Fade-Away of Initial Bias in Longitudinal Surveys

Juha Alho[1], Gerrit Müller[2], Verena Pflieger[3], Ulrich Rendtel[4]

Version: April 2016

Very preliminary! Please do not circulate!

1 Introduction

Longitudinal surveys are plagued by nonresponse not only at their start but also in later phases of the study. For example, in panel surveys the nonresponse at the initial wave may be aggravated by attrition in later panel waves. The attrition can be caused, e.g., by non-cooperation or failure to follow up residential movers (Watson and Wooden 2009). When such losses are documented in a cumulative fashion (as, e.g., for the German Socio Economic Panel (SOEP) by Kroh (2011)), the impression one easily gets is that the cumulative effect of nonresponse on case numbers is most likely accompanied by increased bias, as well.

However, this need not always be the case. The hypothesis of a permanent or aggravating nonresponse bias results from a static view of the variables of interest. To be precise, if nonresponse depends on gender, a nonresponse bias with respect to gender will not vanish in later panel waves. But, many longitudinal surveys are launched to observe and analyse the change or the stability of dynamic variables, such as income or poverty (cf., Atkinson and Marlier (2010) for the European Union Statistics of Income and Living Conditions (EU-SILC)). These characteristics are unstable over time, and there can be considerable exchange between the states "poor" and "non-poor",see for example Rendtel (2013). Therefore, even if there is a substantial overrepresentation of poor people in the first wave of the panel, it will happen that "poor" become "non-poor" and vice versa. This general turnover has the potential to reduce the selective non-response effect observed at the start of the panel. Rendtel (2013) coined the term "fade away effect" for this phenomenon.

There are two essential assumptions of the Markov approach presented here, (1) the state transitions of the respondents and the nonrespondents between panel waves follow the same Markov process, and (2) an individual's probability of responding at later waves must not depend on the state one is currently in.

In order to check for the validity of such assumptions it is necessary to have information about the variables of interest for both the respondents and nonrespondents. The latter information is typically not at hand. However, if participants for a panel survey are sampled from a register, then it may be possible to use the register information on income and labour participation from the register. For example, Sisto (2003) and Rendtel (2013) report a rapid decline of a nonresponse bias on income quintiles (Finnish subsample of the European Community Household Panel (ECHP)) and poverty states (Finnish subsample of EU-SILC) in later panel waves.

We present a general contraction theorem that applies to non-homogeneous

[1] University of Helsinki
[2] IAB, Nürnberg
[3] fu:stat, Berlin
[4] Institute for Statistics and Econometrics, FU Berlin

weakly ergodic Markov chains. Alternatively, the result applies also to deterministic matrix models of population evolution. This gives us two complementary ways to view the time evolution of the panel waves. In the case of time-homogeneous transitions a steady state distribution exists in both cases, and we can use the convergence to the steady state to assess the speed of the fade away effect.

This theoretical framework is applied to the German Panel on Labour Market and Social Security PASS (acronym for Panel Arbeitsmarkt und soziale Sicherung), which is a panel that is linked to the registers of the German Social Security files.

2 Markov chains and related matrix models

We consider the evolution of a population in discrete time, $t = 0, 1, 2, \ldots$, as disaggregated into a finite number of states $S = \{1, \ldots, I\}$. The members of the population are assumed to move from state to state as a Markov process,

$$\begin{aligned} P(Y_t = j | Y_{t-1} = i, Y_{t-2} = s_{t-2}, \ldots, Y_0 = s_0) &= P(Y_t = j | Y_{t-1} = i) \\ &= p_{i,j}(t) \end{aligned}$$

The $I \times I$ matrix of transition probabilities from time $t-1$ to time t is $P(t) = (p_{i,j}(t))$. In the case of panel surveys this refers to state transitions from one wave to the next. If the transition probabilities are time-homogeneous, we write $P(t) \equiv P = (p_{i,j})$, for all t. Transition probabilities from starting state at $t = 0$ to state at time t are given by $P^{(t)} = P(1)P(2)\ldots P(t)$. In the case of time-homogeneous chains we have $P^{(t)} = P^t$.

2.1 A contraction theorem in the case of a time-inhomogeneous Markov chain

Our application of Markov chain theory uses two different starting distributions. Let $\pi_F(0)$ be the starting distribution on the state space for the gross-sample of the panel at first wave. We refer to this sample as the $FULL$-sample. The second Markov chain with a different starting distribution $\pi_R(0)$ applies to those who responded at the first wave. We refer to this sample as the $RESP$-sample.

Our assumption **A** requires that the state transitions of all individuals belonging to the two samples have the same transition matrix $P^{(t)}$. This holds, as long as survey participation has no influence on subsequent state transitions. A priori it seems plausible to assume that the chances on the labour market are not affected by the participation in a survey. Also in interviewer-based surveys it is often the field work setting that influences the participation or nonparticipation. The literature on nonresponse has many indications that paradata which describe field work are powerful indicators for nonresponse (e.g., Watson and Wooden (2009)), and in many instances these paradata are uncorrelated with the state variables of interest. However, there can be situations where a change on the state space interacts with field situations. For example, a change from receiving unemployment benefits to not receiving them may be connected with regional mobility, and the follow-up of such movers may turn out to be a burden for the field-work.

Formally assumption **A** cannot be verified from the respondent data alone, so it must be considered as a Missing at Random (MAR) assumption in the sense of Rubin (1976). But, when the panel sample is recruited from a register and if there is access to key variables for both respondents and nonrespondents, then it may be possible to test assumption **A** directly for variables that are recorded in the register. This approach will be presented in the empirical part below.

Under assumption **A** the state distributions of the two Markov chains at wave t are computed in a sequential fashion from $\pi_F(t) = P'(t)\pi_F(t-1)$ and $\pi_R(t) = P'(t)\pi_R(t-1)$. When all entries of $\pi_R(t)$ are strictly positive, we have the inequalities

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

for all $i = 1, \ldots, I$.

The following contraction theorem states that, under realistic regularity conditions, the two distributions $\pi_F(t)$ and $\pi_R(t)$ converge.

Theorem 2.1. *Suppose that there is lower bound $0 < p_L \leq p_{i,j}(t)$. Then $\pi_F(t)$ and $\pi_R(t)$ converge uniformly in the sense that*

$$\lim_{t \to \infty} (M_t - m_t) = 0. \tag{2}$$

A proof of the result is given in the appendix.

Note, that the assumption of a lower bound would typically be satisfied by social indicators of the type we are interested in. The same is true for the implicit assumption that all states can be reached from all states, in one step.

We should mention that there is an alternative route to our contraction theorem which avoids the Markovian chain model for the individual behaviour. This approach was developed in mathematical demography and serves there as a realistic model to reality when population sizes are large (cf., Cohen 1979). Mathematically, our contraction theorem is actually a result that holds for any sequence of non-negative matrices, say, $A(t)$, and state vectors $X(t)$ that evolve recursively as $X(t) = A(t)X(t-1), t = 1, 2, \ldots$. This matrix model has the weak ergodic property, if, say, $X(0) > 0$ and the elements $a_{ij}(t)$ of matrices $A(t)$ are bounded from below and above, $0 < a \leq a_{ij}(t) \leq A < +\infty$. This approach involves no assumption of individual level stochasticity like the Markov chain model does, and it can provide an approximation to the process even in the presence of absorbing states, like deaths. Note, that in the case of absorbing states, the Markovian approach will not result in a contraction theorem.

Armed with this dual view of the weak egodicity result, we note that the important consequence of this theorem is that a potential nonresponse bias in the first panel wave measured as a difference of the state distributions of the $FULL$-sample and the $RESP$-sample tends to disappear in later panel waves or over time.

2.2 Regularity conditions for the attrition process

The $FULL$- and the $RESP$-samples remain unchanged in later panel waves. What changes, however, are their state distributions $\pi_F(t)$ and $\pi_R(t)$. But the $RESP$-sample is further reduced by panel attrition. The observed sample at t will be denoted by OBS_t. Its state distribution is $\pi_O(t)$.

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. In order to avoid lengthy notation on individual probabilities we display here attrition as a simple matrix multiplication which connects the distributions $\pi_O(t)$ and $\pi_O(t-1)$. Let us define a diagonal matrix of response proportions $R(t) = diag(R_1(t), \ldots, R_I(t))$, where $0 < R_i(t) \leq 1$ for $i = 1, \ldots, I$ at t. Then, under assumption **A** the state distribution of the observed sample satisfies the recursion $\pi_O(t) = n_t R(t) P'(t) \pi_O(t-1), t = 1, 2, \ldots,$ where n_t is a suitable normalisation constant. Our assumption **B** states simply that response frequencies must be equal, or $R_1(t) = \ldots = R_I(t)$ for all t. We see that under this assumption the state distribution is unaffected by the attrition, although case counts may go down. Assumption **B** may be regarded as restrictive, but in the end this is an empirical matter that may be directly decided based on observable data.

There are arguments supporting the hypothesis that survey participation in later panel waves differs substantially from the participation behaviour at the start of the panel. People who have agreed to participate in the first panel wave have shown some interest in the goals of the survey. This shows up in the overall nonresponse rates which usually dramatically decrease after wave one. For example, Junes (2012) reports attrition rates of 8 % (wave 2), 7 % (wave 3) and 5 % (wave 4) which compare to 30 % in wave 1 in the case of the Finnish subsample of EU-SILC. Also field-work related effects come into play. For example, in a face-to-face mode the interviewees expect to see the interviewer of the first round at their door. Behr et al. (2005) report that a change of the interviewer is one of the most important causes of attrition for national subsamples of the ECHP. Similar results are also reported in the analysis of Watson and Wooden (2009). The change of the interviewer is typically related to the recruitment policy of the field-institute and as such independent of the variable of interest. Standard attrition analyses of panel surveys use a lot of variables that potentially explain selective attrition (cf., Kroh (2011) for the SOEP). The attrition analysis of the Australian HILDA panel contains 77 predictors, (cf., Watson and Wooden (2009)). However, when "significant" variables are found, they are seldom stable for the explanation of panel attrition in future waves (Behr et al. (2005)). Moreover, there is a question of magnitude. Often the differences between nonresponse rates are as small as 5 percentage points while the differences due to the Markovian process can amount 50 percentage points or more. Rendtel (2015) simulated different attrition scenarios with Finnish EU-SILC data and reported that in this case differences in attrition propensity up to 10 percentage points do not affect a fade away effect substantially.

2.3 The speed of the fade away effect

The contraction theorem only assures us that two populations with different starting distributions will have similar state distributions eventually. The proof indicates that the convergence is geometric. The rate of convergence depends on the nature of bounds that hold for the elements of the transition matrices. For time-homogeneous chains the geometric convergence also holds, but the situation is much simpler in other ways, as a limiting distribution exists.

Consider an irreducible time-homogeneous chain with $I \times I$ transition matrix P. The largest eigenvalue P is 1, it is simple, and the corresponding eigenvector π^* can be chosen strictly positive, $\pi^* = P'\pi^*$. These results can be proven directly (e.g., Cinlar 1975) or they follow, e.g., from the so-called Perron-Frobenius theorems (e.g., Gantmacher 1959). This key result can be complemented by the following

Theorem 2.2. *Suppose P has the second eigenvalue λ_2, then*

$$|p_{ij}^{(t)} - \pi_j^*| = O(|\lambda_2|^t) \text{ for all } i, j \in S. \tag{3}$$

For a proof, see Seneta (1980, Theorem 4.2).

As is clear from the proof in the Appendix, if P is strictly positive, the speed of convergence is directly related to the minimum entry of P (see also Behrends (2000, p.83 ff)). Thus, processes with low transition probabilities tend to need long time-intervals to reach the steady state.

In the application we will meet a situation where the distribution of the gross sample $\pi_F = (\pi_{F,1}, \ldots, \pi_{F,I})'$ and the net sample of the first wave $\pi_R = (\pi_{R,1}, \ldots, \pi_{R,I})'$ may be far away from the steady state distribution. Yet the differences $D_j(t) = \pi_{F,j}(t) - \pi_{R,j}(t)$ between the two distributions converge to 0 in a geometric fashion.

3 Data Base and Emprical 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.

References

Atkinson, A. and Marlier, E. (eds.) (2010): Income and Living Conditions in Europe. Eurostat doi:10.2785/53320

Behr, A., Bellgardt, E. and Rendtel, U. (2005): Extent and determinants of panel attrition in the European Community Household Panel. European Sociological Review, 21, 489-512.

Behrends, E. (2000): Introduction to Markov Chains. Vieweg Verlag, Heidelberg.

Cinlar, E. (1975): Introduction to stochastic processes. Prentice Hall, Englewood Cliffs.

Cohen, J.E. (1979): Ergodic theorems in demography. Bulletin of the American Mathematical Society, 1, 275-295.

Gantmacher, F.R. (1959): The theory of matrices, Vols. I-II. Chelsea Publishing, New York.

Junes, T. (2012): Initial wave nonresponse and panel attrition in the Finish subsample of EU-SILC, Master Thesis at the Department of Social Statistics, University of Helsinki, Helsinki.

Kroh, M. (2011): Documentation of sample sizes and panel attrition in the German Socio Economic Panel (SOEP) (1984 until 2011), Data Documentation, 59, DIW Berlin

Rendtel, U. (2013): The fade away effect of initial nonresponse in panel surveys: Empirical results for EU-SILC. Methodologies and Working papers (ISSN 1977-0375) Edition 2013, Eurostat

Rendtel, U. (2015): Is there a fade away effect of initial nonresponse bias in EU-SILC? Discussion paper Economics FB Wirtschaftswissenschaft FUB 2015/25. Online available at http://edocs.fu-berlin.de/docs/receive/FUDOCS_document_000000022956

Rubin, D. (1976): Inference and missing data, Biometrica, 63, 581-592.

Seneta, E. (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

Watson, N. and Wooden, M. (2009): Identifying factors affecting longitudinal survey nonresponse. In: P. Lynn (ed.) Methodology of Longitudinal Surveys, 157-181, Wiley.