Survey Research Methods (2021)
Vol. 15, No. 2, pp. 135-146
doi:10.18148/srm/2021.v15i2.7640
European Survey Research Association

Dependent Interviewing: A Remedy or a Curse for Measurement Error in Surveys?

Paulina Pankowska
Department of Sociology, and
Department of Communication Science
Vrije Universiteit Amsterdam
The Netherlands

Bart Bakker
CBS Academy
Statistics Netherlands
The Netherlands

Daniel Oberski
Department of Methodology and Statistics
Utrecht University
The Netherlands

Dimitris Pavlopoulos
Department of Sociology
Vrije Universiteit Amsterdam
The Netherlands

Longitudinal surveys often rely on dependent interviewing (DI) to lower the levels of random measurement error in survey data and reduce the incidence of spurious change. DI refers to a data collection technique that incorporates information from prior interview rounds into subsequent waves. While this method is considered an effective remedy for random measurement error, it can also introduce more systematic errors, in particular when respondents are first reminded of their previously provided answer and then asked about their current status. The aim of this paper is to assess the impact of DI on measurement error in employment mobility. We take advantage of a unique experimental situation that was created by the roll-out of dependent interviewing in the Dutch Labour Force Survey (LFS). We apply hidden Markov modelling (HMM) to linked LFS and Employment Register (ER) data that cover a period before and after dependent interviewing was abolished, which in turn enables the modelling of systematic errors in the LFS data. Our results indicate that DI lowered the probability of obtaining random measurement error but had no significant effect on the systematic component of the error. The lack of a significant effect, particularly in the case of autocorrelated errors, might be driven by the fact that the probability of repeating the same error was extremely high at baseline (i.e. when using standard, independent interviewing); therefore the use of DI could not increase this probability any further.

Keywords: Dependent interviewing (DI); Measurement error; Hidden Markov models (HMM); Panel survey

1 Introduction

Measurement error in survey data is a well-known and well-documented phenomenon. A large volume of literature confirms that, if left unaccounted for, such error often biases estimates and can lead to inaccurate inferences and predictions (Alwin, 2007; Pankowska, Bakker, Oberski, & Pavlopoulos, 2018; Saris & Gallhofer, 2014; West & Blom, 2017). The magnitude of this problem is particularly high when using longitudinal survey data to estimate change or stability over time, as such second-order statistics have been shown to be severely affected by measurement error (Bound, Brown, & Mathiowetz, 2001; Fuller, 2009; Hagenaars, 1990, 1994; Van de Pol & De Leeuw, 1986). More specifically, when the measurement error is random, observed over time changes are often inflated as they not only reflect true changes but also include changes in the error (Jäckle & Eckman, 2019). For this reason, survey methodologists have applied various tools to minimize the occurrence of measurement errors by improving data collection processes in longitudinal surveys (Groves et al., 2011).

One tool in particular that has been widely implemented in various large-scale longitudinal surveys worldwide, such as the British Household Panel Survey (BHPS), the Dutch Labour Force Survey (LFS), and the US Current Population Survey (CPS) (Jäckle, Laurie, & Uhrig, 2007) is *dependent interviewing* (DI). DI a method that uses information from responses provided in previous interview rounds to modify

Contact information: Paulina Pankowska, Vrije Universiteit Amsterdam, Department of Communication Science, De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands (E-Mail: p.k.p.pankowska@vu.nl)

the phrasing and routing of questions in subsequent survey waves, as well as to facilitate within-interview edit checks (Jäckle, 2009; Jäckle et al., 2007; Mathiowetz & McGonagle, 2000). When using proactive dependent interviewing (PDI), the wording of the question is tailored based on the previously provided response(s) (Jäckle, 2009).[1] In this design, interviewees can be asked the question in three distinct manners: in "remind, continue" respondents are reminded of their previous answer and then asked the standard independent question; in "remind, still" they are asked whether the situation described still holds; in "remind, confirm" interviewees are asked to confirm whether their previous response is correct (Hoogendoorn et al., 2004; Jäckle, 2008, 2009; Jäckle & Eckman, 2019; Jäckle et al., 2007; Jäckle & Lynn, 2007; Lugtig & Lensvelt-Mulders, 2014; Mathiowetz & McGonagle, 2000).

PDI is used in longitudinal surveys for two main reasons: (i) it has the potential to improve data quality by achieving higher longitudinal consistency and lower levels of random error (Jäckle, 2009; Mathiowetz & McGonagle, 2000) and (ii) it can increase survey efficiency and reduce respondent burden (Eggs & Jäckle, 2015; Jäckle, 2008). The importance of improving data quality is related to the fact that, as mentioned previously, longitudinal surveys in most cases suffer from random measurement error, which has the potential to severely inflate change estimates (Jäckle, 2009; Jäckle & Lynn, 2007; Lugtig & Lensvelt-Mulders, 2014; Lynn et al., 2006; Van de Pol & De Leeuw, 1986). Previous studies show that PDI has been effective in reducing spurious change and the seam effect[2] in numerous different panel surveys (Jäckle & Eckman, 2019). The need to increase the efficiency of the interview process and to reduce respondent burden is tied to the common complaints made by interviewees about having to answer the same question recurrently even when their circumstances have not changed. PDI reduces the need to repeatedly answer the same question and thus is thought to reduce respondent burden. Furthermore, tailoring the question to the respondents' specific situation and reminding them of their previously provided answers was shown to improve the flow of the interview and simplify the response task (e.g. Sala, Uhrig, & Lynn, 2011). These efficiency gains have also been linked to lower rates of (random) measurement error (Hoogendoorn et al., 2004; Jäckle, 2009; Lynn et al., 2006).

Overall, PDI is potentially an effective technique that allows addressing several challenges faced by survey methodologists when dealing with repeated longitudinal surveys. However, it is not free of shortcomings, as there is some concern that PDI might lead to more systematic measurement error, and particularly autocorrelated error, through two main mechanisms/cognitive processes. First, PDI might increase the incidence of error due to the phenomenon of *(cognitive) satisficing*, wherein respondents, rather than providing a well-thought-out, appropriate answer, tend to opt for the easy, credible response. In the context of PDI, this would imply falsely confirming that the previous answer still holds (Eggs & Jäckle, 2015; Hoogendoorn et al., 2004; Jäckle & Eckman, 2019; Lugtig & Lensvelt-Mulders, 2014). Second, PDI might also have an adverse effect on the error due to the presence of *motivated misreporting*, a phenomenon whereby individuals, to shorten the duration of the interview, provide inaccurate answers that allow them to omit follow-up questions. This implies that when using PDI respondents will be inclined to report that the previous information still holds, as this will likely allow them to skip questions about their current state (Eggs & Jäckle, 2015). Such false reports of 'no change' might lead to spurious stability if a true change did occur. It can also lead to the copying over of an error across waves if no change occurred and the previously provided answer was wrong (Eggs & Jäckle, 2015; Hoogendoorn et al., 2004; Jäckle & Eckman, 2019).

Therefore, the overall effect of PDI on data quality appears uncertain and remains an issue for empirical investigation: on the one hand, this interviewing technique could reduce random error, but, on the other hand, it can increase the incidence of systematic error (as shown e.g. by Lugtig & Lensvelt-Mulders, 2014). From the perspective of substantive researchers, it appears that decreasing spurious change through the use of PDI might come at the expense of increasing spurious stability.

Given the two contradictory effects, and the lack of consensus in the literature regarding the overall utility of DI, this paper aims disentangle the effect of PDI on random and systematic errors, and in this way to assess the overall effect of PDI on measurement error. Specifically, this paper aims to answer the following research questions and test the following hypotheses:

1. What is the effect of PDI on random measurement error?

   H1 The use of PDI, compared to independent interviewing, results in lower levels of random error.

2. What is the effect of PDI on systematic measurement error?

[1] DI can also be used reactively (RDI), whereby respondents are first asked the question independently and then, if an inconsistency is detected between the current and previous answer, a follow-up question is raised to verify whether a change occurred (Jäckle & Eckman, 2019; Uhrig & Sala, 2011). As RDI is primarily applied to numeric responses (Jäckle & Eckman, 2019) and is not expected to have strong implications for systematic error (Lynn, Jäckle, Jenkins, & Sala, 2006), our paper focuses on the effect of PDI on measurement error.

[2] a phenomenon whereby between-wave change is overestimated while within-wave change is underestimated (Jäckle & Eckman, 2019).

H2 The use of PDI, compared to independent interviewing, leads to more systematic error.

For this purpose, rather than conducting our own experiment, we leverage the replacement of PDI with independent interviewing (INDI), which took place at the beginning of 2010 in the Dutch LFS. The questionnaire was changed as the routing in the former version was too complex, leading to mistakes in the interview. As no other major changes in the survey data collection process occurred in the time period under study, this change provides a natural experiment, which allows for the study of the impact of PDI on measurement error while treating independent interviewing (INDI) as the counterfactual.

To assess the magnitude of measurement error in the corresponding survey question, we use hidden Markov models (HMMs), a group of latent class models that allow estimating and correcting for measurement error in categorical, longitudinal data, provided that the model is specified correctly (Biemer, 2004; Pankowska et al., 2018; Pavlopoulos & Vermunt, 2015). The main advantage of these models is that they do not require a "gold-standard", error-free data source, which would serve as a benchmark for the survey data (Biemer, 2011; Vermunt & Magidson, 2002). To model systematic measurement error in the survey data without having to impose unwanted restrictions and risk poor identifiability, we use an extended, two-indicator version of HMMs (Bassi, Hagenaars, Croon, & Vermunt, 2000). These two indicators are obtained by linking data from the Dutch LFS and the Dutch Employment Register (ER).

The remainder of the paper is structured as follows: section 2 provides some context to our study by describing the roll-out of PDI in the Dutch LFS. Section 3 discusses the use of HMMs to assess and correct for measurement error as well as the model and data used in the analysis. Section 4 summarizes the results obtained and, finally, section 5 offers concluding remarks.

2 Dependent interviewing (DI) in the Dutch Labour Force Survey (LFS)

The Dutch Labour Force Survey (LFS) is an address-based sample survey that provides information on labour market characteristics of individuals residing in the Netherlands. It is carried out by Statistics Netherlands and, as of the end of 1999, it is a quarterly rotating panel survey consisting of five waves. Dependent interviewing (DI), and more specifically the "remind, still" style of proactive DI (PDI), was in use in the LFS from the beginning until the end of 2009; at the beginning of 2010 it was replaced by independent interviewing (INDI). Survey respondents were asked about their employment contract using PDI if they met two conditions: (i) they reported in the previous wave that they had a temporary contract and (ii) they indicated that they did not change their job since the previous wave. Respondents who fulfilled both criteria were asked the following question regarding their contract type: *Last time you had a temporary contract. Is this still the case?*. Individuals who changed jobs or those who did not experience a job change but had indicated previously that they had "other" type of contract (i.e. were not in paid employment) were asked the question in an independent fashion as follows: *Do you currently have a permanent contract?*. The contract question was skipped for respondents who in the previous wave reported having a permanent contract and who did not experience a job change; instead, these individuals' responses from the previous wave were copied forward.[3]

This setup, which is summarized in the flowchart of Figure 1, results in three possible scenarios: (i) an individual is subject to INDI if either (a) they indicated that a job change occurred, (b) they reported having "other" type of employment in the previous survey round *or*, (c) they first participated in the LFS after the end of 2009; (ii) an individual is asked the contract question using PDI if they (a) did not change their job since the previous survey wave, (b) they reported being employed on a temporary basis in the previous round *and*, (c) they first took part in the survey before the end of 2009; (iii) the contract question is not asked altogether if (a) no job change occurred *and* (b) the individual previously reported being employed permanently. While all three scenarios occur in our dataset, our analysis focuses on comparing the levels of random and systematic errors when PDI was used with those when INDI was used, but PDI would had been applicable if it was not abolished at the end of 2009. This allows us to take advantage of the natural experiment setup caused by the replacement of PDI with INDI at the beginning of 2010. More specifically, we compare scenario ii (3.31% of cases in our sample), which we refer to as the *treatment group*, to a subset of scenario i (2.28% of cases in our sample)—wherein no job change occurred, a temporary contract was reported in the previous wave, *and* the first round of the LFS was conducted after the end of 2009—which we refer to as the counterfactual or the *control group*.

[3]For more information about the LFS see https://www.cbs.nl/en-gb/our-services/methods/surveys/korte-onderzoeksbeschrijvingen/dutch-labour-force-survey-lfs; the metadata of the LFS (in Dutch) can be found at https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/ebb-enquete-beroepsbevolking-1996-2009 and https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/ebb-enquete-beroepsbevolking-2010-2011.

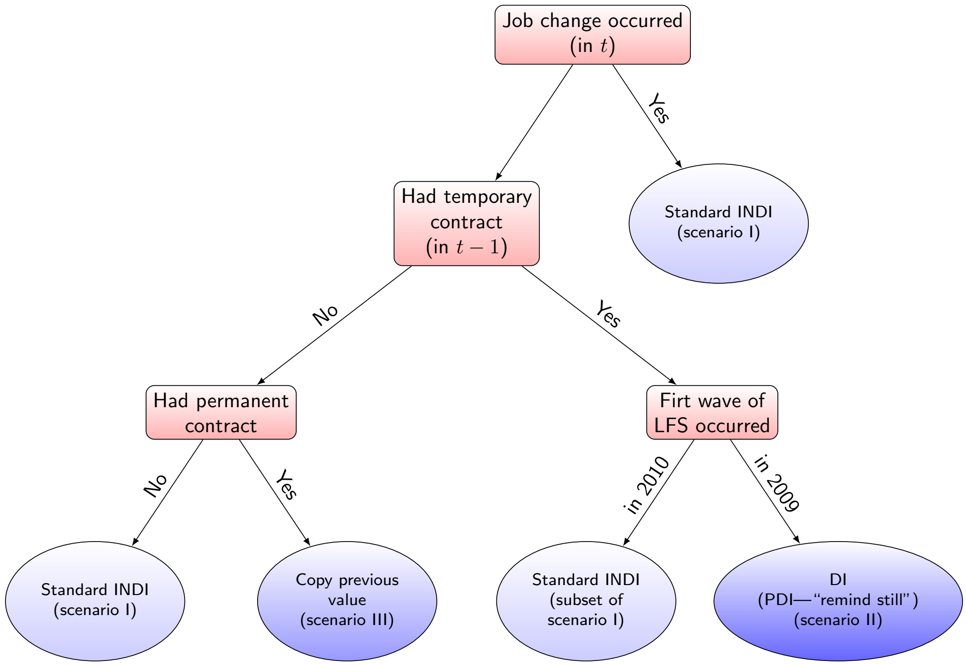

Figure 1. Summary of the interviewing setup in the LFS contract question

3 Methodology

3.1 Assessing and correcting for measurement error using hidden Markov models (HMMs)

Hidden Markov models (HMMs) are a latent variable modelling technique that can be applied to evaluate measurement error in categorical longitudinal survey data (Biemer, 2011; Pankowska et al., 2018; Pavlopoulos & Vermunt, 2015). Their rise in popularity can be attributed to the fact that, unlike other commonly used error assessment methods, they do not require the availability of error-free, 'gold-standard' validation data that are most often unattainable in practice (Biemer & Wiesen, 2002; Pankowska, Bakker, Oberski, & Pavlopoulos, 2019). In this context, HMMs are used when the (dynamic) quantity of interest, e.g. overtime employment transitions, is measured in the panel survey with some degree of error. The models allow separating true change from measurement error which, in turn, allows producing error-corrected estimates of the quantity of interest as well as assessing the level of measurement error in the corresponding survey question (Biemer, 2011; Pankowska et al., 2018).

The standard HMM, which can be fit to surveys with at least three panel waves, consists of two components: (i) the structural component that models the true (latent) initial state probabilities X_0 and the true (latent) transition probabilities between X_{t-1} and X_t, where $t = 1, \ldots, T$; and (ii) the measurement component that models the interactions of the survey observations (which contain error) A_t with the true values X_t at each wave $t = 1, \ldots, T$. The two components are estimated simultaneously. The model relies on two basic assumptions: first, the probability of a specific value of X occurring at time t only depends on its value in the previous time point, X_{t-1}—the so-called *Markov assumption*. This assumption can be stated formally as follows:

$$\Pr(X_t = x_t | X_1 = x_1, \ldots, X_{t-1} = x_{t-1}) = \Pr(X_t = x_t | X_{t-1} = x_{t-1}) \quad (1)$$

where $\Pr(X_t = x_t)$ denotes the probability of the latent state X_t taking on a specific value x_t out of k possible categories. Second, the probability of observing a specific value of A at time t only depends on the true value at the same time point—X_t—the so-called *local independence assumption* or—using a term that is more appropriate for longitudinal data—*independent classification error (ICE) assumption*. This assumption can be stated formally, as follows:

$$\Pr(A_1 = a_1, \ldots, A_T = a_T | X_1 = x_1, \ldots, X_T = x_T) = \prod_{t=1}^{T} \Pr(A_t = a_t | X_t = x_t) \quad (2)$$

where $\Pr(A_1 = a_1, \ldots, A_T = a_T)$ denotes the probability of observing a specific path or sequence of survey states, where each state – $A_1, \ldots, A_T$ – takes on a specific value – $a_1, \ldots, a_T$ – out of k possible categories. Combining the Markov and local independence assumptions leads to the following probability of observing a certain path $A = (A_1, ..., A_T)$ in the survey data:

$$\Pr(A = a) = \sum_{x_0=1}^{k} \cdots \sum_{x_T=1}^{k} \Pr(X_0 = x_0) \prod_{t=1}^{T} \Pr(X_t = x_t | X_{t-1} = x_{t-1}) \prod_{t=1}^{T} \Pr(A_t = a_t | X_t = x_t) \quad (3)$$

where $\Pr(X_0 = x_0)$ represents the initial state latent probabilities and $\Pr(X_t = x_t | X_{t-1} = x_{t-1})$ represents the latent transition probabilities, which follow a first-order Markov process. $\Pr(A_t = a_t | X_t = x_t)$ denotes the classification error (also referred to as emission) probabilities, which satisfy the local independence assumption and are used to estimate question reliability in surveys (Bassi et al., 2000; Biemer, 2011; Pankowska et al., 2018; Pankowska et al., 2019; Pavlopoulos & Vermunt, 2015).

If only three time-points are available, in addition to the two assumptions specified above, further restrictions in the form of time-invariant/constant misclassification (measurement error) rates and latent transitions rates are required to obtain model identification (Biemer, 2011; Pankowska et al., 2018; Van de Pol & De Leeuw, 1986). Given these assumptions and restrictions, which are required to obtain identifiability, the standard, one-indicator HMM can be seen as rather limited in its capacity to model realistic error scenarios. While it is possible to relax some of the assumptions when using richer survey data (i.e. with more than three data points), the practical applicability of the model remains rather limited. To illustrate, even with multiple ($t > 3$) survey waves, one cannot simultaneously model both local dependence, which allows for the occurrence of systematic error, as well as time-varying measurement and/or structural parameters. It is worthwhile noting that, even models that only account for the occurrence of systematic error often suffer from identifiability issues (i.e. are "poorly identifiable"). As a result of these limitations, survey researchers have increasingly started using extended, multiple-indicators versions of the standard HMM, which are more flexible and allow for model specifications that are more reflective of reality (Pankowska et al., 2018; Pankowska et al., 2019; Pavlopoulos & Vermunt, 2015).

A basic two-indicator HMM, which can be obtained, for instance, by linking survey data to register/administrative records, has the following probability of observing certain paths $A = (A_1, \ldots, A_T)$ and $B = (B_1, \ldots, B_T)$:

$$\Pr(A = a, B = b) = \sum_{x_0=1}^{k} \cdots \sum_{x_T=1}^{k} \Pr(X_0 = x_0) \prod_{t=1}^{T} \Pr(X_t = x_t | X_{t-1} = x_{t-1}) \prod_{t=1}^{T} \Pr(A_t = a_t | X_t = x_t) \prod_{t=1}^{T} \Pr(B_t = b_t | X_t = x_t) \quad (4)$$

Where the latent initial state probabilities—$\Pr(X_0 = x_0)$, the latent transition probabilities—$\Pr(X_t = x_t | X_{t-1} = x_{t-1})$, and the survey emission probabilities—$\Pr(A_t = a_t | X_t = x_t)$—are specified in the same way as in the univariate/one-indicator HMM described above. This extended specification also includes the register emission probabilities—$\Pr(B_t = b_t | X_t = x_t)$—that, in a similar way to the survey emission probabilities, also satisfy the local independence assumption. While this is the most basic two-indicator HMM specification, the model can be easily extended further by e.g. (i) accounting for (un)observed heterogeneity and time dependency in the transition and/or emission probabilities and (ii) relaxing the local independence assumption for the survey and/or register data.

3.2 The empirical model

In our analysis, we make use of an extended HMM specification with two indicators that come from two independent data sources (i.e. the Dutch LFS and Employment Register[4]). Such a specification allows us to model the possibility that PDI leads to more systematic error in the survey data and, at the same time, allows the latent transition probabilities to depend on time and personal characteristics (following Pankowska et al., 2018 and Pavlopoulos and Vermunt, 2015).

To obtain a second employment contract type indicator, we linked the LFS data to records from the Dutch Employment Register (ER). The ER is an administrative dataset that combines information from various sources but predominantly consists of tax-related data provided to the Dutch Tax Authorities by employers. It is managed by the Dutch Employee Insurance Agency (UWV) and contains monthly information for all insured employees in the Netherlands on such individual-level characteristics as wages, benefits, and labour relations.[5] The record linkage is performed at the

[4]Both data sources are not freely accessible; access can only be granted by CBS following special arrangements. For more information see https://www.cbs.nl/en-gb/onze-diensten/customised-services-microdata/microdata-conducting-your-own-research.

[5]For more information about the ER see https://www.cbs.nl/nl-nl/achtergrond/2010/35/polisadministratie; the metadata of the ER (in Dutch) can be found at https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/spolisbus-banen-en-lonen-volgens-polisadministratie.

individual level using a unique citizen identification number (in Dutch: *Burgerservicenummer* or, in short, BSN). If the BSN is missing or incomplete, a combination of birth date, sex, postal code, and house number is used instead as a linkage key. The linkage effectiveness of this procedure, i.e. the percentage of linked records in the LFS, is estimated by Statistics Netherlands to be around 98%. In the following we will assume perfect record linkage. Previous research has shown that even if there is linkage error, its effects on the estimates of HMMs is negligible unless this error is large and strongly correlated with the process of interest (Pankowska et al., 2019). This is definitely not the case in our data.

Our linked dataset consists of 86,075 LFS respondents of prime working age (i.e. 25 to 55 years old) who first participated in the survey either in 2009 (PDI in place) or 2010 (PDI abolished). It contains quarterly information on each individual for 5 time points, leading to a total sample size of 430,375 observations. The survey data are subject to unit and item non-response[6], and have relatively high attrition rates. With regards to the main variable of interest—the employment contract type—in the first wave of the survey all 86,075 respondents provided an answer to this question, in the second wave 63,507 provided an answer, in the third one 50,557, in the fourth one 47,088 and in the fifth wave 45,318. The employment contract type variable obtained from the administrative data has about 27% of missing values (a total of 114,500). The submission of reports is mandatory for all employers, therefore officially this dataset cannot suffer from non-response or attrition and the missing values are primarily of individuals who are included in the LFS population but are not a part of the employment register (e.g. self-employed or unemployed individuals). We assume all missing values to be missing at random (MAR) (Little & Rubin, 2019, pp. 118–119). While the MAR assumption might be violated as non-response in the LFS might lead to selection bias, we correct for it to the extent possible by including individual-level covariates in the model. All observations with missing values on the contract type variables are included in the analysis and the model is estimated using full information maximum likelihood; missing values on the covariates are imputed (Vermunt & Magidson, 2013).

Table 1 provides the distribution of observations by the conditions determining PDI eligibility. As can be seen from the table, overall PDI was used in a rather small fraction of the sample. That is, in approx. 3.3% of the cases individuals were asked the question in a PDI fashion (i.e. 3.3% of the observations belong to the treatment group); in around 2.3% of the cases PDI would have been used if it were not abolished (i.e. 2.3% belong to the control group/counterfactual).

In this linked survey and register dataset, the probability of observing particular employment contract paths—A and B—which depend on observed individual-level heterogeneity (Z) and the interviewing regime used (W), according to our two-indicator HMM, can be formalized as follows:

$$\Pr(A = a, B = b|Z, W) = \sum_{x_0=1}^{k} \cdots \sum_{x_T=1}^{k} \Pr(X_0 = x_0|Z) \prod_{t=1}^{T} \Pr(X_t = x_t|X_{t-1} = x_{t-1}, Z)$$

$$\prod_{t=1}^{T} \Pr(A_t = a_t|X_t = x_t, X_{t-1} = x_{t-1}, A_{t-1} = a_{t-1}, W)$$

$$\prod_{t=1}^{T} \Pr(B_t = b_t|X_t = x_t, X_{t-1} = x_{t-1}, B_{t-1} = b_{t-1}) \quad (5)$$

where the (latent) initial state probabilities and transition rates—$\Pr(X_0 = x_0|Z)$ and $\Pr(X_t = x_t|X_{t-1} = x_{t-1}, Z)$—depend on observed individual-level heterogeneity (i.e. the covariates education, gender and ethnicity) and the latent transitions also depend on time (i.e. are time-heterogeneous and depend on t and t^2). The inclusion of covariates in the structural part of the model implies that the Markov assumption holds conditional on these covariates. The emission probabilities for both the survey and register data—$\Pr(A_t = a_t|X_t = x_t, X_{t-1} = x_{t-1}, A_{t-1} = a_{t-1}, W)$ and $\Pr(B_t = b_t|X_t = x_t, X_{t-1} = x_{t-1}, B_{t-1} = b_{t-1})$—relax the local independence assumption allowing for systematic error in both data sources. In more detail, for both the LFS and the ER, we allow the error probabilities to also depend on the lagged true contract—X_{t-1}—and the lagged observed contract—A_{t-1} or B_{t-1}. Additionally, to compare the error levels under PDI and INDI, the LFS emission probabilities also depend on the covariate W, which determines the interviewing regime used and can take 3 values:

- 0 (ref. category) INDI was used but PDI would have been used if it was not abolished;
- 1 INDI was used and would have been used regardless of whether DI had been abolished;
- 2 PDI was used.

In our analysis, we focus on comparing the error levels under PDI to those where PDI would have been used (i.e. category 2 vs. 0).

Following the approach of Manzoni, Vermunt, Luijkx, and Muffels (2010) Manzoni et al. (2010), we use a restricted model for the survey data that only allows for systematic error in situations where the errors are a consequence of the phenomena of *satisficing* and/or *motivated misreporting*. Specifically, we define a logit model for the prob-

[6]In 2009 the response rate of the LFS was 61% and in 2010 it was 53%. The exact response rate of the sample used for the analysis, which was restricted to individuals aged 25 to 55 and excluded those with a temporary contract with intent to be hired permanently, is not available to us.

Table 1
Distribution of observations by DI eligibility (LFS year, job change in t and contract type in $t-1$) ($N = 430,375$)

	LFS contract at $t-1$					
	2009			2010		
Job Change	Permanent %	Temporary %	Other %	Permanent %	Temporary %	Other %
Yes	0.46	0.23	0.62	0.23	0.13	0.44
No	55.15	**3.31**	0.33	36.36	**2.28**	0.46

The percentages correspond to the shares of specific groups in the overall sample and are calculated by dividing the number of individuals who fulfill the respective criteria by the overall sample size; the percentages of treatment and control groups are provided in bold.

ability of having measurement error in the survey data—$\Pr(A_t = a_t | X_t = x_t, X_{t-1} = x_{t-1}, A_{t-1} = a_{t-1}, W)$—in the following form: $\alpha_{a_t,x_t} + \beta_{a_t,a_{t-1},x_t,x_{t-1}} + \alpha_{a_t,x_t,w} + \beta_{a_t,a_{t-1},x_t,x_{t-1},w}$. In this specification, the term $\alpha_{a_t,x_t} + \alpha_{a_t,x_t,w}$ represents the random component of the error while the term $\beta_{a_t,a_{t-1},x_t,x_{t-1}} + \beta_{a_t,a_{t-1},x_t,x_{t-1},w}$ represents the systematic component of the error. In both cases, the first parts of the expression (i.e. α_{a_t,x_t} and $\beta_{a_t,a_{t-1},x_t,x_{t-1}}$) represent the 'baseline' random/systematic error probability while the second parts (i.e. $\alpha_{a_t,x_t,w}$ and $\beta_{a_t,a_{t-1},x_t,x_{t-1},w}$) indicate how the use of different interviewing regimes affects the probabilities of obtaining random/systematic error. Put simply, to compare the effect of using PDI as opposed to standard INDI, we estimate additional random and systematic error parameters for when the contract question was asked in a PDI fashion. The parameters of the systematic error components are freed when the same error can be repeated due to the "remind, still" PDI—i.e. when $A_t = A_{t-1} = \text{temporary} \neq X_t = X_{t-1} = \{\text{permanent}, \text{other}\}$ or when it might cause spurious stability; that is, in a situation where an individual correctly reports having a temporary contract in *t-1*, then experiences a true transition between *t-1* and *t* but erroneously confirms in *t* that she/he is still employed on a temporary basis—i.e. when $A_{t-1} = X_{t-1} = \text{temporary}$ and $A_t = \text{temporary} \neq X_t = \{\text{permanent}, \text{other}\}$. In all other instances the systematic error parameters are set to 0.

For the register data, we only allow for the repetition of the same error over time, as previous research has shown the ER to suffer predominantly from this type of error (Pankowska et al., 2018; Pavlopoulos & Vermunt, 2015). That is, for the error parameters—$\alpha_{b_t,x_t} + \beta_{b_t,a_{t-1},x_t,x_{t-1}}$—we estimate the systematic component—$\beta_{b_t,a_{t-1},x_t,x_{t-1}}$—in situations where $B_t = B_{t-1} \neq X_t = X_{t-1}$; in all other cases we set this component to 0. Table A1 in the Appendix lists all possible systematic error parameters and specifies which ones were freed and which ones were set to 0 in both the LFS and the ER data.

In our model, k runs from 1 to 3 and represents the number of contract type categories {permanent, temporary, other}; T runs from 1 to 5 and corresponds to the months in which the (quarterly) survey took place. The model is estimated in the Latent GOLD software (version 4.5), using the Baum-Welch algorithm, which is an adapted expectation-maximization (EM) procedure (for further details about this process see McLachlan & Krishnan, 2007; Pankowska et al., 2019, pp. 291–292). A path diagram of the model is provided in Figure 2.

4 Results

In this section, we first investigate whether the use of PDI, as shown by previous studies, indeed lowers the occurrence of random measurement error. We then look at whether, as hypothesized, PDI also leads to higher incidence of systematic error. In doing so, we compare the corresponding measurement error parameter estimates obtained when (i) PDI was used in 2009 to (ii) those obtained when INDI was applied in 2010 to cases that would have been eligible for PDI had it not been abolished. To reiterate, both scenarios include observations in which LFS respondents in *t-1* reported having a temporary contract and in *t* stated that they did not change their job. Therefore, all of these individuals fulfilled the criteria for PDI. However, only those who first participated in the survey in 2009 were actually asked the question in a PDI fashion; individuals who started the LFS in 2010 were subject to INDI. Table 2 presents the estimates of the random error parameters under PDI, where the reference category is *INDI would have been PDI*. When investigating the effect of PDI on random error, we estimated four additional error parameter when PDI is used (compared to INDI): reporting temporary in the LFS given that the true contract is permanent or "other", and reporting permanent or "other" given it is temporary. The remaining two parameters (permanent | "other" and "other" | permanent) were restricted to 0 as PDI was specifically applied when a temporary contract was reported and, therefore, should not have any effect in these two error scenarios.

As can be seen from table 2, the use of PDI in the LFS

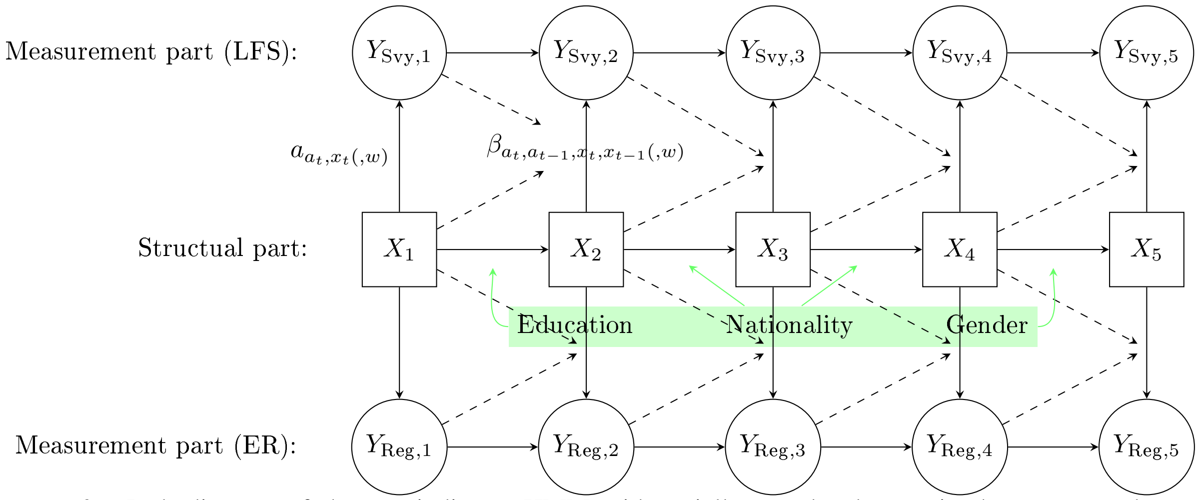

Figure 2. Path diagram of the two-indicator HMM with serially correlated error in the survey and register data and covariate dependent latent initial state and transition probabilities. *Notes:* The structual part refers to $\sum_{x_0=1}^{k} \dots \sum_{x_T=1} \Pr(X_0 = x_0|Z) \prod_{t=1}^{T} \Pr(X_t = x_t|X_{t-1} = x_{t-1}, Z)$; the measurement part of the LFS refers to $\prod_{t=1}^{T} \Pr(A_t = a_t|X_t = x_t, X_{t-1} = x_{t-1}, A_{t-1} = a_{t-1}, W)$; the measurement part of the ER refers to $\prod_{t=1}^{T} \Pr(B_t = b_t|X_t = x_t, X_{t-1} = x_{t-1}, B_{t-1} = b_{t-1})$.

Table 2
Random measurement error parameter estimates.

LFS contract	True contract	Log-linear parameter	S.E.	p
Temporary	Permanent	10.25	12.52	0.41
Permanent	Temporary	−0.64	0.10	0.00
Other	Temporary	−0.47	0.20	0.02
Temporary	Other	18.44	17.90	0.30

reduced the occurrence of random measurement error in instances where respondents erroneously reported to hold a permanent or "other" type of contract while in reality they were employed on a temporary basis ($\beta = -0.64$, $p < 0.05$ and $\beta = -0.47$, $p < 0.05$, respectively). More specifically, when asked the question in a PDI fashion compared to INDI, an LFS respondent, whose true contract at time t is temporary, is almost twice less likely to falsely report having a permanent contract (OR = 1.90) and slightly over 1.5 times less likely to report having "other" type of contract (OR = 1.60).

The probabilities of misreporting a contract type as temporary while in reality it is either permanent or "other" seem unaffected by PDI ($\beta = 10.25$, $p = 0.41$ and $\beta = 18.44$, $p = 0.30$, respectively).[7] The lack of significant effects when the true contract type is either permanent or "other" is to be expected given how this interviewing technique was set up in the LFS and given the eligibility criteria for PDI. More specifically, as individuals are only subject to PDI if they reported having a temporary contract in the previous wave, PDI will only decrease the probability of misreporting a true temporary contract as permanent or "other".

To assess whether DI leads to higher rates of systematic error, we examine the parameter estimates that correspond to situations (i) where the erroneous reporting of a temporary contract can be repeated, and (ii) where the reporting of temporary contract is correct in $t-1$ but then becomes incorrect in t due to a true transition that was not reported. As can be inferred from table 3, which provides the corresponding parameter estimates, PDI does not seem to increase the probability of obtaining systematic error. It appears that PDI leads to neither error autocorrelation nor to spurious stability (i.e. falsely confirming the previously reported answer still holds while a true change occurred).

In more detail, the error parameter estimates corresponding to a situation whereby an individual falsely reports having a temporary contract in $t-1$ and t while in both time points the true contract type is either permanent or "other" are insignificant ($\beta = -9.45$, $p = 0.45$ and $\beta = -19.43$, $p = 0.28$, respectively). Similarly, the probabilities of correctly reporting a temporary contract in $t-1$, but failing to report a true transition to either permanent or temporary employment in t (and confirming to still hold a temporary contract instead) also seem unaffected by PDI ($\beta = 2.23$, $p = 0.79$ and $\beta = 23.03$, $p = 0.69$, respectively).[8] The

[7] It is worthwhile mentioning that the very high coefficient estimates in this case are caused by the fact that the baseline probabilities (i.e. under INDI) of observing temporary given that the true contract type is either permanent or "other" are extremely low. Therefore, even a small increase in these probabilities in absolute terms can have a substantial relative effect.

[8] Again, the large coefficient estimates are caused by the fact that at the baseline (i.e. for INDI) these probabilities are either ex-

Table 3
Systematic measurement error parameter estimates.

LFS contract (in t)	LFS contract (in $t-1$)	True contract (in t)	True contract (in $t-1$)	Log-linear parameter	S.E.	p
Temporary	Temporary	Permanent	Permanent	−9.45	12.53	0.45
Temporary	Temporary	Other	Other	19.43	17.98	0.28
Temporary	Temporary	Permanent	Temporary	2.23	8.37	0.79
Temporary	Temporary	Other	Temporary	23.03	17.90	0.69

lack of an effect on the systematic component of the error, in particular for the scenarios whereby the same error can be repeated, might be due to the fact that even at baseline (i.e. when using standard INDI), there is an extremely high probability of an LFS respondent repeating the same error if no true change occurred (i.e. $\beta = 13.6$, $p < 0.05$ when $\text{LFS}_t = \text{LFS}_{t-1} = \text{temporary} \neq \text{TRUE}_t = \text{TRUE}_{t-1} = \text{permanent}$ and $\beta = 19.5$, $p < 0.05$ when $\text{LFS}_t = \text{LFS}_{t-1} = \text{temporary} \neq \text{TRUE}_t = \text{TRUE}_{t-1} = \text{other}$). These parameter estimates correspond to a probability of over 0.99; therefore, the use of PDI cannot increase the probabilities of repeating the error any further (i.e. there seems to be a ceiling effect). This result is not particularly surprising given the short gaps between the waves in the LFS. That is, any misreporting of a contract due to, for instance, confusion is likely to persist over a relatively short period such as three months, provided that no actual change occurred.

5 Conclusions and Discussion

DI is an interviewing technique that is broadly applied in panel surveys to achieve higher longitudinal consistency and lower levels of random measurement error. The importance of minimizing random error in this context stems from the fact that longitudinal survey data are often used to study over time change or transitions; such second-order statistics are known to be highly sensitive to random measurement error. However, while DI helps to mitigate this problem, it potentially introduces a new one, in particular when used proactively, as it has also been hypothesized to increase the incidence of systematic error due to the phenomena of *cognitive satisficing* and *motivated misreporting*.

Given the potentially conflicting effects of PDI on survey data quality, in this paper we examined the effect of this interviewing technique on both the random and systematic components of the error. Our results confirm that PDI reduces the incidence of random error. On the other hand, we find no confirmation for the claim that systematic measurement error is increased due to PDI. To restate, PDI in the LFS is associated with lower probabilities of misreporting a true temporary contract as permanent or "other" type of contract but it is not associated with higher probabilities of repeating the same error over time and it does not lead to spurious stability (i.e. not reporting a true change).

Thus, overall, in our case PDI appears to have a positive effect on data quality as it allows reducing random error while leaving the systematic component of the error unaffected. It can be seen, therefore, as a useful interviewing technique that allows tackling the problem of spurious change. However, it is important to note that in our analysis the probability of repeating the same error was over 0.99, regardless of the interviewing regime (i.e. also in the absence of DI). These results indicate that the level of this systematic error was already so extreme in the Dutch LFS that the use of PDI could not have increased its magnitude any further (i.e. a ceiling effect had occurred). It is, therefore, possible that DI would have had a significant effect on the systematic component of the error, had the baseline probability not been this high. Despite this limitation or shortcoming, the paper still provides important findings for survey methodologists and designers of survey questionnaires that are valid when the survey data exhibits a high probability of error repetition. In such cases PDI is shown to be an attractive option for obtaining information on categorical characteristics in longitudinal surveys as it reduces random measurement error and, given the high "baseline" probability of error repetition, it does not increase systematic error any further. Therefore, in these scenarios PDI reduces measurement error overall and can be a helpful tool in surveys. It is important to note though that our analysis does not allow one to draw conclusions about the effect of PDI on systematic error in situations wherein the level of autocorrelated error at 'baseline' (when using INDI) is not as high as in this case. Further research should focus on examining the impact of PDI in such situations.

It is also worthwhile mentioning that, in our sample the change from PDI to INDI affected a relatively small percentage of records (i.e. 5%). Therefore, future research should also investigate the impact of changes in the interviewing regime on measurement error, when a greater proportion of the population is affected by these changes. When examining their impact, it is also worth going beyond the specific type of PDI used in our analysis and seeing whether the remaining two types of this interviewing method, i.e. *remind, continue* and *remind, confirm*, have similar effects on the quality of the

tremely high or extremely low.

survey data.

References

Alwin, D. F. (2007). *Margins of error: A study of reliability in survey measurement*. John Wiley & Sons.

Bassi, F., Hagenaars, J. A., Croon, M. A., & Vermunt, J. K. (2000). Estimating true changes when categorical panel data are affected by uncorrelated and correlated classification errors: An application to unemployment data. *Sociological Methods & Research*, *29*(2), 230–268.

Biemer, P. (2004). An analysis of classification error for the revised current population survey employment questions. *Survey Methodology*, *30*(2), 127–140.

Biemer, P. (2011). *Latent class analysis of survey error*. John Wiley & Sons.

Biemer, P., & Wiesen, C. (2002). Measurement error evaluation of self-reported drug use: A latent class analysis of the US National Household Survey on drug abuse. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *165*(1), 97–119.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In K. J. Arrow & I. M. D (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Eggs, J., & Jäckle, A. (2015). Dependent interviewing and sub-optimal responding. *Survey Research Methods*, *9*(1), 15–29.

Fuller, W. A. (2009). *Measurement error models*. John Wiley & Sons.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Hagenaars, J. A. (1990). *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis*. Sage Publications, Inc.

Hagenaars, J. A. (1994). *Latent variables in log-linear models of repeated observations*. Sage Publications, Inc.

Hoogendoorn, A. W. et al. (2004). A questionnaire design for dependent interviewing that addresses the problem of cognitive satisficing. *Journal of Official Statistics*, *20*(2), 219–232.

Jäckle, A. (2008). Dependent interviewing: Effects on respondent burden and efficiency of data collection. *Journal of Official Statistics*, *24*(3), 411–430.

Jäckle, A. (2009). *Dependent interviewing: A framework and application to current research*. Wiley Online Library.

Jäckle, A., & Eckman, S. (2019). Is that still the same? has that changed? on the accuracy of measuring change with dependent interviewing. *Journal of Survey Statistics and Methodology*, *8*(4), 706–725.

Jäckle, A., Laurie, H., & Uhrig, S. (2007). *The introduction of dependent interviewing on the British Household Panel Survey*. ISER Working Paper Series.

Jäckle, A., & Lynn, P. (2007). Dependent interviewing and seam effects in work history data. *Journal of Official Statistics*, *23*(4), 529–551.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Lugtig, P., & Lensvelt-Mulders, G. J. (2014). Evaluating the effect of dependent interviewing on the quality of measures of change. *Field Methods*, *26*(2), 172–190.

Lynn, P., Jäckle, A., Jenkins, S. P., & Sala, E. (2006). *The effects of dependent interviewing on responses to questions on income sources* (tech. rep. No. 3).

Manzoni, A., Vermunt, J. K., Luijkx, R., & Muffels, R. (2010). Memory bias in retrospectively collected employment careers: A model-based approach to correct for measurement error. *Sociological methodology*, *40*(1), 39–73.

Mathiowetz, N. A., & McGonagle, K. A. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics*, *16*(4), 401–418.

McLachlan, G., & Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.

Pankowska, P., Bakker, B., Oberski, D. L., & Pavlopoulos, D. (2018). Reconciliation of inconsistent data sources by correction for measurement error: The feasibility of parameter re-use. *Statistical Journal of the IAOS*, *34*(3), 317–329.

Pankowska, P., Bakker, B. F., Oberski, D. L., & Pavlopoulos, D. (2019). How linkage error affects hidden Markov model estimates: A sensitivity analysis. *Journal of Survey Statistics and Methodology*, *8*(3), 483–512.

Pavlopoulos, D., & Vermunt, J. K. (2015). Measuring temporary employment. do survey or register data tell the truth? *Survey Methodology*, *41*(1), 197–214.

Sala, E., Uhrig, S. N., & Lynn, P. (2011). It is time computers do clever things: The impact of dependent interviewing on interviewer burden. *Field Methods*, *23*(1), 3–23.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons.

Uhrig, S. N., & Sala, E. (2011). When change matters: An analysis of survey interaction in dependent interviewing on the British Household Panel Study. *Sociological Methods & Research*, *40*(2), 333–366.

Van de Pol, F., & De Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods & Research*, *15*(1-2), 118–141.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon

(Eds.), *Applied latent class analysis* (Vol. 11, pp. 89–106). Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2013). Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax. *Belmont, MA: Statistical Innovations Inc*.

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*(2), 175–211.

Appendix
Table

Table A1
List of systematic error parameters (estimated and restricted to 0) in the LFS and ER

Observed contract in $t-1$	Latent contract in t	Latent contract in $t-1$	Observed contract in t: Permanent	Observed contract in t: Temporary	Observed contract in t: Other
Permanent	Permanent	Permanent	-	-	-
Permanent	Permanent	Temporary	-	-	-
Permanent	Permanent	Other	-	-	-
Permanent	Temporary	Permanent	-	-	-
Permanent	Temporary	Temporary	$\beta_{b_t,b_{t-1},x_t,x_{t-1}}$	-	-
Permanent	Temporary	Other	-	-	-
Permanent	Other	Permanent	-	-	-
Permanent	Other	Temporary	-	-	-
Permanent	Other	Other	$\beta_{b_t,b_{t-1},x_t,x_{t-1}}$	-	-
Temporary	Permanent	Permanent	-	$\beta_{a_t,a_{t-1},x_t,x_{t-1}}/\beta_{b_t,b_{t-1},x_t,x_{t-1}}$	-
Temporary	Permanent	Temporary	-	$\beta_{a_t,a_{t-1},x_t,x_{t-1}}$	-
Temporary	Permanent	Other	-	-	-
Temporary	Temporary	Permanent	-	-	-
Temporary	Temporary	Temporary	-	-	-
Temporary	Temporary	Other	-	-	-
Temporary	Other	Permanent	-	-	-
Temporary	Other	Temporary	-	$\beta_{a_t,a_{t-1},x_t,x_{t-1}}$	-
Temporary	Other	Other	-	$\beta_{a_t,a_{t-1},x_t,x_{t-1}}/\beta_{b_t,b_{t-1},x_t,x_{t-1}}$	-
Other	Permanent	Permanent	-	-	$\beta_{b_t,b_{t-1},x_t,x_{t-1}}$
Other	Permanent	Temporary	-	-	-
Other	Permanent	Other	-	-	-
Other	Temporary	Permanent	-	-	-
Other	Temporary	Temporary	-	-	$\beta_{b_t,b_{t-1},x_t,x_{t-1}}$
Other	Temporary	Other	-	-	-
Other	Other	Permanent	-	-	-
Other	Other	Temporary	-	-	-
Other	Other	Other	-	-	-