The Fade-Away Effect of Initial Nonresponse in Panel Surveys: Some Theoretical Ideas and Empirical Results from PASS

Juha Alho*, Gerrit Müller†, Verena Pflieger‡, Ulrich Rendtel§

Version 4th August 2015

Very preliminary! Please do not circulate!

1 Introduction

Panel surveys are plagued by nonresponse at their initial wave, which may be substantial, and also by panel attrition, which is typically smaller but cumulating over time. Besides the loss in terms of case numbers or statistical power, this cumulation is often thought to aggravate initial selective effects. Such a view of a permanent or even worsening nonresponse bias results from a static view on the variables of interest. If, for example, nonresponse depends on gender, which changes only very seldom, an initial nonresponse bias with respect to gender will never vanish in later panel waves; unless of course, attrition is such that it exactly counteracts the selective effects present at the start of the panel. The same view may also hold for variables which change only slowly over time like the highest level of attained education. In these cases calibration methods may help to adjust for nonresponse bias, see Särndal (2007) for a general overview and Rendtel/Harms (2008) for calibration in panel surveys.

However, most panel surveys are launched to observe and analyse the change of variables. Characteristics that are linked to income and poverty, for instance, are much less stable over time and there is a considerable exchange between e.g. the states "poor" and "non-poor", see Rendtel (2013) based on EU-SILC data. So even if there is a substantial over-representation of poor people in the first wave of the panel, it will happen that "poor" become "non-poor" and vice versa. This general turn-over has the potential to let the initial nonresponse bias present at the start of the panel "fade away" over time.

In order to check for such a behaviour empirically, it is necessary to have information about the variables of interest for both respondents and nonrespondents, which are typically not at hand. If, however, a panel survey is sampled from a register, then it is possible to use linked register information. Here, Sisto (2003) and Rendtel (2013) report a rapid decline of initial nonresponse bias between income quintiles (Finnish subsample of the ECHP) and poverty states (Finnish subsample of EU-SILC) in later panel waves. Rendtel (2013) coined the phenomenon as the "fade-away effect" of initial nonresponse.

A Markov chain approach is used here to derive a general contraction theorem in the case of non-homogeneous transitions. In the case of time-homogeneous transitions we can

*University of Helsinki
†IAB, Nürnberg
‡Institute for Statistics and Econometrics, FU Berlin
§Institute for Statistics and Econometrics, FU Berlin

use the convergence to a steady state distribution to assess the strength of the fade-away effect. This theoretical framework is applied for the German Panel on "Labour Market and Social Security" (PASS= Panel Arbeitsmarkt und soziale Sicherung), which is a panel that is linked to the registers of the German Social Security files.

2 The Methodological Framework of Markov Chains

We assume that the characteristic of interest $\{Y_t\}_{t\in\mathbb{N}}$ follows a Markov chain with state space $S = \{1, \dots, I\}$:

$$\begin{aligned} P(Y_t = j|Y_{t-1} = i, Y_{t-2} = s_{t-2}, \dots, Y_1 = s_1) &= P(Y_t = j|Y_{t-1} = i) \\ &= p_{i,j}(t) \end{aligned}$$

Denote the transition law for transitions to wave t by $P(t) = (p_{i,j}(t))_{(i,j=1,\dots,I)}$. The transition law from wave $t-1$ to wave t may be time-inhomogeneous or time-homogeneous. In the time-homogeneous case the transition law is denoted by $P = (p_{i,j})_{(i,j=1,\dots,I)}$. Transitions from wave 1 to wave t_0 are denoted by $P^{(t_0)}$. One obtains $P^{(t_0)} = P(2)P(3)\dots P(t_0)$. In the case of time-homogeneous chains we have $P^{(t_0)} = P^{t_0-1}$.

2.1 A Contraction Theorem in the Case of Time-Inhomogeneous Markov Chains

Our application of Markov chain theory uses two different starting distributions. Let $\pi_I(1)$ be the starting distribution on the state space for the gross-sample of the panel at wave one. We refer to this sample as the $FULL$-sample. There will be a second Markov chain with a different starting distribution $\pi_{II}(1)$, which refers to the net-sample of the panel at wave one. We refer to this sample as the $RESP$ sample.

One essential of our approach is that the transition law for respondents and non-respondents is identical. We refer to this assumption as **Assumption A**. One motivation for this assumption is that the socio-economic laws that drive individual changes over time are not affected by the interviewee's decision to participate in a survey or not. However, one could imagine that the decision on further participation in a panel rests on future views about labour market development. **Assumption A** cannot be verified from the respondent data alone. Therefore it is a typical Missing at Random assumption (MAR) in the sense of Rubin (1976). However, if the panel sample is recruited from a register and if there is access to key variables from the register then it is possible to test **Assumption A**.

The distribution of the two Markov chains at wave t is computed in a sequential fashion by $\pi_I'(t) = \pi_I'(t-1)P(t)$ and $\pi_{II}'(t) = \pi_{II}'(t-1)P(t)$. If all entries of $\pi_{II}(t)$ are positive we may compute:

$$m_t = \min_i \frac{\pi_{I,i}(t)}{\pi_{II,i}(t)} \leq \frac{\pi_{I,i}(t)}{\pi_{II,i}(t)} \leq \max_i \frac{\pi_{I,i}(t)}{\pi_{II,i}(t)} = M_t \tag{1}$$

The following contraction theorem states that the two distributions $\pi_I(t)$ and $\pi_{II}(t)$ converge under some regularity conditions as long as the two chains underlay the same transitions laws $P(t)$:

Theorem 2.1. *Suppose that the positive entries of the $P(t)$ are uniformly bounded away from 0, i.e. there exists a $p_l > 0$ with:*

$$0 < p_l \leq p_{i,j}(t) \leq 1 \text{ for all positive elements of } P(t) \tag{2}$$

Suppose further that there exists a t_0 such that:

$$P(2)\dots P(t_0) > 0 \tag{3}$$

Then $\pi_I(t)$ and $\pi_{II}(t)$ converge uniformly in the sense of

$$\lim_{t\to\infty}(M_t - m_t) = 0 \tag{4}$$

A proof of this theorem is given in the appendix of the full-length paper. Note that the regularity conditions of equations 2 and 3 are not serious restrictions for empirical work.

The important consequence of this theorem is that a potential nonresponse bias in the first panel wave measured as a difference of the $FULL$-sample and the $RESP$-sample tends to disappear in later panel waves.

2.2 Regularity Conditions for the Attrition Process

So far we have fixed a framework under which the distributions π^{FULL} and π^{RESP} converge in later panel waves. However, π^{RESP} is further affected by panel attrition. Denote the sample of observed units at wave t by OBS_t.

We have to make a statement about the attrition process. Intuitively it is clear that panel attrition may counteract the convergence with respect to the Markov chain. Thus panel attrition must not be selective with respect to the variable of interest. This issue is treated more formally for the case of a panel with four waves. From this case the results may be easily extrapolated to longer panels. We use the response indicators R_1, R_2, R_3, R_4, where $R_t = 1$ indicates response and $R_t = 0$ indicates nonresponse at wave t. Furthermore we use the Y_t which indicate the state at wave t with $t = 1, \dots, 4$. The distribution in the OBS_t-sample at wave 4 is $P(Y_4 = j_4|R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1)$.
Now we have:

$$\begin{aligned}
& P(Y_4 = j_4|R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) \\
= & \sum_{j_3} P(Y_4 = j_4|Y_3 = j_3, R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) \\
\times & P(Y_3 = j_3|R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) && (5) \\
= & \sum_{j_3} P(Y_4 = j_4|Y_3 = j_3, R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) \\
\times & \frac{P(R_4 = 1|Y_3 = j_3, R_1 = 1, R_2 = 1, R_3 = 1)}{P(R_4 = 1|R_1 = 1, R_2 = 1, R_3 = 1)} \\
\times & P(Y_3 = j_3|R_1 = 1, R_2 = 1, R_3 = 1) && (6)
\end{aligned}$$

In order to proceed we need **Assumption A** that the transition behavior must not depend on the participation behavior:

$$P(Y_4 = j_4|Y_3 = j_3, R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) = P(Y_4 = i|Y_3 = j_3) \tag{7}$$

Furthermore we need **Assumption B** stating that the previous state does not have a direct effect on the participation in the present wave:

$$P(R_4 = j_4|Y_3 = j_3, R_1 = 1, R_2 = 1, R_3 = 1) = P(R_4 = 1|R_1 = 1, R_2 = 1, R_3 = 1) \tag{8}$$

By using assumptions **A** and **B** one gets:

$$\begin{aligned} & P(Y_4 = j_4 | R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) \\ = & \sum_{j_3} P(Y_4 = i | Y_3 = j_3) P(Y_3 = j_3 | R_1 = 1, R_2 = 1, R_3 = 1) \end{aligned} \tag{9}$$

Using the same kind of analysis for $P(Y_3 = j_3 | R_1 = 1, R_2 = 1, R_3 = 1)$ and inserting into eq. 9 one obtains:

$$\begin{aligned} & P(Y_4 = j_4 | R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) \\ = & \sum_{j_3, j_2} P(Y_4 = i | Y_3 = j_3) P(Y_3 = j_3 | Y_2 = j_2) P(Y_2 = j_2 | R_1 = 1, R_2 = 1) \end{aligned} \tag{10}$$

Finally we arrive at:

$$\begin{aligned} & P(Y_4 = j_4 | R_1 = 1, R_2 = 1, R_3 = 1, R_4 = 1) \\ = & \sum_{j_3, j_2, j_1} P(Y_4 = i | Y_3 = j_3) P(Y_3 = j_3 | Y_2 = j_2) P(Y_2 = j_2 | Y_1 = j_1) \\ \times & P(Y_1 = j_1 | R_1 = 1) \end{aligned} \tag{11}$$

where the last term $P(Y_1 = j_1 | R_1 = 1)$ is the starting distribution for the respondents of wave 1 and the summation is done over 3 cycles of the Markov chain on the state space. According to our contraction theorem the right side of equation 11 converges uniformly to the distribution of the Markov chain which started from the $FULL$-sample.

Assumption **B** and the corresponding expressions for wave 3 and 2 may be regarded as restrictive, as it states that attrition must not be linked to the state of the previous period. It should be noted, however, that assumption **B** can be directly verified from the observed data of the panel. Rendtel (2015) simulated different attrition scenarios with Finnish SILC data and reported that differences up to 10 percentage points in nonresponse propensity do not affect the fade-away effect substantially.

2.3 The "Strength" of the Fade-Away Effect.

The contraction theorem makes no statement about the speed of convergence or the strength of the fade-away effect. However, to be of practical importance it is essential to know whether one has to wait 2 or 3, or 20 panel waves, for a substantial fade-away of an initial nonresponse bias. Generally speaking, the more turn-over there is in the state space, the faster will be the convergence.

In the case of a time-homogeneous Markov chain the famous Perron-Frobenius Theorem allows us to make a statement about the convergence rate. A transition matrix P is called ergodic if there exists a $t_0 \in \mathbb{N}$ such that all $p_{i,j}(t_0) > 0$ for all $i, j = 1, \ldots, I$. Note that Equation 3 in the contraction theorem exactly refers to ergodicity. Also the second precondition of the theorem in equation 2 hold trivially for a time-homogeneous Markov chain. Thus the contraction theorem does hold for any ergodic Markov chain. However, we get even more:

Theorem 2.2. *Let $\{Y_t\}_{t \in \mathbb{N}}$ an ergodic Markov chain with state space S and IxI transition matrix P. If all eigenvalues λ of an ergodic matrix P are disjoint and ordered such that $|\lambda_1| > \ldots > |\lambda_N|$ then it holds:*

There exists a steady state distribution:

$$\pi^* = (\pi_1^*, \ldots, \pi_I^*)' \text{ with } (\pi^*)'P = (\pi^*)' \tag{12}$$

and the convergence to the steady state distribution follows a geometrical pattern:

$$|p_{ij}^{(t)} - \pi_j^*| = O(|\lambda_2|^t) \text{ for all } i, j \in S, t \in \mathbb{N} \tag{13}$$

For a proof, see Seneta (1980).

In the application part we will meet a situation where the distribution of the gross-sample $\pi^{FULL} = (\pi_1^{FULL}, \ldots, \pi_I^{FULL})'$ and the net-sample of wave 1 $\pi^{RESP} = (\pi_1^{RESP}, \ldots, \pi_I^{RESP})'$ will be far away from the steady state distribution. Yet the differences $D_j(t)$ between the two distributions converge to 0 in a geometric fashion. This is due to the above theorem and the triangle inequality (proof in full-length version of the paper). Thus, even if we are far away from the steady state distribution, the contraction property of the Markov chain works with the same rate of convergence.

3 Data Base and Empirical Findings

As mentioned in the introduction, the data used for the empirical examples are from the PASS panel study, which is one of the most comprehensive annual household surveys in Germany in the field of labor market, welfare state and poverty research. PASS is specifically designed to assess the dynamics of a new means-tested welfare benefit scheme, called Unemployment Benefit II (henceforth: UBII), and introduced in 2005 as part of major reform of the German welfare system. We shall focus on the wave 1 “recipient subsample” which is a random sample of benefit units drawn directly from the registry of welfare recipients housed at the Federal Employment Agency (FEA). For both responding and nonresponding cases of this subsample we have available linked register data on UBII receipt covering waves 1-5.

Based on this, the presentation at the NR-Workshop will display some empirical findings on the extent of initial nonresponse bias in UBII receipt and its’ fade-away over time and present tests of assumption A and B.

References

Rendtel, Ulrich (2013): The fade-away effect of initial nonresponse in panel surveys: Empirical results for EU-SILC. Eurostat Methodologies and Workingpapers (ISSN 1977-0375) Edition 2013, doi:10.2785/21863

Rendtel, Ulrich (2015): The Fade-away effect in EU-SILC, Unpublished manuscript

Rendtel, Ulrich; Harms, Torsten (2008): Weighting and calibration for household panels, In: (P. Lynn, ed.) ”Methodology of Longitudinal Surveys”, 265–286, Wiley.

Rubin, Donald (1976): Inference and Missing data, Biometrica, 63, 581–592.

Särndal, Carl-Eric (2007): The calibration approach in survey theory and practice. Survey Methodology, 33, 99–119.

Seneta, Eugene (1980): Non-negative Matrices and Markov Chains, Springer.

Sisto, J.(2003): Attrition Effects on the Design Based Estimates of Disposable Household Income. Chintex Working Paper no. 9 03/2003 `http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/Chintex/`