1 Limited dependent variables: introduction

Up to now, we have been implicitly assuming that the dependent variable, \( y \), was continuous and that its support (range of variation) was the real line. In a number of real world cases, however, the dependent variable \( y \) is not continuous, or is continuous but bounded below by zero. Here’s a list of the most common cases of limited dependent variable models, along with the popular estimators used to estimate these models:

<table>
<thead>
<tr>
<th>Case</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary response variable</td>
<td>( y = 0, 1 ) Probit, logit, lin. prob. model</td>
</tr>
<tr>
<td>Truncated variable or censored variable</td>
<td>( y &gt; 0 ), or ( y \geq 0 ) Tobit</td>
</tr>
<tr>
<td>Count variable</td>
<td>( y = 0, 1, 2, \ldots ) Poisson model</td>
</tr>
<tr>
<td>Discrete choice variable</td>
<td>( y = \text{bus, train, car} ) Multivariate logit or probit</td>
</tr>
<tr>
<td>Ordered (potentially qualitative) variable</td>
<td>( y = a, b, c ) Ordered probit or logit</td>
</tr>
</tbody>
</table>

Most of these estimators can easily be used since they are programmed up in software programs like Stata. Note also that the fact that special estimators exist does not necessarily mean that it is “wrong” to use OLS. Irrespective of the form of \( y \), we can always write the conditional expectation of \( y \), as a function \( g(.) \) of \( x \): \( E(y|x) = g(x) \). The OLS estimator can always be viewed as a linear approximation of \( g(x) \), i.e. \( g(x) \approx x\beta \). The only difference is that when \( y \) is continuous and unbounded, it is possible that the true conditional expectation is indeed linear, \( g(x) = x\beta \). By contrast, the true conditional expectation cannot generally be linear in the case of limited dependent variables. Running OLS to linearly approximate the true function \( g(x) \) may nonetheless
be useful in many settings, for example in the case of the linear probability model. In other settings like the tobit, however, running OLS regressions may not be a useful alternative if we want to correct for selection biases that afflict OLS estimates.

All models but the tobit are discrete response models, i.e. $y$ takes on a limited number of values. These models are covered in this set of lecture notes. The tobit is a different case and will be covered in the next set of notes.

2 Binary response models: probit, logit and linear probability model

This section draws heavily on Wooldridge (2002). Consider models of the form:

$$E(y|x) \equiv \Pr(y = 1|x) = G(x\beta)$$

(1)

This is called an index model as it constrains the effect of $x$ to go through the index $x\beta$. In the case of the linear probability model we saw earlier, the function $G(.)$ is the identity function, i.e. $G(x\beta) = x\beta$. An obvious problem with this model is that nothing prevents $x\beta$ to take on values below zero or above one, which is inconsistent with a probability model. Two popular models that are well defined probability models are the probit and logit models, where $G(.)$ turn out to be cumulative distribution functions (standard normal in the case of the probit).

2.1 Probit and logit

From both an economic and an econometric point of view, it is useful to start with the latent variable $y^*$, where

$$y^* = x\beta + e$$

(2)

For example, say that $y = 1$ if a consumer buys a durable good, and that $y = 0$ otherwise. $y^*$ can be viewed as a propensity to consume the good, where the good is bought whenever $y^*$ exceeds a certain threshold (zero typically):

$$y = 1 \text{ if } y^* > 0, \ y = 0 \text{ otherwise}$$

(3)
Under the assumption that $e$ follows a standard normal distribution, we have

$$\Pr(y = 1|x) = \Pr(y^* > 0) = \Pr(e > -x\beta)$$

$$= 1 - \Phi(-x\beta) = \Phi(x\beta)$$

(4)

i.e. the function $G()$ is simply the cumulative normal distribution. The logit model is similarly defined under the alternative assumption that $e$ follows a logistic (also called “extreme value”) distribution:

$$\Pr(y = 1|x) = \exp(x\beta) /[1 + \exp(x\beta)]$$

(5)

Unlike the probit, the logit has a closed form for the function $G()$, which explains why the logit was popular back when computing power was expensive. Both the logit and probit are usually estimated by maximum likelihood. For each observation $i$, the probability density function of $y_i$ given $x_i$ is:

$$f(y_i|x_i; \beta) = [G(x_i\beta)]^{y_i}[1 - G(x_i\beta)]^{1-y_i}, y_i = 0, 1$$

(6)

The log-likelihood is

$$l_i(\beta) = y_i \log[G(x_i\beta)] + (1 - y_i) \log[1 - G(x_i\beta)]$$

(7)

and $\beta$ can be estimated by numerically maximizing the likelihood function over the observed sample:

$$\text{Max } L(\beta) \text{ w.r.t } \beta, \text{ where } L(\beta) = \sum_{i=1}^{N} l_i(\beta)$$

(8)

Standard tests can be performed on the estimated values of $\beta$ using the standard testing approach of maximum likelihood. For example, standard t-tests (normal to be more precise) can be performed using the estimated variance-covariance matrix of $\beta$ (this type of test is called a Wald test). Constrained and unconstrained models can also be compared using a likelihood ratio test (2 times the difference in log-likelihoods follows a chi-square distribution with $k$ degrees of freedom, where $k$ is the number of restrictions).
2.2 Marginal effects

Unlike in the case of OLS, $\beta$ in the probit or logit model cannot be interpreted as the impact of a small change of $x$ on the outcome variable $y$. Using the chain rule, it follows that the marginal effect (ME) of $x$ is

$$ME = \frac{\partial \Pr(y = 1|x)}{\partial x} = \frac{\partial G(x\beta)}{\partial x} = g(x\beta)\beta,$$

(9)

where $g(z) \equiv \frac{\partial G(z)}{\partial z}$ is the probability density function (standard normal in the case of the probit, i.e. $g(x\beta) = \phi(x\beta)$). So if we want to know how a small increase in $x$ affects the probability of choosing $y = 1$ ($\Pr(y = 1|x) = G(x\beta)$), we need to multiply the estimated $\beta$ by the density evaluated at $x\beta$, $g(x\beta)$. Since $g(x_i\beta)$ takes on different values for different individuals $i$, it is not completely clear what values of $x$ should be picked. One popular method used to compute the marginal effect is to evaluate $g(x\beta)$ at the average value of $x$:

$$ME(1) = g(\bar{x}\beta)\beta$$

(10)

Another possibility is to compute a marginal effect $ME_i = g(x_i\beta)\beta$ for each individual, and compute the average:

$$ME(2) = \left[ \frac{1}{N} \sum_{i=1}^{N} g(x_i\beta) \right] \beta$$

(11)

This is the formula used by default by the command “margins” in Stata (for $ME(1)$ the option “atmeans” can be used). The common practical wisdom is that 1) probit and logit marginal effects tend to be very similar, and 2) that these marginal effects are very close to the linear probability model coefficients. So when we are mostly interested in marginal effects (e.g. how prices, $x$, affect the probability of buying durable goods), using a linear probability model is a highly defendable practice. The model should not be used, however, to predict probabilities for particular subgroups of workers as these probabilities may turn out to be negative or larger than one.

2.3 Endogeneity

Endogeneity of one of the explanatory variable $x$ is tricky in all non-linear models, including probits and logits. One simple solution is to use a linear probability model instead
and just run two-stage least-squares as in any other linear model. One reasonable ap-

proach is to both run an OLS linear probability model and a probit/logit to first show

that marginal effects are similar when the x variables are assumed to be exogenous. If

so, we can then simply run two-stage least-squares on the linear probability model and

argue that doing a proper endogenous probit or logit model (see Wooldridge or Greene

for more detail on these models) would yield similar marginal effects.

It is also possible to use special estimators that have been developed for endogenous

probit or logit models. For example, the “ivprobit” command in Stata can be used

to jointly estimate the probit model and the first-stage equation for the endogenous

regressor.

3 Ordered probit/logit models

The most straightforward extension of the probit or logit model is the case where y takes

on several possible values, but where there is a natural order in the responses. Obviously,

we can think of the number of consumed goods in a given period (e.g. 0, 1, 2, 3,... movies

seen in a month) as an ordered variable, though count models could also be used in such

a case.

A better application of ordered models is thus the case of ordered qualitative variables.
For instance, many surveys including course evaluations ask respondents opinions about

different factors or outcomes (e.g. is the instructor excellent, good, bad or awful...). Say

you order these qualitative responses in four categories: y = 0, 1, 2, 3. The latent variable

framework can then be used to form a model:

\[
\begin{align*}
    y &= 0 \text{ if } y^* \leq 0 \\
    y &= 1 \text{ if } 0 < y^* \leq T_1 \\
    y &= 2 \text{ if } T_1 < y^* \leq T_2 \\
    y &= 3 \text{ if } y^* > T_2
\end{align*}
\]

where the T’s are threshold parameters that can be estimated along with \( \beta \). The

estimation is performed by maximum likelihood again. For example, in the case where
is normal, we can form the likelihood function using the probabilities of each choice:

\[
\begin{align*}
\Pr(y = 0|x) &= \Phi(-x\beta) \\
\Pr(y = 1|x) &= \Phi(-x\beta + T_1) - \Phi(-x\beta) \\
\Pr(y = 2|x) &= \Phi(-x\beta + T_2) - \Phi(-x\beta + T_1) \\
\Pr(y = 3|x) &= 1 - \Phi(-x\beta + T_2)
\end{align*}
\]

An ordered logit model can be defined similarly. Marginal effects can also be computed, just as in the case of the simple probit or logit.

4 Multinomial probit/logit

It is not always possible, however, to rank all possible discrete values of \( y \) in a natural order. A classic example of this is the choice of transportation modes, e.g. taking a bicycle \((y = 1)\), a bus \((y = 2)\), or a car \((y = 3)\) to go to UBC. Daniel McFadden received a Nobel prize for showing (among other things) how to estimate such models in a way that is consistent with an underlying model of utility maximization. This can be done once again using the latent variable formulation, where the latent variables (the \( V \)'s here) can now be interpreted as utilities:

\[
\begin{align*}
V_1 &= x\gamma_1 + u_1 \\
V_2 &= x\gamma_2 + u_2 \\
V_3 &= x\gamma_3 + u_3
\end{align*}
\]

(12)

(13)

(14)

It follows that \( y = 1 \) will be chosen iff \( V_1 > V_2 \) and \( V_1 > V_3 \), \( y = 2 \) will be chosen iff \( V_2 > V_1 \) and \( V_2 > V_3 \), and \( y = 3 \) will be chosen iff \( V_3 > V_1 \) and \( V_3 > V_2 \). As it turns out, it is simpler to redefine the latent variables relative to a base choice, for example \( y = 3 \). We can then define

\[
\begin{align*}
y_1^* &= V_1 - V_3 = x\beta_1 + \epsilon_1 \\
y_2^* &= V_2 - V_3 = x\beta_2 + \epsilon_2
\end{align*}
\]

(15)

(16)
where $\beta_1 = \gamma_1 - \gamma_3, e_1 = u_1 - u_3, \beta_2 = \gamma_2 - \gamma_3, e_2 = u_2 - u_3$. It then follows that 

$y = 1$ will be chosen iff $y_1^* > 0$ and $y_1^* > y_2^*$, etc.

It is not straightforward to estimate this model when the $u$’s (and thus $e$’s) are normally distributed, especially when the number of possible values of $y$ is larger than 3. The numerical problem is that we have to numerically evaluate multivariate integrals, though simulation methods help deal with this problem. By contrast, the probabilities of choice are easily computed in the case of the logit:

$$\Pr(y = j|x) = \exp(x\beta_j) /[1 + \exp(x\beta_1) + \exp(x\beta_2)], \text{ for } j = 1, 2 \tag{17}$$

$$\Pr(y = 3|x) = 1/[1 + \exp(x\beta_1) + \exp(x\beta_2)] \tag{18}$$

This computational simplicity explains the popularity of the multinomial logit model. As in the case of the simple logit, it is advisable to compute marginal effects since those are often the ultimate objects of interest. For example, say that one of the $x$ variable is the price of a bus pass and that we want to know whether subsidizing bus passes would increase bus ridership to UBC (relative to taking a bicycle or a car). The $\beta$’s are parameters of the utility function but they do not provide a direct answer to this question. We have to compute marginal effects instead, just like we usually compute demand elasticities (themselves functions of the utility function parameters) in standard consumer demand problems.

5 Count data

As mentioned above, ordered logits or probits can be used when $y$ is a count variable, i.e. when $y = 0, 1, 2, 3, \ldots$. Using an ordered model is not practical, however, in cases where $y$ can take a large number of values, since we have to estimate a threshold for each value, i.e. estimate a very large number of parameters. A more natural model to use in this case is a Poisson model. A Poisson distribution is the most natural distribution to use for count data. For example, it can be shown that if we flip coins a large number of times, the resulting distribution of the number of heads (or tails) will follow a Poisson distribution.

The Poisson distribution is very convenient to work with because the conditional distribution of $y$ is solely a function of the conditional mean of $y$ (given $x$). Writing the
conditional mean as \( \mu(x) = E(y|x) \), it follows that the density of \( y \) is:

\[
f(y|x) = \exp[-\mu(x)][\mu(x)]^y/y!
\]

In general, this model can be estimated by maximum likelihood. A very convenient specification to use is \( \mu(x) = \exp(x\beta) \). In that case, \( \beta \) can be estimated using several software programs including Stata. The marginal effects are:

\[
\frac{\partial E(y|x)}{\partial x} = \exp(x\beta)\beta
\]

It also follows that:

\[
\beta = \frac{\partial E(y|x)}{\partial x} \cdot \frac{1}{E(y|x)} = \frac{\partial \log E(y|x)}{\partial x}
\]

Thus, \( \beta \) can be interpreted as the semi-elasticity of \( y \) with respect to \( x \).

6 Stata commands

First consider a binary response variable \( y = 0,1 \). Say we have three explanatory variables \( x_1, x_2, \) and \( x_3 \). The stata commands for the probit and logit are simply:

\[
\text{probit y x1 x2 x3}
\]
\[
\text{logit y x1 x2 x3}
\]

Marginal effects (with associated standard errors) are obtained using the “margins” command and the option “dydx” after running the model. The default is version 2 of the marginal effects (ME(2) above). Marginal effects can also be computed at the mean value of the explanatory variables (ME(1)) using the “atmeans” option. The two versions of the command in the example above are:

\[
\text{margins, dydx(*)}
\]
\[
\text{margins, dydx(*) atmeans}
\]

Note that Stata automatically detects whether explanatory variables are dummies or not. In the case of a dummy, we cannot really take a derivative with respect to \( x \). So
Stata computes instead the effect on the predicted probability of switching the dummy from zero to one, holding other x variables at their means.

The commands for ordered logits, ordered probits, and multinomial logits are all similar and are called ologit, oprobit, and mlogit, respectively. Finally, if y is a count variable, just do:

\[ \text{poisson y x1 x2 x3} \]
ECON 594: COMPARING LINEAR PROBABILITY MODEL TO PROBIT AND LOGIT

. reg hwage nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4, vce(robust)

Linear regression
Number of obs = 10000
F(11, 9988) = 318.58
Prob > F = 0.0000
R-squared = 0.1963
Root MSE = 0.38935

------------------------------------------------------------------
 |          Robust
 hwage |        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
 -------------+---------------------------------------------------------------
  nonwhite |  -.0473072    .016463    -2.87   0.004    -.0795781   -.0150364
  covered |   .0203332    .009167     2.22   0.027      .002364    .0383024
    ed0 |  -.1611591   .0135898   -11.86   0.000    -.1877979   -.1345204
    ed1 |  -.0994197   .0103188    -9.63   0.000    -.1196467   -.0791927
    ed3 |   .1019711   .0111448     9.15   0.000     .0801251    .1238171
    ed4 |   .2840649   .0140097    20.28   0.000      .256603    .3115268
    ed5 |   .3139707   .0162058    19.37   0.000     .2822039    .3457374
    age1 |  -.0713636   .0185424    -3.85   0.000    -.1077104   -.0350167
    age2 |   .2724906    .027401     9.94   0.000     .2187791    .3262021
    age3 |  -.1075773   .0119891    -8.97   0.000    -.1310783   -.0840763
    age4 |   .0119867   .0015813     7.58   0.000     .0088871    .0150863
     _cons |   .0054739   .0060027     0.91   0.362    -.0062926    .0172405
------------------------------------------------------------------

. probit hwage nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4

Iteration 0:   log likelihood = -5643.0402
...
Iteration 5:   log likelihood = -4516.9922

Probit regression
Number of obs = 10000
LR chi2(11)   = 2252.10
Prob > chi2   = 0.0000
Log likelihood = -4516.9922
Pseudo R2     = 0.1995

------------------------------------------------------------------
 hwage        |        Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
 -------------+---------------------------------------------------------------
  nonwhite |  -.1742219   .0683698    -2.55   0.011    -.3082241   -.0402196
  covered |   .1126326   .0324429     3.47   0.001     .0500457    .1752196
    ed0 |  -.6932127   .0788324    -8.79   0.000    -.8477212   -.5387041
    ed1 |  -.5412199   .0656299    -8.25   0.000    -.6698522   -.4125876
    ed3 |   .3863531   .0430004     8.98   0.000     .3020739    .4706323
    ed4 |   .879596   .0438801    20.05   0.000      .7935927    .9655993
    ed5 |   .9108847   .0474343    19.20   0.000     .8179151    1.003854
    age1 |   1.833902   .258612     7.09   0.000      1.327032    2.340773
    age2 |  -.6803089   .2309653    -2.95   0.003    -.1132993   -.2472625
    age3 |   .116898   .0787711     1.48   0.138    -.0374906    .2712865
    age4 |  -.0084126   .0089728    -0.94   0.348     -.025999    .0082738
     _cons |  -2.407275   .0960716    -25.06   0.000    -2.595572   -2.219878
------------------------------------------------------------------

. margins, dydx(*)

Average marginal effects
Model VCE : OIM
Expression : Pr(hwage), predict()
dy/dx w.r.t. : nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4
<table>
<thead>
<tr>
<th>Delta-method</th>
</tr>
</thead>
<tbody>
<tr>
<td>dy/dx   Std. Err.  z    P&gt;</td>
</tr>
<tr>
<td>-------------+--------------------------------</td>
</tr>
<tr>
<td>nonwhite</td>
</tr>
<tr>
<td>covered</td>
</tr>
<tr>
<td>ed0</td>
</tr>
<tr>
<td>ed1</td>
</tr>
<tr>
<td>ed3</td>
</tr>
<tr>
<td>ed4</td>
</tr>
<tr>
<td>ed5</td>
</tr>
<tr>
<td>age1</td>
</tr>
<tr>
<td>age2</td>
</tr>
<tr>
<td>age3</td>
</tr>
<tr>
<td>age4</td>
</tr>
</tbody>
</table>

```
.margins, dydx(*) atmeans

Conditional marginal effects
Number of obs = 10000
Model VCE : OIM

Expression : Pr(hwage), predict()
dy/dx w.r.t. : nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4
at           : nonwhite        =       .0553 (mean)
                    covered         =       .3151 (mean)
...```

<table>
<thead>
<tr>
<th>Delta-method</th>
</tr>
</thead>
<tbody>
<tr>
<td>dy/dx   Std. Err.  z    P&gt;</td>
</tr>
<tr>
<td>-------------+--------------------------------</td>
</tr>
<tr>
<td>nonwhite</td>
</tr>
<tr>
<td>covered</td>
</tr>
<tr>
<td>ed0</td>
</tr>
<tr>
<td>ed1</td>
</tr>
<tr>
<td>ed3</td>
</tr>
<tr>
<td>ed4</td>
</tr>
<tr>
<td>ed5</td>
</tr>
<tr>
<td>age1</td>
</tr>
<tr>
<td>age2</td>
</tr>
<tr>
<td>age3</td>
</tr>
<tr>
<td>age4</td>
</tr>
</tbody>
</table>

```
.logit hwage nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4

Iteration 0:  log likelihood = -5643.0402

Iteration 5:  log likelihood = -4512.4795

Logistic regression
Number of obs = 10000
LR chi2(11) = 2261.12
Prob > chi2 = 0.0000

Log likelihood = -4512.4795
Pseudo R2 = 0.2003

<table>
<thead>
<tr>
<th>hwage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef.  Std. Err.  z    P&gt;</td>
</tr>
<tr>
<td>-------------+--------------------------------</td>
</tr>
<tr>
<td>nonwhite</td>
</tr>
<tr>
<td>covered</td>
</tr>
<tr>
<td>ed0</td>
</tr>
</tbody>
</table>
### Average marginal effects

**Model VCE**: OIM  
**Expression**: Pr(hwage), predict()  
**dy/dx w.r.t.**: nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>nonwhite</td>
<td>-0.0481928</td>
<td>0.0177966</td>
<td>-2.71</td>
<td>0.007</td>
<td>-0.0830734</td>
<td>-0.0133121</td>
</tr>
<tr>
<td>covered</td>
<td>0.0223947</td>
<td>0.008313</td>
<td>2.69</td>
<td>0.007</td>
<td>0.0061015</td>
<td>0.0386879</td>
</tr>
<tr>
<td>ed0</td>
<td>-0.185612</td>
<td>0.0225621</td>
<td>-8.04</td>
<td>0.000</td>
<td>-0.231821</td>
<td>-0.139403</td>
</tr>
<tr>
<td>ed1</td>
<td>-0.145642</td>
<td>0.0182926</td>
<td>-7.96</td>
<td>0.000</td>
<td>-0.285212</td>
<td>-0.096076</td>
</tr>
<tr>
<td>ed3</td>
<td>0.1012291</td>
<td>0.0107814</td>
<td>9.39</td>
<td>0.000</td>
<td>0.080998</td>
<td>0.122301</td>
</tr>
<tr>
<td>ed4</td>
<td>0.2200589</td>
<td>0.010121</td>
<td>21.74</td>
<td>0.000</td>
<td>0.200222</td>
<td>0.239857</td>
</tr>
<tr>
<td>ed5</td>
<td>0.225738</td>
<td>0.010206</td>
<td>20.83</td>
<td>0.000</td>
<td>0.204157</td>
<td>0.246518</td>
</tr>
<tr>
<td>age1</td>
<td>0.6341869</td>
<td>0.0792764</td>
<td>8.00</td>
<td>0.000</td>
<td>0.576808</td>
<td>0.691569</td>
</tr>
<tr>
<td>age2</td>
<td>-0.297384</td>
<td>0.067655</td>
<td>-4.40</td>
<td>0.000</td>
<td>-0.429986</td>
<td>-0.164782</td>
</tr>
<tr>
<td>age3</td>
<td>0.0661775</td>
<td>0.0223878</td>
<td>2.96</td>
<td>0.003</td>
<td>0.0222983</td>
<td>0.110568</td>
</tr>
<tr>
<td>age4</td>
<td>-0.0057997</td>
<td>0.0024997</td>
<td>-2.32</td>
<td>0.020</td>
<td>-0.010699</td>
<td>-0.000904</td>
</tr>
</tbody>
</table>

### Conditional marginal effects

**Model VCE**: OIM  
**Expression**: Pr(hwage), predict()  
**dy/dx w.r.t.**: nonwhite covered ed0 ed1 ed3 ed4 ed5 age1 age2 age3 age4  
**at**: nonwhite = 0.0553 (mean)  
covered = 0.3151 (mean)

<table>
<thead>
<tr>
<th></th>
<th>Delta-method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>nonwhite</td>
<td>-0.0432834</td>
<td>0.0160364</td>
<td>-2.70</td>
<td>0.007</td>
<td>-0.0741742</td>
<td>-0.0118526</td>
</tr>
<tr>
<td>covered</td>
<td>0.0201134</td>
<td>0.0074858</td>
<td>2.69</td>
<td>0.007</td>
<td>0.0054415</td>
<td>0.0347853</td>
</tr>
<tr>
<td>ed0</td>
<td>-0.1670174</td>
<td>0.0205736</td>
<td>-8.12</td>
<td>0.000</td>
<td>-0.2073409</td>
<td>-0.1266939</td>
</tr>
<tr>
<td>ed1</td>
<td>-0.1308059</td>
<td>0.0163984</td>
<td>-7.96</td>
<td>0.000</td>
<td>-0.1629461</td>
<td>-0.0986657</td>
</tr>
<tr>
<td>ed3</td>
<td>0.0909169</td>
<td>0.0104433</td>
<td>9.05</td>
<td>0.000</td>
<td>0.0712305</td>
<td>0.110634</td>
</tr>
<tr>
<td>ed4</td>
<td>0.1976416</td>
<td>0.0111838</td>
<td>17.67</td>
<td>0.000</td>
<td>0.1757218</td>
<td>0.2195614</td>
</tr>
<tr>
<td>ed5</td>
<td>0.2024151</td>
<td>0.0120441</td>
<td>16.81</td>
<td>0.000</td>
<td>0.1788091</td>
<td>0.2260211</td>
</tr>
<tr>
<td>age1</td>
<td>0.5695827</td>
<td>0.0600572</td>
<td>9.48</td>
<td>0.000</td>
<td>0.4518728</td>
<td>0.6872926</td>
</tr>
<tr>
<td>age2</td>
<td>-0.2670899</td>
<td>0.0590281</td>
<td>-4.78</td>
<td>0.000</td>
<td>-0.3766574</td>
<td>-0.1575223</td>
</tr>
<tr>
<td>age3</td>
<td>0.0594361</td>
<td>0.0191162</td>
<td>3.11</td>
<td>0.002</td>
<td>0.021969</td>
<td>0.096931</td>
</tr>
<tr>
<td>age4</td>
<td>-0.0052089</td>
<td>0.0021654</td>
<td>-2.41</td>
<td>0.016</td>
<td>-0.0094529</td>
<td>-0.009649</td>
</tr>
</tbody>
</table>