

In Search of Lost Time Aggregation

By EDMUND CRAWLEY*

Draft: September 29, 2019

In 1960, Working noted that time aggregation of a random walk induces serial correlation in the first difference that is not present in the original series. This important contribution has been overlooked in a recent literature analyzing income and consumption in panel data. I examine Blundell, Pistaferri and Preston (2008) as an important example for which time aggregation has quantitatively large effects. Using new techniques to correct for the problem, I find the estimate for the partial insurance to transitory shocks, originally estimated to be 0.05, increases to 0.24. This larger estimate resolves the dissonance between the low partial consumption insurance estimates of Blundell, Pistaferri and Preston (2008) and the high marginal propensities to consume found in the natural experiment literature.

JEL: C18, D12, D31, D91, E21

Keywords: Income, Consumption, Time Aggregation

* Federal Reserve Board, 20th Street and Constitution Avenue N.W., Washington, DC 20551, edmund.s.crawley@frb.gov. The analysis and conclusions set forth are those of the author alone and do not indicate concurrence by other members of the research staff or the Board of Governors. Many thanks to Chris Carroll for support and guidance.

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 time aggregated series are common¹ and a small literature has grown around how to account for time aggregation in various settings.²

However, the effect of time aggregation has been overlooked in much of the literature studying the covariance structure of household income and consumption dynamics.³ This oversight can result in significant bias. I examine Blundell, Pistaferri and Preston (2008) (henceforth BPP) not only as a way to demonstrate new techniques to overcome the bias, but also because the consumption responses to transitory and permanent income shocks are of significant economic interest in themselves. Indeed, Kaplan and Violante (2010) argue that “the BPP insurance coefficients should become central in quantitative macroeconomics”. Using the same Panel Study of Income Dynamics (PSID) data as in BPP, I update their underlying model to account for time aggregation. I find the estimate for partial insurance to transitory shocks, originally estimated in BPP to be 0.05, to be 0.24 when time aggregation is accounted for. This new estimate resolves the dissonance between BPP’s “full insurance of transitory shocks” and a parallel literature that, using natural experiments, finds large consumption responses to transitory income shocks.⁴ While this paper will focus on the implications of time aggregation for the methodology in BPP, the techniques can be applied to a broad swath of the literature.

¹For an example see Campbell and Mankiw (1989)

²A sample of this literature includes Amemiya and Wu (1972), Weiss (1984) and Drost and Nijman (1993).

³The literature goes back to early work such as Hause (1973), Weiss and Lillard (1979) and MaCurdy (1982) that look at the covariance structure of the income process. Following BPP, a number of papers have looked at income and consumption together, for example Arellano, Blundell and Bonhomme (2017)

⁴A small sample of this literature includes Parker et al. (2013), Agarwal and Qian (2014) and Sahm, Shapiro and Slemrod (2010). Consumers also answer that they have a high marginal propensity to consume when asked, see Fuster, Kaplan and Zafar (2018) and Jappelli and Pistaferri (2014). For an overview of the entire literature on consumption responses to income shocks, see Jappelli and Pistaferri (2010). Note the dissonance between BPP and the natural experiment literature is also addressed by Commault (2017). In contrast to this paper, her approach makes structural changes to the underlying model but does not address time aggregation.

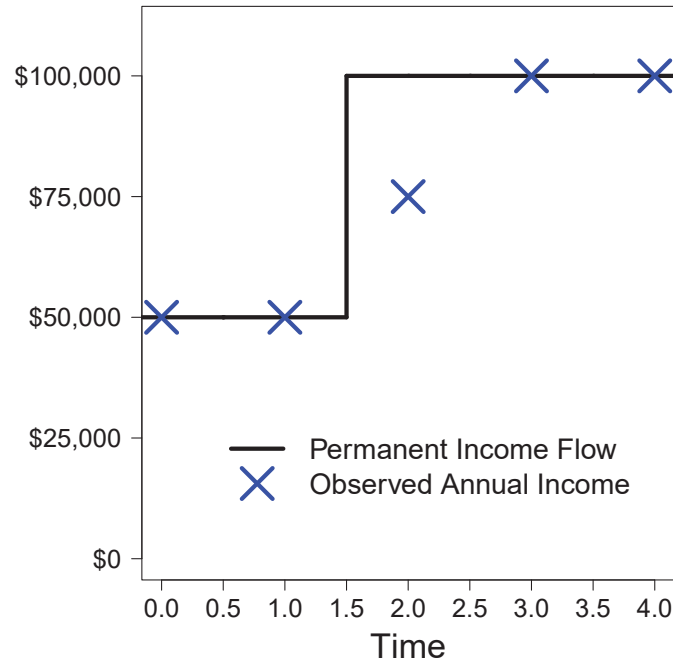


FIGURE 1. INCOME FLOW AND OBSERVED INCOME

I. What is Time Aggregation?

Time aggregation occurs when a time series is observed at a lower frequency than the underlying data that generates it. 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. To transform income into an annual frequency, we sum up all the income that was received by a household during the year. The key insight of Working (1960) is that even if there is no correlation between changes in income at the underlying frequency, changes in the resulting time aggregated series will show positive autocorrelation. The intuition behind this can be seen in figure 1, showing the income process of a household that begins with an annual salary of \$50,000 and receives a permanent pay rise to \$100,000 mid-way through the second year. The solid line shows this jump in income

flow occurring just once. The crosses show the income we actually observe in annual data. During the second year the household receives an annual \$50,000 salary for six months, followed by \$100,000 in the second six months, resulting in a reported income of \$75,000 for the entire year. The single shock to income therefore appears in the time aggregated data as two increases. In this way, an income change in one year is positively correlated with an income change in the following year, even if the underlying income process follows a random walk.

II. Modelling Time Aggregation in Blundell, Pistaferri and Preston (2008)

A. The Model in Discrete Time Without Time Aggregation

Here I briefly describe the method used by Blundell, Pistaferri and Preston (2008) to estimate household consumption responses to permanent and transitory income shocks. The model described here is a simplified version of the original in order to highlight the role played by time aggregation.⁵

The core of the model is the assumptions made on the income and consumption processes. Unexplained log income growth for household i follows the process:

$$\Delta y_{i,t} = \zeta_{i,t} + \Delta \nu_{i,t}$$

where $\zeta_{i,t}$ (the change in permanent income) and $\nu_{i,t}$ (transitory income) are each mean zero, finite variance, i.i.d. and independent of each other.

The unexplained change in log consumption is modeled as a random walk that moves in response to permanent and transitory income shocks:

$$\Delta c_{i,t} = \phi \zeta_{i,t} + \psi \nu_{i,t}$$

where ϕ and ψ are the *partial insurance* parameters. A value of zero implies full

⁵In this simplified model I assume insurance parameters are constant across both time and households, that the transitory component of income has no persistence, and that there are no taste shocks. These elements are reintroduced in section III in which I show the quantitative effect of time aggregation.

insurance (consumption does not respond at all to the income shock), while a value of one implies no insurance. The core of the empirical methodology is to identify these insurance parameters in the data from the following identities:

$$(1) \quad \phi = \frac{\text{Cov}(\Delta c_t, \Delta y_{t-1} + \Delta y_t + \Delta y_{t+1})}{\text{Cov}(\Delta y_t, \Delta y_{t-1} + \Delta y_t + \Delta y_{t+1})}$$

$$(2) \quad \psi = \frac{\text{Cov}(\Delta c_t, \Delta y_{t+1})}{\text{Cov}(\Delta y_t, \Delta y_{t+1})}$$

B. The Model in Continuous Time with Time Aggregation

In this section I show how time aggregation can significantly bias the partial insurance parameter estimates obtained by equations 1 and 2. The model in this section will be the exact analog of the discrete time model just described, but embedded in continuous time where shocks are spread uniformly throughout the year.⁶ The main result does not hinge on the use of continuous time, and similar estimates would be obtained by dividing the year into quarters or months.⁷

Time is continuous and one time unit represents one year. For the income process we will assume two underlying martingale processes, P_t and Q_t such that for all $s_1 > s_2 > s_3 > s_4 > 0$:

$$\text{Var}(P_{s_1} - P_{s_2}) = (s_1 - s_2)\sigma_P^2$$

$$\text{Cov}(P_{s_1} - P_{s_2}, P_{s_3} - P_{s_4}) = 0$$

$$P_s = 0 \quad \text{if } s < 0$$

and similarly for Q_t . Brownian motion fits these assumptions, but the slightly more general definition allows for jumps in the income process, such as getting

⁶There is little formal evidence on the distribution of shocks throughout the year. While this assumption is unlikely to be strictly true, it is more reasonable than the implicit assumption of BPP that shocks all occur 1st January each year.

⁷The autocorrelation of a time aggregated random walk is 0.25 in continuous time, compared to 0.23 for a discrete quarterly model and almost indistinguishable from a discrete monthly model. The theoretical moments are however significantly more elegant in continuous time. See online appendix B.B1

promoted. Instantaneous income in a period dt is given by:⁸

$$(3) \quad dy_t = P_t dt + dQ_t$$

that is they receive their permanent income flow ($P_t = \int_0^t dP_s$) multiplied by time dt in addition to a one-off transitory income dQ_t .

Keeping with the assumption that consumption is a random walk with insurance parameters ϕ and ψ , instantaneous consumption is given by:

$$(4) \quad dc_t = \phi P_t dt + \psi Q_t dt$$

that is, they consume a proportion ϕ of their permanent income and a proportion ψ of the cumulation of all the transitory income they have received in their lifetime ($Q_t = \int_0^t dQ_s$).

In the Panel Study of Income Dynamics (PSID) data, we observe the total income received over the previous calendar year at time T :

$$y_T^{obs} = \int_{T-1}^T dy_t$$

Consumption is measured by a survey at the beginning of the following calendar year, which I map to a snapshot of consumption exactly at the end of the calendar year:⁹

$$(5) \quad c_T^{obs} = \phi P_T + \psi Q_T$$

⁸A more formal treatment of how to relate this to the log income process is given in online appendix B.B1.

⁹BPP use data on food consumption to impute total annual consumption. The questionnaire asks about food consumption in a typical week, but unfortunately the timing of this ‘typical week’ is not clear. The questionnaire is usually given at the end of March in the following year. See Altonji and Siow (1987) and Hall and Mishkin (1982) for differing views. In online appendix B.B3 I show that the timing of the ‘typical’ week can have a large effect on the results. This is an important drawback to using this method with the PSID data. In Crawley and Kuchler (2018) we use expenditure data imputed from Danish administrative records in which the timing of expenditure is very clearly defined.

The BPP method makes use of the changes in observable income and consumption, which in the time aggregated model relate to:

$$\Delta y_T^{obs} = \left(\int_{T-2}^{T-1} (s - (T-2)) dP_s + \int_{T-1}^T (T-s) dP_s \right) + \left(\int_{T-1}^T dQ_t - \int_{T-2}^{T-1} dQ_t \right) \quad (6)$$

$$\Delta c_T^{obs} = \phi \int_{T-1}^T dP_s + \psi \int_{T-1}^T dQ_s \quad (7)$$

We see that these observable income and consumption changes in equations 1 and 2 recover the permanent, but not the transitory insurance parameter:

$$(8) \quad \frac{\text{Cov}(\Delta c_T^{obs}, \Delta y_{T-1}^{obs} + \Delta y_T^{obs} + \Delta y_{T+1}^{obs})}{\text{Cov}(\Delta y_T^{obs}, \Delta y_{T-1}^{obs} + \Delta y_T^{obs} + \Delta y_{T+1}^{obs})} = \phi$$

$$(9) \quad \frac{\text{Cov}(\Delta c_T^{obs}, \Delta y_{T+1}^{obs})}{\text{Cov}(\Delta y_T^{obs}, \Delta y_{T+1}^{obs})} = \psi - \frac{(3\phi - \psi)\sigma_P^2}{6\sigma_Q^2 - \sigma_P^2}$$

Indeed the transitory insurance coefficient bears little relation to the true value of ψ . For example, if permanent and transitory variances are equal, and households follow the permanent income hypothesis ($\phi = 1, \psi = 0$), the estimate for ψ using this method will be *negative* 0.6.

III. Revised BPP Estimates

In this section I repeat the BPP estimation procedure, but with the model moments coming from the continuous time model with time aggregated income. While the core identification in BPP is illustrated in equations 1 and 2, the full estimation procedure minimizes the distance between all the observable covariances ($\text{Cov}(\Delta y_T^{obs}, \Delta y_S^{obs})$, $\text{Cov}(\Delta c_T^{obs}, \Delta c_S^{obs})$ and $\text{Cov}(\Delta c_T^{obs}, \Delta y_S^{obs})$) and their model implied equivalents.¹⁰ The full set of these model implied moments for the continuous time model, extended to include time varying coefficients, transitory

¹⁰I follow the exact same diagonally weighted minimum distance procedure in BPP as described in online appendix D of Blundell, Pistaferri and Preston (2008)

persistence and taste shocks, can be found in appendix A.A1 and online appendix B.B2.

TABLE 1—MINIMUM-DISTANCE PARTIAL INSURANCE AND VARIANCE ESTIMATES

Persistence Type:	BPP		Time Agg.		
	None	MA(1)	None	Uniform	Linear Decay
ψ	0.0503	0.0501	0.2421	0.2510	0.2403
(Partial insurance tran. shock)	(0.0505)	(0.0430)	(0.0431)	(0.0428)	(0.0417)
ϕ	0.4692	0.6456	0.3384	0.3287	0.3516
(Partial insurance perm. shock)	(0.0598)	(0.0941)	(0.0471)	(0.0580)	(0.0627)

Table 1 shows the estimates for the transitory and permanent insurance parameters, first using BPP’s original method and then with time aggregation. As there is no equivalent to an MA(1) process in continuous time, I consider two alternative ways to introduce persistence in the transitory shock, as well as reporting results assuming no persistence. First I assume a transitory shock provides a stream of income uniformly distributed over a short period of time (to be estimated). Second I assume the stream of income decays linearly over a short period.¹¹ The time aggregated results are not very sensitive to these assumptions.

The top row of table 1 gives the main result showing the transitory insurance parameter increases from 0.05 in BPP to 0.24 with time aggregation. This new estimate is much more in line with the literature that estimates MPCs using natural experiments.

Note that this large consumption response to transitory shocks, along with a lifetime budget constraint, is incompatible with the assumption common in both models that consumption moves as a random walk. This misspecification perhaps explains why the permanent insurance parameter appears to be relatively small in the time aggregated model. In Crawley and Kuchler (2018) we suggest ways to approach this.

¹¹See online appendix B.B2 for details.

IV. Conclusion

This paper highlights the importance of time aggregation when working with panel data, especially when analyzing the covariance matrix of income and consumption growth. It also resolves the dissonance between BPP's estimates of transitory income insurance and the natural experiment literature on marginal propensity to consume. Going forward, I hope the methods used here to correct for the time aggregation problem can be useful for researchers, especially as more and more high quality panel datasets on income and consumption become available.

REFERENCES

- Agarwal, Sumit, and Wenlan Qian.** 2014. "Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore." *American Economic Review*, 104(12): 4205–4230.
- Altonji, Joseph G., and Aloysius Siow.** 1987. "Testing the Response of Consumption to Income Changes with (Noisy) Panel Data." *The Quarterly Journal of Economics*, 102(2): 293–328.
- Amemiya, Takeshi, and Roland Y. Wu.** 1972. "The Effect of Aggregation on Prediction in the Autoregressive Model." *Journal of the American Statistical Association*, 67(339): 628–632.
- Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme.** 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica*, 85(3): 693–734.
- 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." National Bureau of Economic Research, Inc NBER Chapters.
- Commault, Jeanne.** 2017. "How Does Consumption Respond to a Transitory Income Shock? Reconciling Natural Experiments and Structural Estimations."
- Crawley, Edmund, and Andreas Kuchler.** 2018. "Consumption Heterogeneity: Micro Drivers and Macro Implications."
- Drost, Feike C., and Theo E. Nijman.** 1993. "Temporal Aggregation of Garch Processes." *Econometrica*, 61(4): 909–927.
- Fuster, Andreas, Greg Kaplan, and Basit Zafar.** 2018. "What Would You Do With \$500? Spending Responses to Gains, Losses, News and Loans." National Bureau of Economic Research Working Paper 24386.
- Hall, Robert, and Frederic Mishkin.** 1982. "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica*, 50(2): 461–81.
- Hause, John.** 1973. "The Covariance Structure of Earnings and the On the Job Training Hypothesis." National Bureau of Economic Research Working Paper 25.
- Jappelli, Tullio, and Luigi Pistaferri.** 2010. "The Consumption Response to Income Changes." *Annual Review of Economics*, 2(1): 479–506.
- Jappelli, Tullio, and Luigi Pistaferri.** 2014. "Fiscal Policy and MPC Heterogeneity." *American Economic Journal: Macroeconomics*, 6(4): 107–136.
- Kaplan, Greg, and Giovanni L. Violante.** 2010. "How Much Consumption Insurance beyond Self-Insurance?" *American Economic Journal: Macroeconomics*, 2(4): 53–87.

- MaCurdy, Thomas E.** 1982. “The use of time series processes to model the error structure of earnings in a longitudinal data analysis.” *Journal of Econometrics*, 18(1): 83 – 114.
- Parker, Jonathan A, Nicholas S Souleles, David S Johnson, and Robert McClelland.** 2013. “Consumer spending and the economic stimulus payments of 2008.” *The American Economic Review*, 103(6): 2530–2553.
- Sahm, Claudia R., Matthew D. Shapiro, and Joel B. Slemrod.** 2010. “Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications.” *Tax Policy and the Economy*, 24: 69–110.
- Weiss, Andrew A.** 1984. “Systematic sampling and temporal aggregation in time series models.” *Journal of Econometrics*, 26(3): 271–281.
- Weiss, Yoram, and Lee Lillard.** 1979. “Components of Variation in Panel Earnings Data: American Scientists 1960-70.” *Econometrica*, 47: 437–54.
- Working, Holbrook.** 1960. “Note on the Correlation of First Differences of Averages in a Random Chain.” *Econometrica*, 28(4): 916–918.

MATHEMATICAL APPENDIX

A1. Identification in the Full Model

In this appendix I calculate the full set of identifying equations for the non-stationary model with measurement error in consumption and taste shocks. Online appendix B.B2 extends these to add persistence in the transitory shock.

I am interested in the full set of observable covariances:

$$\text{Cov}(\Delta y_T^{obs}, \Delta y_S^{obs})$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta c_S^{obs})$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta y_S^{obs})$$

for all T and S in $\{1, 2, \dots\}$. I further make the assumption that while the variance of the permanent and transitory shocks and insurance coefficients can change from year to year, within each year these are constant. The variance the permanent shock in year T is $\sigma_{P,T}^2$ and the transitory shock $\sigma_{Q,T}^2$. I use equation 6 for the change in observable log income, and extend equation 7 for the change in observable log consumption to include taste shocks (ξ_t) and measurement error (u_T):

$$\Delta c_T^{obs} = \phi \int_{T-1}^T dP_s + \psi \int_{T-1}^T dQ_s + \int_{T-1}^T d\xi_s + u_T - u_{T-1}$$

These two equations allow for the calculation of all the required identification equations:

$$\begin{aligned} \text{Var}(\Delta y_T^{obs}) &= \mathbb{E} \left(\int_{T-2}^{T-1} (s - (T-2))^2 dP_s dP_s + \int_{T-1}^T (T-s)^2 dP_s dP_s \right) \\ &\quad + \mathbb{E} \left(\int_{T-1}^T dQ_t dQ_t + \int_{T-2}^{T-1} dQ_t dQ_t \right) \\ (A1) \quad &= \frac{1}{3} \sigma_{P,T}^2 + \frac{1}{3} \sigma_{P,T-1}^2 + \sigma_{Q,T}^2 + \sigma_{Q,T-1}^2 \\ \text{Cov}(\Delta y_T^{obs}, \Delta y_{T+1}^{obs}) &= \mathbb{E} \left(\int_{T-1}^T (T-s)(s - (T-1)) dP_s dP_s \right) - \mathbb{E} \left(\int_{T-1}^T dQ_t dQ_t \right) \\ &= \frac{1}{6} \sigma_{P,T}^2 - \sigma_{Q,T}^2 \\ \text{Cov}(\Delta y_T^{obs}, \Delta y_{T-1}^{obs}) &= \frac{1}{6} \sigma_{P,T-1}^2 - \sigma_{Q,T-1}^2 \\ \text{Cov}(\Delta y_T^{obs}, \Delta y_S^{obs}) &= 0 \quad \forall S, T \text{ such that } |S - T| > 1 \end{aligned}$$

$$\begin{aligned}
\text{Var} \Delta c_T^{obs} &= \phi^2 \mathbb{E} \left(\int_{T-1}^T dP_s dP_s \right) + \psi^2 \mathbb{E} \left(\int_{T-1}^T dQ_s dQ_s \right) + \mathbb{E} \left(\int_{T-1}^T d\xi_s d\xi_s \right) + \sigma_{u,T}^2 + \sigma_{u,T-1}^2 \\
&= \phi^2 \sigma_{P,T}^2 + \psi^2 \sigma_{Q,T}^2 + \sigma_{\xi,T}^2 + \sigma_{u,T}^2 + \sigma_{u,T-1}^2
\end{aligned}$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta c_{T+1}^{obs}) = -\sigma_{u,T}^2$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta c_{T-1}^{obs}) = -\sigma_{u,T-1}^2$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta c_S^{obs}) = 0 \quad \forall S, T \text{ such that } |S - T| > 1$$

$$\begin{aligned}
\text{Cov}(\Delta c_T^{obs}, \Delta y_T^{obs}) &= \mathbb{E} \left(\phi_T \int_{T-1}^T (T-s) dP_s dP_s + \psi_T \int_{T-1}^T dQ_s dQ_s \right) \\
&= \frac{1}{2} \phi_T \sigma_{P,T}^2 + \psi_T \sigma_{Q,T}^2
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\Delta c_T^{obs}, \Delta y_{T+1}^{obs}) &= \mathbb{E} \left(\phi_T \int_{T-1}^T (s - (T-1)) dP_s dP_s - \psi_T \int_{T-1}^T dQ_s dQ_s \right) \\
&= \frac{1}{2} \phi_T \sigma_{P,T}^2 - \psi_T \sigma_{Q,T}^2
\end{aligned}$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta y_{T-1}^{obs}) = 0$$

$$\text{Cov}(\Delta c_T^{obs}, \Delta y_S^{obs}) = 0 \quad \forall S, T \text{ such that } |S - T| > 1$$