Boston University

Institute for Economic Development

Heteroscedastic Sample Selection and Developing Country Wage Equations

Julie Anderson Schaffner

IED Discussion Paper Series
Number 101
June 1999

Heteroscedastic Sample Selection and Developing Country Wage Equations

Julie Anderson Schaffner*
Stanford University
June 1999

Abstract: It has become common to deal with potential selectivity bias in developing country wage equations by employing Heckman's (1979) two-step method or related techniques, despite the potential for such methods to produce misleading results if the assumptions on which they are based are incorrect. This paper highlights the importance of allowing for (the nonlinearities implied by) selection rule heteroscedasticity, and of employing model selection tests and sensitivity analysis, in producing results that inspire more confidence than those of standard applications. There is economic reason to suspect heteroscedasticity and econometric reason to believe that the nonlinearities it introduces into the first stage will improve the performance of two-stage estimators. In an application to urban Peru, homoscedasticity is strongly rejected, and in models allowing for heteroscedasticity, selection rule normality is no longer rejected and estimates of key parameters become more robust to changes in other statistical assumptions. Because the nonlinearities appear to be captured well by the inclusion of quadratic terms in the first stage, the results suggest that researchers may have much to gain by including quadratic terms in standard probit selection rule estimation.

JEL Codes: J31,O15
Keywords: selectivity bias, developing country wages

*Department of Economics, Stanford University, Stanford, CA 94305-6072. The author gratefully acknowledges helpful comments from Joseph Hotz, Anjini Kochar, Anne Royalty, Frank Wolak and two anonymous referees.

I. Introduction

High rates of self employment in developing countries generate widespread suspicion of selectivity bias in OLS estimation of wage equations. Many researchers have dealt with potential selectivity bias by adopting Heckman's (1979) two-step method or maximum likelihood methods employing similar statistical assumptions. [1] Unfortunately, standard applications of these methods inspire little confidence, because such methods can produce misleading results when the statistical assumptions on which they are based are incorrect (see, e.g. Goldberger, 1983).

This paper suggests that the results produced even by parametric, easy-to-implement selectivity bias correction methods can inspire more confidence when model selection testing is used to select (from a specified, diverse set) assumptions for which there is support in the data, and when sensitivity analysis is used to identify parameters whose estimates are robust across a wide range of assumptions. In particular, it highlights the importance of allowing for selection rule heteroscedasticity. It sets out economic reason to suspect heteroscedasticity and econometric reason to believe that the nonlinearities introduced by heteroscedasticity into selection rule estimation can improve performance of two-step estimators, and suggests simple ways of allowing for such heteroscedasticity. After reviewing ways of allowing also for non-normality in selection rule and wage offer equation errors, it then presents an application to urban Peru, in which selection rule homoscedasticity is strongly rejected, in which allowing for heteroscedasticity appears much more important than allowing for selection rule non-normality, and in which allowing for selection rule heteroscedasticity causes estimates of some parameters to become much more robust across changes in other statistical assumptions. The salient features of these nonlinearities appear to be captured well

[1]Developing country examples applying Heckman's method to control for selection into wage employment are Sahn and Alderman (1988), Tannen (1991), Sumner (1981), King (1990), Deolalikar (1993), and Ashraf and Ashraf (1993). U.S. examples applying Heckman's method include Blau and Beller (1988) and Wellington (1993). Examples applying related techniques controlling for selection across more than two activities include Glewwe (1991), Khandker (1991), Strauss and Thomas (1991), Gindling (1991), Appleton, Collier and Horsnell (1990), Alderman and Kozel (1989), Vijverberg (1993), and Magnac (1991).

by the inclusion of quadratic terms in selection rule estimation. The practical implication is that researchers have much to gain by including quadratic terms in standard probit selection rule estimation.

Section II presents the model underlying standard estimation of wage equations in endogenously selected samples. Section III discusses reasons for, implications of and methods for introducing, selection rule heteroscedasticity, while Section IV reviews ways of handling departures from normality in selection rule and wage offer equation errors. Together these sections describe a wide range of alternative parametric statistical models that admit two-stage (and in some cases maximum likelihood) estimation, and that serve as the basis for model selection testing and sensitivity analysis. The application to urban Peru is described in Section V and Section VI concludes.

II. The Standard Model

In standard selectivity-adjusted wage equation applications, the natural logarithm of the wage offer (w_i) is assumed to be determined according to:

$$w_i = X_i \beta + \epsilon_i \tag{1}$$

and is observed only if the individual is a wage employee. Selection into wage employment (denoted by $I_i = 1$) follows the rule:

$$I_i = 1 \ \ iff \ I_i^* > 0$$

$$I_i = 0 \ \ iff \ I_i^* \leq 0$$

where

$$I_i^* = Z_i \gamma + \eta_i, \tag{2}$$

X_i and Z_i are exogenous vectors of the individual's characteristics, β and γ are vectors of parameters to be estimated, and ϵ_i and η_i are normal random variables with variances σ^2 and 1, respectively, and correlation coefficient ρ. In what follows, I will refer to equation (1) as the wage offer equation and equation (2) as the selection rule.

Under these assumptions, the regression of w_i on X_i in the wage-earning subsample is

$$E(w_i|X_i, I_i = 1) = X_i\beta + E(\epsilon_i|I_i^* \geq 0) = X_i\beta + \sigma\rho M(Z_i\gamma) \tag{3}$$

where $M(v) = \phi(v)/\Phi(v)$, and where $\phi(\ )$ and $\Phi(\ )$ are the standard normal pdf and cdf. As long as ρ is nonzero, and X_i and Z_i are correlated, then OLS regression coefficients are biased estimates of the true β's, which describe ceteris paribus effects of the X's on log earnings. The popular two-stage method of "correcting" for this bias involves first estimating the Inverse Mills' Ratio, $M(Z_i\gamma)$, using probit estimates of γ, and then including it in an OLS regression of w_i on X_i in the selected sample. The model may also be estimated by full maximum likelihood methods.

III. Heteroscedasticity and Selection Rule Nonlinearity

The selection rule error, η_i, summarizes the effects of unobserved characteristics on the propensity to be a wage earner. Its variance will depend on the observed characteristics Z_i when Z_i and the unobserved characteristics in the error term have interactive effects on wages. For example, consider the simple case in which there is just one unobserved characteristic ν_i, with $E(\nu_i|Z_i) = 0$ and $Var(\nu_i|Z_i) = \sigma_\nu^2$, and in which

$$I_i^* = Z_i\gamma + h(Z_i)\nu_i. \tag{4}$$

$h(Z_i)$ is the effect of an additional unit of the unobserved characteristic on the propensity to be a wage earner(I_i^*), for individuals with observed characteristics Z_i. In this model the Z_i's have random coefficients, the means of which are the elements of γ, and the variance of the selection rule error η_i is heteroscedastic, related to the Z_i's according to

$$Var(\eta_i) = h(Z_i)^2\sigma_\nu^2.$$

A variety of potentially important interactions suggest themselves. For example, if capital requirements are important for entering self-employment, then unobserved characteristics influencing individuals' borrowing constraints (such as unobserved dimensions of wealth and connections) might affect the choice of wage over self-employment, but their effect might be stronger for younger individuals, who have had less opportunity to save up for entry into self-employment. Thus observed age and unobserved borrowing constraints might have a (negatively) interactive effect on selection into wage employment. For another example, spending time in schools with unobservedly higher quality or better links to wage employing establishments might have a greater effect on the propensity to be wage employees than spending time in lower quality and less well connected schools. Thus unobserved school qualities and observed years of schooling might be expected to have (positively) interactive effects on the propensity to be wage earners.

Interaction effects between observed and unobserved characteristics are a possibility in any model, but it is especially important to allow for the possibility of the implied heteroscedasticity in selection rule models, which are nonlinear. As is well known, ignoring heteroscedasticity leads to inconsistency in nonlinear models (see, e.g. Hurd, 1979).

Consider generalizing the standard model of the previous section to allow simply, but flexibly for selection rule heteroscedasticity. Let the selection rule error have variance

$$\sigma_{\eta i}^{2} = \exp(2 Z_i^{-} \delta) \tag{5}$$

where Z_i^{-} is a vector containing all the elements of Z_i except the intercept term, so that

$$\eta_i' = \frac{\eta_i}{\exp(Z_i^{-} \delta)} \tag{6}$$

rather than η_i is standard normal. This functional form for the dependence of σ_η^2 on Z_i is convenient, because it constrains the standard deviation to be positive and nests the case of homoscedasticity ($\delta = 0$). This formulation could be derived from a structural model such as the one described in (4), setting $h(Z_i) = \exp(Z_i^{-} \delta)$, but it may also represent the reduced form of a model with multiple unobserved characteristics. Under such heteroscedasticity, equation (3) must be modified by replacing $M(Z_i \gamma)$ by

$$M(g(Z_i)) = \frac{\phi(Z_i \gamma / \exp(Z_i^{-} \delta))}{\Phi(Z_i \gamma / \exp(Z_i^{-} \delta))},$$

where the modified selection rule index $g(Z_i)$ equals $Z_i\gamma/\exp(Z_i^{-}\delta)$. Consistent estimates of the wage offer equation may still be obtained using two-stage (as well as maximum likelihood) methods, but the first stage is now a heteroscedastic probit, from which estimates of δ as well as γ are obtained. Likelihood ratio tests of the null hypothesis of homoscedasticity are straightforward to perform.[2]

Heteroscedasticity introduces nonlinearity into the selection rule index (i.e. the argument of the selection correction term M(.)), which may be useful for improving identification of the second stage wage offer equation. As pointed out by Nelson (1984), Nawata (1994) and others, two-step estimation of standard sample selection models may be hampered by high collinearity between the M(.) term and the second stage regressors, because in certain ranges (clarified by Leung and Yu, 1996) the M(.) function is nearly linear. Heteroscedasticity leads to the replacement of $M(Z_i\gamma)$ by $M(g(Z_i)) = M(Z_i\gamma/\exp(Z_i^{-}\delta))$, which will often be less collinear with Z_i. Furthermore, heteroscedastic selection rule models may be identified even in one case, discussed by Olsen (1980), in which homoscedastic models are not identified: the case in which $X_i = Z_i$, the selection rule error is uniformly distributed, and the conditional mean of the wage offer equation error is linear in η_i. In this case $M(g(Z_i)) = g(X_i)$, and the second stage regression suffers from perfect multicollinearity unless the selection rule index g(.) is nonlinear. Selection rule heteroscedasticity can thus be seen as a functional form assumption aiding identification in much the same way that the normality assumption does. Having more economic content, however, researchers may find it a more satisfying sort of functional form assumption to exploit in identification.

It will be useful to compare performance of heteroscedastic probit selection models to homoscedastic selection models in which nonlinearity is introduced into the index more simply by

[2] For ease of exposition, I refer to rejection of the null as rejection of homoscedasticity in favor of heteroscedasticity. Note, however, that the likelihood function for the model described in (2) together with heteroscedasticity as in (5) is the same as the likelihood function for a model that replaces (2) by the nonlinear expression $I_i^* = Z_i\gamma/\exp(Z_i\delta) + \eta_i$, with η_i being i.i.d. standard normal. Thus rejection of model (2) with homoscedasticity in favor of the more general model of this section could result either from heteroscedasticity in the selection rule error or from greater nonlinearity in the deterministic portion of the selection rule.

augmenting Z_i with quadratic terms. Probits with quadratic terms are easier to implement in common statistical packages than are heteroscedastic probits. Thus it would be useful for applied researchers to know whether quadratic models capture the salient nonlinearities as well as the heteroscedastic models. The (homoscedastic) quadratic and heteroscedastic models may be compared by employing Vuong's (1989) nonnested likelihood ratio tests (discussed more below) to discern which is closer to the unknown true model.

IV. Departures from Normality

Two additional assumptions made in standard applications of selectivity bias correction methods have been the subject of more attention than the homoscedasticity assumption: the assumptions that the selection rule error and wage offer equation error are normally distributed.[3] This section collates from the literature a set of easily implementable model specifications that introduce departures from both normality assumptions, and that allow use of model selection testing and sensitivity analysis in evaluating the assumptions. In the interest of facilitating sensitivity analysis, non-normalities are introduced in ways that admit estimation with only straightforward modifications of parametric, two-stage estimation.[4]

For the parametric methods employed here, the distribution of the selection rule error, $F(.)$, must be fully specified. In order to assess the potential role for departures from normality in both kurtosis and skew (as in Mroz, 1987), I compare the performance of first-stage models employing the normality assumption, to those in which the first stage error is assumed logistic and lognormal. Likelihood ratio tests proposed by Vuong (1989) for nonnested hypotheses are especially attractive

---

[3] For studies questioning the normality assumption in estimation of selectivity-adjusted wage equations for developing countries, see Appleton, et al., 1990 and Vijverberg, 1991. For a study questioning the normality assumption in the estimation of female labor supply functions in the U.S., see Newey, et al. (1990).

[4] If easy-to-implement sensitivity analysis were not desired, approaches to specification testing other than the one pursued below might be preferable. For example, specification tests based on score or likelihood ratio tests of moment conditions have the advantage that their small sample properties are better understood (Chesher and Smith, 1997; Chesher and Spady, 1991) than those of the Vuong nonnested likelihood ratio tests employed below.

for selecting among the $F(.)$ specifications in the present context, because they are compatible with the intention of identifying which among several tractable but imperfect models is closest to the unknown true model. The pairwise tests concern the null hypothesis that the two models are equally far from the true model according to the Kullback-Liebler Information Criterion (KLIC), versus the alternatives that one or the other model is closer to the true model. They are also easy to implement.[5] They may be performed assuming either homoscedastic or heteroscedastic specifications.[6] As these and additional tests mentioned below are not independent, relatively high significance levels are called for in choosing critical values.

The assumption of wage offer error normality may be made compatible with any of the three assumptions regarding the shape of $F(.)$, and with the possibility of heteroscedasticity in the selection rule error, by assuming (following Lee, 1982, part 3) that the wage equation error is joint normally distributed (with correlation coefficient ρ) with a transformation of the selection rule error:

$$\eta_i^* = \Phi^{-1}\left(F\left(\frac{\eta_i}{\exp(Z_i^- \delta)}\right)\right), \tag{7}$$

where $F(.)$ denotes the normal, logistic or lognormal cdf. Allowing for non-normality as well as heteroscedasticity, $M(Z_i\gamma)$ in the second stage regression (3) must be replaced by

$$M^n(g(Z_i)) = \frac{\phi(\Phi^{-1}(F(g(Z_i))}{1 - F(-g(Z_i))} \tag{8}$$

---

[5] The statistics may be computed as ((N-1)/N)t, where N is the number of observations and t is the t-statistic on the constant term in the artificial regression of the first model's log-likelihood contribution minus the second model's log-likelihood contribution on a constant term. The statistic is asymptotically normal. Unusually high values indicate model one is closer to the true model, and unusually low values indicate model two is closer to the true model. If the value is between the standard normal critical values, one cannot discriminate between the models in the given data.

[6] The heteroscedastic models are specified such that $\eta_i' = \eta_i / \exp(Z_i^- \delta)$ has either the standard normal, logistic or lognormal distribution.

where again the selection rule index $g(Z_i)$ is $Z_i\gamma/\exp(Z_i^{-}\delta)$. Under the assumption that the wage offer error is normal, both two-stage and maximum likelihood methods produce consistent estimates.

Various two-stage methods yield consistent estimates under weaker assumptions regarding the wage offer equation error, under which maximum likelihood estimation may not be consistent and efficient. The two-stage estimates of the previous paragraph produce consistent estimates as long as the mean of ϵ_i conditional on η^*_i is assumed linear in η^*_i (Lee, 1982). Consistent two-stage estimates are also possible under the alternative assumption that the mean of ϵ_i conditional on the error $\eta_i' = \eta_i/\exp(Z_i^{-}\delta)$ is linear in η_i' (as in Olsen, 1980). This produces the same two-stage method as linearity in η^*_i when $F(.)$ is normal, but different two-stage methods when F(.) is logistic or lognormal. In those cases (allowing also for possible heteroscedasticity), the correction term $M(Z_i\gamma)$, in equation (3) must be replaced by $M^t(g(Z_i))$, where

$$M^t(g(Z_i)) = \frac{1}{\sigma_F}E(\eta_i' \mid \eta_i' > -g(Z_i)),$$

in which the expectation term is the expectation of the truncated $F(\eta_i')$ distribution.[7] Unfortunately, Vuong tests, which are based on likelihood functions, cannot be used to select from among these second stage models the ones that are closest to the true model. Analysis of the sensitivity of wage offer equation estimates across these models employing equally plausible assumptions is, however, useful for identifying coefficient estimates that inspire the most confidence.

More substantial departures from wage offer normality may be introduced following Lee (1982, part 4), by assuming that the standardized wage equation error

$$\epsilon_i^* = \frac{\epsilon_i}{\sigma} \qquad (9)$$

---

[7] For the logistic distribution, $M^t(v) = -[\frac{\ln(v) - vF(-v)}{F(v)}]\frac{\sqrt{(3)}}{\pi}$ and for the lognormal distribution

$$M^t(v) = \frac{\sqrt{(e)}}{\sqrt{(e^2-e)}}[\frac{\Phi(1-\ln(\sqrt{(e)}-v))}{\Phi(-\ln(\sqrt{(e)}-v))} - 1] \text{ (for } v \le \sqrt{(e)}.$$

and the transformed selection rule error η_i^* (as defined in (7)) have the Type AA distribution, which nests the joint normal distribution as a special case. Consistent estimation is possible with two-stage methods including three correction terms. Allowing for heteroscedasticity as well as this, Equation (3) must be replaced by one that contains two additional correction terms:[8]

$$E(w_i|X_i,P_i^*>0)=X_i\beta+\sigma\rho M''(g(Z_i))+\tau_1 M''(g(Z_i))J(g(Z_i))+\tau_2 M''(g(Z_i))(J(g(Z_i))^2-1) \quad (10)$$

where $M''(v)$ is as defined in (8), $J(g(Z_i))=\Phi^{-1}(F(-g(Z_i)))$, and τ_1 and τ_2 involve cross moments of ϵ_i^* and η_i^*. Wald tests (of the joint insignificance of the two additional terms) may be used to test the null hypothesis that $E(\epsilon_i|\eta_i^*)$ is linear in η_i^*, versus the alternative of these more substantial departures from wage offer normality, while maintaining any of the three assumptions regarding $F(.)$.[9]

V. Application to Urban Peru

In this section the selection models, model selection tests and sensitivity analysis suggested above are applied in the estimation of wage offer equations for men and women in urban Peru. Over half of the urban labor force in Peru is self employed, raising the possibility of important selection biases. Peruvian wage equations may be estimated using the 1985-86 Peruvian Living Standards Measurement Survey (Grootaert and Arriagada, 1986), which is similar to high quality datasets of modest size collected by the World Bank in many countries (Grosh and Glewwe, 1995).

---

[8] Though Lee derives this form from the parametric distributional assumptions set out in the text, such equations have been called semiparametric estimators. See Newey, et al. (1990) for an application in which their semiparametric "series approximation" estimator includes only the first two of the three correction terms included in equation (10).

[9] While I have allowed for non-normality in wage offer as well as selection rule errors, I have allowed for heteroscedasticity only in the selection rule errors. Concern about heteroscedasticity in the wage equation error is reduced by being able to transform the dependent variable, using the log of the wage rather than the wage, and by the observation that heteroscedasticity in the linear second stage introduces inefficiency but not inconsistency. Some sorts of heteroscedasticity are "allowed for" in the two-stage estimators by calculating standard errors that allow for the presence of arbitrary heteroscedasticity. Exploratory work using simple forms of wage offer heteroscedasticity that imply differences across groups in the coefficient on the selection correction terms did not find evidence of significant differences.

The model selection tests and sensitivity analysis presented here pertain to a "bare bones" specification of X_i and a rich (by the standards of developing country wage literature) specification of Z_i. X_i contains only years of completed schooling, potential labor market experience and its square, while Z_i contains those variables plus indicators of marital status, numbers of the individual's own children in the 0-9 and 10-14 age ranges, household structure (numbers of children in the household in those age ranges, numbers of adult males and females in the household), household nonlabor income per capita and family background variables (years of education of the individual's father and mother). Detailed variables definitions and descriptive statistics, as well as additional results derived from models with fewer exclusion restrictions, either as the result of adding variables to X_i or eliminating variables from Z_i, are reported in Schaffner (1997). All parameter estimates and standard errors are available from the author.[10]

Specification tests relating to the selection rule are reported in Tables 1, 2 and 3. Table 1 reports the results of likelihood ratio tests that pit the standard "linear homoscedastic" specification (which includes the square of experience, but otherwise incorporates the Z's only linearly) against three more flexible alternatives (in the three sections of the table), and that are performed for each of three maintained assumptions regarding the shape of F(.) (in the three rows of each section of the table).[11] The first alternative maintains the use of only linear terms (plus the square in experience) but allows for heteroscedasticity as in (5). The second and third alternatives maintain the simpler homoscedastic specification, but introduce flexibility into the selection rule index by introducing quadratic terms. The "full quadratic" model includes all non-redundant squares and cross-products of the Z's. This model having 90 parameters, it will not be surprising to find that it appears to be

[10] All results are based on treating I_i as an indicator of whether or not the individual is a wage earner. Whether or not the individual reports a wage is treated as random. Exploratory work suggested that using wage reporting status rather than wage earning status as the selection rule dependent variable has little effect on the results.

[11] Wald tests constructed using misspecification-consistent covariance matrix estimates following White (1982) yield similar results.

closer to the true model than the linear heteroscedastic model with only 27 parameters. To shed light on the relative merits of the functional form associated with heteroscedasticity, the table also includes comparisons with the "parsimonious quadratic" specification, which adds to the linear homoscedastic specification only the quadratic terms that entered the full quadratic specification with t ratios of at least 2.0 (1.8) for males (females) and has numbers of parameters much closer to that of the heteroscedastic specifications. Tables 2 and 3 present the results of Vuong non-nested likelihood ratio tests, which pit the null hypothesis that the two models under the headings Model 1 and Model 2 are equally far from the true model (by the KLIC) versus the alternatives that one or the other model is closer to the true model. Large positive values of the statistic indicate that Model 1 is closer to the true model, while large negative numbers indicate that Model 2 is closer to the true model. Table 2 presents results comparing normal, logistic and lognormal distributional assumptions, while maintaining assumptions regarding possible heteroscedasticity and the possible importance of quadratic terms.[12] Table 3 presents results comparing heteroscedastic and quadratic specifications, in most cases while maintaining the assumptions regarding the nature of F(.). Both because of daunting computational problems with quadratic lognormal specifications, and because testing suggested the superiority of normal and logistic distributional assumptions, quadratic specifications are introduced only in normal and logistic models.

These three tables lead to the following observations regarding the specification of the selection rule:

• The null hypothesis of homoscedasticity is resoundingly rejected in favor of models allowing for heteroscedasticity as in (5), for both males and females, whether F(.) is assumed normal, logistic or lognormal (see the first section of Table 1).

• While non-nested tests lead to the rejection of normality in favor of the logistic assumption,

[12] A fourth specification, also lognormal but allowing a skew in the opposite direction was also attempted. For males, such a model was found inferior to the ones reported by the Vuong test. For females, the model could not be estimated, even when a large number of outliers were eliminated.

for both males and females, when the linear homoscedastic specification is maintained (first section of Table 2), normality is not rejected in heteroscedastic specifications (second section of Table 2).

• When selection rule flexibility is allowed for by adding quadratic terms rather than allowing for the more structural treatment of heteroscedasticity, the standard linear homoscedastic model is again resoundingly rejected (second and third sections of Table 1).

• In the quadratic specifications, normality is rejected in favor of the logistic distribution (last two sections of Table 2).

• Non-nested tests indicate that whether F(.) is assumed normal or logistic, the full quadratic model (with 90 parameters) is closer to the true model than the heteroscedastic model (with 27 parameters) (rows 1 and 3 of Table 3).

• When the heteroscedastic models are pitted against parsimonious quadratic models, however, the heteroscedastic models appear no worse, and in some cases slightly better, than the quadratic specifications (rows 2 and 4 of Table 3).[13]

The good performance of the heteroscedastic formulation relative to a quadratic with similar number of parameters lends some credence to the notion that the important nonlinearities arise out of heteroscedasticity. But the introduction of quadratic terms into homoscedastic specifications appears to provide for an easy and more comprehensive treatment of the nonlinearities.

The remaining tables pertain to estimation of male and female wage offer equations. As discussed above, each choice of selection rule error distribution F(.), and each choice of linear homoscedastic, heteroscedastic, or quadratic selection rule specification, may be combined with one of four assumptions regarding the wage offer equation error ϵ_i (where ϵ_i^*, η_i^* and η_i' are defined in (9), (7) and (6)): (1) ϵ_i^* is joint normally distributed with η_i^*, (2) $E(\epsilon_i^*|\eta_i^*)$ is linear in η_i^*, (3) $E(\epsilon_i^*|\eta_i')$ is linear in η_i', and (4) ϵ_i^* has a Type AA distribution η_i^*. Table 4 reports the results of Wald tests of a null hypothesis consistent with the first or second assumptions, which are the most standard and

[13] The last row of Table 3 is included because the previous table shows that if a heteroscedastic specification is maintained, one would select a normal distribution for F(.), while if a quadratic specification is maintained, one would select a logistic distribution for F(.).

require the inclusion of only one selection correction term, versus an alternative hypothesis associated with the fourth assumption, which allows for the most flexible departures from normality and calls for the inclusion of the three selection correction terms in equation (10). They are calculated using estimates of the covariance matrix that are consistent in the presence of both pre-estimated regressors and heteroscedasticity. Table 4 leads to the observations that:

- When linear homoscedastic selection rule specifications are maintained, the null hypothesis that the additional selection correction terms are unnecessary cannot be rejected for either males or females, and whether F(.) is assumed normal or logistic (rows 1 and 4 of Table 4).

- However, once flexibility is introduced into the estimation of the selection rule with either heteroscedastic or full quadratic specifications, the null hypothesis is rejected in favor of models incorporating three selection correction terms for males (rows 2, 3, 5 and 6 of Table 4).[14]

Tables 5, 6 and 7 illustrate the sensitivity of wage offer equation estimates to changes in the choice of F(.) (seen by comparing corresponding cells across top and bottom sections of the table), choice of homoscedastic, heteroscedastic and full quadratic selection rule specifications (seen by comparing columns within rows and sections), and the choice of assumptions regarding the distribution of the wage offer equation error (seen by comparing rows within columns and sections). Estimates of standard errors for the two-stage (2S) estimates, which are consistent in the face of pre-estimated regressors and heteroscedasticity, are calculated following Heckman and MaCurdy (1986, p. 1944). Heteroscedasticity-consistent estimates of the standard errors for maximum likelihood (ML) estimates are calculated following White (1982). They are nearly identical to standard estimates of ML standard errors (employing analytical second derivatives) for the schooling coefficient (Table 5), but tend to be fifty percent larger than those estimates for the covariance parameter estimates (Table 6) in the linear homoscedastic selection rule models, and for most

[14] Wald tests of the joint significance of three additional correction terms $(M^2, M^2J, M^2(J^2-1))$ failed to reject the null at any level of significance. Thus there is no evidence that additional departures from normality should be introduced.

estimates of male-female log wage differences (Table 7).

Table 5 presents estimates of the schooling effect, which are often interpreted as the rate of return on investments in additional years of schooling and are considered important for the assessment of education policy. The table leads to the following observations:

- Point estimates of the effect of schooling on log wages are highly robust to changes in statistical assumptions, especially for men. Selectivity adjusted estimates for men vary across statistical models in the range .101 to .112, all only slightly lower than the OLS estimate of .116.

- Selectivity-adjusted estimates for women are somewhat less robust, but still vary only from .110 to .137 across all selectivity bias correction models.

The statistical specification of the selectivity model does not appear important to the estimation of the schooling coefficient. In contrast, Schaffner (1997) demonstrates that estimates of this coefficient are not robust to changes in X_i, especially to the inclusion of family background variables, which brings the estimated returns to schooling down significantly.

Table 6 presents estimates of $\sigma\rho$, the covariance between the selection rule and wage offer equation errors. Estimation of this term and its standard error is considered important both for assessing the significance of the selectivity problem and for shedding light on the nature of labor markets. For example, Gindling (1991) draws from the significance of this term conclusions about whether selection into wage employment primarily reflects the whims of premium-paying employers segmented labor markets, or the income-enhancing choices of workers in perfectly competitive markets. Table 6 generates the following observations:

- Point estimates of $\sigma\rho$ and related tests of the significance of selectivity effects are much less robust to changes in statistical assumptions than estimates of the schooling effect on log wages. Point estimates range from -2.207 to +.633 for men and from -.607 to +.329 for women.

- Introducing flexibility into selection rule estimation, using either the heteroscedastic or full quadratic models, renders estimation of this coefficient much more robust in sign and statistical significance. All point estimates in these models are negative for both men and women. The null hypothesis of no selectivity effects is always rejected in heteroscedastic and quadratic specifications for men and never rejected for women.[15]

- Among heteroscedastic selection rule specifications, the only modification to the standard statistical assumptions that has a large impact on the estimate of $\sigma\rho$ is the inclusion of the two additional selection correction terms that become relevant in models that allow for a Type AA distribution of ϵ_i^* and η_i^*. Adding these terms renders the point estimate of $\sigma\rho$ much more negative.

Estimates produced by heteroscedastic and full quadratic selection rule models inspire more confidence, being more robust across the other statistical assumptions. A strong indication of the inadequacy of the linear homoscedastic selection rule models are the large differences in point estimates across ML and two-stage estimates that should both provide consistent estimates under the same sets of statistical assumptions. Schaffner(1997) shows that the beneficial effect of selection rule flexibility on the robustness of estimation appears even in models in which X and Z are identical, and thus in which identification rests entirely on functional form assumptions.

Finally, Table 7 presents estimated male-female log wage offer differences for individuals with two sets of observed characteristics: those with six years of schooling (primary complete) and no potential labor market experience, and those with six years of schooling and ten years of potential labor market experience.[16] In order to illustrate the range of predicted wage differences that researchers would produce when maintaining each set of statistical assumptions indicated by column, row and section of the table, the predicted differences are based on predicted log wages for males (females) from models containing the relevant selectivity correction terms when hypothesis tests reject the null hypothesis of no selectivity effects for males (females), and from OLS when such hypothesis

[15] The coefficient may be better identified in the heteroscedastic and full quadratic models as a result of reduced multicollinearity between the M(.) term and the other wage offer equation regressors. The R-squared for regression of the M(.) terms on schooling, experience, experience squared and a constant are .88, .64 and .53 (.76, .66 and .57) in the homoscedastic, heteroscedastic and full quadratic models for males (females) when F(.) is normal.

[16] Predicted differences for individuals with additional sets of characteristics are available from the author. Mean years of schooling in the entire sample (wage earning subsample) are 9.5 (10.1) years for men and 8.3 (10.9) years for females.

tests fail to reject the null. Selectivity adjustments are applied in all but five cases for males and in no cases for females. Table 7 generates the following observations.

- Estimated male-female log wage differences vary tremendously as the statistical assumptions vary. Most of the estimated differences are also unbelievably large.[17]

- The range of estimated log wage differences is somewhat smaller for models allowing for selection rule heteroscedasticity or for full quadratic selection rules, but it is still large and the point estimates are very high.

- Under all sets of statistical assumptions estimated male-female wage differences fall as potential labor market experience rises.[18]

Estimates of the level of male-female wage differences do not inspire confidence.[19] Researchers may need to focus instead on what they can learn about gender in labor markets from the dependence of estimated differences on observed characteristics.

VI. Conclusion

Even without abandoning parametric, easy-to-implement methods, researchers can take a much more flexible approach to the estimation of wage equations in endogenously selected samples, allowing for departures from three standard statistical assumptions: normality of the selection rule error, normality of the wage equation error, and homoscedasticity of the selection rule error. Selection rule error normality, and to a lesser extent, wage offer equation error normality, have been the subject of considerable attention in the literature. This paper emphasizes departures from the

---

[17] For example, maximum likelihood estimation of selectivity adjusted wage equations under the most standard assumptions (of homoscedastic selection rule and joint normality of the errors) produces an estimated difference in log wages of 1.709 for individuals with six years of schooling and no experience, implying a wage level difference of 450 percent. Two-stage estimation employing the same statistical assumptions leads to an estimated difference in log wages of .310, or a percent wage difference of 36 percent.

[18] Additional calculations (available from the author) indicate that under most sets of assumptions estimated differences vary little by level of schooling.

[19] The instability of the male-female wage difference estimates seems to stem largely from the instability of the regression intercept estimate, which is not unique to this application. Vella (1988) presents results on estimated U.S. black-white wage differences derived from selectivity-adjusted models. He finds the slope coefficients stable and the intercept estimate unstable across the two selection models he considers.

assumption of homoscedasticity in the selection rule error, which may be much more important but has received very little attention in the literature. The paper provides economic reasons to suspect selection rule heteroscedasticity and econometric reason to believe that the nonlinearities introduced by heteroscedasticity into selection rule estimation might aid identification, and suggests ways of introducing it into parametric models that admit two-stage estimation. In the application to urban Peru, the hypothesis of selection rule homoscedasticity was resoundingly rejected in favor of more flexible alternatives, and allowing for selection rule heteroscedasticity appeared more important than allowing for selection rule non-normality. Wage offer equation non-normality also appears important for men, but evidence for this would have been missed if attention had been restricted to models assuming selection rule homoscedasticity.

Happily for researchers interested in estimating the returns to schooling in developing countries, none of these statistical considerations appear to matter for estimates of the return to schooling in urban Peru. Given the maintained assumptions regarding which variables belong in the selection rule and the wage offer equations, estimates of the ceteris paribus effect of years of schooling on log wages vary between .101 to .111 for males and in the slightly wider range of .110 to .137 for females as the statistical assumptions are varied, the ranges lying just below the OLS estimates. The ranges are even narrower among the preferred models, which allow for selection rule heteroscedasticity. This result frees researchers to focus more attention on the specification of what belongs on the right hand side of the wage equation, which does have a large impact on schooling coefficient estimates (Schaffner, 1997)

Statistical specification matters more for estimation of the covariance between selection rule and wage offer equation errors. Among models assuming homoscedastic selection rule errors, variation in assumptions regarding selection rule and wage offer equation normality causes covariance point estimates to swing from large positive to large negative values, and leads to very different conclusions about the hypothesis that there is no selectivity effect. Among models allowing for selection rule heteroscedasticity, all covariance estimates for males and females are negative, statistically significantly different from zero for men but not for women.

Even among the preferred models, which allow for selection rule heteroscedasticity, male-female log wage difference estimates are discouragingly variable and often of unbelievable magnitude. This leads to two conclusions. First, estimated gender differences derived from selectivity-adjusted wage equations should be greeted with skepticism. Second, while allowance for greater flexibility in selection rule estimation is of great practical utility for research on developing country labor markets, it is not a panacea.

References

Alderman, H. and V. Kozel (1989), "Formal and Informal Sector Wage Determination in Urban Low-Income Neighborhoods in Pakistan," Living Standards Measurement Study Working Paper No. 65, World Bank.

Appleton, S., P. Collier and P. Horsnell (1990), "Gender, Education and Employment in Côte d'Ivoire," Social Dimensions of Adjustment in Sub-Saharan Afric Working Paper No. 8, World Bank.

Ashraf, J. and B. Ashraf (1993), "Estimating the Gender Wage Gap in Rawalpindi City," Journal of Development Studies 29(2), pp.365-376.

Blau, F. D. and A. H. Beller (1988), "Trends in Earnings Differentials by Gender, 1971-1981," Industrial and Labor Relations Review 41(4), pp.513-529.

Chesher, A. and R. J. Smith (1997), "Likelihood Ratio Specification Tests," Econometrica 65(3), pp.627-646.

Chesher, A. and R. Spady (1991), "Asymptotic Expansions of the Information Matrix Test Statistic", Econometrics 59(3), pp787-815.

Deolalikar, A. (1993), "Gender Differences in the Returns to Schooling and School Enrollment Rates in Indonesia," Journal of Human Resources 28(4), pp.899-932.

Gindling, T. H. (1991), "Labor Market Segmentation and the Determination of Wages in the Public, Private-Formal and Informal Sectors in San José, Costa Rica," Economic Development and

Cultural Change 39(3), pp.585-606.

Glewwe, P. (1991), "Schooling, Skills and the Returns to Education: An Econometric Exploration Using Data from Ghana," Living Standards Measurement Study Working Paper No. 76, World Bank.

Goldberger, A. (1983), "Abnormal Selection Bias," in S. Karlin, T. Amemiya, and L.A. Goodman (eds.), Studies in Econometrics, Time Series and Multivariate Statistics, Academic Press, pp.67-84.

Grootaert, C. and A.-M. Arriagada (1986), "The Peruvian Living Standards Survey: An Annotated Questionnaire," World Bank, processed.

Grosh, M. and P. Glewwe (1995), "A Guide to Living Standards Measurement Study Surveys and Their Data Sets," World Bank, processed.

Heckman, J. (1979), "Sample Selection as a Specification Error," Econometrica 47(1), pp.153-61.

Heckman, J. J. and T. E. MaCurdy (1986), "Labor Econometrics," Chapter 32 in Z. Griliches and M. D. Intriligator (eds.), Handbook of Econometrics, Volume 3, pp.1917-1977.

Hurd, M. (1979), "Estimation in Truncated Samples When There is Heteroskedasticity," Journal of Econometrics 11, pp.247-258.

Khandker, S. R. (1991), "Labor Market Participation, Returns to Education and Male-Female Wage Differences in Peru," in B. K. Herz and S. R. Khandker (eds.), "Women's Work, Education, and Family Welfare in Peru, World Bank Discussion Papers No. 116, pp.37-66.

King, E. (1990), "Does Education Pay in the Labor Market? The Labor Force Participation, Occupation and Earnings of Peruvian Women," Living Standards Measurement Study Working Paper No. 67, World Bank.

Lee, L.F. (1982), "Some Approaches to the Correction of Selectivity Bias," Review of Economic Studies 49, pp.355-372.

Leung, S.F. and S. Yu (1996), "On the Choice Between Sample Selection and Two-Part Models," Journal of Econometrics 72, pp.197-229.

Magnac, Th. (1991), "Segmented or Competitive Labor Markets?" Econometrics 59(1), pp.165-187.

Mroz, T. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Econometrica 55(4), pp.765-799.

Nawata, K. (1994), "Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator," Economics Letters 45, pp.33-40.

Nelson, F. D. (1984), "Efficiency of the Two-Step Estimator for Models with Endogenous Selection," Journal of Econometrics 24, p.181-196.

Newey, W. K., J. L. Powell, and J. R. Walker (1990), "Semiparametric Estimation of Selection Models: Some Empirical Results," American Economic Review 80(2), pp.324-328.

Olsen, R. J. (1980), "A Least Squares Correction for Selectivity Bias," Econometrica 48(7), pp.1815-1820.

Sahn, D. E. and H. Alderman (1988), "The Effects of Human Capital on Wages, and the Determinants of Labor Supply in a Developing Country," Journal of Development Economics 29)2), pp.157-184.

Schaffner, J. A. (1997), "The Sensitivity of Wage Equation Estimates for a Developing Country to Changes in Sample Selection Model Specification," Stanford University, processed.

Strauss, J. and D. Thomas (1991), "Wages, Schooling and Background: Investments in Men and Women in Urban Brazil," Yale Economic Growth Center Working Paper no. 649.

Sumner, D.A. (1981), "Wage Functions and Occupational Selection in a Rural Less Developed Country Setting," Review of Economics and Statistics 63(4), pp.513-519.

Tannen, M. (1991), "Labor Markets in Northeast Brazil: Does the Dual Market Model Apply?" Economic Development and Cultural Change 39(43), pp.567-584.

Vella, F. (1988), "Generating Conditional Expectations from Models with Selectivity Bias," Economics Letters 28, pp.97-974.

Vijverberg, W. (1993), "Educational Investments and Returns for Women and Men in Côte d'Ivoire," Journal of Human Resources 23(4), pp.933-974.

Vijverberg, W. (1991), "Selectivity and Distributional Assumptions in Static Labor Supply Models," Southern Economic Journal 57(3), pp.822-840.

Vuong, Q. H. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," Econometrics 57(2), pp.307-333.

Wellington, A. J. (1993), "Changes in the Male/Female Wage Gap, 1976-85," Journal of Human Resources 28(2), pp.383-411.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," Econometrica 50(1), pp.1-25.

Table 1

Likelihood Ratio Tests of Linear Homoscedastic Specifications ($\delta = \gamma^a = \gamma^b = 0$)
Versus Three Alternatives
in Selection Rule Models of the Form*

$$I_i* = Z_i\gamma + Z_i^a\gamma^a + Z_i^b\gamma^b + \eta_i, \quad \frac{\eta_i}{\exp(Z_i^- \delta)} \sim F(.)$$

	Males			Females		
	χ^2 Stat.	d.f.	P-Value	χ^2 Stat.	d.f.	P-Value
Alternative Hypothesis of Heteroscedastic Specification ($\delta \neq 0, \gamma^a = \gamma^b = 0$)						
F(.) Normal	142.28	13	.000	116.56	13	.000
F(.) Logistic	130.98	13	.000	97.28	13	.000
F(.) Lognormal**	136.54	13	.000	117.52	13	.000
Alternative Hypothesis of Full Quadratic Specification ($\delta = 0, \gamma^a \neq 0, \gamma^b \neq 0$)						
F(.) Normal	219.22	76	.000	170.86	76	.000
F(.) Logistic	214.74	76	.000	165.64	76	.000
Alternative Hypothesis of Parsimonious Quadratic Specification ($\delta = 0, \gamma^a \neq 0, \gamma^b = 0$)						
F(.) Normal	135.88	17	.000	81.32	13	.000
F(.) Logistic	135.54	18	.000	77.56	11	.000

* Z_i, Z_i^a, and Z_i^b are mutually exclusive and exhaustive subsets of the variables included in the full quadratic specification. Z_i contains only linear terms plus the square of experience, Z_i^a contains the additional quadratic terms that entered with t statistics above 2 (1.8) for males (females) in the full quadratic specification. Z_i^b contains the remaining quadratic terms.

** Tests performed on samples excluding two (eight) outliers that created computational problems in the lognormal model for males (females).

Table 2
Vuong Non-Nested Likelihood Ratio Tests Comparing Distributional Assumptions
for Models of the Form*

$$I_{\varnothing}^{*} = Z_{\varnothing}\gamma + Z_{\varnothing}^{`}\gamma^{`} + Z_{\varnothing}^{\prime}\gamma^{\prime} + \eta_{\varnothing} \quad \frac{\eta_{\varnothing}}{\exp(Z_{\varnothing}^{-}\delta)} \sim F(.)$$

Model 1	Model 2	Males		Females	
		N(0,1) Statistic	P-Value	N(0,1) Statistic	P-Value
Comparisons of Distributional Assumptions in Homoscedastic Specifications ($\delta = 0$, $\gamma^{`} = 0$, $\gamma^{\prime} = 0$)**					
F(.) Normal	F(.) Logistic	-3.067	.002	-2.976	.001
F(.) Normal	F(.) Lognormal	.535	.296	.003	.499
F(.) Logistic	F(.) Lognormal	1.430	.076	1.266	.103
Comparisons of Distributional Assumptions in Heteroscedastic Specifications ($\delta \neq 0$, $\gamma^{`} = 0$, $\gamma^{\prime} = 0$)***					
F(.) Normal	F(.) Logistic	1.236	.108	.746	.228
F(.) Normal	F(.) Lognormal	.942	.173	-.067	.473
F(.) Logistic	F(.) Lognormal	.860	.195	.134	.467
Comparison of Distributional Assumption in Full Quadratic Specifications ($\delta = 0$, $\gamma^{`} \neq 0$, $\gamma^{\prime} \neq 0$)					
F(.) Normal	F(.) Logistic	-1.377	.084	-2.980	.001
Comparisons of Distributional Assumptions in Parsimonious Quadratic Specifications ($\delta = 0$, $\gamma^{`} \neq 0$, $\gamma^{\prime} = 0$)****					
F(.) Normal	F(.) Logistic	-2.080	.019	-2.907	.002

* See notes to Table 1 for definitions of Z, $Z_{\varnothing}^{`}$ and $Z_{\varnothing}^{\prime}$.
** Test for women performed on sample excluding one outlier that created computational problems in the lognormal model.
*** Tests performed on samples excluding two (eight) outliers that created computational problems in the lognormal model for males (females).

Table 3
Vuong Non-Nested Likelihood Ratio Tests Comparing Heteroscedastic and Quadratic Specifications for Models of the Form*

$$I_{\varnothing}^{*} = Z_{\varnothing}\gamma + Z_{\varnothing}^{`}\gamma^{`} + Z_{\varnothing}^{\prime}\gamma^{\prime} + \eta_{\varnothing} \quad \frac{\eta_{\varnothing}}{\exp(Z_{\varnothing}^{-}\delta)} \sim F(.)$$

Model 1	Model 2	Males		Females	
		N(0,1) Statistic	P-Value	N(0,1) Statistic	P-Value
Comparisons of Heteroscedastic and Quadratic Specifications in Models with F(.) Normal					
Heteroscedastic $\delta \neq 0, \gamma^{`} = 0, \gamma^{\prime} = 0$	Full Quadratic $\delta = 0, \gamma^{`} \neq 0, \gamma^{\prime} \neq 0$	-2.586	.005	-2.025	.031
Heteroscedastic $\delta \neq 0, \gamma^{`} = 0, \gamma^{\prime} = 0$	Pars. Quadratic $\delta = 0, \gamma^{`} \neq 0, \gamma^{\prime} = 0$	.234	.408	1.656	.049
Comparisons of Heteroscedastic and Quadratic Specifications in Models with F(.) Logistic					
Heteroscedastic $\delta \neq 0, \gamma^{`} = 0, \gamma^{\prime} = 0$	Full Quadratic $\delta = 0, \gamma^{`} \neq 0, \gamma^{\prime} \neq 0$	-2.718	.003	-2.634	.004
Heteroscedastic $\delta \neq 0, \gamma^{`} = 0, \gamma^{\prime} = 0$	Pars. Quadratic $\delta = 0, \gamma^{`} \neq 0, \gamma^{\prime} = 0$	-.161	.436	1.027	.152
Comparison of Heteroscedastic Normal with Parsimonious Quadratic Logistic					
F(.) Normal Heteroscedastic $\delta \neq 0, \gamma^{`} = 0, \gamma^{\prime} = 0$	F(.) Logistic Pars. Quadratic $\delta = 0, \gamma^{`} \neq 0, \gamma^{\prime} = 0$	-.111	.445	.953	.170

* See notes to Table 1 for definitions of Z, $Z_{\varnothing}^{`}$ and $Z_{\varnothing}^{\prime}$.

Table 4

Wald Tests of $E(\epsilon_i | \eta_i^*) = \rho\sigma\eta_i^*$

For Models of the Form*

$$I_i^* = Z_i\gamma + Z_i^q\gamma^q + \eta_i, \quad \frac{\eta_i}{\exp(Z_i^{-}\delta)} \sim F(.), \quad w_i = X_i\beta + \epsilon_i, \quad \epsilon_i^*, \eta_i^* \sim TypeAA,$$

$$\text{where } \epsilon_i^* = \epsilon_i/\sigma_\epsilon \text{ and } \eta_i^* = \Phi^{-1}\left(F\left(\frac{\eta_i}{\exp(Z_i^{-}\delta)}\right)\right)$$

Maintained Assumptions	Males		Females	
	χ^2 Stat.	P-Value	χ^2 Stat.	P-Value
Specifications Assuming F(.) Normal				
Linear Homoscedastic $\delta = \gamma^q = 0$	1.043	.594	0.483	.786
Heteroscedastic $\delta \neq 0, \gamma^q = 0$	9.415	.009	2.680	.262
Full Quadratic $\delta = 0, \gamma^q \neq 0$	7.178	.028	3.108	.211
Specifications Assuming F(.) Logistic				
Linear Homoscedastic $\delta = \gamma^q = 0$	0.930	.628	1.213	.545
Heteroscedastic $\delta \neq 0, \gamma^q = 0$	4.282	.118	2.803	.246
Full Quadratic $\delta = 0, \gamma^q \neq 0$	5.247	.073	4.794	.09

* Z_i and Z_i^q are mutually exclusive and exhaustive subsets of the variables included in the full quadratic specification. Z_i contains only the linear terms plus the square of experience. Z_i^q contains the remaining quadratic terms.

Table 5

Estimated Schooling Coefficients (Standard Errors in Parentheses)

For Models of the Form

$$I_i^* = Z_i\gamma + Z_i^q\gamma^q + \eta_i,\ \eta_i' = \frac{\eta_i}{\exp(Z_i^-\delta)} \sim F(.),\ w_i = X_i\beta + \epsilon_i,\ where\ \epsilon_i^* = \frac{\epsilon_i}{\sigma_\epsilon}\ \text{and}\ \eta_i^* = \Phi^{-1}(F(\frac{\eta_i}{\exp(Z_i^-\delta)}))$$

Case	Method	Males			Females		
		Homoscedastic $\delta=\gamma^q=0$	Heteroscedastic $\delta\neq0,\gamma^{\otimes}=0$	Full Quadratic $\delta=0,\gamma^{\otimes}\neq0$	Homoscedastic $\delta=\gamma^q=0$	Heteroscedastic $\delta\neq0,\gamma^{\otimes}=0$	Full Quadratic $\delta=0,\gamma^{\otimes}\neq0$
No Selectivity	OLS	.116 (.005)			.115 (.007)		
Selectivity with F(.) Normal							
ϵ^*,η^* joint normal	ML	.102 (.006)	.105 (.007)	.102 (.006)	.137 (.024)	.111 (.013)	.110 (.012)
$E(\epsilon\vert\eta^*)$ linear	2S	.111 (.007)	.108 (.006)	.109 (.006)	.129 (.012)	.112 (.010)	.111 (.010)
ϵ^*,η^* Type AA	2S	.112 (.008)	.103 (.007)	.108 (.006)	.129 (.012)	.112 (.010)	.111 (.010)
Selectivity with F(.) Logistic							
ϵ^*,η^* joint normal	ML	.101 (.006)	.105 (.006)	.102 (.006)	.131 (.023)	.112 (.012)	.111 (.012)
$E(\epsilon\vert\eta^*)$ linear	2S	.108 (.007)	.108 (.006)	.109 (.006)	.125 (.012)	.112 (.010)	.112 (.010)
$E(\epsilon\vert\eta')$ linear	2S	.109 (.007)	.109 (.006)	.109 (.006)	.127 (.012)	.114 (.010)	.114 (.010)
$\epsilon*,\eta*$ Type AA	2S	.109 (.008)	.103 (.007)	.108 (.006)	.127 (.012)	.112 (.011)	.112 (.010)

* See note to Table 4 for definitions of Z and $Z_{\oslash}^{\otimes}$.

Table 6

Estimates of $\sigma\rho$ (Standard Errors in Parentheses)

For Models of the Form

$$I_i^* = Z_i\gamma + Z_i^q\gamma^q + \eta_i,\ \eta_i' = \frac{\eta_i}{\exp(Z_i^-\delta)} \sim F(.),\ w_i = X_i\beta + \epsilon_i,\ where\ \epsilon_i^* = \frac{\epsilon_i}{\sigma_\epsilon} \text{ and } \eta_i^* = \Phi^{-1}(F(\frac{\eta_i}{\exp(Z_i^-\delta)}))$$

Case	Method	Males			Females		
		Homoscedastic $\delta=\gamma^q=0$	Heteroscedastic $\delta\neq 0,\gamma^{\otimes}=0$	Full Quadratic $\delta=0,\gamma^{\otimes}\neq 0$	Homoscedastic $\delta=\gamma^a=\gamma^b=0$	Heteroscedastic $\delta\neq 0,\gamma^{\otimes}=0$	Full Quadratic $\delta=0,\gamma^{\otimes}\neq 0$
No Selectivity	OLS	n.a.			n.a.		
Selectivity with F(.) Normal							
ϵ^*,η^* joint normal	ML	-.509 (.110)	-.387 (.129)	-.477 (.110)	.329 (.343)	-.050 (.152)	-.062 (.113)
$E(\epsilon\vert\eta^*)$ linear	2S	-.168 (.211)	-.326 (.117)	-.254 (.091)	.205 (.151)	-.036 (.114)	-.044 (.092)
ϵ^*,η^* Type AA	2S	.633[a] (.854)	-2.067[b] (.930)	-1.051[b] (.376)	-.204[a] (.666)	-.357[a] (.745)	-.520[a] (.365)
Selectivity with F(.) Logistic							
ϵ,η^* joint normal	ML	-.516 (.103)	-.394 (.124)	-.484 (.107)	.238 (.326)	-.046 (.149)	-.057 (.135)
$E(\epsilon\vert\eta^*)$linear	2S	-.278 (.208)	-.332 (.140)	-.248 (.090)	.151 (.141)	-.038 (.113)	-.040 (.085)
$E(\epsilon\vert\eta')$ linear	2S	-.285 (.211)	-.303 (.114)	-.242 (.092)	.164 (.136)	-.015 (.106)	-.018 (.082)
$\epsilon*,\eta*$ Type AA	2S	.487[a] (.889)	-2.207[b] (2.248)	-.921[b] (.392)	-.367[a] (.742)	-.385[a] (.744)	-.607[a] (.404)

[*] See note to Table 4 for definitions of Z and $Z_{\varnothing}^{\otimes}$.

[a] A Wald test of the null hypothesis of no selectivity effects (zero coefficients on three selection correction terms) fails to reject the null.

[b] A Wald test of the null hypothesis of no selectivity effects (zero coefficients on three selection correction terms) rejects the null.

Table 7

Estimates of Male Minus Female Log Wage Offers (Standard Errors in Parentheses)

For Models of the Form

$$I_i^* = Z_i\gamma + Z_i^a\gamma^q + \eta_i,\ \eta_i' = \frac{\eta_i}{\exp(Z_i^-\delta)} \sim F(.),\ w_i = X_i\beta + \epsilon_i,\ where\ \epsilon_i^* = \frac{\epsilon_i}{\sigma_\epsilon}\ \text{and}\ \eta_\varnothing^* = \Phi^{-1}(F(\frac{\eta_\varnothing}{\exp(Z_\varnothing^-\delta)}))$$

Case	For Individuals with School=6,Experience=0			For Individuals with School=6,Experience=10		
	Homoscedastic $\delta=\gamma^q=0$	Heteroscedastic $\delta\neq 0,\gamma^{\otimes}=0$	Full Quadratic $\delta=0,\gamma^{\otimes}\neq 0$	Homoscedastic $\delta=\gamma^q=0$	Heteroscedastic $\delta\neq 0,\gamma^{\otimes}=0$	Full Quadratic $\delta=0,\gamma^{\otimes}\neq 0$
No Selectivity	.310 (.109)			.143 (.063)		
Selectivity with F(.) Normal						
ϵ^*,η^* joint normal	1.071 (.199)	.855 (.217)	.978 (.190)	.672 (.127))	.533 (.142)	.618 (.123)
$E(\epsilon\|\eta^*)$ linear	.310 (.109)	.748 (.197)	.661 (.163)	.143 (.063)	.458 (.015)	.394 (.109)
ϵ^*,η^* Type AA	.310 (.109)	1.877 (.586)	1.072 (.251)	.143 (.063)	1.660 (.586)	.867 (.231)
Selectivity with F(.) Logistic						
ϵ^*,η^* joint normal	1.083 (.190)	.846 (.210)	.990 (.187)	.676 (..119)	.527 (.137)	.623 (.121)
$E(\epsilon\|\eta^*)$ linear	.310 (.109)	.753 (.193)	.652 (.165)	.143 (.063)	.461 (.127)	.386 (.108)
$E(\epsilon\|\eta')$ linear	.310 (.109)	.708 (.190)	.638 (.165)	.143 (.063)	.425 (.124)	.372 (.107)
$\epsilon*,\eta*$ Type AA	.310 (.109)	1.878 (1.327)	.992 (.262)	.143 (.063)	1.660 (1.342)	.786 (.240)

* See note to Table 4 for definitions of Z and $Z_\varnothing^{\otimes}$