Dependent Interviewing

A Remedy or a Curse for Measurement Error in Surveys?

ITSEW 2018
International Total Survey Error Workshop
June 4 to 6 2018, in Durham, North Carolina, US

Paulina Pankowska, p.k.p.pankowska@vu.nl

Vrije Universiteit Amsterdam, Statistics Netherlands

VU VRIJE UNIVERSITEIT AMSTERDAM

Presentation outline

- Background
  - – Longitudinal surveys & measurement error
  - – Dependent interviewing (DI)
- Our research
  - – Evaluating the effect of DI on:
    - Reliability of questions
    - Model estimates

Longitudinal surveys

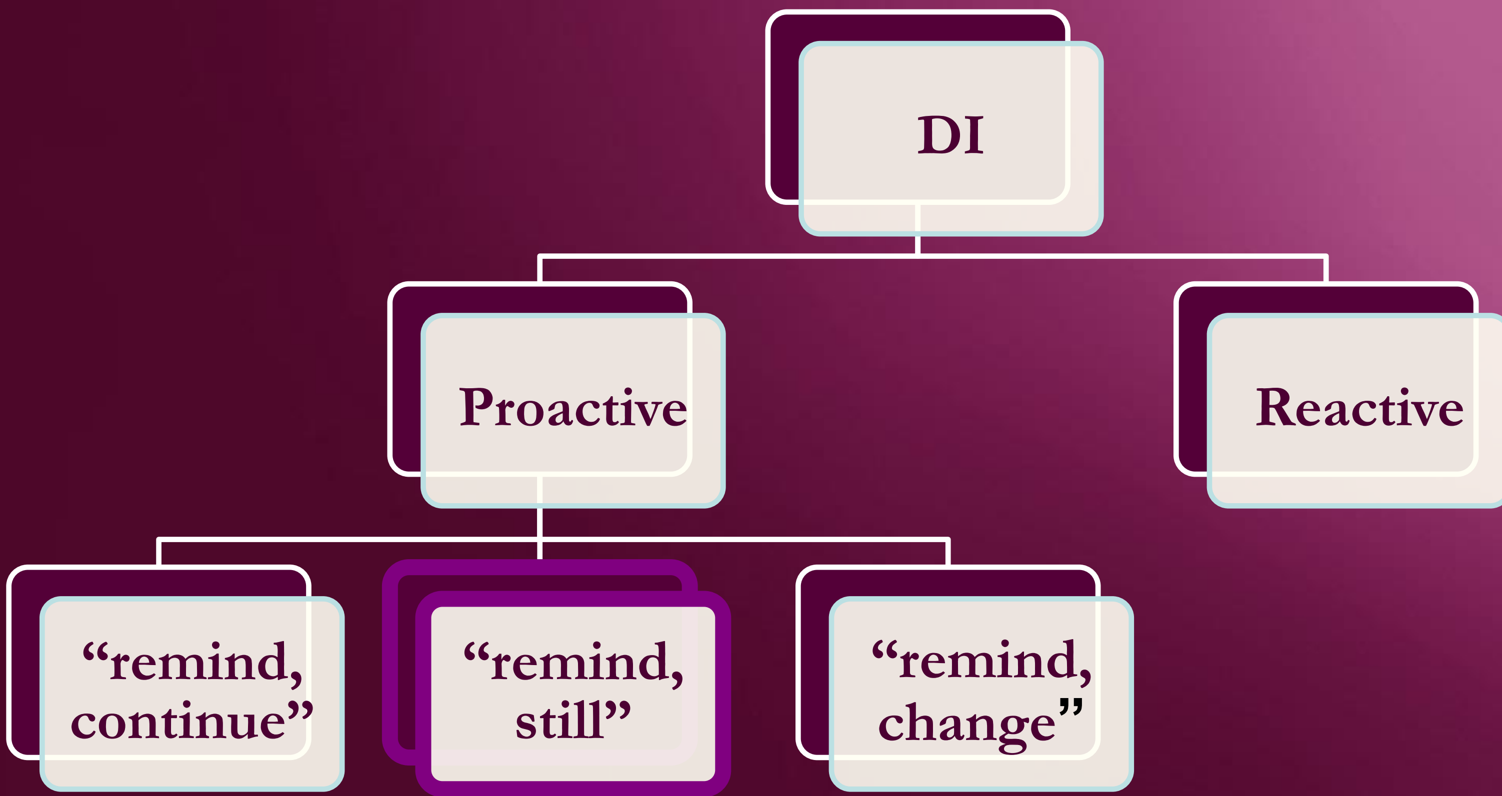
- Collect data for same individuals over time
- Allow studying over-time change/ transitions

But…

- Suffer from measurement error which affects transition estimates
  - Spurious change
- (Potential) solution- **DI**

Dependent interviewing (DI)

- Interviewing technique in which respondents are reminded of previous answers



DI & measurement error

- Expected to reduce (random) error
  - Assists recall ⟶ less spurious change
  - Reduces cognitive burden
- Might lead to systematic error
  - Satisficing ⟶ error carry-over
- Overall effect on reliability and estimates **uncertain**

Our research

- **Focus:** Evaluating the effect of DI on measurement error
- **Method:** Using latent class modelling to estimate misclassification rates and bias under DI and INDI
- **Data:**
  - Linked data from the Dutch LFS and ER
  - Main variable of interest- individuals' contract type (3 categories: permanent, temporary, other)

Our research

- **Temporary contracts**
  - **DI:**
    - In place until end of 2009
    - PDI- “remind, still”
    - Only if no job change occurred
  - **INDI:**
    - In place since 2010
    - Also in 2009 if job change occurred

“Last time you had a temporary contract. Is this still the case?”

“Do you currently have a permanent contract?”

Our research

- **Permanent contracts**
  - No job change occurred- value carried over
  - Job change occurred- question asked
- **A total of 3 scenarios:**

Job change occurred	**2009**	**2010**
Yes	**(I)** IND for temp and perm	
No	**(II)** DI for temp and value copied for perm	**(III)** INDI for temp and value copied for perm

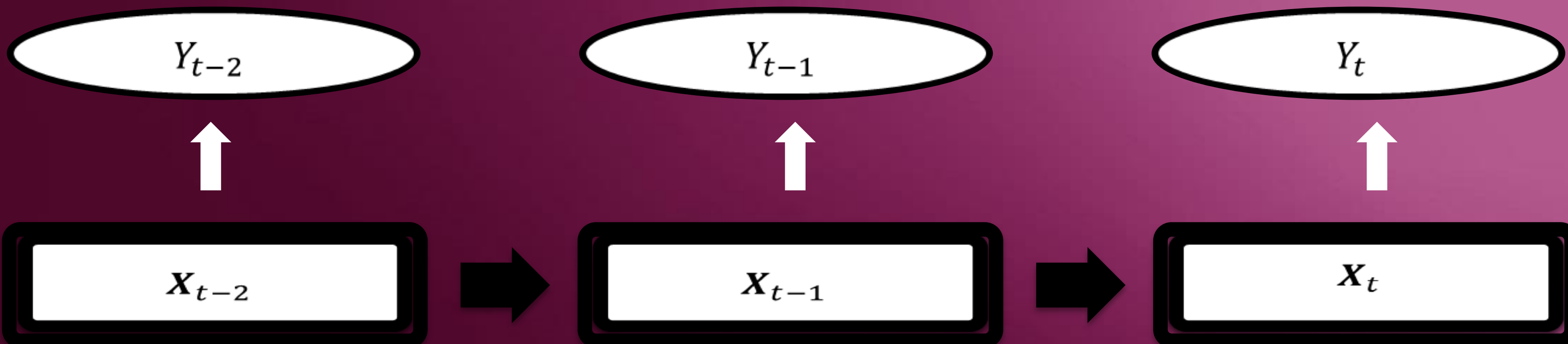
Latent Class Modelling (LCM)

- **Latent class analysis (LCA):**
  - Applied to categorical, cross-sectional data
  - Uses multiple conditionally independent indicators
  - Separates true value from error

- **Hidden Markov Models (HMMs):**
  - Applied to categorical, longitudinal data
  - Uses repeated measures over time
  - Separates true change from error

Hidden Markov Models (HMMs)

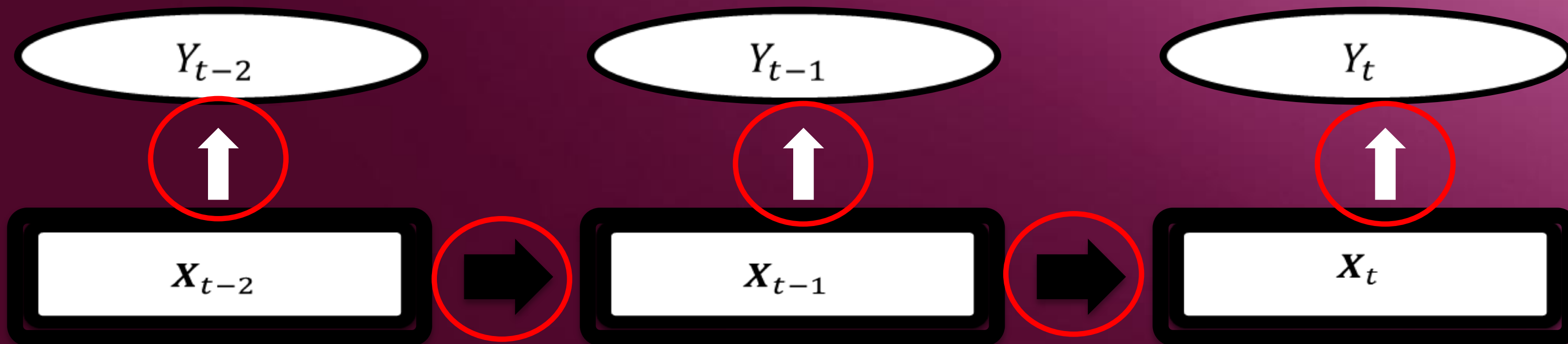
- HMMs **consist of:**
  - a structural part (latent/ true)
  - a measurement part (observed)



$$\boldsymbol{P}(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \sum_{x_0=1}^{k} \sum_{x_1=1}^{k} \dots \sum_{x_T=1}^{k} P(X_{i0} = x_0) \prod_{t=1}^{T} P\left(X_{it} = x_t \middle| X_{i(t-1)} = x_{t-1}\right) \prod_{t=1}^{T} P(Y_{it} = y_t | X_{it} = x_t)$$

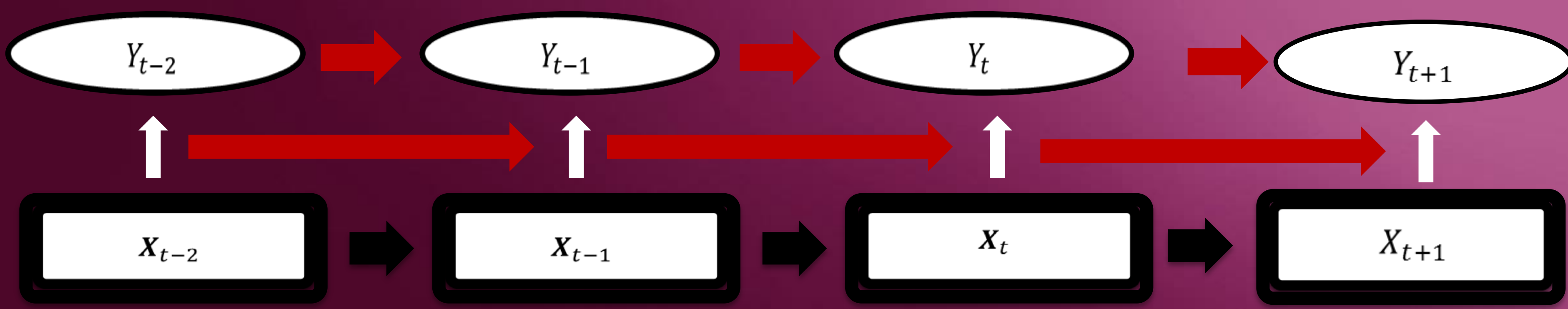
Hidden Markov Models (HMMs)

- HMMs assume that:
  - The latent state at time t only depends on the latent state at time t-1 - ***the Markov assumption***
  - The observed state at time t only depends on the latent state at time t - ***the local independence assumption***



HMMs & DI

- DI (might) lead to autocorrelation of error
- Need to relax the local independence assumption

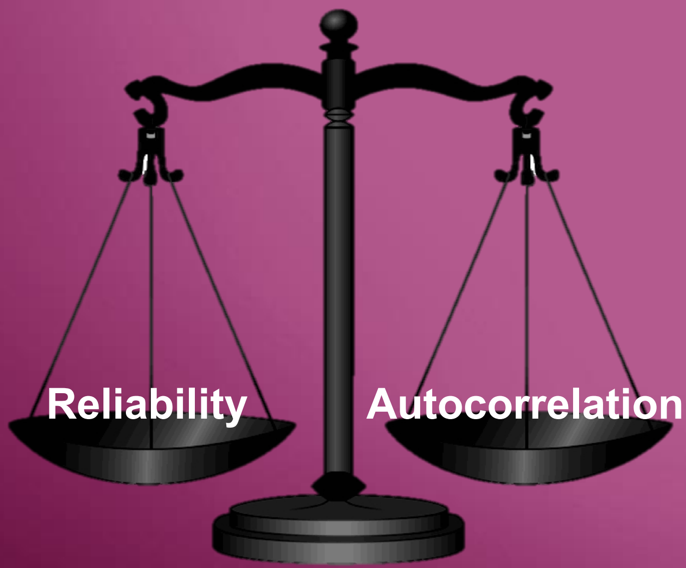


HMMs & DI

- Impossible to model autocorrelation **and** time-variant misclassification rates

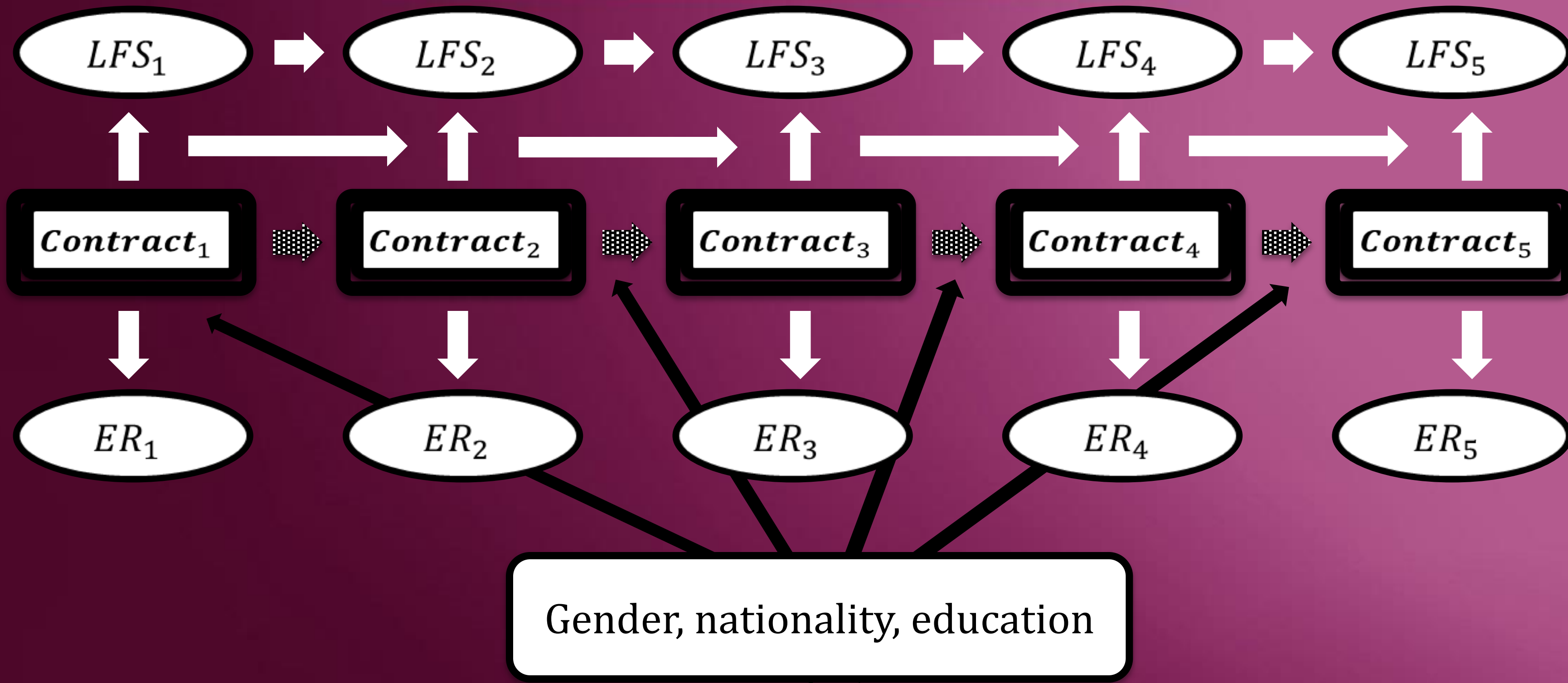
↓

- Identifiability requires constant reliability



- **Unless** there are multiple indicators per time point (allow modeling both)

Our model



$$\boldsymbol{P(C_i = c_i,\ E_i = e_i |\ Z_i)} = \sum_{x_0=1}^{3}\sum_{x_1=1}^{3}\dots\sum_{x_T=1}^{3} P(X_{i0} = x_0 |\ Z_i)$$

$$\prod_{t=1}^{T} P\left(X_{it} = x_t \middle| X_{i(t-1)} = x_{t-1}, Z_i\right)$$

$$P(E_{i0} = e_0 | X_{i0} = x_0)$$

$$\prod_{t=1}^{T} P(E_{it} = e_t | X_{it} = x_t,\ X_{i(t-1)} = x_{t-1},\ E_{i(t-1)} = e_{t-1})$$

$$\prod_{t=1}^{T} P(C_{it} = c_t | X_{it} = x_t)$$

Data

- Linked dataset with information from LFS and ER

- Sample consists of 20,760 individuals aged 25 to 55 who
  I. Started LFS in first quarter of 2009 (8,536)- ***DI***
  II. Started LFS in last quarter of 2010 (12,080)- ***INDI***

- Dataset contains quarterly information on each individual for 5 time points

Results

- Misclassification (measurement error) rates

Scenario		Overall	Temporary contracts	% of cases
(I) Job change (2009/ 2010)-	**INDI**	0,13	0,22	20%
(II) No job change & 2009-	**DI**	**0,11**	**0,28**	**24%**
(III) No job change & 2010-	**INDI**	0,11	0,25	17%

- Probability of carrying-over the same error for DI

Observed at t-1	Latent at t-1	Latent at t	Observed at t		
			Perm	Temp	Other
Temporary	Permanent	Permanent	0,10	**0,90**	0,00
Temporary	Other	Other	0,00	**0,87**	0,13

Results

- Misclassification rates for those who had or would have had DI (overall and by wave)

Scenario	Overall	Wave				
		1	2	3	4	5
(II) No job change & 2009- DI	**0,301**	-	0,301	0,290	0,307	0,306
(III) No job change & 2010- INDI	**0,297**	-	0,291	0,295	0,300	0,305

Results

- Latent and observed contract ditributions

Type of contract	2009		2010	
	Observed	Latent	Observed	Latent
Permanent	0,69	0,61	0,68	0,60
Temporary	0,05	0,14	0,05	0,14
Other	0,26	0,25	0,27	0,26

- Latent and observed contract transition rates

Transition rate	2009		2010	
	Observed	Latent	Observed	Latent
Temp to perm	0,07	0,05	0,11	0,05

Conclusions

- Overall, DI appears to have a marginal effect on data quality
- DI associated with high probability of the copying of errors
- DI reduces random error (spurious change)

Thank you!

- Paulina Pankowska, p.k.p.pankowska@vu.nl
- Bart Bakker, bfm.bakker@cbs.nl
- Peter Lugtig, p.lugtig@uu.nl
- Daniel Oberski, d.l.oberski@uu.nl
- Dimitris Pavlopoulos, d.pavlopoulos@vu.nl