1 Data generating process

The data generating process follows a first-order Markov process.

- $y^*_{i,t} = \alpha + \beta x_{i,t} + \gamma_1 y_{i,t-1} + c_i + u_{i,t}$
- $i = 1, ..., N$ und $t = 1, ..., T$
- $y = 1$, if $y^*_{i,t} > 0$

with:

1. $u_{i,t} \overset{iid}{\sim} N(0, \sigma_e^2)$
2. $c_i \sim N(0, 1)$
3. $x_{i,t} \sim N(0, 1)$
4. $y^*_{i,0} == \alpha + \beta x_{i,t} + \gamma_1 y_{i,t-1} + 0.5 * c_i + 0.5 * z_i + u_{i,t}, z_i \overset{iid}{\sim} N(0, 1)$

- Parameter values:
  - $\gamma_1 = 0.5$, $\beta = -0.5$, $T = 10$, $N = 1000$, number of repetitions $= 1000$

2 Study

- **Model 1**: Wooldridge - Estimator, balanced sample
  - $\hat{\gamma}_1 = 0.5000074$
- **Model 2**: Wooldridge - Estimator, unbalanced sample, all observations for $t = 6$ are dropped (for $t = 7$, the information on y_{t-1} is available)
  - $\hat{\gamma}_1 = 0.5826112$
- **Model 3**: Wooldridge - Estimator, unbalanced sample, 50 percent of the observations are dropped if $t >= 6$
  - $\hat{\gamma}_1 = 0.4988575$
- The results show that using the Wooldridge-Estimator, an unbalanced samples leads to reliable parameter estimates when the individual drops out of the panel after attrition (Model 3). In contrast, if the observation reenters the unbalanced panel, the estimator is biased (Model 2).
- For questions on the analysis please contact alexander.mosthaf(at)googlemail.com