Multinomial logistic regression with fixed effects

Klaus Pforr

GESIS – Leibniz-Institute for the Social Sciences

July 16, 2015
Motivation

Why fixed effects?

- Reduce omitted variable bias
- Unobserved heterogeneity can be related with observed covariates

Why multinomial logit?

- fixed effects models implemented for continuous, binary, count data dependent variables
- polytomous categorical dependent variables in all sub-disciplines of social sciences
Statistical model by Chamberlain (1980)

What is the femlogit?

mlogit across T with unobserved time-constant tendency towards each alternative

Assumptions

• Mlogit-Link: \( \Pr(y_{it} = o_j) = \frac{\exp(\alpha_{ij} + x_{it}\beta_j)}{\sum_{k=1}^{J} \exp(\alpha_{ik} + x_{it}\beta_k)} \) mit \( \alpha_{iB} = \beta_B = 0 \)

• Strict exogeneity: \( f_{y_{it}|x_{i1},...,x_{iT1},\alpha_i} = f_{y_{it}|x_{it},\alpha_i} \)

• Conditional independence across time: \( \forall s, t: f_{y_{is}|x_{is},\alpha_i} \perp f_{y_{it}|x_{it},\alpha_i} \)

No assumption on relationship between unobserved heterogeneity and covariates \( f_{\alpha_i|x_{i1},...,x_{iT1}} \)
Estimation

Problem of unobserved heterogeneity $\alpha_i$:

$\Rightarrow$ Solution by Chamberlain (1980)

- Frequency of alternative $j$ is sufficient statistic for individual tendency $\alpha_{ij}$ towards alternative $j$

- Probability of complete time series $(y_{i1}, \ldots, y_{iT_i})$ conditional on sufficient statistic of inclinations towards alternatives

$$
\Pr \left( y_i | \sum_t \delta_{y_it, o_1}, \ldots, \sum_t \delta_{y_it, o_J} \right) = \frac{\prod_{t=1}^{T_i} \prod_{j \neq B} \exp(x_{it/B}) \delta_{y_{it}, o_j}}{\sum_{v_i \in \tau_i} \left( \prod_{t=1}^{T_i} \prod_{j \neq B} \exp(x_{it/B}) \delta_{v_{it}, o_j} \right)}
$$

$\Rightarrow$ Unobserved heterogeneity is canceled out
Estimation – cont.

Log-likelihood function

$$E(\ln \ell_i(\beta)) = \frac{1}{N} \sum_{i=1}^{N} \ln \frac{\exp(\sum_{t=1}^{T_i} \sum_{j \neq B} \delta_{y_{itj}} x_{itj} \beta_j)}{\sum_{\nu_i \in \gamma_i} \exp(\sum_{t=1}^{T_i} \sum_{j \neq B} \delta_{\nu_{itj}} x_{itj} \beta_j)}$$

Estimation with maximum likelihood algorithm

$$\hat{\beta}_{ML} = \max_{\beta}(E(\ln \ell_i(\beta)))$$
Implementation

Estimation until now

Workaround solution with data transformation trick and binary fixed effects logit by Börsch-Supan (1987)

⇒ Only feasible for small N, short T, and few alternatives

Now available: femlogit

• First general implementation of femlogit model
• Easy and ready-to-use implementation in widely used software Stata

\texttt{femlogit depvar [indepvars] [if] [in], group(varlist) /*}
\texttt{*/ [baseoutcome(#) constraints(clist) difficult /*}
\texttt{*/ or robust]}

Application 1: Effect of Social Class Status on Party Identification

Data & Model

- Inspired by Kohler (2002)
- SOEP 2007–2012
- Information about
  - Party identification
  - Social class (EGP)
  - Employment status, business size, civil service, gross earnings, family status, # kids in hh, age, education, country of birth
  - Effect of EGP class status on party identification (alternatives: Soc. Democ., Christ. Democ., Liberal, Greens, Socialist, Radical Right, Other, No Ident.)

- Advantage of femlogit: Implicit control for all variables at voter-level constant across waves
Application 1: Effect of Social Class Status on Party Identification

Results

- Controls: Employment status, business size, civil service, gross earnings, marital status, #kids in hh, age, education, country of birth
- Date unweighted
Application 2: Effect of Smoking during Pregnancy on Length of Gestation

Data & Model

- Inspiration and data by Abrevaya (2006)
- Multi-level data: children nested in mothers
- Information about
  - gestation age
  - mothers’ smoking behavior during pregnancy
  - prenatal care (Kessner index, # doctoral visits)
  - mothers’ sociodemographic background
- Effect of Smoking on odds of pre-term birth vs. full term birth vs. post-term birth
- Advantage of \texttt{femlogit}: Implicit control for all variables at mother-level constant across children
Application 2: Effect of Smoking during Pregnancy on Length of Gestation

Results

- Controls: Prenatal care, # doctor visits, marital status, education, race
- Date unweighted
Conclusion

• First implementation of multinominal logit with fixed effects in widely used software
• Implementation works good with large N and small T
• Problem of unobserved heterogeneity in many applications in social sciences
  • Effect of social class of party identification partly overestimated
  • Effect of smoking on gestation age partly overestimated
Thank you for your attention!

Download ado with: findit femlogit
Literature


