fhetprob: A fast QMLE Stata routine for fractional probit models with multiplicative heteroskedasticity

Richard Bluhm*

May 26, 2013

Introduction

Stata can easily estimate a binary response probit models with modeled heteroskedasticity (`hetprob`) or without heteroskedasticity (`probit` or `glm`). Nevertheless, it only allows for estimation of fractional response models *without* heteroskedasticity via the GLM suite. The reason behind this restriction is purely computational. The official implementations of probit models take advantage of several mathematical simplifications that are only available when the dependent variable is either strictly zero or unity.

Cutting out "unnecessary" computations positively affects runtime, especially on larger datasets, but sacrifices generality. Thanks to Stata's comprehensive and easy to use maximum likelihood suite, writing a simple linear form MLE for fractional response models with heteroskedasticity is near trivial (Gould, Pitblado, and Poi 2010). However, estimation relying only on numerical derivatives may be computationally expensive. The log-likelihood for (fractional) probit models with heteroskedasticity is difficult to maximize and may take a substantial amount of time if both the gradients and Hessian are computed numerically. Yet, speed may matter a lot for larger data sets.

The program described in this note (`fhetprob`) extends Stata's own `hetprob` command to allow for fractional response variables and computes all likelihood derivatives analytically in order to realize significant speed gains over a simple linear form ML estimator. While `fhetprob` can estimate binary response models as a special case, it is by no means a replacement for Stata's own method as it will run slower than `hetprob` even with moderately large N.[1]

Fractional response models have several important applications and are gaining popularity in econometrics. They can be applied to estimate models of proportions in cross-sectional data (Papke and Wooldridge 1996; Wooldridge 2010a) and balanced panels which may be subject to unobserved heterogeneity and endogeneity (Papke and Wooldridge 2008). However, many panel data sets used in applied research are unbalanced, sometimes heavily so. To estimate fractional response models with unbalanced

*Maastricht University, Graduate School of Governance/ UNU-MERIT. Maastricht, The Netherlands. *Email:* richard.bluhm -at- maastrichtuniversity.nl or bluhm -at- merit.unu.edu.

[1] Simulations with binary responses show that with $N = 1000$ they are just about equally fast, but with $N = 10^4$ `fhetprob` is $\approx 25\%$ slower. However, the loss is not increasing in N beyond that point, it's still about 25% with $N = 10^5$ and $N = 10^6$).

panels the conditional variance should be allowed to vary with the nature of the unbalancedness and thus requires models that explicitly allow for certain forms of heteroskedasticity (Wooldridge 2010b). This note outlines the methods behind `fhetprob` and provides examples for cross-section and unbalanced panel data.

Method

A classic probit DGP in index model notation supposes that $y = \mathbf{1}[\mathbf{x}_i'\boldsymbol{\beta} + \epsilon]$ with a constant error variance ($\mathrm{Var}[\epsilon] = \sigma_\epsilon^2$). If we relax the constant variance assumption and instead assume the error variance depends on $\mathbf{z}_i$ as follows $\mathrm{Var}[\epsilon|\mathbf{x}_i, \mathbf{z}_i] = (\exp(\mathbf{z}_i'\boldsymbol{\gamma}))^2$, we obtain a probit model with multiplicative heteroskedasticity (Harvey 1976).

Fractional probit models are defined analogously but instead of the index model, assume that the conditional expectation of the outcome is defined by a probit 'link' function such that $\mathrm{E}[y|\mathbf{x}_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta})$. Then, similarly to the binary case, a fractional response model with multiplicative heteroskedasticity can be written as $\mathrm{E}[y|\mathbf{x}_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta} \times \exp(-\mathbf{z}_i'\boldsymbol{\gamma}/2))$, see for example Wooldridge (2010b). In both cases, the typical Bernoulli likelihood is

$$\mathcal{L} = \prod_{i=1}^{N} \left(G(\cdot)^{y_i} + (1 - G(\cdot))^{1-y_i} \right) \tag{1}$$

where $G(\cdot) = \Phi(\cdot)$ is the probit link function (or standard normal c.d.f.).

Since the Bernoulli distribution is in the linear exponential family (LEF), the corresponding quasi-maximum likelihood estimator (QMLE) solves

$$\arg\max_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{N} \left[y_i \ln \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})}\right) + (1-y_i) \ln\left(1 - \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\exp(\mathbf{z}_i'\boldsymbol{\gamma})}\right)\right)\right] \tag{2}$$

To accommodate the fractional case, no simplifications are used that rely on assuming that y_i can only take on unity or zero (e.g. see, Greene 2011, 690–692), hence the derivation involves a little more algebra than otherwise.

The likelihood equations are

$$\frac{\partial \ln \mathcal{L}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left(y_i \frac{\phi(\omega_i)}{\Phi(\omega_i)} + (1-y_i)\frac{-\phi(\omega_i)}{\Phi(-\omega_i)} \right) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})\mathbf{x}_i = \mathbf{0} \tag{3}$$

$$\frac{\partial \ln \mathcal{L}}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{N} \left(y_i \frac{\phi(\omega_i)}{\Phi(\omega_i)} + (1-y_i)\frac{-\phi(\omega_i)}{\Phi(-\omega_i)} \right) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})(-\mathbf{x}_i'\boldsymbol{\beta})\mathbf{z}_i = \mathbf{0} \tag{4}$$

where $\omega_i = \mathbf{x}_i'\boldsymbol{\beta} \times \exp(-\mathbf{z}_i'\boldsymbol{\gamma})$.

The Hessian terms are

$$\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \sum_{i=1}^{N} \Bigg(y_i \left[\frac{-\omega_i \phi(\omega_i)}{\Phi(\omega_i)} - \frac{\phi^2(\omega_i)}{\Phi^2(\omega_i)} \right] \\
&\quad + (1-y_i) \left[\frac{\omega_i \phi(\omega_i)}{\Phi(-\omega_i)} - \frac{\phi^2(\omega_i)}{\Phi^2(-\omega_i)} \right] \Bigg) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})^2 \mathbf{x}_i \mathbf{x}_i' \\
\mathbf{H}_{11} &= \sum_{i=1}^{N} \left(y_i[-\omega_i s_i - s_i^2] + (1-y_i)[\omega_i q_i - q_i^2] \right) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})^2 \mathbf{x}_i \mathbf{x}_i' \qquad (5)
\end{aligned}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} = \sum_{i=1}^{N} \left(y_i \left[\frac{\phi(\omega_i)(\omega_i^2 - 1)}{\Phi(\omega_i)} + \frac{\omega_i \phi^2(\omega_i)}{\Phi^2(\omega_i)} \right] + (1 - y_i) \left[\frac{\phi(\omega_i)(1 - \omega_i^2)}{\Phi(-\omega_i)} + \frac{\omega_i \phi^2(\omega_i)}{\Phi^2(-\omega_i)} \right] \right) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})\mathbf{x}_i\mathbf{z}_i'$$

$$\mathbf{H}_{12} = \sum_{i=1}^{N} \left(y_i[s_i(\omega_i^2 - 1) + \omega_i s_i^2] + (1 - y_i)[q_i(1 - \omega_i^2) + \omega_i q_i^2] \right) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})\mathbf{x}_i\mathbf{z}_i' \tag{6}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \sum_{i=1}^{N} \left(y_i \left[\frac{\phi(\omega_i)(\omega_i^2 - 1)}{\Phi(\omega_i)} + \frac{\omega_i \phi^2(\omega_i)}{\Phi^2(\omega_i)} \right] + (1 - y_i) \left[\frac{\phi(\omega_i)(1 - \omega_i^2)}{\Phi(-\omega_i)} + \frac{\omega_i \phi^2(\omega_i)}{\Phi^2(-\omega_i)} \right] \right) \exp(-\mathbf{z}_i'\boldsymbol{\gamma})(-\mathbf{x}_i'\boldsymbol{\beta})\mathbf{z}_i\mathbf{z}_i'$$

$$\mathbf{H}_{22} = \sum_{i=1}^{N} \left(y_i[s_i(\omega_i^2 - 1) + \omega_i s_i^2] + (1 - y_i)[q_i(1 - \omega_i^2) + \omega_i q_i^2] \right) \times \exp(-\mathbf{z}_i'\boldsymbol{\gamma})(-\mathbf{x}_i'\boldsymbol{\beta})\mathbf{z}_i\mathbf{z}_i' \tag{7}$$

where $s_i = \phi(\omega_i)/\Phi(\omega_i)$ and $q_i = \phi(\omega_i)/\Phi(-\omega_i)$. The Hessian is then just the collection of the Hessian terms.

Equations (2) to (7) are then used to define an `e2` (formerly `d2`) type ML evaluator in Stata which supports equation-level scores (Gould, Pitblado, and Poi 2010). Robust variance-covariance estimation is essential because this is a QLME estimator for which we are assuming a correctly-specified conditional mean but allow all other features of the distribution to be misspecified (Gourieroux, Monfort, and Trognon 1984). As Papke and Wooldridge (1996) point out, regular standard errors based on the inverse information matrix will be too large and do *not* approximate the asymptotic standard errors if the GLM variance assumption $\text{Var}[y|\mathbf{x}, \mathbf{z}] = \sigma^2 \text{E}[y|\mathbf{x}, \mathbf{z}] - (1 - \text{E}[y|\mathbf{x}, \mathbf{z}])$ is violated or holds with underdispersion ($\sigma^2 < 1$). By the same reasoning pooled QMLE for panel models also necessarily implies clustering. Stacking the estimated coefficients as $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$, the fully robust variance estimator using the empirical Hessian is

$$\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}) = \left(-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) \right)^{-1} \left(\frac{N}{N-1} \sum_{i}^{N} \mathbf{s}_i'(\hat{\boldsymbol{\theta}})\mathbf{s}_i(\hat{\boldsymbol{\theta}}) \right) \left(-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) \right)^{-1} \tag{8}$$

where $\mathbf{s}_i$ is the individual contribution to the log-likelihood (or equation-level *score*).

Partial Effects

Similar to the case of binary response probit, the estimated coefficients are scaled by a common factor that varies from specification to specification. However, depending on the variance equation, *a single coefficient may no longer reveal the sign or relative magnitude of the estimated effect.* To obtain the partial (marginal) effect of a particular continuous variable (w_k), we take derivatives with respect to the corresponding elements in the outcome and/or variance equation. The key complication is that w_k can be in $\mathbf{x}_i$, $\mathbf{z}_i$ or both, hence we define the estimated partial effect as follows.

$$\widehat{\mathrm{PE}}(w_k)_i = \phi\left(\frac{\mathbf{x}_i'\hat{\boldsymbol{\beta}}}{\exp(\mathbf{z}_i'\hat{\boldsymbol{\gamma}})}\right) \times \frac{\mathbf{1}[w_k \in \mathbf{x}_i]\hat{\beta}_k - \mathbf{1}[w_k \in \mathbf{z}_i]\hat{\gamma}_k(\mathbf{x}_i'\hat{\boldsymbol{\beta}})}{\exp(\mathbf{z}_i'\hat{\boldsymbol{\gamma}})} \tag{9}$$

where the two indicator functions toggle the cases: 1) $w_k \in \mathbf{x}_i$ and $w_k \notin \mathbf{z}_i$, 2) $w_k \notin \mathbf{x}_i$ and $w_k \in \mathbf{z}_i$, and 3) $w_k \in \mathbf{x}_i$ and $w_k \in \mathbf{z}_i$. In the case of dummy variables, discrete differences should be used instead.

The individual partial effects can the be averaged across the sample population to obtain the average partial effects (APEs), partial effects estimated at the mean of the covariates, or partial effects estimated at other interesting values (say, quantiles). The standard errors of these quantities can be bootstrapped or derived explicitly using the delta method. Stata's `margins` command conveniently implements the delta method using numerical approximations for all estimation commands that can recover the conditional expectation of the outcome variable. Thus, `margins` uses `fhetprob`'s predictions for $\mathrm{E}[y|\mathbf{x},\mathbf{z}]$ to estimate the desired partial effects of the conditional mean. The second example below illustrates this numerically.

Examples

Cross-section Data: Absent panel data, applications of fractional or binary response variables with heteroskedasticity are rare and require a strong prior knowledge or hypotheses about the underlying data generation process. The fudamental problem is that in these models it's impossible to distinguish between a mispecified mean and variance equation. The following example is taken from the Stata manual for `hetprob` and mainly serves to illustrate the unusual behavior when calling `fhetprob` with a binary dependent variable instead of a fractional response (for additional details see the **[R] hetprob** section of the Stata manuals).

```
. clear
. set obs 1000
obs was 0, now 1000
. set seed 1234567
. gen x = 1-2*runiform()
. gen xhet = runiform()
. gen sigma = exp(1.5*xhet)
. gen p = normal((0.3+2*x)/sigma)
. gen y = cond(runiform()<=p,1,0)

. fhetprob y x, het(xhet) nolog

The dependent variable is binary and not a fractional response. Consider using
the official 'hetprob' command instead. The fhetprob program does not verify
if the outcome variable is specified correctly for the binary response case.

Heteroskedastic fractional probit model          Number of obs      =       1000
                                                 Wald chi2(1)       =      65.23
Log pseudolikelihood = -569.4783                 Prob > chi2        =     0.0000
```

```
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y            |
           x |    2.22803   .2758597     8.08   0.000     1.687355    2.768705
       _cons |   .2493821   .0843367     2.96   0.003     .0840853    .4146789
-------------+----------------------------------------------------------------
lnsigma2     |
        xhet |   1.602537   .2671326     6.00   0.000     1.078967    2.126107
------------------------------------------------------------------------------
Wald test of lnsigma2=0:             chi2(1) =    35.99   Prob > chi2 = 0.0000

For fractional responses oim SEs are too big, robust option implied to correct
bias. For binary responses non-robust SEs can be obtain with option vce(oim).
```

First, notice how the program warns the user that `fhetprob` is not designed for binary outcomes. It offers no corresponding data checks, less options and runs slower on binary outcomes. Second, since the program assumes a fractional response outcome, it will automatically act as if the user intended to specify the `robust` option. In all other aspects, the results are identical to invoking `hetprob` with `robust` and, as expected, we cannot reject that the estimated coefficients are equal to their true value.

Unbalanced Panel Data with Correlated Random Effects: This example is taken from Jeffrey Wooldridge's 2011 presentation at the Chicago Stata Users group meeting. The data is from Papke's (2005) paper "The effects of spending on test pass rates: evidence from Michigan" published in the *Journal of Public Economics*. An updated version of this data is also used by Papke and Wooldridge (2008).

The dependent variable is the fraction of fourth graders passing the math test of the Michigan Education Assessment Program (MEAP) in a particular school. The coefficient of interest is on `lavgrexp` (log of average expenditure per student). Additional controls are the fraction of students eligible for the free or reduced-price lunch programs (`lunch`), the log of the number of students enrolled in each school (`lenrol`), and a set of time dummies (`y95` to `y98`). To allow for unobserved heterogeneity in the form of Correlated Random Effects (CRE), the time averages of all time-varying covariates are included and given the suffix `b` (for details see Wooldridge 2010b). Further, both the outcome and variance equation are allowed to depend on the number of observations within each subpanel (T_i), denoted `tobs3` and `tobs4`, relative to $T_i = 5$. There are no obsvervations with $T_i = 1$, but in other application these would need to be dropped.

The script below first downloads several datasets, unzips and then loads the MEAP data. Alternatively, this can be done manually beforehand.

```
. global url ///
  "http://mitpress.mit.edu/sites/default/files/titles/content/wooldridge/"
. copy $url/statafiles.zip woold2nd.zip, replace
. unzipfile woold2nd.zip

. use meap94_98, clear
. tab tobs
```

```
  number of |
       time |
    periods |      Freq.     Percent        Cum.
------------+-----------------------------------
          3 |      1,512       21.15       21.15
          4 |      1,028       14.38       35.52
          5 |      4,610       64.48      100.00
------------+-----------------------------------
      Total |      7,150      100.00

. gen tobs3 = (tobs == 3)
. gen tobs4 = (tobs == 4)
. replace math4 = math4/100
. fhetprob math4 lavgrexp lunch lenrol y95 y96 y97 y98 lavgrexpb ///
    lunchb lenrolb  y95b y96b y97b y98b tobs3 tobs4, ///
    het(tobs3 tobs4) vce(cluster schid) nolog

Heteroskedastic fractional probit model         Number of obs    =       7150
                                                Wald chi2(16)    =    3367.03
Log pseudolikelihood = -4414.841                Prob > chi2      =     0.0000

                               (Std. Err. adjusted for 1683 clusters in schid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
math4        |
    lavgrexp |   .1142198   .0735598     1.55   0.120    -.0299547    .2583943
       lunch |  -.0013961    .001221    -1.14   0.253    -.0037891    .0009969
      lenrol |   -.067624   .0561521    -1.20   0.228    -.1776802    .0424321
         y95 |   .3241894   .0150181    21.59   0.000     .2947545    .3536243
         y96 |   .3724917   .0203004    18.35   0.000     .3327036    .4122797
         y97 |   .2830853   .0217498    13.02   0.000     .2404566     .325714
         y98 |   .7162732   .0239386    29.92   0.000     .6693543     .763192
   lavgrexpb |   .1622915   .0957332     1.70   0.090    -.0253421    .3499251
      lunchb |  -.0126246   .0012652    -9.98   0.000    -.0151044   -.0101448
     lenrolb |  -.0029271   .0610953    -0.05   0.962    -.1226718    .1168175
        y95b |   .8794233   .5371531     1.64   0.102    -.1733774    1.932224
        y96b |   .7270717   .2073896     3.51   0.000     .3205955    1.133548
        y97b |   .6338043   .4187646     1.51   0.130    -.1869593    1.454568
        y98b |    .273375   .4579277     0.60   0.551    -.6241467    1.170897
       tobs3 |   .0222168   .0562549     0.39   0.693    -.0880408    .1324744
       tobs4 |   .0884656   .0891879     0.99   0.321    -.0863394    .2632706
       _cons |  -1.856402   .6052343    -3.07   0.002    -3.042639   -.6701641
-------------+----------------------------------------------------------------
lnsigma2     |
       tobs3 |   .2007709   .0566528     3.54   0.000     .0897335    .3118083
       tobs4 |   .5504932   .1162986     4.73   0.000     .3225522    .7784343
------------------------------------------------------------------------------
Wald test of lnsigma2=0:             chi2(2) =    32.52   Prob > chi2 = 0.0000
```

The coefficients are scaled and cannot be interpreted directly. However, we can easily obtain the average partial effects either manually using case 1 of formula (9) or automatically via `margins`:

```
. predictnl double pe=normalden(xb(#1)/exp(xb(#2)))*_b[lavgrexp]/exp(xb(#2))
. summarize pe

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         pe  |      7150    .0359899    .0072119   .0174126   .0455671

. margins, dydx(lavgrexp)

Average marginal effects                          Number of obs   =       7150
Model VCE    : Robust

Expression   : E[math4|x], predict()
dy/dx w.r.t. : lavgrexp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .0359899   .0231872     1.55   0.121    -.0094561    .0814359
------------------------------------------------------------------------------
```

Given that these are CRE models, the APEs are obtained by averaging over the unobserved heterogeneity in the cross-section dimension. In addition, we average over the time dimension to obtain a single scale factor. It's crucial to understand the underlying assumptions and the circumstances in which this device works, as it's easy to obtain effects that are actually not identified. For an in-depth treatment of these issues please refer to Papke and Wooldridge (2008) and Wooldridge (2010a, chap. 18).

References

Gould, W., J. Pitblado, and B. Poi. 2010. *Maximum Likelihood Estimation With Stata.* 4 ed.. Stata Press.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. Pseudo Maximum Likelihood Methods: Theory. *Econometrica*: 681–700.

Greene, W.H. 2011. *Econometric Analysis.* 7 ed.. Prentice Hall.

Harvey, A.C. 1976. Estimating Regression Models With Multiplicative Heteroscedasticity. *Econometrica: Journal of the Econometric Society*: 461–465.

Papke, L.E. 2005. The Effects of Spending on Test Pass Rates: Evidence From Michigan. *Journal of Public Economics* 89, no. 5-6: 821–839.

Papke, L.E., and J.M. Wooldridge. 1996. Econometric Methods for Fractional Response Variables With an Application To 401 (k) Plan Participation Rates. *Journal of Applied Econometrics* 11, no. 6: 619–632.

———. 2008. Panel Data Methods for Fractional Response Variables With an Application To Test Pass Rates. *Journal of Econometrics* 145, no. 1–2: 121–133.

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

———. 2010b. Correlated Random Effects Models With Unbalanced Panels. *Manuscript.*