

# A Capture-Recapture Model with Temporary Emigration for Estimating Population Size from Incomplete Registers

**Lucy Brown<sup>1</sup>**: lyb3@kent.ac.uk

Eleni Matechou<sup>1</sup>

Bruno Santos<sup>2</sup>

Eleonora Mussino<sup>3</sup>

*1 University of Kent, UK*

*2 University of Lisbon, Portugal*

*3 Stockholm University, Sweden*



# Motivating Case Study

## Overcoverage:

- Due to imperfect emigration and/or death registration
- Leads to serious bias in population estimates
- Negatively influences policy-making and research

## Overcoverage Estimation:

- Existing approaches<sup>[1][2]</sup> in Sweden rely on multiple systems estimation (MSE) and only consider annual snapshots of the register data
- Instead, we have employed a longitudinal approach, following individuals, and hence registers, over different years

[1] Andrea Monti, Sven Drefahl, Eleonora Mussino, and Juho Härkönen. Over-coverage in population registers leads to bias in demographic estimates. *Population Studies*, 74(3):451–469, 2020

[2] Eleonora Mussino, Bruno Santos, Andrea Monti, Eleni Matechou, and Sven Drefahl. Multiple systems estimation for studying over-coverage and its heterogeneity in population registers. *Quality & Quantity*, pages 1–24, 2023

# Swedish Register Data

- Provided by the Swedish National Institute of Statistics (Statistics Sweden - SCB)
- All foreign-born residents who first entered Sweden between the years 2003 and 2016 as adults

(This is not real data)

Nine Registers

Emigration, (Re-)Immigration and Death records

emp	stud	intmove	faminc	amf	child
0	0	0	0	1	0
1	0	0	1	0	0
1	0	0	1	0	0
1	0	0	1	0	0
1	0	0	1	0	0

immig	emig	reimmig	marr	div	swecit
1	0	0	0	0	0
0	1	0	0	0	0
1	0	1	0	0	0
0	0	0	1	0	0
0	1	0	0	0	0

id	year	age	cob	sex	death	
1	XXX1	2003	27	France	1	0
2	XXX1	2004	28	France	1	0
3	XXX1	2007	31	France	1	0
4	XXX1	2008	32	France	1	0
5	XXX1	2009	33	France	1	0
6	XXX2	2014	68	India	0	0
7	XXX2	2015	69	India	0	1

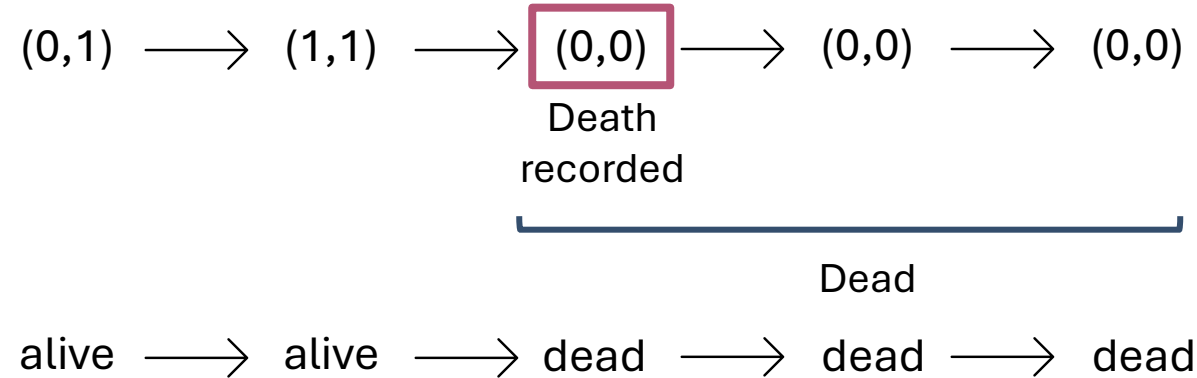
Covariates (treated as categorical):

- Sex
- Country of birth
- Age
- Time Since first entering Sweden

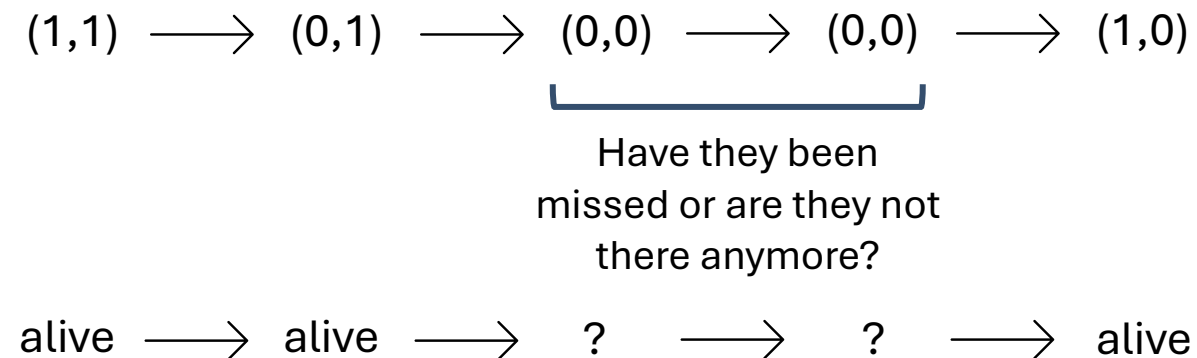
# Observations

An individual is observed on some **combination of R registers**, on which they can be observed (1) or not observed (0), i.e. 2 possible outcomes for each register  $\Rightarrow 2^R$  register combinations

Example Person #1



Example Person #2



# Capture-Recapture & HMM Formulation

## Capture Recapture (CR) Models<sup>[3]</sup>:

1. An initial capture is made
2. Unmarked individuals are marked in some unique way (tags, rings etc)
3. Individuals are released back into the population
4. Subsequent captures are made, allowing individuals to be tracked over time

[3] Kenneth H. Pollock. Capture-recapture models. Journal of the American Statistical Association, 95(449):293–296, 2000. ISSN 01621459, 1537274X.  
URL <http://www.jstor.org/stable/2669550>

- CR models are commonly used in ecology to estimate population size for wild animals
- An individual's true state is unobservable (latent)
- Hidden Markov model (HMM) formulation marginalises over the latent state to calculate the marginal likelihood
- This iterative method is very efficient and allows latent states to be inferred

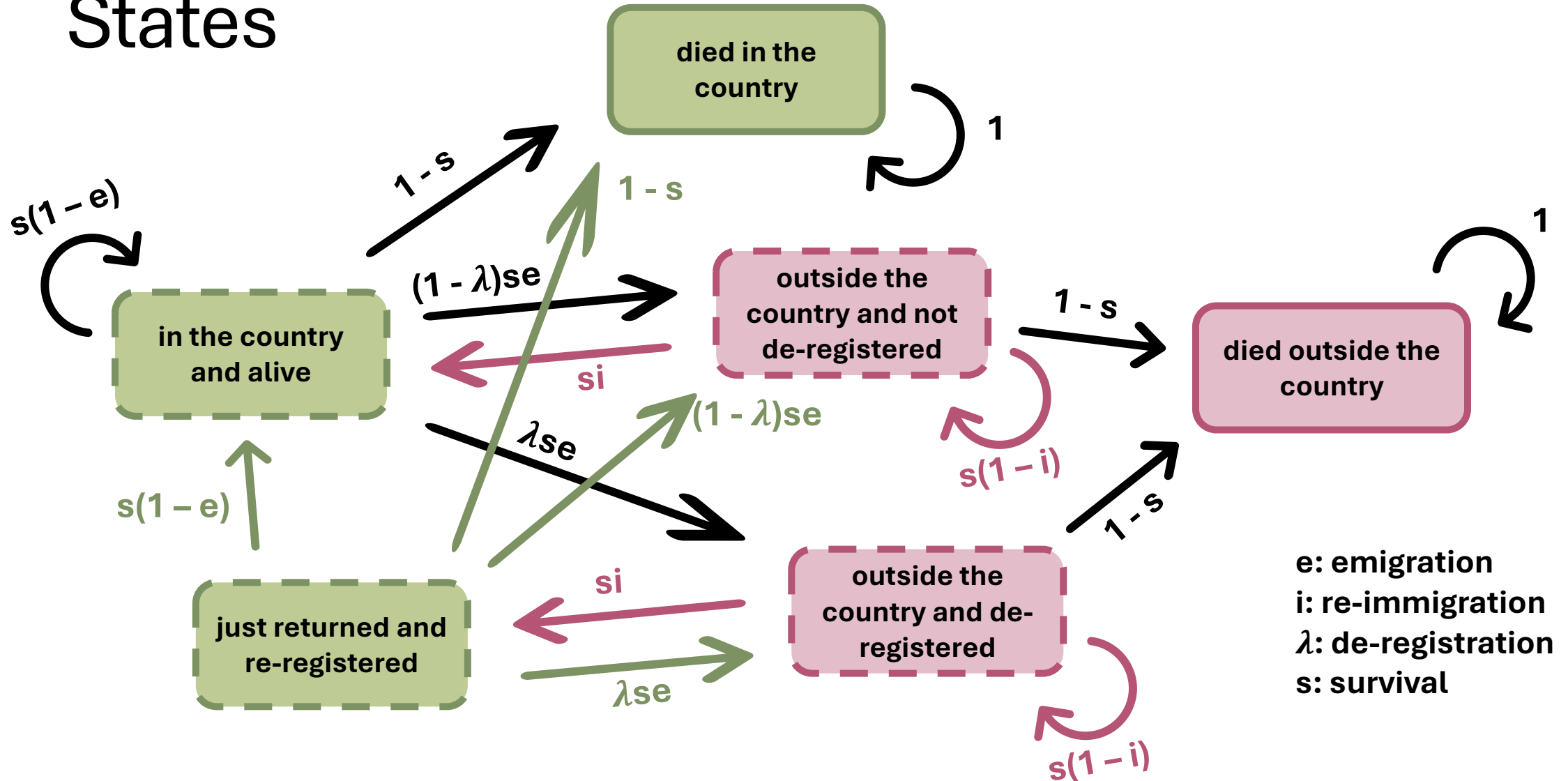
Probability of history at the previous time point

Probability of transitioning to each new state given the state at  $t - 1$

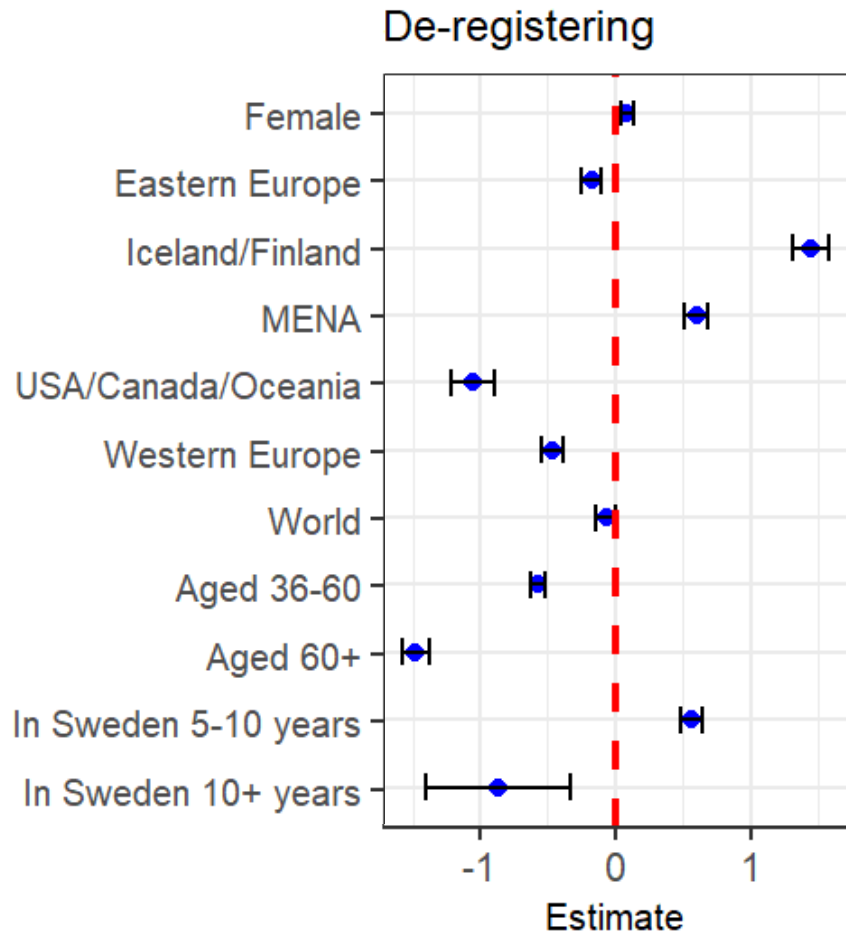
Observation probabilities dependent on the current state

$$\alpha_1 = \delta P(y_1)$$
$$\alpha_t = \alpha_{t-1} \Gamma_{t-1} P(y_t)$$
$$\Rightarrow L_T = \alpha_T \mathbf{1}'$$

# States



# De-registering Probability



**Baseline:** men aged 18-35, born in Denmark/Norway who first entered Sweden less than 5 years ago

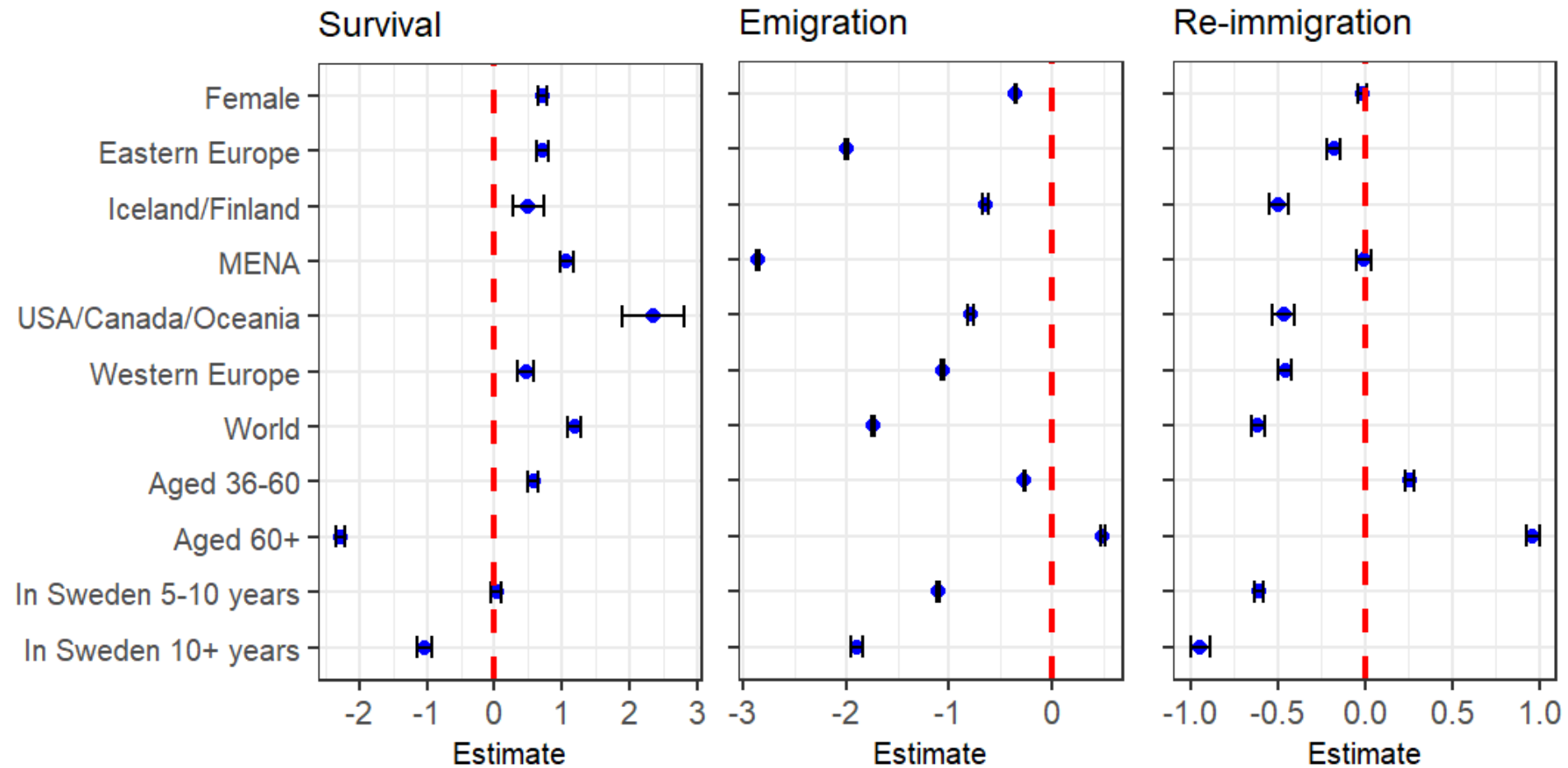
$$\text{logit}(\lambda) = -0.332 \Rightarrow \lambda = 0.418$$

**Women** aged 18-35, born in **Iceland/Finland** who first entered Sweden **5-10 years ago**

$$\text{logit}(\lambda) = -0.332 + 0.088 + 1.452 + 0.568 \Rightarrow \lambda = 0.855$$

Men **aged 60+**, born in **USA/Canada/Oceania** who first entered Sweden **10+ years ago**

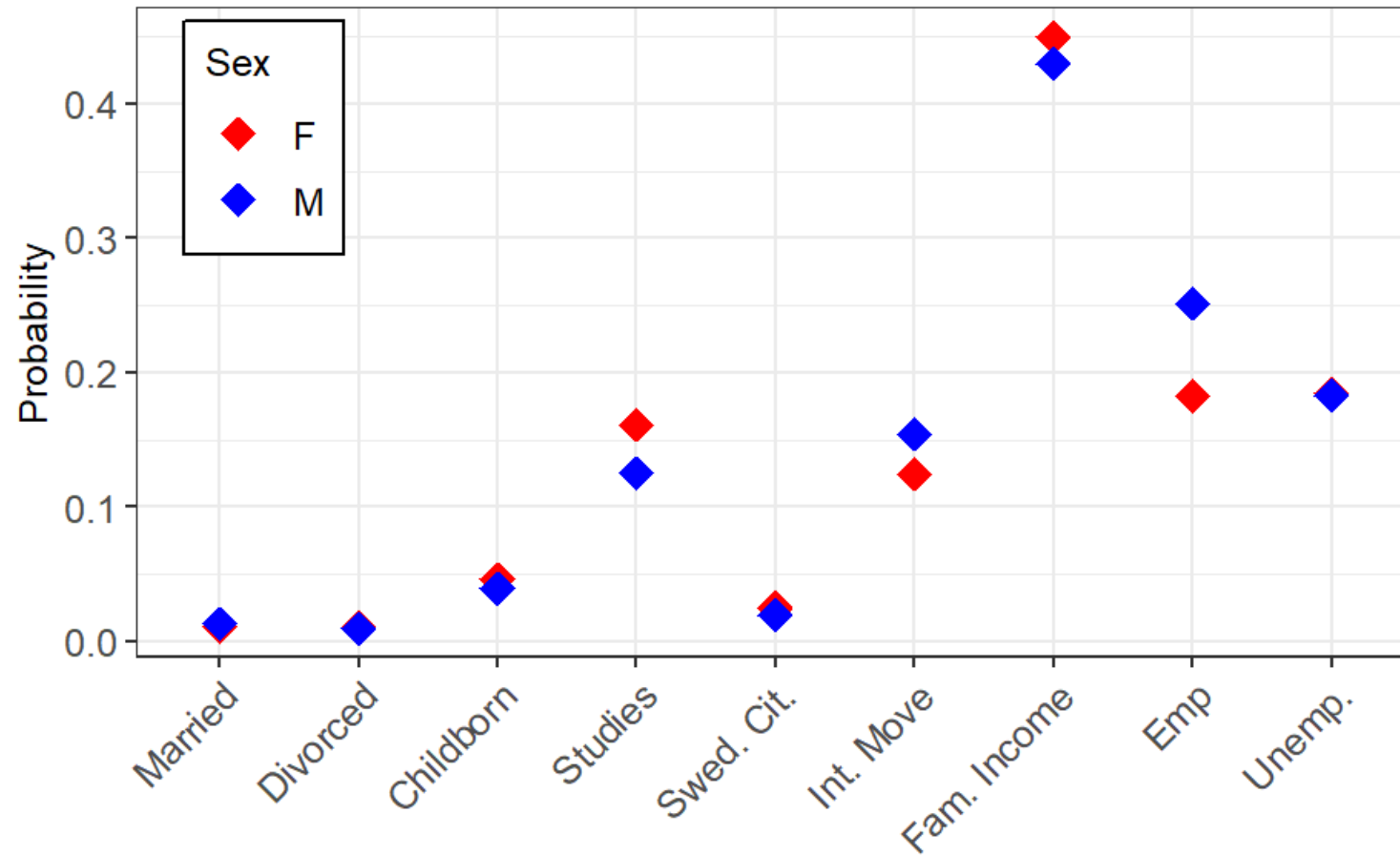
$$\text{logit}(\lambda) = -0.332 - 1.053 - 1.474 - 0.865 \Rightarrow \lambda = 0.024$$



**Baseline:** men aged 18-35, born in Denmark/Norway who first entered Sweden less than 5 years ago

- Survival: 0.996 (0.996, 0.997)
- Emigration: 0.395 (0.391, 0.398)
- Re-immigration: 0.053 (0.052, 0.055)

## Register Observation Probability by Sex



We obtain observation probability estimates for each register and each combination of registers.

Probability of being unobserved:

- Male: 0.0251
- Female: 0.0220



Probability of being observed in at least one register:

- Male: 0.9749
- Female: 0.9780

# Future Work

- ★ Experiment with methods to calculate population size estimates e.g. Viterbi Algorithm
- ★ Considering further extensions such as dependence between family units
- ★ Application to equivalent data from Norway, provided by Statistics Norway
  - Additional complexity that checks are regularly done and individuals manually removed if not present in the country



Thank you for your attention!

Email: [lyb3@kent.ac.uk](mailto:lyb3@kent.ac.uk)

