Down the Rabbit Hole: Habit-formation in Internet Use among Unemployed Workers*

Tristan Potter†
Drexel University

November 24, 2021

Abstract

This letter tests for habit-formation in leisure-related internet use (LIU) using time diary data from a panel of unemployed workers. Drawing on insights from the consumption-habit literature, I use a simple model of intertemporal time allocation to derive a test for habit-formation in leisure activities. The data reveal robust evidence of habit-formation in LIU. Contrary to a popular narrative, however, this effect is driven by Generation-X rather than by digitally native Millennials. With the exception of reading, I find no evidence of habit-formation across various *offline* leisure activities.

Keywords: Habit-formation; leisure; internet use
JEL Classification: D91, J22

*Thanks to Sebastien Bradley, Ioannis Kospentaris, Andre Kurmann, Irina Murtazashvili, Maria Olivero, Ricardo Serrano-Padial, Mark Stehr and Yoto Yotov for helpful comments and suggestions.

†School of Economics, LeBow College of Business, Drexel University. Email: tristan.l.potter@drexel.edu. Phone: 215-895-2540. Mailing address: 3220 Market St., Philadelphia, PA 19104.

1 Introduction

There is increasing concern about the large—and growing—amount of free time that we spend online: Congress recently considered legislation intended to reduce time spent on social media;[1] the WHO now includes digital gaming as an official addictive disorder;[2] and the New York Times' podcast "Rabbit Hole" chronicles how immersive YouTube viewing behavior can lead to political radicalization.[3] These examples all lead to a fundamental question about online leisure: Is internet use habit-forming? Existing evidence concerning this question is limited and typically based on neurological studies or associational cross-sectional analysis.[4] This letter provides causal evidence on habit-formation in leisure-related internet use (LIU) using weekly time-diary data from a panel of unemployed workers.

The analysis proceeds as follows: Section 2 considers a simple model of time allocation and derives an estimating equation in the spirit of the consumption-habit literature. Section 3 discusses data, measurement and identification. Section 4 estimates the model and documents evidence of habit-formation in LIU. Section 5 concludes with suggestions for future research.

2 Model

To derive an estimating equation, I consider a simple model of intertemporal leisure-time allocation. The model is motivated by the observation that intertemporal leisure choice is, in many respects, analogous to intertemporal consumption choice: Utility maximization subject to time constraints implies that we should smooth leisure over time—indeed, this helps explain why we leave household chores for the weekend (thus "borrowing" leisure) and spend Sundays preparing meals for the week (thus "saving" leisure). Accordingly, the model and derivations below extend Dynan (2000) and Deaton (1992)—who study habit-formation in consumption—to habit-formation in leisure.

Formally, consider an unemployed worker who allocates time either to a leisure activity or to stochastic time commitments (household chores, errands, etc.) that can be transferred intertemporally at cost ζ but must eventually be completed. In each period t, worker i chooses leisure time, $l_{i,t}$, to maximize[5]

$$E_t\left[\sum_{d=0}^{D}\beta^d u(\tilde{l}_{i,t+d};\xi_{i,t+d})\right] \tag{1}$$

where β is the discount factor, $\xi_{i,t+d}$ is a taste shock to leisure, and

$$\tilde{l}_{i,t+d} \equiv l_{i,t+d} - \rho l_{i,t+d-1} \tag{2}$$

is habit-adjusted leisure where ρ governs the extent of habit-formation. Optimization is subject to three constraints,

$$l_{i,t} \leq T \tag{3}$$

$$l_{i,t} + \tau_{i,t} \leq T + b_{i,t} - (1+\zeta)b_{i,t-1} \tag{4}$$

[1] The 2019 Social Media Addiction Reduction Technology (SMART) Act proposed, among other measures, a default 30-minute daily limit on social media apps.

[2] https://www.who.int/news-room/q-a-detail/addictive-behaviours-gaming-disorder.

[3] https://www.nytimes.com/column/rabbit-hole.

[4] Researchers have recently begun to study this question in an experimental setting. See, e.g., Allcott et al. (2021).

[5] See Appendix A for a complete derivation in a more general framework with multiple leisure activities.

$$b_{i,D} \leq 0 \tag{5}$$

where T denotes the total time available in any period, $\tau_{i,t} \geq 0$ denotes an idiosyncratic time commitment shock in period t which detracts from total (lifetime) time available for leisure, and $b_{i,t}$ denotes "borrowed" time in period t. The constraint in (3) indicates that leisure cannot exceed total time in a day, (4) indicates that leisure can be borrowed (or saved) in a manner largely analogous to consumption, reflecting the intuition described above, and (5) indicates that all time commitments must eventually be completed. In Appendix A, I show that a generalized version of this model implies

$$\Delta \ln(l_{i,t}) = \gamma_0 + \rho \Delta \ln(l_{i,t-1}) + \gamma_1 \Delta \ln(\xi_{i,t}) + e_{i,t} \tag{6}$$

where γ_0 and γ_1 are constants and $e_{i,t}$ is a mean-zero and serially uncorrelated error under rational expectations. This is the main estimating equation.

3 Data and Estimation

3.1 Data

I estimate (6) on a panel of unemployed workers from the Survey of Unemployed Workers in New Jersey (SUWNJ).[6] The SUWNJ is a weekly survey of unemployment insurance benefit recipients in New Jersey beginning in the fall of 2009 and continuing through early 2010. The survey covers 6,025 respondents for up to 24 weeks for a total of 39,201 interviews. Critically, the survey contains a detailed time diary component in which respondents provide an hour-by-hour account of the preceding day, selecting up to two choices from 21 activity categories for each hour.

3.2 Measuring LIU

I construct a measure of LIU using six of the 21 activity categories in the SUWNJ: "Using the Computer/Internet/Email" (CIE), "Watching TV" (TV), "On the phone" (Phone), "Working" (Work), "Searching for a job" (Search), "Attending job training program" (Training), and "Preparing for/taking course" (Course). Specifically, for any hour of the day $\tau \in \{$7am, 8am,..., 10pm$\}$, let $a_{\tau,1}$ and $a_{\tau,2}$ denote the two reported activities.[7] Without loss of generality, order activities such that CIE is activity $a_{\tau,1}$ if selected. Then I define LIU as:

$$\text{LIU} \equiv \sum_{\tau=\text{7am}}^{\text{10pm}} \begin{cases} 60 & \text{if } a_{\tau,1} = \text{CIE} \quad \& \quad a_{\tau,2} \in \{\text{CIE, TV, Phone}, \cdot\} \\ 30 & \text{if } a_{\tau,1} = \text{CIE} \quad \& \quad a_{\tau,2} \notin \{\text{Work, Search, Training, Course}\} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where '$\cdot$' indicates non-response. This precise formulation is not essential so long as CIE is included. Based on this construction, the average worker spent 68 minutes/day on LIU in 2009/2010.[8]

3.3 Empirical strategy

As in the consumption-habit literature, time-averaging bias and measurement error could potentially threaten consistent estimation of (6). Time-averaging bias—which arises when the decision interval

[6] The SUWNJ is publicly available from Princeton University: https://opr.princeton.edu/archive/njui/.

[7] The subscripts 1 and 2 are arbitrary and do not indicate the order in which activities were completed.

[8] This value is similar to 2010 data from Zenith Media in which the average respondent spent 85 minutes/day online. See Appendix B.

is short relative to the observation interval—is not likely in the present context given the use of time diary data. Measurement error—which will impart an MA(2) structure to the error term—will attenuate estimates of ρ, thus biasing results *away* from a finding of habit-formation.[9] Accordingly, I adopt two instrumenting strategies to estimate (6): One that does not attempt to correct for measurement error and thus yields estimates that should be understood as lower bounds on the extent of habit-formation, and one that attempts to mitigate the possibility of measurement error.

First, following Anderson and Hsiao (1981), I instrument $\Delta \ln(l_{i,t-1})$ with $\ln(l_{i,t-2})$ ("A-H"). The virtue of this approach is that $\ln(l_{i,t-2})$ is highly correlated with $\Delta \ln(l_{i,t-1})$ but uncorrelated with $\Delta \ln(\xi_{i,t})$ and will thus alleviate the correlation between the lagged dependent variable and any unobserved component of the preference shocks in the error term.[10] Measurement error, however, will attenuate these estimates towards zero.

Second, to address possible attenuation due to measurement error, I propose a second instrument: Twice-lagged *total* leisure time ("TLT"). This instrument is correlated with $\Delta \ln(l_{i,t-1})$ by construction because LIU is a component of total leisure time. Furthermore, under an assumption that coarser divisions of time are easier for respondents to recall than more granular ones—for example, if it is easier for a respondent to remember that an online course ended at 8pm after which they enjoyed leisure until going to bed than it is to remember exactly how that leisure time was divided—then aggregating various leisure activities will reduce measurement error. Thus, under this assumption, the proposed instrument accepts a higher variance (due to a weaker correlation with $\Delta \ln(l_{i,t-1})$) in exchange for reduced bias, implying that we should expect higher but less precise point estimates.

4 Results

4.1 Leisure-related internet use

Table 1 reports results from estimation of equation (6) on the full sample. The first two columns correspond to a specification that includes controls for jobless duration, whether the interview was on a weekend, and calendar-week and survey-cohort dummies. The second two columns correspond to a specification that adds demographic controls.

The first row of Table 2 reports the coefficient on the lagged growth rate of LIU, ρ, indicating the extent of habit-formation. Across specifications, the coefficient is positive and significant. Consistent with the discussion of measurement error and the different instrumenting strategies above, the point estimates from the TLT instrument are higher, and also less precise, than those from the A-H instrument. For context, note that in the presence of habit-formation, a one-time $g\%$ shock to the growth rate of LIU will result in LIU asymptoting towards a value that is approximately $g\left(\frac{\rho}{1-\rho}\right)\%$ greater than it would be absent habit-formation. Concretely, the midpoint of the estimates in Table 1 implies that a shock that doubles LIU will eventually result in individuals permanently spending 25% more time online than they otherwise would.

The A-B tests are consistent with the presence of measurement error and thus the interpretation of the A-H estimates as lower bounds on ρ.[11] There is weaker evidence of second-order autocorrelation with the A-H instrument because the A-B test will tend to (erroneously) reject the null when $\hat{\rho}$ is close to zero, as it is under A-H. Furthermore, the Kleinbergen-Paap test suggests that under-identification is not a problem with either instrument. In Appendix D, I show that these results are robust to

[9] See Appendix C.1 for further discussion.

[10] I also estimate the model using the Arellano and Bond (1991) difference-GMM estimator. See Appendix D.2.

[11] Measurement error will generate second- but not third-order autocorrelation in the residuals. See Appendix C.2.

Table 1: Habit-formation in LIU

	Basic controls		*Full controls*	
	A-H	TLT	A-H	TLT
Habit (ρ)	0.07***	0.33***	0.06***	0.30***
	(0.02)	(0.09)	(0.02)	(0.09)
Duration	-0.00	0.00	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Weekend	-0.26	-0.22	-0.27	-0.23
	(0.37)	(0.47)	(0.37)	(0.47)
Female			0.02	0.02
			(0.05)	(0.05)
Millennial (20-29)			0.01	0.01
			(0.11)	(0.13)
Baby Boomer (50-69)			-0.05	-0.04
			(0.04)	(0.05)
College grad.			-0.15	-0.08
			(0.19)	(0.19)
Savings			-0.04	-0.06
			(0.04)	(0.05)
Married			-0.08	-0.08
			(0.08)	(0.08)
Kids			0.04	0.04
			(0.06)	(0.06)
Constant	0.07	0.14	0.25	0.24
	(0.38)	(0.48)	(0.44)	(0.52)
Dummies:				
Time	×	×	×	×
Cohort	×	×	×	×
A-B test: AR(2)†	0.38	0.01	0.30	0.02
A-B test: AR(3)†	0.80	0.86	0.81	0.86
K-P test‡	0.00	0.00	0.00	0.00
Observations	15347	15347	15139	15139

† p-values; H_0: No order-i autocorrelation in residuals.

‡ p-values; H_0: Model underidentified.

Robust standard errors in parentheses.

Source: SUWNJ

Notes: Regressions use survey weights. Standard errors are robust and clustered at the person level. Sample includes all respondents between ages 20 and 70. AH: Anderson-Hsiao estimator; TLT: 2SLS instrumenting with twice-lagged total leisure time.

Table 2: Habit-formation by generational cohort

	Millennials		*Generation-X*		*Baby Boomers*	
	A-H	TLT	A-H	TLT	A-H	TLT
Habit (ρ)	0.05	0.58	0.08***	0.29**	0.05	0.26**
	(0.07)	(0.63)	(0.03)	(0.13)	(0.04)	(0.13)
Full controls:	×	×	×	×	×	×
A-B test: AR(2)†	0.18	0.06	0.83	0.21	0.27	0.50
A-B test: AR(3)†	0.62	0.64	0.81	0.76	0.72	0.61
K-P test‡	0.00	0.04	0.00	0.00	0.00	0.00
Observations	931	931	5437	5437	8771	8771

† p-values; H_0: No order-i autocorrelation in residuals.
‡ p-values; H_0: Model underidentified.
Robust standard errors in parentheses.

alternative definitions of LIU, the use of additional lags (via difference-GMM), and excluding outlying observations of the dependent variable.

4.2 A generational divide

Table 2 reports results for three generational cohorts:[12] Millennials (20-29), Generation-X (30-49) and Baby Boomers (50-69). The results reveal an interesting finding: The habit-formation observed in Table 1 appears to be driven by middle-aged workers rather than by Millennials. This result contrasts with the popular narrative that Millennials are uniquely addicted to social media, etc. However, it is consistent with recent data indicating that Generation-X spends roughly 10% more time on social media than do Millennials.[13] This result may reflect the fact that Millennials are the first generation to be truly "digitally native," thus rendering them less susceptible to the predictive algorithms, digital marketing, etc. that could explain the apparent habit-formation in LIU among older generations.

4.3 Offline leisure

An important advantage of the SUWNJ is that it provides information on forms of leisure other than LIU. This allows us to test whether the habit-formation observed in LIU in Table 1 is also a feature of *offline* leisure, or unique to internet use. Accordingly, Table 3 estimates (6) replacing LIU with several common types of offline leisure: watching TV, reading/writing, socializing and exercise.

Table 3 reveals two interesting results: (i) there is no evidence of habit-formation in watching TV, socializing or exercise, and (ii) there is some tentative evidence of habit-formation in reading. These two results are of independent interest, but are also important for interpreting the results for LIU in Tables 1 and 2: Specifically, they suggest that the results for LIU are not due to some mechanical feature of the data or estimation procedure (in which case we would expect to consistently find evidence of habit-formation in other activities), and they also highlight that it is not screen time (such as watching TV), but rather *online* time, that is habit-forming.

[12] Categorization based on respondents' ages in 2009.
[13] "Generation-X More Addicted to Social Media Than Millennials, Report Finds" (*New York Times*, 2017)

Table 3: Habit-formation in offline leisure

	TV		*Reading*		*Socializing*		*Exercise*	
	A-H	TLT	A-H	TLT	A-H	TLT	A-H	TLT
Habit (ρ)	-0.01	0.10	0.08	0.79**	0.04	0.12	0.06	-0.23
	(0.04)	(0.14)	(0.06)	(0.37)	(0.04)	(0.41)	(0.04)	(0.17)
Full controls:	×	×	×	×	×	×	×	×
A-B test: AR(2)†	0.47	0.03	0.08	0.97	0.94	0.10	0.63	0.93
A-B test: AR(3)†	0.22	0.22	0.72	0.71	0.53	0.40	0.69	0.70
K-P test‡	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Observations	5437	5437	5437	5437	5437	5437	5437	5437

† p-values; H_0: No order-i autocorrelation in residuals.

‡ p-values; H_0: Model underidentified.

Robust standard errors in parentheses.

Notes: Results based on the sample of middle-aged (30-49) respondents.

5 Conclusion

The results documented in this letter lead to a number of important questions for future research: Does habit-formation in internet use affect labor supply decisions? How does habit-formation interact with other behavioral mechanisms such as present bias? And what are the implications for policy? Future research should address these questions to further our understanding of the nature of time spent online.

References

Allcott, H., M. Gentzkow, and L. Song (2021). Digital addiction. *Working Paper*.

Anderson, T. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*.

Deaton, A. (1992). *Understanding Consumption*. Oxford: Clarendon Press.

Dynan, K. (2000). Habit formation in consumer preferences: Evidence from panel data. *American Economic Review*.

Hayashi, F. (1985). The permanent income hypothesis and consumption durability: Analysis based on japanese panel data. *Quarterly Journal of Economics*.

Appendices

A Derivation of estimating equation

This appendix derives the main estimating equation in (6), adapting the derivations in Dynan (2000) and Deaton (1992) to the case of leisure. The model and derivation below generalize the model described in the body of the text to the case of multiple leisure activities.

In each period t, worker i chooses how much time to devote to leisure activity a, $l_{i,t}^a$, for $a = 1, .., A$, to maximize

$$E_t \left[\sum_{d=0}^{D} \beta^d u(\tilde{l}_{i,t+d}^1, \tilde{l}_{i,t+d}^2, ..., \tilde{l}_{i,t+d}^A; \xi_{i,t+d}^1, \xi_{i,t+d}^2, ..., \xi_{i,t+d}^A) \right] \tag{A.1}$$

where β is the discount factor, $\xi_{i,t+d}^a$ is a taste shock for $a = 1, ..., A$, and

$$\tilde{l}_{i,t+d}^a \equiv l_{i,t+d}^a - \rho^a l_{i,t+d-1}^a \tag{A.2}$$

is habit-adjusted leisure where ρ^a governs the extent of habit-formation for activity a. Optimization is subject to three constraints,

$$\textstyle\sum_{a=1}^{A} l_{i,t}^a \leq T \tag{A.3}$$

$$\textstyle\sum_{a=1}^{A} l_{i,t}^a + \tau_{i,t} \leq T + b_{i,t} - (1+\zeta) b_{i,t-1} \tag{A.4}$$

$$b_{i,D} \leq 0 \tag{A.5}$$

where T denotes the total time available in any period, $\tau_{i,t} \geq 0$ denotes an idiosyncratic time commitment shock in period t which detracts from total (lifetime) time available for leisure, $b_{i,t}$ denotes "borrowed" time in period t, and ζ is the intertemporal price of leisure, analogous to the interest rate for consumption. The constraint in (A.3) indicates that leisure cannot exceed total time in a day, (A.4) indicates that leisure can be borrowed (or saved) in a manner largely analogous to consumption, reflecting the intuition described in Section 2,[14] and (A.5) indicates that all time commitments must eventually be completed.[15]

Letting $\Gamma_{i,t}$ denote the Lagrange multiplier on the constraint in (A.3), the first-order condition for the optimal choice of time devoted to leisure activity a is given by:

$$\frac{\partial u}{\partial \tilde{l}_{i,t}^a} - \rho\beta E_t \left\{ \frac{\partial u}{\partial \tilde{l}_{i,t+1}^a} \right\} - \Gamma_{i,t} = \beta(1+\zeta) E_t \left\{ \frac{\partial u}{\partial \tilde{l}_{i,t+1}^a} - \rho\beta \frac{\partial u}{\partial \tilde{l}_{i,t+2}^a} - \Gamma_{i,t+1} \right\}. \tag{A.6}$$

I maintain two assumptions throughout the analysis: First, in keeping with the consumption-habit literature, I assume that the felicity function is separable in the various (habit-adjusted) leisure

[14] For more intuition on (A.4), let $b_{i,t-1} = 0$, rearrange, and note that the constraint will always bind, so we can write: $T = \sum_{a=1}^{A} l_{i,t}^a + \tau_{i,t} - b_{i,t}$. Thus, total time in a day (T) is divided between leisure ($\sum_{a=1}^{A} l_{i,t}^a$) and non-leisure activities ($\tau_{i,t} - b_{i,t}$) where the latter is just total new time commitments (e.g. needing to clean the house) net of borrowed time (e.g. cleaning put off until the future).

[15] Note that so long as $\tilde{l}_{i,t}^a = 0$ yields infinite marginal utility (as is the case with the standard CRRA utility function considered below), $\tilde{l}_{i,t}^a$ will always be positive. This implies that, for any weakly positive initial level of leisure $l_{i,0}^a$, $l_{i,t}^a$ will also always be positive for $t > 0$. Thus, there is no need to impose a non-negativity constraint on $l_{i,t}^a$.

activities.[16] Second, I assume that the time constraint in (A.3) is slack, so that the associated multipliers are zero: $\Gamma_{i,t} = \Gamma_{i,t+1} = 0$. This assumption reflects the fact that, in the SUWNJ, virtually no respondents spend all of their time on leisure.[17]

Under these assumptions, we can follow Hayashi (1985), who shows (in the context of an analogous model of consumption habit-formation) that when D is large, (A.6) simplifies to

$$E_t \left\{ \beta(1+\zeta) \frac{\partial u / \partial \tilde{l}^a_{i,t+1}}{\partial u / \partial \tilde{l}^a_{i,t}} \right\} = 1 \tag{A.7}$$

which implies

$$\beta(1+\zeta) \frac{\partial u / \partial \tilde{l}^a_{i,t}}{\partial u / \partial \tilde{l}^a_{i,t-1}} = 1 + \epsilon_{i,t} \tag{A.8}$$

where $\epsilon_{i,t}$ is the individual's forecast error, reflecting innovations to the "permanent" level of time available for leisure (for example, unexpectedly having to take care of a sick or aging family member). Under rational expectations, $E_{t-1}[\epsilon_{i,t}] = 0$ and $\epsilon_{i,t}$ is serially uncorrelated.

Assuming the felicity function is isoelastic and of the form

$$u(\tilde{l}^1_{i,t+d}, ..., \tilde{l}^A_{i,t+d}; \xi^1_{i,t+d}, ..., \xi^A_{i,t+d}) = \textstyle\sum_{a=1}^{A} \xi^a_{i,t} \frac{\left(\tilde{l}^a_{i,t}\right)^{1-\sigma}}{1-\sigma}, \tag{A.9}$$

we can write (A.8) as

$$\beta(1+\zeta) \frac{\xi^a_{i,t}}{\xi^a_{i,t-1}} \left[\frac{\tilde{l}^a_{i,t}}{\tilde{l}^a_{i,t-1}} \right]^{-\sigma} = 1 + \epsilon_{i,t}. \tag{A.10}$$

Taking the natural logarithm and using (A.2), we obtain

$$\Delta \ln(l^a_{i,t} - \rho^a l^a_{i,t-1}) = \frac{\ln \beta + \ln(1+\zeta)}{\sigma} + \frac{1}{\sigma} \Delta \ln(\xi^a_{i,t}) - \frac{1}{\sigma} \ln(1 + \epsilon_{i,t}). \tag{A.11}$$

Finally, following Dynan (2000), I approximate $\Delta \ln(l^a_{i,t} - \rho^a l^a_{i,t-1})$ with $\Delta \ln(l^a_{i,t}) - \rho^a \Delta \ln(l^a_{i,t-1})$, which allows (A.11) to be written as

$$\Delta \ln(l^a_{i,t}) = \gamma_0 + \rho^a \Delta \ln(l^a_{i,t-1}) + \gamma_1 \Delta \ln(\xi^a_{i,t}) + e_{i,t} \tag{A.12}$$

where $\gamma_0 \equiv [\ln \beta + \ln(1+\zeta)]/\sigma$, $\gamma_1 \equiv \frac{1}{\sigma}$ and $e_{i,t} \equiv -\ln(1+\epsilon_{i,t})/\sigma$. This is equation (6) in the text.

[16] The consumption-habit literature typically assumes separability of the felicity function in various types of consumption goods (and implicitly leisure) in order to estimate models using data on, e.g., food expenditures. See, for example, the discussion at the beginning of Section 2 in Dynan (2000).

[17] 99% of time diaries show fewer than 10 hours of leisure, out of a total possible 16 hours, suggesting that the vast majority of the sample devotes a significant amount of time each day to non-leisure activities. This is consistent with the constraint in (A.3) not binding.

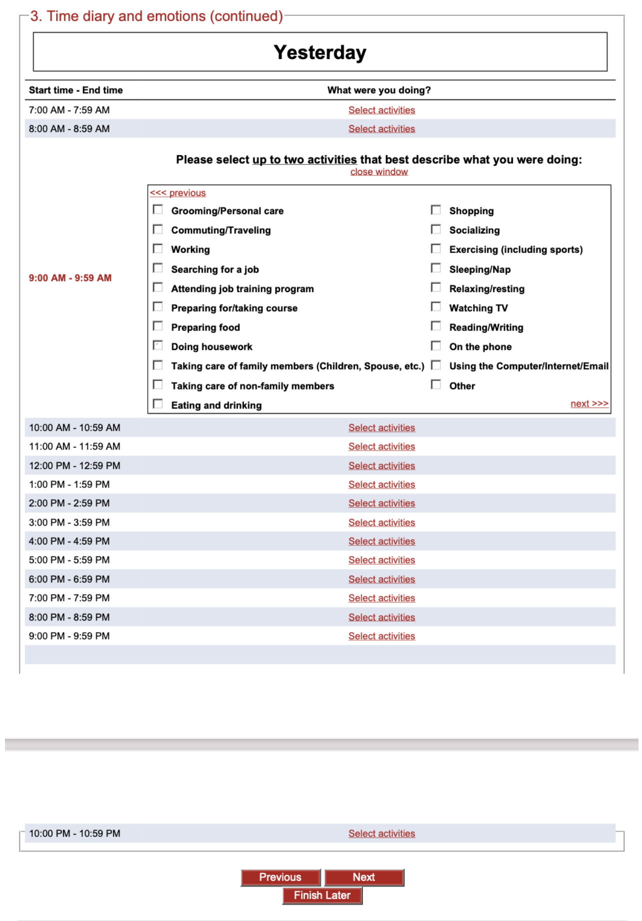
B Data

This appendix provides further details on construction of time-use variables used for the empirical analysis.

B.1 Survey of Unemployed Workers in New Jersey

Each weekly survey of the SUWNJ asked respondents to complete a time diary, accounting for their activities between 7am and 11pm on the previous day. Figure A1 contains a graphic of what respondents would see in this section of the survey.[18]

Figure A1: Survey of Unemployed Workers in New Jersey: Time Diary



B.2 Variable construction

Because all time-use variables are defined in such a way that they will contain zeros, I add one minute to the total number of minutes before taking logs in order to estimate (6). The "offline" time use variables used in the empirical analysis in Table 3 are computed identically to LIU, except that if an individual reports concurrently engaging in a leisure activity and internet use, the time spent on the activity is recorded as a zero in order to restrict attention to offline activities.

[18]Complete survey data and documentation may be obtained from: https://dss.princeton.edu/catalog/resource1350.

B.3 Instrument construction: Total leisure time

The instrument proposed to mitigate measurement error in Section 3 is constructed as follows. Let $\mathbb{L}$ denote the set of leisure-related activities in the SUWNJ and $\mathbb{N}$ the set of non-leisure activities. Then, ordering activities such that leisure activities are always the first activity if selected (as in the text), I define the total leisure time instrument, T, as:

$$\mathrm{T} \equiv \sum_{\tau=7\mathrm{am}}^{10\mathrm{pm}} \begin{cases} 60 & \text{if } a_{\tau,1} \in \mathbb{L} \ \& \ a_{\tau,2} \in \{\mathbb{L}, \cdot\} \\ 30 & \text{if } a_{\tau,1} \in \mathbb{L} \ \& \ a_{\tau,2} \notin \{\mathbb{L}, \cdot\} \ \& \ a_{\tau,2} \notin \mathbb{N} \\ 0 & \text{otherwise.} \end{cases}$$

This is the instrument corresponding to columns denoted by "TLT" (total leisure time) in the empirical analysis in Section 4.

B.4 Summary statistics

Table A1 reports summary statistics for the constructed time-use variables described in the main text.

Table A1: Summary statistics for constructed time-use variables

	Mean	Std. Dev.	Min.	P_{25}	P_{50}	P_{75}	Max.
LIU	68.1	89.9	0	0	30	120	960
TV	100.9	100.9	0	0	90	150	960
Reading/Writing	23.3	49.5	0	0	0	30	840
Exercise	16.1	38.2	0	0	0	0	540
Socializing	30.6	70.1	0	0	0	30	840

Source: Survey of Unemployed Workers in New Jersey

Notes: All times are reported in minutes per day. Sample consists of all respondents.

C Measurement error

This appendix shows that measurement error will tend to bias the A-H instrument and discusses implications for the Arellano-Bond autocorrelation tests in the main text.

C.1 Inconsistency of the A-H instrument

In the text I argue that the Anderson-Hsiao instrument will lead to attenuated point estimates in the presence of measurement error, and will thus understate the extent of habit-formation. To see this, suppose that various categories of leisure are measured with error. Formally, letting $l_{i,t}$ denote true leisure time and $l^*_{i,t}$ denote measured leisure time, suppose

$$\ln(l^*_{i,t}) = \ln(l_{i,t}) + \nu_{i,t} \tag{C.1}$$

where $\nu_{i,t}$ is classical measurement error. Using this, the estimating equation in (6) becomes[19]

$$\Delta \ln(l^*_{i,t}) - \nu_{i,t} + \nu_{i,t-1} = \gamma_0 + \rho \left[\Delta \ln(l^*_{i,t-1}) - \nu_{i,t-1} + \nu_{i,t-2}\right] + e_{i,t} \tag{C.2}$$

which can be rearranged to obtain

$$\Delta \ln(l^*_{i,t}) = \gamma_0 + \rho \Delta \ln(l^*_{i,t-1}) + z_{i,t} \tag{C.3}$$

where $z_{i,t}$ is the composite error term, defined as

$$z_{i,t} \equiv \nu_{i,t} - (1+\rho)\nu_{i,t-1} + \rho\nu_{i,t-2} + e_{i,t}. \tag{C.4}$$

To see that the Anderson-Hsiao instrument (that is, instrumenting $\Delta \ln(l^*_{i,t-1})$ with $\ln(l^*_{i,t-2})$) will lead to attenuation, take the probability limit of the 2SLS estimator of ρ,

$$\text{plim}(\hat{\rho}^{AH}) = \rho + \frac{Cov(\ln(l^*_{i,t-2}), z_{i,t})}{Cov(\ln(l^*_{i,t-2}), \Delta \ln(l^*_{i,t-1}))}. \tag{C.5}$$

To determine the direction of the inconsistency, we consider the sign of the numerator and denominator of the second term on the right-hand side of (C.5). For the numerator, we have

$$\begin{aligned} Cov(\ln(l^*_{i,t-2}), z_{i,t}) &= Cov(\ln(l_{i,t-2}) + \nu_{i,t-2}, \nu_{i,t} - (1+\rho)\nu_{i,t-1} + \rho\nu_{i,t-2} + e_{i,t}) \\ &= \rho Var(\nu_{i,t-2}) > 0 \end{aligned}$$

where the inequality holds if there is habit-formation ($\rho > 0$).[20] For the denominator, we have

$$\begin{aligned} Cov(\ln(l^*_{i,t-2}), \Delta \ln(l^*_{i,t-1})) &= Cov(\ln(l^*_{i,t-2}), \ln(l^*_{i,t-1}) - \ln(l^*_{i,t-2})) \\ &= Cov(\ln(l^*_{i,t-2}), \ln(l^*_{i,t-1})) - Var(\ln(l^*_{i,t-2})) \\ &\leq Var(\ln(l^*_{i,t-2})) - Var(\ln(l^*_{i,t-2})) \\ &= 0 \end{aligned}$$

[19] For ease of notation, I neglect the preference shock term $\gamma_1 \Delta \ln(\xi_{i,t})$.

[20] $\rho < 0$ would reflect "durability" in leisure, and would imply the opposite sign. In this case, negative point estimates would be attenuated towards zero.

where the second-to-last line follows from the Cauchy-Schwarz inequality. Thus, the second term on the right-hand side of (C.5) must be negative, implying that $\text{plim}(\hat{\rho}^{AH}) < \rho$.

C.2 Discussion of serial correlation tests

Inspection of equation (C.4) reveals that, in the presence of measurement error, we should expect both first- and second-order autocorrelation in the residuals, but no third-order autocorrelation. Reassuringly, as noted in the main text, we detect no evidence of third-order autocorrelation in the residuals. However, the results for the tests for second-order autocorrelation (see, for example, Table 1) appear to be sensitive to the instrumenting strategy: Specifically, the results using the A-H instrument (and also the GMM instruments in Appendix D.2) suggests that there is no second-order autocorrelation, whereas the results using the TLT instrument suggest that there is. In fact, these apparently conflicting results are both consistent with the presence of measurement error. To understand why, note that the extent to which we should expect to observe second-order autocorrelation in the residuals depends critically on the value of $\hat{\rho}$: For values of $\hat{\rho}$ close to zero, such as those I find using the A-H instrument, the correlation between the residuals $\hat{z}_{i,t}$ and $\hat{z}_{i,t-2}$ will also be close to zero, leading to an erroneous failure to reject the null of no second-order autocorrelation. Furthermore, this explains why the test for second-order autocorrelation using the TLT instrument—which, as I have agued, can reasonably be expected to at least partially mitigate measurement error and thus yield a higher estimate $\hat{\rho}$—yields a rejection of the null and thus evidence of second-order autocorrelation in the residuals. Importantly, this does not invalidate the TLT instrument under the assumption (described in the main text) that aggregation of individual leisure activities tends to reduce measurement error, implying that the TLT instrument itself is not significantly contaminated with measurement error and is thus uncorrelated with the error term.

D Robustness

This appendix considers several robustness checks for the main results.

D.1 Alternative variable definitions

The results in the main text are based on a particular definition of LIU. Table A2 reports results for two alternative definitions of LIU. The first ("Excl. Phone/TV") excludes time spent online during which individuals also report using the phone or TV. The second ("Incl. R/W") includes time spent online during which individuals also report reading or writing.

Table A2: Alternative definitions of LIU

	Baseline		*Excl. Phone/TV*		*Incl. R/W*	
	A-H	TLT	A-H	TLT	A-H	TLT
Habit (ρ)	0.06***	0.30***	0.06***	0.32***	0.09***	0.31***
	(0.02)	(0.09)	(0.02)	(0.10)	(0.03)	(0.10)
Duration	-0.00	0.00	-0.00	0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Weekend	-0.27	-0.23	-0.27	-0.23	-0.54	-0.52
	(0.37)	(0.47)	(0.37)	(0.48)	(0.38)	(0.47)
College grad.	-0.15	-0.08	-0.18	-0.11	-0.06	-0.03
	(0.19)	(0.19)	(0.19)	(0.18)	(0.20)	(0.19)
Savings	-0.04	-0.06	-0.04	-0.06	0.01	-0.01
	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Married	-0.08	-0.08	-0.07	-0.07	-0.08	-0.07
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.08)
Kids	0.04	0.04	0.04	0.04	-0.01	-0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
Constant	0.25	0.24	0.29	0.28	0.24	0.25
	(0.44)	(0.52)	(0.45)	(0.52)	(0.46)	(0.53)
Dummies:						
Time	×	×	×	×	×	×
Cohort	×	×	×	×	×	×
A-B test: AR(2)†	0.31	0.02	0.40	0.13	0.19	0.03
A-B test: AR(3)†	0.81	0.86	0.92	0.95	0.59	0.65
K-P test‡	0.00	0.00	0.00	0.00	0.00	0.00
Observations	15139	15139	15139	15139	15139	15139

† p-values; H_0: No order-i autocorrelation in residuals.

‡ p-values; H_0: Model underidentified.

Robust standard errors in parentheses.

Notes: Baseline: Definition of LIU in main text; Excl. Phone/TV: Definition of LIU excludes time spent online also using phone or TV; Incl. R/W: Definition of LIU includes time spent online also reading or writing.

D.2 Difference GMM

Table A3 reports results from the baseline A-H specification in the main text as well as the Arellano-Bond difference GMM estimator which uses additional lags of $\ln(l_{i,t-1})$ as instruments.

Table A3: Difference GMM

	Basic controls		*Full controls*	
	A-H	GMM	A-H	GMM
Habit (ρ)	0.07***	0.07***	0.06***	0.06***
	(0.02)	(0.02)	(0.02)	(0.02)
Duration	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Weekend	-0.26	-0.02	-0.27	-0.02
	(0.37)	(0.08)	(0.37)	(0.08)
College grad.			-0.15	-0.26*
			(0.19)	(0.15)
Savings			-0.04	-0.02
			(0.04)	(0.04)
Married			-0.08	-0.07
			(0.08)	(0.07)
Kids			0.04	0.07
			(0.06)	(0.06)
Constant	0.07	0.05	0.25	0.37
	(0.38)	(0.19)	(0.44)	(0.26)
Dummies:				
Time		×		×
Cohort		×		×
Hansen test‡		0.13		0.19
Observations	15347	15347	15139	15347

‡ p-values; H_0: Over-identifying restrictions valid.
Two-step standard errors in parentheses.
Notes: Results based on instrumenting with the first five lags of $\ln(l_{i,t-1})$.

D.3 Excluding outlying observations

Table A4 reports results after excluding observations for which $\Delta \ln(l_{i,t})$ is below the 1^{st} percentile or in excess of the 99^{th} percentile of the distribution of $\Delta \ln(l_{i,t})$.

Table A4: Trimming extreme observations

	Basic controls		*Full controls*	
	A-H	TLT	A-H	TLT
Habit (ρ)	0.12***	0.42***	0.11***	0.39***
	(0.03)	(0.11)	(0.03)	(0.11)
Duration	-0.00	0.01	-0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Weekend	-0.30	-0.24	-0.29	-0.23
	(0.37)	(0.48)	(0.37)	(0.47)
College grad.			-0.15	-0.07
			(0.20)	(0.20)
Savings			-0.02	-0.06
			(0.05)	(0.05)
Married			-0.11	-0.10
			(0.09)	(0.10)
Kids			0.13	0.14
			(0.08)	(0.09)
Constant	-0.00	0.07	0.11	0.07
	(0.38)	(0.49)	(0.44)	(0.53)
Dummies:				
Time	×	×	×	×
Cohort	×	×	×	×
A-B test: AR(2)†	0.05	0.00	0.04	0.01
A-B test: AR(3)†	0.85	0.71	0.85	0.72
K-P test‡	0.00	0.00	0.00	0.00
Observations	14694	14694	14494	14494

† p-values; H_0: No order-i autocorrelation in residuals.

‡ p-values; H_0: Model underidentified.

Robust standard errors in parentheses.

Notes: Sample excludes outliers, defined as observations below the 1^{st} percentile or above the 99^{th} percentile of the distribution of $\Delta \ln(l_{i,t})$.