Task completion without commitment

David J. Freeman and Kevin Laughren*

March 2, 2022

Abstract

We conduct an experiment where participants make choices between completing a task now or waiting to complete it in the future. We vary the future dates at which the task can be completed and the amount of effort required to complete it at each date. We infer both a participant's preferences over when to complete the task and their expectations about how their future preferences differ from their current preferences. We find high rates of time consistency, driven by participants' strong tendency to complete tasks immediately – even when it requires exerting more effort than waiting.

March 2022

Keywords: task completion, present bias, time inconsistency, procrastination, preproperation.

*Freeman: Department of Economics, Simon Fraser University, david_freeman@sfu.ca. Laughren: Smith School of Business, Queen's University, kevinlaughren@gmail.com. This research was funded by SSHRC IDG 430-2016-00193 and conducted under SFU REB certificate #20160245.

Many economic decisions involve a trade-off between current and future benefits and costs: how much to consume versus save, to exercise or not, and whether to complete an onerous task today or delay. In such situations a person makes choices at multiple points of time and cannot commit their future choices. However, most experiments that study intertemporal decision-making study choices made at a single point in time over delayed monetary rewards (e.g. Coller & Williams 1999; Harrison *et al.* 2002; Andreoni & Sprenger 2012). Such experiments are uninformative of three crucial axioms of intertemporal choice.

One normative axiom of intertemporal choice that cannot be tested with choices at a single point in time is *time consistency*: if a person makes an initial choice between two options, that person should not wish to revise their initial choice at any later time if the consequences of those actions do not change. This introduces a second axiom of intertemporal choice: a person correctly forecasts their future choices, an axiom known as *sophistication*. A third axiom of intertemporal choice is *time invariance* (Halevy, 2015): if a person chooses an option over another today, they would make the same choice between these options tomorrow if the consequences of each action were shifted one day in the future. Little existing experimental work addresses sophistication and time invariance.

We introduce a new experimental design that observes participant task-completion decisions for real-effort tasks to measure violations of time consistency, sophistication, and time invariance at the individual level. A participant faces multiple different choice sets that list two or three days and a number of chores associated with each day. For each choice that includes the current day, the participant must indicate their choice to either do the task "today" or "not today". If a participant chooses "today", they must complete the specified number of real-effort chores by the end of today. If they choose "not today", tomorrow they will face all schedules for which they previously selected "not today" plus those schedules for which "today" was not previously available. Crucially, their initial choices cannot commit their future choices except by completing the task. Thus, we can elicit each participant's preference to complete or not complete the task each day for each effort schedule that the participant faces. We vary the effort required on each available day across schedules as well as the days available to complete the task, which allows us to identify violations of time consistency, sophistication about time inconsistency, and time invariance.

Our participants exhibit a high degree of time consistency in spite of the considerable power to detect violations with the experimental design. We find that 50 of 82 participants are time consistent in every test, and across all participants 84% of tests satisfy time consistency. We find that this time consistency arises from a strong tendency to choose to complete a task immediately, even when delaying would have reduced the number of required chores. We find that 29 of the 50 time consistent participants always choose to complete the task on day 1 when it is available. Overall, 70% of two-date choices are resolved in favour of completing the task immediately. including half of all two-date choices in which delaying reduces the number of chores that must be completed. In two-date choices, a participant's beliefs about their own future behavior are trivial – and thus the immediate completion tendency we document for real effort tasks does not arise from sophistication about inconsistent preferences.

These findings are broadly inconsistent with the dominant intuition that people tend to prefer to delay unpleasant tasks (e.g. O'Donoghue & Rabin (1999)) and thus contradicts the common modeling assumption that bads like aversive and effortful chores[1] entail negative flow payoffs that are discounted if they occur in the future. Our findings thus pose a challenge to the literature on intertemporal choice. Our experiment was designed to be similar to existing real-effort experiments that study intertemporal trade-offs (Augenblick *et al.* 2015; Augenblick & Rabin 2019). Our design features are primarily intended to distinguish sophistication from naivete about present bias – and should have no bearing on the presence of present bias. Yet we do not find the choice patterns that similar work attributes to present bias – we find an opposing pattern of behavior. This stark finding stands apart from most of the experimental economics literature (although we note some exceptions, including Aycinena *et al.* 2020). However, a recent literature in psychology finds evidence of "precrastination": in experimental bucket-carrying tasks, participants exhibit a tendency to start the task too soon even when it increases the total amount of effort required (Rosenbaum *et al.* 2014). Rosenbaum et al. attribute this to a desire to "reduce working memory loads". Haushofer (2015) formalizes this idea through a "cost of keeping track" and shows it can explain diverse evidence on intertemporal choice through a completely different mechanism than present-biased discounting. Our findings suggest this may be a more important determinant of intertemporal choice in standard economic decisions than previously recognized.

[1] We both tested out our real effort tasks and found them extremely aversive.

Related Literature There is a long experimental literature on intertemporal choice that studies preferences over delayed monetary rewards revealed at one point in time (e.g. Coller & Williams 1999; Harrison *et al.* 2002). Since money can be saved and borrowed, in principle, such experiments should not reveal intertemporal preferences if participants broadly bracket their experimental choices with opportunities outside of the lab (Cubitt & Read 2007). Thus, some intertemporal choice experiments use less-fungible rewards like snacks that will be consumed immediately (Read & Van Leeuwen 1998) and real-effort tasks (Augenblick *et al.* (2015); Augenblick & Rabin (2019)). Our design uses a real-effort task from Augenblick *et al.* (2015).

Much of the literature on intertemporal choice only studies choices made at one point in time and thus cannot directly test time consistency or time invariance – with some exceptions. Read & Van Leeuwen (1998) have their participants choose a post-lunch snack both a week in advance, and then ask them to choose again the day they consume the snacks – and find that participants choose unhealthy snacks more frequently when choosing same-day than in-advance. Augenblick *et al.* (2015) study participants' allocations of real-effort chores between earlier and later dates made both when the earlier date is in the future and when it is in the present – and find that participants tend to delay effortful chores, exhibiting present bias. Augenblick & Rabin (2019) elicit participant preferences over quantities of delayed real-effort chores at different points in time and their beliefs about their future such preferences. These two real-effort experiments both find that participants tend to prefer to delay effortful chores, and exhibit present bias by exhibiting a disproportionate preference to delay immediate effort. Halevy (2015) uses a design in which participants report their preferences between smaller-sooner versus larger-later monetary payments in successive weeks, and tests time invariance and time consistency of preferences for delayed monetary rewards. He finds that over half of participants are time consistent, and roughly half of all participants satisfy time invariance. Compared to Halevy, we study choices involving real-effort tasks – for which previous work has demonstrated higher rates of present bias than for delayed monetary payments (Augenblick *et al.*, 2015).

Existing papers test for a person's sophistication or naivete about their own time inconsistency by measuring demand for commitment devices (e.g. Ashraf *et al.* 2006; Augenblick *et al.* 2015; see a review in Bryan *et al.* 2010) or comparing elicited beliefs to actual future choices (Augenblick & Rabin 2019). In contrast, our design elicits choices at different points of time in an environment in which a participant cannot commit their future choices, which allows us to employ Freeman's (2021) approach to test both sophistication and naivete about time inconsistency for each participant. Much of our design is motivated by a literature that commonly finds evidence of time-inconsistent present bias and that has found widely different degrees of sophistication about it (Ashraf *et al.* 2006; Augenblick *et al.* 2015; Augenblick & Rabin 2019). Yet unlike this literature, we find little evidence of time-inconsistency – and thus have little to say about sophistication and naivete.

After finding that participants frequently prefer to complete the task today even when delaying would reduce the number of chores required, we discovered a conceptual precedent for our findings in the psychology literature (Rosenbaum *et al.* 2014; Wasserman 2019). Rosenbaum et al. find that when faced with the choice of two buckets to carry to a fixed endpoint, participants tend to pick up whichever is closer to them – even if it entails a longer carrying distance. Fournier *et al.* (2019a) find evidence that suggests this effect arises from a desire to start a task sooner, rather than a desire to complete a subgoal sooner. Fournier *et al.* (2019b) show these effects are mediated by working memory – suggesting that precrastination may at least in part arise from a desire to reduce working memory loads. Within economics, existing work has shown that formal models that incorporate memory can explain facets of intertemporal choice that are difficult to explain as aspects of intertemporal preferences alone (Ericson 2017; Haushofer 2015).

1 Theoretical Framework

We design our experiment to study how future options to complete a task affect participant decisions to complete the task. Consider a participant who must complete a real-effort task exactly once. When first confronted with the task, they are informed of the two or three different days on which they can do the task and how much effort they must exert to complete it on each day. On each day until the last day with an option available, they can either complete the task or not, but they cannot commit their future behaviour except by completing the task.

Each *effort schedule* can be represented as a vector of effort requirements, one for each possible date. We write $(e_1, e_2, e_3, \emptyset)$ to denote the effort schedule in which e_t is the effort required to complete the task on day t, and the task cannot be completed at

$t = 4$. We consider effort schedules with three consecutive completion options of the form $(e_1, e_2, e_3, \emptyset)$ and $(\emptyset, e_2, e_3, e_4)$, as well as effort schedules with that are derived from them by removing one option.[2] Let c denote a *completion function* that returns the time, from among those available, at which the person would complete the task given an effort schedule. That is, $t = c(e_1, e_2, e_3, \emptyset)$ denotes that the person would complete the task at time t if they faced $(e_1, e_2, e_3, \emptyset)$, where t must be either 1, 2, or 3 in this case.

A participant faces a trade-off between a desire to not exert effort today and a desire to avoid effort later, and must forecast their future choices to assess this trade-off. Specifically, when making a today or not-today decision at time $t = 1$ when facing the effort schedule $(e_1, e_2, e_3, \emptyset)$, the participant must consider their intertemporal preferences to make trade-offs between earlier versus later effort. They must also consider how they would behave in the future should they wait today, since they cannot commit their future behavior. Freeman (2021) shows that if a participant is time inconsistent, their choices in $(e_1, e_2, e_3, \emptyset)$ can reveal their sophistication or naivete about their time inconsistency if their actions change when removing one option.

Consider a participant who delays at $t = 1$ when facing the effort schedule $(e_1, e_2, e_3, \emptyset)$ and ultimately completes the task at $t = 3$, that is, $3 = c(e_1, e_2, e_3, \emptyset)$. If that participant would complete the task at $t = 1$ when facing $(e_1, \emptyset, e_3, \emptyset)$, that is, $1 = c(e_1, \emptyset, e_3, \emptyset)$, then the participant exhibits what Freeman calls a *doing-it-later reversal*, and reveals themself naive. This is because their choice to do it at $t = 1$ when facing $(e_1, \emptyset, e_3, \emptyset)$ reveals that at $t = 1$, they prefer completing it at $t = 1$ to waiting until $t = 3$. Thus, they would only initially delay when facing $(e_1, e_2, e_3, \emptyset)$ and then complete it at $t = 3$ if they were to incorrectly believe that they would instead complete it at $t = 2$. By the same logic, they would also exhibit a doing-it-later reversal and reveal themself naive if instead $2 = c(e_1, e_2, e_3, \emptyset)$ and $1 = c(e_1, e_2, \emptyset, \emptyset)$.

If instead they complete the task at $t = 1$ when facing $(e_1, e_2, e_3, \emptyset)$, that is, $1 = c(e_1, e_2, e_3, \emptyset)$, but also would delay at $t = 1$ when facing $(e_1, e_2, \emptyset, \emptyset)$, that is, $2 = c(e_1, e_2, \emptyset, \emptyset)$, they exhibit a *doing-it-earlier reversal* and reveal themself sophisticated. This is because their choice to delay at $t = 1$ when facing $(e_1, e_2, \emptyset, \emptyset)$

[2] We note that a statement about effort schedules derived from $(e_1, e_2, e_3, \emptyset)$ will apply to analogous statements about effort schedules derived from $(\emptyset, e_2', e_3', e_4')$ when $e_1 = e_2'$, $e_2 = e_3'$, and $e_3 = e_4'$ by shifting all references to days by one.

reveals that at $t = 1$, they would rather wait to complete the task at $t = 2$. Thus they would only complete the task at $t = 1$ when facing $(e_1, e_2, e_3, \emptyset)$ if they expect that they would not complete it at $t = 2$ and would instead complete it at the currently-less-preferred time $t = 3$. That is, completing the task at $t = 1$ reveals that they expect their future choice to be inconsistent with their current preferences. Similar logic applies if instead $1 = c(e_1, e_2, e_3, \emptyset)$ and $3 = c(e_1, \emptyset, e_3, \emptyset)$: this would be a doing-it-earlier reversal and reveal sophistication about future inconsistency.

Fixing the three-option schedule $(e_1, e_2, e_3, \emptyset)$, we wish to observe how the participant would choose from each possible pair of completion options in that schedule as well as in the original schedule. We refer to each such quadruple of choice observations as a *quad*. Observing completion times for all effort schedules in a quad allows us the possibility to observe time inconsistency and a doing-it-later or doing-it-earlier reversal. To observe whether a participant would exhibit a possible doing-it-earlier or a doing-it-later reversal involving $(e_1, e_2, e_3, \emptyset)$, we need to also observe how they would behave in $(e_1, e_2, \emptyset, \emptyset)$ and $(e_1, \emptyset, e_3, \emptyset)$. By also observing a participant's choice from $(\emptyset, e_2, e_3, \emptyset)$, our experiment can reveal their time inconsistency directly. A participant reveals themself time inconsistent if either (i) $2 = c(e_1, e_2, \emptyset, \emptyset)$, $1 = c(e_1, \emptyset, e_3, \emptyset)$, and $3 = c(\emptyset, e_2, e_3, \emptyset)$, or (ii) $1 = c(e_1, e_2, \emptyset, \emptyset)$, $3 = c(e_1, \emptyset, e_3, \emptyset)$, and $2 = c(\emptyset, e_2, e_3, \emptyset)$.

Standard models of intertemporal choice also restrict how choices vary across quads. Economic models of intertemporal choice assume that choices between pairs of completion opportunities are *monotonic* in effort: a person has a greater tendency to complete the task on a day if the required effort that day is weakly less, and the amount of effort required to complete the task later is weakly more. For example, if $e'_1 \leq e_1$, $e'_2 \geq e_2$, and $1 = c(e_1, e_2, \emptyset, \emptyset)$, then monotonicity requires $1 = c(e'_1, e'_2, \emptyset, \emptyset)$. Economic models also typically make a less fundamental assumption about preferences known as time invariance: if all effort requirements are shifted by one day, then the completion time also shifts by one day. For example, if $e_1 = e'_2$ and $e_2 = e'_3$, then time invariance requires that $1 = c(e_1, e_2, \emptyset, \emptyset)$ if and only if $2 = (\emptyset, e'_2, e'_3, \emptyset)$. Our design allows us to test both of these assumptions.

2 Experiment Design

We design a real-effort experiment to obtain data on participants' task-completion decisions. After an initial orientation session, our experiment is four days long and took place Monday-Thursday ($t = 1, 2, 3, 4$ respectively). On each day of the experiment a participant was required to sign-in to our online interface and make relevant decisions. On one day that depends on their choices, a participant must complete several units of our real-effort task. We refer to each unit of our real-effort task as a *chore*. Each chore requires the participant to transcribe 40 blurry Greek characters by clicking them from their screen; this real effort task was adapted from Augenblick *et al.* (2015). Participants were also required to complete one mandatory chore on each day of the experiment after making any required decisions. All participants who completed the entire experiment were paid $25 by electronic transfer on Sunday.

To observe a participant's decisions in multiple different effort schedules, we employ a variation on the random incentive system, providing incentive for a participant to report their true preferences of whether to work or not on each day. On the first day of the experiment, the participant is presented with all effort schedules for the experiment and is informed that one of these has been randomly chosen and will be implemented – the *schedule that counts*. The participant then must choose to complete chores "today" or "not today" for every effort schedule in which a $t = 1$ option is available. If they choose "today" in the schedule that counts, then they complete the required number of extra chores today. Otherwise, when they log in the next day, they face a "today" or "not today" decision for those effort schedules with two or more options remaining.

Table 2 displays the specific effort values participants saw in each version of the experiment. In all cases, we study quads with completion options on two or three out of either $1, 2, 3$ or $2, 3, 4$; the latter effort schedules are obtained by shifting the former by one day.[3] Shifting quads by one day also enables us to test time invariance. We observe each participant make up to 8 choices over the values in one effort profile: one quad with opportunities to work on $1, 2$, and 3 and one quad with options on $2, 3$, and 4. Some quads are partially censored when participants complete their extra chores after making $t = 1$ or $t = 2$ choices.

A choice quad takes one of 24 possible values. We identify four categories of choice using these values: time consistent, reversal, non-Strotzian, and censored. The categorization of each choice quad is displayed in Table 1. In four of the quads there are two choices which are endogenously censored from observation and we do not attempt

[3]The presence of the latter quads, combined with the random incentive system and $(\emptyset, e_2, e_3, \emptyset)$ choice sets, implies that we have at least a $\frac{5}{8}$ chance of observing $t = 2$ choices.

to classify those quads. Only 7 of the 24 possible values of quads can be rationalized by time consistent preferences, but the majority of our observations are time consistent. Choice quads which are not time consistent either represent a preference reversal as defined above, or can be categorized as non-Strotzian. Preference reversals can be either doing-it-earlier (identifying sophistication) or doing-it-later (identifying naivete). Non-Strotzian choice quads cannot be rationalized by a single $t = 1$ utility function over when to complete the task, and are divided into those which identify a preference for flexibility versus those that identify a preference for commitment, with neither type being generated by a perception-perfect strategy (Freeman, 2021).

Implementation We recruited 101 participants from the SFU Experimental Economics Laboratory Research Participation System. Experiments were conducted entirely online, with a live introductory Zoom session followed by four days of asynchronous participation at any time of the participants' choosing. Participants were paid a $7 (CAD) show up fee for the introductory session and an all-or-nothing completion payment of $25, both paid by email transfer on the Sunday following the final experiment deadline (Thursday at 11:59pm). The completion payment requires the participant to sign in and complete at least one chore on four consecutive days, complete extra chores on one of those days, and make task-completion decisions when available. Participants were recruited to attend a 30 minute introductory session, held on Fridays via Zoom. Informed consent was collected, instructions were read aloud by an experimenter, and questions were taken via confidential chat. After questions were answered, participants were asked to demonstrate that they were able to sign-in to the online experiment interface and complete one chore, with technical support provided by the experimenter until all participants were successful. Participants were sent a reminder email on Monday morning with a link to the experiment. Participants were required to sign in to the experiment through the University's centralized authentication service. Upon signing in Monday, participants chose their own email reminder time for the daily reminders for the remainder of the experiment.

101 participants completed the online introduction and received the show up fee. 82 participants completed all of the experiment requirements and received full payment by email transfer. The remaining 19 participants missed a day of participation or chose not to complete their chores on one or more day, so they received only the

$c(e_1, e_2, \emptyset, \emptyset)$	$c(e_1, \emptyset, e_3, \emptyset)$	$c(\emptyset, e_2, e_3, \emptyset)$	$c(e_1, e_2, e_3, \emptyset)$	Quad Type	Preference
1	1	$\emptyset$	1	Time Consistent	t = 1
1	1	2	1		t = 1
1	1	3	1		t = 1
1	3	3	3		t = 3
2	1	2	2		t = 2
2	3	2	2		t = 2
2	3	3	3		t = 3
1	3	$\emptyset$	1	Reversal	earlier
1	3	2	1		earlier
1	3	3	1		earlier
2	1	$\emptyset$	1		earlier
2	1	2	1		earlier
2	1	3	1		earlier
1	3	2	2		later
2	1	3	3		later
1	1	2	2	Non-Strotzian	flexibility
1	1	3	3		flexibility
2	3	$\emptyset$	1		commitment
2	3	2	1		commitment
2	3	3	1		commitment
1	1	$\emptyset$	$\emptyset$	Censored	censored
1	3	$\emptyset$	$\emptyset$		censored
2	1	$\emptyset$	$\emptyset$		censored
2	3	$\emptyset$	$\emptyset$		censored

Table 1: Identification using observable choice quads

The above table extends to quads derived from $(\emptyset, e_2, e_3, e_4)$ by adding 1 to every integer in the table and shifting all efforts and $\emptyset$s in the header one position to the right.

Effort Profile	# Participants Observed	# Quads Observed	Versions
14, 20, 28	*45*	*76*	V1, V2
16, 20, 25	*82*	*144*	All
18, 20, 22	*82*	*144*	All
19, 20, 21	*22*	*37*	V2
20, 20, 20	*82*	*144*	All
22, 20, 18	*37*	*68*	V3
25, 20, 16	*37*	*68*	V3
	82	*681*	

Table 2: Experiment Effort Profiles.
Each triple describes the number of chores required if working on day 1, day 2, or day 3 (or for working on day 2, day 3, or day 4).

show-up fee and are excluded from analysis.

The baseline number of chores (20) and length of chore (40 characters) were chosen so the session would require less than one hour of a participant's time over the four days to complete all chores and make all decisions required for the $25 completion pay. A 40-character chore requires 40 button clicks with 100% accuracy; we required 30-45 seconds to complete each chore on our earliest attempts. It is difficult to estimate an average hourly wage because participants had until 11:59pm each day to complete all decisions and chores and could take breaks before this deadline at no cost. This results in a wide range of observed time spent on extra chores (00:12:37 - 13:38:12), however the median participant completes 20 extra chores in under 25 minutes.

We varied the effort schedules faced across three versions of the experiment (Table 2). In each version, the relevant effort profiles listed on the left-hand column of Table 2 were used to form two quads, one quad with the listed efforts faced on $t = 1, 2, 3$ and the other quad with the listed efforts faced on $t = 2, 3, 4$. In the first two versions of the experiment we used quads designed to have power to detect a participant's present bias and their sophistication or naivete about said bias. For our third version, we also included control quads designed to test whether some participants exhibit a negative discount rate by choosing to exert more effort and complete the task at an earlier date.

We conducted Version 1 starting on July 20, 2020 with 23 participants. After observing many "today" choices in Version 1, we added an additional schedule in Version 2 to allow us to detect even small degrees of present bias and conducted that session starting July 27, 2020, with 22 participants. Still observing many "today" choices, we changed the set of effort schedules for Version 3 to enable us to detect whether participants would make such choices if doing so increased the number of chores required, which would indicate an opposing preference to those generated by discounting and present bias. We conducted Version 3 with 15 participants starting March 8, 2021 and with 22 participants starting March 29, 2021.

Data Censoring We do not always observe two full choice quads from each effort schedule because the day on which a participant completes their extra chores is endogenous. When a participant completes their payoff-relevant extra chores at $t = 1$ (Monday), they make no further task-completion decisions. In these cases, we obtain no data for $t = 2, 3, 4$ effort schedules nor do we obtain $t = 2$ decisions from $t = 1, 2, 3$ effort schedules. This partial censorship also occurs on $t = 2, 3, 4$ effort schedules when extra chores are completed at $t = 2$.

This endogenous censoring is inherent when studying any incentivized when-to-do-it choices. However, our design has a $1/2$ probability that a $t = 2, 3, 4$ schedule is the schedule that counts, and a $1/8$ probability that a $t = 1, 2, 3$ schedule with no option to do it on Monday is the schedule that counts. This design results in a $5/8$ exogenous probability that a participant makes payoff relevant choices on at least two days. Our software randomly assigned 52 of 82 participants such an effort schedule, which we refer to as the *non-endogenous subsample.* We highlight this subsample when discussing results which could be subject to endogeneity. The remaining 30 participants generate data that is subject to endogenous censoring, including 5 participants who (endogenously) generate only censored quads.

3 Results

Recall that a quad represents four choice problems from one effort profile. We start by categorizing a participant's choice quads into time consistent, doing-it-earlier reversal, doing-it-later reversal, or Non-Strotzian (Table 1).

RESULT 1: Overall, choices in 84% of uncensored quads are time consistent. At the individual level, 50 of 82 participants are time consistent in all of their uncensored quads.

When all quads are considered regardless of censoring or endogeneity, 500 of 681 observations are Time Consistent. After removing the 84 censored observations, 500

Effort Profile	Time Consistent			Reversal		Non-	Censored	Total Quads
	1st day	2nd day	3rd day	earlier	later	Strotz		Observed
14, 20, 28	72%	3%	7%	4%	3%	7%	5%	76
16, 20, 25	69%	2%	6%	8%	1%	3%	10%	144
18, 20, 22	65%	4%	5%	8%	1%	7%	11%	144
19, 20, 21	65%	5%	5%	5%	3%	5%	11%	37
20, 20, 20	53%	6%	8%	8%	1%	7%	17%	144
22, 20, 18	37%	7%	24%	7%	0%	9%	16%	68
25, 20, 16	38%	7%	26%	10%	0%	3%	15%	68
TOTAL uncensored	67%	5%	12%	9%	1%	7%		597
TOTAL	59%	5%	10%	8%	1%	6%	12%	681
RANDOM uncensored	14%	11%	11%	28%	11%	25%		
RANDOM	13%	10%	10%	25%	10%	23%	9%	

Table 3: Classifying Observed Choice Quads by Effort Schedule (All data)
RANDOM is the expected proportion if all completion decisions are random and independent.
"1st day" is $t = 1$ in $(e_1, e_2, e_3, \emptyset)$ effort schedules and $t = 2$ in $(\emptyset, e_2, e_3, e_4)$ effort schedules.

of 597 (84%) uncensored quads are time consistent, 399 (67%) of which exhibit a consistent preference to complete the extra chores on day 1. Table 3 provides the classifications by effort schedule and remarkably over two-thirds (67%) of all quads are time consistent in each effort profile; the level or interest rate on effort does not appear to affect overall rates of time consistency.[4]Among the time inconsistent quads, there are similar quantities of Reversals and Non-Strotzian observations (both in the high single digits). Comparatively if choice data were generated randomly by independently mixing at each choice, 36% of uncensored choice profiles would be time consistent, and 39% would exhibit a reversal.

The full set of data in Table 3 are subject to endogenous sampling and censoring. Table 4 displays results for the non-endogenous subsample, and further drops the quads derived from $(\emptyset, e_2, e_3, e_4)$ effort schedules since they are subject to endogenous observation of choices at $t = 3$ (Wednesday). These data have zero censored observations by construction, yet still exhibit a very consistent mix of decisions to the "Total Uncensored" data from Table 3.

Within the non-endogenous subsample 28 of 52 participants are time consistent

[4]There is no statistical difference in the rate of time consistency between the schedule with the highest rate of time consistency (16,20,25) and the lowest rate of time consistency (20,20,20) using a Fisher exact test (p=0.24). All participants completed these two schedules, regardless of which version they faced.

Effort Profile	Time Consistent			Reversal		Non-Strotz	Censored	Quads
	1st day	2nd day	3rd day	earlier	later			Observed
14, 20, 28	79%	4%	0%	4%	4%	8%	0%	24
16, 20, 25	75%	2%	6%	12%	0%	6%	0%	52
18, 20, 22	73%	4%	4%	10%	2%	8%	0%	52
19, 20, 21	75%	0%	0%	8%	8%	8%	0%	12
20, 20, 20	60%	10%	2%	12%	4%	13%	0%	52
22, 20, 18	36%	14%	29%	4%	0%	18%	0%	28
25, 20, 16	32%	14%	36%	14%	0%	4%	0%	28
TOTAL	63%	7%	10%	10%	2%	9%		248

Table 4: Classifying Observed Choice Quads by Effort Schedule (Non-endogenous subsample)

This table only uses only the $(e_1, e_2, e_3, \emptyset)$ effort schedules, and only the 52 participants whose randomly assigned schedule that counts does not include e_1

in 100% of observed choices. 18 of these 28 time consistent participants exclusively generated quads with a time consistent preference to complete the chores on day 1, even on those schedules for which waiting on day 1 would entail completing fewer total chores. Since we observe this subsample make choices on at least two days, all of these tests of time consistency are non-trivial. All remaining tables in the main text of results include only this non-endogenous subsample – though a comparison of Tables 3 and 4 suggests that data censoring does not appear drive our results on time consistency.

The remaining 30 of 82 participants were randomly assigned a schedule that counts which allowed them to complete their extra chores on Monday, and this subsample is subject to endogenous selection. On Monday, the participants in this set who chose to work "today" for their schedule that counts do not make any more decisions at future dates. Over half (17/30) of these participants exclusively generate quads that are time consistent, and another 5 of the 30 only generate censored observations, and thus satisfy time consistency trivially.[5] The remaining 8 of these 30 endogenous participants generated at least one reversal or Non-Strotzian quad.

Time consistency is tested within a single quad of choices, but an additional consideration is whether a participant's full set of choices are collectively sensible

[5]For example, suppose on Monday we observe "today" for $(e_1, e_2, \emptyset, \emptyset)$, "not today" for $(e_1, \emptyset, e_3, \emptyset)$, and "not today" for $(e_1, e_2, e_3, \emptyset)$. If $(e_1, e_2, \emptyset, \emptyset)$ is the choice that counts, the participant completes extra chores Monday. Thus $c(e_1, e_2, e_3, \emptyset)$ and $c(\emptyset, e_2, e_3, \emptyset)$ are never observed, and the quad is censored.Table 1 shows the categorization of all possible quad observations.

across effort schedules.

RESULT 2: 90% of participants who are time consistent *within* every quad also demonstrate monotonicity *across* all quads. Overall, fewer than 5% of observations need to be dropped to make every participant consistent with monotonicity.

Monotonicity links preferences across effort values and requires participants to consistently prefer exerting less effort while controlling for time. We evaluate whether a subject violates monotonicity, considering every binary choice in all quads in the experiment.

We count the total number of monotonicity violations for each participant, and 58 of 82 participants (71%) demonstrate no violations of monotonicity in their dataset. Of the 50 participants who were time consistent in 100% of their classified quads, only 5 made a choice violating monotonicity, thus 45 of 82 participants were both time consistent and monotonic.

For those participants who do violate monotonicity, we use the Houtman-Maks Index (HMI) to represent the maximal proportion of data which can be collectively consistent with monotonicity (Houtman & Maks, 1985; Heufer & Hjertstrand, 2015; Demuynck & Hjertstrand, 2019). This involves a simple linear optimization for each participant, minimizing the number of observations removed subject to the constraint that there are zero monotonicity violations in the remaining dataset. In total, 76 of 1726 observations are removed for a weighted mean HMI of 0.955, and the mean HMI among those with at least one violation is 0.86. The distribution of HMI by participant in Figure 1 further demonstrates that monotonicity violations are rare and concentrated in a minority of individuals.

RESULT 3: Choices show a strong immediate completion tendency.

This pattern is evident whether the 1st day is $t = 1$ (as in Table 4) or is $t = 2$ (see Table 8 in the Appendix), as both show over 60% of quads are time consistent with a preference for day 1; this suggests the pattern is not simply a day of the week preference. Even for quads based around the effort profiles (22, 20, 18) and (25, 20, 16) the consistent preference to work today is the modal choice profile, even though waiting would require less total work for the same payment. Table 5 shows that over 75% of the individual binary choices are choices to work "today", including approximately half of binary choices from the decreasing effort profiles (22, 20, 18) and (25, 20, 16). Given the median participant required 1.2 minutes per chore, this implies

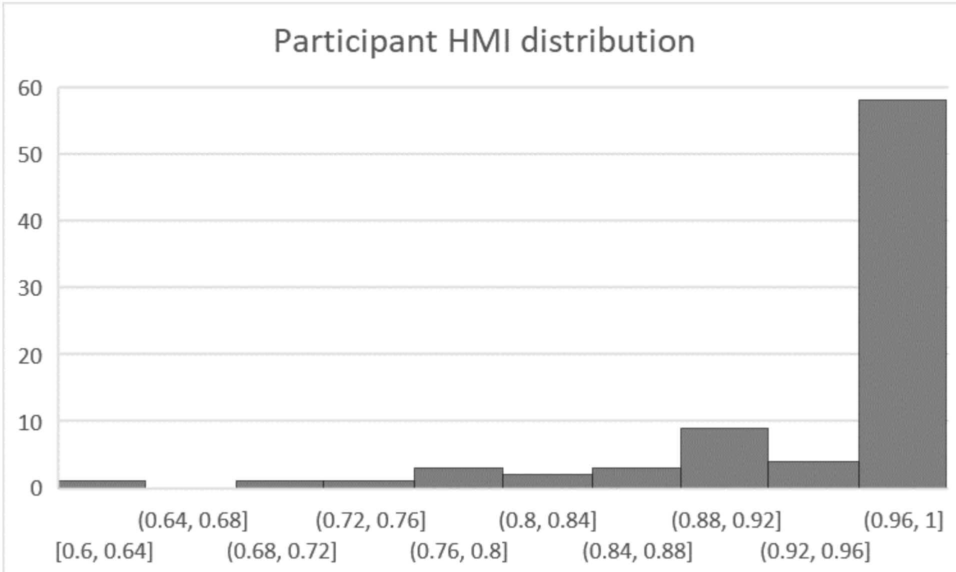

Figure 1: Participant Houtman-Maks Index - Monotonicity

a willingness to exert an extra 6 minutes of effort to complete the extra chores early. The median participant completes 24 total chores (20 extra chores, plus a minimum of one each day) implying this preference results in a 20% increase in effort for fixed pay. Remarkably, 18 of 28 participants in the non-endogenous subsample who were time consistent in every effort schedule chose to work on day 1 in every choice it was available.

When the same effort tradeoff is observed on different days, a participant who makes a different today/not today choice has violated time invariance. A violation of time invariance could suggest an unobserved preference to complete the task on a specific day.

RESULT 4: 79% of comparable decision pairs satisfy time invariance.

Time invariance requires us to compare individual choices in $t = 1, 2, 3$ quads to their analogous choice in $t = 2, 3, 4$ quads. Restricting attention to the non-endogenous subsample of participants who were predetermined to make choices twice, there are 725 total possible tests of time invariance.[6] Time invariance is satisfied in

[6]We conduct three tests per effort profile, including the tripleton schedule; however we exclude $(\emptyset, e_2, e_3, \emptyset)$ to $(\emptyset, \emptyset, e'_3, e'_4)$ choices since observations of the latter require observing choices on $t = 3$, which is subject to endogenous selection. 19 of 248 possible tests comparing $(e_1, e_2, e_3, \emptyset)$ to $(\emptyset, e'_2, e'_3, e'_4)$ are excluded because the participant chooses "not today" at both $t = 1$ and $t = 2$, but the final choice between $(e'_3, e'_4)$is (endogenously) not revealed.

Proportion choosing to work Today

Effort profile	$(e_1, e_2, \emptyset, \emptyset)$	$(e_1, \emptyset, e_3, \emptyset)$	$(\emptyset, e_2, e_3, \emptyset)$
14, 20, 28	96%	88%	88%
16, 20, 25	90%	85%	90%
18, 20, 22	81%	88%	88%
19, 20, 21	92%	92%	83%
20, 20, 20	77%	77%	85%
22, 20, 18	57%	50%	61%
25, 20, 16	46%	43%	57%
TOTAL	77%	76%	81%

Table 5: Proportion choosing to work Today when a choice is available (Non-endogenous subsample)

1st Day Choice, 2nd Day Choice

Choice set	Time Invariant	Today, Not Today	Not Today, Today
$c(e_1, e_2, \emptyset, \emptyset)$	79%	8%	14%
$c(e_1, \emptyset, e_3, \emptyset)$	78%	8%	13%
$c(e_1, e_2, e_3, \emptyset)$	80%	8%	12%
TOTAL	79%	8%	13%

Table 6: Time Invariance Violations by Choice Set (Non-endogenous subsample)

79% of tests, and 23 of 52 participants in the non-endogenous subsample satisfy time invariance in every possible test.

Table 6 displays the proportion of pairwise choices that violate time invariance when we observe a participant make choices from comparable effort schedules on two different days. The relative scarcity of violations of time invariance suggest specific day-of-the-week preferences are driving choice in at most 21% of tests. The number of chores and interest rate on effort do not appear to systematically affect the rate of failure of time invariance across schedules (Appendix Table 9).

The small response to a negative interest rate on effort apparent in Table 4 indicates that choices are not well represented by a standard model of intertemporal preferences in which participants discount costly future relative to immediate effort. We conduct a structural estimation to facilitate a comparison of behaviour in our experiment to existing work.

RESULT 5: Structural estimation of a model of quasi-hyperbolic discounting yields $\beta > 1$ – capturing a strong tendency to complete real effort

tasks immediately.

We model the probability of choosing "today" as resulting from a latent utility model. Consider only the binary decisions, and let e_t, e_{t+k} denote the effort requirements at the two available periods, t and $t+k$. Let $Y_t = 1$ denote a "today" choice at t and $Y_t = 0$ denote a "not today" choice. We assume that $Y_t = 1 \iff Y_t^* \geq 0$, where Y_t^* represents the time-t (unobserved) utility difference between choosing "today" and "not today". We specify a structural quasi-hyperbolic discounting model with a linear disutility-of-effort: $Y_t^* = U_t(Y_t = 1, e_t, e_{t+k}) - U(Y_t = 0, e_t, e_{t+k})$ where $U(Y_t = 1, e_t, e_{t+k}) = -\lambda e_t$ and $U(Y_t = 0, e_t, e_{t+k}) = -\beta\delta^k\lambda e_{t+k}$ with β and δ scalar time preference parameters to be estimated.

The net utility of working today can be written as $Y_t^* = -\lambda e_t + \beta\delta\lambda e_{t+k}\mathbb{I}_{\{k=1\}} + \beta\delta^2\lambda e_{t+k}\mathbb{I}_{\{k=2\}}$. We assume there is some variation in individual values of Y_t^* due to individual preference shocks, and specify a logit regression $Y_t^* = x'b + \epsilon_t$, where $x'b = b_0 e_t - b_1 e_{t+k}\mathbb{I}_{\{k=1\}} + b_2 e_{t+k}\mathbb{I}_{\{k=2\}}$and $\epsilon_t \sim \Lambda()$, a standard binary logit model with no intercept term.[7] We recover estimates of (b_0, b_1, b_2) and use them to estimate $\hat{\beta} = \frac{(b_1)^2}{b_0 b_2}$; $\hat{\delta} = \frac{b_2}{b_1}$; and $\hat{\lambda} = -b_0$.[8] We cluster standard errors by participant, and recover asymptotic standard errors for the parameter estimates using the delta method. Parameter estimates and their asymptotic standard errors are in Table 7. We provide the underlying logit regression estimates of (b_0, b_1, b_2) and further details in Appendix Tables 10, 11, 12, and **??**.

Previous studies of intertemporal preference consistently estimate values of $\beta < 1$, with the interpretation being that there is additional (non-geometric) discounting of all future periods relative to the present (Augenblick *et al.* 2015). The participants in this experiment had a clear disposition to complete the extra chores today, and this is reflected in the estimate of $\beta > 1$, as caring more about future utility than today's utility would result in the observed participant disposition to complete the task today (with 580/743 binary choices resulting in the choice to work "today"). The disutility

[7]Forcing the regression to an intercept at zero is equivalent to assuming that $Prob(Y_t = 1|e_{t=0}, e_{t+k} = 0) = 0.5$, which is true in this structural utility model.

[8]We caution against overly interpreting our point estimates. The parameter λ can be viewed as controlling sensitivity to effort or alternatively the degree of stochasticity, and our estimation cannot separate these interpretations. Similarly, notice that $-e_t > -\beta\delta^{t+k}e_{t+k}$ if and only if $-e_t^\gamma > -\beta^\gamma\delta^{\gamma t+\gamma k}e_{t+k}^\gamma$ for every $\gamma > 0$. Our design does not identify the curvature-of-disutility-of-effort parameter γ, so our point estimates of δ and β cannot be directly compared to those in existing work that does attempt to identify γ (e.g. Augenblick *et al.* 2015). However, this does not affect our tests of the null hypotheses (i) $\beta = 1$ and (ii) $\delta = 1$.

		Confidence Intervals ($\alpha = 0.05$)	
Parameter	Estimate (Std. Error)	Lower Bound	Upper Bound
Present Bias β	1.53 (0.22)	1.11	1.96
Discount Factor δ	0.93 (0.050)	0.83	1.03
Disutility of Effort λ	0.14 (0.042)	0.06	0.22
Observations	743		
Clusters	52		

Table 7: Results of Structural Logit Estimation (non-endogenous subsample)
Estimated using binary effort schedules only. Standard errors of logit regression are clustered by individual participant. Asymptotic standard errors estimated using the delta method (derivation in Appendix)

of effort has the expected sign, and geometric discounting which is identified from the difference in choices when delay is $k = 2$ days versus $k = 1$, does not appear to be a significant driver of choice since $\delta \approx 1$.

4 Discussion

Our novel experimental design is ideally-suited to measure a person's sophistication or naivete about their own time inconsistency. Yet we discovered far more time consistency than we expected based on prior work from economics experiments (e.g. Thaler 1981; Read & Van Leeuwen 1998; Ashraf *et al.* 2006), including other experiments that use designs with similar real-effort tasks (Augenblick *et al.* 2015; Augenblick & Rabin 2019). This appears to be driven by a strong tendency to complete tasks immediately – even when this requires additional effort. In our structural model, this leads us to estimate that our subjects tend to have future biased preferences. While some previous studies have found evidence of future bias (e.g. Sayman & Öncüler 2009; Attema *et al.* 2010; Takeuchi 2011; Montiel Olea & Strzalecki 2014), it is the opposite of the present bias found in much prior experimental work and commonly assumed in applications of quasi-hyperbolic discounting (Laibson 1997; O'Donoghue & Rabin 1999). Taken together with prior work, our findings suggest that quasi-hyperbolic discounting with present bias may not be a good descriptive model of task completion. Recent work in psychology that uses a different research design has also found that people tend to exhibit a preference to start tasks sooner (Rosenbaum *et al.* 2014). Our findings may result from a behavioural bias distinct from present bias that merits new theory. Haushofer's (2015) cost of keeping track model is one such approach.

We also explored two other standard properties assumed in most models of intertemporal choice – time invariance and monotonicity – and we do not detect systematic failures of either property. This suggests that these are both descriptively reasonable properties to retain.

References

Andreoni, James, & Sprenger, Charles. 2012. Estimating time preferences from convex budgets. *American Economic Review*, **102**(7), 3333–56.

Ashraf, Nava, Karlan, Dean, & Yin, Wesley. 2006. Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, **121**(2), 635–672.

Attema, Arthur E, Bleichrodt, Han, Rohde, Kirsten IM, & Wakker, Peter P. 2010. Time-tradeoff sequences for analyzing discounting and time inconsistency. *Management Science*, **56**(11), 2015–2030.

Augenblick, Ned, & Rabin, Matthew. 2019. An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, **86**(3), 941–975.

Augenblick, Ned, Niederle, Muriel, & Sprenger, Charles. 2015. Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, **130**(3), 1067–1115.

Aycinena, Diego, Blazsek, Szabolcs, Rentschler, Lucas, & Sprenger, Charles. 2020. *Intertemporal Choice Experiments and Large-Stakes Behavior*. Tech. rept. Chapman University, Economic Science Institute.

Bryan, Gharad, Karlan, Dean, & Nelson, Scott. 2010. Commitment devices. *Annu. Rev. Econ.*, **2**(1), 671–698.

Coller, Maribeth, & Williams, Melonie B. 1999. Eliciting individual discount rates. *Experimental Economics*, **2**(2), 107–127.

Cubitt, Robin P, & Read, Daniel. 2007. Can intertemporal choice experiments elicit time preferences for consumption? *Experimental Economics*, **10**(4), 369–389.

Demuynck, Thomas, & Hjertstrand, Per. 2019. Samuelson's Approach to Revealed Preference Theory: Some Recent Advances. *Chap. Ch.9 of:* Cord, Robert A, Anderson, Richard G, & Barnett, William A (eds), *Paul Samuelson Master of Modern Economics*. Springer.

Ericson, Keith Marzilli. 2017. On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association*, **15**(3), 692–719.

Fournier, Lisa R, Stubblefield, Alexandra M, Dyre, Brian P, & Rosenbaum, David A. 2019a. Starting or finishing sooner? Sequencing preferences in object transfer tasks. *Psychological research*, **83**(8), 1674–1684.

Fournier, Lisa R, Coder, Emily, Kogan, Clark, Raghunath, Nisha, Taddese, Ezana, & Rosenbaum, David A. 2019b. Which task will we choose first? Precrastination and cognitive load in task ordering. *Attention, Perception, & Psychophysics*, **81**(2), 489–503.

Freeman, David J. 2021. Revealing Naïveté and Sophistication from Procrastination and Preproperation. *American Economic Journal: Microeconomics*, **13**(2), 402–38.

Halevy, Yoram. 2015. Time consistency: Stationarity and time invariance. *Econometrica*, **83**(1), 335–352.

Harrison, Glenn W, Lau, Morten I, & Williams, Melonie B. 2002. Estimating individual discount rates in Denmark: A field experiment. *American economic review*, **92**(5), 1606–1617.

Haushofer, Johannes. 2015. *The cost of keeping track*. Tech. rept. Citeseer.

Heufer, Jan, & Hjertstrand, Per. 2015. Consistent subsets: Computationally feasible methods to compute the Houtman–Maks-index. *Economics Letters*, **128**, 87–89.

Houtman, Martijn, & Maks, J. 1985. Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden*, **19**(1), 89–104.

Laibson, David. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, **112**(2), 443–478.

Montiel Olea, José Luis, & Strzalecki, Tomasz. 2014. Axiomatization and measurement of quasi-hyperbolic discounting. *The Quarterly Journal of Economics*, **129**(3), 1449–1499.

O'Donoghue, Ted, & Rabin, Matthew. 1999. Doing it now or later. *American economic review*, **89**(1), 103–124.

Read, Daniel, & Van Leeuwen, Barbara. 1998. Predicting hunger: The effects of appetite and delay on choice. *Organizational behavior and human decision processes*, **76**(2), 189–205.

Rosenbaum, David A, Gong, Lanyun, & Potts, Cory Adam. 2014. Pre-crastination: Hastening subgoal completion at the expense of extra physical effort. *Psychological Science*, **25**(7), 1487–1496.

Sayman, Serdar, & Öncüler, Ayse. 2009. An investigation of time inconsistency. *Management Science*, **55**(3), 470–482.

Takeuchi, Kan. 2011. Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, **71**(2), 456–478.

Thaler, Richard. 1981. Some empirical evidence on dynamic inconsistency. *Economics letters*, **8**(3), 201–207.

Wasserman, Edward A. 2019. Precrastination: The fierce urgency of now. *Learning & behavior*, **47**(1), 7–28.

Effort Profile	Time Consistent			Reversal		Non-Strotz	Censored	Quads Observed
	1st day	2nd day	3rd day	earlier	later			
14, 20, 28	83%	0%	0%	8%	4%	4%	0%	24
16, 20, 25	77%	2%	0%	8%	2%	2%	10%	52
18, 20, 22	73%	4%	2%	6%	0%	2%	13%	52
19, 20, 21	75%	8%	0%	0%	0%	0%	17%	12
20, 20, 20	67%	2%	6%	10%	0%	0%	15%	52
22, 20, 18	50%	0%	18%	11%	0%	4%	18%	28
25, 20, 16	57%	0%	18%	7%	0%	4%	14%	28
TOTAL	69%	2%	6%	8%	1%	2%	13%	248

Table 8: Classifying Observed Choice Quads by Effort Schedule (Non-endogenous subsample)
This table only uses only the $(\emptyset, e_2, e_3, e_4)$ effort schedules, and only the 52 participants whose randomly assigned choice that counts does not include e_1

	First Day Choice, Second Day Choice		
Effort Profile	Time Invariant	Today, Not Today	Not Today, Today
14, 20, 28	88%	3%	10%
16, 20, 25	86%	7%	7%
18, 20, 22	82%	9%	9%
19, 20, 21	86%	9%	6%
20, 20, 20	70%	13%	17%
22, 20, 18	70%	9%	21%
25, 20, 16	76%	1%	23%
Grand Total	79%	8%	13%

Table 9: Time Invariance by Effort Schedule (Non-endogenous subsample)

Table 10: Structural Logit Estimates (Non-endogenous subsample)

	Dependent variable:
	Today1
EffortToday	-0.140^{***}
	(0.042)
EffortNotTodayK1	0.199^{***}
	(0.041)
EffortNotTodayK2	0.185^{***}
	(0.043)
Observations	743
Log Likelihood	-358.575
Akaike Inf. Crit.	723.150
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Variance-Covariance Matrix of Structural Logit Estimates (Non-endogenous subsample)

	EffortToday	EffortNotTodayK1	EffortNotTodayK2
EffortToday	0.001803	-0.001716	-0.001747
EffortNotTodayK1	-0.001716	0.001716	0.001725
EffortNotTodayK2	-0.001747	0.001725	0.001826

Table 12: Variance-Covariance Matrix of Parameter Estimates using Delta Method

	β	δ	λ
β	0.046933	0.070261	0.007851
δ	0.070261	0.163917	0.016764
λ	0.007851	0.016764	0.001803

Appendix

Deriving Parameter Estimates and Variance-Covariance Matrix for Delta Method

We estimate the logistic regression:

$y_t = b_0 e_t + b_1 \mathrm{I}_{k=1} e_{t+k} + b_2 \mathrm{I}_{k=2} e_{t+k} + \epsilon_t$

For our structural model, we are interested in the values of $\frac{b_1^2}{b_0 b_2}$ $(= \beta)$, $\frac{b_2}{b_1}$ $(= \delta)$, and b_0 $(= \lambda)$

If $b \sim N(b^{\star}, \Sigma)$ then the distribution of $f(b)$ is $N(f(b^{\star}), C\Sigma C')$ where $C = \nabla f(b)$.

Let $f(b) = \begin{bmatrix} \frac{b_1^2}{b_0 b_2} \\ \frac{b_2}{b_1} \\ b_0 \end{bmatrix}$. Then, $C = \nabla f(b) = \begin{bmatrix} -\frac{b_1^2}{b_0^2 b_2} & \frac{2b_1}{b_0 b_2} & -\frac{b_1^2}{b_0 b_2^2} \\ 0 & -\frac{b_2}{b_1^2} & \frac{1}{b_1} \\ 1 & 0 & 0 \end{bmatrix}$.

The estimated variance matrix of interest is $C\hat{\Sigma}C'$