ln\sigma$ is negative.

SA.8 Demand curve for individuals who has an extreme $\ln \sigma$

Fig. S28 shows demand curves for individuals with very large or small values of the curvature parameter $\ln \sigma$. Both individuals with $\ln \sigma = 6$ and 7 allocate all amount to the sooner period at lower prices than the switching point; otherwise, they allocate all to the later period. There is no difference in behaviour between the two. For $\ln \sigma = 5$, since they allocate an amount greater than 0 to the sooner period at a price $1 + r = 1.44$, which is higher than the switching point, there is a difference in the behaviour compared with $\ln \sigma = 6$. However, there is no difference in behaviour other than at a price $1 + r = 1.44$.

For $\ln \sigma = -2$, -3 and -4, the demand curves do not strictly coincide, so there is a difference in behaviour. However, the differences are slight.

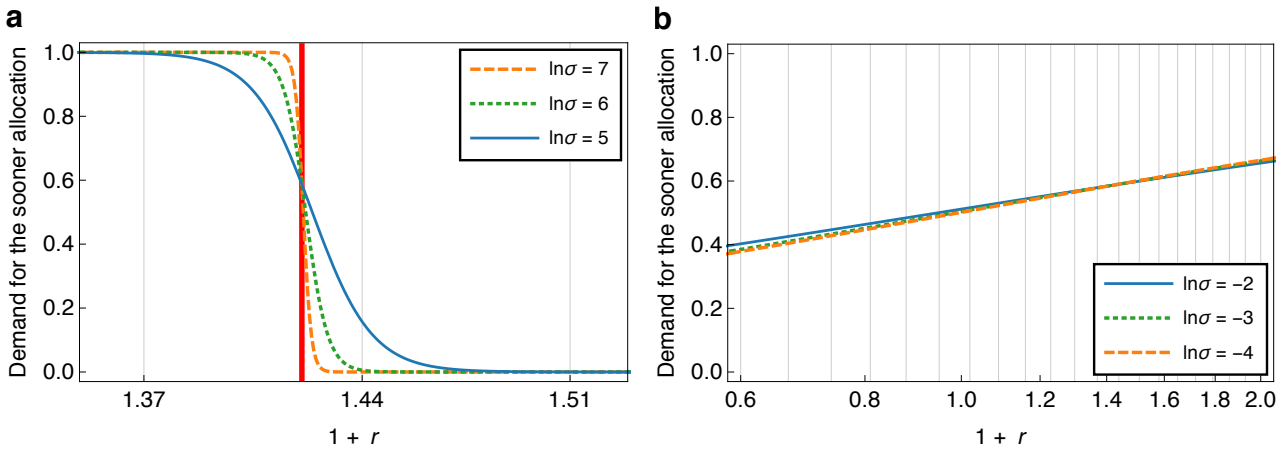

Figure S28: Demand curve for an individual whose curvature parameter is **a**) $\ln \sigma \geq 5$ and **b**) $\ln \sigma \leq -2$, where $\delta = 0.9950$ and $\beta = 1$. The horizontal axis representing price $1 + r$ is on a logarithmic scale. Note that the individual faces the decision problem of allocating between now ($t = 0$) and $k = 70$ days later with the prices indicated by the vertical lines in the figure. For Fig. **a**, only the neighbourhood of the switching point (indicated as a red line) is shown.