The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	\$115	1-year subscription	\$145
2-year subscription	\$210	2-year subscription	\$270
3-year subscription	\$285	3-year subscription	\$375
1-year student subscription	\$ 85	1-year student subscription	\$115
1-year institutional subscription	\$345	1-year institutional subscription	\$375
2-year institutional subscription	\$625	2-year institutional subscription	\$685
3-year institutional subscription	\$875	3-year institutional subscription	\$965
Electronic only		**Electronic only**	
1-year subscription	\$ 85	1-year subscription	\$ 85
2-year subscription	\$155	2-year subscription	\$155
3-year subscription	\$215	3-year subscription	\$215
1-year student subscription	\$ 55	1-year student subscription	\$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

femlogit—Implementation of the multinomial logit model with fixed effects

Klaus Pforr
GESIS—Leibniz–Institute for the Social Sciences
Mannheim, Germany
klaus.pforr@gesis.org

Abstract. Fixed-effects models have become increasingly popular in social-science research. The possibility to control for unobserved heterogeneity makes these models a prime tool for causal analysis. Fixed-effects models have been derived and implemented for many statistical software packages for continuous, dichotomous, and count-data dependent variables. Chamberlain (1980, *Review of Economic Studies* 47: 225–238) derived the multinomial logistic regression with fixed effects. However, this model has not yet been implemented in any statistical software package. Possible applications would be analyses of effects on employment status, with special consideration of part-time or irregular employment, and analyses of effects on voting behavior that implicitly control for long-time party identification rather than measuring it directly. This article introduces an implementation of this model with the new command `femlogit`. I show its application with British election panel data.

Keywords: st0362, femlogit, multinomial logit, fixed effects, panel data, multilevel data, unobserved heterogeneity, discrete choice, random effects, conditional logit

1 Introduction

Fixed-effects models have become increasingly popular in sociology. The possibility to control for unobserved heterogeneity makes these models a prime tool for causal analysis (Gangl 2010; Brüderl and Ludwig 2015). Fixed-effects models for continuous, dichotomous, and count dependent variables are widely used and available in Stata and many other software packages. However, a fixed-effects estimator for polytomous discrete dependent variables is not yet available for any statistical software package (Allison 2009, 44). The available alternatives for such dependent variables are the pooled multinomial logistic or probit regression (Wooldridge 2010, 609; Rabe-Hesketh and Skrondal 2012, 653–658) and the multinomial logistic or probit regression with random effects (Wooldridge 2010, 619ff.; Rabe-Hesketh and Skrondal 2012, 659ff.). For both models, we must assume that any unobserved heterogeneity is independent of the observed covariates.

In this article, I present an implementation of the multinomial logistic regression with fixed effects (`femlogit`) in Stata. The `femlogit` command implements an estimator by Chamberlain (1980). The implementation draws on the native Stata multinomial logit and conditional logit model implementations. The actual `ml` evaluator uses Mata functions to implement the conditional likelihood function.

 st0362

Possible applications of the fixed-effects estimator include analyses of effects on employment status, with special consideration of part-time or irregular employment, and analyses of the effects on voting behavior that implicitly control for stable individual differences in party preference rather than measuring it directly.

After explaining the mathematical background and the implementation of the model, I will discuss the syntax of `femlogit`. Then I will show the application of the ado-file and the interpretation of its results with a model of voting behavior with British election panel data.

2 Statistical model

The statistical model was first proposed by Chamberlain (1980, 231). More extensive expositions are found in Lee (2002, 143ff.) and Pforr (2013). Here I assume a sample of individuals $i = 1, \ldots, N$ with observations across time $t = 1, \ldots, T_i$.[1] The outcome variable, o_j with $j = 1, \ldots, J$, is a polytomous categorical variable with J identical levels for all individuals and observation times. The values of the outcome levels are unrestricted: $\forall j : o_j \in \mathcal{R}$. For each individual i and each observation time t, the chosen outcome, y_{it}, is measured as the dependent variable and a vector of M independent variables $\mathbf{x}_{it} = (x_{it1}, \ldots, x_{itM})$. Next to the realized choices, I define y^*_{itj} as the latent propensity for each individual i at time t to choose outcome j. With this notation, I assume the following relation between the propensities, y^*_{itj}, and the independent variables, $\mathbf{x}_{itj}$:

$$\forall j \in (1, \ldots, J) \colon y^*_{itj} = \alpha_{ij} + \mathbf{x}_{it}\boldsymbol{\beta}_j + \epsilon_{itj} \tag{1}$$

In this equation, $\boldsymbol{\beta}_j$ is the coefficient vector, which must be estimated. α_{ij} is a random variable. The error term, ϵ_{itj}, is a type I (Gumbel-type) extreme-value random variable, independent and identically distributed across all outcomes j. The link to the chosen outcome is defined by

$$\forall j \in (1, \ldots, J) \colon \Pr\left(y_{it} = o_j \middle| \boldsymbol{\alpha}_i, \boldsymbol{\beta}, \mathbf{x}_{it}\right) = \Pr\left(\max_{k \in (1,\ldots,J)} y^*_{itk} = y^*_{itj} \middle| \boldsymbol{\alpha}_i, \boldsymbol{\beta}, \mathbf{x}_{it}\right)$$

With these assumptions, I can derive the probabilities of each outcome. To guarantee identifiability, I define an arbitrarily chosen outcome $B \in (1, \ldots, J)$ as the base outcome, and I restrict the respective coefficients to 0: $\alpha_{iB} = 0$, $\boldsymbol{\beta}_B = \mathbf{0}$. From this, I obtain

$$\Pr\left(y_{it} = o_j \middle| \boldsymbol{\alpha}_i, \boldsymbol{\beta}, \mathbf{x}_{it}\right) = \begin{cases} \dfrac{\exp\left(\alpha_{ij} + \mathbf{x}_{it}\boldsymbol{\beta}_j\right)}{1 + \sum\limits_{k \neq B} \exp\left(\alpha_{ik} + \mathbf{x}_{it}\boldsymbol{\beta}_k\right)} & j \neq B \\ \dfrac{1}{1 + \sum\limits_{k \neq B} \exp\left(\alpha_{ik} + \mathbf{x}_{it}\boldsymbol{\beta}_k\right)} & j = B \end{cases} \tag{2}$$

1. The subscript i at T_i means that the model allows for analyzing unbalanced panel data. However, attrition must be at least at random; that is, attrition is completely at random when conditioning for the independent variables (Wooldridge 2010, 828).

So far, I have set up the assumptions for the pooled multinomial logistic regression, which can be consistently estimated if there is no unobserved heterogeneity: $\forall j : \alpha_{ij} = \alpha_j$.

The advantage of the multinomial logit model with fixed effects is that it allows for individual unobserved heterogeneity with respect to the intercepts. The heterogeneity terms, α_{ij}, are random variables with no restrictions on the joint distribution with the independent variables, $\mathbf{x}_{it}$. Directly estimating the individual α_{ij} creates an incidental parameter problem, which leads to inconsistent estimators with asymptotics solely based on $N \to \infty$. However, with additional assumptions, it is possible to consistently estimate the coefficient vector $\boldsymbol{\beta}$. First, we assume that the observed covariates are strictly exogenous conditional on the unobserved heterogeneity.

$$\forall t \in (1, \ldots, T_i), \quad j \in (1, \ldots, J) \colon f_{y_{it}|\alpha_{ij},\mathbf{x}_i} \equiv f_{y_{it}|\alpha_{ij},\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT_i}} = f_{y_{it}|\alpha_{ij},\mathbf{x}_{it}}$$

Second, we assume that the error terms are independent across time. That is, autocorrelation is ruled out.

$$\forall s, \quad t \in (1, \ldots, T_i), \quad j \in (1, \ldots, J) \colon \epsilon_{isj} \perp \epsilon_{itj} \tag{3}$$

Chamberlain (1980) states that under these additional assumptions, the term $\theta_{ij} \equiv \sum_{t=1}^{T_i} \delta_{y_{it}o_j}$, where δ denotes the Kronecker delta function with respect to y_{it} and o_j, is a sufficient statistic for the unobserved heterogeneity, α_{ij}. This relation means that the sum of occurrences of an outcome j for an individual i across time is a sufficient statistic for inclination toward that outcome.

Because there is a sufficient statistic for the unobserved heterogeneity, one can reformulate the likelihood function so that the estimands, α_{ij}, disappear. The probability mass function for the sequence of chosen outcomes across time for individual i conditional on the sufficient statistic is

$$f_{\mathbf{y}_i|\boldsymbol{\alpha}_i,\boldsymbol{\beta},\mathbf{x}_i,\boldsymbol{\theta}_i} = \frac{\prod_{t=1}^{T_i} \prod_{j=1}^{J} \Pr\left(y_{it} = o_j \middle| \boldsymbol{\alpha}_i, \boldsymbol{\beta}, \mathbf{x}_i, \boldsymbol{\theta}_i\right)^{\delta_{y_{it}o_j}}}{\sum_{\boldsymbol{\upsilon}_i \in \Upsilon_i} \left[\prod_{t=1}^{T_i} \prod_{j=1}^{J} \Pr\left(\upsilon_{it} = o_j \middle| \boldsymbol{\alpha}_i, \boldsymbol{\beta}, \mathbf{x}_i, \boldsymbol{\theta}_i\right)^{\delta_{\upsilon_{it}o_j}} \right]} \tag{4}$$

The summation in the denominator is taken over all "potential" sequences of chosen outcomes $\boldsymbol{\upsilon}_i \equiv (\upsilon_{i1}, \ldots, \upsilon_{iT_i})$ that fulfill the condition of the sufficient statistic $\boldsymbol{\theta}_i$. The set Υ_i contains all sequences $\boldsymbol{\upsilon}_i$ for which the sum of occurrences of each outcome j is the same as it is for the realized sequence y_i. Formally, this means

$$\Upsilon_i \equiv \left\{ (\upsilon_{i1}, \ldots, \upsilon_{iT_i}) \middle| \forall j \in (1, \ldots, J) \colon \sum_{t=1}^{T_i} \delta_{\upsilon_{it}o_j} = \sum_{t=1}^{T_i} \delta_{y_{it}o_j} = \theta_{ij} \right\} \tag{5}$$

Technically, Υ_i is the set of all permutations of the realized sequence of chosen outcomes $\mathbf{y}_i$. Taking into account the assumptions and definitions above, we can write (4) as

$$f_{\mathbf{y}_i|\boldsymbol{\alpha}_i,\boldsymbol{\beta},\mathbf{x}_i,\boldsymbol{\theta}_i} = \frac{\exp\left(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it}o_j}\mathbf{x}_{it}\beta_j\right)}{\sum_{\boldsymbol{\upsilon}_i\in\Upsilon_i}\exp\left(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{\upsilon_{it}o_j}\mathbf{x}_{it}\beta_j\right)} \tag{6}$$

Having derived the probability mass function, we see that the simplified expression of the log-likelihood function of the multinomial logit model with fixed effects follows its definition. The contribution to the log likelihood of individual i is

$$\begin{aligned}\ln\ell_i\left(\boldsymbol{\beta}|\,\mathbf{y}_i,\mathbf{x}_i\right) &= \ln f_{\mathbf{y}_i|\boldsymbol{\alpha}_i,\boldsymbol{\beta},\mathbf{x}_i,\boldsymbol{\theta}_i}\\ &= \sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it}o_j}\mathbf{x}_{it}\beta_j - \ln\sum_{\boldsymbol{\upsilon}_i\in\Upsilon_i}\exp\left(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{\upsilon_{it}o_j}\mathbf{x}_{it}\beta_j\right)\end{aligned}$$

Therefore, the overall log-likelihood function for the sample—given a simple random sample of panel groups—is

$$\ln L\left(\boldsymbol{\beta}|\,\mathbf{y},\mathbf{x}\right) = \sum_{i=1}^{N}\ln\ell_i\left(\boldsymbol{\beta}|\,\mathbf{y}_i,\mathbf{x}_i\right) \tag{7}$$

The maximum likelihood (ML) estimator of (7) is a consistent asymptotically normal estimator of the coefficient vector $\boldsymbol{\beta}$ (Wooldridge 2010, 473–481).

2.1 Special case: Binary logit with fixed effects

The binary logit with fixed effects is a special case of the multinomial logit model with fixed effects with $J = 2$. Usually, the outcome variable o_j is coded as $o_1 = 0$ and $o_2 = 1$. Also the base outcome is commonly defined as $B = 1$. This simplifies (2) to

$$\begin{aligned}\Pr\left(y_{it} = 1|\,\alpha_i,\boldsymbol{\beta},\mathbf{x}_{it}\right) &= \frac{\exp\left(\alpha_i + \mathbf{x}_{it}\boldsymbol{\beta}\right)}{1+\exp\left(\alpha_i + \mathbf{x}_{it}\boldsymbol{\beta}\right)}\\ \Pr\left(y_{it} = 0|\,\alpha_i,\boldsymbol{\beta},\mathbf{x}_{it}\right) &= \frac{1}{1+\exp\left(\alpha_i + \mathbf{x}_{it}\boldsymbol{\beta}\right)}\end{aligned}$$

Note that the heterogeneity term, α_i, is now a scalar because it reflects only the contrast between outcome $o_2 = 1$ and $o_1 = 0$. Similarly, the remaining coefficient vector $\boldsymbol{\beta}$ also reflects only this contrast. Furthermore, (6), which is the basis of the log-likelihood function, is simplified to

$$f_{\mathbf{y}_i|\alpha_i,\boldsymbol{\beta},\mathbf{x}_i,\boldsymbol{\theta}_i} = \frac{\exp\left(\sum_{t=1}^{T_i}y_{it}\mathbf{x}_{it}\boldsymbol{\beta}\right)}{\sum_{\boldsymbol{\upsilon}_i\in\Upsilon_i}\exp\left(\sum_{t=1}^{T_i}\upsilon_{it}\mathbf{x}_{it}\boldsymbol{\beta}\right)}$$

Note that the simplification of $\delta_{y_{it}o_j}$ to y_{it} rests on the specific dummy coding of y_{it}. For more details on this model and its implementation in Stata, see [R] **clogit**.

2.2 Interpretation

Usually, the estimates of binary and multinomial response models are interpreted as odds-ratio or logit effects or as effects on the predicted probabilities and related constructs (for example, average marginal effects).

Regarding the first class, odds-ratio and logit effects are criticized as unintuitive. Moreover, with this interpretation approach based on arbitrary restriction assumption of the variance of the error term ϵ in (1), effects across nested models or across different groups cannot be easily compared (Allison 1999; Kohler, Karlson, and Holm 2011; Best and Wolf 2015; Breen, Karlson, and Holm 2013).

Therefore, for nonlinear cross-sectional models, the second class of interpretation constructs is recommended (Long and Freese 2006, 157ff.). This option is not given for the fixed-effects model. The probability expression in (2) cannot be evaluated, because the unobserved heterogeneity vector $\boldsymbol{\alpha}$ is not estimated. Even if plausible values for $\boldsymbol{\alpha}$ are inserted in the equation, to conduct significance tests, one has to find plausible values for their variances and covariances with the other independent variables. Cameron and Trivedi (2005, 797) suggest using the binary logistic regression with fixed effects to interpret predicted probabilities of the estimation (6), which can be generalized to the multinomial case. However, although this circumvents the problem of finding a plausible conditional distribution for the unobserved heterogeneity $f_{\boldsymbol{\alpha}|\mathbf{x}}$, the object of interpretation here is more unintuitive than with the odds-ratio and logit effects. With this approach, one interprets the effects of a unit or marginal change in the independent variables at a specific time x_t on the probability that a specific time series of outcomes $y_1, \ldots, y_T$ is realized conditional on the probability of all permutations of the time series. For realistic applications, any choice of the time series of outcomes is arbitrary. Furthermore, the interpretation of the conditional probability remains imprecise, because the permutation can be understood only as an analogue for the tendency to choose each outcome. The odds-ratio effects interpretation as shown above is the only viable option for the binary and multinomial fixed-effects logistic regression.

2.3 Robust standard errors

For other models, specifically those with panel data, it is common to report Huber–White or sandwich-estimator standard errors. These standard errors are robust to specific violations of model assumptions. For linear panel-data models, sandwich-estimator standard errors are robust to heteroskedasticity and serial correlation (see Cameron and Trivedi [2005, 705ff.]). For multilevel models with continuous dependent variables, sandwich-estimator standard errors can be robust to heteroskedasticity and correlation within higher-level units across lower-level units.

However, for a nonlinear model with fixed effects as described here, the robustness of the sandwich estimator is restricted to violation of homoskedasticity on the level of the panel groups (Wooldridge 2010, 608–625). The assumption of error independence across time, (3), must be maintained; that is, the sandwich estimator is not robust to violation of this assumption. However, the sandwich estimator is robust to violation of independence across panel groups i, (1). This is equivalent to heteroskedasticity robustness. Note that this implies robustness to varying error variances within and between clusters of panel groups.

If the assumption of independence across time is violated, the ML estimator of (7) is inconsistent and can be interpreted only as a quasi-ML estimator, where the sandwich-estimator standard errors can be used to "test hypotheses about the best approximation to the true density" (Wooldridge 2010, 503). Note that `xtlogit, fe` also precludes robust standard errors.

3 Implementation

To implement `femlogit`, I use the `moptimize()` Mata suite because it allows me to implement the evaluator as a `gf2` type. This increases precision and computational speed (Gould, Pitblado, and Poi 2010, 20–24). Moreover, the `gf2`-type evaluator enables a straightforward consideration of the panel-data structure and an easier integration into the `svy` command suite.[2] The evaluator is implemented as a Mata function. Besides being the natural choice with `moptimize()`, this allows a straightforward integration of the Mata function `cvpermute()`, which is used to loop over the set Υ_i in (5). The `gf2`-type evaluator expects arguments to be the dependent variable column vector $\mathbf{y}\colon \sum_{i=1}^{N} T_i \times 1$, the independent variable matrix $\mathbf{x}\colon \sum_{i=1}^{N} T_i \times M$, and an initial coefficient row vector $\boldsymbol{\beta} : 1 \times (J-1)\,M$. The evaluator returns the column vector $\{\ln \ell_i\,(\boldsymbol{\beta})\} : N \times 1$ of the individual contributions for all panel groups, the gradient matrix $\mathbf{g}\colon N \times (J-1)\,M$, and the Hessian matrix $\mathbf{H}\colon (J-1)\,M \times (J-1)\,M$. For the latter two function outcomes, the first and second partial derivatives, with respect to all coefficients β_{jm}, are derived analytically and inserted in the Mata evaluator.

The `moptimize()` call is embedded in an ado-wrapper, following the structure of the implementations of `mlogit` and `clogit`. Worth mentioning here are the definition of the estimation sample and the initial values of the coefficient vector β. For the estimation sample, first, observations with missing values on the dependent, independent, or panel-group indicator variables are deleted. Second, collinear independent variables are excluded. Finally, panel groups without variance across time in the dependent variables, as well as independent variables without variance across time in all panel groups, are dropped. The initial values for the coefficient vector are the estimated coefficients of a pooled multinomial logit model. This follows the implementation of `clogit`, where the initial values are taken from the pooled binary logit model. The implemented command identifies panel groups by using the panel-group indicators set by `xtset`.

2. Note that there is no support for weights and the `svy` command suite in the current version of the implementation.

3.1 Data structure

The implementation expects the data to be organized in long format—that is, from the panel-data perspective, each observation represents a time point of one person. The following is a modified version of the example data used in [R] **clogit**:[3]

```
. use femlogitid
. list in 1/11
```

	id	y	x1	x2
1.	1014	3	0	4
2.	1014	0	1	4
3.	1014	2	1	6
4.	1014	1	1	8
5.	1017	0	0	1
6.	1017	2	0	7
7.	1017	1	1	10
8.	1019	0	0	1
9.	1019	2	1	7
10.	1019	1	1	7
11.	1019	1	1	9

The first four observations belong to the person with the `id` = 1014. The independent variables are `x1` and `x2`, and `y` is the categorical dependent variable with four levels $(0, 1, 2, 3)$. Note that the different levels of the categorical dependent variable are stored in one variable and one case, similarly to `mlogit`. In contrast, the implementation of `clogit` expects the outcomes of the dependent variable for each time point to be stored in long format.

3.2 Computational problems

The current implementation enumerates the sum over all permutations of the individual sequences y_i in the denominator of (6). This means that computation time increases with the number of permutations in the dependent variable. In practice, this will rise with T_i. The computation time can be very high, even if T_i is large for only a small subset of individuals $i = 1, \ldots, N$. If computation becomes unwieldy, a random subset of available measurements of all observation units should be analyzed. This selection should not depend on the number of available measurements for each observation. Increasing N should not increase the computation time severely.

3. The data `femlogitid.dta` and syntax `femlogit_example1.do` can be found in the online appendix.

4 The femlogit command

4.1 Syntax

The command `femlogit` is called with the following syntax:

`femlogit` *depvar* [*indepvars*] [*if*] [*in*] [, `group(`*varlist*`)` `baseoutcome(`#`)` `constraints(`*clist*`)` `difficult` `or` `robust`]

depvar and *indepvars* may not contain factor variables or time-series operators. No prefix commands are allowed. Weights and `vce()` are not allowed at this point.

4.2 Options

`group(`*varlist*`)` specifies one or more identifier variables (numeric or string) for the matched groups. It overrides the default group indicator that is specified with `xtset`.

`baseoutcome(`#`)` specifies the value of *depvar* to use as the base outcome. The default is to choose the mode outcome.

`constraints(`*clist*`)` specifies the linear constraints to be applied during estimation. The default is to perform unconstrained estimation. *clist* has the form # [-#] [, # [-#] ...].

`difficult` specifies that the "hybrid" method be used in nonconcave regions of the likelihood function instead of the default "modified Marquardt" method (Gould, Pitblado, and Poi 2010, 15–17).

`or` reports the estimated coefficients transformed to odds ratios, that is, $\exp(b)$ rather than b. Confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated.

`robust` uses the robust or sandwich estimator of variance. This is valid only for quasi-ML interpretation (Wooldridge 2010, 502ff.). It can be interpreted only as heteroskedasticity robustness, not as panel robustness.

4.3 Stored results

femlogit stores the following in e():

Scalars

e(N)	number of observations	e(r2_p)	pseudo-R-squared
e(N_drop)	number of observations dropped because of invariant dependent variable	e(ll)	log likelihood
		e(ll_0)	log likelihood, constant-only model
		e(chi2)	χ^2
e(N_group_drop)	number of groups dropped because of invariant dependent variable	e(p)	significance
		e(rank)	rank of e(V)
		e(ic)	number of iterations
e(k)	number of parameters	e(rc)	return code
e(k_eq)	number of equations in e(b)	e(converged)	1 if converged, 0 otherwise
e(k_eq_model)	number of equations in overall model test	e(baseout)	value of *depvar* to be treated as the base outcome
e(k_dv)	number of dependent variables	e(ibaseout)	index of the base outcome
e(df_m)	model degrees of freedom	e(k_out)	number of outcomes

Macros

e(cmd)	femlogit	e(user)	femlogit_eval_gf2()
e(cmdline)	command as typed	e(technique)	nr
e(depvar)	name of dependent variable	e(crittype)	log likelihood or log pseudolikelihood
e(title)	title in estimation output		
e(chi2type)	Wald or LR; type of model χ^2 test	e(properties)	b V
		e(predict)	_predict
e(vce)	oim or robust	e(marginsok)	xb
e(vcetype)	Robust	e(marginsnotok)	stdp stddp
e(opt)	moptimize	e(eqnames)	names of equations
e(which)	max	e(group)	name of group() variable
e(ml_method)	gf2		

Matrices

e(b)	coefficient vector	e(V)	variance–covariance matrix of the estimator
e(Cns)	constraints matrix		
e(ilog)	iteration log (up to 20 iterations)	e(V_modelbased)	model-based variance
		e(out)	outcome values
e(gradient)	gradient vector		

Functions

e(sample)	marks estimation sample

5 Application: Effect of ideological distance on voting behavior with British election panel data

In this section, I demonstrate the femlogit command and explain how to interpret the results. My example follows the one that Skrondal and Rabe-Hesketh (2003) use to illustrate the application of multilevel random-effects models for polytomous and ordinal dependent variables. They analyze data from the 1987–1992 panel of British Election Study (Heath et al. 1992) to fit a model of the recalled vote choice for the Conservative, Labour, or Liberal party and a model of the rank order of the parties. Here I concentrate on the recalled vote choice and use the femlogit command to estimate the effect of the distance on the left–right policy dimension between the voter and the party on the vote choice. I control for the time-varying rating of perceived inflation and implicitly for all time-variant factors at the voter level. The analysis syntax for this example is found in `femlogit_example2.do`, which is provided in the online appendix.

The raw data are taken from Rabe-Hesketh and Skrondal (2012, 680ff.). Cleaning and preparation leads to the following analysis data:

```
. describe

Contains data
  obs:         2,458
 vars:             9
 size:        46,702
--------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
serialno        int     %8.0g                 Respondent number
rldist2         float   %9.0g                 Dist(Labour)-Dist(Conservative)
rldist3         float   %9.0g                 Dist(Liberal)-Dist(Conservative)
male            byte    %8.0g                 Male
manual          byte    %8.0g                 Manual worker
inflation       byte    %8.0g                 Perceived inflation
age             float   %9.0g                 Age in 10 yr units
yr92            byte    %8.0g                 1992 election indicator
choice          byte    %12.0g     choice     Recalled vote for party
--------------------------------------------------------------------------------
Sorted by:  serialno
     Note:  dataset has changed since last saved
```

The dependent variable `choice` is a discrete variable with three alternatives—"Conservative", "Labour", and "Liberal". In the multinomial logit model with fixed effects, the following four independent variables are used: the difference of the distance between the voter and the Labour party and the distance between the voter and the Conservative party (`rldist2`); the difference of the distance between the voter and the Liberal party and the distance between the voter and the Conservative party (`rldist3`); a rating of the perceived inflation (`inflation`); and a wave dummy (`yr92`).

The data are in long format. As the summary command for panel data, `xtdescribe`, shows, the dataset contains information on 1,344 persons across both elections. For 1,114 persons, the time series across both waves is complete. For the remaining 230 persons, information is missing for at least one wave.

```
. xtset serialno yr92
       panel variable:  serialno (unbalanced)
        time variable:  yr92, 0 to 1
                delta:  1 unit

. xtdescribe

serialno:  2, 11, ..., 5997                                  n =       1344
    yr92:  0, 1, ..., 1                                      T =          2
           Delta(yr92) = 1 unit
           Span(yr92)  = 2 periods
           (serialno*yr92 uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       2         2         2       2       2

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
     1114     82.89   82.89 |  11
      121      9.00   91.89 |  1.
      109      8.11  100.00 |  .1
 ---------------------------+---------
     1344    100.00         |  XX
```

The differences in the policy distances vary not only across voters and waves but also across alternatives. This allows us to specify the model as a mixed-logit model (Cameron and Trivedi 2005, 495).[4] That is, I estimate one coefficient for the alternative-varying policy distances and alternative-specific coefficients for the alternative-invariant voters' rating of inflation and the wave dummy. To do this, I define the following constraints for the effects of the policy distances:

```
. constraint 1 [Labour]rldist3=0
. constraint 2 [Liberal]rldist2=0
. constraint 3 [Labour]rldist2=[Liberal]rldist3
```

With these constraints, the effect of the relative policy distance between the voter and the Liberal party plays no role in the propensity to vote for labor in comparison with the Conservative party and vice versa. The relative policy distance between the voter and the Labour party is irrelevant in the propensity to vote for the Liberal party instead of the Conservative party. The third constraint guarantees that the relative policy distances have the same effect on both propensities.

4. This specification should not be confused with logistic regression with random slopes or random covariate effects, which is implemented as `mixlogit` by Hole (2007).

The estimation output of `femlogit` for this model is as follows:

```
. femlogit choice rldist2 rldist3 inflation yr92, group(serialno) const(1/3)
> b(1)
note: 1097 groups (1964 obs) dropped because of all positive or
      all negative outcomes.
Iteration 0:   log likelihood = -156.16844
Iteration 1:   log likelihood = -139.49392
Iteration 2:   log likelihood = -138.19403
Iteration 3:   log likelihood = -138.19006
Iteration 4:   log likelihood = -138.19006
Fixed-effects multinomial logistic regression     Number of obs   =        494
                                                  Wald chi2(5)    =      45.69
Log likelihood = -138.19006                       Prob > chi2     =     0.0000
 ( 1)  [Labour]rldist3 = 0
 ( 2)  [Liberal]rldist2 = 0
 ( 3)  [Labour]rldist2 - [Liberal]rldist3 = 0
```

choice	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conservative	(base outcome)					
Labour						
rldist2	-.0590691	.0145332	-4.06	0.000	-.0875536	-.0305846
rldist3	(omitted)					
inflation	.8354586	.3692285	2.26	0.024	.111784	1.559133
yr92	.6791261	.2734095	2.48	0.013	.1432534	1.214999
Liberal						
rldist2	(omitted)					
rldist3	-.0590691	.0145332	-4.06	0.000	-.0875536	-.0305846
inflation	.5786913	.305657	1.89	0.058	-.0203854	1.177768
yr92	-.2315669	.2188483	-1.06	0.290	-.6605018	.1973679

The output header shows that 1,097 voters and 1,964 observations are dropped, because there is no variance in the dependent variable across waves for these voters. That is, the model is fit with 247 voters and 494 observations. The iteration log shows that the ML algorithm converged after four steps. The log likelihood for the first step is derived from the initial coefficient vector, which is the result of pooled multinomial logit with the same variable structure. The header also shows the Wald test statistic of 45.69. The five degrees of freedom reflect the reduced number of a free number of parameters. Note that the command returns a Wald test instead of a likelihood-ratio test because constraints were specified.

The coefficient table shows the logarithm of the relative-risk ratios for a one-unit change in the corresponding variables. That is, with an increase in the relative distance between a voter and the Labour party by one unit ceteris paribus, the logarithm of the probability to vote for labor divided by the probability to vote for the Conservative party decreases by 0.059. Equivalently, ceteris paribus, this relative distance increases by one unit, and the odds to vote for labor versus voting conservative increase by a factor of $\exp(0.059) = 0.943$; that is, they decrease by 6.7%. Similarly, with each unit increase in the inflation rating ceteris paribus, the odds to vote for labor versus voting conservative increase by 130.6%, and the odds to vote liberal versus voting conservative increase by 78.4%. One can interpret the odds effects for other contrasts by looking at the respective coefficient or variable differences. For example, if the inflation rating increases by one unit ceteris paribus, the odds to vote labor versus voting liberal increase by a factor of $\exp(0.835 - 0.579) = 1.293$ or 29.3%.

As mentioned previously, the multinomial logit model with fixed effects allows for possibly confounding unobserved heterogeneity at the level of the voter with respect to the preferences for a specific party. Alternative models have to rule this out or have to measure the heterogeneity. In table 1, I show the respective effects for the pooled multinomial logistic regression and the multinomial logistic regression with random effects. For the first model, panel–robust standard errors are used to account for possible correlation across waves. The latter model is fit with `gsem`, as described in [SEM] **example 41g**. In the alternative models, heterogeneity is captured in the time-invariant variables `male`, `age`, and `manual`.

Table 1. Pooled, random-effects, and fixed-effects models for voting example

	POMLOGIT exp(β)/se	REMLOGIT exp(β)/se	FEMLOGIT exp(β)/se
Labour			
Relat. policy dist.	0.896***	0.818***	0.943***
	(0.005)	(0.011)	(0.014)
Inflation	2.134***	3.812***	2.306*
	(0.236)	(0.815)	(0.851)
1992 election	1.153	1.564*	1.972*
	(0.112)	(0.346)	(0.539)
Male	0.452***	0.261***	
	(0.068)	(0.082)	
Age	0.702***	0.499***	
	(0.037)	(0.056)	
Manual worker	1.952***	5.188***	
	(0.302)	(1.767)	
Constant	0.059***	0.007***	
	(0.029)	(0.007)	
Liberal			
Relat. policy dist.	0.896***	0.818***	0.943***
	(0.005)	(0.011)	(0.014)
Inflation	1.735***	2.938***	1.784
	(0.185)	(0.584)	(0.545)
1992 election	0.808*	0.771	0.793
	(0.080)	(0.159)	(0.174)
Male	0.493***	0.304***	
	(0.073)	(0.092)	
Age	0.810***	0.632***	
	(0.039)	(0.066)	
Manual worker	0.900	1.235	
	(0.132)	(0.393)	
Constant	0.102***	0.013***	
	(0.048)	(0.012)	
Var($\alpha_{\text{Lab.}}$)		14.672***	
		(2.988)	
Var($\alpha_{\text{Lib.}}$)		13.915***	
		(2.325)	
Cov($\alpha_{\text{Lab.}}, \alpha_{\text{Lib.}}$)		11.441***	
		(2.377)	
log likelihood	−1946.269	−1764.331	−138.190
N obs.	2458	2458	494
N groups	1344	1344	247

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$; base outcome: Conservative party;
reference categories: 1987 election, female, not manual worker;
FEMLOGIT: multinomial logit model with fixed effects;
POMLOGIT: pooled multinomial logistic regression;
REMLOGIT: multinomial logistic regression with random effects.

6 Conclusion

In this article, I introduce an implementation of multinomial logistic regression with fixed effects as derived by Chamberlain (1980). With this model, it is possible to consistently estimate effects of time-varying regressors on the log-odds of multinomial outcomes when time-invariant unobserved heterogeneity is present. In particular, time-invariant unobserved heterogeneity may be correlated with predictor variables. The implemented ado `femlogit` is applied to real data. In an example with British election panel data, I estimate the effect of perceived distance in the left–right political dimension between a candidate and a voter on voting behavior. The specific advantage of the multinomial logit model with fixed effects in this example is that the effect of policy distance on vote intention is estimated net of all time-invariant voter characteristics that may affect vote intention, perceived policy distance, or both.

7 Acknowledgments

I thank Josef Brüderl, Jeffrey Pitblado, Henning Best, Stephanie Eckman, Michael Herrmann, Andreas Horr, Reinhard Schunck, the participants of the 2011 German Stata Users Group meeting in Bamberg, and an anonymous reviewer for useful comments and suggestions. I thank Sophia Rabe-Hesketh and Anders Skrondal for providing the data used in the examples.

8 References

Allison, P. D. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods and Research* 28: 186–208.

———. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.

Best, H., and C. Wolf. 2015. Logistic regression. In *Regression Analysis and Causal Inference*, ed. H. Best and C. Wolf, 57–82. London: Sage.

Breen, R., K. B. Karlson, and A. Holm. 2013. Total, direct, and indirect effects in logit and probit models. *Sociological Methods and Research* 42: 164–191.

Brüderl, J., and V. Ludwig. 2015. Fixed-effects panel regression. In *Regression Analysis and Causal Inference*, ed. H. Best and C. Wolf, 327–358. London: Sage.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* 47: 225–238.

Gangl, M. 2010. Causal inference in sociological research. *Annual Review of Sociology* 36: 21–47.

Gould, W., J. Pitblado, and B. Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, TX: Stata Press.

Heath, A., R. Jowell, J. K. Curtice, J. A. Brand, and J. C. Mitchell. 1992. British General Election Panel Survey, 1987–1992. Ann Arbor, MI. Inter-university Consortium for Political and Social Research.

Hole, A. R. 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* 7: 388–401.

Kohler, U., K. B. Karlson, and A. Holm. 2011. Comparing coefficients of nested nonlinear probability models. *Stata Journal* 11: 420–438.

Lee, M.-J. 2002. *Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables*. San Diego: Academic Press.

Long, J. S., and J. Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.

Pforr, K. 2013. *femlogit: Implementation und Anwendung der multinominalen logistischen Regression mit „fixed effects“*, vol. 11 of *GESIS-Schriftenreihe*. Köln: GESIS—Leibniz-Institut für Sozialwissenschaften.

Rabe-Hesketh, S., and A. Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata. Volume II: Categorical Responses, Counts, and Survival*. 3rd ed. College Station, TX: Stata Press.

Skrondal, A., and S. Rabe-Hesketh. 2003. Multilevel logistic regression for polytomous data and rankings. *Psychometrika* 68: 267–287.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

About the author

Klaus Pforr is a project consultant and researcher at GESIS—Leibniz–Institute for the Social Sciences in Mannheim, Germany.