

# Estimating employment transitions while accounting for classification error<sup>1</sup>

## Evidence from South Africa <sup>2</sup>

Zander Prinsloo <sup>1</sup>    Rulof Burger <sup>2</sup>    Cobus Burger <sup>2</sup>    Dean Jolliffe <sup>1</sup>

<sup>1</sup>Development Economics, Data Group  
World Bank

<sup>2</sup>Department of Economics  
Stellenbosch University

December 8, 2025

---

<sup>1</sup>We gratefully acknowledge financial support from the UK government through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Programme.

<sup>2</sup>The findings, interpretations, and conclusions expressed in this presentation are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

# Key messages

- Survey data are measured with error
- Classification error in employment status upwardly biases estimates of employment transition rates
- Small error rates cause large biases
- It is important to know true churn
- Classic transition matrices exaggerate job entry and exit rates and underestimate chronic unemployment

# Our approach

- Develop novel structural estimator that identifies true transition rate in presence of misclassification error from three waves of panel data
- Extend model to allow alternative explanations for violation of Markov property:
  - ▶ duration dependence
  - ▶ individual heterogeneity
- Present further evidence of misclassification exaggerating churn:
  - ▶ More restrictive matching algorithms produce lower transition rates
  - ▶ Respondents who provide unreliable responses to other questions have higher transition rates

- Quarterly Labour Force Survey (QLFS)
- Estimated transition rate (across successive quarters):
  - ▶ Job entry rate: 8.63%
  - ▶ Job exit rate: 8.20%

## Effects of misclassification: violation of Markov property

- Markov assumption implies that only current employment matters for future transitions
- Under this assumption, high observed two-wave transition rates predict very high three-wave transition rates
- However, observed three-wave transition rates are much lower than predicted: violation of Markov property

Table: Transition rates

	Observed 2-wave	Predicted 3-wave	Observed 3-wave
Job Entry	8.63%	15.80%	11.10%
Job Exit	8.20%	15.02%	10.20%

# Simple Markov model of misclassified employment process

- Binary but unobservable period  $t$  true employment status,  $s_t^*$  (1 if employed, 0 otherwise) generated as:

$$s_t^* \equiv \theta_1 s_{t-1}^* + \theta_2 (1 - s_{t-1}^*) + e_t^* \quad (1)$$

- $\theta_2$  is job entry rate and  $1 - \theta_1$  is job exit rate
- $e_t^*$  is unobservable Bernoulli distributed error term such that  $E(e_t^* | S_{t-1}^*) = 0$
- Binary observable employment status determined as

$$s_t \equiv \pi s_t^* + (1 - \pi)(1 - s_t^*) + e_t \quad (2)$$

- $\pi$  is misclassification probability

# Structural estimator: results

Table: Implied probabilities

	(1)	(2)	(3)
Entry rate	8.42*** (0.14)	2.92*** (0.19)	2.92*** (0.19)
Exit rate	8.32*** (0.14)	2.88*** (0.19)	2.89*** (0.19)
Misclassification rate		2.99*** (0.13)	
Misclass. rate: non-employed			2.92*** (0.16)
Misclass. rate: employed			3.06*** (0.16)
Misclassification	No	Symmetric	Asymmetric
LL	-29126	-28479.4	-28479.1
Observations	22,963	22,963	22,963

Note:

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

# Alternative explanations for Markov violation

- Reasons other than misclassification error for failure of Markov property:
  - 1 Worker heterogeneity
  - 2 Transitions can be duration-dependent

# Alternative explanations: observed heterogeneity

Table:

	(1)	(2)
Entry rate: period 2	4.44*** (0.002)	5.02*** (0.003)
Entry rate: period 3	4.44*** (0.002)	5.02*** (0.003)
Exit rate: period 2	5.08*** (0.003)	5.54*** (0.004)
Exit rate: period 3	5.01*** (0.003)	5.48*** (0.004)
Misclassification rate	2.38*** (0.11)	2.22*** (0.10)
Covariates	Educ + Age	+ Race + Gender
LL	-25279	-24811.8
Observations	22,245	22,245

Note:

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

## Alternative explanations: AR(2) employment process

- Use four waves of QLFS data
- Tendency for recent entries and exits to reverse
- Significant but smaller misclassification rate
- Classification error still makes churn appear 25% higher

## Additional evidence: matching algorithms

- Mismatching can cause misclassification
- If estimated employment transitions are less frequent when using more stringent matching criteria, this would support hypothesis of upwardly biased transition rate estimates due to misclassification error
- Investigate three matching criteria:
  - 1 Same household ID and person number
  - 2 Same household ID, race, gender, and age
  - 3 Same household ID, person number, race, gender, age

## Additional evidence: matching algorithms

**Table:** Implied probability estimates from structural model on data from different matching algorithms

	(1) First	(2) Second	(3) Third
Miscl. rate	5.30%*** (0.011%)	3.38%*** (0.012%)	3.32%*** (0.012%)
Job entry	4.87%*** (0.019%)	3.70%*** (0.022%)	3.97%*** (0.022%)
Job exit	3.82%*** (0.016%)	2.77%*** (0.020%)	2.36%*** (0.019%)
Empl. rate	51.73%*** (0.018%)	50.00% (0.030%)	52.71%*** (0.032%)
Descriptive entry	13.74	9.25	9.58
Descriptive exit	13.45	9.14	8.28
Log likelihood	-13,062,960	-4,195,539	-3,467,521
Misclassification	Yes	Yes	Yes

Note: '\*\*\*' as 0.001, '\*\*' as 0.01, '\*' as 0.05

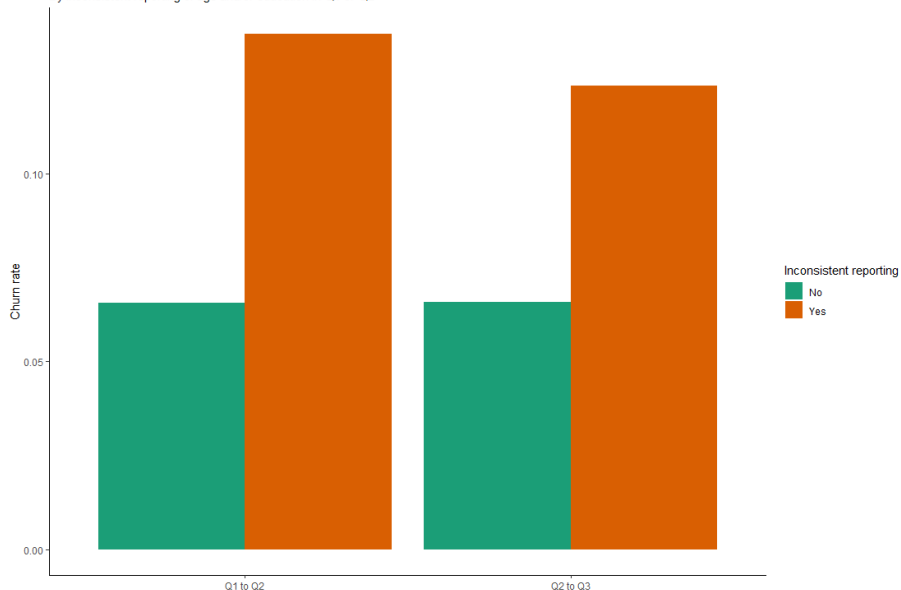
## Additional evidence: inconsistent responses

- Unreliable responses in these questions may be correlated to unreliable responses in employment status
- Inconsistency when:
  - 1 worker's age decreases, or increases by more than one year between quarters,
  - 2 worker's years of completed schooling years decreases, or increases by more than one year between quarters

# Additional evidence: inconsistent responses

## Churn Rates

By inconsistent reporting of age and/or education in Q1 or Q2



## Additional evidence: inconsistent responses

Table: Structural inconsistencies regression results: implied probabilities

	(1)
Miscl. Rate	2.77%*** (0.07%)
Miscl. Rate if Inconsistent	13.98%*** (0.45)
Job Entry	3.08%*** (0.12%)
Job Exit	2.35%*** (0.12%)
Empl Rate	49.01%*** (0.20%)
Log likelihood	-88 313
Misclassification	Yes
Age Inconsistencies	Yes

Note: '\*\*\*' as 0.001, '\*\*' as 0.01, '\*' as 0.05

# Additional evidence: matching algorithms and inconsistent responses

- Observed employment transition rates are lower:
  - ▶ when using more stringently matched panel data
  - ▶ for individuals who provide consistent responses to age and education question across successive waves
- In both cases, estimated transition rates are still higher than suggested by structural estimates

# Conclusion

- Strong evidence of small but significant misclassification error in employment outcome
- Even small misclassification rate can severely bias estimated transition rates
- Structural model uses three wave of panel data to estimate misclassification rate of 3%
- Estimate is robust to various identifying assumptions
- South African job entry and exit rates are overestimated by nearly a factor of 3
- This also has obvious implications for other estimated transition rates (e.g. poverty transitions)