

Analyzing Proportions: Fractional Response and Zero One Inflated Beta Models

Richard Williams, University of Notre Dame, <https://academicweb.nd.edu/~rwilliam/>

Last revised November 18, 2024

Sources:

Buis, M. L. (2020). Analysis of Proportions. In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), SAGE Research Methods Foundations. <https://dx.doi.org/10.4135/9781526421036852590>

Jeffrey Wooldridge, 2011, "Fractional response models with endogenous explanatory variables and heterogeneity", http://www.stata.com/meeting/chicago11/materials/ch11_wooldridge.pdf.

"Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates" Leslie E. Papke and Jeffrey M. Wooldridge, Journal of Applied Econometrics, Vol. 11, No. 6 (Nov. - Dec., 1996), pp. 619-632. [https://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1099-1255\(199611\)11:6%3C619::AID-JAE418%3E3.0.CO;2-1](https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1255(199611)11:6%3C619::AID-JAE418%3E3.0.CO;2-1).

"How do you fit a model when the dependent variable is a proportion?" <http://www.stata.com/support/faqs/statistics/logit-transformation/>.

"How does one do regression when the dependent variable is a proportion?" <https://stats.oarc.ucla.edu/stata/faq/how-does-one-do-regression-when-the-dependent-variable-is-a-proportion/>

"Stata Tip 63: Modeling Proportions" Kit Baum, The Stata Journal, Volume 8 Number 2: pp. 299-303 <http://www.stata-journal.com/article.html?article=st0147>

"Stata command for fractional logit with endogenous regressor?" <https://www.statalist.org/forums/forum/general-stata-discussion/general/1410304-stata-command-for-fractional-logit-with-endogenous-regressor#post1410304>

NOTE: Material in handout is current as of April 3, 2022. Since `fracglm` and `fracivp` are still in beta form, there may be changes in the future. (But it won't surprise me if they remain beta forever!)

In many cases, the dependent variable of interest is a proportion, i.e. its values range between 0 and 1. Wooldridge (1996, 2011) gives the example of the proportion of employees that participate in a company's pension plan. Baum (2008) gives as examples the share of consumers' spending on food, the fraction of the vote for a candidate, or the fraction of days when air pollution is above acceptable levels in a city. This handout will discuss a few different ways for analyzing such dependent variables: fractional response models (both heteroskedastic and non-heteroskedastic), zero one-inflated beta models, and fractional ivprobit models.

Fractional Response Models. As Wooldridge notes, many Stata commands (`logit`, `probit`, `hetprobit`) could analyze DVs that are proportions, but they impose the data constraint that the dependent variable must be coded as either 0 or 1, i.e. you can't have a proportion as the dependent variable even though the same formulas and estimation techniques would be appropriate with a proportion. Wooldridge offers his own short programs that relax this limitation; but a more flexible solution is offered by Richard Williams' user-written routine, `fracglm`, currently in (perpetual) beta testing. `fracglm` is adapted from `oglm`, and is probably easier to use than `oglm` when the dependent variable is a dichotomy (rather than an ordinal variable with 3 or more categories.)

To get `fracglm`, from within Stata type

```
net install fracglm, from(https://academicweb.nd.edu/~rwilliam/stata)
```

This is usually the best way to install. Files are placed in the right locations, and `adoupdate` will capture any updates.

That doesn't always work though. If it doesn't work for you, try pointing your browser to

<https://academicweb.nd.edu/~rwilliam/stata/fracglmbeta.zip>

Download the file (it may download automatically), unzip it, and follow the directions for installing that are in the `Readme.txt` file.

The following is adapted from the help for `fracglm`:

`fracglm` estimates Fractional Response Generalized Linear Models (e.g. Fractional Probit, Fractional Logit) with or without heteroskedasticity. Fractional response variables range in value between 0 and 1. An example of a fractional response variable would be the percentage of employees covered by an employer's pension plan.

`fracglm` also works with binary 0/1 dependent variables. `fracglm` supports multiple link functions, including logit (the default), probit, complementary log-log, log-log, log and cauchit. When these models include equations for heteroskedasticity they are also known as heterogeneous choice/ location-scale / heteroskedastic regression models.

`fracglm` fills gaps left by other Stata commands. Commands like `logit`, `probit` and `hetprob` do not allow for fractional response variables. `glm` can estimate some fractional response models but does not allow an equation for heteroskedasticity.

Several special cases of generalized linear models can also be estimated by `fracglm`, including the binomial generalized linear models of `logit`, `probit` and `cloglog` (which also assume homoskedasticity), `hetprob`, as well as similar models that are not otherwise estimated by Stata. This makes `fracglm` particularly useful for testing whether constraints on a model (e.g. homoskedastic errors) are justified, or for determining whether one link function is more appropriate for the data than are others.

In addition, Stata 14 introduced the `fracreg` command. It isn't quite as flexible as `fracglm` (e.g. it doesn't support as many link functions) but if you have Stata 14 it may be fine for your needs.

Example. Papke and Wooldridge (1996) give an example of participation rates in employer 401(k) pension plans. "Pension plan administrators are required to file Form 5500 annually with the Internal Revenue Service, describing participation and contribution behavior for each plan offered. Papke (1995) uses the plan level data to study, among other things, the relationship between the participation rate and various plan characteristics, including the rate at which a firm matches employee contributions."

In Wooldridge's (2011) version of this example, data are from 4,075 companies in 1987. The key variables used in this analysis are:

```
. use https://academicweb.nd.edu/~rwilliam/statafiles/401kpart, clear
. codebook prate mrate ltotemp age sole, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
prate	4075	2597	.840607	.0036364	1	partic/employ
mrate	4075	3521	.463519	0	2	plan match rate, per \$
ltotemp	4075	2147	6.97439	4.65396	13.00142	log(totemp)
age	4075	50	8.186503	1	71	age of the plan
sole	4075	2	.3693252	0	1	=1 if only pension plan

Wooldridge (2011) gives an example of a fractional probit model with heteroskedasticity. He recommends using robust standard errors (otherwise the standard errors are too large; you can confirm this by rerunning the following example with `vce(oim)`; you will see dramatic differences in the test statistics and standard errors.) He wrote his own program for this but `fracglm` can easily reproduce his results.

```
. fracglm prate mrate ltotemp age i.sole, het(mrate ltotemp age i.sole) link(p)
```

```
Heteroskedastic Fractional Probit Regression      Number of obs   =      4075
                                                    Wald chi2(4)    =      152.29
                                                    Prob > chi2     =      0.0000
Log pseudolikelihood = -1674.5212                Pseudo R2       =      0.0632
```

	prate	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
prate						
mrate		1.384694	.223861	6.19	0.000	.9459349 1.823454
ltotemp		-.1495096	.013966	-10.71	0.000	-.1768825 -.1221367
age		.0670722	.0100639	6.66	0.000	.0473474 .086797
1.sole		-.11827	.0932336	-1.27	0.205	-.3010046 .0644645
_cons		1.679377	.1058994	15.86	0.000	1.471818 1.886936
lnsigma						
mrate		.240357	.0537812	4.47	0.000	.1349479 .3457662
ltotemp		.0375185	.0144217	2.60	0.009	.0092525 .0657845
age		.0171714	.0027289	6.29	0.000	.0118229 .0225199
1.sole		-.1627546	.0631069	-2.58	0.010	-.2864417 -.0390674

[NOTE: `vce(robust)` is the default for both `fracglm` and `fracreg`. If you are using `fracglm` with a binary dependent variable, you may wish to specify `vce(oim)` instead.]

Wooldridge (2011) notes that a simple Wald test can be used to determine whether the coefficients in the heteroskedasticity equation are significantly different from zero. (I believe this is better than a likelihood ratio test because LR tests are problematic when using robust standard errors).

```

. test [lnsigma]

( 1) [lnsigma]mrate = 0
( 2) [lnsigma]ltotemp = 0
( 3) [lnsigma]age = 0
( 4) [lnsigma]0b.sole = 0
( 5) [lnsigma]1.sole = 0
    Constraint 4 dropped

      chi2( 4) = 109.26
    Prob > chi2 = 0.0000

```

You interpret these results pretty much the same way you would interpret the results from a hetprobit model. A higher match rate, an older fund, and having fewer employees all increase participation rates.

If you want to make results more tangible, you can use methods like we have used before. For example,

```
. margins, dydx(*)
```

```

Average marginal effects              Number of obs =      4075
Model VCE      : Robust

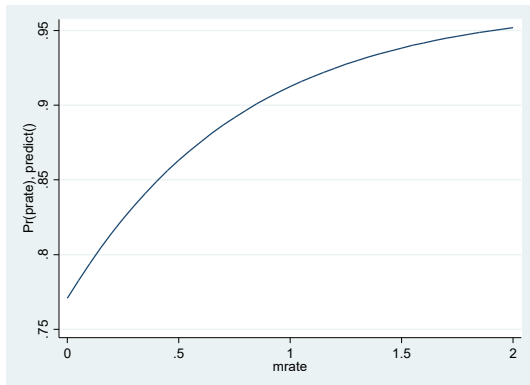
Expression      : Pr(prate), predict()
dy/dx w.r.t.    : mrate ltotemp age 1.sole

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
mrate	.159632	.0114214	13.98	0.000	.1372466	.1820175
ltotemp	-.0306262	.0020032	-15.29	0.000	-.0345524	-.0267
age	.0065659	.0006323	10.38	0.000	.0053266	.0078052
1.sole	.016089	.0061307	2.62	0.009	.0040729	.028105

Note: dy/dx for factor levels is the discrete change from the base level.

```
. mcp mrate, at1(0 (.05) 2)
```



Wooldridge (2011) also says you “should do a comparison of average partial effects [aka average marginal effects] between ordinary fractional probit and heteroskedastic fractional probit.” Non-heteroskedastic models can also be estimated with `fracglm`:

```
. fracglm prate mrate ltotemp age i.sole, link(p)
```

```
Fractional Probit Regression          Number of obs   =      4075
                                      Wald chi2(4)     =      695.89
                                      Prob > chi2     =      0.0000
Log pseudolikelihood = -1681.9607      Pseudo R2      =      0.0591
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
prate						
mrate	.5955726	.038756	15.37	0.000	.5196123	.6715329
ltotemp	-.1172851	.0080003	-14.66	0.000	-.1329655	-.1016048
age	.0180259	.0014218	12.68	0.000	.0152392	.0208126
1.sole	.0944158	.0271696	3.48	0.001	.0411645	.1476672
_cons	1.428854	.0593694	24.07	0.000	1.312493	1.545216

```
. margins, dydx(*)
```

```
Average marginal effects          Number of obs   =      4075
Model VCE      : Robust
```

```
Expression      : Pr(prate), predict()
dy/dx w.r.t.    : mrate ltotemp age 1.sole
```

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
mrate	.1362769	.0088064	15.47	0.000	.1190167	.1535372
ltotemp	-.0268368	.0018454	-14.54	0.000	-.0304537	-.0232199
age	.0041246	.0003277	12.59	0.000	.0034824	.0047669
1.sole	.0213349	.0060421	3.53	0.000	.0094927	.0331771

Note: dy/dx for factor levels is the discrete change from the base level.

Based on the earlier Wald test we would prefer the heteroskedastic model. You can also see that there are some modest differences in the Average Marginal Effects estimated by the two models.

Using Stata 14's `fracreg` instead, the heteroskedastic and non-heteroskedastic models are estimated by

```
. use https://academicweb.nd.edu/~rwilliam/statafiles/401kpart, clear
. fracreg probit prate mrate ltotemp age i.sole, het(mrate ltotemp age i.sole) nolog
```

```
Fractional heteroskedastic probit regression  Number of obs   =      4,075
                                                Wald chi2(5)    =      152.30
                                                Prob > chi2     =      0.0000
Log pseudolikelihood = -1674.5212              Pseudo R2      =      0.0088
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
prate						
mrate	1.384675	.2238372	6.19	0.000	.9459625	1.823388
ltotemp	-.1495096	.0139658	-10.71	0.000	-.1768822	-.1221371
age	.0670714	.0100629	6.67	0.000	.0473485	.0867943
1.sole	-.1182733	.0932298	-1.27	0.205	-.3010003	.0644537
_cons	1.679372	.1058965	15.86	0.000	1.471819	1.886926
lnsigma						
mrate	.2403557	.0537805	4.47	0.000	.1349478	.3457635
ltotemp	.0375172	.01442	2.60	0.009	.0092545	.0657799
age	.0171714	.0027289	6.29	0.000	.0118229	.0225199
1.sole	-.1627574	.063104	-2.58	0.010	-.2864389	-.0390759

```
. fracreg probit prate mrate ltotemp age i.sole, nolog
```

```
Fractional probit regression          Number of obs   =      4,075
                                      Wald chi2(5)     =      695.89
                                      Prob > chi2      =      0.0000
Log pseudolikelihood = -1681.9607      Pseudo R2       =      0.0591
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
prate						
mrate	.5955726	.038756	15.37	0.000	.5196123	.6715329
ltotemp	-.1172851	.0080003	-14.66	0.000	-.1329655	-.1016048
age	.0180259	.0014218	12.68	0.000	.0152392	.0208126
1.sole	.0944158	.0271696	3.48	0.001	.0411645	.1476672
_cons	1.428854	.0593694	24.07	0.000	1.312493	1.545216

Other Comments on Fractional Response Models:

1. Other than the fact that the heteroskedastic model fits better in this case, what is the rationale for it? I asked Jeffrey Wooldridge about this and he emailed me the following:

I think of it [the heteroskedastic model] mainly as a simple way to get a more flexible functional form. But this can also be derived from a model where, say, $y(i)$ is the fraction of successes out of $n(i)$ Bernoulli trials, where each binary outcome, say $w(i,j)$, follows a heteroskedastic probit. Then $E(y(i)|x(i))$ would have the form estimated by your Stata command.

Or, we could start with an omitted variable formulation: $E[y(i)|x(i),c(i)] = \text{PHI}[x(i)*b + c(i)]$, where the omitted variable $c(i)$ is distributed as Normal with mean zero and variance $h(x(i))$. As an approximation, we might use an exponential function for $1 + h(x(i))$, and then that gives the model, too.

2. As noted in Wooldridge (2011) and in the Stata FAQ cited above, the `glm` command can also be used to estimate non-heteroskedastic models. Specify `family(binomial)` and either `link(p)` or `link(l)`. These are the same results that `fracglm` gave for the non-heteroskedastic model.

```
. glm prate mrate ltotemp age i.sole, vce(robust) link(p) family(binomial) nolog
note: prate has noninteger values
```

```
Generalized linear models          No. of obs   =      4075
Optimization      : ML             Residual df   =      4070
                                      Scale parameter =      1
Deviance          = 885.9205448      (1/df) Deviance = .2176709
Pearson           = 896.7484978      (1/df) Pearson = .2203313

Variance function: V(u) = u*(1-u/1)      [Binomial]
Link function      : g(u) = invnorm(u)    [Probit]

Log pseudolikelihood = -1289.354251      AIC           = .6352659
                                      BIC           = -32946.47
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
prate					
mrate	.5955726	.038756	15.37	0.000	.5196123 .6715329
ltotemp	-.1172851	.0080003	-14.66	0.000	-.1329655 -.1016048
age	.0180259	.0014218	12.68	0.000	.0152392 .0208126
1.sole	.0944158	.0271696	3.48	0.001	.0411645 .1476672
_cons	1.428854	.0593694	24.07	0.000	1.312493 1.545216

Zero Inflated Beta Models. The Stata FAQ (<http://www.stata.com/support/faqs/statistics/logit-transformation/>) warns that other types of models may be advisable depending on why the 0s or 1s exist. From the FAQ (it talks about a logit transformation but the same is true for probit):

A traditional solution to this problem [the dependent variable is a proportion] is to perform a logit transformation on the data. Suppose that your dependent variable is called y and your independent variables are called X . Then, one assumes that the model that describes y is

$$y = \text{invlogit}(XB)$$

If one then performs the logit transformation, the result is

$$\ln\left(\frac{y}{1-y}\right) = XB$$

We have now mapped the original variable, which was bounded by 0 and 1, to the real line. One can now fit this model using OLS or WLS, for example by using `regress`. Of course, one cannot perform the transformation on observations where the dependent variable is zero or one; the result will be a missing value, and that observation would subsequently be dropped from the estimation sample.

A better alternative is to estimate using `glm` with **family(binomial)**, **link(logit)**, and **robust**; this is the method proposed by Papke and Wooldridge (1996).

In either case, there may well be a substantive issue of interpretation. Let us focus on interpreting zeros: the same kind of issue may well arise for ones. Suppose the y variable is proportion of days workers spend off sick. There are two extreme possibilities. The first extreme is that all observed zeros are in effect sampling zeros: each worker has some nonzero probability of being off sick, and it is merely that some workers were not, in fact, off sick in our sample period. Here, we would often want to include the observed zeros in our analysis and the `glm` route is attractive. The second extreme is that some or possibly all observed zeros must be considered as structural zeros: these workers will not ever report sick, because of robust health and exemplary dedication. These are extremes, and intermediate cases are also common. In practice, it is often helpful to look at the frequency distribution: a marked spike at zero or one may well raise doubt about a single model fitted to all data.

A second example might be data on trading links between countries. Suppose the y variable is proportion of imports from a certain country. Here a zero might be structural if two countries never trade, say on political or cultural grounds. A model that fits over both the zeros and the nonzeros might not be advisable, so that a different kind of model should be considered.

Baum (2008) elaborates on the problem of structural zeros and 1s. He notes “the managers of a city that spends none of its resources on preschool enrichment programs have made a discrete choice. A hospital with zero heart transplants may be a facility whose managers have chosen not to offer certain advanced services. In this context, the glm approach, while properly handling both zeros and ones, does not allow for an alternative model of behavior generating the limit values.”

He suggests alternatives such as the “zero-inflated beta” model, which allows for zero values (but not unit values) in the proportion and for separate variables influencing the zero and nonzero values (i.e. something similar to the zero-inflated or hurdle models that you have for count data). A one-inflated beta model allows for separate variables influencing the one and non-one values

Both the zero and one inflated beta models can be estimated via Maarten Buis’s `zoib` program, available from SSC. The help file for `zoib` says

```
zoib fits by maximum likelihood a zero one inflated beta distribution to a distribution of a variable
depar. depar ranges between 0 and 1: for example, it may be a proportion. It will estimate the
probabilities of having the value 0 and/or 1 as separate processes. The logic is that we can often
think of proportions of 0 or 1 as being qualitatively different and generated through a different
process as the other proportions.
```

Here is how we can apply the one-inflated beta model to the current data. In these data, no company has a value of zero, but about a third of the cases have a value of 1, so we use the `oneinflate` option to model the 1s separately.

```
. zoib prate mrate ltotemp age i.sole, oneinflate( mrate ltotemp age i.sole)

Iteration 0:  log likelihood = -1350.3099
Iteration 1:  log likelihood = -881.01326
Iteration 2:  log likelihood = -860.4238
Iteration 3:  log likelihood = -860.34541
Iteration 4:  log likelihood = -860.34541

ML fit of oib                                Number of obs =          4075
Log likelihood = -860.34541                   Wald chi2(4) =          438.05
                                              Prob > chi2 =          0.0000

-----+-----
      prate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
proportion
  mrate |   .7549614   .0570378    13.24   0.000    .6431693   .8667535
  ltotemp |  -.1138555   .0111967   -10.17   0.000   -1.1358006  -.0919104
    age |   .0236683   .0022057    10.73   0.000    .0193452   .0279915
  1.sole |   .0035928   .0397814     0.09   0.928   -1.0743772   .0815629
    _cons |   1.507286   .0870497    17.32   0.000    1.336672   1.677901
-----+-----
oneinflate
  mrate |   .9482935   .0859375    11.03   0.000    .7798591   1.116728
  ltotemp |  -.2918532   .027768   -10.51   0.000   -1.3462776  -.2374288
    age |   .0190046   .0038664     4.92   0.000    .0114266   .0265826
  1.sole |   .6041419   .0762853     7.92   0.000    .4546255   .7536583
    _cons |   .4242109   .2017944     2.10   0.036    .0287012   .8197205
-----+-----
ln_phi
  _cons |   1.621576   .0262855    61.69   0.000    1.570057   1.673094
-----+-----
```

The one-inflate equation shows that companies with higher match rates, fewer total employees, older plans, and that have only one pension plan available are more likely to have 100% participation in their plans. When participation is not 100%, these same variables (except sole) also are associated with higher participation rates.

Other Models & Programs. I am not familiar with most of these, but the help for `zoib` suggests that some of these programs may also sometimes be helpful when modeling proportions:

`betafit` fits by maximum likelihood a two-parameter beta distribution to a distribution of a variable `depvar`. `depvar` ranges between 0 and 1: for example, it may be a proportion.

`dirifit` fits by maximum likelihood a Dirichlet distribution to a set of variables `depvarlist`. Each variable in `depvarlist` ranges between 0 and 1 and all variables in `depvarlist` must, for each observation, add up to 1: for example, they may be proportions.

`fmlogit` fits by quasi maximum likelihood a fractional multinomial logit model. Each variable in `depvarlist` ranges between 0 and 1 and all variables in `depvarlist` must, for each observation, add up to 1: for example, they may be proportions. It is a multivariate generalization of the fractional logit model proposed by Papke and Wooldridge (1996).

For the latter two programs, the help files give as examples models where the dependent variables are the proportions of a municipality's budget that are spent on governing, public safety, education, recreation, social work, and urban planning. Independent variables include whether or not there are any left-wing parties in city government. If I understand the models correctly, the coefficients tell you how the independent variables increase or decrease the proportion of spending in each area. For example, the results show that when there is no left wing party in city government, less of the city budget tends to get spent on education.

Fractional ivprobit commands. `fracivp` is a beta program adapted from Stata 12's `ivprobit` program. It relaxes the assumption that the dependent variable be coded 0/1 and allows it to be a proportion instead. `fracivp` estimates Fractional Response Probit models with continuous endogenous regressors. This is a use at your own risk program; it seems to work ok but I haven't fully tested it yet. `cmp` (discussed next) may be a better (or at least more proven) choice. Comments are welcome. To get `fracivp`, from within Stata type

```
net install fracivp, from(https://academicweb.nd.edu/~rwilliam/stata)
```

That doesn't always work though. If it doesn't work for you, try pointing your browser to

<https://academicweb.nd.edu/~rwilliam/stata/fracivpbeta.zip>

Download the file (it may download automatically), unzip it, and follow the directions for installing that are in the Readme.txt file.

Another choice for fractional ivprobit (and lots of other things) is Dennis Roodman's `cmp` (Conditional mixed process estimator with multilevel random effects and coefficients) command (available from SSC). `cmp` is incredibly powerful. Among other things, it can estimate fractional ivprobit models. See <https://www.statalist.org/forums/forum/general-stata-discussion/general/1410304-stata-command-for-fractional-logit-with-endogenous-regressor> for a discussion. I'll give an example but read the thread and the `cmp` help file if you want to understand it better.

```
. use https://academicweb.nd.edu/~rwilliam/statafiles/401kpart, clear

. cmp setup
$cmp_out      = 0
$cmp_missing  = .
$cmp_cont     = 1
$cmp_left     = 2
$cmp_right    = 3
$cmp_probit   = 4
$cmp_oprobit  = 5
$cmp_mprobit  = 6
$cmp_int      = 7
$cmp_trunc    = 8 (deprecated)
$cmp_roprobit = 9
$cmp_frac     = 10

. cmp (prate = mrate ltotemp i.sole age) (age= mrate ltotemp i.sole agesq),
ind($cmp_frac $cmp_cont)
```

Note: fractional probit models imply `vce(robust)`.

Fitting individual models as starting point for full model fit.

Note: For programming reasons, these initial estimates may deviate from your specification.

For exact fits of each equation alone, run `cmp` separately on each.

Source	SS	df	MS	Number of obs	=	4,075
Model	25.1282708	4	6.2820677	F(4, 4070)	=	216.54
Residual	118.073925	4,070	.029010793	Prob > F	=	0.0000
Total	143.202196	4,074	.035150269	R-squared	=	0.1755
				Adj R-squared	=	0.1747
				Root MSE	=	.17033

	prate	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mrate		.1072729	.0066685	16.09	0.000	.0941991 .1203468
ltotemp		-.0281719	.0018764	-15.01	0.000	-.0318507 -.0244931
1.sole		.0177024	.0060337	2.93	0.003	.0058732 .0295317
age		.0037	.0002979	12.42	0.000	.0031159 .0042841
_cons		.9505378	.0143984	66.02	0.000	.9223091 .9787665

(4,075 real changes made)


```
. * Test only the first equation, since that is what fracivp does
. test [prate]
```

```
( 1) [prate]mrate = 0
( 2) [prate]ltotemp = 0
( 3) [prate]0b.sole = 0
( 4) [prate]1.sole = 0
( 5) [prate]age = 0
      Constraint 3 dropped
```

```
      chi2( 4) = 674.30
      Prob > chi2 = 0.0000
```

```
. fracivp prate mrate ltotemp i.sole (age=agesq), vce(robust) nolog
```

```
Probit model with endogenous regressors      Number of obs      =      4,075
                                             Wald chi2(4)       =      674.30
Log pseudolikelihood = -12351.342           Prob > chi2         =      0.0000
```

```
-----+-----
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0162198	.0014474	11.21	0.000	.0133829	.0190567
mrate	.6009217	.0386194	15.56	0.000	.5252291	.6766143
ltotemp	-.1160397	.0079939	-14.52	0.000	-.1317074	-.1003721
1.sole	.0961139	.0271793	3.54	0.000	.0428435	.1493842
_cons	1.431082	.0592794	24.14	0.000	1.314896	1.547267
-----+-----						
/athrho	.0319314	.0113532	2.81	0.005	.0096795	.0541832
/lnsigma	1.199452	.0643687	18.63	0.000	1.073292	1.325612
-----+-----						
rho	.0319205	.0113416			.0096792	.0541303
sigma	3.318298	.2135944			2.924991	3.764489

```
-----+-----
```

```
Instrumented: age
Instruments: mrate ltotemp 1.sole agesq
```

```
-----+-----
Wald test of exogeneity (/athrho = 0): chi2(1) = 7.91 Prob > chi2 = 0.0049
```

Both `fracivp` and `cmp` produce identical results, which makes me feel good about `fracivp`. `fracivp` may have more (untested) post-estimation options that might make it a better choice in some cases.