Journal of Time Use Research
International Association for Time Use Research (IATUR)
jtur.iatur.org

Getting Started with Time-Use Data

Jay Stewart[1]

1. U.S. Bureau of Labor Statistics, Stewart.Jay@bls.gov

Abstract

Time-use data have unique characteristics that make it different from other types of household survey data. Single-day time use surveys provide a detailed snapshot of a person's activities on the diary day. But the large amount of day-to-day variation in the amount of time spent in various activities means that activities done on the diary day do not reflect the person's long-run time use. Thus, time-use data is a sample of person-days, not a sample of people. This feature of time-use data has implications for its analysis.

Keywords: Time-use data, methodology, sample design, multivariate analysis, 'short-run' and 'long-run' time use

https://doi.org/10.32797/jtur-2018-2

1 Introduction

There is still a lot of discussion in the community of time-use researchers regarding the proper way to analyze time-use data. I was an author of two articles (Stewart, 2013; Frazis and Stewart, 2012) that examined the unique features of time-diary data and the implications of these features for analysis. These technical papers are like a User's Guide for Time-Diary Data. This article covers much of the same material but in a less detailed and less technical way, hence the title.

I will start out talking about the unique features of time-use data (I will use the terms time-use and time-diary interchangeably). Next, I will illustrate how these features affect the types of statistics that can be estimated using time-diary data and the importance of using sample weights. Finally, I will discuss other analytical issues that are relevant to the analysis of time-diary data. Throughout the paper, I will present examples based on a single-day time-use survey such as the American Time Use Survey (ATUS), which I am most familiar with. I recognize that many countries conduct time-use surveys (most for longer than the US) that may collect data for two or more days, but the principles I lay out here are applicable to time-use surveys that collect two days of data.

At different places in this article, I will mention how other researchers have approached estimation, but I will not cite any of these articles. The interested reader can refer to the two papers referenced above.

This work is licensed under a Creative Commons Attribution 4.0 International License

2 How are Time-Use Data Different from Other Household Data?

Probably the most distinctive feature of time-diary data is the short reference period—usually one or two days—compared to reference periods of a week or more for most household surveys. Many time-use surveys ask respondents to report about the previous day, which reduces the recall period. The shorter recall period results in more accurate data, because it is easier to recall the activities of the previous day than the previous week. And because time diaries ask respondents to report all of their activities rather than asking about a few select activities, they are relatively free of social desirability bias and aggregation bias since the time spent in all activities on the diary day must sum to 1440 minutes. Time-diary data would seem to be ideal for answering a number of questions about the time spent in non-market activities such as childcare, household work, and job search that are of interest to policymakers but not captured by most household surveys.

However, for most policy questions, it is long-run time use—the amount of time that individuals spend in an activity over the course of a month or a year—that is relevant. Thus, there is a mismatch between the one-day reference period of most time-use surveys and the period of interest to policymakers. This mismatch would not matter if people did the same thing every day. But there is a lot of day-to-day variation in how people spend their time. Because of this variation, a single-day diary (or even a two-day diary) is not representative of any individual's long-run time use. This feature of time-diary data has important implications for its analysis.

A convenient, and I think sensible, way to think about individuals' decision making is to assume that people decide how much time they want to spend on an activity over a long period of time, say a month. How that time is allocated across the days of the month will depend on many factors. Some will be things that the researcher can observe (day of week, weather, etc.), while others (illness, personal preference, etc.) are not observed. From the researcher's point of view, the distribution across days of time spent in the activity can be thought of as being random.

3 Implications for Analysis

Given that single-day diaries (or even two-day diaries) are not representative of how individuals spend their time, one should think of a single-day time-use dataset as a sample of person-days—not as a sample of people. To illustrate how time-diary data differ from other household data, consider the following example. Suppose that everybody in the economy spends 7 hours per week doing household work. In this simple example, the mean and median time spent doing household work is 7 hours per week or 1 hour per day. The variance, across individuals, in the amount of time spent doing household work is 0, because everybody spends the same amount of time in that activity. Now let's compare the estimates that we would obtain from a single-day time-use survey and a standard household survey with a one-week reference period.

3.1 Mean Time Use

For this example, suppose that half of respondents spend 7 hours doing household work every Sunday, and no time during the other six days of the week, and the other half of the sample spends 3.5 hours each on Tuesdays and Saturdays. Assuming that everybody reports their activities correctly,[1] both the time-diary and the household survey will generate the same estimate of the average amount of time spent doing household work—7 hours per week or 1 hour per day. For the time-diary survey, the calculation is as follows:

$$0.5\times(0+3.5+0+0+0+3.5+0) + 0.5\times(0+0+0+0+0+0+7) = 7$$

Thus, the time-diary survey data can be used to estimate the average amount of time spent (per person) doing household work.

3.2 Median Time Use

Here the story changes—the household and time-diary data give different answers. In the household data, the median time spent doing household work (calculated over individuals) is 7 hours per week or 1 hour per day. But in the time-diary data, the median is calculated over a sample of person days. In the hypothetical example above, people spend 0 hours doing household work on more than half of the person days in the sample ($^{1}/_{14}$ of the sample spends 7 hours on the diary day, $^{2}/_{14}$ of the sample spends 3.5 hours on the diary day, and the remaining $^{11}/_{14}$ of the sample spend no time on the diary day), which implies that the median is 0.

It is worth noting that this is also true for any percentile rank. Thus, graphs that purport to show the distribution of the time individuals spend in a particular activity using one- or two-day time-diary data are probably capturing a fair amount of day-to-day variation. The more day-to-day variation there is, the less representative the distribution is of individuals' time use. Comparisons over time are also problematic because changes in the distribution could be at least partially due to changes in how people distribute time over the days of the month.

To illustrate, if one were to generate the distribution of time spent in leisure activities, the people at the higher percentiles were likely interviewed about a weekend day. Changes in this distribution could be caused by people shifting leisure time either to or away from weekends as well as real changes in their long-run time use. Needless to say, this type of analysis is not just uninformative, it is very misleading.

3.3 Variance of Time Use

As with percentiles, the day-to-day variation in time use makes it impossible to estimate the variance of time spent in an activity across individuals. The time spent in an activity by person i on day d, t_{id}, can be thought of as the sum of the individual's average long-run time use, m_i, and a random term representing day-to-day variation, e_{id}:

$$t_{id} = m_i + e_{id} \tag{1}$$

where m_i and e_{id} are uncorrelated by construction. The variance of daily time use t_{id} is:

$$Var(t_{id}) = Var(m_i) + Var(e_{id}) \geq Var(m_i) \tag{2}$$

Unless the variance of e_{id} equals zero (no day-to-day variation), the variance of daily time use will overestimate the variance across individuals, and therefore is an upper bound for $Var(m_i)$.

[1] It is also necessary to assume that the weights are correct. In particular, that the weights in the time-diary data give each day of the week equal representation. I also assume that there is no sampling error.

3.4 Multivariate Analysis

Many time-use studies use regression analysis and, in most of these, time-use is the dependent variable although a few studies have included time use as an independent variable.

Time Use as the Dependent Variable

The short one- or two-day reference period in most time-use surveys results in a large number of zero-value observations for many activities. Because of the large number of zeroes in time-diary data, a common approach has been to estimate these regressions using Tobit. Some papers have reported Tobit coefficients, but most report Tobit marginal effects, which account for the probability that the time spent in the activity is zero. More recently—following my critique of different estimation methods—researchers have stopped using Tobit in favor of OLS or the Cragg two-part model.

The Tobit model was originally introduced to address situations where zero-value observations represent a “corner solution” in an individual’s utility maximization calculation. That is, to address situations where the individual never purchases a good or never does the activity, which implies that the zero represents the individual’s long-run time use.[2] But there are two ways that a zero observation can arise in time-diary data:

(1) The individual never does the activity. For example, most non-parents never do childcare.
(2) The individual does the activity, but did not do it on the diary day. For example, employed people do not work every day, the unemployed do not look for work every day, and parents do not do childcare every day.

The second reason is more common, mainly because researchers typically restrict their sample to people who do the activity.[3]

In my paper “Tobit or Not Tobit?” I examine the bias associated with three commonly-used estimation procedures: Tobit, OLS, and the Cragg two-part model.[4] To compare these procedures, I constructed a simulated dataset where the average time spent on an activity is a linear function of three covariates plus a normally-distributed random term (i.e., the assumptions underlying the Tobit model). The daily time spent on the activity is randomly determined, as is the fraction of zero-value observations.

I found that Tobit marginal effects were biased downward and that the extent of the bias increases as the fraction of zero observations in the data increases (holding mean time use constant). The two-part model generates unbiased results unless the fraction of zeroes in the data is a function of one of the covariates. In that case, the two-part model generates biased results. This is unfortunate and a little surprising, because the main reason that a researcher might want to use the two-part model is to shed light on the tradeoff between how often people do the activity

[2] Under this interpretation, individuals would like to purchase negative amounts of the good or spend negative amounts of time in the activity. So reporting coefficients is appropriate.

[3] Researchers identify “doers” using other questions in the survey. In some cases, it is not possible to identify doers, because the necessary question was not asked. As noted below, it is not possible to identify doers by looking at time spent on the diary day.

[4] In the Cragg two-part model, a probit equation is estimated over the entire sample and a separate truncated regression on the non-zero observations, both using maximum likelihood. The coefficients from these two equations are combined to estimate unconditional marginal effects.

and how much time they spend in the activity conditional on doing the activity on the diary day. Only OLS produced unbiased results in all of the simulations. However, it is worth noting that as the fraction of zero-value observations in the sample increases, standard errors on the coefficients increase and the regression R^2 decreases. This occurs because more of the variation comes from extreme-valued observations—zeroes and large positive values.

The main reason that Tobit performs poorly is that it assumes that the process that determines whether an individual does the activity is the same process that determines how much time the individual spends on the activity conditional on doing the activity. The two-part model is an improvement, in that it allows these processes to differ.[5] However, if individuals make decisions about their time use as described above, then the independent variables in the probit equation should include day-specific variables that affect how the individual allocates time across days (for example, rainfall and temperature). But typically the right hand side variables in the probits are long-run variables (demographic and household characteristics).

A key insight of the paper is that the zero-value observations do not convey any special information. That is, given that an individual does the activity, there is little difference between spending one minute doing the activity on the diary day and spending zero minutes. In a regression context, the random variation around long-run time use is easily handled by the error term in an OLS regression.

There are a couple of other points worth noting. First, estimating an OLS regression is analogous to estimating means in that both are linear functions of the data. In a simple regression that includes only 0-1 (dummy) variables, the OLS coefficients are conditional sample means. If the dummy variables uniquely identify all groups, and are exhaustive and mutually exclusive (and there are no other variables in the regression), then the coefficients are unconditional sample means.

Related to the previous point, it is common to include a day-of-week (usually weekend or weekday) dummy variable in regressions. But if the researcher is interested in the association between the right-hand-side variables and long-run time use, then day-specific variables (such as a weekday dummy) need not be included as independent variables. While it is true that time use is very different on weekdays vs. weekend days, including a weekday dummy complicates the interpretation of the coefficients and does not really add any insight as long as the data are weighed so that each day of the week receives equal representation in the sample. If one is interested in how the association between the right-hand-side variables and the dependent variable differ by day of week, there are better approaches (see the section on Reporting Results below).

And second, as I note in the paper, the presence of zero-value observations in the data are likely to affect standard errors by introducing heteroscedasticity into the residuals. Thus standard errors should be estimated using a robust procedure, which should be easy to implement in most statistical packages.

---

[5] The generalized Tobit model also allows the two processes to differ. I did not examine that model in my paper, but the results are likely to be similar to those from the two-part model. The generalized Tobit model is a little more complicated to estimate than the two-part model, but it is more efficient because the two parts are estimated jointly.

Time Use as an Independent Variable

Including time use as an independent variable is more complicated. The correct approach depends on the question to be answered and whether the researcher is interested in the effects of long-run time use or the effects of time use on the diary day. For example, time spent exercising on the diary day would have a negligible effect on a person's body mass index (BMI). Rather, it is long-run time use that matters, and a single-day time diary provides only an imperfect estimate of long-run time use as shown in equation (1).

Even though daily time use has a negligible effect on the dependent variable, one could view daily time use as a proxy for long-run time use. From equation (1), we know that daily time use is the sum of long-run time use plus an error term. We can therefore view daily time use as measuring long-run time use with error, which is an example of classical measurement error. It is well known that classical measurement error biases coefficient estimates toward zero. However, it is worth nothing that this result holds only for a single independent variable measured with error. If more than one time-diary variable is included in the regression, then in general nothing can be said about the direction of the bias, which can be bounded only in specific circumstances.[6]

There are two solutions to errors in variables. The standard approach is to use instrumental variables (IV) estimation.[7] For example, weather has been used as an instrument for exercise time and for time spent watching television. Here, the researcher must be careful to determine whether the instrument predicts long-run or short-run time use. For example, temperature and rainfall on the diary day would predict short-run time use, whereas average temperature and rainfall would predict long-run time use. Another advantage of using IV is that it is also addresses endogeneity. In the obesity example, one could imagine that, in addition to exercise time affecting BMI, BMI may affect time spent exercising.

The second approach is to divide the sample into groups based on demographic characteristics and calculate group means for the independent and dependent variables. For example, married women age 25-34 with one or more children under five might constitute a group. Each group is an observation, and the group mean of the dependent variable is regressed on the group means of each of the independent variables. The groups should be mutually exclusive and exhaustive. In forming the groups, it is important to make the groups small enough so that there is enough variation and large enough to ensure proper day-of-week representation (see below). It is also important to correct the standard errors. This approach is simpler than IV, but does not address any potential endogeneity issues.

3.5 Sample Weights and Day-of-Week Representation

The discussions above implicitly assume that we have a simple random sample and that the sample is distributed approximately equally across the days of the week. But most surveys are not simple random samples. Rather they are stratified random samples, and sample weights account for the

[6] For a full discussion, see Klepper and Leamer (1984).

[7] A full treatment of IV is beyond the scope of this paper. Briefly, IV is a technique where the researcher finds variables (the instruments) that are correlated with the independent variable, but are not correlated with the outcome variable. For example, if the outcome is body mass index, and the independent variable is exercise, then the researcher could use average temperature and rainfall as instruments. The researcher uses the instruments to generate predicted values for the time-use variable, which are then used in the regression.

survey's sample design and non-response.[8] Therefore it is important to use the survey's sample weights to ensure that the sample properly reflects the characteristics of the population. In most household surveys, the weights account only for the demographic characteristics of the population. However, as noted above, time-diary data differ from other types of household data, because they are a sample of person-days.[9] Because time use varies considerably by day of week, it is important that the seven days of the week are equally represented—or at least that weekdays and weekend days are correctly represented.

Ideally, day of week is included as a stratifying variable when constructing sample weights. For example, the ATUS oversamples weekend days. About 10 percent of the sample is interviewed about each of the five weekdays and about 25 percent of the sample is interviewed on each of the two weekend days. The average sample weight for weekday observations is about 2.5 times as large as the weights of the weekend observations, which results in each day of the week being about $^1/_7$ of the weighted sample. If researchers do not use the sample weights when estimating statistics, such as the mean, their estimates will be seriously biased. Because weekend days represent about half the sample, but only $^2/_7$ of actual days, time spent working would be underestimated and time spent in leisure and household work would be overestimated.

It is important to note that, even though the ATUS weights are constructed to ensure correct day-of-week representation at the aggregate level and for major demographic groups, they may not be correct for smaller, more-detailed demographic groups. It is always a good idea to check the weighted distribution of observations across days and adjust the weights if necessary. To adjust the weights so that each day of the week represents $^1/_7$ of the sample, generate the distribution of the weighted sample across days of the week and calculate the adjustment factor for each day of the week. To illustrate, suppose the distribution of the weighted sample is: 0.12, 0.14, 0.16, 0.10, 0.20, 0.15 and 0.13. If each day were equally represented, then each day would represent about $1/7^{th}$ (or about 14.29 percent of the sample). The adjustment factor for Monday (0.12) would be approximately 1.19 (more precisely, the adjustment factor is equal to $1/(0.12\times7)$).[10] This adjustment is not perfect, because it does not use the same methodology used to generate the sample weights. But it is a significant improvement over unadjusted weights for small demographic groups.

Another approach is to construct synthetic weeks. That is, generate an estimate for each day of the week (using sample weights), and add the seven estimates to arrive at a weekly total (or average the seven estimates to arrive at average daily time use). This approach can be used to estimate means and should yield similar estimates to the reweighting approach, although it is a little more complicated to calculate standard errors. This approach cannot be used when estimating regressions.

[8] Typically, household surveys are stratified by demographic characteristics and often geographic location. In addition, many surveys oversample smaller demographic groups to ensure sufficient sample to generate estimates for these groups. Observations in oversampled groups have lower weights. Along the same lines, observations in demographic groups that have lower response rates are assigned larger weights.

[9] An intuitive way to think about weights is "what do they add up to?" In BLS's household survey, the Current Population Survey, the weights are equal to the number of people that the observation represents and they add up to the population. In contrast, the weights in the ATUS add up to the number of person days in a year (365×population).

[10] To arrive at the adjustment factor, solve $^1/_7 = 0.12x$ for x. Factors for others days are calculated similarly.

4 Other Issues

Below, I discuss several issues that come up when working with time-use data.

4.1 Classification Based on Time Diaries

It is tempting to use time diary-data to examine topics like shift work or time poverty. However, given the considerable day-to-day variation in the amount of time that people spend in various activities, classifying people into groups based on a single day's activities is problematic.

To illustrate, let us return to the example above, where people spend 7 hours per week doing household work, and suppose that we consider someone to be time poor if they spend more than 3 hours per day doing household work. In the household survey, none of the respondents would be classified as time poor, because they only spend an average of one hour per day on household work. The time-diary estimate of the fraction of people who are time poor is ${}^{3}/_{14}$ or about 21 percent. Thus, as with medians, estimating the fraction of people who are time poor using time-diary data is misleading because we are implicitly assuming that people spend the same amount of time on each activity every day.

Identifying shift workers using time diaries is equally problematic. Shift workers may have worked more-standard work hours on their diary day or may not have worked at all. Also, non-shift workers may have worked shift hours on their diary day. That said, it is possible to examine the time use of shift workers, but only if the time-use survey asks a retrospective question to identify this type of work arrangement. For this type of analysis, it is crucial to include both work and non-work days to get a complete picture of their time use.

4.2 Reporting Results by Day of Week

It is common to see researchers report regression results separately for weekdays and weekend days, or to include a weekend dummy variable in a regression. The rationale is that time spent on weekend days and weekdays is very different, especially for people who are employed. While this is true, both approaches make it difficult to interpret the results.

Let me start by saying that there is nothing inherently wrong with running separate regressions for weekends and weekdays. The main issue with this approach is that researchers simply report the two regressions without providing any additional context. This makes the results difficult to interpret because, for most questions, researchers and policy-makers are interested in the total effect. What does it mean to say that working mothers spend one hour less per day with their children on weekdays, compared with non-working mothers, and two hours more per day on weekends? It is possible to figure out the total effects (and their standard errors). But why not make it easier on the reader and report the results in a more intuitive way?

In Frazis and Stewart (2012), we offered a suggestion: in addition to reporting the separate regressions, also report the total effect. This could be done in one of two ways. Under the first approach, the researcher aggregates the results from the weekday and weekend regressions to estimate the total effect. In the above example, the combined effect is that employed mothers spend 1 hour less per week doing childcare compared to non-employed mothers (5 weekdays × −1 hour + 2 weekend days × 2 hours). By adding the third column showing the total effect, the researcher could report something like this: "Working mothers spend one hour less per week doing childcare, compared with mothers who are not employed. They spend one hour less with their children on weekdays and make up most of that time by spending two hours more per day on weekends. Thus, while the overall effect of maternal employment on childcare time is relatively small, working mothers shift much of their childcare time to weekends."[11]

The second approach is to estimate a regression over the entire sample—both weekend days and weekdays (without day-of-week dummy variables)—to estimate the total effect. If the sample is weighted so that the day-of-week representation is correct, then this approach should generate the same results as the first approach. Depending on the question at hand, the researcher may still want to show the results from the separate weekday and weekend regressions, because they provide insight as to how employed mothers arrange their time compared with non-employed mothers.

4.3 Time Use of Couples

Some time-use surveys collect data from multiple household members. Typically, the diary day is the same for all household members, which allows the researcher to get a complete picture of the household's time use on the diary day. When these data are available, it is tempting to analyze how household work and other responsibilities are allocated between husbands and wives. But it is important to keep in mind that there are two ways that husbands and wives can substitute their time when allocating time to household work—within-day and between-day—and that we observe only within-day substitution. For example, it may be possible to examine tag-team parenting by looking at when the two parents spend time with their children on the diary day and whether they spent that time together. But it is impossible to determine which spouse spends more time with their children over longer periods of time, because parents substitute time between days as well as within a day.

It should be noted that most questions about couples' time use can be answered using time-use surveys that collect only one diary per household. For example, we can determine how much time mothers and fathers spend with their children by estimating mean childcare time separately for mothers and fathers.[12] For this type of exercise, one would restrict the sample to mothers and fathers that have the same family characteristics (for example, number and ages of children). To look at couples' together-time, the researcher could use information about who the respondent was with. It is worth noting that estimates based on husbands' time diaries will generally be different from estimates based on wives' diaries. These types of comparisons could also be done in a regression framework, which would make it easier to control for family characteristics.

5 Concluding Remarks

Time-diary data are valuable for exploring many aspects about how people use their time. But time-dairy data are different from most other household surveys because of the short reference period relative to the period of interest to most researchers and policymakers. In this paper, I have tried to discuss the implications of the unique features of time-diary data in an intuitive and non-

---

[11] This is an example. When looking at maternal employment, it might make more sense to run separate regressions for workdays and non-work days. In this case, the two sets of regression coefficients would be weighted by the respective (weighted) fractions of workdays and non-work days.

[12] In surveys like the ATUS, that have detailed information on who else was present during each episode, one can also determine whether their spouses are present during childcare time.

technical way.[13] The main thing to keep in mind when analyzing time-diary data is that they are not a sample of people, but of person-days. This feature of time-diary data has not been fully appreciated by many researchers. As a result, many research studies have reported incorrect or misleading results, missed opportunities to put their results in context, or have attempted to answer questions that really cannot be answered with time-diary data. Our goal in writing the original papers was to clear up some of the misconceptions about time-diary data, and to guide researchers on what can and cannot be estimated with time-diary data.

Acknowledgements

This article draws heavily on my research with Harley Frazis. I thank Harley Frazis and Sabrina Pabilonia for helpful comments on earlier drafts of this paper, but any remaining errors are mine alone. The views expressed here are mine and do not necessarily reflect those of the US Bureau of Labor Statistics.

References

Frazis, Harley and Jay Stewart (2012) "How to Think About Time-Use Data: What Inferences Can We Make About Long- and Short-Run Time Use from Time Diaries?" *Annales d'Economie et Statistique* (*Annals of Economics and Statistics*) 105/106, pp. 231-246. https://doi.org/10.2307/23646463

Klepper, Steven and Edward E. Leamer (1984) "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica* 52(1), pp. 162-183. https://doi.org/10.2307/1911466

Stewart, Jay (2013) "Tobit or Not Tobit?" *Journal of Economic and Social Measurement* 38(3), pp. 263-290. https://doi.org/10.3233/JEM-130376

[13] Interested readers may want to refer to the original papers.