Estimating Heteroskedastic and Instrumental Variable Models for Binary Outcome Variables in R

by Mauricio Sarrias

Abstract The objective of this article is to introduce the package **Rchoice** which provides functionality for estimating heteroskedastic and instrumental variable models for binary outcomes, whith emphasis on the calculation of the average marginal effects. To do so, I introduce two new functions of the **Rchoice** package using widely known applied examples. I also show how users can generate publication-ready tables of regression model estimates.

1 Introduction

Often, applied researchers in different fields deal with binary (probit and logit) models that exhibit heteroskedasticity (the error variance is not homogeneous across individuals), or with endogenous variables.[1] In both cases, the standard binary logit and probit estimator will be inconsistent, which can lead to misleading conclusions (Yatchew and Griliches 1985; Wooldridge 2010).[2]

One widely used estimator to address heteroskedastic disturbances in the realm of binary outcomes is the fully parametric multiplicative heteroskedastic binary model (Keele and Park 2006). This model assumes that the error term's variance depends on specific known covariates. For example, Alvarez and Brehm (1995) use a heteroskedastic probit model to show that policy choices about abortion are heterogeneous due to unequal variances.[3]

If some of the regressor is endogenous, approaches such as the control function (CF, Wooldridge 2015) or the maximum likelihood estimator (MLE, Newey 1987; Rivers and Vuong 1988) allow to remediate the inconsistent estimates using an instrumental variables (IV) approach.

Routines for heteroskedastic and IV models exist in commercial software such as Stata (StataCorp 2019) and LIMDEP (Greene 2002). One advantage of Stata is that its command `margins` allows such models to quickly and flexibly compute marginal effects. This is very attractive for users who need to produce and export tables of estimates in Latex or other formats.

In this article, I review the main approaches and functions in R to estimate heteroskedastic and IV models for binary outcomes, with a special focus on applied examples and the computation of the marginal effects. Additionally, this article introduces two new functions of the **Rchoice** package (Sarrias 2016) that allow estimating both types of models. The first function, `hetprob()`, estimates binary dependent variable models assuming a parametric form for the heteroskedasticity. The model can be either the probit or logit model and the parameters are estimated by Maximum Likelihood (ML), which find the parameter values that make the observed data most probable under the assumptions of the statistical model.

The second function, `ivpml()`, estimates binary probit models with endogenous continuous variables using also the ML approach. As an additional feature, **Rchoice** also provides functions to compute the average marginal effects for both models under different modelling approaches: categorical variables, interactions terms, and quadratic variables. The package can also be used in concert with the **memisc** package (Elff 2012), which produces publication-ready tables of regression model estimates. Finally, I show that both functions produce the same estimates as the corresponding Stata commands.[4]

The function `hetprob()` is intended to complement other related packages in R. For example, the packages **glmx** (Zeileis, Koenker, and Doebler 2015) and **oglmx** (Carroll 2018) also allow to estimate heteroskedastic binary models using MLE. The latter has the advantage of being able to compute the marginal effects. However, the current version does not allow to identify functions of variables that enter the equations for the mean and standard equations, interaction terms, or polynomials. The `ivpml()` function provides the MLE for the probit model and hence complements the R package **ivprobit** (Zaghdoudi 2018) which provides a two-step procedure. Another is the **LARF** package (An

[1]In econometrics, endogeneity refers to situations in which an explanatory variable is correlated with the error term. The common sources of endogeneity are omitted variables, simultaneity, and measurement error.

[2]Inconsistency means that the estimator will not converge in probability to the true parameter.

[3]For other applications see Knapp and Seaks (1992) and Williams (2009).

[4]Stata codes for replicating the main results of this article are presented in **Appendix C** and **Appendix D**. Do files are available in the supplemental material.

and Wang 2016), which estimates local averages response functions for binary treatments and binary instruments.

2 Models

2.1 Heterokedastic binary model

The multiplicative heterokedastic binary model (also known as the location-scale binary model) for cross-sectional data has the following structure (Williams 2009):[5]

$$y_i^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{1}$$

$$\text{Var}(\epsilon_i|\mathbf{z}_i) = \sigma_i^2 = \sigma_\epsilon^2 \left[\exp\left(\mathbf{z}_i^\top \delta\right)\right]^2, \tag{2}$$

where y_i^* is the latent (unobserved) response variable for individual $i = 1, ..., n$, $\mathbf{x}_i$ is a k-dimensional vector of explanatory variables determining the latent variable y_i^*, $\boldsymbol{\beta}$ is the vector of parameters, and ϵ_i is the error term distributed either normally or logistically with $\mathbf{E}(\epsilon_i|\mathbf{z}_i, \mathbf{x}_i) = 0$ and multiplicative heterokedastic variance $\text{Var}(\epsilon_i|\mathbf{z}_i) = \sigma_i^2, \forall i = 1, ..., n$ (Harvey 1976). The variance for each individual is modeled parametrically assuming that it depends on a p-dimensional vector of observed variables $\mathbf{z}_i$, whereas δ is the vector of coefficients associated with each variable. It is important to emphasize that $\mathbf{z}_i$ does not include a constant, otherwise the parameters are not identified (Greene and Hensher 2010).

Since we do not observe y_i^*, we need a rule that relates the binary variable that we actually observe, y_i, to the latent variable. As it is standard, we use the following rule:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Using Equations (1), (2) and (3), the probability of observing $y_i = 1$ is:

$$\Pr(y_i = 1|\mathbf{x}_i, \mathbf{z}_i) = F\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\exp(\mathbf{z}_i^\top \delta)}\right), \tag{4}$$

where $F(\cdot)$ is either $\Phi(\cdot)$, that is, the cumulative distribution function (CDF) for the standard normal distribution, such that $\sigma_\epsilon^2 = 1$, or $\Lambda(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$, where $\Lambda(\cdot)$ represents the CDF for the standard logistic distribution, so that $\sigma_\epsilon^2 = \pi^2/3$.

Let $\boldsymbol{\theta}$ be the $(k + p)$-dimensional vector of all parameters. The vector $\boldsymbol{\theta}$ can be estimated using the Maximum Likelihood procedure. Using Equation (4), the MLE is the value of the parameters that maximizes the following log-likelihood function:[6]

$$\widehat{\boldsymbol{\theta}}_{ML} \equiv \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argmax}} \sum_{i=1}^n \ln\left\{\left[1 - F\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\exp(\mathbf{z}_i^\top \delta)}\right)\right]^{1-y_i} \left[F\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\exp(\mathbf{z}_i^\top \delta)}\right)\right]^{y_i}\right\}.$$

As in any non-linear model, the estimated coefficients alone cannot be interpreted as marginal changes on $\Pr(y_i = 1|\mathbf{x}_i, \mathbf{z}_i)$. Let w_k be a continuous variable appearing in both $\mathbf{x}$ and $\mathbf{z}$, then the partial effect is (see Greene 2003):

$$\frac{\partial \Pr(y_i = 1|\mathbf{x}_i, \mathbf{z}_i)}{\partial w_{ik}} = f\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\exp\left(\mathbf{z}_i^\top \delta\right)}\right)\left(\frac{\beta_k - (\mathbf{x}_i^\top \boldsymbol{\beta})\delta_k}{\exp\left(\mathbf{z}_i^\top \delta\right)}\right), \tag{5}$$

where $f(\cdot)$ is the probability density function (PDF) for the standard normal or standard logistic distribution. The average partial effect (APE) can be consistently estimated as follows:

$$\widehat{\text{APE}}_k = \frac{1}{n}\sum_{i=1}^n f\left(\frac{\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}}{\exp\left(\mathbf{z}_i^\top \widehat{\delta}\right)}\right)\left(\frac{\widehat{\beta}_k - (\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})\widehat{\delta}_k}{\exp\left(\mathbf{z}_i^\top \widehat{\delta}\right)}\right), \tag{6}$$

and their standard error can be estimated either by delta method or bootstrap. The delta method

[5] Multiplicative exponential heteroskedasticity was first proposed by Harvey (1976) for linear models. For identification of the multiplicative heterokedastic binary model see Carlson (2019).

[6] The analytic gradient and Hessian for the multiplicative heterokedastic binary model used by **Rchoice** are presented in **Appendix A**.

provides an analytic approximation for the standard errors based on the asymptotic variance-covariance matrix of the MLE. The bootstrap is non-parametric resampling technique, which involves generating a large number of resampled datasets (bootstrap samples) and estimating (6) for each sample. For further details see Wooldridge (2010).

Finally, a likelihood-ratio (LR) or Wald test can be performed to test the null hypothesis of homoskedasticity: $H_0 : \delta = \mathbf{0}$.

2.2 Probit models with endogenous continous variable

Consider the following two-equation model:

$$y_{1i}^* = \mathbf{x}_{1i}^\top \boldsymbol{\beta}_1 + \gamma y_{2i} + \epsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad (7)$$

$$y_{2i} = \mathbf{x}_{1i}^\top \delta_1 + \mathbf{x}_{i2}^\top \delta_2 + v_i = \mathbf{z}_i^\top \delta + v_i, \quad (8)$$

$$y_{1i} = \mathbf{1}\left[y_{1i}^* > 0\right], \quad (9)$$

where $i = 1, ..., n$, y_{1i}^* is a latent (unobserved) response variable for individual i and we observe $y_{1i} = 1$ if and only if $\mathbf{1}\left[y_{1i}^* > 0\right]$, y_{2i} is the **continuous endogenous** variable, $\mathbf{x}_{i1}$ is a k_1-dimensional vector of predetermined (exogenous) variables, $\mathbf{x}_{i2}$ is a k_2-dimensional vector of additional (exogenous) instruments, $\mathbf{x}_i = \left(\mathbf{x}_{1i}^\top, y_{2i}\right)^\top$ is a $k \times 1$ column vector such that $k = k_1 + 1$, and $\mathbf{z}_i = \left(\mathbf{x}_{1i}^\top, \mathbf{x}_{2i}^\top\right)^\top$ is a $p \times 1$ vector where $p = k_1 + k_2$. Equation (7) is the structural equation, whereas Equation (8) is the first-stage equation. Further, assume that (ϵ, v) are distributed as bivariate normal with zero mean.

Two-step approach

The simplest approach for estimating the parameters of Equation (7) and (8) is using a two-step procedure (Rivers and Vuong 1988) also known as Control Function (CF) approach (Wooldridge 2015). Under joint normality of (ϵ, v), we can write ϵ as a function of v as follows:[7]

$$\epsilon_i | v_i = \frac{\sigma_\epsilon}{\sigma_v}\rho v_i + \eta_i, \quad (10)$$

where $\text{Var}(\epsilon_i) = \sigma_\epsilon^2$, $\text{Var}(v_i) = \sigma_v^2$, $\eta_i \sim N\left[0, (1-\rho^2)\sigma_\epsilon^2\right]$ and $\rho = \text{Cov}(\epsilon_i, v_i)/(\sigma_\epsilon \cdot \sigma_v)$. If $\rho = 0$, y_2 is exogenous and the traditional probit model will deliver consistent estimates. For identification, we need to set $\text{Var}(\epsilon_i) = 1$. Then Equation (10) can be re-written as:

$$\epsilon_i = \lambda v_i + \eta_i, \quad (11)$$

where $\eta_i \sim N\left[0, (1-\rho^2)\right]$ and $\lambda = \text{Cov}(\epsilon_i, v_i)/\sigma_v^2$. Inserting Equation (11) in the latent Equation (7) yields:

$$y_{1i}^* = \mathbf{x}_{1i}^\top \boldsymbol{\beta}_1 + \gamma y_{2i} + \lambda v_i + \eta_i,$$

and the probability of observing $y_{1i} = 1$ is:

$$\Pr(y_{1i} = 1 | y_{2i}, \mathbf{z}_i, v_i) = \Pr(y_{1i}^* > 0 | y_{2i}, \mathbf{z}_i, v_i) = \Phi\left(\mathbf{x}_{1i}^\top \boldsymbol{\beta}_1^* + \gamma^* y_{2i} + \lambda^* v_i\right). \quad (12)$$

Thus, if we knew v_i, a probit of y_1 on $\mathbf{x}$ and v would consistently estimate the scaled parameters $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1/\sqrt{1-\rho^2}$, $\gamma^* = \gamma/\sqrt{1-\rho^2}$, and $\lambda^* = \lambda/\sqrt{1-\rho^2}$. Using this idea, the estimation procedure is as follows (see Wooldridge 2010, sect. 15.7.2):

- Run an OLS regression of y_2 on $\mathbf{z}$ (Equation (8)) and compute the residuals $\widetilde{v}_i = y_{2i} - \mathbf{z}_i^\top \widetilde{\delta}$. Both $\widetilde{\delta}$ and $\widetilde{v}$ are consistently estimated.
- Run the probit y_1 on $\mathbf{x}_1$, y_2 and $\widetilde{v}$ to get consistent estimators of the scaled coefficients $\boldsymbol{\beta}^*$, γ^* and λ^*.

Note that the term control function comes from the fact that the inclusion of $\widetilde{v}$ in the second step controls for the correlation between ϵ_i and v_i.

Some of the structural parameters can be recovered after the two-step procedure. Since $\sigma_\epsilon = 1$, $\rho = \text{Cov}(\epsilon_i, v_i)/\sigma_v = \lambda \cdot \sigma_v$. Thus, an estimate of ρ can be recovered from:

$$\widehat{\rho} = \widehat{\lambda}^* \cdot \widetilde{\sigma}_v, \quad (13)$$

[7] If $x \sim N(\mu, \sigma^2)$, then we can write $x_i = \mu + \sigma u_i$, where $u_i \sim N(0, 1)$.

where $\widehat{\lambda}^*$ is the probit estimate of λ^* and $\widetilde{\sigma}_v$ is the square root of the usual error variance estimator from the first-stage regression. The unscaled parameters can also be recovered using the two-stage estimates. For instance, since $\gamma^* = \gamma/\sqrt{(1-\rho^2)}$, and using our result in Equation (13), then $\widehat{\gamma} = \widehat{\gamma}^*\left[1-\left(\widehat{\lambda}^*\cdot\widetilde{\sigma}_v\right)^2\right]^{1/2}$.

As explained by Wooldridge (2010), the usual probit z-statistic on $\widetilde{v}$ is a valid test of the null hypothesis that y_2 is exogenous: $H_0 : \lambda^* = 0$.[8] However, the estimated variance-covariance matrix of the probit model does not deliver correct standard errors for the rest of the parameters since it does not include the sampling variability of $\widehat{\delta}$ when $\lambda \neq 0$.

Following Wooldridge (2015), the APEs are obtained by taking either derivatives or differences (depending on whether the explanatory variable is continuous or discrete) of the Average Structural Function (ASF) given by:

$$\text{ASF}(\mathbf{x}_1, y_2) = \mathbf{E}_v\left[\Phi\left(\mathbf{x}_{1i}^\top\boldsymbol{\beta}_1^* + \gamma^* y_{2i} + \lambda^* v_i\right)\right]. \tag{14}$$

This function averages out the first-stage residuals v_i, purging the model of endogeneity. Under the weak law of large numbers, a consistent estimator for $\text{ASF}(\mathbf{x}_1, y_2)$ in Equation (14) is:

$$\widehat{\text{ASF}} = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\mathbf{x}_{1i}^\top\widehat{\boldsymbol{\beta}}_1^* + \widehat{\gamma}^* y_{2i} + \widehat{\lambda}^* v_i\right), \tag{15}$$

which incorporates the estimated unobservables from the first stage without perturbing them. Hence, to estimate the APE for y_2 we can compute:

$$\widehat{\text{APE}}_{y_2} = \widehat{\gamma}^*\frac{1}{n}\sum_{i=1}^{n}\phi\left(\mathbf{x}_{1i}^\top\widehat{\boldsymbol{\beta}}_1^* + \widehat{\gamma}^* y_{2i} + \widehat{\lambda}^* v_i\right). \tag{16}$$

where $\phi(\cdot)$ is the standard normal density function. A standard error for this $\widehat{\text{APE}}$ can be obtained via the delta method or bootstrap.

2.3 Maximum Likelihood approach

We can also estimate the parameters using the MLE. To derive the log-likelihood function, we need to find the joint distribution $f(y_{1i}, y_{2i}|\mathbf{z}) = f(y_{1i}|y_{2i}, \mathbf{z}_i)f(y_{2i}|\mathbf{z}_i)$. Under the joint normality, $y_{2i}|\mathbf{z}_i \sim N(\mathbf{z}_i^\top\delta, \sigma_v^2)$ and its conditional marginal density is (Wooldridge 2014):

$$f(y_{2i}|\mathbf{z}_i) = \frac{1}{\sigma_v}\phi\left(\frac{y_{2i} - \mathbf{z}_i^\top\delta}{\sqrt{1-\rho^2}}\right). \tag{17}$$

Using the fact that the normal distribution is symmetric, the conditional density of y_{2i} given $(y_{2i}, \mathbf{z}_i)$ can be written as:

$$f(y_{1i}|y_{2i}, \mathbf{z}_i) = \Phi\left[q_i\cdot\left(\frac{\mathbf{x}_i^\top\boldsymbol{\beta} + \frac{\rho}{\sigma_v}\left(y_{2i} - \mathbf{z}_i^\top\delta\right)}{\sqrt{1-\rho^2}}\right)\right], \tag{18}$$

where $q_i = 2y_{2i} - 1$ (see Greene 2003). Using Equations (17) and (18), the joint probability for each individual i is:

$$f(y_{1i}, y_{2i}|\mathbf{z}_i;\boldsymbol{\theta}) = \Phi\left[q_i\cdot\left(\frac{\mathbf{x}_i^\top\boldsymbol{\beta} + \frac{\rho}{\sigma_v}\left(y_{2i} - \mathbf{z}_i^\top\delta\right)}{\sqrt{1-\rho^2}}\right)\right]\frac{1}{\sigma_v}\phi\left(\frac{y_{2i} - \mathbf{z}_i^\top\delta}{\sqrt{1-\rho^2}}\right). \tag{19}$$

The MLE is a value of the parameter vector that maximizes the following expression:[9]

$$\widehat{\boldsymbol{\theta}}_{ML} \equiv \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argmax}}\sum_{i=1}^{n}\ln\left\{\Phi\left[q_i\cdot\left(\frac{\mathbf{x}_i^\top\boldsymbol{\beta} + \frac{\rho}{\sigma_v}\left(y_{2i} - \mathbf{z}_i^\top\delta\right)}{\sqrt{1-\rho^2}}\right)\right]\frac{1}{\sigma_v}\phi\left(\frac{y_{2i} - \mathbf{z}_i^\top\delta}{\sqrt{1-\rho^2}}\right)\right\}.$$

[8] Under the null $H_0 : \lambda^* = 0$ it is true that $\epsilon = v$ and therefore the distribution of v does not play any role under the null.

[9] The analytic gradient and Hessian for the MLE used by **Rchoice** are presented in **Appendix B**.

After the parameters are estimated, the APE for the endogenous variable can be estimated as:

$$\widehat{\text{APE}}_{y_2} = \frac{\widehat{\gamma}}{\sqrt{1-\widehat{\rho}^2}}\frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}} + \frac{\widehat{\rho}}{\widehat{\sigma}_v}\widehat{v}_i}{\sqrt{1-\widehat{\rho}^2}}\right). \tag{20}$$

A second option would be to compute the effect for the structural model assuming that endogeneity does not exist (the values of the covariates are given and fixed). In this case, the APE for the endogenous variable is computed as:

$$\widehat{\text{APE}}_{y_2} = \widehat{\gamma}\frac{1}{n}\sum_{i=1}^{n}\phi\left(\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}\right). \tag{21}$$

3 Applications

3.1 Heteroskedastic binary models

Promotion of scientists

To show how R can be used to fit heteroskedastic binary response models, I first use Allison (1999)'s dataset called "`tenure.cvs`" (see also Williams 2010). The data consists of observations of the careers of university professors over time, tracking multiple cross-sectional and longitudinal indicators including gender, the number of published article, and quality of department, among others.

We can load the dataset into R as follows:

```
tenure_data <- read.csv(file = 'tenure.csv')
```

Following Allison (1999) and Williams (2009) I focus on whether women get a lower payoff from their published work than men. First, I estimate a binary logit model using the `glm()` function for men and women separately, where the structural model is given by

$$\begin{aligned} \text{tenure}^* &= \beta_0 + \beta_1\text{year} + \beta_2\text{year}^2 + \beta_3\text{select} + \beta_4\text{articles} + \beta_5\text{prestige} + \epsilon, \\ \text{tenure} &= \mathbf{1}\left[\text{tenure}^* > 0\right], \end{aligned}$$

where ϵ is distributed logistically with mean 0 and variance $\pi^2/3$. The dependent variable, `tenure`, is whether an assistant professor was promoted in that year, and 0 otherwise, `year` is the number of years since the beginning of the assistant professorship, `select` is a measure of undergraduate selectivity of the colleges where scientists received their bachelor's degree, `articles` is the cumulative number of articles published by the end of each person-year, and `prestige` is a measure of prestige of the department in which scientist was employed. To obtain similar results as Allison (1999), I restrict the sample to `year <= 10`. Thus, each person has one record per year of service as an assistant professor, for as many as ten years.

```
sub_data <- subset(tenure_data, year <= 10)
logit_m <- glm(tenure ~  year + I(year^2) + select + articles + prestige,
               subset = (female == 0),
               data   = sub_data,
               family = binomial(link = "logit"))
logit_w <- glm(tenure ~  year + I(year^2) + select + articles + prestige,
               subset = (female == 1),
               data   = sub_data,
               family = binomial(link = "logit"))
```

To present the results I use the `mtable()` function from **memisc** package (Elff 2012).

```
library("memisc")
mtable("Logit for men" = logit_m,
       "Logit for women" = logit_w,
       summary.stats = c("Log-likelihood", "AIC", "BIC", "N"))

#>
#> Calls:
#> Logit for men: glm(formula = tenure ~ year + I(year^2) + select + articles +
```

```
#>     prestige, family = binomial(link = "logit"), data = sub_data,
#>     subset = (female == 0))
#> Logit for women: glm(formula = tenure ~ year + I(year^2) + select + articles +
#>     prestige, family = binomial(link = "logit"), data = sub_data,
#>     subset = (female == 1))
#>
#> ==================================================
#>                   Logit for men  Logit for women
#> --------------------------------------------------
#>   (Intercept)       -7.680***      -5.842***
#>                     (0.681)        (0.866)
#>   year               1.909***       1.408***
#>                     (0.214)        (0.257)
#>   I(year^2)         -0.143***      -0.096***
#>                     (0.019)        (0.022)
#>   select             0.216***       0.055
#>                     (0.061)        (0.072)
#>   articles           0.074***       0.034**
#>                     (0.012)        (0.013)
#>   prestige          -0.431***      -0.371*
#>                     (0.109)        (0.156)
#> --------------------------------------------------
#>   Log-likelihood   -526.545       -306.191
#>   AIC              1065.090        624.382
#>   BIC              1097.863        654.155
#>   N                1741           1056
#> ==================================================
#>   Significance: *** = p < 0.001; ** = p < 0.01;
#>                 * = p < 0.05
```

From previous output, it can be noticed that the coefficient of `articles` for men is approximately twice as large as for women: `0.074` vs `0.034`. One possible conclusion we could draw from this result is that women suffer from discrimination. That is, the return per additional article on the propensity to get a promotion is on average lower for women, holding other things constant. However, Allison (1999) notes that this result might be due to variance error term differences. For example, women might have more heterogeneous career patterns than men due to unobserved factors affecting promotion. In particular, assume that we have the following model for men (M) and women (W):

$$
\begin{aligned}
y^*_{iM} &= \mathbf{x}_{iM}^\top\boldsymbol{\beta} + \epsilon_{iM}, \\
y^*_{iW} &= \mathbf{x}_{iW}^\top\boldsymbol{\beta} + \epsilon_{iW}, \\
\epsilon_{iM} &\sim \Lambda(0, \sigma^2_M), \\
\epsilon_{iW} &\sim \Lambda(0, \sigma^2_W),
\end{aligned}
$$

where $\Lambda(\cdot)$ is the logistic CDF. Both men and women have the same coefficients, $\boldsymbol{\beta}$, in the propensity to be promoted, but different scales, $\sigma^2_M \neq \sigma^2_W$. Note that the logit model identifies $\beta = \frac{\alpha}{\sigma}$. Thus, if women have greater variance than men, $\sigma_W > \sigma_M$, their coefficient will be smaller, assuming similar return to productivity. To allow for such possibility, Williams (2009) suggests fitting a heteroskedastic logit (HET-Logit) model where the standard deviation of the error term is modeled as

$$
\sigma_i = \exp(\delta \cdot \texttt{female}_i).
$$

This model can be estimated in R using the `hetglm()` function from **glmx** package or `hetprob()` function from **Rchoice** package. The syntax to fit the model using `hetprob()` is the following

```
library("Rchoice")
het_logit <- hetprob(tenure ~ factor(female) + year + I(year^2) + select +
                       articles + prestige | factor(female),
                     data = sub_data,
                     link = "logit")
```

Similarly to `hetglm()` function, the `formula` argument of `hetprob()` has the form `y ~ x | z`, where `y` is the binary response variable, `x` are the explanatory covariates, and `z` are the covariates affecting the variance of the error term. The argument `link` indicates whether a logit (`link = "logit"`) or probit (`link = "probit"`) model should be fitted.

The output is the following:

```
summary(het_logit)
```

```
#> --------------------------------------------------------------------
#> Maximum Likelihood estimation of Heteroskedastic Binary model
#> Newton-Raphson maximisation, 4 iterations
#> Return code 8: successive function values within relative tolerance limit (reltol)
#> Log-Likelihood: -836.2824
#> 8  free parameters
#>
#> Estimates for the mean:
#>                  Estimate Std. error  z value   Pr(> z)
#> (Intercept)     -7.490505   0.659663 -11.3551 < 2.2e-16 ***
#> factor(female)1 -0.939190   0.370524  -2.5348 0.0112524 *
#> year             1.909544   0.199694   9.5624 < 2.2e-16 ***
#> I(year^2)       -0.139687   0.016943  -8.2448 < 2.2e-16 ***
#> select           0.181920   0.052657   3.4548 0.0005507 ***
#> articles         0.063534   0.010219   6.2173 5.059e-10 ***
#> prestige        -0.446207   0.096904  -4.6046 4.132e-06 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Estimates for lnsigma:
#>                     Estimate Std. error z value Pr(> z)
#> het.factor(female)1  0.30223    0.14618  2.0675 0.03868 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> LR test of lnsigma = 0: chi2 4.5 with 1 df. Prob > chi2 =  0.0339
#> --------------------------------------------------------------------
```

The results using `hetglm()` are the following

```
library("glmx")
het_glmx <- hetglm(tenure ~ factor(female) + year + I(year^2) + select +
                       articles + prestige | factor(female),
                     data = sub_data,
                    family = binomial(link = "logit"))
summary(het_glmx)
```

```
#>
#> Call:
#> hetglm(formula = tenure ~ factor(female) + year + I(year^2) + select +
#>     articles + prestige | factor(female), data = sub_data, family = binomial(link = "logit"))
#>
#> Deviance residuals:
#>     Min      1Q  Median      3Q     Max
#> -1.8473 -0.5666 -0.2926 -0.1149  3.3397
#>
#> Coefficients (binomial model with logit link):
#>                  Estimate Std. Error z value Pr(>|z|)
#> (Intercept)     -7.490489   0.648517 -11.550  < 2e-16 ***
#> factor(female)1 -0.939174   0.364357  -2.578 0.009948 **
#> year             1.909540   0.199095   9.591  < 2e-16 ***
#> I(year^2)       -0.139686   0.016762  -8.334  < 2e-16 ***
#> select           0.181919   0.051916   3.504 0.000458 ***
#> articles         0.063534   0.009884   6.428  1.3e-10 ***
#> prestige        -0.446207   0.097083  -4.596  4.3e-06 ***
#>
#> Latent scale model coefficients (with log link):
#>                 Estimate Std. Error z value Pr(>|z|)
#> factor(female)1   0.3022     0.1433   2.109   0.0349 *
#> ---
```

```
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Log-likelihood: -836.3 on 8 Df
#> LR test for homoscedasticity: 4.501 on 1 Df, p-value: 0.03387
#> Dispersion: 1
#> Number of iterations in nlminb optimization: 7
```

Although the coefficients estimated by both functions are very similar, their standard errors are somewhat different. One potential explanation for this difference is the optimization algorithm used by each function. `hetprob()` uses Newton-Raphson algorithm available in `maxLik()` function from **maxLik** package (Henningsen and Toomet 2011), whereas `hetglm()` uses `nlminb` algorithm as default.

Now, I compare the logit and Het-Logit estimates using `mtable()` function.[10] The following output presents the estimates.

```
mtable("Logit for men"   = logit_m,
       "Logit for women" = logit_w,
       "Heteroskedastic" = het_logit,
       summary.stats = c("Log-likelihood", "AIC", "BIC", "N"))

#>
#> Calls:
#> Logit for men: glm(formula = tenure ~ year + I(year^2) + select + articles +
#>     prestige, family = binomial(link = "logit"), data = sub_data,
#>     subset = (female == 0))
#> Logit for women: glm(formula = tenure ~ year + I(year^2) + select + articles +
#>     prestige, family = binomial(link = "logit"), data = sub_data,
#>     subset = (female == 1))
#> Heteroskedastic: hetprob(formula = tenure ~ factor(female) + year + I(year^2) +
#>     select + articles + prestige | factor(female), data = sub_data,
#>     link = "logit", method = "nr")
#>
#> =====================================================================
#>                    Logit for men  Logit for women    Heteroskedastic
#>                     -----------     -----------    -------------------
#>                       tenure          tenure          mean     lnsigma
#> ---------------------------------------------------------------------
#>   (Intercept)        -7.680***       -5.842***      -7.491***
#>                      (0.681)         (0.866)        (0.660)
#>   year                1.909***        1.408***       1.910***
#>                      (0.214)         (0.257)        (0.200)
#>   I(year^2)          -0.143***       -0.096***      -0.140***
#>                      (0.019)         (0.022)        (0.017)
#>   select              0.216***        0.055          0.182***
#>                      (0.061)         (0.072)        (0.053)
#>   articles            0.074***        0.034**        0.064***
#>                      (0.012)         (0.013)        (0.010)
#>   prestige           -0.431***       -0.371*        -0.446***
#>                      (0.109)         (0.156)        (0.097)
#>   factor(female)1                                   -0.939*    0.302*
#>                                                     (0.371)   (0.146)
#> ---------------------------------------------------------------------
#>   Log-likelihood   -526.545        -306.191       -836.282
#>   AIC              1065.090         624.382       1688.565
#>   BIC              1097.863         654.155       1736.055
#>   N                1741            1056           2797
#> =====================================================================
#>   Significance: *** = p < 0.001; ** = p < 0.01; * = p < 0.05
```

The estimated coefficients for the HET-Logit model indicate that being a woman increases the variance of the error term ($\widehat{\delta}$ = `0.302`) and decreases the propensity to be promoted ($\widehat{\beta}_6$ = `-0.939`).

Using the estimate $\widehat{\delta}$, we can also compute how much the disturbance standard deviation differ by gender. Note that the standard deviation of the error term for women is $\sigma_W = \exp(0.302)$, whereas for men is $\sigma_M = \exp(0) = 1$. Then,

[10] `mtable()` does not support objects of class `hetglm`.

```
sigma_w <- exp(coef(het_logit)["het.factor(female)1"])
(1 -  sigma_w) / sigma_w

#> het.factor(female)1
#>          -0.2608322
```

This result implies that the standard deviation of the disturbance for men is 26% lower than the standard deviation for women. Conversely, this also means that the standard deviation of the residuals is $\exp(0.302) = 1.35$ times larger for women compared to men (Williams 2009, 2010). The 95%-CI for this ratio can be computed using the delta method by `deltaMethod()` function from **car** package (Fox, Friendly, and Weisberg 2013):

```
library("car")
sharef <- "(1 - exp(`het.factor(female)1`)) / exp(`het.factor(female)1`)"
deltaMethod(het_logit, sharef)

#>                                                              Estimate       SE
#> (1 - exp(`het.factor(female)1`))/exp(`het.factor(female)1`) -0.26083  0.10805
#>                                                                 2.5 %  97.5 %
#> (1 - exp(`het.factor(female)1`))/exp(`het.factor(female)1`) -0.47261 -0.0491
```

So far, the HET-Logit estimates suggest that there are gender differences in both the dependent variable and in the variance of the error term. However, the estimated coefficients do not allow us to conclude whether women have a lower return than men for productivity. To give some insights about this question, I estimate a HET-Logit model including the interaction between `female` and `articles` in the choice equation:

```
het_logit2 <- hetprob(tenure ~ factor(female) + year + I(year^2) + select +
                        articles + prestige + factor(female)*articles |
                        factor(female),
                      data = sub_data,
                      link = "logit")
print(summary(het_logit2), digits = 3)

#> ------------------------------------------------------------------
#> Maximum Likelihood estimation of Heteroskedastic Binary model
#> Newton-Raphson maximisation, 4 iterations
#> Return code 1: gradient close to zero (gradtol)
#> Log-Likelihood: -835.1335
#> 9  free parameters
#>
#> Estimates for the mean:
#>                         Estimate Std. error z value Pr(> z)
#> (Intercept)              -7.3653     0.6547  -11.25 < 2e-16 ***
#> factor(female)1          -0.3781     0.4500   -0.84   0.401
#> year                      1.8383     0.2029    9.06 < 2e-16 ***
#> I(year^2)                -0.1343     0.0170   -7.89 3.1e-15 ***
#> select                    0.1700     0.0517    3.29   0.001 **
#> articles                  0.0720     0.0114    6.31 2.8e-10 ***
#> prestige                 -0.4205     0.0961   -4.37 1.2e-05 ***
#> factor(female)1:articles  -0.0305     0.0187   -1.63   0.104
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Estimates for lnsigma:
#>                    Estimate Std. error z value Pr(> z)
#> het.factor(female)1    0.177      0.163    1.09    0.28
#>
#> LR test of lnsigma = 0: chi2 1.22 with 1 df. Prob > chi2 =  0.2684
#> ------------------------------------------------------------------
```

The coefficient for `female * articles` is not statistically significant when residual variation by gender is involved. As argued by Allison (1999), this result proposes dissimilarities in productivity

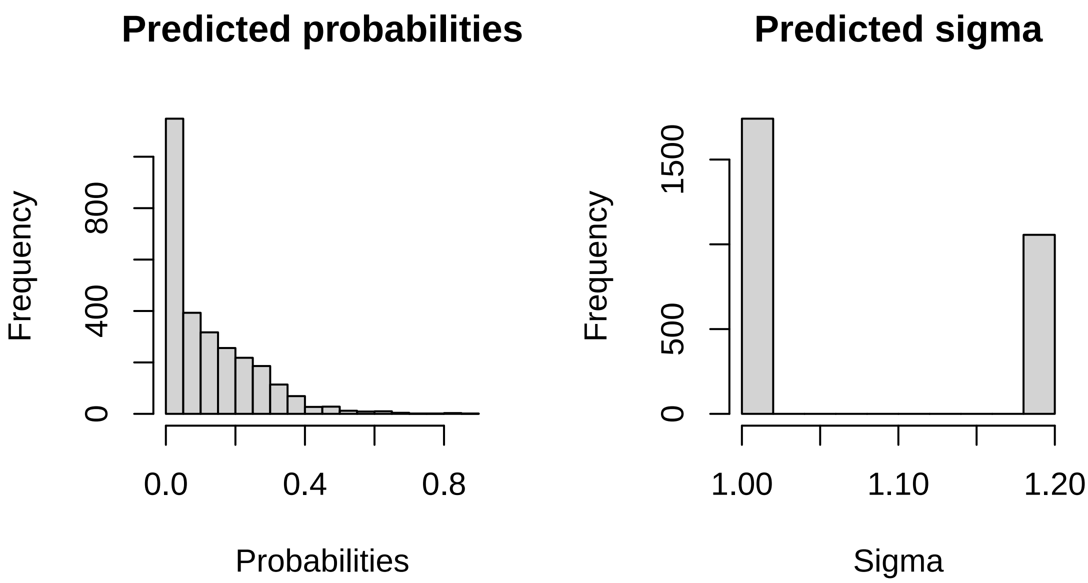

Figure 1: Distribution of predicted probability and predicted sigma

returns between males and females resulting from variability in unobserved factors rather than discriminatory influences.

Once we have fitted a model, we can use the `predict()` command to obtain the predicted probability and the predicted scale factor, $\hat{\sigma}_i$, which can be readily used for visualization as shown in Figure 1. The following lines plots the distribution of both measures:

```
par(mfrow = c(1, 2))
hist(predict(het_logit2, type = "pr"),
     main = "Predicted probabilities",
     xlab = "Probabilities")
hist(predict(het_logit2, type = "sigma"),
     main = "Predicted sigma",
     xlab = "Sigma")
```

An additional feature of **Rchoice** package is that it allows to estimate the APEs for heteroskedastic binary models, as in Equation (6), using `effect()` function. Similarly to command `margins()` from **margins** package (Leeper 2021) or `avg_slopes()` from **marginaleffects** package (Arel-Bundock 2023), this function takes into account whether the variables are continuous, categorical or both. The user must specify categorical variables using `factor()` in the `formula` argument; otherwise, the `effect()` function will assume that the variable is continuous, when the variable may already be a `factor` in the dataset. In the following lines, we compute the APEs for a HET-Probit and HET-Logit model.[11] The results are the following:

```
eff_logit  <- effect(het_logit2)
het_probit <- hetprob(tenure ~ factor(female) + year + I(year^2) + select +
                        articles + prestige + factor(female)*articles |
                        factor(female),
                      data = sub_data,
                      link = "probit")
eff_probit <- effect(het_probit)
mtable(eff_probit,
       eff_logit)
```

```
#>
#> Calls:
#> eff_probit: hetprob(formula = tenure ~ factor(female) + year + I(year^2) +
#>     select + articles + prestige + factor(female) * articles |
#>     factor(female), data = sub_data, link = "probit", method = "nr")
```

[11]The Jacobian matrix is computed numerically using `jacobian()` function from **numDeriv** package (Gilbert and Varadhan 2019).

```
#> eff_logit: hetprob(formula = tenure ~ factor(female) + year + I(year^2) +
#>     select + articles + prestige + factor(female) * articles |
#>     factor(female), data = sub_data, link = "logit", method = "nr")
#>
#> =============================================
#>                    eff_probit    eff_logit
#> ---------------------------------------------
#>   factor(female)1    -0.031**     -0.031**
#>                      (0.012)      (0.012)
#>   year                0.032***     0.032***
#>                      (0.002)      (0.002)
#>   select              0.014***     0.015***
#>                      (0.004)      (0.004)
#>   articles            0.006***     0.005***
#>                      (0.001)      (0.001)
#>   prestige           -0.035***    -0.036***
#>                      (0.008)      (0.008)
#> ---------------------------------------------
#>   Log-likelihood   -832.478     -835.133
#>   N                2797         2797
#> =============================================
#>   Significance: *** = p < 0.001;
#>                 ** = p < 0.01; * = p < 0.05
```

The APEs are very close to each other and statistically significant. According to the HET-Probit estimates, one additional published article increases the probability of being promoted by 0.6 percent points, whereas being a woman decreases the probability of promoted by 3.1%.

Labor participation

Our second example is a replication of Greene (2003)'s example 17.7 based on the dataset "`mroz.cvs`". This dataset is based on a cross-section data on the wages of 428 working, married women, originating from the 1976 Panel Study of Income Dynamics (PSID), which can be loaded as follows:

```
mroz <- read.csv(file = 'mroz.csv')
mroz$kids <- with(mroz, factor((kidslt6 + kidsge6) > 0,
                               levels = c(FALSE, TRUE),
                               labels = c("no", "yes")))
mroz$finc <-  mroz$faminc / 10000
```

Using this data, Greene (2003) estimates the following HET-Probit model for women labor participation:

$$\text{inlf}^* = \beta_0 + \beta_1\text{age} + \beta_2\text{age}^2 + \beta_3\text{finc} + \beta_4\text{educ} + \beta_5\text{kids} + \epsilon, \tag{22}$$

$$\epsilon \sim N(0, \sigma_i^2), \tag{23}$$

$$\sigma_i = \exp(\delta_1\text{kids} + \delta_2\text{finc}), \tag{24}$$

where `inlf` is a dummy variable indicating whether the woman participates in labor force, `age` is age in year, `finc` is family income in 1975 dollars divided by 10,000, `educ` is education in year and `kids` indicates whether children under 18 are present in the household. It is further assumed that `kids` and `finc` affect the variability of the error term.

The probit, Het-Probit and average marginal effects are estimated as follows:[12]

```
labor_hom <- glm(inlf ~  age + I(age^2) + finc + educ + factor(kids),
                 data = mroz,
                 family = binomial(link = "probit"))
labor_het <- hetprob(inlf ~  age + I(age^2) + finc + educ + factor(kids) |
                       factor(kids) + finc,
                     data = mroz,
                     link = "probit")
eff_labor_het <- effect(labor_het)
```

[12]Greene (2003) computes the marginal effects at the mean instead of the average marginal effects.

```
mtable(labor_hom,
       labor_het,
       eff_labor_het)
```

```
#>
#> Calls:
#> labor_hom: glm(formula = inlf ~ age + I(age^2) + finc + educ + factor(kids),
#>     family = binomial(link = "probit"), data = mroz)
#> labor_het: hetprob(formula = inlf ~ age + I(age^2) + finc + educ + factor(kids) |
#>     factor(kids) + finc, data = mroz, link = "probit", method = "nr")
#> eff_labor_het: hetprob(formula = inlf ~ age + I(age^2) + finc + educ + factor(kids) |
#>     factor(kids) + finc, data = mroz, link = "probit", method = "nr")
#>
#> ========================================================================
#>                          labor_hom       labor_het       eff_labor_het
#>                         -----------  ------------------   -----------
#>                            inlf         mean    lnsigma      inlf
#> ------------------------------------------------------------------------
#>   (Intercept)             -4.157**     -6.030*
#>                           (1.404)      (2.498)
#>   age                      0.185**      0.264*              -0.009***
#>                           (0.066)      (0.118)              (0.003)
#>   I(age^2)                -0.002**     -0.004*
#>                           (0.001)      (0.001)
#>   finc                     0.046        0.424    0.313*      0.069**
#>                           (0.043)      (0.222)  (0.123)     (0.024)
#>   educ                     0.098***     0.140**              0.030***
#>                           (0.023)      (0.052)              (0.009)
#>   factor(kids): yes/no    -0.449***    -0.879** -0.141      -0.161***
#>                           (0.130)      (0.303)  (0.324)     (0.043)
#> ------------------------------------------------------------------------
#>   Log-likelihood        -490.848     -487.636             -487.636
#>   N                      753          753                  753
#> ========================================================================
#>   Significance: *** = p < 0.001; ** = p < 0.01; * = p < 0.05
```

The results show that family income does not play any role in the choice equation. However, it increases the variability of the error term. APE indicates that an increase of $10,000 of family income increases the probability of labor force involvement by 6.9%. There is not enough statistical evidence that proves having children under 18 in the household produces heteroskedasticity.

We can also use the Wald test provided by `linearHypothesis()` function from car package to test the null hypothesis of homoskedasticity:

```
coefs <- names(coef(labor_het))
linearHypothesis(labor_het, coefs[grep("het", coefs)])
```

```
#> Linear hypothesis test
#>
#> Hypothesis:
#> het.factor(kids)yes = 0
#> het.finc = 0
#>
#> Model 1: restricted model
#> Model 2: inlf ~ age + I(age^2) + finc + educ + factor(kids) | factor(kids) +
#>     finc
#>
#>   Df  Chisq Pr(>Chisq)
#> 1
#> 2  2 6.5331    0.03814 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis of homoskedasticity is rejected at the 5% with a $\chi^2_2 = 6.533$.

Supplementary materials provide Stata code (version 16.1) to replicate all the results in this Section. The log file is presented in **Appendix C**. Overall, the results using Stata are exactly the same to those reported by `hetprob()` function from **Rchoice** package.

3.2 Instrumental variable probit model

Control function approach

In this example, and similar to Wooldridge (2010), we use the `mroz` sample and assume the following slightly modified model for married women's labor force participation from previous Section:

$$
\begin{aligned}
\text{inlf}^* =& \beta_0 + \beta_1\text{educ} + \beta_2\text{exper} + \beta_3\text{exper}^2 + \beta_4\text{age} + \beta_5\text{kidslt6} + \\
& \beta_6\text{kidsge6} + \beta_7\text{nwifeinc} + \epsilon, \\
\text{nwifeinc} =& \delta_0 + \delta_1\text{educ} + \delta_2\text{exper} + \delta_3\text{exper}^2 + \delta_4\text{age} + \delta_5\text{kidslt6} + \\
& \delta_6\text{kidsge6} + \delta_7\text{huseduc} + \upsilon, \\
\text{lfp} =& \mathbf{1}\left[\text{lfp}^* > 0\right]
\end{aligned}
$$

where `nwifeinc` is the other sources of income (divided by 1,000) and assumed to be endogenous. A just identified IV model is estimated by using husband's education, (`huseduc`), as an instrument for `nwifeinc`. The strong identification assumption here is that husband's schooling is unrelated to factors that affect a married woman's labor force decision once `nwifeinc` and the other variables are accounted for (Wooldridge 2010).

When interpreting the results from an IV model, it is important to compare its magnitude with a model that assumes exogeneity. In this example, our benchmark APE for `nwifeinc` is obtained by the standard probit model:

```
probit <- glm(inlf ~  educ + exper + I(exper^2) + age + kidslt6 + kidsge6 + nwifeinc,
              data = mroz,
              family = binomial(link = "probit"))
ape.probit <- mean(dnorm(predict(probit, type = "link"))) * coef(probit)["nwifeinc"]
ape.probit

#>     nwifeinc
#> -0.003616176
```

Accordingly, an increase of $1,000 in other sources of income reduces the labor force participation probability by 0.4%, holding all other factors constant. Note that the same APE, along with its standard error, can also be obtained using `avg_slopes()` command:

```
library("marginaleffects")
avg_slopes(probit, variables = "nwifeinc")

#>
#>      Term Estimate Std. Error     z Pr(>|z|)   S    2.5 %    97.5 %
#>  nwifeinc -0.00362    0.00147 -2.46   0.0139 6.2 -0.0065 -0.000736
#>
#> Columns: term, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high
#> Type:  response
```

I proceed to estimate the model using the CF approach. First, I estimate the first-step equation, which is a linear model, and obtain the residuals $\widetilde{\upsilon}$:

```
fstep   <- lm(nwifeinc ~ educ + exper + I(exper^2) + age + kidslt6 + kidsge6  + huseduc,
            data = mroz)
mroz$res.hat <- fstep$residuals
```

We can also test the power of the instrument using `linearHypothesis()` function:

```
linearHypothesis(fstep, "huseduc")
```

```
#> Linear hypothesis test
#>
#> Hypothesis:
#> huseduc = 0
#>
#> Model 1: restricted model
#> Model 2: nwifeinc ~ educ + exper + I(exper^2) + age + kidslt6 + kidsge6 +
#>     huseduc
#>
#>   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
#> 1    746 86955
#> 2    745 81120  1    5834.8 53.586 6.427e-13 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The first-stage F statistic on huseduc is substantially above the traditional cut-off of ten suggesting that the instrument is not weak.

The second-step is computed using `glm()` function and adding the residuals (`res.hat`) as an additional explanatory variable:

```
sstep  <- glm(inlf ~  educ + exper + I(exper^2) + age + kidslt6 + kidsge6 + nwifeinc + res.hat,
            data   = mroz,
            family = binomial(link = "probit"))
summary(sstep)
```

```
#>
#> Call:
#> glm(formula = inlf ~ educ + exper + I(exper^2) + age + kidslt6 +
#>     kidsge6 + nwifeinc + res.hat, family = binomial(link = "probit"),
#>     data = mroz)
#>
#> Deviance Residuals:
#>     Min       1Q   Median       3Q      Max
#> -2.2523  -0.9078   0.4204   0.8566   2.2803
#>
#> Coefficients:
#>               Estimate Std. Error z value Pr(>|z|)
#> (Intercept)  0.0171183  0.5380339   0.032  0.97462
#> educ         0.1702142  0.0377615   4.508 6.56e-06 ***
#> exper        0.1163118  0.0193869   6.000 1.98e-09 ***
#> I(exper^2)  -0.0019458  0.0005999  -3.244  0.00118 **
#> age         -0.0449529  0.0101351  -4.435 9.19e-06 ***
#> kidslt6     -0.8444319  0.1197268  -7.053 1.75e-12 ***
#> kidsge6      0.0477912  0.0449431   1.063  0.28761
#> nwifeinc    -0.0368639  0.0183848  -2.005  0.04495 *
#> res.hat      0.0267092  0.0191539   1.394  0.16318
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for binomial family taken to be 1)
#>
#>     Null deviance: 1029.75  on 752  degrees of freedom
#> Residual deviance:  800.61  on 744  degrees of freedom
#> AIC: 818.61
#>
#> Number of Fisher Scoring iterations: 4
```

Since the z-statistic for `res.hat` is 1.4, we cannot reject the null hypothesis that `nwifeinc` is exogenous: $H_0 : \lambda = 0$.

An estimate of ρ can be obtained using Equation (13) and the following syntax:

```
lambda.hat    <- coef(sstep)["res.hat"]
k             <- length(fstep$coefficients)
```

```
SSE           <- sum(fstep$residuals^2)
n             <- length(fstep$residuals)
sigma.upsilon <- sqrt(SSE/(n - k))
rho.hat       <- lambda.hat * sigma.upsilon
rho.hat

#>   res.hat
#> 0.2787068
```

Thus, the estimated correlation using the CF approach is $\widehat{\rho} = 0.279$. It is important to recall that the estimated coefficients for the sstep model represent the coefficients scaled by a factor of $1/\sqrt{1-\rho^2}$. Moreover, the standard errors from the sstep model are biased since they do not consider the sampling error of the first stage. However, we can use ivprobit() function from **ivprobit** package (Zaghdoudi 2018) to get the correct standard errors:[13]

```
library("ivprobit")
twostep.probit <- ivprobit(inlf ~  educ + exper + I(exper^2) + age + kidslt6 + kidsge6 |
                             nwifeinc | educ + exper + I(exper^2) + age + kidslt6 +
                             kidsge6 + huseduc,
                           data = mroz)
summary(twostep.probit)

#>                    Coef        S.E.  t-stat     p-val
#> Intercep    0.01711834  0.54865782  0.0312  0.975118
#> educ        0.17021419  0.03848938  4.4224 1.121e-05 ***
#> exper       0.11631183  0.01976301  5.8853 6.001e-09 ***
#> I(exper^2) -0.00194584  0.00061195 -3.1798  0.001535 **
#> age        -0.04495285  0.01032548 -4.3536 1.526e-05 ***
#> kidslt6    -0.84443188  0.12176581 -6.9349 8.818e-12 ***
#> kidsge6     0.04779117  0.04578807  1.0437  0.296940
#> nwifeinc   -0.03686390  0.01874338 -1.9668  0.049580 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimates of the sstep and twostep.probit models are the same, while their standard errors are slightly different.

The APE for nwifeinc—and any other continuous variable—can be computed using Equation (16) and its standard error via bootstrap method. Below, I use package **boot** (Canty 2002) to perform the simulation. First, a function called ape() is created which returns the APE. The first argument of this function is the dataset, whereas the second argument can be an index vector of the observations in the dataset.

```
ape <- function(data, indices){
  d <- data[indices, ]
  # Compute the first-stage regression
  fstep    <- lm(nwifeinc ~ educ + exper + I(exper^2) + age + kidslt6 + kidsge6 +
                   huseduc,
                 data = d)
  # Obtain the residuals
  d$res.hat <- fstep$residuals
  # Compute the second-stage regression
  sstep   <- glm(inlf ~  educ + exper + I(exper^2) + age + kidslt6 + kidsge6 +
                   nwifeinc + res.hat,
                 data   = d,
                 family = binomial(link = "probit"))
  # Compute APE for nwincome
  out <- mean(dnorm(predict(sstep, type = "link"))) * coef(sstep)["nwifeinc"]
  return(out)
}
```

Once we have defined the function ape(), we can use the boot() function to perform the bootstrap procedure. In the following syntax, R = 500 resamplings are used and the 90%-CI interval is obtained using boot.ci() function.

[13] ivprobit() uses a minimum chi-squared estimator (Newey 1987).

```
library("boot")
set.seed(666)
results <- boot(data = mroz, statistic = ape, R = 500)
results
```

```
#>
#> ORDINARY NONPARAMETRIC BOOTSTRAP
#>
#>
#> Call:
#> boot(data = mroz, statistic = ape, R = 500)
#>
#>
#> Bootstrap Statistics :
#>       original        bias    std. error
#> t1* -0.0110576 -0.0005050597 0.005877061
```

```
boot.ci(results, type = "norm", conf = 0.90)
```

```
#> BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
#> Based on 500 bootstrap replicates
#>
#> CALL :
#> boot.ci(boot.out = results, conf = 0.9, type = "norm")
#>
#> Intervals :
#> Level      Normal
#> 90%   (-0.0202, -0.0009 )
#> Calculations and Intervals on Original Scale
```

The results show that another $1,000 in other sources of income reduces the labor force participation probability by 1.1 percent points with 90%-CI $[-2, -.09]$. This estimate, which is marginally statistically significant, is about three times larger than the probit estimate that treats `nwifeinc` as exogenous: `-0.04`.

Finally, we can recover the unscaled parameters by multiplying the coefficients by $\sqrt{(1-\hat{\rho}^2)}$ as follows:

```
coef(sstep) * sqrt(1 - rho.hat^2)
```

```
#>  (Intercept)         educ        exper   I(exper^2)          age      kidslt6
#>  0.016440046  0.163469667  0.111703116 -0.001868741 -0.043171653 -0.810972324
#>      kidsge6     nwifeinc      res.hat
#>  0.045897507 -0.035403215  0.025650873
```

Maximum likelihood estimator

In this Section I estimate the model from previous Section using the MLE. To do so, I use the `ivpml()` function from **Rchoice** package. The syntax is as follows:

```
fiml.probit <- ivpml(inlf ~  educ + exper + I(exper^2) + age + kidslt6 + kidsge6 +
                       nwifeinc | huseduc +  educ + exper + I(exper^2) + age +
                       kidslt6 + kidsge6,
                     data = mroz)
```

```
#>
#> Estimating a just identified model....
#>
#> Obtaining starting values from probit and linear model...
```

The syntax of `ivpml()` is similar to that of `ivreg()` function from **AER** package. The `formula` has two part in the right-hand side, that is, `y ~ x | z` where `y` is the binary response variable, `x` are the regressors ($\mathbf{x}$ in Equation (7)), and `z` are the exogenous variables ($\mathbf{x}_1$ and $\mathbf{x}_2$ in Equation (8)).

During the optimization procedure, `ivpml()` displays several messages which can be turned-off by setting `messages = FALSE`. The output indicates that the model is just-identified and that the initial values for the optimization procedure are obtained from the traditional probit and linear models for the structural and first-stage equation, respectively. Similarly to `hetprob()` function, the optimization algorithm can be managed using the argument `method`, which is passed on to the `maxLik()` function. Currently, the default algorithm is the Newton-Raphson, `method = "nr"`.

```
summary(fiml.probit)
```

```
#> --------------------------------------------
#> Maximum Likelihood estimation of IV Probit model
#> Newton-Raphson maximisation, 3 iterations
#> Return code 8: successive function values within relative tolerance limit (reltol)
#> Log-Likelihood: -3230.642
#> 18  free parameters
#> Estimates:
#>                         Estimate  Std. error z value   Pr(> z)
#> inlf:(Intercept)      1.6499e-02  5.3008e-01  0.0311 0.9751702
#> inlf:educ             1.6403e-01  3.1225e-02  5.2531 1.495e-07 ***
#> inlf:exper            1.1209e-01  2.1199e-02  5.2873 1.241e-07 ***
#> inlf:I(exper^2)      -1.8751e-03  5.9150e-04 -3.1701 0.0015237 **
#> inlf:age             -4.3319e-02  1.1331e-02 -3.8230 0.0001319 ***
#> inlf:kidslt6         -8.1375e-01  1.2994e-01 -6.2623 3.794e-10 ***
#> inlf:kidsge6          4.6054e-02  4.3139e-02  1.0676 0.2857141
#> inlf:nwifeinc        -3.5524e-02  1.6190e-02 -2.1941 0.0282247 *
#> nwifeinc:(Intercept) -1.4720e+01  3.7672e+00 -3.9076 9.322e-05 ***
#> nwifeinc:huseduc      1.1782e+00  1.6009e-01  7.3594 1.847e-13 ***
#> nwifeinc:educ         6.7469e-01  2.1254e-01  3.1744 0.0015016 **
#> nwifeinc:exper       -3.1299e-01  1.3752e-01 -2.2760 0.0228480 *
#> nwifeinc:I(exper^2)  -4.7756e-04  4.4955e-03 -0.1062 0.9153983
#> nwifeinc:age          3.4015e-01  5.9390e-02  5.7274 1.020e-08 ***
#> nwifeinc:kidslt6      8.2627e-01  8.1402e-01  1.0151 0.3100812
#> nwifeinc:kidsge6      4.3553e-01  3.2027e-01  1.3599 0.1738728
#> lnsigma               2.3398e+00  2.5768e-02 90.8016 < 2.2e-16 ***
#> atanhrho              2.7379e-01  1.9296e-01  1.4189 0.1559361
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Instrumented: nwifeinc
#> Instruments: (Intercept) huseduc educ exper I(exper^2) age kidslt6 kidsge6
#> Wald test of exogeneity (corr = 0): chi2 2.01 with 1 df. Prob > chi2 =  0.1559
#> --------------------------------------------
```

During the optimization procedure the parameters σ_v and ρ might tend to the boundary points of the parameter space, generating identifiability problems of the MLE. To avoid this issue, `ivpml()` re-parametrizes the parameters.[14] First, to ensure $\sigma_v > 0$, `ivpml()` instead estimates $\log \nu_v$ such that:

$$\sigma_v = \exp(\log \nu_v). \tag{25}$$

Second, `ivpml()` forces the correlation to remain in the $(-1, +1)$ range by using the inverse hyperbolic tangent:

$$\text{atanh}(\rho) = \tau = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right),$$

where τ is unrestricted, and ρ can be obtained using the inverse of τ:

$$\tau^{-1} = \rho = \tanh(\tau). \tag{26}$$

In the following syntax, we recover σ_v and ρ using Equations (25) and (26), respectively, and their standard errors are computed using delta method approach by `deltaMethod()` function:

```
deltaMethod(fiml.probit, "exp(lnsigma)")
```

[14] This re-parametrization is also used by `ivprobit` function in Stata.

```
#>              Estimate       SE    2.5 % 97.5 %
#> exp(lnsigma) 10.37928  0.26746  9.85508 10.903
```

```
deltaMethod(fiml.probit, "tanh(atanhrho)")
```

```
#>                 Estimate        SE     2.5 % 97.5 %
#> tanh(atanhrho)  0.267145  0.179190 -0.084061 0.6184
```

Again, the FIML estimate of ρ is close to that found using the CF approach which was `0.279`. If significant, a positive ρ would indicate that there is a positive correlation between ϵ and υ. That is, the unobserved factors that make it more likely for a woman to have a higher income from other sources also make it more likely that the woman will be participating in the labor force.

For those users who are more familiar with Stata (see **Appendix D**), it is important to mention that its function `ivprobit` estimates the 95%-CI for $\widehat{\rho}$ and $\widehat{\sigma}_{\upsilon}$ as follows:

```
cbind(exp(coef(fiml.probit)["lnsigma"] - qnorm(0.975) * stdEr(fiml.probit)["lnsigma"]),
      exp(coef(fiml.probit)["lnsigma"] + qnorm(0.975) * stdEr(fiml.probit)["lnsigma"]))
```

```
#>             [,1]     [,2]
#> lnsigma 9.868094 10.91695
```

```
cbind(tanh(coef(fiml.probit)["atanhrho"] - qnorm(0.975) * stdEr(fiml.probit)["atanhrho"]),
    tanh(coef(fiml.probit)["atanhrho"] + qnorm(0.975) * stdEr(fiml.probit)["atanhrho"]))
```

```
#>                [,1]      [,2]
#> atanhrho -0.1040317 0.5730038
```

The APEs can be estimated using the function `effect()`. The main argument of this function is `asf`. If `asf = TRUE` (the default), then the APEs are computed using Equation (20). On the other hand, if `asf = FALSE` the APEs are computed using Equation (21).

```
summary(effect(fiml.probit))
```

```
#> ------------------------------------------------------
#> Marginal effects for the IV Probit model:
#> ------------------------------------------------------
#>               dydx Std. error z value  Pr(> z)
#> educ      0.051057   0.011101   4.599 4.24e-06 ***
#> exper     0.023071   0.002952   7.816 5.44e-15 ***
#> age      -0.013484   0.002986  -4.516 6.31e-06 ***
#> kidslt6  -0.253295   0.033077  -7.658 1.89e-14 ***
#> kidsge6   0.014335   0.013520   1.060   0.2890
#> nwifeinc -0.011058   0.005550  -1.992   0.0463 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Note: Marginal effects computed as the average for each individual
```

```
summary(effect(fiml.probit, asf = FALSE))
```

```
#> ------------------------------------------------------
#> Marginal effects for the IV Probit model:
#> ------------------------------------------------------
#>               dydx Std. error z value  Pr(> z)
#> educ      0.048777   0.008733   5.585 2.33e-08 ***
#> exper     0.021997   0.003723   5.908 3.46e-09 ***
#> age      -0.012882   0.003322  -3.878 0.000105 ***
#> kidslt6  -0.241982   0.036594  -6.613 3.78e-11 ***
#> kidsge6   0.013695   0.012792   1.071 0.284373
#> nwifeinc -0.010564   0.004736  -2.230 0.025724 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Note: Marginal effects computed as the average for each individual
```

The results show that both APEs are close to each other. Note also that the estimated APE for `nwifeinc` using the CF approach is very similar to that ones obtained by MLE. **Appendix D** also shows that the Stata function `ivprobit()` provides the same estimates and marginal effects as `ivpml()` function.

4 Summary

The aim of the article was to provide a primer on estimating heteroskedastic and IV model for binary outcomes in R. I also show that the current version of **Rchoice** package (available at https://cran.r-project.org/web/packages/Rchoice/index.html) allows to estimate such models in a flexible way and provides accurate average marginal effects that are very similar to those provided by Stata's `margins` command. **Rchoice** can be used in concert with other packages. For example, one can format the summary output from **Rchoice** with **memisc** to produce well-formatted tables for regression estimates

5 Appendix A: Gradient and Hessian for binary response models with heteroskedasticity

In this section, I provide the analytic gradient and Hessian used by `hetprob()` function in **Rchoice**. The log-likelihood function for the binary choice model with exponential heteroskedasticity can be written as:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln F(a_i),$$

where $F(\cdot)$ is either the CDF of the standard normal or standard logistic distribution, $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top\right)^\top$ is the full $(k+p)$-dimensional vector of parameters, and:

$$\begin{aligned} a_i &= q_i \left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\exp(\mathbf{z}_i^\top \boldsymbol{\delta})}\right), \\ q_i &= 2(y_i - 1). \end{aligned}$$

The gradient is:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^{n} \left[\frac{f(a_i)}{F(a_i)} \frac{\partial a_i}{\partial \boldsymbol{\theta}}\right], \\ &= \sum_{i=1}^{n} \left[m(a_i)\mathbf{g}_i\right], \end{aligned}$$

where $m(\cdot) = f(\cdot)/F(\cdot) = \phi(\cdot)/\Phi(\cdot)$ for the probit model and $m(\cdot) = 1 - \Lambda(\cdot)$ for the logit model, and $\partial a_i / \partial \boldsymbol{\theta} = \mathbf{g}_i$ such that:

$$\mathbf{g}_i = \begin{pmatrix} \frac{\partial a_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial a_i}{\partial \boldsymbol{\delta}} \end{pmatrix} = \frac{q_i}{\exp(\mathbf{z}_i^\top \boldsymbol{\delta})} \begin{pmatrix} \mathbf{x}_i \\ -\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right) \mathbf{z}_i \end{pmatrix}.$$

The Hessian is given by:

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{\partial}{\partial \boldsymbol{\theta}^\top} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right), \\ &= \sum_{i=1}^{n} \left[h(a_i)\mathbf{g}_i\mathbf{g}_i^\top + m(a_i)\mathbf{H}_i\right], \end{aligned}$$

where $h(a_i) = \partial m(a_i)/\partial a_i = -a_i m(a_i) - m(a_i)^2$ and:

$$\mathbf{H}_i = \frac{\partial a_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \frac{\partial a_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\partial a_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\delta}^\top} \\ \frac{\partial a_i}{\partial \boldsymbol{\delta} \partial \boldsymbol{\beta}^\top} & \frac{\partial a_i}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^\top} \end{pmatrix} = \begin{pmatrix} \mathbf{O} & -\frac{q_i}{\exp(\mathbf{z}_i^\top \boldsymbol{\delta})} \mathbf{x}_i \mathbf{z}_i^\top \\ -\frac{q_i}{\exp(\mathbf{z}_i^\top \boldsymbol{\delta})} \mathbf{z}_i \mathbf{x}_i^\top & \frac{q_i(\mathbf{x}_i^\top \boldsymbol{\beta})}{\exp(\mathbf{z}_i^\top \boldsymbol{\delta})} \mathbf{z}_i \mathbf{z}_i^\top \end{pmatrix}.$$

6 Appendix B: Gradient and Hessian for binary response models with endogeneity

In this section, I provide the analytic gradient and Hessian used by `ivpml` function in **Rchoice**. The log-likelihood function can be written as:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\ln\left(\Phi(a_i)\right) + \ln(1) - \ln\left(\sigma_v\right) + \ln\left[\phi(b_i)\right]\right],$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \delta^\top, \sigma_v, \rho\right)^\top$ is an $(k+p+2)$-dimensional vector and:

$$\begin{aligned}
a_i &= q_i \left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta} + \frac{\rho}{\sigma_v}\left(y_{2i} - \mathbf{z}_i^\top \delta\right)}{\sqrt{1-\rho^2}}\right), \\
b_i &= \frac{y_{2i} - \mathbf{z}_i^\top \delta}{\sigma_v}, \\
q_i &= 2(y_i - 1), \\
\sigma_v &= \exp(\ln \nu_v), \\
\rho &= \tanh(\tau).
\end{aligned}$$

The first derivatives of the log-likelihood function are:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n}\left[m(a_i)\left(\frac{q_i}{\sqrt{1-\rho^2}}\right)\mathbf{x}_i\right], \\
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \delta} &= \sum_{i=1}^{n}\left[-m(a_i)\left(\frac{q_i\left(\rho/\sigma_v\right)}{\sqrt{1-\rho^2}}\right) + b_i\left(\frac{1}{\sigma_v}\right)\right]\mathbf{z}_i, \\
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \ln \nu_v} &= \sum_{i=1}^{n}\left[-m(a_i)\frac{q_i\rho}{\sqrt{1-\rho^2}}b_i + b_i^2 - 1\right], \\
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \tau} &= \sum_{i=1}^{n}\left[m(a_i)q_i\left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta}\rho + b_i}{\sqrt{\operatorname{sech}^2(\tau)}}\right)\right],
\end{aligned}$$

where $m(a_i) = \phi(a_i)/\Phi(a_i)$, $d\tanh(\tau)/d\tau = \operatorname{sech}^2(\tau)$, and we use the fact that $\phi'(b_i) = -b_i\phi(b_i)$ so that $\phi'(b_i)/\phi(b_i) = -b_i$.

The Hessian is given by:

$$\mathbf{H} = \begin{pmatrix}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \delta^\top} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \ln \nu_v} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \tau} \\
. & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \delta\partial \delta^\top} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \delta\partial \partial \ln \nu_v} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \delta\partial \tau} \\
. & . & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\ln \nu_v)^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \ln \nu_v \partial \tau} \\
. & . & . & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \tau^2}
\end{pmatrix}.$$

The second derivatives are:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \sum_{i=1}^n \left[h(a_i) \left(\frac{q_i}{\sqrt{1-\rho^2}} \right)^2 \mathbf{x}_i \mathbf{x}_i^\top \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\delta}^\top} = \sum_{i=1}^n \left[-h(a_i) \left(\frac{q_i}{\sqrt{1-\rho^2}} \right)^2 \left(\frac{\rho}{\sigma_v} \right) \mathbf{x}_i \mathbf{z}_i^\top \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \ln \nu_v} = \sum_{i=1}^n \left[-h(a_i) \left(\frac{q_i}{\sqrt{1-\rho^2}} \right)^2 (\rho b_i)\, \mathbf{x}_i \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \tau} = \sum_{i=1}^n \left[h(a_i) \left(\frac{q_i}{\sqrt{1-\rho^2}} \right) \left(q_i \left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta} \rho + b_i}{\sqrt{\operatorname{sech}^2(\tau)}} \right) \right) \mathbf{x}_i \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^\top} = \sum_{i=1}^n \left[h(a_i) \left(\frac{q_i\,(\rho/\sigma_v)}{\sqrt{1-\rho^2}} \right)^2 - \frac{1}{\sigma_v^2} \right] \mathbf{z}_i \mathbf{z}_i^\top,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\delta} \partial \ln \nu_v} = \sum_{i=1}^n \left(\frac{b_i}{\sigma_v} \right) \left[h(a_i) \left(\frac{q_i \rho}{\sqrt{1-\rho^2}} \right)^2 - 2 \right] \mathbf{z}_i,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\delta} \partial \tau} = \sum_{i=1}^n \left[-h(a_i) \left(\frac{q_i\,(\rho/\sigma_v)}{\sqrt{1-\rho^2}} \right) \left(q_i \left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta} \rho + b_i}{\sqrt{\operatorname{sech}^2(\tau)}} \right) \right) \right] \mathbf{z}_i,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\ln \nu_v)^2} = \sum_{i=1}^n \left[h(a_i) \left(\frac{q_i \rho b_i}{\sqrt{1-\rho^2}} \right)^2 + m(a_i) \left(\frac{q_i \rho b_i}{\sqrt{1-\rho^2}} \right) - 2b_i^2 \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \ln \nu_v \partial \tau} = \sum_{i=1}^n \left\{ -b_i \left[h(a_i) \frac{q_i \rho}{\sqrt{1-\rho^2}} q_i \left(\frac{\mathbf{x}_i^\top \boldsymbol{\beta} \rho + b_i}{\sqrt{\operatorname{sech}^2(\tau)}} \right) + m(a_i) \frac{q_i}{\sqrt{\operatorname{sech}^2(\tau)}} \right] \right\},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \tau^2} = \sum_{i=1}^n \left[h(a_i) \left(q_i \frac{\mathbf{x}_i^\top \boldsymbol{\beta} \rho + b_i}{\sqrt{\operatorname{sech}^2(\tau)}} \right)^2 + q_i m(a_i) \frac{\mathbf{x}_i^\top \boldsymbol{\beta} + b_i \rho}{\sqrt{\operatorname{sech}^2(\tau)}} \right],$$

where $h(a_i) = -a_i m(a_i) - m(a_i)^2$.

7 Appendix C: Stata code for heteroskedastic binary response models

```
 *==========================================
 *** Example 1: Promotion of scientists ***
 *==========================================
. import delimited "$dir/tenure.csv", clear
(23 vars, 2,945 obs)

. ** Logit models for men and women
. quietly eststo logit_m: logit tenure year c.year#c.year select ///
                          articles prestige if (year <= 10 & female == 0)

. quietly eststo logit_w: logit tenure year c.year#c.year select ///
                          articles prestige if (year <= 10 & female == 1)

. esttab logit_m logit_w, b(3) se(3)
------------------------------------------------
                      (1)             (2)
                   tenure          tenure
------------------------------------------------
tenure
year                1.909***        1.408***
                  (0.214)         (0.257)
c.year#c.y~r       -0.143***       -0.096***
                  (0.019)         (0.022)
```

```
select             0.216***        0.055
                  (0.061)         (0.072)
articles           0.074***        0.034**
                  (0.012)         (0.013)
prestige          -0.431***       -0.371*
                  (0.109)         (0.156)
_cons             -7.680***       -5.842***
                  (0.681)         (0.866)
--------------------------------------------
N                   1741            1056
--------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. ** Heterokedastic logit model
. quietly ssc install oglm

. quietly eststo het_logit: oglm tenure i.female year c.year#c.year select ///
                         articles prestige if (year <= 10), hetero(i.female) link(logit)

. esttab logit_m logit_w het_logit, b(3) se(3)
------------------------------------------------------------
                     (1)             (2)             (3)
                  tenure          tenure          tenure
------------------------------------------------------------
tenure
year               1.909***        1.408***        1.910***
                  (0.214)         (0.257)         (0.200)
c.year#c.y~r      -0.143***       -0.096***       -0.140***
                  (0.019)         (0.022)         (0.017)
select             0.216***        0.055           0.182***
                  (0.061)         (0.072)         (0.053)
articles           0.074***        0.034**         0.064***
                  (0.012)         (0.013)         (0.010)
prestige          -0.431***       -0.371*         -0.446***
                  (0.109)         (0.156)         (0.097)
0.female                                           0.000
                                                     (.)
1.female                                          -0.939*
                                                  (0.371)
_cons             -7.680***       -5.842***
                  (0.681)         (0.866)
------------------------------------------------------------
lnsigma
0.female                                           0.000
                                                     (.)
1.female                                           0.302*
                                                  (0.146)
------------------------------------------------------------
cut1
_cons                                              7.491***
                                                  (0.660)
------------------------------------------------------------
N                   1741            1056            2797
------------------------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. ** Testing how much the disturbance standard deviation differ by gender
. margins, expression((1 - exp([lnsigma]_b[1.female])) / exp([lnsigma]_b[1.female]))
Warning: expression() does not contain predict() or xb().
Warning: prediction constant over observations.

Predictive margins                              Number of obs     =      2,797
```

```
Model VCE    : OIM
Expression   : (1 - exp([lnsigma]_b[1.female])) / exp([lnsigma]_b[1.female])

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -.2608323   .1080501    -2.41   0.016    -.4726065   -.0490581
------------------------------------------------------------------------------

. ** Heterokedastic logit model 2
. eststo het_logit2: oglm tenure i.female year c.year#c.year select ///
        articles prestige i.female#c.articles if (year <= 10), hetero(i.female) link(logit)

Heteroskedastic Ordered Logistic Regression     Number of obs     =      2,797
                                                LR chi2(8)        =     415.39
                                                Prob > chi2       =     0.0000
Log likelihood = -835.13347                     Pseudo R2         =     0.1992

-----------------------------------------------------------------------------------
           tenure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------+----------------------------------------------------------------
tenure            |
         1.female |  -.3780598   .4500207    -0.84   0.401    -1.260084    .5039645
             year |   1.838257   .2029491     9.06   0.000     1.440484     2.23603
                  |
    c.year#c.year |  -.1342828    .017024    -7.89   0.000    -.1676492   -.1009165
                  |
           select |   .1699659   .0516643     3.29   0.001     .0687057    .2712261
         articles |   .0719821   .0114106     6.31   0.000     .0496178    .0943464
         prestige |  -.4204742   .0961206    -4.37   0.000     -.608867   -.2320813
                  |
female#c.articles |
                1 |  -.0304836   .0187427    -1.63   0.104    -.0672185    .0062514
------------------+----------------------------------------------------------------
lnsigma           |
         1.female |   .1774193   .1627087     1.09   0.276    -.1414839    .4963226
------------------+----------------------------------------------------------------
            /cut1 |   7.365286   .6547121    11.25   0.000     6.082074    8.648498
-----------------------------------------------------------------------------------

. ** Plot predicted probability and sigma
. predict phat, pr outcome(1)
. predict sigmahat, sigma

. hist phat
(bin=34, start=2.232e-12, width=.02503351)

. hist sigmahat
(bin=34, start=1, width=.00570976)

. ** Average Marginal Effects for logit and probit heterokedastic models
. quietly oglm tenure i.female year c.year#c.year select ///
           articles prestige i.female#c.articles if (year <= 10), hetero(i.female) link(probit)

. eststo eff_probit: margins, dydx(*) predict(outcome(1)) post
Average marginal effects                        Number of obs     =      2,797
Model VCE    : OIM
Expression   : Pr(tenure==1), predict(outcome(1))
dy/dx w.r.t. : 1.female year select articles prestige
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
    1.female |   -.031161   .0115614    -2.70   0.007     -.053821   -.0085011
        year |    .031839   .0019586    16.26   0.000     .0280002    .0356779
      select |   .0142546   .0041796     3.41   0.001     .0060626    .0224465
    articles |     .00559   .0007685     7.27   0.000     .0040838    .0070962
    prestige |  -.0350608   .0077056    -4.55   0.000    -.0501635   -.0199581
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. quietly oglm tenure i.female year c.year#c.year select ///
      articles prestige i.female#c.articles if (year <= 10), hetero(i.female) link(logit)

. eststo eff_logit: margins, dydx(*) predict(outcome(1)) post
Average marginal effects                        Number of obs     =      2,797
Model VCE    : OIM
Expression   : Pr(tenure==1), predict(outcome(1))
dy/dx w.r.t. : 1.female year select articles prestige
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -.0312105   .0115836    -2.69   0.007     -.053914    -.008507
        year |   .0319057   .0019277    16.55   0.000     .0281275    .0356839
      select |   .0145808   .0042388     3.44   0.001      .006273    .0228886
    articles |   .0053378   .0007523     7.09   0.000     .0038632    .0068123
    prestige |  -.0360711    .007934    -4.55   0.000    -.0516215   -.0205207
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. esttab eff_probit eff_logit, b(3) se(3)
--------------------------------------------
                      (1)             (2)

--------------------------------------------
0.female            0.000           0.000
                      (.)             (.)
1.female           -0.031**        -0.031**
                  (0.012)         (0.012)
year                0.032***        0.032***
                  (0.002)         (0.002)
select              0.014***        0.015***
                  (0.004)         (0.004)
articles            0.006***        0.005***
                  (0.001)         (0.001)
prestige           -0.035***       -0.036***
                  (0.008)         (0.008)
--------------------------------------------
N                    2797            2797
--------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. *=========================================
. *** Example 2: Labor Participation ***
. *=========================================
.
. * Open dataset and create variables
. import delimited "$dir/mroz.csv", clear
. gen kids = (kidslt6 + kidsge6) > 0
. gen finc = faminc/10000

. * Hetekedastic binary probit model
. quietly eststo labor_hom: probit inlf age c.age#c.age finc educ kids
. quietly eststo labor_het: oglm inlf age c.age#c.age finc educ i.kids, ///
                            hetero(finc i.kids) link(probit)
```

```
. quietly eststo eff_labor_het: margins, dydx(*) predict(outcome(1)) post
. esttab labor_hom labor_het eff_labor_het, b(3) se(3)

------------------------------------------------------------
                      (1)             (2)             (3)   
                     inlf            inlf                   
------------------------------------------------------------
main                                                        
age                 0.185**         0.264*         -0.009***
                  (0.066)         (0.118)         (0.003)   
c.age#c.age        -0.002**        -0.004*                  
                  (0.001)         (0.001)                   
finc                0.046           0.424           0.069** 
                  (0.042)         (0.222)         (0.024)   
educ                0.098***        0.140**         0.030***
                  (0.023)         (0.052)         (0.009)   
kids               -0.449***                                
                  (0.131)                                   
0.kids                              0.000           0.000   
                                      (.)             (.)   
1.kids                             -0.879**        -0.161***
                                  (0.303)         (0.043)   
_cons              -4.157**                                 
                  (1.402)                                   
------------------------------------------------------------
lnsigma                                                     
finc                                0.313*                  
                                  (0.123)                   
0.kids                              0.000                   
                                      (.)                   
1.kids                             -0.141                   
                                  (0.324)                   
------------------------------------------------------------
cut1                                                        
_cons                               6.030*                  
                                  (2.498)                   
------------------------------------------------------------
N                     753             753             753   
------------------------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. * Wald test
. estimates restore labor_het
(results labor_het are active now)

. quietly oglm
. test [lnsigma]: finc 1.kids

 ( 1)  [lnsigma]finc = 0
 ( 2)  [lnsigma]1.kids = 0

           chi2(  2) =    6.53
         Prob > chi2 =    0.0381
```

8 Appendix D: Stata code for binary response models with endogeneity

```
. ************************************
. *** IV Probit ***
. ************************************

. *========================================
. *** Control function approach ***
```

```
. *=======================================

. import delimited "$dir/mroz.csv", clear
(22 vars, 753 obs)

. * Probit estimates and marginal effect
. probit inlf educ exper c.exper#c.exper age kidslt6 kidsge6 nwifeinc

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -402.06651
Iteration 2:   log likelihood = -401.30273
Iteration 3:   log likelihood = -401.30219
Iteration 4:   log likelihood = -401.30219

Probit regression                               Number of obs     =        753
                                                LR chi2(7)        =     227.14
                                                Prob > chi2       =     0.0000
Log likelihood = -401.30219                     Pseudo R2         =     0.2206
---------------------------------------------------------------------------------
           inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
          exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
                |
c.exper#c.exper |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
                |
            age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
        kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
        kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
       nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
          _cons |   .2700768    .508593     0.53   0.595    -.7267473    1.266901
---------------------------------------------------------------------------------

. margins, dydx(nwifeinc)
Average marginal effects                        Number of obs     =        753
Model VCE    : OIM
Expression   : Pr(inlf), predict()
dy/dx w.r.t. : nwifeinc

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0036162   .0014414    -2.51   0.012    -.0064413   -.0007911
------------------------------------------------------------------------------

. * Control function approach
. eststo fstep: reg nwifeinc educ exper c.exper#c.exper age kidslt6 kidsge6 huseduc

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(7, 745)       =     27.13
       Model |  20676.7705         7  2953.82436   Prob > F        =    0.0000
    Residual |  81120.3451       745  108.886369   R-squared       =    0.2031
-------------+----------------------------------   Adj R-squared   =    0.1956
       Total |  101797.116       752  135.368505   Root MSE        =    10.435
---------------------------------------------------------------------------------
       nwifeinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           educ |   .6746951   .2136829     3.16   0.002     .2552029    1.094187
          exper |  -.3129877   .1382549    -2.26   0.024    -.5844034   -.0415721
                |
c.exper#c.exper |  -.0004776   .0045196    -0.11   0.916    -.0093501     .008395
                |
            age |   .3401521   .0597084     5.70   0.000     .2229354    .4573687
```

```
       kidslt6 |   .8262719   .8183785     1.01   0.313    -.7803305    2.432874
       kidsge6 |   .4355289   .3219888     1.35   0.177    -.1965845    1.067642
       huseduc |   1.178155   .1609449     7.32   0.000     .8621956    1.494115
         _cons |  -14.72048   3.787326    -3.89   0.000    -22.15559   -7.285383
------------------------------------------------------------------------------

. predict res_hat, resi

. test huseduc
 ( 1)  huseduc = 0
       F(  1,   745) =   53.59
            Prob > F =    0.0000

. eststo sstep: probit inlf educ exper c.exper#c.exper age kidslt6 kidsge6 nwifeinc res_hat

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -401.13728
Iteration 2:   log likelihood = -400.30361
Iteration 3:   log likelihood = -400.30301
Iteration 4:   log likelihood = -400.30301

Probit regression                               Number of obs     =        753
                                                LR chi2(8)        =     229.14
                                                Prob > chi2       =     0.0000
Log likelihood = -400.30301                     Pseudo R2         =     0.2225
------------------------------------------------------------------------------
           inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           educ |   .1702153   .0376718     4.52   0.000     .0963798    .2440507
          exper |   .1163123   .0193312     6.02   0.000     .0784239    .1542007
                |
c.exper#c.exper |  -.0019459   .0006009    -3.24   0.001    -.0031235   -.0007682
                |
            age |   -.044953   .0101367    -4.43   0.000    -.0648206   -.0250855
        kidslt6 |  -.8444363   .1198154    -7.05   0.000     -1.07927   -.6096025
        kidsge6 |   .0477905   .0443204     1.08   0.281    -.0390758    .1346568
       nwifeinc |  -.0368641   .0182706    -2.02   0.044    -.0726738   -.0010543
        res_hat |   .0267093   .0189352     1.41   0.158    -.0104031    .0638217
          _cons |   .0171187   .5392914     0.03   0.975    -1.039873     1.07411
------------------------------------------------------------------------------

. * Two-step IV-probit
. ivprobit inlf educ exper c.exper#c.exper age kidslt6 kidsge6 (nwifeinc = huseduc), twostep
Checking reduced-form model...

Two-step probit with endogenous regressors        Number of obs   =        753
                                                  Wald chi2(7)    =     173.79
                                                  Prob > chi2     =     0.0000
------------------------------------------------------------------------------
                |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
       nwifeinc |  -.0368641   .0186314    -1.98   0.048    -.0733809   -.0003472
           educ |   .1702153   .0384014     4.43   0.000     .0949499    .2454806
          exper |   .1163123   .0197084     5.90   0.000     .0776846      .15494
                |
c.exper#c.exper |  -.0019459   .0006129    -3.17   0.001    -.0031471   -.0007446
                |
            age |   -.044953    .010327    -4.35   0.000    -.0651936   -.0247125
        kidslt6 |  -.8444363   .1218529    -6.93   0.000    -1.083264    -.605609
        kidsge6 |   .0477905    .045177     1.06   0.290    -.0407549    .1363359
          _cons |   .0171187   .5498911     0.03   0.975    -1.060648    1.094885
------------------------------------------------------------------------------
Instrumented:  nwifeinc
Instruments:   educ exper c.exper#c.exper age kidslt6 kidsge6 huseduc
```

```
------------------------------------------------------------------------------
Wald test of exogeneity: chi2(1) = 1.99                  Prob > chi2 = 0.1584


. *=========================================
. *** MLE ***
. *=========================================

. ivprobit inlf educ exper c.exper#c.exper age kidslt6 kidsge6 (nwifeinc = huseduc)

Fitting exogenous probit model

Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -401.13728
Iteration 2:   log likelihood = -400.30361
Iteration 3:   log likelihood = -400.30301
Iteration 4:   log likelihood = -400.30301

Fitting full model

Iteration 0:   log likelihood = -3230.6635
Iteration 1:   log likelihood = -3230.6421
Iteration 2:   log likelihood = -3230.6421

Probit model with endogenous regressors        Number of obs     =        753
                                               Wald chi2(7)      =     200.50
Log likelihood = -3230.6421                    Prob > chi2       =     0.0000
-----------------------------------------------------------------------------------------
                       |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+-----------------------------------------------------------------
              nwifeinc |  -.0355243   .0161904    -2.19   0.028    -.0672569   -.0037916
                  educ |   .1640289   .0312249     5.25   0.000     .1028293    .2252285
                 exper |    .112085   .0211991     5.29   0.000     .0705356    .1536344
                       |
       c.exper#c.exper |  -.0018751   .0005915    -3.17   0.002    -.0030345   -.0007158
                       |
                   age |  -.0433193   .0113314    -3.82   0.000    -.0655284   -.0211101
               kidslt6 |  -.8137458   .1299442    -6.26   0.000    -1.068432   -.5590599
               kidsge6 |   .0460536   .0431386     1.07   0.286    -.0384966    .1306037
                 _cons |   .0164965   .5300821     0.03   0.975    -1.022445    1.055438
-----------------------+-----------------------------------------------------------------
 corr(e.nwifeinc,e.inlf)|   .2671475   .1791903                      -.1040303    .5730063
         sd(e.nwifeinc)|   10.37928   .2674576                       9.868095    10.91695
-----------------------------------------------------------------------------------------
Instrumented:  nwifeinc
Instruments:   educ exper c.exper#c.exper age kidslt6 kidsge6 huseduc
-----------------------------------------------------------------------------------------
Wald test of exogeneity (corr = 0): chi2(1) = 2.01       Prob > chi2 = 0.1559

. eststo me1: margins, dydx(*) predict(pr) post
Average marginal effects                       Number of obs     =        753
Model VCE    : OIM
Expression   : Average structural function probabilities, predict(pr)
dy/dx w.r.t. : nwifeinc educ exper age kidslt6 kidsge6
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0110576   .0055497    -1.99   0.046    -.0219348   -.0001805
        educ |   .0510572   .0111011     4.60   0.000     .0292994     .072815
       exper |   .0230711   .0029517     7.82   0.000     .0172858    .0288563
         age |   -.013484    .002986    -4.52   0.000    -.0193365   -.0076314
     kidslt6 |  -.2532945   .0330766    -7.66   0.000    -.3181235   -.1884655
     kidsge6 |   .0143351   .0135204     1.06   0.289    -.0121644    .0408345
```

```
------------------------------------------------------------------------------

. qui ivprobit inlf educ exper c.exper#c.exper age kidslt6 kidsge6 (nwifeinc = huseduc)
. eststo me2: margins, dydx(*) predict(pr fix(nwifeinc)) post

Average marginal effects                        Number of obs     =        753
Model VCE    : OIM

Expression   : Average structural function probabilities, predict(pr fix(nwifeinc))
dy/dx w.r.t. : nwifeinc educ exper age kidslt6 kidsge6
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0105638   .0047364    -2.23   0.026    -.0198469   -.0012807
        educ |   .0487769   .0087333     5.59   0.000       .03166    .0658937
       exper |   .0219965   .0037232     5.91   0.000     .0146992    .0292939
         age |  -.0128817   .0033216    -3.88   0.000     -.019392   -.0063714
     kidslt6 |  -.2419815   .0365941    -6.61   0.000    -.3137047   -.1702583
     kidsge6 |   .0136948   .0127924     1.07   0.284    -.0113777    .0387674
------------------------------------------------------------------------------
Warning: The chosen prediction can result in estimates of derivatives or
         contrasts that do not have a structural function interpretation.

. esttab  me1 me2, b(3) se(3)
--------------------------------------------
                      (1)             (2)   

--------------------------------------------
nwifeinc           -0.011*         -0.011*  
                  (0.006)         (0.005)   
educ                0.051***        0.049***
                  (0.011)         (0.009)   
exper               0.023***        0.022***
                  (0.003)         (0.004)   
age                -0.013***       -0.013***
                  (0.003)         (0.003)   
kidslt6            -0.253***       -0.242***
                  (0.033)         (0.037)   
kidsge6             0.014           0.014   
                  (0.014)         (0.013)   
--------------------------------------------
N                     753             753   
--------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

9 Acknowledgments

I express my thanks to FONDECYT, Chile, Grant 1200522 for full financial support.

References

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods & Research* 28 (2): 186–208. https://doi.org/10.1177/0049124199028002003.

Alvarez, R Michael, and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science*, 1055–82. https://doi.org/10.2307/2111669.

An, Weihua, and Xuefu Wang. 2016. "**LARF**: Instrumental Variable Estimation of Causal Effects Through Local Average Response Functions." *Journal of Statistical Software* 71 (1): 1–13. https://doi.org/10.18637/jss.v071.c01.

Arel-Bundock, Vincent. 2023. **marginaleffects***: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. https://vincentarelbundock.github.io/marginaleffects/.
Canty, Angelo J. 2002. "Resampling Methods in R: The **boot** Package." *The Newsletter of the R Project Volume* 2 (3). https://journal.r-project.org/articles/RN-2002-017/.
Carlson, Alyssa. 2019. "Parametric Identification of Multiplicative Exponential Heteroscedasticity." *Oxford Bulletin of Economics and Statistics* 81 (3): 686–96. https://doi.org/10.1111/obes.12280.
Carroll, Nathan. 2018. **oglmx***: Estimation of Ordered Generalized Linear Models*. https://CRAN.R-project.org/package=oglmx.
Elff, Martin. 2012. **memisc***: Tools for Management of Survey Data, Graphics, Programming, Statistics, and Simulation*. http://CRAN. R-project. org/package= memisc.
Fox, John, Michael Friendly, and Sanford Weisberg. 2013. "Hypothesis Tests for Multivariate Linear Models Using the **car** Package." *The R Journal* 5 (1): 39. https://doi.org/10.32614/RJ-2013-004.
Gilbert, Paul, and Ravi Varadhan. 2019. **numDeriv***: Accurate Numerical Derivatives*. https://CRAN.R-project.org/package=numDeriv.
Greene, William H. 2002. *LIMDEP Econometric Modelling Guide: Version 8.0*. New York: Econometric Software. https://www.limdep.com/.
———. 2003. *Econometric Analysis*. 7th ed. Pearson Education India.
Greene, William H, and David A Hensher. 2010. *Modeling Ordered Choices: A Primer*. Cambridge University Press.
Harvey, Andrew C. 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica: Journal of the Econometric Society*, 461–65. https://doi.org/10.2307/1913974.
Henningsen, Arne, and Ott Toomet. 2011. "**maxLik**: A Package for Maximum Likelihood Estimation in R." *Computational Statistics* 26 (3): 443–58. https://doi.org/10.1007/s00180-010-0217-1.
Keele, Luke, and David K Park. 2006. "Difficult Choices: An Evaluation of Heterogeneous Choice Models." In *Paper for the 2004 Meeting of the American Political Science Association*, 2–5.
Knapp, Laura Greene, and Terry G Seaks. 1992. "An Analysis of the Probability of Default on Federally Guranteed Student Loans." *The Review of Economics and Statistics*, 404–11. https://doi.org/10.2307/2109484.
Leeper, Thomas J. 2021. **margins***: Marginal Effects for Model Objects*. https://CRAN.R-project.org/package=margins.
Newey, Whitney K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36 (3): 231–50. https://doi.org/10.1016/0304-4076(87)90001-7.
Rivers, Douglas, and Quang H Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39 (3): 347–66. https://doi.org/10.1016/0304-4076(88)90063-2.
Sarrias, Mauricio. 2016. "Discrete Choice Models with Random Parameters in R: The **Rchoice** Package." *Journal of Statistical Software* 74 (1): 1–31. https://doi.org/10.18637/jss.v074.i10.
StataCorp. 2019. *Stata Statistical Software, Release 16*. College Station, TX: StataCrop LLC. https://www.stata.com/.
Williams, Richard. 2009. "Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups." *Sociological Methods & Research* 37 (4): 531–59. https://doi.org/10.1177/0049124109335735.
———. 2010. "Fitting Heterogeneous Choice Models with `oglm`." *The Stata Journal* 10 (4): 540–67. https://doi.org/10.1177/1536867X1101000402.
Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.
———. 2014. "Quasi-Maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables." *Journal of Econometrics* 182 (1): 226–34. https://doi.org/10.1016/j.jeconom.2014.04.020.
———. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources* 50 (2): 420–45. https://doi.org/10.3368/jhr.50.2.420.
Yatchew, Adonis, and Zvi Griliches. 1985. "Specification Error in Probit Models." *The Review of Economics and Statistics*, 134–39. https://doi.org/10.2307/1928444.
Zaghdoudi, Taha. 2018. "**ivprobit**: An R Package to Estimate the Instrumental Variables Probit." *Journal of Open Source Software* 3 (22): 523. https://doi.org/10.21105/joss.00523.
Zeileis, Achim, Roger Koenker, and Philipp Doebler. 2015. **glmx***: Generalized Linear Models Extended*. https://CRAN.R-project.org/package=glmx.

Mauricio Sarrias
Universidad de Talca
Facultad de Economía y Negocios (FEN)
Universidad de Talca, Talca, Chile
https://www.msarrias.com
ORCiD: 0000-0001-5932-4817

mauricio.sarrias@utalca.cl