# Identification and Estimation of the Triangular Simultaneous Equations Model in the Absence of Exclusion Restrictions Through the Presence of Heteroskedasticity

Roger Klein
Rutgers University

Francis Vella
European University Institute

Preliminary and Incomplete

**Abstract**

This paper provides a control function type estimator to adjust for endogeneity in the triangular simultaneous equations model where there are no available exclusion restrictions to generate suitable instruments. Our approach is to exploit the presence of heteroskedasticity in the model to adjust the conventional control function estimator. In the presence of heteroskedasticity this adjustment takes a non-linear form and this provides the necessary source of identification. We are able to provide such an estimator without any additional assumptions about the heteroskedasticity other than it takes a multiplicative form. In addition to providing the estimator and deriving its properties we present some simulation evidence which indicates the estimator works well. We also provide an empirical example which investigates the return to education.

## 1 Introduction

The estimation of linear models with endogenous regressors via the use of instrumental variables (IV) is one of the most commonly employed methodologies in empirical economic investigations. However, while there is general agreement that IV is the appropriate form of estimation for a large class of models with endogeneity there is frequently very little agreement about the exclusion restrictions which are imposed in specific empirical applications. In fact, so relatively infrequently is there an obvious exclusion restriction which can be employed, that many papers are now equally motivated by the availability of such a restriction as much as by an interest in the substantial issue under investigation. Moreover, in an attempt to avoid the criticism associated with the use of instruments which can not be justifiably excluded from the conditional mean of the dependent variable of interest there are frequently cases where the search for such exclusions has led to the use of instruments which violate the other condition for instrument validity, namely that the instrument is sufficiently correlated with the endogenous regressor. This has lead to a large and growing literature related to the use of weak instruments.

It is well known that IV can be implemented in a number of ways. One method is to employ the residual from the reduced form from the endogenous regressor as an additional variable in the primary equation of interest. This approach is commonly referred to as a control function procedure. Irrespective of how it is implemented, however, the presence of endogeneity is generally tackled via statements about the conditional means of the endogenous regressors. For example, the control function procedure operates by subtracting the appropriate component of the reduced form error from the primary equation. The quantity that is deducted clearly is reflected by the value of the control function and its estimated coefficient where the latter captures the correlation between the reduced form and primary equation error normalized by the variance of the reduced form error. When the errors are homoskedastic the mapping of one error to the other is constant and the variation in the quantity deducted across observations depends purely on the value of the control function. Moreover as the control function is a linear function of the endogenous regressor and the explanatory variables the model is only identified in the presence of an exclusion restriction. However, consider the case where the errors in each equation are heteroskedastic. The quantity now deducted still relies on the correlation between the errors and the variance of the reduced form error but the values of the correlation and variance depend on whereabouts in the data they are evaluated. If one can isolate the form of the heterosedasticity one is then able to estimate the constant mapping between the homoskedastic errors by appropriately adjusting the value of the control function. Thus the model is now identified even in the absence of exclusion restrictions. A model with heteroskedasticity suggests that the unobservables in the reduced form are "priced" differently in the primary equation depending on whereabouts in the sample space they are evaluated. The ability to identify this mapping provides a form of model identification.

This logic indicates that a potential way to identify models is to make some assumptions about the variances and covariances of the errors. Clearly, and as is shown below, it is relatively simple to estimate the form of heteroskedasticity in the reduced form equation without making particularly strong assumptions. However, in order to make the appropriate form of adjustment to the control function it is necessary to estimate the variance function from the primary equation and this is more difficult given that one cannot immediately estimate the primary equation residual. Thus, in previous attempts to exploit heteroskedasticity as an identification device it has generally been necessary to state that the heteroskedasticity in the primary equation has specific and somewhat undesirable properties (see Vella and Verbeek 199? and Rummery, Vella and Verbeek 199?) or that the form of the heteroskedasticity is known up to some unknown parameters (see Rigobon 2003). One exception is Klein and Vella (2003) which examines the identification of a binary treatment effect in models with heteroskedasticity.

In this paper we derive a consistent estimator for the parameters of the triangular model in the absence of exclusion restrictions and in the presence of multiplicative heteroskedasticity. We do so by making relatively minimal assumptions about the form of the heteroskedasticity. We also allow the heteroskedasticity in each equation to be unknown functions of the same exogenous variables which appear in the conditional means of the equations. The approach of the estimator is based on the logic discussed in the previous paragraph. That is, we develop a control function estimator which is adjusted for the presence of heteroskedasticity. To do so we show we are able to estimate the form of heteroskedasticity in each of the equations and make the appropriate adjustments.

In the following section we outline the model in which we are interested. In sections 3 and 4 we discuss estimation strategies and how they might be implemented. Formal results are contained in section 5. Section 6 provides some simulation evidence and section 7 provides an empirical investigation. Concluding comments are contained in section 8. We note that this draft is preliminary and incomplete. The Appendix, which contains all proofs, will be included in the final draft.[1]

# 2 Model

Consider the following triangular model

$$\begin{aligned} Y_{1i} &= X_i\beta + \theta Y_{2i} + u_i & (1) \\ Y_{2i} &= X_i\pi + v_i & (2) \\ u_i &= S_u(X_i\delta_u) * u_i^* & (3) \\ v_i &= S_v(X_i\delta_v) * v_i^* & (4) \end{aligned}$$

where $Y_{1i}$ and $Y_{2i}$ are continuous endogenous variable; $X_i$ is a vector of variables that are independent of the error components $u_i^*$, $v_i^*$ ; $S_u$ and $S_v$ are unknown functions; $\beta$, $\theta$, $\pi$, $\delta_u$ and $\delta_u$ are unknown parameters; and $u_i^*$ and $v_i^*$ are zero mean random variables with non zero covariance and each with constant variance. Moreover, we assume that the $u_i^*$ and $v_i^*$ are correlated. The primary objective of estimation is to conduct inference on the parameters $\beta$ and $\theta$. Note that the model allows the same $X's$ in the both the reduced form and the primary equation without imposing any restrictions on the parameter values. We also allow the exact same $X's$ to appear in the functions underlying the process generating the heteroskedasticity in each equation.

It is immediately apparent from the structure of the model that least squares estimation of the main equation 2 will lead to inconsistent estimates of $\beta$ and $\theta$ due to the endogeneity of $Y_{2i}$ operating through the correlation of the errors across the two equations. Moreover, given the absence of exclusion restrictions, in that we make no assumptions about the elements of $\beta$ and $\pi$, there are no available instruments in this model. In the absence of making some additional statements about the form of the heteroskedasticity, which would then provide some additional moments one could exploit in estimation, it is not clear how one could produce consistent estimates of $\beta$ and $\theta$.

We show below that it is possible to identify all the features of the above model that are of interest by exploiting the presence of heteroskedasticity while making relatively few assumptions about its form. To provide some intuition for our approach consider the case where the errors are jointly normal. In the conventional control function approach, and in the absence of heteroskedasticity, the endogeneity of the $Y_{2i}$ is controlled for by including an estimate of $v_i$ in 1 as an additional regressor. This approach works because the inclusion of $v_i$ captures the component of $Y_{2i}$ that is correlated with $u_i$. In the absence of heteroskedasticity the mapping from the $v_i$ to $u_i$ is linear so one requires an exclusion restriction for the model to be identified. That is, the mapping is simply $E[u_i|v_i] = (\sigma_{uv}/\sigma_v^2)\, v_i$. Thus the mapping captured by $(\sigma_{uv}/\sigma_v^2)$ does not

[1] The complete paper, including the appendix, will be available upon request in approximately 1-2 months.

depend on $i$. However, when the model contains heteroskedasticity the $E[u_i|v_i] = (\sigma_{uv,i}/\sigma^2_{vi})\, v_i$ where the covariance and variance terms which capture the manner in which $v$ maps into $u$ varies across $i$. Provided that the mapping is not linear, in that $(\sigma_{uv,i}/\sigma^2_{vi})$ does not equal $C * (\sigma_{uv}/\sigma^2_v)$ for some constant $C$, this nonlinearity provides a form of identification which can be exploited. The challenge we confront in this paper is to show that we are able to estimate $(\sigma_{uv,i}/\sigma^2_{vi})$ while making as few assumptions as possible about the process generating the heteroskedasticity.

To begin it is useful to think about the estimation of the model also under the assumption that the error terms, $u^*$ and $v^*$ are jointly normal. In this case the log of the likelihood conditioned on $X$ and $v$ has the following form (up to an additive constant):

$$\ln\ L = -\frac{1}{2}\sum_{i=1}^{N}\left\{\frac{Y_{1i} - X_i\beta - \theta Y_{2i} - \alpha(S_{ui} * v_i^*)}{S_{ui}\sqrt{1-\rho^2}}\right\}^2 - (N/2)\sum_{i=1}^{N} Ln\left(S^2_{ui}\left(1-\rho^2\right)\right) \tag{5}$$

where $v_i^*$ is equal to $v_i/S_v(X_i\delta_v)$ and where $\rho$ is the correlation coefficient between $u_i^*$ and $v_i^*$The form of this likelihood function immediately reveals a number of the features of the estimation problem we confront. First, given that $v_i^*$ can be obtained from estimation of the reduced form (in a manner described below), the estimation of the primary equation would be straightforward if one knew $S_{ui}$. Second, when $S_{ui}$ is known it is possible to consistently estimate the parameters from the primary equation by linear least squares procedures. Third, given that the estimator of the main equation is a least squares procedure, identification requires that the matrix $W = [X_i : Y_{2i} : S_{ui} * v_i]$ is of full rank. This will be satisfied provided that the $S_{ui} * v_i$ term is not recoverable from any linear combination of $X_i$ and $Y_{2i}$. This will generally be true in the presence of heteroskedasticity.

It is useful to examine this correction term $S_{ui} * v_i^*$ as this provides some insight into why the heteroskedasticity identifies the model but also why the estimation of the model is complicated by its presence. Recall that the correction has the form $S_{ui} * v_i^*$ which is equal to $\frac{S_{ui}}{S_{vi}} * v_i^*$. Again a number of points are worth noting. First, the presence of heteroskedasticity generally ensures that $\frac{S_{ui}}{S_{vi}}$ varies across $i$ and this variation ensures that $W$ is of full rank. Second, while heteroskedasticity will generally produce a $W$ of full rank, the heteroskedaticity is insufficient if $\frac{S_{ui}}{S_{vi}}$ is equal to some constant. Thus the model requires that the heteroskedasticity is not of the same form in both equations. Finally, the estimation of the model requires that one estimates $S_{ui}$ and this is the feature of the estimation which complicates the process.

Previous papers have attempted to exploit the presence of heteroskedasticity as a form of identification. Vella and Verbeek (1997), and subsequently Vella et al (1999) develop an estimation procedure based on the rank order of the reduced form residuals for various subsets of the data. The variable determining the selection of subsets is also assumed to be responsible for the heteroskedasticity. The papers then define various estimators where the rank is used as an instrument or control function, or as the basis for various data transformations, as a means for controlling for endogeneity. The estimators all require, however, that the heteroskedasticity in the main equation is uncorrelated with that in the reduced form. Thus while the proposed estimator(s) were all relatively simple to implement, and the assumptions regarding the heteroskedasticity may seem reasonable in some empirical applications, the general structure of the model is quite limited. Thus while the estimator we propose below is somewhat more complicated we feel that the added generality it brings outweighs the additional costs of computation.

Another approach which might be employed would be to parameterize the $S_u$ and $S_v$ functions. If this is done in the appropriate manner it would then be possible to derive additional moments which can be exploited in a GMM framework and under certain conditions this will also identify the model. This methodology is employed by Rigobon (2003). Our approach is more general in that we do not parameterize the heteroskedasticity although our methodology is restricted to the triangular structure whereas the approach adopted by Rigobon is not. Finally note that the idea "creating" instruments or control functions in the absence of exclusion restrictions is not limited to cases where the model is contaminated with heteroskedasticity. Dagenais (199?) and Lewbel (1999) also discuss estimation of models where there are endogenous regressors and no exclusion restrictions. They show that when there is measurement of a specific form one is able to use instruments based on the higher powers of the included variables. The estimator which we present is not related to the approach adopted there.

# 3 Estimation Strategy

If one ignores the presence of the heteroskedasticity the obvious approach would be to simply employ the residual from estimating the reduced form and including this as an additional regressor in the primary equation. This however, does not work for the obvious reason that $W$ is not of full rank. The conditional likelihood function given above indicates that the appropriate control function to be employed is $S_{ui}v_i^*$. If this is what we employ the condition that we require for consistency is that the new error term is uncorrelated with $Y_{2i.}$ • That is, with $\alpha \equiv cov\left(u^*, v^*\right)/Var\left(v^*\right):$

$$\begin{aligned} E[Y_{2i} * (S_{ui}u_i^* - \alpha S_{ui}v_i^*)] &= 0 \\ E[Y_{2i} * \{S_{ui}(u_i^* - \alpha v_i^*)\}] &= 0 \\ E[S_{vi}v_i^* * \{S_{ui}(u_i^* - \alpha v_i^*)\}] &= 0 \end{aligned}$$

noting this is satisfied by the fact that the component of $Y_{2i}$ that is correlated with the error is that operating through $v_i^*$. Thus the inclusion of $S_{ui}v_i^*$ in the model accounts for the endogeneity although, as noted above, there is the problem that $S_{ui}$ is unknown and needs to be estimated.

An alternative control function procedure • would be to include that component of $S_{vi}v_i^*$ that is related to $\mathrm{s}_{u_i}u_i^*$. Namely, with $\overline{S}_{vi} = E[S_{vi}|X_i\delta_u)]$, include the variable $\overline{S}_{vi}v_i^*$ in the model to obtain:

$$Y_{1i} = X_i\beta + \theta Y_{2i} + \bar{\alpha}\overline{S}_{vi}v_i^*$$

Consistency now requires

$$E\left[\left[S_{vi}v_i^*\right]\left[S_{ui}u_i^* - \bar{\alpha}\overline{S}_{vi}v_i^*\right]\right] = 0.$$

If we take expectations of this over the informations set $u^*, v^*$ and $X_i\delta_u$ we get

$$\begin{aligned} & E\left[\left[\overline{S}_{vi}v_i^*\right]\left[S_{ui}u_i^* - \bar{\alpha}\overline{S}_{vi}v_i^*\right]\right] \\ = \ & E\left[\left(S_{ui}\overline{S}_{vi}\right)\right]E\left[u_i^*v_i^*\right] - \bar{\alpha}E\left(\overline{S}_{vi}^2\right)E\left(v_i^{*2}\right). \end{aligned}$$

Thus for

$$\bar{\alpha} = \left(\frac{cov(u_i^* v_i^*)}{var(v_i^*)}\right) \frac{E(S_{ui}\overline{S}_{vi})}{E\left(\overline{S}_{vi}^2\right)}$$

we find that the condition is satisfied. This naturally requires that • $\delta_1$ is known. • Thus the above shows that we have two controlled equations that we can exploit in estimation to derive consistent estimates of the primary equation parameters. These are

$$Y_{1i} = X_i\beta + \theta Y_{2i} + \alpha S_{ui} v_i^* \tag{6}$$
$$Y_{1i} = X_i\beta + \theta Y_{2i} + \alpha \overline{S}_{vi} v_i^*. \tag{7}$$

A major advantage of the second formulation, involving the use of $\overline{S}_{vi}$ is that it does not require the estimation of $S_{ui}$. That is, for any candidate value of the index parameters $\delta_u$ we can compute the remaining parameters of the model. In this manner, we can employ this equation to concentrate out all parameters other than the index parameters $\delta_u$. In the following section we discuss how to exploit this concentration and equation (7) to obtain consistent estimates of all the parameters of the model.

Before proceeding one should note that the presence of the heteroskedasticity in the model suggests that the use of GLS will lead to more efficient estimates. We explore this possibility below in developing an estimator.

# 4 The Estimator: Implementing Strategies

We first obtain consistent estimates of the reduced form slope parameters by regressing $Y_{2i}$ on $X_i$ to get $\widehat{\pi}$. We then estimate the residuals as

$$\widehat{v}_i = Y_{2i} - X_i\widehat{\pi}$$

To obtain an estimate of the $v_i^*$ we also require an estimate of the $S_v$ function which also requires $\delta_v$. To obtain this estimate we perform a semi-parametric least squares regression of $\ln(\widehat{v}_i^2)$ on $X_i$ to obtain $\widehat{\delta}_v$. We then estimate $S_v$ from:

$$\widehat{S^2}_{vi} = \widehat{E}\left(\widehat{v}_i^2 | X_i\widehat{\delta}_v\right)$$

where the $\widehat{E}$ is a non-parametric estimator based on the use of kernels. Using this estimate $\widehat{S}_v$ one can then produce an estimate of $v_i^*$ as $\widehat{v}_i^* = \widehat{v}_i/\widehat{S}_v$.

Once we have an estimate of $v_i^*$ we can define estimates of the other parameters of the model as functions of the index parameters $\delta_u$. That is, once we have an estimate of $\delta_u$ it is relatively straightforward to compute the other parameters using the relationship in 7. With the true index parameters given by $\delta_u$ employ OLS on (6) to obtain the estimates $\beta(\delta_u)$ and $\theta(\delta_u)$. We can then obtain an estimate of the residual from 1 as

$$u_i(\delta_u) = Y_{1i} - X_i\beta(\delta_u) - \theta(\delta_u)Y_{2i} = S_{ui}(\delta_u)u_i^* \tag{8}$$

We then obtain an estimate of $S_{ui}(\delta_u)$ from

$$S_{ui}^2(\delta_u) = \widehat{E}\left((u_i(\delta_u))_i^2 | X_i\delta_u\right)$$

where the $\widehat{E}$ again represents a non-parametric estimator based on the use of kernels.

One might now think of the two following optimization problems, based on our discussion above, to recover $\delta_u$ and $\beta(\delta_u)$, and $\theta(\delta_u)$ :

$$\min_{\delta_u} \sum \{Y_{1i} - X_i\beta(\delta_u) - \theta(\delta_u)Y_{2i} - \bar{\alpha}(\delta_u)S_{ui}(\delta_u)v_i^*\}^2 \tag{9}$$

$$\min_{\delta_u} \sum \left\{Y_{1i} - X_i\beta(\delta_u) - \theta(\delta_u)Y_{2i} - \alpha(\delta_u)\overline{S}_{vi}(\delta_u)v_i^*\right\}^2 \tag{10}$$

An interesting feature of the model under examination is that the estimates produced by 9 produces consistent estimates while those produced by 10 do not. The reason for this is discussed below. Nevertheless the estimates of all the parameters from 9 are consistent.

However, in examining the finite sample performance of the estimators produced by 9 we noted that while all the estimates displayed only a small degree of bias, the variance for the estimator of $\delta_u$ was unacceptably large. Intuitively, the information in (9) is "weak" for purposes of estimating index parameters. Instead, by focusing on the residuals, $\hat{u}_i$, it would seem preferable to estimate the $u$-index in the same way that we estimated the $v$-index. Accordingly, to obtain better behaved estimates of $\delta_u$ we employ the following procedure. Given an estimator of $\delta_u$ we can obtain an estimate of $u_i(\delta_u)$ as is done in 8. We can then define the following optimization problem as a means to obtaining $\delta_u$

$$\min_{\delta_u} \sum \{f(u_i(\delta_u)) - f(u_i(\delta_u)|X_i\delta_u)\}^2$$

for any function $f(.)$. Given the form of the heteroskedasticity that we assume, the function we choose is the log of the squared residual. [2]Thus we estimate $\delta_u$ through the following optimization problem

$$\min_{\delta_u} \sum \left\{(u_i(\delta_u)^2) - (u_i(\delta_u)^2|X_i\delta_u)\right\}^2 \tag{11}$$

This will lead to consistent estimates providing we have a consistent starting value. Accordingly, we employ the value from 9 as the starting value for 11. We then use the relationship in 7 to define the other parameters for the estimated $\delta_u$.

# 5 Properties of Proposed Estimator

# 6 Simulation Evidence

To examine the performance of our proposed procedure under a control setting we performed a number of simulation exercises. To make the setting as least favorable we simulated the following model where the same exogenous variables appear in the conditional means and the conditional

[2]Here, the residual must be appropriately modified and trimmed so that it is finite for every N and so that it can not become "too large" as N increases.

variances of both endogenous variables. The model has the form

$$
\begin{aligned}
Y_{1i} &= 2 + 2 * x_{1i} + 2 * x_{2i} + 2 * Y_{2i} + u_i \\
Y_{2i} &= 2 + 2 * x_{1i} + 2 * x_{2i} + v_i \\
u_i &= \{\gamma * (1 + \exp(.25 * x_{1i} + .75 * x_{2i}) + (1 - \gamma)\} * u_i^* \\
v_i &= \{\gamma * (1 + \exp(.75 * x_{1i} + .25 * x_{2i}) + (1 - \gamma)\} * v_i^* \\
u^* &= .75 * v_i^* + N(0,1) \text{ and } v_i^* \sim N(0,1).
\end{aligned}
$$

In the simulations we examine we employ both normal and chi-squared $x's$ and we also experimented with different distributions for the disturbances. The parameter $\gamma$ controls the level of heteroskedasticity in the model so we examine the impact of different values of this parameter for the estimates. Accordingly we experimented with different values for this parameter. The simulation results for $n = 1000$ and 100 replications are reported in Table 1. Before discussing the simulation results it is worth highlighting the structure of the model that we simulate. We employ the same exogenous variables in both the conditional means and the conditional variances. Moreover, we use the same functional forms for the heteroskedasticity in each equation. Given that the estimator is required to purge $E[S_v|X_i\delta_u]$ from the main equation this is likely to be more difficult when the $S_v$ and $S_u$ are highly correlated. Thus the design we examine is constructed to be one which is difficult for our procedure to work well.

An examination of Table 1 reveals a number of interesting features of the simulations. First consider the first two columns for the case when $\gamma$ is equal to 1. The OLS estimates for the main equation's parameters in this specification are a long way from the true values of 2 indicating that there is a large degree of endogeneity in this model. While the estimates for the $x's$ are less than 1, indicating a bias of over 50 percent, the coefficient for the endogenous regressor is approximately 27 percent. Consider now the lower panel of columns 1 and 2. First focus on the estimates of $\delta_u$. Given the form of the model and the parametrization that we employ, the true value of this parameter .33. The estimate $\delta_{u1}$ corresponds to the parameter which is estimated by minimizing 9 noting that we expect that this estimate is likely to have some difficulty in obtaining an accurate value given that it does not directly exploit the fact that the variance of the residuals are a function of $\delta_{u1}$. The average point estimate of .431 is not bad but the standard error of the estimate is high at .395. Along $\delta_{u2}$ we provide the estimate of $\delta_u$ which comes from 11 where we note that is this a more natural estimate in the sense that it more directly exploits the role of $\delta_u$ in determining the variance of $u$. The average point estimate from this specification is .389 and, in contrast to $\delta_{u1}$, there is a relatively large decrease in the magnitude of the standard error which is now .177. Finally in this lower panel we examine the point estimates from both the least squares and the generalized least squares procedures. Both the procedures appear to work well and represent an effective way to eliminate the bias in the model. The adjusted ordinary least squares estimates all display a bias of approximately 2 to 3 percent while the GLS estimates have a bias in the order of 3.5 percent. If we contrast this to the unadjusted OLS estimates the procedures we have suggested provide a remarkable improvement. Also note that the simulations indicate that there are some reasonable efficiency gains in employing the GLS procedure over the least squares method.

One would expect that the performance of the estimator depends on the form and degree of heteroskedasticity. Accordingly, it is useful to consider other designs where there is less heteroskedasticity by decreasing the value of $\gamma$. We reduced this parameter to .75 and the results are also shown in table 1 noting that these results are only for 72 replications. First note that in the upper part of Table 1 that the change in the design has slightly increased the level of bias in the OLS estimates. The bias for the exogenous variables remains over 50 percent while the bias for the endogenous regressor is still around 25 percent. Now consider the lower panel. The reduction in the level of heteroskedasticity has had very little effect on the ability of the procedure to reduce the bias. For all of the estimates the bias is of the order of 3 to 4 percent. There appears to be a slight increase in the standard error of the estimated coefficients.

The simulation evidence we report here is very supportive of our procedure. Note, however, we examined a number of variations on the design that, while we do not report the results in detail here, are worth mentioning. First, we generate the identical design to that used here where the only change is that we employed exogenous variables generated as $\chi$ squareds. This had no notable impact on the results indicating that the favorable features of the normal $X's$, which are often seen as enhancing the performance of estimation procedures, is not important here. Second, we also experiment with a sightly different design which reduced the correlation between the $S_v$ and $S_u$. We did this by changing the dgp's for the heteroskedasticity to

$$\begin{aligned} u_i &= \{\gamma * (1 + \exp(.15 * x_{1i} + .85 * x_{2i}) + (1-\gamma)\} * u_i^* \\ v_i &= \left\{\gamma * (1 + \exp(.75 * x_{1i} + .05 * x_{2i}^2) + (1-\gamma)\right\} * v_i^* \end{aligned}$$

and we simulated the model for 500 observations. The results are shown in Table 1 under the heading design 2. The results are reported for 100 replications. Even with the smaller sample size we see that the estimator is able to exploit the heteroskedasticity sufficiently well to substantially reduce the bias. Note that the smaller sample size is more than compensated by the reduction in the correlation between the two forms of heteroskedasticity. One might expect that this lower level of correlation is a more accurate description of what is likely to be found in practice. For this reason the simulation results are very encouraging. Note that in this design the true value for $\delta_u$ is .176. Table 1 indicates that even our preferred estimator has some difficulty estimating this parameter with precision.

# 7 Empirical Example: Estimating the Returns to Education

We now focus on applying our procedure to an existing data set. Given that a great deal of the discussion of lack of suitable instruments and the presence of endogeneity is related to the return to schooling literature, we choose to focus on this type of example to illustrate how our procedure can be appropriate for tackling such an issue. We employ data taken from the 1985 wave of Australian Longitudinal Survey. This is a data set which contains labor market and background information on a sample of Australian youth. In this paper we examine the determinants of wages for a sample of working individuals. Note that in this paper we focus on the wage determination process conditional on working and thus we do not address the issue of the endogeneity of the working decision. The model we estimate has the following form

$$\begin{aligned} \ln\ wage_i \ &= \ \beta_0 + \sum_{j=1} \beta_{1j} * family + \sum_{j=1}^{2} \beta_{2j} * parent's\ education + \beta_3 * siblings + \\ &\sum \beta_4 * state\ of\ school + \sum \beta_5 * school\ type + \sum_{j=1}^{2} \beta_{6j} * work\ chars \\ &\beta_7 * Gender + \beta_8 * Australian + \beta_9 * Age + \beta_{10} * Age^2 + \theta * schooling \end{aligned} \tag{12}$$

$$\begin{aligned} schooling_i \ &= \ \pi_0 + \sum_{j=1} \pi_{1j} * family + \sum_{j=1}^{2} \pi_{2j} * parent's\ education + \pi_3 * siblings + \\ &\sum \pi_4 * state\ of\ school + \sum \pi_5 * school\ type + \pi_6 * Gender \\ &+ \pi_7 * Australian + \pi_8 * Age \end{aligned} \tag{13}$$

where the $family$ variables capture the composition of the family when the individual was aged 14 and also if the mother was employed in market work when the individual was aged 14; $parent's$ $education$ are dummy variables indicating whether each of the parents had college degree; $siblings$ denotes number of siblings; $state\ of\ school$ and $school\ type$ are dummy variables indicating the region and type of school the individual attended; $Australian$ denotes Australian Born; $work$ $chars$ captures whether the individual is engaged in union and government employment and the remaining variables are self explanatory.

Note that the specification of the education and wage equations are different. More importantly, the wage equation includes all of the variables included in the education equation plus a quadratic in age and some work related variables. There appears to be no justification for including these variables in the educational attainment equation. We employ the same variables for the conditional variances for each equation as we do for the conditional means. Most importantly, note that the model is not identified as there are no variables included in the education equation which do not appear in the wage equation.

A feature of the literature devoted to the returns to education is that it is not clear what factors affect the individual's propensity to invest and the subsequent rate of return. For example, many of the background variables are included to capture the possibility that the individuals environment during his/her youth not only directly affect the individual's level of investment but also directly the individual's performance in the labor market, including his/her wage. As a result there appear to be no obvious exclusion restrictions which one can employ to control for the endogeneity. Accordingly we employ the control function procedure outlined above.

It is useful to consider why heteroskedasticity might appear in this model. First consider the schooling equation. Rummery, Vella and Verbeek (1999) argue that one source of heteroskedasticity in the schooling equation might be captured by the regional variables. That is, one might consider that the location of schools within a region may be capable of generating heteroskedasticity. For example, consider the case where the distance to the nearest school influenced the schooling decision. Then, one can see that various allocations may produce the same expected level of schooling but drastically different variances. The same is also true of many of the other variables. For example, consider the variable which captures whether or not the school attended was Roman Catholic. It is well established that attendance at a Catholic school increases educational attainment but there is a large amount of heterogeneity across Catholic schools in Australia and it seems almost implausible that one would expect the same impact irrespective of the quality. Similar logic applies to the presence of heteroskedasticity in the wage equation. In short, while many of the variables would be expected to influence the wage rate it does not seem likely that the impact is constant.

We now employ our procedure using a sample of approximately 4000 individuals from the Australian Longitudinal Survey. Note that we only report the estimates for the wage equation and this is done in Table 2. The standard errors for the estimates are reported in parentheses under the estimates. First consider the estimation of the OLS equation noting that we employ age and age squared in place of the more conventionally employed experience and experience squared. This approach generates quite different estimates of the return to education and we discuss this below. A number of features are worth noting from the OLS estimates in Table 2. First, there is little evidence that the background variables exercise much influence on the wage level. Second, with the exception of one the regional dummies there is no indication that the school type or region influences the wage level. This latter result is somewhat unexpected since one might expect that there is some direct return to attending a private school. Note, however, that the absence of any such effect in these data might reflect the endogeneity of the schooling decision. This naturally is what makes the decision to exclude the background or school type variables unjustified in the absence of additional information. Finally, the return to education in these data, on the basis of the OLS estimates, is .022 percent. This number seems very low but two factors should be taken into account. First, these individuals are very young (15 to 26 years old) and the returns to their investment. Second, the return to education is sensitive to the inclusion of age in the place of experience where experience is calculated as age-schooling minus 6. When we replace age and age squared with experience and experience squared the return to schooling is approximately 7 percent.

Now focus on the adjusted control function procedures which are reported alongside the OLS estimates in Table 2. A number of remarkable features are note worthy. First, the correction term has a statistically significant coefficient indicating that the estimator is able to identify the presence of endogeneity in this particular setting. Second, a number of the point estimates are quite different to the OLS estimates. In particular, the estimates on the school types are now both statistically significant. Noting that the control (excluded group) is state financed school we see that attendance at Roman catholic schools and other private schools directly increase wages by around 5 percent. There continues to be no direct effect from the background variables with the exception of the number of siblings but this effect appears to be small in magnitude. Finally consider the coefficient on the variable of primary interest, namely education. The coefficient has now increased to suggest a rate of return of 6 percent. Also, while it is less precise than the OLS estimate it is still reasonably precisely estimated. Although it is impossible to draw conclusions on the validity of this increase, given we do not know the true estimate, the increase to 6 percent is consistent with many previous studies that find OLS underestimates the return to schooling. The estimate that the control function procedure produces is also very reasonable in magnitude.

**Table 1: Simulation Results**

| | $\gamma = 1$ | $\gamma = 1$ | $\gamma = .75$ | $\gamma = .75$ | **Design 2** | |
|---|---|---|---|---|---|---|
| | **OLS** | | | | | |
| **constant** | .925 | | .906 | | .981 | |
| | (.099) | | (.095) | | (.131) | |
| $\mathbf{x}_1$ | .922 | | .903 | | .976 | |
| | (.103) | | (.098) | | (.138) | |
| $\mathbf{x}_2$ | .932 | | .911 | | .987 | |
| | (.095) | | (.096) | | (.141) | |
| $\mathbf{Y}_{2i}$ | 2.539 | | 2.550 | | 2.511 | |
| | (.048) | | (.045) | | (.062) | |
| | **CF** | **GLS CF** | **CF** | **GLS CF** | **CF** | **GLS CF** |
| **constant** | 2.055 | 2.075 | 2.078 | 2.076 | 2.020 | 2.053 |
| | (.300) | (.263) | (.305) | (.282) | (.346) | (.255) |
| $\mathbf{x}_1$ | 2.052 | 2.073 | 2.076 | 2.073 | 2.018 | 2.053 |
| | (.298) | (.261) | (.307) | (.283) | (.346) | (.248) |
| $\mathbf{x}_2$ | 2.058 | 2.073 | 2.079 | 2.073 | 2.022 | 2.051 |
| | (.304) | (.264) | (.310) | (.285) | (.347) | (.247) |
| $\mathbf{Y}_2$ | 1.976 | 1.964 | 1.965 | 1.965 | 1.991 | 1.974 |
| | (.150) | (.130) | (.154) | (.141) | (.167) | (.121) |
| $\boldsymbol{\delta}_{u1}$ | .431 | | .454 | | .211 | |
| | (.395) | | (.473) | | (.299) | |
| $\boldsymbol{\delta}_{u2}$ | .389 | | .391 | | .199 | |
| | (.177) | | (.202) | | (.154) | |

Table 2: Returns to Education

| Variable | OLS | CF | Variable | OLS | CF |
|---|---|---|---|---|---|
| **Constant** | -5.309 | -5.604 | **Roman Cath** | .019 | .055 |
| | (.297) | (.321) | | (.020) | (.023) |
| **Australian** | -.017 | -.025 | **Private** | .024 | .043 |
| | (.014) | (.014) | | (.023) | (.023) |
| **Neither parent** | -.001 | .001 | **Age Squared** | .584 | .576 |
| | (.016) | (.016) | | (.028) | (.029) |
| **Mother Only** | -.034 | -.017 | **Age Squared** | -.012 | -.012 |
| | (.033) | (.032) | | (.001) | (.001) |
| **Father Only** | -.007 | .005 | **Female** | -.044 | -.059 |
| | (.040) | (.040) | | (.009) | (.009) |
| **Mother Working** | .006 | .004 | **Union** | .070 | .072 |
| | (.090) | (.090) | | (.010) | (.010) |
| **Mother with Degree** | -.007 | -.014 | **Government** | .084 | .079 |
| | (.024) | (.024) | | (.095) | (.095) |
| **Father with Degree** | .004 | -.015 | **Education** | .022 | .060 |
| | (.017) | (.019) | | (.002) | (.013) |
| **No. of Siblings** | .002 | .006 | **Control Function** | | -.038 |
| | (.003) | (.003) | | | (.014) |
| **State 1** | .054 | .066 | | | |
| | (.017) | (.018) | | | |
| **State 2** | .018 | .031 | | | |
| | (.017) | (.018) | | | |
| **State 3** | .014 | .022 | | | |
| | (.019) | (.019) | | | |
| **State 4** | .025 | .028 | | | |
| | (.022) | (.022) | | | |
| **State 5** | .008 | .043 | | | |
| | (.026) | (.029) | | | |