Constant Discounting, Temporal Instability and Dynamic Inconsistency in Denmark

A Longitudinal Field Experiment

Harrison, Glenn W.; Lau, Morten; Yoo, Hong Il

Document Version
Final published version

Published in:
International Economic Review

DOI:
10.1111/iere.12729

Publication date:
2025

INTERNATIONAL ECONOMIC REVIEW
Vol. 00, No. 0, August 2024 DOI: 10.1111/iere.12729

CONSTANT DISCOUNTING, TEMPORAL INSTABILITY, AND DYNAMIC INCONSISTENCY IN DENMARK: A LONGITUDINAL FIELD EXPERIMENT*

By Glenn W. Harrison, Morten I. Lau, and Hong Il Yoo

Georgia State University, U.S.A.; Copenhagen Business School, Denmark, and Durham University, U.K.; Loughborough University, U.K.

Claims that individuals have dynamically inconsistent preferences are usually made by studying individual discount rates over different time delays, but where those discount rates are elicited at a single point in time. However, to test dynamic inconsistency one has to know if the same subject has a different discounting function *at a later point in time*. We evaluate data from a longitudinal field experiment undertaken with a nationally representative sample of the adult Danish population. We cannot reject the hypothesis of constant discounting at the population level, but we reject the hypotheses of temporal stability and dynamic consistency.

1. INTRODUCTION

Dynamic inconsistency is often cited as a behavioral trait, which highlights the importance of considering alternative formulations of intertemporal choice behavior that do not rely on constant discount rates. A simple instance of dynamic inconsistency, also known as time inconsistency, arises when the decision maker prefers a larger, later (LL) payment to a smaller, sooner (SS) payment when both payments are delayed, but prefers the SS payment to the LL payment once enough time has passed to make the sooner payment immediately accessible. It is possible to attribute this type of preference reversal to genuine shifts in the decision maker's preferences between decision dates, but it is also possible to find theoretical explanations that do *not* require temporal instability in preferences. Prominent examples of the latter approach are models of present-biased preferences, in which discount rates are declining over longer time delays between payments.

Inferences on dynamic inconsistency are usually made by studying individual discount rates over different horizons, but where those discount rates are elicited at a single point in time.[1] Thus, individuals might be prompted to reveal their discount rate over a one-month period starting at some reference point in time, and then reveal their discount rate over a 24-month period starting from the same reference point.[2] If the elicited discount rates vary over different time horizons, the individual is typically claimed to have preferences that imply dynamic inconsistency, by holding and acting on preferences at one point in time that contradict the preferences and decisions of the same individual at a later date. However, evaluation of dynamic inconsistency requires a longitudinal design that elicits the same individual's choices

*Manuscript received May 2023; revised May 2024.

We are grateful to Iourii Manovskii (editorial board member) and three referees for valuable comments. We thank the Danish Social Science Research Council for research support under project 275-08-0289 and 12-130950, and the Carlsberg Foundation under Grant 2008-01-0410. Harrison and Lau are also affiliated with the School of Commerce, University of Cape Town. Please address correspondence to: Hong Il Yoo, Loughborough Business School, Loughborough University, LE11 3TU, Leicestershire, U.K. E-mail: *h.i.yoo@lboro.ac.uk*.

[1] See Frederick et al. (2002) and Cohen et al. (2020) for literature reviews on the measurement and elicitation of time preferences.

[2] An alternative approach in the literature is to vary the time delay to the sooner payment and keep the time period between the sooner and later payments constant, see Halevy (2015) for an example.

at two different points in time. Although cross-sectional discrepancy between the two sets of choices may support the hypothesis of nonconstant discounting, one cannot infer dynamic inconsistency from these choices without invoking the assumption of temporal stability in time preferences. The need to distinguish between the notions of dynamic consistency, constant discounting and temporal stability has been well-recognized in the literature on time preferences,[3] and Halevy (2015) provides a particularly clear statement of the theoretical interplay between the three.[4] Despite the importance of temporal stability for inferences about dynamic consistency, there is very little data available from longitudinal experiments or other primary sources to *directly* test these properties of intertemporal preferences, and none that reflects a nationally representative population.[5]

We evaluate and test the hypotheses of constant discounting, temporal stability, and dynamic consistency using data from a longitudinal field experiment with a *nationally representative sample of the adult Danish population*. The experiment is designed to elicit discount rates for assets of different time delays at different decision dates. Our econometric approach is structural in the sense that we directly estimate latent preference parameters that characterize a potentially nonconstant discounting function, and use the estimated coefficients to draw statistical inferences. Estimating structural models and undertaking formal hypothesis tests with respect to underlying latent primitives raises an important point about statistical tests of the underlying theory. It is possible that two of the three properties each *individually* hold in a statistical sense, but that they do not *jointly* hold statistically. We demonstrate an instance of this issue as a special case of our analysis.

We model time preferences using a quasi-hyperbolic (QH; also known as "β-δ") discounting function, which nests constant and nonconstant discounting depending on the value of the present bias parameter β. We evaluate constant discounting by estimating the present bias parameter from choices in the initial and repeat experiments, respectively, and we evaluate temporal stability by comparing the estimated discounting functions from the initial and repeat experiments. To evaluate dynamic consistency, we consider choices between two payments delivered at or after the repeat experiment. We then ask whether the discounting function estimated from the initial experiment predicts the same choice as the discounting function estimated from the repeat experiment. The degree of present bias in the discounting function from the initial experiment is therefore irrelevant to our evaluation of dynamic consistency, since neither payment option is immediate from the perspective of the initial experiment. However, the present bias parameter is directly relevant to our evaluation of temporal stability, which requires that the elicited discounting functions from the initial and repeat experiments display the same degree of present bias.

To control for utility curvature on inferred discount rates, we jointly estimate discounting and utility functions, following the approach by Andersen et al. (2008), and extend their modeling framework to our longitudinal data set by accounting for within-individual correlation in risk and time preferences over time. Our objective of evaluating discounting functions at two different decision dates provides a distinct reason to stress the importance of controlling for utility curvature. In a cross-sectional analysis, incorrectly assuming linear utility may bias the estimation of discounting functions since neglected utility curvature may be mistaken for the effects of discounting. In a longitudinal analysis, it is thus possible that temporal instability in the utility function is mistaken for temporal instability in the discounting function, unless one jointly estimates both functions at both decision dates. We control for temporal instability in utility curvature under Rank-Dependent Utility (RDU) Theory following Quiggin (1982), as well as utility curvature under Expected Utility Theory (EUT).

[3] For example, see Bommier (2006, p. 1236) and Meier and Sprenger (2015, p. 273).

[4] Halevy (2015) nicely clarifies the distinction and relation between three properties that characterize time preferences: stationarity, time invariance, and time consistency. These properties correspond to our definitions of constant discounting, temporal stability, and dynamic consistency, respectively.

[5] We review the existing literature in Subsection 5.1.

We recruited a nationally representative sample of Danish adults between 19 and 75 years of age, and draw inferences relevant to the broad population. Although the experimenter may invite a random representative sample of subjects from a field population of interest, whether those subjects show up to the experiment is their own decision. Selection bias arises when subjects with certain types of (latent) preferences are more likely to self-select into the observed sample. The longitudinal design also raises concerns about possible sample attrition bias, which arises when unobserved individual preferences are correlated with the decision to select into the repeat experiment.[6] To control for nonrandom selection and attrition effects that may bias our inferences with respect to the adult Danish population, we apply the direct likelihood approach of Harrison et al. (2020) and construct a statistical model that explicitly recognizes participants in the longitudinal experiment as self-selected members of the population in *both* waves.

Our structural model controls for endogenous sample selection and attrition bias by allowing the error terms in the selection and attrition equations to be correlated with unobserved heterogeneity in individual preferences. The model combines the random coefficient specification of population heterogeneity with joint estimation of risk and time preferences akin to Andersen et al. (2008), Heckman's (1979) celebrated correction for selection bias, and correction for attrition bias in the style of Capellari and Jenkins (2004). Estimation of selection and attrition processes requires data on those who accepted our invitations to the experiments as well as those who did not. We targeted a random sample of adult Danes obtained from the Danish Civil Registry and observe some basic sociodemographic characteristics on both groups of individuals, regardless of their decision to participate in the experiments. Our design builds in explicit randomization of incentives to participate by varying the recruitment fee in the initial experiment, which provides an exclusion restriction for the empirical identification of endogenous sample selection bias.

We generally find dynamic inconsistency in the Danish population between 2009 and 2010. This finding is remarkable since we find constant discounting in the *first* wave in 2009, which empirically illustrates the importance of distinguishing constant discounting from dynamic consistency. The estimated population means in the first wave for the baseline (annual) discount rate, δ, and the present bias parameter, β, are, respectively, equal to 0.109 and 1.002. We do *not* find comparable results on constant discounting in the *second* wave in 2010, and that is what is required, on one leg of the tripod of properties, to claim dynamic consistency. The estimated population means in the second wave for δ and β are equal to 0.075 and 0.989, respectively. We also reject the hypothesis of temporal stability of time preferences over the two waves. Hence, we infer dynamic inconsistency in the general population. At a methodological level, we also show that flexible specifications of risk preferences, and corrections for sample selection and attrition, matter for our inferences on constant discounting, temporal stability, and dynamic consistency.

In Section 2, we discuss our longitudinal field experiment with a focus on sampling procedures and identification of latent risk and time-preference parameters. In Section 3, we formalize hypotheses of constant discounting, temporal stability, and dynamic consistency from our structural model of time preferences. Section 4 reports our findings, based on estimation of the joint distribution of individual-specific preference parameters within and between waves of the experiment. The approach allows us to make inferences at the population level as well as the individual level. Section 5 connects our analysis to previous literature. We conclude in Section 6 and provide a numerical example that illustrates the implications of our finding on dynamic inconsistency.

[6] It is common to claim that there is no selection or attrition bias if the sample and population means of some key observed characteristics are similar. However, such difference-in-means comparisons do not address the econometric origin of selection or attrition bias, namely, *unobservable characteristics* that are correlated with selection decisions and preferences. Harrison et al. (2020) show that correction for nonrandom selection and attrition can have significant economic effects on inferred risk attitudes, even when difference-in-means tests may (mis)lead one to expect otherwise.

2. DATA

Our data originate from a longitudinal, artifactual field experiment conducted in Denmark. The first wave was designed to study alternative specifications of discount functions (Andersen et al., 2014) and intertemporal risk aversion (Andersen et al., 2018b). The second wave was designed to study dynamic consistency, temporal stability of risk and time preferences, and sample selection and attrition in risk and time preferences, which we address here.

2.1. *Sampling Procedures.* Between September 28 and October 22, 2009, we conducted an artifactual field experiment with 413 Danes. The sample was drawn to be representative of the adult population aged between 18 and 75 years as of January 1, 2009. We obtained a random sample of 50,000 Danes from the Danish Civil Registration Office, stratified the sample by geographic area, and sent out 1,996 invitations to a randomly selected subsample. The information from the Danish Civil Registration Office includes sex at birth, age, residential location, marital status, and whether the individual is an immigrant. Thus, we are in the fortunate, and rare, position of knowing some basic demographic characteristics of the individuals who do *not* select into our experiment. At a broad level, our final sample is representative of the population in terms of observable characteristics: the sample of 50,000 Danes had an average age of 49.8, 50.1% of them were married, and 50.7% were female; our final sample of 413 subjects had an average age of 48.7, 56.5% of them were married, and 48.2% were female.

The initial recruitment letter clearly explained that there would be fixed and stochastic earnings from participating in the experiment. The fixed amount is 500 kroner in one treatment and 300 kroner in another treatment, and subjects were randomly assigned to one of these two treatments.[7] The stochastic earnings in the recruitment letter refer to the risk aversion and discounting tasks. There were 40 risk-aversion tasks and 130 discounting tasks in the experiment, where the risk-aversion tasks preceded the discounting tasks in one treatment, and vice versa in another treatment. Between April 2010 and October 2010, we repeated the decision tasks with a sample of 182 subjects from the 413 subjects who participated in the first experiment.[8] Each subject was interviewed in private in the second experiment, and the meeting was conducted at a convenient location for them (e.g., their private residence or the hotel where the first experiment took place). All subjects were paid a fixed fee of 300 kroner for their participation in the second experiment.[9]

2.2. *Decision Tasks.* Individual discount rates are evaluated by asking subjects to make a series of choices over two outcomes that are paid at different dates. For example, the sooner option could be 3,000 kroner now, and the later option could be 3,300 kroner in one year. If the subject with a linear utility function chooses the sooner option, we can infer that the discount rate is above 10% for a one-year time delay. If the same subject picks the later option instead, we can infer that the annual discount rate is below 10%. By varying the amount of the later option, we can identify the discount rate of the individual, conditional on knowing the utility of those amounts to the individual. One can also vary the time delay between the

[7] The exchange rate at the time was close to 5 kroner per U.S. dollar.

[8] From the sample of 413 subjects in the first experiment, a random subsample of 354 subjects was invited to the second experiment. The sample response rate in the first wave was 24.1% with the high recruitment fee, and 18.1% with the low recruitment fee. This reduction in sample response rates in the first wave is statistically significant according to a Fisher Exact test, with a p-value less than 0.001. After participating in the first wave, the sample response rate in the second wave was slightly lower for those recruited into the first wave with the high recruitment fee compared to those recruited with the low fee. The sample response rates were 48.4% and 54.7% in the second wave, respectively, and are not statistically different according to a Fisher Exact test with a two-sided p-value of 0.24. The sampling procedures are documented in Harrison et al. (2020).

[9] We did not vary the recruitment fee in the second experiment because we offered to interview the subjects at home or the hotel where the first experiment was conducted. The experiments were very time-consuming and expensive to conduct, and we paid subjects the low recruitment fee of 300 kroner in the second experiments to keep costs down, although we see the value in varying recruitment fees in the second stage as well.

sooner and later options to identify the discounting function, and of course one can vary the delay to the sooner option. This method has been widely employed (e.g., Coller and Williams, 1999; Harrison et al., 2002; Eckel et al., 2005; Andersen et al., 2008; and Dohmen et al., 2010).

We consider time delays between the sooner and later options from two weeks to one year. Each subject was presented with choices over four different time delays in ascending or descending order, and those time delays were drawn at random from a set of 13 intervals (two weeks, and 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 months).[10] We also varied the delay to the sooner option on a between-subjects basis: roughly half of the sample had decision tasks with no delay to the sooner option, and the other half had a delay of 30 days. Similarly, we varied the provision of implied annual interest rates for each choice on a between-subjects basis. Finally, we employed two different principals on a between-subjects basis (1,500 and 3,000 kroner) to assess the significance of magnitude effects on elicited discount rates. The four sets of treatments, the order of time delay, the delay to the sooner option, information on implied interest rates, and the level of the principal, give a $2 \times 2 \times 2 \times 2$ design. Each subject was randomly assigned to one of these 16 combinations in each wave of the experiment.

The subjects were presented with 40 binary choices, in four sets of 10 with the same time delay between the sooner and later option. The annual interest rate varied between 5% and 50%, in increments of 5%, on the principal of 1,500 or 3,000 kroner. We randomly selected one decision task for each subject using numbered dice and the subjects were paid their preferred sooner or later option in that task.[11]

Utility functions are evaluated from data in which subjects made a series of choices over two risky lotteries. For example, lottery A might give the individual a 10–90 chance of receiving 2,000 or 1,600 kroner to be paid today, and lottery B might have a 10–90 chance of receiving 3,850 or 100 kroner today. We gave subjects 40 choices, in four sets of 10 with the same prize combinations. The prize sets employed were: [A1: 2,000 and 1,600; B1: 3,850 and 100], [A2: 1,125 and 750; B2: 2,000 and 250], [A3: 1,000 and 875; B3: 2,000 and 75], and [A4: 2,250 and 1,000; B4: 4,500 and 50]. The order of the four prize sets was randomized for each subject, with probabilities of high prizes varying in ascending order between 0.1 and 1 in increments of 0.1 within each set. We asked each subject to select their preferred option in each of the 40 decision tasks and then randomly selected one task for payment using numbered dice.[12]

[10] We oversampled the first three time delays, since short time delays are important with respect to identification of alternative discounting functions. The shorter time delays were each chosen with probability 2/16 = 0.125, compared to the 1/16 = 0.0625 probability for each of the others. The subjects were randomly assigned to four different time delays in each wave of the experiment.

[11] The following language explained the payment procedures: "You will receive the money on the date stated in your preferred option. If you receive some money today, then it is paid out at the end of the experiment. If you receive some money to be paid in the future, then it is transferred to your personal bank account on the specified date. In that case you will receive a written confirmation from Copenhagen Business School which guarantees that the money is reserved on an account at Danske Bank. You can send this document to Danske Bank in a prepaid envelope, and the bank will transfer the money to your account on the specified date." Payments by way of bank transfer are common in Denmark, Copenhagen Business School is a well-known institution in Denmark, and Danske Bank is the largest financial enterprise in the country as measured by total assets. All payments to subjects were made by Danske Bank on the specified dates of the subjects' preferred options, and we treat all future payments as credible and certain in the intertemporal decision model. Halevy (2008) and Chakraborty et al. (2020), on the other hand, treat present payments as certain and future payments as uncertain in their evaluation of the interplay between risk and time preferences. They introduce a hazard rate on future payments, which means that the correction for probability weighting under RDU enters the additive intertemporal utility function since future payments are uncertain.

[12] The experimental procedures were standard: Appendix A in Andersen et al. (2014) documents an English translation of the instructions, and shows typical screen displays for the discounting and risk-aversion tasks, as well as a list of parameter values for all choices. After all choices had been made the subject was asked a series of standard sociodemographic questions. Subsection 5.2 discusses alternative procedures for eliciting time preferences. The large incentives and budget constraints precluded us from paying all subjects, so each subject was given a 10% chance of being paid for one choice in each set of 40 choices. The average payment was 242 kroner for the risk attitude choices and 201 kroner for the discounting choices, for a combined average of 443 kroner, or $91, in addition to the fixed fee of $60 or $100.

3. HYPOTHESES

Each discounting task presents a choice between option A that pays Y_t in period t and option B that pays $Y_{t+\tau}$ in period $t + \tau$, where $\tau > 0$. Given an atemporal utility function $U(m)$ and a discounting function $D(t)$, the discounted utility of each option is specified as

$$PV_A = D(t) \times U(Y_t + \omega) + D(t + \tau) \times U(\omega), \tag{1}$$

$$PV_B = D(t) \times U(\omega) + D(t + \tau) \times U(Y_{t+\tau} + \omega), \tag{2}$$

where ω is a measure of background consumption, which is set exogenously in our model. We follow Andersen et al. (2014) and set it to the average daily consumption of private non-durable goods per capita, which was 130 kroner at the time of our experiments in 2009. Andersen et al. (2008) show that the addition of background consumption $\omega > 0$ is a sufficient condition to avoid negative discount rates under exponential discounting, popularizing its inclusion in structural estimation over the traditional assumption of $\omega = 0$.[13] They also report that when the estimation procedure accounts for nonlinear curvature of $U(m)$ over levels of monetary payment comparable to our design, the structural estimates of discount rates are robust to assumptions concerning the specific value of background consumption: for example, in their exponential discounting model, the inferred discount rate only changes from 10.2% to 9.8% as ω is varied from 50 to 200 kroner.

Our experiment includes discounting tasks with variation in the delay to the sooner option. The resulting basket of choices between an immediate payment and a future payment, and between two future payments, allows us to identify and estimate a QH discounting function

$$\begin{aligned} D(t) &= 1 && \text{if } t = 0 \\ &= \beta \times 1/(1+\delta)^t && \text{if } t > 0, \end{aligned} \tag{3}$$

where β is a present bias parameter and δ is a *baseline* discount rate for someone with no present bias ($\beta = 1$).[14] More generally, δ can be interpreted as a long-run discount rate, regardless of the value of β.[15] We denote time delay t in years (e.g., $t = 0.5$ for a six-month horizon), and specify δ on an annualized basis. We assume a priori that decision makers display long-run delay aversion ($\delta > 0$), and let β take values on either side of 1 to allow for present bias ($\beta < 1$) as well as future bias ($\beta > 1$). We assume an additive intertemporal utility function, following convention.

Consider an individual who at time t_1 makes a choice between a sooner reward delivered at time $t_1 + t$ and a later reward delivered at time $t_1 + t + \tau$, where $t \geq 0$ is the delay to the sooner option, and $\tau > 0$ is the time delay between the sooner and later payments. To facilitate the discussion, let t_1 be the date of the initial experiment and t_2 be the date of the second experiment. Assume for the moment that the sooner and later payments are constant nominal

[13] Andreoni and Sprenger (2012) use convex budget sets to elicit discounting and utility functions, and treat background consumption as an endogenous parameter that is estimated along with the $\{\beta, \delta\}$ parameters in the QH discounting function and a relative risk-aversion parameter α under EUT. They do not place any sign restrictions on background consumption and report both positive and negative values for this parameter. To allow negative background consumption, one would need to restrict the relative risk-aversion parameter $\alpha > 0$ in their power specification, which is equivalent to $r < 1$ in our utility function. This inequality constraint may be innocuous if one is interested in estimating the population mean of the r parameter, which happens to be smaller than 1. However, when estimating the population distribution of risk attitudes, as long as there are *some* individuals whose r parameters exceed 1, the constraint is not innocuous, regardless of whether the mean is smaller or greater than 1.

[14] The QH specification was introduced by Phelps and Pollak (1968) for a social planning problem, and applied to model individual behavior by Elster (1979, p. 71) and then Laibson (1997).

[15] The discount rate under the QH specification is $d^*(t)$ that solves $1/(1 + d^*)^t = \beta \times 1/(1 + \delta)^t$. The solution is given by $d^*(t) = (1 + \delta) / \beta^{(1/t)} - 1$. For any $\beta > 0$, $d^*(t)$ converges to δ as the horizon t increases, and in this sense δ can be interpreted as a long-run discount rate.

values regardless of the payment date and time delay. Denote by $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\}$ a binary indicator of whether the later payment is preferred at time t_1. The indicator $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_2}\}$ is similarly defined with respect to time t_2.

The QH discounting function $D(t)$ in Equation (3) takes values of β_1 and δ_1 in the first wave (t_1) and β_2 and δ_2 in the second wave (t_2) of the longitudinal experiment. We subscript $D(t)$ as $D_w(t)$ to emphasize that it is based on β_w and δ_w, where $w \in \{1, 2\}$. Define the relative discounting factor, $R_w(t, t + \tau) \equiv D_w(t + \tau)/D_w(t)$. The individual is more likely to choose the later payment when the present value favors the later instead of the sooner payment. The evaluation of sooner and later payments in the structural model is invariant to using the discount factors of $D_w(t)$ over time delay t and $D_w(t + \tau)$ over time delay $t + \tau$ or alternative discount factors of 1 over horizon t and $R_w(t, t + \tau)$ over horizon $t + \tau$. It follows that when one considers two different pairs of numerator and denominator discount factors, both pairs favor the same payment as long as the implied relative discount factors are identical. We use $R_w(t_1 + t, t_1 + t + \tau|\mathbf{t_1})$ to denote the relative discount factor associated with the choice $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\}$, and vary the time indices to address other cases.

In terms of binary preference relations, *constant discounting* at the time of the initial experiment refers to the property that choices made at time t_1 depend only on the time delay between the two payments, so $y\{t_1, t_1 + \tau|\mathbf{t_1}\} = y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\}$ where $t > 0$. In our structural model, this property can be stated as a single hypothesis H_0: $\beta_1 = 1$. With a delay to the sooner option $t > 0$, choice $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\}$ is between two future payments and the relative discounting factor is given by $R_1(t_1 + t, t_1 + t + \tau|\mathbf{t_1}) = 1/(1 + \delta_1)^\tau$. With no delay to the sooner option, choice $y\{t_1, t_1 + \tau|\mathbf{t_1}\}$ is between an immediate reward and a future reward such that $R_1(t_1, t_1 + \tau|\mathbf{t_1}) = \beta_1/(1 + \delta_1)^\tau$. The two relative discounting factors are identical and favor the same payment when $\beta_1 = 1$, which is when the discounting function displays neither present nor future bias.

Temporal stability refers to the property that choices do not vary from one point in time to another, as long as each payment date is adjusted to maintain the same time delay relative to the decision date, that is, $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\} = y\{t_2 + t, t_2 + t + \tau|\mathbf{t_2}\}$. This can be structurally stated as a joint hypothesis H_0: $\beta_1 = \beta_2$ and $\delta_1 = \delta_2$. To focus on the pure effects of discounting, we hold constant (or "partial out," to use a linear regression analogy) the effects of the utility function that may vary between the evaluation points, as the joint estimation of discounting and utility functions allows us to do. Consider first discounting tasks where $t > 0$. Choice $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\}$ in the initial experiment and choice $y\{t_2 + t, t_2 + t + \tau|\mathbf{t_2}\}$ in the repeat experiment give relative discounting factors of $R_1(t_1 + t, t_1 + t + \tau|\mathbf{t_1}) = 1/(1 + \delta_1)^\tau$ and $R_2(t_2 + t, t_2 + t + \tau|\mathbf{t_2}) = 1/(1 + \delta_2)^\tau$, respectively. Both relative discounting factors favor the same reward when $\delta_1 = \delta_2$. Without any delay to the sooner payment, the relative discounting factor is $R_1(t_1, t_1 + \tau|\mathbf{t_1}) = \beta_1/(1 + \delta_1)^\tau$ for the initial choice $y\{t_1, t_1 + \tau|\mathbf{t_1}\}$, and $R_2(t_2, t_2 + \tau|\mathbf{t_2}) = \beta_2/(1 + \delta_2)^\tau$ for the repeat choice $y\{t_2, t_2 + \tau|\mathbf{t_2}\}$. On its own, equation $R_1(t_1, t_1 + \tau|\mathbf{t_1}) = R_2(t_2, t_2 + \tau|\mathbf{t_2})$ has two unknown parameters on each side, and may hold for some configuration of $\{\beta_1, \delta_1\}$ and $\{\beta_2, \delta_2\}$ despite $\beta_1 \neq \beta_2$ and $\delta_1 \neq \delta_2$. But given the condition $\delta_1 = \delta_2$ for tasks without a front-end delay, the two relative discounting factors favor the same reward when $\beta_1 = \beta_2$.

Finally, *dynamic consistency* refers to the property that choices do not vary from one point to another when each payment date is fixed, that is, $y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1}\} = y\{t_1 + t, t_1 + t + \tau|\mathbf{t_1 + t}\}$. This can be structurally stated as a joint hypothesis H_0: $\beta_2 = 1$ and $\delta_1 = \delta_2$. While this property also pertains to comparisons of choices made at two points in time, each payment date is now fixed in time. Consider first decisions with a delay to the sooner payment $t > 0$ in the repeat experiment. Since the choice is between two future payments from the perspective of the initial and repeat experiment, one may expect the present bias parameters β_1 and β_2 to be irrelevant. Indeed, the relative discounting factors, $R_1(t_2 + t, t_2 + t + \tau|\mathbf{t_1}) = 1/(1 + \delta_1)^\tau$ for choice $y\{t_2 + t, t_2 + t + \tau|\mathbf{t_1}\}$ in the initial experiment and $R_2(t_2 + t, t_2 + t + \tau|\mathbf{t_2}) = 1/(1 + \delta_2)^\tau$ for choice $y\{t_2 + t, t_2 + t + \tau|\mathbf{t_2}\}$ in the repeat experiment, favor the same payment when $\delta_1 = \delta_2$. Now suppose that the delay to the sooner payment in the repeat experiment is removed.

This makes the sooner payment immediate from the perspective of the repeat experiment, but it remains delayed from that of the initial experiment. As one may expect, β_1 remains irrelevant but β_2 becomes relevant to relative discounting factors in the repeat experiment: $R_1(t_2, t_2 + \tau|\mathbf{t_1}) = 1/(1 + \delta_1)^\tau$ for the initial choice $y\{t_2, t_2 + \tau|\mathbf{t_1}\}$, and $R_2(t_2, t_2 + \tau|\mathbf{t_2}) = \beta_2/(1 + \delta_2)^\tau$ for the repeat choice $y\{t_2, t_2 + \tau|\mathbf{t_2}\}$. On its own, $R_1(t_2, t_2 + \tau|\mathbf{t_1}) = R_2(t_2, t_2 + \tau|\mathbf{t_2})$ may be satisfied by some configuration of $\beta_2 \neq 1$ and $\delta_1 \neq \delta_2$. But given the condition $\delta_1 = \delta_2$ in the case with a positive delay to the sooner payment in the repeat experiment, the two relative discounting factors favor the same reward when $\beta_2 = 1$.

Our structural approach fully accommodates the interplay of constant discounting, temporal stability, and dynamic consistency: Each of our three hypotheses may hold without the other two, and a combination of any two hypotheses implies the third one. Applying the logic behind our approach to a wider range of nonconstant discounting functions shows that the QH discounting function is a rare, if not unique, functional form that allows one to test each hypothesis of constant discounting, temporal stability, and dynamic consistency separately. For example, a hyperbolic discounting function $D_w(t) = 1/(1 + \kappa_w t)$ allows one to test for temporal stability in the κ_w parameter and nothing else. Given any nonzero value of κ_w, the relative discount factor, $R_w(t, t + \tau) = [1 + \kappa_w(t + \tau)]/(1 + \kappa_w t)$, implies nonconstant discounting and dynamic inconsistency. A Weibull discounting function $D_w(t) = \exp\{-\delta_w t^{\zeta w}\}$ leads to a relative discount factor of $R_w(t, t + \tau) = \exp\{-\delta_w[(t + \tau)^{\zeta w} - t^{\zeta w}]\}$, which does not allow one to test for constant discounting and dynamic consistency separately. When applied to a choice between payments dated t_2 and $t_2 + \tau$, the relative discount factor becomes $R_1(t_2, t_2 + \tau|\mathbf{t_1}) = \exp\{-\delta_1[(t_2 - t_1 + \tau)^{\zeta 1} - (t_2 - t_1)^{\zeta 1}]\}$ for the initial choice at t_1, and $R_2(t_2, t_2 + \tau|\mathbf{t_2}) = \exp\{-\delta_2[\tau^{\zeta 2}]\}$ for the repeat choice at t_2. Thus, in the Weibull case, constant discounting at the time of the initial experiment, $\zeta_1 = 1$, is a necessary condition for dynamic consistency, $R_1(t_2, t_2 + \tau|\mathbf{t_1}) = R_2(t_2, t_2 + \tau|\mathbf{t_2})$.

Although the three hypotheses of constant discounting, temporal stability, and dynamic consistency pertain to discounting functions, explicit controls for utility curvature in each wave of the experiment are important to avoid biased evaluations of those hypotheses. We use the risk-aversion tasks to identify the utility function $U(m)$ and assume that it has the functional form

$$U(m) = m^{(1-r)}/(1-r), \tag{4}$$

where the r parameter is an index of utility curvature, with concave utility when $r > 0$ and convex utility when $r < 0$. Under EUT, the r parameter is interpreted as the coefficient of relative risk aversion.

Our identification strategy employs the more general RDU model that attributes risk preferences to the effects of rank-dependent probability weighting as well as utility curvature. We jointly estimate the utility function in (4) with the probability weighting function (PWF)

$$w(P) = \exp\{-(-\ln P)^{\varphi}\}, \tag{5}$$

where the φ parameter determines the shape of the PWF, which follows an inverse-S shape over probabilities if $\varphi < 1$ and an S shape if $\varphi > 1$. An inverse-S shape overweights probabilities ($w(P) > P$) when P is relatively small, and underweights probabilities ($w(P) < P$) when P is relatively large.[16] The order of overweighting and underweighting is reversed for an S-shaped function. EUT is a special case of RDU which assumes $\varphi = 1$, hence $w(P) = P$ everywhere. As with the discounting parameters, we allow the values of the risk-preference parameters r and φ to vary between waves.

[16] Given our functional form assumption, the small and large probabilities are defined relative to $P = 0.368$, which is the fixed point where $w(P) = P$ for any value of φ.

4. RESULTS

We are interested in evaluating and testing hypotheses with respect to three properties of individual time preferences: constant discounting, temporal stability, and dynamic consistency. We use maximum simulated likelihood to estimate the full statistical model that captures behavioral noise in decision making,[17] unobserved preference heterogeneity, endogenous selection into the first wave of the longitudinal field experiment, and endogenous panel attrition between the two waves of the same experiment. The statistical model is documented in Appendix A.

Our model accommodates unobserved preference heterogeneity by specifying the time and risk-preference parameters for each wave $w \in \{1, 2\}$ as random parameters $\{\beta_{nw}, \delta_{nw}, r_{nw}, \varphi_{nw}\}$, where n indexes different individuals in the population. Then, we estimate population means, medians, and standard deviations of the individual-specific preference parameters as well as within-wave and between-wave correlations in those parameters. This approach of estimating the joint distribution of the preference parameters allows us to evaluate discounting functions at both the population and the individual levels in a coherent manner. At the population level, we test whether discounting functions evaluated at the population means of the preference parameters display each property of interest. At the individual level, we derive the population share of individuals who have discounting functions that display each property of interest, based on the estimated joint distribution function. We also use the estimated correlation matrix to study within-individual correlation in the same parameter between the two waves, and within-individual correlation in different parameters at the same wave.

The statistical procedure addresses the panel dimension of the data set at both the modeling and inferential stages. At the modeling stage, our random parameter specification induces panel correlation across repeated observations on the same individual, in an analogous fashion to mixed logit models for repeated choice data. At the inferential stage, we adjust all standard errors and test statistics for clustering at the individual level.

We use Johnson's (1949) "S_B" distribution to capture the population distribution of each time-preference parameter. An S_B distribution is a logit transformation of a normal distribution. This transformation produces a flexible parametric distribution that is capable of approximating a wide range of shapes in population distributions (e.g., uniformity, unimodality, bimodality, and left and right skewness) without requiring us to impose any shape restriction a priori. The primary trade-off is that we cannot obtain analytic solutions for the population means, standard deviations, and correlation coefficients describing the S_B-distributed time-preference parameters.[18] We simulate these population moments using 10,000 draws from the estimated joint distribution of the underlying normal parameters, and compute associated standard errors and confidence intervals using the bootstrapping procedure of Krinsky and Robb (1986). All moments that we report for the time-preference parameters, β_{nw} and δ_{nw}, refer to their S_B distributions instead of the underlying normal distributions.

4.1. *Constant Discounting, Temporal Stability, and Dynamic Consistency.*

The upper-left panel of Figure 1 displays the estimated population distribution of the baseline discount rate δ_{nw} for each wave w, based on the empirical model that corrects for endogenous selection and attrition bias. The estimated coefficients in the model are reported in Table 1. The estimated distribution for wave 1 is unimodal and right-skewed, neither of which has been imposed a priori. There is an evident degree of preference heterogeneity, although most individuals seem to have baseline discount rates in the interval between 0% and 10% per annum. This informal description is comparable with the estimated mean, median, and standard deviation, which

[17] Our stochastic specification adopts the contextual utility model of Wilcox (2011) for risk-aversion tasks and a Fechner error specification for discounting tasks. We discuss alternatives in the literature in Subsection 5.2.

[18] We choose more restrictive but analytically tractable distributions for the risk-preference parameters r_{nw} and φ_{nw}, whose roles in our analyses are akin to control variables in everyday regression contexts. Specifically, we use normal and lognormal distributions to capture the population distributions of r_{nw} and φ_{nw}, respectively.

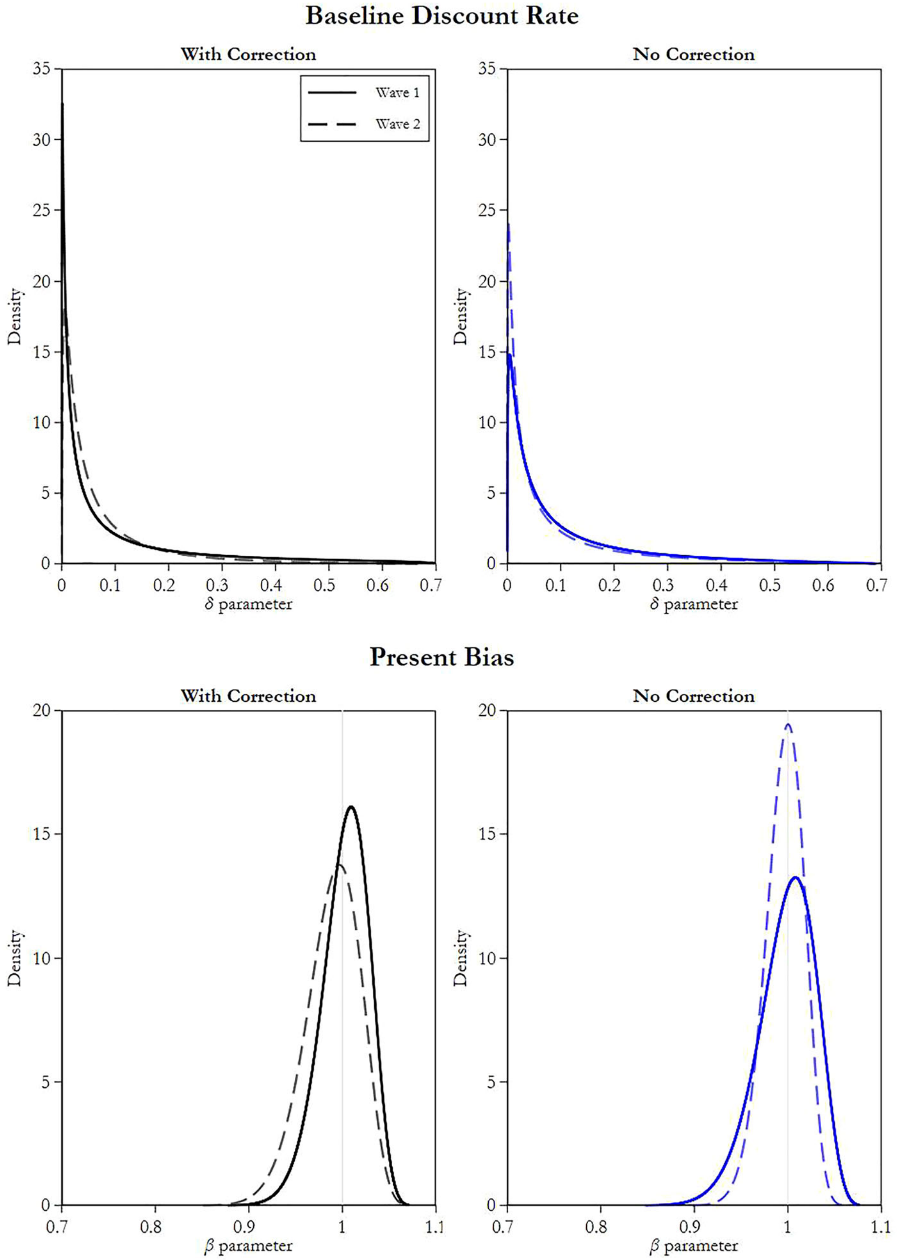

FIGURE 1

POPULATION DISTRIBUTIONS OF DISCOUNTING PARAMETERS

are 10.9%, 4.4%, and 14.5%, respectively. The estimated population distribution for wave 2 takes a similar shape with slightly higher density in the interval between 0% and 10%, and the estimated mean, median, and standard deviation are 7.5%, 4.1%, and 9.2%, respectively. All the estimated means, medians, and standard deviations are significantly greater than 0, with p-values < 0.001. Henceforth, unless we say otherwise, all point estimates in our discussion have p-values less than 0.001.

The lower-left panel of Figure 1 displays the estimated population distribution of the present bias parameter β_{nw} for each wave, based on the same empirical model that corrects for endogenous selection and attrition bias. The estimated distribution for wave 1 is unimodal

TABLE 1

ESTIMATES OF STRUCTURAL PARAMETERS WITH FULL CONTROLS FOR SAMPLE SELECTION AND ATTRITION

Variable	Estimate	Standard Error	p-Value	95% Confidence Interval	
		Selection equation: $\beta_1/\sqrt{\text{var}(u_{n1})}$			
Female	−0.046	0.063	0.464	−0.171	0.078
Young	0.187	0.117	0.112	−0.044	0.417
Middle	0.326	0.111	0.003	0.108	0.544
Old	0.395	0.102	<0.001	0.195	0.595
High_fee	0.189	0.066	0.004	0.060	0.318
Dist	−0.033	0.006	<0.001	−0.044	−0.021
Dist^2	0.001	<0.001	<0.001	<0.001	0.001
Constant	−0.827	0.104	<0.001	−1.032	−0.623
		Attrition equation: $\beta_2/\sqrt{\text{Var}(u_{n2})}$			
Female	−0.074	0.125	0.552	−0.319	0.171
Young	−0.308	0.236	0.191	−0.770	0.154
Middle	−0.269	0.217	0.216	−0.694	0.156
Old	−0.221	0.204	0.278	−0.621	0.178
IncLow	−0.192	0.170	0.258	−0.525	0.141
IncHigh	−0.122	0.141	0.387	−0.399	0.155
Earnings	0.054	0.037	0.147	−0.019	0.126
Constant	0.678	0.230	0.003	0.228	1.129
		Means of δ and β parameters in wave 1 and wave 2			
δ_1	0.109	0.007	<0.001	0.096	0.124
δ_2	0.075	0.009	<0.001	0.062	0.097
β_1	1.002	0.002	<0.001	0.999	1.006
β_2	0.989	0.003	<0.001	0.983	0.995
		Medians of δ and β parameters in wave 1 and wave 2			
δ_1	0.044	0.008	<0.001	0.029	0.059
δ_2	0.041	0.006	<0.001	0.029	0.053
β_1	1.005	0.002	<0.001	1.001	1.008
β_2	0.992	0.002	<0.001	0.987	0.996
		Standard deviations and correlation coefficients of δ and β parameters in wave 1 and wave 2			
$\sigma_{\delta 1}$	0.145	0.007	<0.001	0.131	0.159
$\sigma_{\delta 2}$	0.092	0.009	<0.001	0.078	0.113
$\sigma_{\beta 1}$	0.025	0.004	<0.001	0.018	0.033
$\sigma_{\beta 2}$	0.030	0.004	<0.001	0.022	0.039
$\rho_{\delta 1\ \delta 2}$	0.354	0.059	<0.001	0.246	0.474
$\rho_{\beta 1\ \beta 2}$	0.273	0.131	0.036	−0.006	0.502
		Means of r and φ parameters in wave 1 and wave 2			
r_1	0.951	0.066	<0.001	0.821	1.081
r_2	1.076	0.089	<0.001	0.902	1.250
φ_1	2.171	0.195	<0.001	1.787	2.554
φ_2	2.091	0.236	<0.001	1.627	2.554
		Standard deviations and correlation coefficients of r and φ parameters in wave 1 and wave 2			
σ_{r1}	0.725	0.056	<0.001	0.616	0.835
σ_{r2}	0.597	0.067	<0.001	0.466	0.728
$\sigma_{\varphi 1}$	3.579	0.870	<0.001	1.875	5.283
$\sigma_{\varphi 2}$	2.864	0.876	0.001	1.147	4.581
$\rho_{r1\ r2}$	0.668	0.078	<0.001	0.515	0.821
$\rho_{\varphi 1\ \varphi 2}$	0.871	0.063	<0.001	0.747	0.995

NOTE: Table B1 in Online Appendix B reports other correlation coefficients.

and left-skewed, and the interval between 0.95 and 1.05 captures most decision makers in the population. While the estimated mean and median is 1.002 and 1.005, respectively, the population distribution displays a statistically significant standard deviation of 0.025. The estimated distribution for wave 2 is more skewed to the left, with an estimated mean, median, and standard deviation of 0.989, 0.992, and 0.030, respectively.[19]

[19] Formally, the marginal distribution of the baseline discount rate parameter, $\delta_{nw} = 0.7/[1 + \exp(-\delta^*{}_{nw})]$, is characterized by the mean and standard deviation of the underlying normal parameter $\delta^*{}_{nw}$. Similarly, the marginal distri-

We now move to more formal statistical tests of properties describing time preferences at the population level, and consider first *the hypothesis of constant discounting*, H_0: $E(\beta_{nw}) = 1$. The estimated population mean of the present bias parameter for wave 1 is equal to 1.002, and we cannot reject H_0: $E(\beta_{n1}) = 1$, despite the small standard error of 0.002 (p-value = 0.317). The estimated population mean for wave 2 is equal to 0.989 and close to unity, but we nevertheless reject H_0: $E(\beta_{n2}) = 1$, with a p-value < 0.001. Hence, we cannot reject the null hypothesis that discount rates are constant in wave 1, but we formally reject the hypothesis in wave 2, although the estimated population mean of the present bias parameter is very close to 1.

Consider next *the hypothesis of temporal stability*, H_0: $E(\beta_{n1}) = E(\beta_{n2})$ and $E(\delta_{n1}) = E(\delta_{n2})$. The estimated difference in the population means of the present bias parameter, $E(\beta_{n2}) - E(\beta_{n1})$, is equal to –0.013, with a 95% confidence interval of [–0.019, –0.007], and the estimated difference in the population means of the baseline discount rate, $E(\delta_{n2}) - E(\delta_{n1})$, is equal to –0.033, with a 95% confidence interval of [–0.048, –0.012]. The confidence intervals are consistent with relatively small standard errors for the estimated differences in populations means of the two discounting parameters, and we reject both H_0: $E(\beta_{n1}) = E(\beta_{n2})$ and H_0: $E(\delta_{n1}) = E(\delta_{n2})$ with a p-value < 0.001. We also reject the joint null hypothesis of temporal stability (p-value < 0.001).[20]

Finally, consider *the hypothesis of dynamic consistency*, H_0: $E(\beta_{n2}) = 1$ and $E(\delta_{n1}) = E(\delta_{n2})$. This null hypothesis entails constant discounting in wave 2 and temporal stability in the baseline discount rate across the two waves. The joint test of the two constraints does not entail constant discounting in wave 1, since the relevant payment options are delayed from the perspective of wave 1, which implies that β_{n1} does not influence the ordering of discounted utilities. Although the estimated coefficients roughly satisfy $E(\beta_{n2}) = 1$ and $E(\delta_{n1}) = E(\delta_{n2})$, we reject each null hypothesis, as well as the joint hypothesis of dynamic consistency, with a p-value < 0.001.[21]

Our econometric approach also allows us to undertake a more individualistic evaluation of constant discounting, temporal stability, and dynamic consistency, by considering the population shares of decision makers whose discounting functions agree with each hypothesis. Since we have estimated the population distributions of all preference parameters, including all relevant correlation coefficients, we can derive population shares for each type of discounting behavior by computing the following marginal and joint probabilities: $\Pr(|\beta_{nw} - 1| < \varepsilon_\beta)$ for constant discounting at wave w; $\Pr(|\beta_{n2} - \beta_{n1}| < \varepsilon_\beta$ and $|\delta_{n2} - \delta_{n1}| < \varepsilon_\delta)$ for temporal stability between the two waves; and $\Pr(|\beta_{n2} - 1| < \varepsilon_\beta$ and $|\delta_{n2} - \delta_{n1}| < \varepsilon_\delta)$ for dynamic consistency. For hypotheses involving present bias parameters, the absolute tolerance $\varepsilon_\beta > 0$ defines small deviations from rigid predictions that we will allow for when classifying discounting behavior into different types.[22] For hypotheses involving baseline discount rates, the tolerance is given

bution of the present bias parameter, $\beta_{nw} = 0.7 + 0.4/[1 + \exp(-\beta^*_{nw})]$, is characterized by the mean and standard deviation of the underlying normal parameter β^*_{nw}. Using $\chi^2(2)$-distributed Wald tests, we reject the hypothesis that the two marginal distributions in wave 1 and 2 are identical for each of parameters δ_{nw} and β_{nw}, with p-values less than 0.001.

[20] Two correlation coefficients are directly relevant to temporal stability in time preferences: corr(δ_{n1}, δ_{n2}) and corr(β_{n1}, β_{n2}) capture within-individual correlation over time in the baseline discount rate and present bias parameter, respectively. Each correlation coefficient may be used as a measure of individual-level temporal stability in the respective parameter. With corrections for endogenous selection and attrition bias, the estimated value of corr(δ_{n1}, δ_{n2}) is 0.354 with a p-value < 0.001, and the estimated value of corr(β_{n1}, β_{n2}) is 0.273 with a p-value = 0.018. Thus temporal stability holds for each parameter in the sense that if some individuals display greater long-run delay aversion or present bias than the average person in wave 1, they also tend to do so in wave 2. However, this directional prediction is not deterministic, since we reject the hypothesis of perfect positive correlation for either parameter (p-value < 0.001). We discuss comparisons of these results with previous literature in Subsection 5.3.

[21] We obtain the same conclusions when we consider stronger tests of temporal stability and dynamic consistency that also include utility curvature. In that case, the test for temporal stability is H_0: $E(\beta_{n1}) = E(\beta_{n2})$, $E(\delta_{n1}) = E(\delta_{n2})$ and $E(r_{n1}) = E(r_{n2})$; and the test for dynamic consistency is H_0: $E(\beta_{n2}) = 1$, $E(\delta_{n1}) = E(\delta_{n2})$ and $E(r_{n1}) = E(r_{n2})$.

[22] For instance, our evaluation classifies individual n's behavior as constant discounting in wave w if $(1 - \varepsilon_\beta) < \beta_{nw} < (1 + \varepsilon_\beta)$, whereas the rigid prediction requires that $\beta_{nw} = 1$.

TABLE 2
INDIVIDUAL-LEVEL DISCOUNTING BEHAVIOR WITH FULL CONTROLS FOR SAMPLE SELECTION AND ATTRITION

Variable	Estimate	Standard Error	p-Value	95% Confidence Interval	
		A. Constant discounting			
$\Pr(\lvert\beta_1 - 1\rvert < 0.025)$	0.669	0.074	<0.001	0.538	0.823
$\Pr(\lvert\beta_2 - 1\rvert < 0.025)$	0.598	0.066	<0.001	0.481	0.739
		B. Temporal stability			
$\Pr(\lvert\Delta\beta\rvert < 0.025)$	0.530	0.058	<0.001	0.430	0.659
$\Pr(\lvert\Delta\delta\rvert < 0.050)$	0.530	0.025	<0.001	0.477	0.575
$\Pr(\lvert\Delta\beta\rvert < 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.279	0.040	<0.001	0.209	0.364
$\Pr(\lvert\Delta\beta\rvert < 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.251	0.026	<0.001	0.206	0.308
$\Pr(\lvert\Delta\beta\rvert > 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.251	0.032	<0.001	0.182	0.306
$\Pr(\lvert\Delta\beta\rvert > 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.219	0.032	<0.001	0.153	0.280
		C. Dynamic consistency			
$\Pr(\lvert\beta_2 - 1\rvert < 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.315	0.041	<0.001	0.245	0.406
$\Pr(\lvert\beta_2 - 1\rvert < 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.283	0.032	<0.001	0.225	0.349
$\Pr(\lvert\beta_2 - 1\rvert > 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.215	0.035	<0.001	0.136	0.274
$\Pr(\lvert\beta_2 - 1\rvert > 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.187	0.035	<0.001	0.119	0.257

by $\varepsilon_\delta > 0$ instead.[23] We consider $\varepsilon_\beta = 0.025$ and $\varepsilon_\delta = 0.05$ as benchmark tolerances, and examine the sensitivity of our results to alternative configurations. Setting $\varepsilon_\delta = 0.05$ is arguably a natural benchmark for the baseline discount rate, since the smallest increment in annual interest rates across our discounting tasks for a particular time horizon is 5%. A benchmark tolerance for ε_β is less obvious and we set it equal to 0.025, which is close to the estimated standard deviation of the population distribution for the present bias parameter in both waves.

Table 2 reports the estimated population shares of decision makers that fall into different types of discounting behavior, based on the benchmark tolerance configuration. We find that a majority of Danes display *constant discounting* in both waves of the experiment: 66.9% of the decision makers have constant discount rates in wave 1, and 59.8% have constant discount rates in wave 2. The estimated population share is significantly different from 50% in wave 1 (p-value $= 0.022$), but we cannot reject that it is equal to 50% in wave 2 (p-value $= 0.138$).

The decision makers who display *temporal stability* in the baseline discount rate and the present bias parameter are estimated to make up 27.9% of the population, with a 95% confidence interval of [20.9%, 36.4%]. Hence, a majority of the population (72.1%) displays temporal instability in at least one of the two parameters. When we look at each parameter separately, those with temporally stable baseline discount rates make up 53.0% of the population, and those with temporally stable present bias parameters also make up 53.0%. The two parameters in the QH discounting function thus show a similar degree of temporal (in)stability. The population share of those with temporal instability in both parameters is estimated to be 21.9%.

Finally, the estimated share of decision makers who display *dynamic consistency* is equal to 31.5%, with a 95% confidence interval of [24.5%, 40.6%]. Dynamic consistency entails constant discounting in wave 2 along with temporally stability in baseline discount rates between the two waves. The share of decision makers with constant discounting in wave 2 is 59.8%, and that of those with temporally stable baseline discount rates is 53.0%. Constant discounting thus seems to be a more important source of dynamic consistency than temporal stability in baseline discount rates, although our inference concerning dynamic consistency is based on a joint probability that takes into account the estimated correlation coefficients for different preference parameters.

[23] Since we adopt continuous population distributions to model preference heterogeneity, setting $\varepsilon_\beta = \varepsilon_\delta = 0$ leads to the trivial conclusion that all three population shares are equal to 0.

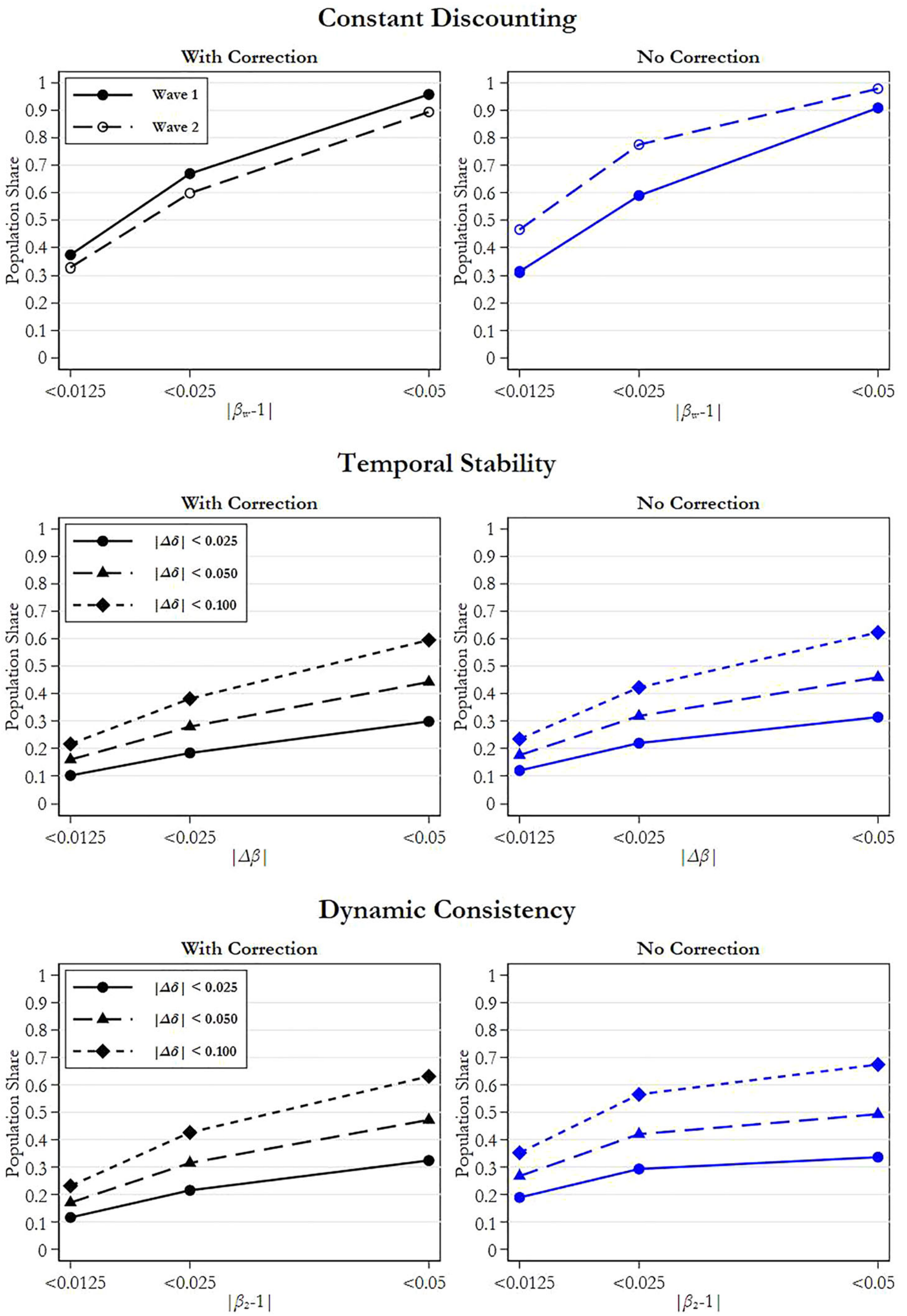

Figure 2

cumulative probabilities

Figure 2 illustrates the sensitivity of these findings to different values of the two noise tolerances. The three panels display the estimated shares of decision makers who reveal constant discounting, temporal stability, and dynamic consistency, respectively. The estimated population shares increase with the size of the tolerances, by construction, but the tendency is the same as before: when we look at the shares of decision makers with these traits we find more empirical support for constant discounting than temporal stability and dynamic consistency.[24]

4.2. *Are Risk and Time Preferences Correlated?* Turning to risk preferences, the upper-left panel in Figure 3 shows the estimated population distributions of the utility curvature parameter r_{nw} across the two waves, with controls for nonrandom selection and attrition bias. The population distribution of the utility curvature parameter shifts to the right in wave 2 compared to wave 1, but the apparent increase in risk aversion is *not* statistically significant (p-value $= 0.097$) at the 5% significance level. The estimated population mean is equal to 0.951 with a p-value of 0.066 in wave 1, and equal to 1.076 with a p-value of 0.089 in wave 2. We also observe that the population distribution in wave 2 has a smaller standard deviation than the distribution in wave 1; the estimated standard deviation is 0.725 in wave 1 and 0.597 in wave 2, and we cannot reject the null hypothesis that the estimated difference in the two coefficients is equal to 0 at the 5% level (p-value $= 0.088$). Hence, we find *temporal stability* with respect to the population mean, and also with respect to the standard deviation of the utility curvature parameter.[25]

The estimated population distributions of the shape parameter φ_{nw} in the PWF are displayed in the lower-left panel of Figure 3. The distributions control for selection and attrition bias, and we observe insignificant differences in the estimated population distributions of the shape parameter between the two waves. The estimated difference in the population mean between the two waves is statistically insignificant (p-value $= 0.758$), and we also find that the standard deviation of the population distribution is temporally stable (p-value $= 0.520$).[26]

Turning to the association between risk and time preferences, we find that the baseline discount rate is negatively correlated with utility curvature in both waves: the estimated values of corr(δ_{n1}, r_{n1}) and corr(δ_{n2}, r_{n2}) are -0.378 and -0.349, respectively, and both coefficients are significantly smaller than 0 (p-value < 0.001).[27] The negative correlation coefficients thus suggest that those with higher baseline discount rates tend to be less risk-averse, ceteris paribus.[28] The present bias parameter is significantly correlated with utility curvature in wave 1, but not in wave 2. In wave 1, more present-biased individuals have less concave utility: the estimated value of corr(β_{n1}, r_{n1}) is 0.183 and significantly greater than 0 (p-value < 0.001). In wave 2, however, the estimated value of corr(β_{n2}, r_{n2}) is 0.005 and we cannot reject that the two distributions are independent (p-value $= 0.949$).[29]

The estimated correlation coefficient between the baseline discount rate and the shape parameter for the PWF in wave 1, corr$(\delta_{n1}, \varphi_{n1})$, is 0.268. The corresponding coefficient in wave 2, corr$(\delta_{n2}, \varphi_{n2})$, is 0.342. Both coefficients are significantly greater than 0. The implications of these positive correlation coefficients are less straightforward to evaluate. When the

[24] Table C2 in Appendix C reports estimated population shares of decision makers who reveal constant discounting, temporal stability, and dynamic consistency, based on an alternative structural model with EUT to identify utility curvature. Using the same tolerance configurations as above, we find smaller population shares for each of those three types of discounting behavior compared to the model with RDU.

[25] The estimated correlation coefficient between the two wave-specific population distributions of the utility curvature parameter, corr(r_{n1}, r_{n2}), is equal to 0.668, and we reject the hypothesis that they are independent (p-value $<$ 0.001).

[26] The estimated between-wave correlation of the shape parameter, corr$(\varphi_{n1}, \varphi_{n2})$, is 0.871 with a standard error of 0.063, which suggests a strong degree of temporal stability at the individual level.

[27] Table B1 in Appendix B reports the estimated correlation coefficients.

[28] This finding is not an algebraic artifact. The discounted utility difference between sooner and later payment options changes in favor of the sooner option, as δ_{nw} (i.e., long-run delay aversion) increases or r_{nw} (i.e., concavity of the utility function) decreases. But this algebraic property says nothing about whether individuals with $\delta_{nw} > E(\delta_{nw})$ tend to have $r_{nw} < E(r_{nw})$.

[29] The estimated pattern of correlation coefficients between risk and time preferences under EUT is generally the same as the results under RDU, in terms of sign and statistical significance. The only exception is the estimate of corr(β_{n2}, r_{n2}), which is 0.210 and significantly greater than 0 under EUT (p-value $= 0.002$). The estimate of corr(β_{n1}, r_{n1}) remains positive and significant, and we thus find that more present-biased individuals tend to be less risk-averse in both waves.

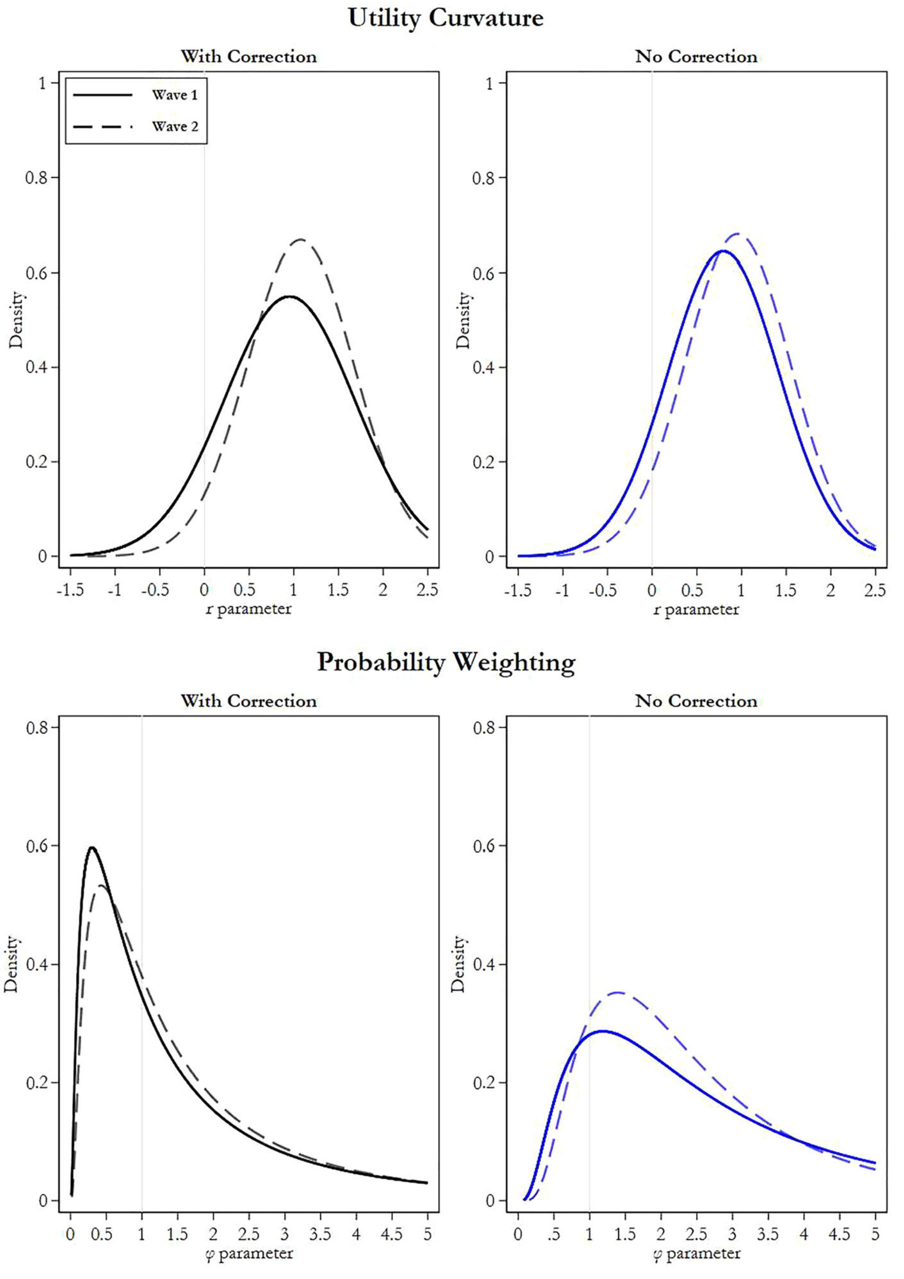

FIGURE 3

POPULATION DISTRIBUTIONS OF RISK PARAMETERS

probability of the best outcome increases, the PWF switches from being concave to convex when $\varphi_{nw} < 1$, and vice versa when $\varphi_{nw} > 1$. We find that $E(\varphi_{n1})$ and $E(\varphi_{n2})$ are equal to 2.171 and 2.091, respectively, which means that the average person has an S-shaped PWF. If the average decision maker underweights (overweights) a particular probability, someone with $\varphi_{nw} > E(\varphi_{nw})$ will underweight (overweight) it even further. The positive correlation coefficients thus suggest that the effects of S-shaped probability weighting tend to be more pronounced for those with higher baseline discount rates. The results are more mixed with respect to the correlation between the present bias and shape parameters. The value of corr(β_{n1}, φ_{n1}) is −0.213 and significantly smaller than 0, but corr(β_{n2}, φ_{n2}) is equal to 0.098 and insignificant (p-value = 0.189). Hence, in wave 1, more present-biased individuals tend to have PWFs that display more pronounced S shapes. This association, however, almost vanishes in wave 2.

4.3. *Controlling for Endogenous Selection and Attrition.* Our model accounts for endogenous sample selection and panel attrition by specifying selection and attrition equations in the style of Heckman (1979) and Capellari and Jenkins (2004), and allowing the error terms in those equations to be correlated with each other as well as each of the time- and risk-preference parameters. Our experimental design provides natural candidates for exclusion restrictions to assist empirical identification of the selection and attrition equations. The initial invitation letter randomized individuals to different recruitment fees, and the longitudinal design allows us to observe each subject's additional earnings from the first experiment. We assume that the recruitment fees influence the initial decision to accept the first invitation, but not the decision to accept the second invitation once we control for additional earnings from the first experiment. We maintain the usual hypothesis that the recruitment fees do not affect the subject's evaluation of time-dated payments and lottery pairs directly. Finally, subjects had to travel to hotel meeting rooms to participate in the first experiment, whereas each subject chose their own preferred venue for the second experiment.

These features of our experimental design allow us to distinguish the selection equation from the attrition equation. We include the recruitment fee only in the selection equation, and the actual earnings from the first experiment only in the attrition equation.[30] In addition, we augment the selection equation with each subject's home-to-hotel distance (in miles) and its square value, and attrition equation with self-reported income that is only available for those who participated in the first experiment; both equations also include the individual's age and gender. The distinctive theoretical structures that we have placed on the discounting choice model (namely, discounted utilities with QH discounting functions) and the risky choice model (RDU) immediately differentiate the two models from each other, as well as from the selection and attrition equations.

The sample selection equation reported in Table 1 shows that recruitment fees and the distance from an individual's home to a session venue significantly affect the decision to participate in the experiments.[31] The law of demand effectively applies to the participation decisions, with the propensity to self-select into the experiments increasing significantly when the recruitment fee is raised from 300 to 500 kroner for participation in wave 1. We also find a statistically significant and U-shaped association between the self-selection index and the home-to-hotel distance, with a negative and diminishing marginal effect of distance up to a turning point at 30.6 miles. As one may expect, people who live farther away from the session venues are less likely to participate, and people who live closer are more sensitive to a small increase in distance.[32] Looking at personal characteristics, middle-aged and older subjects are more likely to participate in the first wave than younger age groups.[33] Table 1 also shows that it is

[30] Adding the recruitment fee to the attrition equation does not change our results and conclusions with respect to constant discounting, temporal stability, and dynamic consistency. Table D1 in Appendix D reports the estimated parameters for the model where the recruitment fee is included in both the selection and attrition equation. Additional earnings in the initial experiment include payments for choices in three sets of decision tasks which elicit individual risk attitudes, discount rates, and correlation aversion, respectively. Additional earnings are thus negatively correlated with individual discount rates in the first experiment. We obtain virtually the same results when we restrict the measure of additional earnings to payments for choices in the risk-aversion tasks. Table D2 in Appendix D reports the estimated parameters for the model where additional earnings in the attrition equation refer to payments for choices in the risk-aversion tasks.

[31] The coefficient estimates for the sample selection and attrition equations have been rescaled so that the reported results can be interpreted in the same manner as the coefficient estimates for the usual censored probit models that standardize the variance of each component equation to unity.

[32] The marginal effect of distance is positive after the turning point, but this is an artifact of the quadratic specification that has limited practical significance, since only 20 of the 1996 invitees lived outside a 30.6-mile radius from a venue.

[33] Most variables have self-evident definitions. The "young" age group is defined as those less than 30 years, the "middle" age group is defined as those between 40 and 50 years, and the "old" age group is those over 50 years.

difficult to explain panel retention in terms of the subject's observed characteristics, though the point estimates suggest that those who earned more from the initial experiment are more likely to return to the second experiment if invited.[34]

The hypothesis of no endogenous selection bias involves nine parametric constraints: it states that the error term in the selection equation is uncorrelated with the error term in the attrition equation and the population distributions of the eight preference parameters in the structural model. We reject this hypothesis with a p-value < 0.001, so the result is consistent with significant sample selection. The hypothesis of no endogenous attrition bias implies that the error term in the attrition equation is uncorrelated with the eight preference parameters, which is also rejected (p-value < 0.001). This result is also consistent with significant sample attrition. The estimated correlation coefficient between the error terms in the selection and attrition equations is equal to -0.273 with a standard error of 0.055, which further implies that one cannot correct for attrition bias without also correcting for selection bias.[35]

From Figure 1, we can see the effects of controlling for selection and attrition bias on baseline discount rates and present bias parameters. The left panels show the estimated population distributions from a model with corrections for endogenous selection and attrition, and the right panels show the corresponding results from a model without such corrections. Although we model the time-preference parameters using S_B distributions that can display a variety of shapes, the corrected and uncorrected population distributions happen to take on similar shapes. The corrected and uncorrected point estimates are also similar. The estimated coefficients in the model without corrections for sample selection and attrition bias are reported in Table 3. The estimated mean and standard deviation of the uncorrected distribution for the baseline discount rate are 11.1% and 12.8% in wave 1, and 9.5% and 12.3% in wave 2. For the present bias parameter, the estimated mean and standard deviation is 0.998 and 0.030 in wave 1, and 0.997 and 0.021 in wave 2.

Incorrectly assuming away endogenous selection and attrition bias changes our population-level inferences on constant discounting in wave 2, but not in wave 1. Based on the estimated population means, we do not reject constant discounting in wave 1 (p-value $= 0.317$) or in wave 2 (p-value $= 0.317$). Moreover, we cannot reject the hypothesis of temporal stability at the 10% significance level (p-value $= 0.051$), whereas temporal stability was rejected at the 1% significance level in the model with corrections. Although we do not reject constant discounting in wave 2 and cannot reject temporal stability at the 10% significance level, we continue to reject the hypothesis of dynamic consistency in the model without corrections (p-value $= 0.009$).[36]

The results illustrate the important methodological point that inferences on constant discounting and temporal stability must be made jointly to infer dynamic consistency. In this instance, the two t-tests of single properties lead us to conclude that we have constant discounting in wave 2 and temporal stability between the two waves, but the joint Wald test of both properties leads us to conclude that we have dynamic *inconsistency*.

Lower income is defined in variable "IncLow" by a household income in 2008 below 300,000 kroner, and higher income is defined in variable "IncHigh" by a household income of 500,000 kroner or more.

[34] Removing personal characteristics from the selection and attrition equations does not change our overall results. Table D3 in Appendix D reports the estimated parameters for the model where the recruitment fee and home-to-hotel distance are included in the selection equation, and the recruitment fee and revised measure of additional earnings are included in the attrition equation.

[35] The attrition equation can display selection bias because an individual's attrition outcome is observed only if the individual participated in the initial experiment, just as the individual's discounting and risky choices are.

[36] Incorrectly assuming away sample selection, but still correcting for sample attrition is relatively easy to do. In a longitudinal experiment, there is often considerable "baseline" data collected on individuals, which can form the basis for corrections for attrition even if one does not, typically by definition, know anything about those that do not even turn up to the baseline. However, this special case also leads to some important differences in inferences compared to correcting for both selection and attrition jointly. It leads to incorrectly rejecting the null of constant discounting in wave 1, and incorrectly failing to reject the null of temporal stability. We do *correctly* reject the joint hypothesis of dynamic consistency at the 5% level, but *incorrectly* fail to also reject it at the 1% level.

TABLE 3
ESTIMATES OF STRUCTURAL PARAMETERS WITHOUT CONTROLS FOR SAMPLE SELECTION AND ATTRITION

Variable	Estimate	Standard Error	p-Value	95% Confidence Interval	
		Means of δ and β parameters in wave 1 and wave 2			
δ_1	0.111	0.008	<0.001	0.095	0.128
δ_2	0.095	0.007	<0.001	0.080	0.109
β_1	0.998	0.002	<0.001	0.994	1.002
β_2	0.997	0.003	<0.001	0.991	1.001
		Medians of δ and β parameters in wave 1 and wave 2			
δ_1	0.059	0.005	<0.001	0.049	0.068
δ_2	0.042	0.008	<0.001	0.025	0.058
β_1	1.001	0.002	<0.001	0.997	1.006
β_2	0.997	0.002	<0.001	0.993	1.002
		Standard deviations and correlation coefficients of δ and β parameters in wave 1 and wave 2			
$\sigma_{\delta 1}$	0.128	0.009	<0.001	0.111	0.147
$\sigma_{\delta 2}$	0.123	0.011	<0.001	0.100	0.143
$\sigma_{\beta 1}$	0.030	0.004	<0.001	0.023	0.039
$\sigma_{\beta 2}$	0.021	0.003	<0.001	0.015	0.028
$\rho_{\delta 1\ \delta 2}$	0.373	0.034	<0.001	0.318	0.453
$\rho_{\beta 1\ \beta 2}$	0.292	0.146	0.045	−0.062	0.502
		Means of r and φ parameters in wave 1 and wave 2			
r_1	0.797	0.072	<0.001	0.656	0.938
r_2	0.955	0.073	<0.001	0.811	1.099
φ_1	3.337	0.307	<0.001	2.736	3.938
φ_2	2.653	0.272	<0.001	2.120	3.186
		Standard deviations and correlation coefficients of r and φ parameters in wave 1 and wave 2			
σ_{r1}	0.620	0.039	<0.001	0.543	0.696
σ_{r2}	0.586	0.071	<0.001	0.447	0.725
$\sigma_{\varphi 1}$	3.328	0.779	<0.001	1.802	4.855
$\sigma_{\varphi 2}$	1.945	0.664	0.003	0.643	3.247
$\rho_{r1\ r2}$	0.679	0.077	<0.001	0.528	0.830
$\rho_{\varphi 1\ \varphi 2}$	0.248	0.160	0.120	−0.065	0.562

NOTE: Table B2 in Online Appendix B reports other correlation coefficients.

In Table 4, we use the results from the model without corrections for sample selection and attrition to evaluate the population shares of decision makers who reveal constant discounting, temporal stability, and dynamic consistency. Using the same tolerance configuration used for the model with corrections ($\varepsilon_\beta = 0.025$ and $\varepsilon_\delta = 0.05$), we continue to find constant discounting for a majority of the population in both waves. But in contrast to the corrected estimates, the uncorrected estimates suggest that constant discounting is more prevalent in wave 2 than wave 1. The estimated population share is 58.9% in wave 1 and 77.5% in wave 2, with confidence intervals of [47.6%, 72.1%] and [63.6%, 90.1%], respectively. The population share of those with temporally consistent discounting functions continues to be relatively small; the estimated population share is 31.8%, with a confidence interval of [21.3%, 42.8%]. Finally, the population share of those with dynamically consistent time preferences is equal to 42.0%. This estimate is 10.5 percentage points larger than the corresponding share in the corrected model, and we cannot reject the null hypothesis that half of the population have dynamically consistent preferences (p-value = 0.089).

4.4. *Restrictions on Probability Weighting.* The best known theory of decision making under risk is EUT, which allows one to estimate a nonlinear utility function and use it to control for the effects of utility curvature on the inferred discounting function. Our modeling framework employs RDU, which generalizes EUT, and we find statistically significant probability weighting in the population as a whole. Specifically, our model nests EUT as a special case that constrains the shape parameter of the PWF, φ_{nw}, to 1 for every person n and wave w. As Table 1 shows, we reject this hypothesis: the estimated standard

TABLE 4
INDIVIDUAL-LEVEL DISCOUNTING BEHAVIOR WITHOUT CONTROLS FOR SAMPLE SELECTION AND ATTRITION

Variable	Estimate	Standard Error	p-Value	95% Confidence Interval	
		A. Constant discounting			
$\Pr(\lvert\beta_1 - 1\rvert < 0.025)$	0.589	0.063	<0.001	0.476	0.721
$\Pr(\lvert\beta_2 - 1\rvert < 0.025)$	0.775	0.068	<0.001	0.636	0.901
		B. Temporal stability			
$\Pr(\lvert\Delta\beta\rvert < 0.025)$	0.577	0.072	<0.001	0.434	0.717
$\Pr(\lvert\Delta\delta\rvert < 0.050)$	0.496	0.026	<0.001	0.449	0.552
$\Pr(\lvert\Delta\beta\rvert < 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.318	0.055	<0.001	0.213	0.428
$\Pr(\lvert\Delta\beta\rvert < 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.260	0.021	<0.001	0.216	0.298
$\Pr(\lvert\Delta\beta\rvert > 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.179	0.039	<0.001	0.107	0.256
$\Pr(\lvert\Delta\beta\rvert > 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.244	0.040	<0.001	0.164	0.324
		C. Dynamic consistency			
$\Pr(\lvert\beta_2 - 1\rvert < 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.420	0.047	<0.001	0.330	0.512
$\Pr(\lvert\beta_2 - 1\rvert < 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.354	0.034	<0.001	0.276	0.406
$\Pr(\lvert\beta_2 - 1\rvert > 0.025; \lvert\Delta\delta\rvert < 0.05)$	0.076	0.026	0.003	0.031	0.130
$\Pr(\lvert\beta_2 - 1\rvert < 0.025; \lvert\Delta\delta\rvert > 0.05)$	0.149	0.044	0.001	0.067	0.240

deviation of φ_{nw} is significantly greater than 0 in each wave, which suggests that φ_{nw} is heterogeneous.

Nevertheless, given the prominence of EUT, Appendix C reports the EUT-based estimation results for comparisons with the RDU-based results that we have presented so far. In the EUT-based results, the estimated population means of the discounting parameters are very close to the null hypothesis values under constant discounting, temporal stability, and dynamic consistency: the point estimates of $E(\delta_{n1})$, $E(\delta_{n2})$, $E(\beta_{n1})$, and $E(\beta_{n2})$ are 12.4%, 13.3%, 0.993, and 1.003, respectively. But we obtain relatively small standard errors for these estimates. We reject constant discounting in wave 1 (p-value < 0.001), although that result for wave 1 does not matter for the test of dynamic consistency. We do not reject constant discounting in wave 2 (p-value $= 0.262$). We do reject temporal stability (p-value $= 0.007$). However, despite rejecting temporal stability, we do not reject dynamic consistency (p-value $= 0.372$). Hence, one would come to the wrong conclusion about dynamic consistency if one constrained our specification by imposing EUT instead of RDU.

5. RELATED LITERATURE

5.1. *Dynamic Consistency.* Halevy (2015) provides a particularly clear statement of the *theoretical* interplay between constant discounting, temporal stability, and dynamic consistency. He *evaluates* that interplay using a very different approach than we use. His identification and modeling strategy is based on binary preference relations over pairs of sooner and later payments that are relevant to analysts who are interested in making *nonstructural* inferences by studying aggregate and individual-level *choice patterns*. The time delay between sooner and later payments was one week in all pairwise choices, and identification is based on variation in decision dates and time delays to sooner payments.[37] He used undergraduate students from a Canadian university as subjects in his experiments, an appropriate convenience sample to initially examine behavior in this, and other, settings.

Halevy (2015) recruited 149 students in the first wave of the main experiment, of which 130 also participated in the second wave. The sooner payment was either 10 or 100 Canadian

[37] The one week time delay between sooner and later payments may not be sufficient to detect present bias if subjects do not discriminate between the two payments over the brief time interval (e.g., a subject might view the "present" as today, this week or this month, at least under some models of discounting other than the QH model, as noted by Takeuchi, 2011). This issue pertains to the delay between the two payment dates, and is not related to, or mitigated by, the two treatments with variation in the front-end delay to the sooner payment.

dollars. Later payments varied between \$9.90 and \$10.00 in increments of \$0.10 in the low-stakes treatment; and between \$99 and \$110 in increments of \$1 in the high-stakes treatment.[38] The observed patterns of intertemporal choices by his subjects do *not* reject constant discounting, *do* reject temporal stability, and *partly* reject dynamic consistency. Inferences about time *preferences* are not the same as inferences about binary *choice patterns*, however, unless one assumes away "behavioral noise." It is perfectly possible for observed choice patterns to be driven by behavioral errors that are correlated with latent time preferences.

Studying choice patterns does not require explicit parameterization of the underlying discounting model that our analysis entails. This nonparametric implementation is an attractive feature if the goal is to study the incidence of dynamic consistency over a specific time horizon in the experimental design. To draw more design-free inferences based on discounting parameters, however, it is crucial to introduce appropriate variations in time horizons to trace the shape of the discounting function. Our design therefore presented subjects with various time horizons, between two weeks and one year, for receipt of the later option. We also varied receipt of the earlier option between 0 and 30 days. This design allows estimation of a rich set of discounting models (e.g., Exponential, QH, Smoothly Hyperbolic), which define the discount rates of the subjects, at a particular point of time, over horizons between today and 13 months into the future. What matters for structural estimation is that the revisit to the subject occurs within that interval, not at any particular point in that interval. Admittedly, by having thus relaxed the temporal sequencing of the experimental design, we are not able to engage in a nonparametric analysis of choice patterns. This highlights that the structural and nonstructural approaches have complementary roles to play, and it would be erroneous to claim that one is superior to the other.

Some studies use willingness to pay for commitment devices as indicators of subjective awareness of behavioral tendencies to act in a dynamically inconsistent manner (e.g., Augenblick et al., 2015; and Giné et al., 2018). Although some of those devices may be useful in restricting intertemporal choice, O'Donoghue and Rabin (2015, p. 277ff.) correctly note that they are not measures of dynamic inconsistency. The effects of commitment devices are, at best, suggestive of the effects of dynamic consistency, but could also reflect other behavioral traits such as low willpower.

Only a handful of experimental studies with nonstudent subjects have longitudinal designs that enable statistical tests of temporal stability alongside tests for constant discounting. Even then, those studies focus on relatively narrow segments of the general population; do not account for the interplay of interpersonal heterogeneity, endogenous sample selection, and endogenous panel attrition; and, with the exception of Janssens et al. (2017), do not consider formal or informal tests of dynamic consistency.[39]

Janssens et al. (2017) studied adult members of low-income farming households in Kwara, Nigeria, who were already participating in a monthly health and finances survey. The temporal sequencing of their experimental design was directly inspired by Halevy (2015), though they presented a time delay of one month between sooner and later payments instead of one week. They observe response patterns exhibiting constant discounting (56%), temporal

[38] To facilitate comparisons of implied interest rates in decision tasks with other studies and field alternatives, it is convenient to state these on an annual basis. Weekly interest rates between −1% and 10% are equivalent annual interest rates between −67.8% and 14,104%.

[39] Although Sadoff et al. (2020) allude to dynamic inconsistency, their approach falls into a different strand of literature, since they do not elicit discounting behavior. The population in their study refers to customers at two grocery stores in low-income areas of Chicago and Los Angeles, and the notion of dynamic consistency refers to disparities in the demand for a basket of groceries from the order date to the delivery date. They test this hypothesis by examining whether the taste coefficients in a rank-ordered logit model change over time. One cannot disentangle dynamic consistency from temporal stability in this setting. Indeed, in the stated preference literature where the rank-ordered logit model originates from, the same type of parametric restriction is used to test temporal stability (see Doiron and Yoo, 2017).

instability (58%), or dynamic consistency (57%) for a majority of subjects.[40] Meier and Sprenger (2015) recruited a sample of low-to-moderate income earners who obtained a free tax preparation service at a particular site run by the City of Boston. They consider QH discounting and find empirical support for nonconstant discounting and temporal stability. The population of interest in Kirby et al. (2002) is residents in two villages in a tropical rain forest area of Bolivia. They assume away behavioral noise and algebraically derive hyperbolic discount rates, which preclude statistical tests of constant discounting, and find that within-subject correlation in the hyperbolic discounting parameter is 0.32 or less over the four quarters in their longitudinal design. Finally, Dean and Sautmann (2021) recruited a sample of household heads who participated in a children's health-care program in a periurban area of Bamako, Mali. The sooner option was either paid immediately or delayed by one week, and the time horizon between the sooner and later options was one week in all decision tasks. Summarizing the choice patterns observed at weekly intervals for three consecutive weeks, they report that 70–76% of the subjects display constant discounting in each week, and the within-subject correlation in choices ranges from 0.67 to 0.72 across the weekly waves of the experiment.[41]

In general, we contribute to this literature with the first set of longitudinal findings based on a nationally representative sample of the general population; a modeling framework which jointly addresses preference heterogeneity, selection bias, and attrition bias; and a coherent inferential approach that allows one to use the same set of preference parameter estimates to test for constant discounting, temporal stability, and dynamic consistency at both individual and population levels.

5.2. *General Issues in the Elicitation of Time Preferences*

5.2.1. *Time preferences for populations or individuals?* As documented in Appendix A, our estimation method does not require that each subject responds to the same set of questions in both waves of the experiment. We fit a joint structural model to a pooled sample of every individual's selection outcome and, where applicable, attrition outcome and choices during the experiment. Our model incorporates a random coefficient specification to account for interpersonal preference heterogeneity, along with selection and attrition equations to distinguish subjects from nonparticipants. An alternative approach to addressing interpersonal heterogeneity is to fit a structural model to observations on each subject separately.[42] This approach precludes formal statistical inferences about the population from which the subjects are drawn as well as corrections for selection and attrition biases.

5.2.2. *Time preferences and the modeling of behavioral noise.* Our stochastic specification adopts the contextual utility model of Wilcox (2011) for risk-aversion tasks and a Fechner error specification for discounting tasks. Apesteguia and Ballester (2018) argue that the random preference (RP) model is a more attractive stochastic specification to use in structural estimation since it generates choice probabilities which vary monotonically in a risk- or time-preference parameter. As Wilcox (2008, subsection 4.1.1) already noted, however, the RP

[40] We caution that their experiment employed the "convex budgeting" procedure of Andreoni and Sprenger (2012), and that they encountered (p. 88) a now-familiar problem with the results. So many tokens were allocated to "corners" that their model of time preferences implied risk-*loving* behavior. They reject that inference as a priori implausible, and just impose risk neutrality in order to draw inferences about discount rates.

[41] Annual interest rates vary between 0% and 4,096% in Janssens et al. (2017), between 4% and 1,899,088% in Meier and Sprenger (2015), between 15.14% and $10^{24} \times 6.94\%$ in Kirby et al. (2002), and between −99.99% and $10^{42} \times 2.91\%$ in Dean and Sautmann (2021).

[42] This approach is very demanding of the size of the sample of choices that each individual makes, and of course the extra demands of joint estimation of risk preferences and time preferences. A better approach would be to use Bayesian Hierarchical Models, which allow inferences from the pooled sample of individuals to provide informative prior distributions for estimation of individual-level parameters (e.g., Gao et al., 2023). This approach would also allow individual subjects to have different or smaller choice sets, and to have between-subject treatments.

model generates implausible predictions in other respects. For example, the RP model predicts that the probability of choosing a lottery instead of its mean-preserving spread remains constant regardless of their variance difference, as long as both alternatives are defined over the same set of three outcomes. Moreover, the stochastic monotonicity property of the RP model is irrelevant to our analysis which jointly estimates the QH discounting function with RDU. This property only applies to the stand-alone, nonjoint, analysis of a decision model with a one-dimensional preference parameter (Lau and Yoo, 2023), such as an exponential discounting function or EUT with a CRRA utility function.[43]

5.2.3. *Time preferences and arbitrage opportunities.* There are several issues that arise when one considers the opportunities that subjects have when they arbitrage their choices in experiments with choices they might make in the field or with their income and wealth outside the experiment. It is not appropriate to consider all of these issues here, other than to consider their effects on risk or time preferences that might be considered as sources of temporal instability and hence as confounds to our inferences about dynamic stability.

One possible source of arbitrage was identified by Coller and Williams (1999) and studied extensively in the field by Harrison et al. (2002). It arises when a subject faces borrowing and savings interest rates that are identical, even if they might differ from individual to individual: in the extreme characterization of "perfect capital markets" they are the same for everyone. It follows from the Fisher Separation Theorem that we would then be unable to identify individual time preferences from observed choices in an experiment. However, we do not live in that perfect world. It has been common for experiments on time preferences, such as Coller and Williams (1999) for university students, and Harrison et al. (2002), Andersen et al. (2008, 2014, 2018b) for adult Danes, to collect information on individual borrowing and savings interest rates. The gap between them makes it possible to infer time preferences without concerns for arbitrage.[44]

A second possible source of arbitrage arises from the possibility that individuals might perfectly integrate their (actual or expected) income from experiments with their personal wealth. This would have an effect on the appropriate argument for the utility function, in turn serving as a possible confound for inferences about risk preferences and/or time preferences. Of course, to the extent that one can control for asset integration by proxies for income or wealth, they can be corrected for. An opportunity for a direct test arises in Denmark, due to the ability to link our experimental data directly to the Danish Registry, which contains administrative data on various types of income and (financial) wealth of the subject. Andersen et al. (2018a) undertake this direct test and find that subjects do not perfectly integrate experimental income with field wealth, and that one can safely ignore that potential confound, at least for the Danish population.

A third possible source of arbitrage mentioned in the literature, such as Cubitt and Read (2007) and Cohen et al. (2020), arises from the relationship between choices over time-dated money and intertemporal consumption. When one thinks of *utility* discount rates, it becomes more natural for some economists to think in terms of the utility of consumption flows instead of stocks of money. In that case, Cubitt and Read (2007) show that, even if one assumes linear utility functions for simplicity of exposition, discount rates over money do not necessarily elicit discount rates over consumption.

[43] The two-parameter RP model reported by Apesteguia et al. (2024) jointly estimates a one-parameter RP model for choice data under risk and a separate one-parameter RP model for choice data over time delay. In their study, risk preferences are modeled as EUT with a CRRA utility function and time preferences are modeled as exponential discounting.

[44] When the borrowing and savings rates might influence responses in experiments, or inferences about standard errors of estimated discount rates, standard econometric methods allow one to "censor" observed choices in those experiments since one knows the censoring threshold from the elicited data on those rates. Coller and Williams (1999) document how one does this.

One can make structural assumptions about the link between time-dated money and consumption flows, and hence "translate" inferences about the former into inferences about the latter. This is what Andersen et al. (2008) did, with a "dual-self" model of decision making following Fudenberg and Levine (2006). This approach does not *test* the proposition in question. Instead it makes assumptions, whether or not they are a priori plausible, that are hard to test, and that allow one to infer discount rates over consumption flows if true.

It is also possible to ask if the "worst case" preferences that generate differences between discount rates over money and consumption are plausible. This worst case requires that the elasticity of substitution between consumption flows in two periods is "sufficiently" low. Assumptions about this elasticity allow one to bound the possible difference between the two discount rates, and those bounds can become very tight for plausible elasticities. A more direct response is to ask if this "worst case" behavior is indeed observed, when one modifies the basic experimental design to allow it to show itself. A simple modification of the canonical experimental task we used channels a constructive suggestion by Cubitt and Read (2007, p. 384). The constructive suggestion referred to is to simply allow each subject to form a portfolio between the SS amount of money and the LL amount of money, instead of having to choose one or the other. In other words, to form an *interior* portfolio from the two choice outcomes. There is evidence[45] that one "almost never" observes the preferences, at the individual subject level, that generate the problems posed by Cubitt and Read (2007, p. 384). This suggestion was popularized by Andreoni and Sprenger (2012), with some other extensions.[46] If individual subjects chose an interior solution, the worst-case conditions required for arbitrage between monetary allocations and consumption allocations to be an issue arise; if individual subjects choose one of the extreme solutions, this arbitrage issue does not apply as a matter of theory.

5.2.4. *Time preferences over primary or monetary rewards?* The use of experimental decisions concerning money, as in our design, to study intertemporal choice models such as Equations (1) and (2) raises questions about the possible implications of mismatches between the timing of monetary payment and actual consumption. One response is that the argument of the utility function need not be seen as actual consumption, and that it is sensible to define the utility function over income, an approach taken in many economic applications including the analysis of risk preferences. Another response comes from a strand of literature which argues that time preferences are more appropriately studied using decisions involving goods (e.g., ice cream) or activities (e.g., real effort task) that must be consumed or performed during the experiment. Cohen et al. (2020, section 3) review studies involving such *primary rewards*, and note that these tend to report higher discount rates and more frequent violations of constant discounting. Their review also highlights that it is debatable whether primary rewards indeed offer better controls over intertemporal utility flows than monetary rewards. For example, subjects might compensate for their participation in the experiment by altering their behavior outside the lab, such as skipping a regular snack after consuming ice cream in the experiment, or resting after engaging in strenuous effort tasks, all scenarios suggested by Augenblick et al. (2015). One might also point out that primary rewards come with potential confounds that would be expected a priori to vary from subject to subject (e.g., dietary preferences or intrinsic enjoyment of video games used as effort tasks). In that respect, the fungibility of money provides greater control. We concur with the broader view that both primary and monetary rewards are of interest in different applications, and that there is no formal or informal basis to exclude one in favor of the other.

Augenblick et al. (2015) use a longitudinal design and evaluate constant discounting over money and work effort. They pool intertemporal choices from both waves of their

[45] Presented in Harrison and Swarthout (2011).

[46] Andreoni and Sprenger (2012) extended the method proposed here by giving the subject 100 tokens to allocation between the sooner and later time period, and then *varying the exchange rate* between tokens and money for sooner or later amounts.

longitudinal experiment with undergraduate students and structurally estimate QH discounting functions over money and work effort.[47] They find that subjects are more susceptible to present bias in work effort tasks compared to monetary decision tasks. However, they do not utilize the longitudinal dimension of their data to test for temporal stability and dynamic consistency, since they explicitly *assume* temporal stability.

Augenblick and Rabin (2019) use a longitudinal design with similar work effort tasks (identification of Greek letters) to study intertemporal choice. They asked subjects to state how many computer-based tasks the subjects were willing to complete at different dates and piecemeal payments. All payments to the subjects were made at specific dates, and the QH discounting function in their intertemporal model is identified from variation in decision and work effort dates whereas payment dates are kept fixed. They do not evaluate dynamic consistency but test for constant discounting assuming nonlinear (dis)utility over work effort and linear utility over income from the work effort tasks.

5.3. *Comparison of Results.* Chuang and Schechter (2015) report in a literature review that the "correlation of time preferences over time" in longitudinal experiments with real incentives varies between 0.004 and 0.75. In each of those experiments, subjects were asked to complete a structured list of choices between sooner and later payments, and an algebraic transformation of each subject's switching point from sooner to later payments was used as a measure of their time preferences. Thus, those correlation coefficients are effectively descriptive statistics for raw data.

It is difficult to compare the magnitudes of those correlation coefficients across studies, because the algebraic transformation in question varies from study to study. Dean and Sautmann (2021) compute correlation coefficients for switching points observed in two different periods directly, without taking any further transformation of the data. Before computing the correlation coefficients, Kirby et al. (2002) and Kirby (2009) transform each switching point into the parameter κ_w from the hyperbolic discounting function $D_w(t) = 1/(1 + \kappa_w t)$ reviewed in Section 3. Wölbert and Riedl (2014) instead transform each switching point into an exponential discount rate. Meier and Sprenger (2015) also undertake this type of algebraic calculation, because their structural model does not incorporate correlated preference parameters. Exploiting variations in switching points across tasks with and without front-end delays, they compute that $\text{corr}(\delta_{n1}, \delta_{n2})$ and $\text{corr}(\beta_{n1}, \beta_{n2})$ are equal to 0.246 and 0.364, respectively, which happen to be similar to our structural estimates. Similarly, Yoon (2020) implements a cross-sectional experiment, using algebraic transforms to infer discount rates assuming linear utility. These attempts to derive discounting functions from switching points rely on assumptions that the utility function is linear, that there is no sample selection into the experiment, and that there is no behavioral error in the subject's decision making.

6. CONCLUSIONS

At a substantive level, we find dynamic inconsistency in the Danish population between 2009 and 2010. In an analysis that allows for small classification errors, we find that 68.5% of the decision makers display dynamic inconsistency, by violating constant discounting in the

[47] Augenblick et al. (2015) elicited time preferences from choice allocations in convex budget sets similar to those in Andreoni and Sprenger (2012). There were four sets of monetary decision tasks in their design. In the first week of the experiment, the subjects were asked to allocate \$20 between sooner payments now and/or later payments in four weeks; sooner payments now and/or later payments in seven weeks; and sooner payments in four weeks and/or later payments in seven weeks. Four weeks later the subjects were asked to repeat the first set of decision tasks with sooner payments now and/or later payments in four weeks. These tasks were repeated with different interest rates. In the work effort tasks, the subjects were endowed with 50 replications of specific computer-based tasks that involved identification of Greek letters and completion of partial Tetris games. In the first week of the experiment, the subjects were asked to allocate the computer-based tasks between work effort in one week and/or work effort in two weeks at various intertemporal exchange rates, and one week later they were asked to allocate the computer-based tasks between work effort now and/or work effort in one week.

second wave of the experiment and/or temporal stability in baseline discount rates between the first and second waves. The results refer to a stratified sample of the adult population in Denmark, and the statistical analysis recognizes preference heterogeneity within that population while controlling for endogenous sample selection and attrition.

A simple numerical example illustrates this finding. Imagine an SS outcome of 1,000 kroner in one year and an LL outcome of 1,100 kroner in two years. Using the estimated population means for β and δ from Table 1 for wave 1, we calculate present values of the SS (LL) outcome today of 903.5 (896.2) kroner, making the SS outcome more attractive. If wave 2 occurs six months from the present, and using the estimated population means for wave 2, these present values would be 953.9 (976.1) kroner, making the LL outcome more attractive, and reversing the preferences in wave 1. If wave 2 occurs one year from the present instead, these present values for wave 2 become 1,000 (1,012.0), again leading to the preference reversal. Figure 4 illustrates this simple example. The horizontal line is the present value of the SS outcome of 1,000 kroner and the upward sloping line is the present value of LL outcome. The dashed, vertical line indicates the indifference point between the SS and LL outcomes, which moves from 1,109 kroner when the two options are evaluated today, to 1,075 kroner when the two options are evaluated after six months, and to 1,087 kroner when they are evaluated after one year. Dynamic inconsistency in this simple example thus occurs at annual interest rates between 7.5% and 10.9%.[48]

At a methodological level, we demonstrate the careful interplay of theory, experimental design, and econometric inference needed to draw our substantive conclusion on dynamic inconsistency. We know from the literature on individual discounting that one needs to pay attention to issues such as utility over income (or goods) to draw reliable inferences about time preferences, whether that issue was attended to by experimental design (e.g., Laury et al., 2012; and Andreoni and Sprenger, 2012) or joint estimation across several experimental tasks (e.g., Andersen et al., 2008). We also need to consider ways in which discount rates may be nonconstant, such as QH discounting functions (e.g., Phelps and Pollak, 1968; Elster, 1979; and Laibson, 1997), and we need to consider corrections for sample selection and attrition in analyses of longitudinal data to infer nonbiased preferences of a population (e.g., Harrison et al., 2020).

Inferences about dynamic consistency require careful consideration of all these issues. The need to test for constant discounting and temporal stability in order to draw inferences about dynamic consistency is stressed by Halevy (2015), which serves as a warning against claiming that nonconstant discounting necessarily implies dynamic consistency. Our results show that it is critical to test joint hypotheses of nonconstant discounting and temporal stability, which requires longitudinal data and, for now, a structural model of time preferences.

Our econometric analysis is facilitated by access to a remarkable combination of Danish civil registry and experimental data. However, this is not to say that our structural approach to address selection and attrition bias is applicable only in this rich data environment. Longitudinal data sets normally provide sufficient information to correct for endogenous attrition bias, and correction for endogenous selection bias does not necessarily require access to administrative data. In some cases, experiments on risk and time preferences have been combined with large household surveys, such as the Living Standard Survey in Vietnam (Tanaka et al., 2010), the U.K. Household Longitudinal Study (Galizzi et al., 2016) and the American Life Panel (Dimmock et al., 2021). One may then merge the sample of experimental

[48] One can of course look at other numerical examples where the preference reversal moves in the opposite direction. Suppose the baseline discount rate is temporally stable with $\delta_1 = \delta_2 = 0.109$ and the present bias parameter in wave 2, β_2, is equal to 0.989, such that constant discounting is violated in wave 2. The decision maker will make the same choice when the two delayed outcomes are evaluated today and after six months, and select the LL outcome when the annual interest rate is higher than 10.9%. However, the decision maker will select the SS outcome at annual interest rates below 12.1% when the two outcomes are evaluated after one year, leading to the classical example of preference reversal in intertemporal choice at annual interest rates between 10.9% and 12.1%.

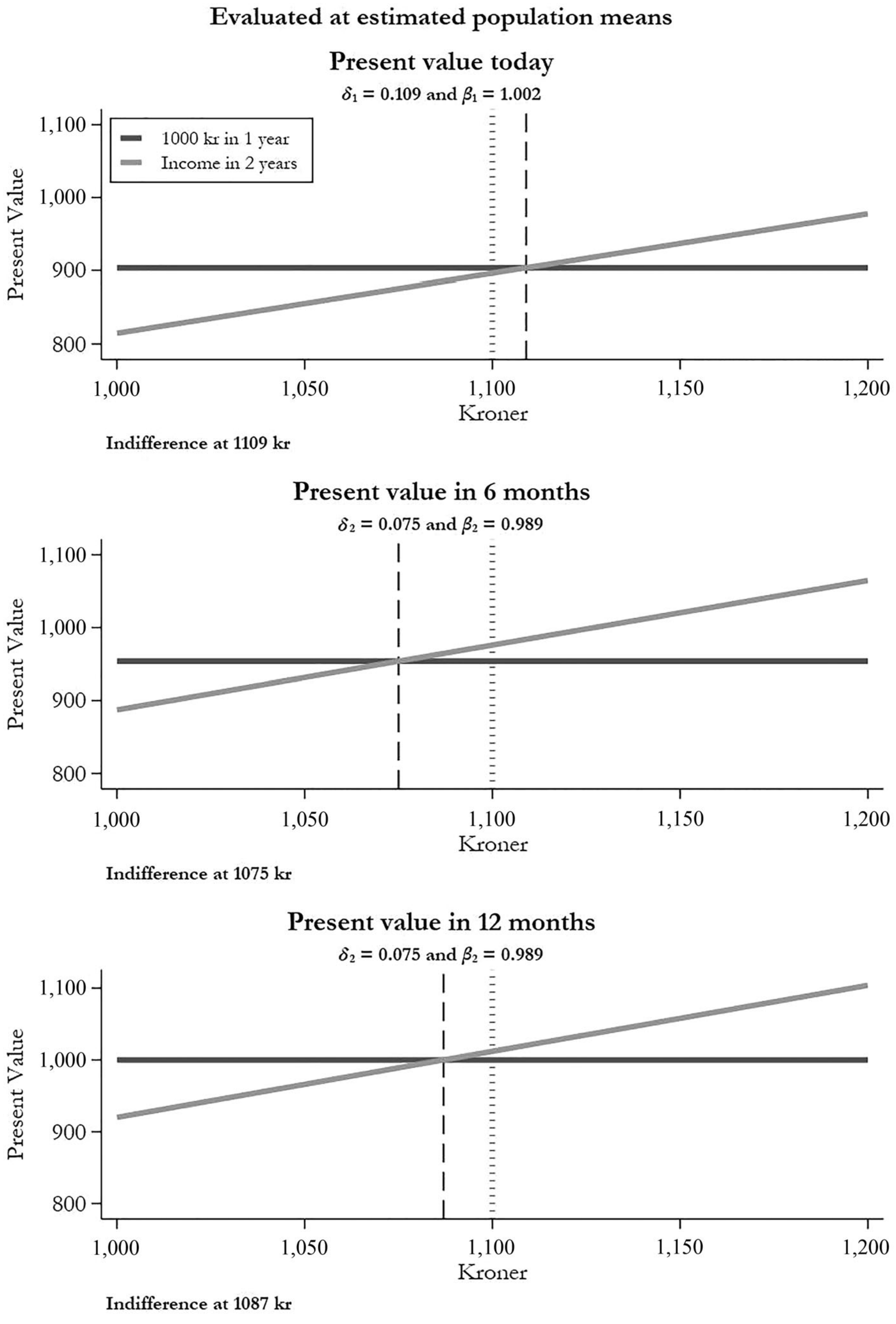

FIGURE 4

INTERTEMPORAL CHOICE UNDER RISK NEUTRALITY

participants with the sample of nonparticipants in the main survey, and evaluate the effects of self-selection into the experiment relative to the sampling frame of the main survey.

Even when experimental data are not directly linked to survey data, it is customary to comment on potential selection bias by comparing the experimental sample to an unrelated household survey sample in terms of average sociodemographic characteristics. If one is willing to accept the validity of such informal comparisons, one may as well consider the same survey sample as a sample of nonparticipants and formally test for endogenous selection. A unique feature of our experimental design is exogenous variation in recruitment fees that we use as an exclusion restriction to estimate the selection equation: without a deliberate design, it is difficult to find an equally attractive restriction. This challenge has not deterred the use of sample selection models in other empirical studies of economic behavior. In our view, formal correction for endogenous selection and attrition bias is an important step that one should consider in any field experiment when those studies intend to make inferences for broader segments of the population.

DATA AVAILABILITY STATEMENT The data that support the findings of this study are openly available in openCPSR at https://www.openicpsr.org/openicpsr/project/208322, reference number 208322.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Online Appendices

REFERENCES

Andersen, S., J. Cox, G. W. Harrison, M. I. Lau, E. E. Rutström, and V. Sadiraj, “Asset Integration and Attitudes to Risk: Theory and Evidence,” *Review of Economics & Statistics* 100 (2018a), 816–30.

———, G. W. Harrison, M. I. Lau, and E. E. Rutström, “Eliciting Risk and Time Preferences,” *Econometrica* 76 (2008), 583–618.

———, ———, ———, and ———, “Discounting Behavior: A Reconsideration,” *European Economic Review* 71 (2014), 15–33.

———, ———, ———, and ———, “Multiattribute Utility, Intertemporal Utility and Correlation Aversion,” *International Economic Review* 59 (2018b), 537–55.

Andreoni, J., and C. Sprenger, “Estimating Time Preferences from Convex Budgets,” *American Economic Review* 102 (2012), 3333–56.

Apesteguia, J., and M. A. Ballester, “Monotone Stochastic Choice Models: The Case of Risk and Time Preferences,” *Journal of Political Economy* 126 (2018), 74–106.

———, ———, and A. Gutierrez-Daza, “Random Discounted Expected Utility,” Working Paper, Universitat Pompeu Fabra, July 2024, Available at https://jose-apesteguia.github.io/R-DEU.pdf.

Augenblick, N., M. Niederle, and C. Sprenger, “Working over Time: Dynamic Inconsistency in Real Effort Tasks,” *Quarterly Journal of Economics* 130 (2015), 1067–115.

———, and M. Rabin, “An Experiment on Time Preference and Misprediction in Unpleasant Tasks,” *Review of Economic Studies* 86 (2019), 941–75.

Bommier, A., “Uncertain Lifetime and Intertemporal Choice: Risk Aversion as a Rationale for Time Discounting,” *International Economic Review* 47 (2006), 1223–46.

Capellari, L., and S. P. Jenkins, “Modelling Low Income Transitions,” *Journal of Applied Econometrics* 19 (2004), 593–610.

Chakraborty, A., Y. Halevy, and K. Saito, “The Relation between Behavior under Risk and Time,” *American Economic Journal: Insights* 2 (2020), 1–16.

Chuang, Y., and L. Schechter, “Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results,” *Journal of Development Economics* 117 (2015), 151–70.

Cohen, J., K. M. Ericson, D. Laibson, and J. M. White, “Measuring Time Preferences,” *Journal of Economic Literature* 58 (2020), 299–347.

Coller, M., and M. B. Williams, “Eliciting Individual Discount Rates,” *Experimental Economics* 2 (1999), 107–27.

Cubitt, R. P., and D. Read, “Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption?” *Experimental Economics* 10 (2007), 369–89.

Dean, M. R. W., and A. Sautmann, “Credit Constraints and the Measurement of Time Preferences,” *Review of Economics and Statistics* 103 (2021), 119–35.

DIMMOCK, S. G., R. KOUWENBERG, O. S. MITCHELL, and K. PEIJNENBURG, "Household Portfolio Underdiversification and Probability Weighting: Evidence from the Field," *Review of Financial Studies* 34 (2021), 4524–63.

DOHMEN, T., A. FALK, D. HUFFMAN, and U. SUNDE, "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (2010), 1238–60.

DOIRON, D., and H. I. YOO, "Temporal Stability of Stated Preferences: The Case of Junior Nursing Jobs," *Health Economics* 26 (2017), 802–9.

ECKEL, C., C. JOHNSON, and C. MONTMARQUETTE, "Savings Decisions of the Working Poor: Short- and Long-term Horizons," in J. Carpenter, G. W. Harrison, and J. A. List, eds., *Field Experiments in Economics*, Research in Experimental Economics, Volume 10 (Greenwich, CT: JAI Press, 2005).

ELSTER, J., *Ulysses and the Sirens: Studies in Rationality and Irrationality* (Cambridge: Cambridge University Press, 1979).

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

FUDENBERG, D., and D. K. LEVINE, "A Dual-Self Model of Impulse Control," *American Economic Review* 96 (2006), 1449–76.

GALIZZI, M. M., S. R. MACHADO, and R. MINIACI, "Temporal Stability, Cross-Validity, and External Validity of Risk Preference Measures: Experimental Evidence from a UK Representative Sample," LSE Research Online Documents on Economics 67554, London School of Economics and Political Science, 2016,

GAO, X. S., G. W. HARRISON, and R. TCHERNIS, "Behavioral Welfare Economics and Risk Preferences: A Bayesian Approach," *Experimental Economics* 26 (2023), 273–303.

GINÉ, X., J. GOLDBERG, D. SILVERMAN, and D. YANG, "Revising Commitments: Field Evidence on the Adjustment of Prior Choices," *Economic Journal* 128 (2018), 159–88.

HALEVY, Y., "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect," *American Economic Review* 98 (2008), 1145–62.

———, "Time Consistency: Stationarity and Time Invariance," *Econometrica* 83 (2015), 335–52.

HARRISON, G. W., M. I. LAU, and M. B. WILLIAMS, "Estimating Individual Discount Rates for Denmark: A Field Experiment," *American Economic Review* 92 (2002), 1606–17.

———, ———, and H. I. YOO, "Risk Attitudes, Sample Selection and Attrition in a Longitudinal Field Experiment," *Review of Economics and Statistics* 102 (2020), 552–68.

———, and J. T. SWARTHOUT, "Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption? Yes," CEAR Working Paper 2011-14, Center for the Economic Analysis of Risk, Georgia State University, 2011.

HECKMAN, J. J., "Sample Selection Bias as a Specification Error," *Econometrica* 47 (1979), 153–62.

JANSSENS, W., B. KRAMER, and L. SWART, "Be Patient When Measuring Hyperbolic Discounting: Stationarity, Time Consistency and Time Invariance in a Field Experiment," *Journal of Development Economics* 126 (2017), 77–90.

JOHNSON, N. L., "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika* 36 (1949), 149–76.

KIRBY, K. N., "One-Year Temporal Stability of Delay-Discount Rates," *Psychonomic Bulletin & Review* 16 (2009), 457–62.

———, R. GODOY, V. REYES-GARCÍA, E. BYRON, L. APAZA, W. LEONARD, E. PÉREZ, V. VADEZ, and D. WILKIE, "Correlates of Delay-Discount Rates: Evidence from Tsimane' Amerindians of the Bolivian Rain Forest," *Journal of Economic Psychology* 23 (2002), 291–316.

KRINSKY, I., and A. L. ROBB, "On Approximating the Statistical Properties of Elasticities," *Review of Economics and Statistics* 68 (1986), 715–19.

LAIBSON, D., "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 112 (1997), 443–77.

LAU, M. I., and H. I. YOO, "A General Approach to Estimating Random Preference Models of Risk Attitudes," CEAR Working Paper 2020-18, Center for the Economic Analysis of Risk, Georgia State University, 2023.

LAURY, S. K., M. M. MCINNES, and J. T. SWARTHOUT, "Avoiding the Curves: Direct Elicitation of Time Preferences," *Journal of Risk & Uncertainty* 44 (2012), 181–217.

MEIER, S., and C. D. SPRENGER, "Temporal Stability of Time Preferences," *Review of Economics and Statistics* 97 (2015), 273–86.

O'DONOGHUE, T., and M. RABIN, "Present Bias: Lessons Learned and To Be Learned," *American Economic Review (Papers & Proceedings)* 105 (2015), 273–79.

PHELPS, E. S., and R. A. POLLAK, "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies* 35 (1968), 185–99.

QUIGGIN, J., "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization* 3 (1982), 323–43.

Sadoff, S., A. Samek, and C. Sprenger, "Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts," *Review of Economic Studies* 87 (2020), 1954–88.

Takeuchi, K. "Non-parametric test of time consistency: Present bias and future bias," *Games and Economic Behavior* 71 (2011), 456–78.

Tanaka, T., C. F. Camerer, and Q. Nguyen, "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *American Economic Review* 100 (2010), 557–71.

Wilcox, N. T., "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Stochastic Modeling Primer and Econometric Comparison," in J.C. Cox and G.W. Harrison, eds., *Risk Aversion in Experiments*, Research in Experimental Economics, Volume 12 (Bingley, UK: Emerald, 2008).

———, "'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice under Risk," *Journal of Econometrics* 162 (2011), 89–104.

Wölbert, E., and A. Riedl, "Measuring Time and Risk Preferences: Reliability, Stability, Domain Specificity," CESifo Working Paper No. 4339, 2014.

Yoon, H., "Impatience and Time Inconsistency in Discounting Models," *Management Science* 66 (2020), 5850–60.