

# Time Aggregation in Panel Data on Income and Consumption

May 23, 2018

Edmund Crawley<sup>1</sup>

JHU and Danmarks Nationalbank

---

## Abstract

The time aggregation problem. of [Working \(1960\)](#)

---

**Keywords** Income, Consumption, Saving

**JEL codes** D12, D31, D91, E21

---

<sup>1</sup>Crawley: Department of Economics, Johns Hopkins University, [ecrawle2@jhu.edu](mailto:ecrawle2@jhu.edu)

Thanks to Danmarks Nationalbank for providing financial support

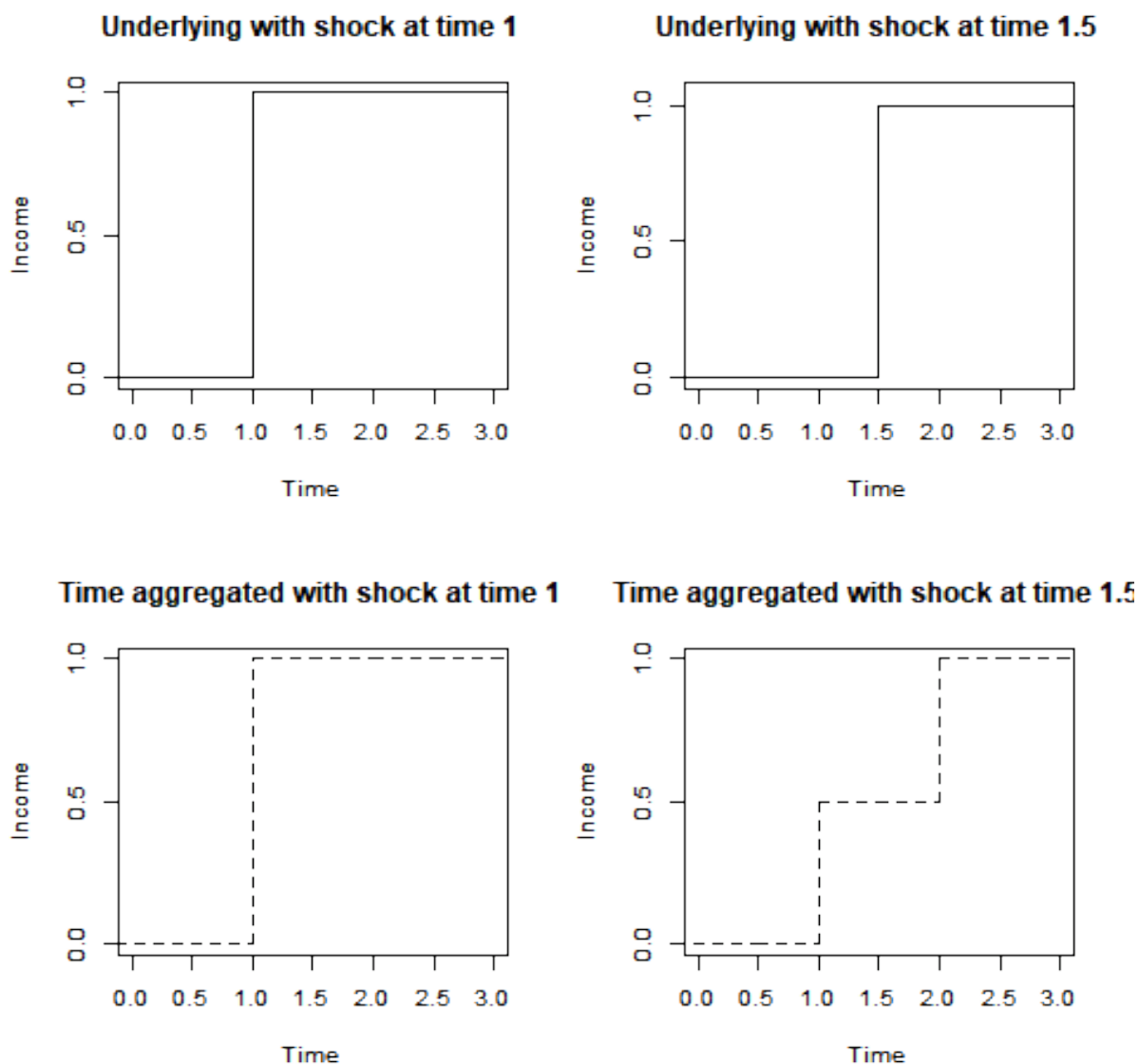
# 1 Introduction

In a short note in *Econometrica*, Working (1960) made the simple but important point that time aggregation can induce serial correlation that is not present in the original series. This fact was readily absorbed by the macroeconomic literature, where such time aggregated series are common (for an example see Campbell and Mankiw (1989)). Recently, by studying the covariance structure of panel data, much progress has been made in understanding household income and consumption dynamics. However, this literature has not accounted for the serial correlation induced by the time aggregated nature of observed income and consumption data. This oversight can result in significant bias. This paper will focus on the implications of time aggregation for the methodology in Blundell, Pistaferri, and Preston (2008) (henceforth BPP), but it applies to a broad swath of the literature. I show that the pass through from transitory income to consumption, originally estimated by BPP to be 5%, is close to 20% when the serial correlation in the data induced by time aggregation is accounted for.

## 1.1 What is Time Aggregation?

Many observed time series in economics are given at a lower frequency than the underlying data that generates them. For example, income is often observed at an annual frequency when it may in fact consist of paychecks arriving at a monthly, biweekly or irregular timetable. The transform income into an annual frequency we sum up all the income that was received by a household during the year, a process known as time aggregation. The key insight of Working (1960) is that even if there is no correlation between changes in income at the underlying frequency, the time resulting time aggregated series will show positive autocorrelation. The intuition behind this can be seen in figure 1 showing an income process that begins at zero and increases to one in the second year. The top left graph shows the path of income if the shock occurs exactly at the start of the second year, and the bottom left graph shows the time aggregated process exactly mirrors that. There is no income in the first year and one unit of income in each of the second and third years. The top right shows an alternative income process in which the shock occurs half way through the second year. Now the resulting time aggregated process (bottom right) does not mirror the underlying. As before there is no income in the first year, but in the second year the individual receives an income of one for half the year, resulting in a time aggregated income of 0.5. In the third year the individual receives an income of one for the entire year, and hence a time aggregated income of one. If we can only see the time aggregated process, when we observe income increasing from year one to year two we do not know if the shock occurred at the beginning of the year or half way through. If it occurred at the beginning of the year, as in the left hand graphs of figure 1, then we would not expect to see any further increase in the time aggregated process associated with it. However, if it occurred half way through, as in the right hand graphs of figure 1, we would only have observed half the total increase in income and would expect the time aggregated

**Figure 1** Time Aggregation Induces Serial Correlation



process to continue to increasing in the following period. Therefore, assuming there is some positive probability that the shock occurred half way through the second period, we would expect to see further increases in the observed process. This is how time aggregation induces serial correlation in the first difference of an observed process even when the underlying process is a random walk. Section 2 lays this out formally and shows that this autocorrelation tends to  $\frac{1}{4}$  as the number of time subdivisions increases to infinity.

## 1.2 Relation to the Literature

# 2 Time Aggregated Random Walk

In this section I formally prove that a time aggregated random walk is autocorrelated and show that this autocorrelation tends to  $\frac{1}{4}$  as the number of time subdivisions increases to infinity. I will also introduce continuous time notation that will be used for the underlying model in section 3.

## 2.1 The two sub-division case

I begin with the two-subdivision case. The underlying income process follows a random walk at discrete time periods  $t \in \{0, 1, 2, 3, \dots\}$ :

$$y_t = \begin{cases} 0 & \text{if } t = 0 \\ y_{t-1} + \varepsilon_t & \text{otherwise} \end{cases}$$

where  $\varepsilon_t$  is i.i.d. and has variance  $\sigma^2$ . The time aggregated process is observed every two periods at  $T \in \{2, 4, 6, \dots\}$  and is equal to the sum of income over the two periods leading up to it:

$$y_T^{obs} = y_T + y_{T-1}$$

The observed income change is given by:

$$\begin{aligned} \Delta^2 y_T^{obs} &= y_T^{obs} - y_{T-2}^{obs} \\ &= \varepsilon_T + 2\varepsilon_{T-1} + \varepsilon_{T-2} \end{aligned}$$

This allows for easy calculation of the serial correlation:

$$\begin{aligned} \text{Cov}(\Delta^2 y_T^{obs}, \Delta^2 y_{T-2}^{obs}) &= \sigma^2 \\ \text{Var}(\Delta^2 y_T^{obs}) &= \sigma^2 + 4\sigma^2 + \sigma^2 \\ &= 6\sigma^2 \\ \text{Corr}(\Delta^2 y_T^{obs}, \Delta^2 y_{T-2}^{obs}) &= \frac{1}{6} \end{aligned}$$

## 2.2 The N sub-division case

The two sub-division case easily extends to N sub-divisions. Using the same underlying income process, the observable income process is now aggregated over N periods:

$$y_T^{obs} = \sum_{t=T-N+1}^T y_t$$

So that the observed change in income is:

$$\Delta^N y_T^{obs} = \sum_{t=T-N+1}^T y_t - y_{t-N}$$

$$\begin{aligned}
&= \varepsilon_T + \varepsilon_{T-1} + \dots + \varepsilon_{T-N+2} + \varepsilon_{T-N+1} \\
&\quad + \varepsilon_{T-1} + \varepsilon_{T-2} + \dots + \varepsilon_{T-N+1} + \varepsilon_{T-N} \\
&\quad + \varepsilon_{T-2} + \varepsilon_{T-3} + \dots \\
&\quad \dots \\
&\quad \quad \quad + \varepsilon_{T-N+1} + \dots + \varepsilon_{T-2N+2} \\
&= N\varepsilon_{T-N+1} + \sum_{i=1}^{N-1} i \left( \varepsilon_{T-i+1} + \varepsilon_{T-2N+i+1} \right)
\end{aligned}$$

We can now calculate the autocorrelation:

$$\begin{aligned}
\text{Cov}(\Delta^N y_T^{obs}, \Delta^N y_{T-N}^{obs}) &= \sum_{i=1}^{N-1} i(N-i)\sigma^2 \\
&= \frac{N(N^2-1)}{6}\sigma^2 \\
\text{Var}(\Delta^N y_T^{obs}) &= N^2\sigma^2 + 2 \sum_{i=1}^{N-1} i^2\sigma^2 \\
&= \frac{N(2N^2+1)}{3}\sigma^2 \\
\text{Corr}(\Delta^N y_T^{obs}, \Delta^N y_{T-N}^{obs}) &= \frac{N^2-1}{2(2N^2+1)} \rightarrow \frac{1}{4} \text{ as } N \rightarrow \infty
\end{aligned}$$

## 2.3 The Continuous Time Case

It will turn out to be significantly simpler to work with a model in which shocks can occur at any point in continuous time. Here I introduce some notation for such a model, and show that it gives a good approximation even if the actual underlying process is discrete (say quarterly or monthly).

The underlying income process will be modeled as a martingale process in continuous time,  $y_t$ , where for all  $s_1 > s_2 > s_3 > s_4 > 0$ :

$$\begin{aligned}
\text{Var}(y_{s_1} - y_{s_2}) &= (s_1 - s_2)\sigma^2 \\
\text{Cov}(y_{s_1} - y_{s_2}, y_{s_3} - y_{s_4}) &= 0 \\
y_s &= 0 \quad \text{if } s < 0
\end{aligned}$$

The process has independent increment increments. A Brownian motion would satisfy these criteria, but although in continuous time there is no restriction that it is a continuous process (it may have jumps).<sup>1</sup> The observed income process is the sum of income

---

<sup>1</sup>Note that such a process will take both positive and negative values, and therefore may not be a good choice for an income process. In appendix \*\*\*\*\* I show that under certain assumptions the same results approximately hold when shocks are multiplicative rather than additive.

over a year:

$$\begin{aligned}\bar{y}_T &= \int_{T-1}^T y_t dt \\ &= \int_{T-1}^T \int_0^t dy_s dt\end{aligned}$$

So that:

$$\begin{aligned}\Delta \bar{y}_T &= \int_{T-1}^T \int_0^t dy_s dt - \int_{T-2}^{T-1} \int_0^t dy_s dt \\ &= \int_{T-1}^T \int_{t-1}^t dy_s dt \\ &= \int_{T-1}^T (T-s) dy_s + \int_{T-2}^{T-1} (s-(T-2)) dy_s\end{aligned}$$

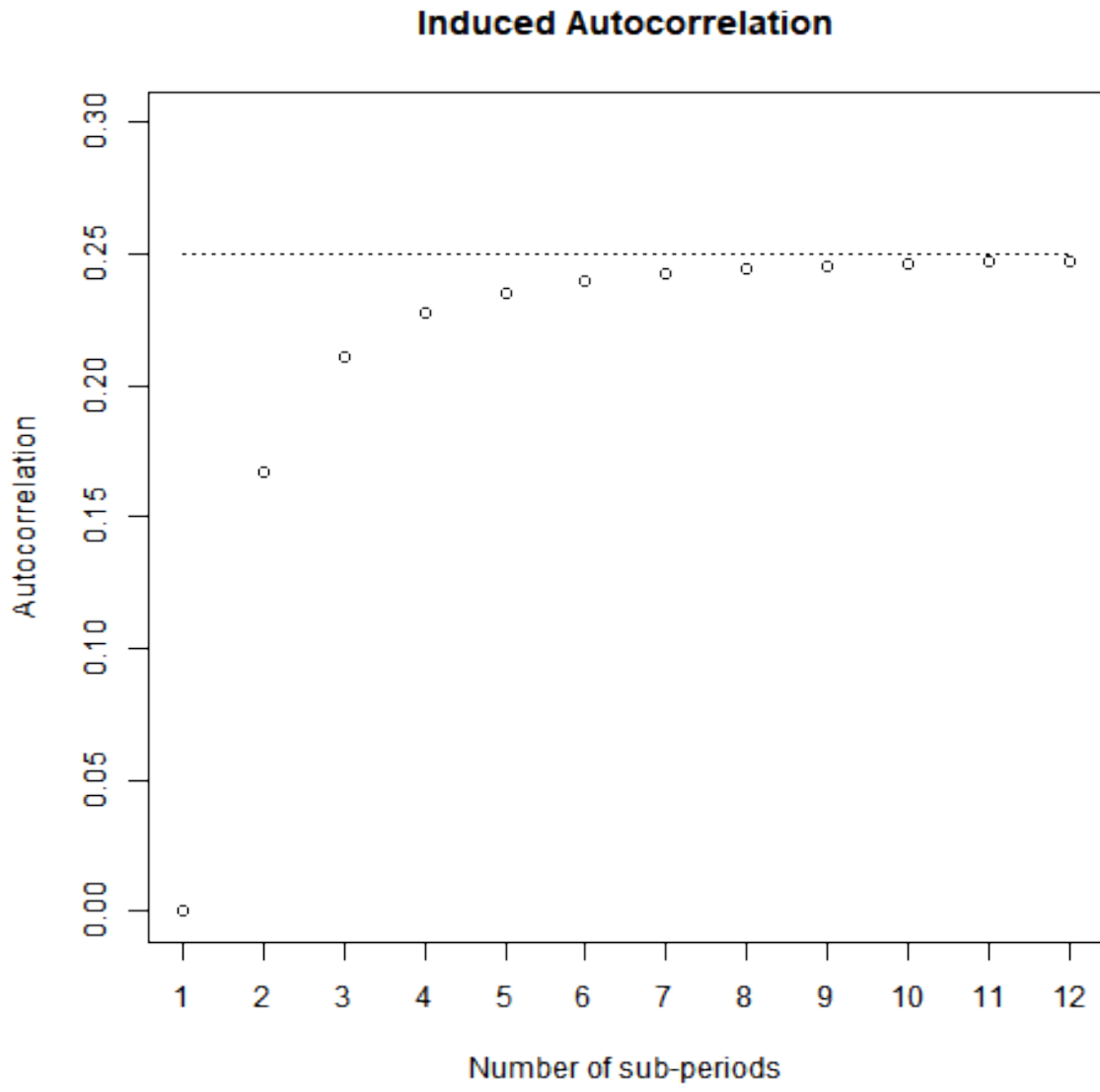
The autocorrelation can now be calculated:

$$\begin{aligned}\text{Cov}(\Delta \bar{y}_T, \Delta \bar{y}_{T-1}) &= \int_{T-2}^{T-1} (T-1-s)(s-(T-2)) \sigma^2 dt \\ &= \frac{1}{6} \sigma^2 \\ \text{Var}(\Delta \bar{y}_T) &= \int_{T-1}^T (T-s)^2 \sigma^2 dt + \int_{T-2}^{T-1} (s-(T-2))^2 \sigma^2 dt \\ &= \frac{2}{3} \sigma^2 \\ \text{Corr}(\Delta \bar{y}_T, \Delta \bar{y}_{T-1}) &= \frac{1}{4}\end{aligned}$$

Which is unsurprisingly the same as the limit of the autocorrelation in the  $N$  sub-periods case. Figure 2 shows how fast the  $N$  sub-period case converges towards the continuous time case. When  $N = 1$  the time aggregated process is the same as the underlying random walk so the autocorrelation is zero. When income is quarterly ( $N = 4$ ) the autocorrelation is 0.23 and is closely approximated by the continuous time model. With monthly ( $N = 12$ ) or higher frequency for income shocks the discrete and continuous models are almost indistinguishable.

### 3 Time Aggregation in Blundell, Pistaferri, and Preston (2008)

**Figure 2** Induced Autocorrelation for different N



## References

- BLUNDELL, RICHARD, LUIGI PISTAFERRI, AND IAN PRESTON (2008): “Consumption Inequality and Partial Insurance,” *American Economic Review*, 98(5), 1887–1921.
- CAMPBELL, JOHN Y., AND N. GREGORY MANKIW (1989): “Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence,” NBER Chapters, National Bureau of Economic Research, Inc.
- WORKING, HOLBROOK (1960): “Note on the Correlation of First Differences of Averages in a Random Chain,” *Econometrica*, 28(4), 916–918.



## Appendix