Standard Error Correlation in Two-Stage Optimization Models: A Quasi-Maximum Likelihood Estimation Aproach

Fernando Rios-Avila, Gustavo J. Canavire-Bacarreza

Standard Error Correction in Two-Stage Optimization Models:
A Quasi-Maximum Likelihood Estimation Approach

Fernando Rios-Avila
Levy Economics Institute,
Annandale-on-Hudson, NY USA
friosavi@levy.org

Gustavo J. Canavire-Bacarreza
Universidad EAFIT, Medellin Colombia
gcanavir@eafit.edu.co

Abstract: Following Wooldridge (2014), we discuss and implement in Stata an efficient maximum likelihood approach to the estimation of corrected standard errors of two-stage optimization models. Specifically, we compare the robustness and efficiency of this estimate using different non-linear routines already implemented in Stata such as ivprobit, ivtobit, ivpoisson, heckman, and ivregress.

Keywords: Maximum Likelihood Estimation, non-linear models, endogeneity, two-step models, standard errors

1. Introduction

Endogeneity in non-linear models has not been easy to handle given the probabilistic nature of the endogenous explanatory variable. Applying two-stage linear methods produce inconsistent results for the structural parameters as well as the standard errors. Although the properties and formulation of the correct standard errors associated with two-stage optimization models have been available in the literature for decades (Murphy and Topel, 1985; Newey and MacFadden, 1994 and White, 1982), outside of already packaged commands, practical implementation is not yet standard (Terza, 2016). Despite the efforts of many authors to provide strategies and programming codes to facilitate the implementation of standard error corrections in Stata (Hardin, 2002; Hole, 2006 and Terza, 2016), the majority of applied researchers implement bootstrap methods at best or ignore the problem by reporting the uncorrected errors at worst (Terza, 2016).

In this article, we suggest a different and easier to implement approach to the estimation of corrected standard errors of two-stage optimization models, based on Wooldridge (2014), compared to

those already applied in other packaged commands within Stata. The method relies on the maximum likelihood estimation available in Stata, which is used to estimate a joint function that characterizes the data- generating process of the first-stage and second-stage systems, under the assumption of independent errors. The main use of this model is to facilitate the estimation of endogenous non-linear models using a control function approach such as the ones described in Terza et al. (2008), with corrected standard errors.

The rest of the paper is organized as follows. In Section 2, we review the framework of quasi-likelihood function in non-linear models. Section 3 establishes the general framework of the estimation based on maximum likelihood functions. Section 4 provides examples of the estimation compared to existing methods in Stata and of the control function approach, which is also known as the two-stage residual imputation approach (Terza, 2008).

2. Two-Stage Estimations

Systems in which results from one estimation are used in a second model are common in the applied literature. The most prominent examples are the estimation of the two-stage Heckman selection models (Heckman, 1979), two-stage least squares (2SLS) for the treatment of endogeneity in linear models, and the control function approach for the robust estimation of non-linear models with endogenous variables (Terza et al., 2008; Wooldridge, 2014). Although these types of models can be estimated jointly, two-step procedures are often easier to be estimated because they require fewer restrictions on the joint distribution of the data-generating functions (Greene, 2003).

While easy to implement, the main drawback of two-stage models has been that the standard errors obtained from the second stage alone are incorrect, as they ignore the measurement error that carries over from using the predictions of one model in the next model. Hardin (2002) and Hole (2006) provide guidance for implementing the variance estimator suggested by Murphy and Topel (1985) when one includes the predictions of a first-stage model into a second model. More recently, Terza (2016) suggests an additional simplification of the estimation of standard errors in two-stage models, with

emphasis on the application of the two-stage residual inclusion approach and the handling of endogeneity in nonlinear models.

The original model described in Hardin (2002) is characterized as follows:

$$Model\ 1: E(y_1|x_1, \theta_1)$$

$$Model\ 2: E\big(y_2|x_2, \theta_2, E(y_1|x_1, \theta_1)\big)$$

where in Model 1 we are modeling the conditional mean of the endogenous variable y_1 as a function of exogenous variables x_1, and model 2 models the conditional mean of the variable y_2 as a function of exogenous variables x_2, and some form of the predicted values of the first model.

According to Hardin (2002), two approaches can be used to estimate the models. The first approach is a full information maximum likelihood, where we start by specifying a joint distribution $f(y_1, y_2|x_1, x_2, \theta_1, \theta_2)$ which is then maximized jointly. A second approach is the estimation of a limited information maximum likelihood (LIML), a two-step procedure. Because the first model only depends on the parameter θ_1, it can be consistently estimated. The estimated parameters of the first model can be directly included in the second model, which can be estimated using a conditional log likelihood function:

$$\max_{\theta_2} \mathcal{L}_2(\theta_2|\theta_1) = \ln \sum ln\left(f\left(y_2|x_2, \theta_2, g(x_1, \hat{\theta}_1)\right)\right)$$

In this framework, Hardin (2002) and Hole (2006) indicate that while the raw variance obtained from the second stage model is incorrect, it can easily be corrected using the Murphy-Topel approach, which states that:

$$\hat{V}(\theta_2) = \hat{V}^*(\theta_2) + \hat{V}^*(\theta_2)\big(\hat{C}\hat{V}_1\hat{C}' + \hat{R}\hat{V}_1\hat{C}' - \hat{C}\hat{V}_1\hat{R}'\big)\hat{V}^*(\theta_2)$$

where $\hat{V}^*(\theta_2)$ is the incorrect estimation of the variance obtained from the second-stage maximization problem, $\hat{V}_1$ is the correct estimation of the variance of the first-stage estimate. $\hat{C}'$ and $\hat{R}$ are the matrices products defined as:

$$\hat{C} = E\left\{\left(\frac{\partial \mathcal{L}_2}{\partial \theta_2}\right)\left(\frac{\partial \mathcal{L}_2}{\partial \theta_1'}\right)\right\} \ and \ \hat{R} = \left\{\left(\frac{\partial \mathcal{L}_2}{\partial \theta_2'}\right)\left(\frac{\partial \mathcal{L}_2}{\partial \theta_1}\right)\right\}$$

Which Hardin (2002) indicates can be traced back to represent the sandwich estimate of the variance given two maximum likelihood functions $\mathcal{L}_1(\theta_1)$ and $\mathcal{L}_2(\theta_2|\theta_1)$. (see Hardin (2002) for details).

3. Quasi-Maximum Likelihood Approach

According to Wooldridge (2014), the model described above can also be estimated using what he describes as a quasi-limited information maximum likelihood (QLIML). As described in his paper, for cases in which the models are linear (two-stage least squares), and under the assumption that the errors of the models distribute as normal and independent, the joint maximum likelihood function can be written as:

$$\mathcal{L}(\theta_1, \theta_2) = \mathcal{L}_1(\theta_1) + \mathcal{L}_2(\theta_2|\theta_1)$$

As he indicates, the estimation of this type of model can be done using readily available statistical software. Further, inferences can be drawn by reporting the White (1982) type of sandwich variance estimations, to account for miss-specified likelihood functions. Furthermore, Wooldridge (2014) indicates that the estimation of this joint quasi-maximum likelihood estimation (QMLE) can also be applied to a larger set of maximum likelihood models that belong to the linear exponential family, because they remain consistent even when the density function is partially miss-specified (Cameron and Trivedi, 2005). This suggests that a range of models, such as the models described in Terza et al. (2008) and Wooldridge (2015), can also be estimated consistently using quasi-maximum likelihood functions.

Notably, this type of estimation is not foreign to the processes already built into statistical software such as Stata. According to the reference manual, the estimation of Probit and Tobit models with endogenous variables follows this approach (Newey, 1987). A few examples will be presented regarding how the estimation of two-stage models can be performed using QMLE.

3.1. Estimation of QMLE in Stata

The estimation of models using Stata Maximum likelihood functions is simple, but requires some programing and knowledge of the appropriate density functions the researcher assumes determines the data generating process of the data in hand. In this section, we provide the general setup of how the program can be written using the *lf* method, which is the simplest method with the Stata ML estimation commands.

As described in Cameron and Trivedi (2009), the *lf* method can be used for the especial case where the objective function is an *m* estimator. Namely, the sum or average of some function $q(.)$ over the sample of observations. As the authors indicate, this method can be applied for any function $q(.)$, but that robust standard errors need to be reported if the objective function is not Likelihood based. In the framework of a two-step program, the corresponding program can be written as follows:

```
program myols_ols
        args lnf xb1 s1 xb2 s2
        qui {
                tempvar lnf1 lnf2
                * OLS ML Component L2
                gen double `lnf2'=ln(normalden($ML_y2,`xb2',exp(`s2')))
                * OLS ML Component L1
                gen double `lnf1'=ln(normalden($ML_y1,`xb1',exp(`s1')))
                replace `lnf'=`lnf1'+`lnf2'
        }
end
```

This program can be used for the simultaneous estimation of two linear models under the assumption of normal and independent distributed errors. Calling for the estimation of this model is rather simple. It requires providing two left hand variables, two sets of explanatory variables. Other options can be used for the maximization methods or initial values:

```
ml model lf myols_ols (xb1:y1=x1 x2 x3) (s1:) (xb2:y2=x1 x2 x3 x4) (s2:), maximize robust
```

Now, in order to allow for the results from the one of the models to enter into the log likelihood of the other model, we need to specify an additional parameter to be estimated, which will introduced in the log likelihood function. In this example, we will call this link parameter `g'. The program and the maximization command then are modified as follows:

```
program myols_ols
        args lnf xb1 g s1 xb2 s2
        qui {
                tempvar lnf1 lnf2
                * OLS ML Component L2
                gen double `lnf2'=ln(normalden($ML_y2,`xb2',exp(`s2')))
                * OLS ML Component L1, including the prediction of the first model xb2 into the
                second one.
                gen double `lnf1'=ln(normalden($ML_y1,`xb1'+`g'*`xb2',exp(`s1')))
                replace `lnf'=`lnf1'+`lnf2'
        }
end
```

```
ml model lf myols_ols (xb1:y1=x1 x2 x3) (g:) (s1:) (xb2:y2=x1 x2 x3 x4) (s2:), maximize robust
```

Notice that in this case we are implementing something similar to a two stage least squares, where the predictions from the first model `xb2' is used as an additional variable determining the conditional mean of the second model. Also notice that for this particular example, the endogenous variable y2, is not included in the list of explanatory variables in the main model. If we would be using a residual inclusion/Control Function approach, however, the endogenous variable would be included in the list of explanatory variables.

The above code can be easily extended to allow for multiple equations, for example, a model with two endogenous variables, by just including an additional link parameter from the first stage to the second stage:

```
program myols_ols
        args lnf xb1 g1 g2 s1 xb2 s2 xb3 s3
        qui {
                tempvar lnf1 lnf2 lnf3
                * OLS ML Component L2
                gen double `lnf2'=ln(normalden($ML_y2,`xb2',exp(`s2')))
                gen double `lnf3'=ln(normalden($ML_y3,`xb3',exp(`s3')))

                * OLS ML Component L1, including the prediction of the first model xb2 into the
                second one.
```

```
            gen double `lnf1'=ln(normalden($ML_y1,`xb1'+`g1'*`xb2'+`g2'*`xb3',exp(`s1')))
            replace `lnf'=`lnf1'+`lnf2'+`lnf3'
        }
end
```

```
ml model lf myols_ols (xb1:y1=x1 x2) (g1:) (g2:) (s1:) (xb2:y2=x1 x2 x3 x4) (s2:) (xb3:y3=x1 x2
x3 x4) (s3:), maximize robust
```

In the next section, we show a few examples of how this simple set up can be used for the estimation of

two step estimation procedures, comparing the results with Stata built-in procedures.

4.   Applications

4.1.   Probit and Tobit models with endogenous continuous variables

According to Terza et al (2008), a consistent estimation of models with endogeneity in the context of

nonlinear outcome functions uses the two-stage residual approach, also known as the control function

approach (CF; Wooldridge 2015). Because the underlying models behind the estimation of IV Probit and

IV Tobit follow a setup similar to the 2SRI approach, we begin by providing the implementation of the

proposed method in the framework of these two models. In the framework of a continuous endogenous

variable, the first stage involves the estimation of a linear model, while the second step estimates a probit

model, using the predicted residuals of the first stage as explanatory variables.

        To estimate this model in Stata, the first step is to write a program that specifies this objective

function. We will use a simplified program based on the one presented in section 3.1:


```
program myivprobit
args lnf xb1 g xb2 s2
* xb1 will be the parameters associated with the probit model
* g will be the parameter associated with the predicted error
* xb2 and s2 are correspond to the conditional mean and variance of 1st stage OLS
qui {
* OLS ML Component L1
replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`s2')))
* probit ML component L2
replace `lnf'=`lnf'+ln(normal(`xb1'+`g'*($ML_y2-`xb2'))) if $ML_y1==1
replace `lnf'=`lnf'+ln(1-normal(`xb1'+`g'*($ML_y2-`xb2'))) if $ML_y1==0
}
end
```

Using this program, we estimate a model equivalent to an IVProbit and compare it to the outcomes from

the built-in command -ivprobit-.

```
use http://www.stata-press.com/data/r14/laborsup, clear
ivprobit fem_work fem_educ kids (other_inc = male_educ), vce(robust)
est sto m1
ivprobit fem_work fem_educ kids (other_inc = male_educ), two
est sto m2
* for this particular model, we need to set up an initial value for the variance, to correct for
the overdispertion of "other_inc".
matrix lns2=2.9
matrix colname lns2=_cons
matrix coleq lns2=lns2
ml model lf myivprobit (fem_work:fem_work= other_inc fem_educ kids )  (g:=) (other_inc:other_inc=
fem_educ kids male_educ ) (lns2:), init(lns2)   maximize robust
est sto m3
```

This particular example took 15 iterations to be estimated using the NR optimization method. Below we

show the outputs of all the models:

Table 1 IVprobit Results

	ivprobit-ml	ivprobit-two	QMLE
main			
other_inc	-0.05428	-0.05847	-0.05847
	(0.0062)	(0.0093)	(0.0092)
fem_educ	0.21111	0.22744	0.22744
	(0.0251)	(0.0282)	(0.0264)
kids	-0.18209	-0.19617	-0.19617
	(0.0487)	(0.0496)	(0.0513)
_cons	0.36721	0.39561	0.39561
	(0.4410)	(0.4983)	(0.4896)
g			0.02405
			(0.0097)
lns2	2.81338		2.81338
	(0.0327)		(0.0327)
N	500	500	500

Note: Robust Standard errors in parenthesis, except for ivprobit-two steps.
First Stage results available upon request.

As can be seen, except for differences caused by the maximum likelihood estimation of the OLS

component (it is, as Wooldridge (2015) describes, consistent but biased), the results of the two-step model

and the QML point estimates are identical, with small differences in the standard errors.

A second command that explicitly uses the joint quasi-maximum likelihood function is the estimation of the IVTobit. The first stage here also involves the estimation of a linear model, and the second uses the residuals of the first as an explanatory variable assuming the outcome variable follows as a censored normal distribution. As indicated earlier, we must specify the objective function. We can again estimate the model using maximum likelihood and the two-step procedure based on Stata example datasets:

```
program myivtobit
args lnf xb1 g s1 xb2 s2
* Similar to before, we first specify the ML corresponding to the OLS
qui {
replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`s2')))
* And we add the corresponding contribution coming from the Tobit Maximum Likelihood
replace `lnf'=`lnf'+ln(normalden($ML_y1,`xb1'+`g'*($ML_y2-`xb2'),exp(`s1'))) if $ML_y1>0
replace `lnf'=`lnf'+ln(1-normal(( `xb1'+`g'*($ML_y2-`xb2'))/exp(`s1')))      if $ML_y1==0
}
end
```

We can again estimate the model using maximum likelihood and two-step procedure base some Stata example datasets:

```
webuse laborsup, clear
* we modify the outcome function just to assume the information is censored at 0
replace fem_inc=fem_inc-10
ivtobit fem_inc fem_educ kids (other_inc = male_educ), ll robust
est sto m1
ivtobit fem_inc fem_educ kids (other_inc = male_educ), ll  twostep
est sto m2
ml  model  lf  myivtobit  (fem_inc:fem_inc=  other_inc  fem_educ  kids  )      (g:=)   (lns1:)
(other_inc:other_inc= fem_educ kids male_educ ) (lns2:) , init(lns2)  robust maximize
est sto m3
```

Table 2 IVtobit estimation

	ivtobit-ML	ivtobit-two	QMLE
main			
other_inc	-0.90454	-0.90454	-0.90454
	(0.1322)	(0.1330)	(0.1322)
fem_educ	3.27239	3.27239	3.27239
	(0.3788)	(0.3969)	(0.3788)
kids	-3.31236	-3.31236	-3.31236
	(0.7543)	(0.7220)	(0.7543)
_cons	9.24735	9.24735	9.24734

	(7.5886)	(7.3739)	(7.5886)
alpha/g	0.29077		0.02405
	-0.1297		(0.0097)
lns1	2.87403		2.87403
	(0.0497)		(0.0497)
N	500	500	500

Note: Robust Standard errors in parenthesis, except for ivtobit-two steps. First Stage results available upon request.

In this case, the program takes longer and requires 36 iterations for its estimation. The number of required estimations can be reduced by providing better initial values and alternative maximization techniques. Nevertheless, because this model is exactly identified, the point estimates for the IV Tobit ML, IV Tobit two-step, and the QMLE are identical, with small differences in the standard errors compared to the IV Tobit two-step.

4.2. Heckman Selection Model

Another model that is readily available in statistical software such as Stata is the classic Heckman selection model. The two-stage procedure suggested for the estimation of this model (Heckman, 1979) can also be adjusted and implemented in the framework of the QMLE. The two stage procedure, also known as heckit procedure, involves the estimation of a probit model in the first stage. The parameters of this first stage are used to estimate a selection term or inverse mills ratio, which are included in the second stage with is assumed to be a linear model.

The program associated with the estimation of this model can then be written as:

```
program myheckman
args lnf xb1 g s1 xb2
qui {
  *Here the first stage is a Probit Model
  replace `lnf'=ln(normal(`xb2')) if $ML_y2==1
  replace `lnf'=ln(1-normal(`xb2')) if $ML_y2==0
  *And the second is a linear model, which we assume has normal distributed errors.
  replace  `lnf'=`lnf'+ln(normalden($ML_y1,`xb1'+`g'*normalden(`xb2')/normal(`xb2'),exp(`s1'))) if
  $ML_y2==1
}
end
```

Once again, we can compare these results with the standard Stata commands:

```
webuse womenwk, clear
gen wseen=wage!=.
replace wage=0 if wage==.
heckman wage educ age, select(wseen=married children educ age)
est sto m1
heckman wage educ age, select(wseen=married  children educ age) two
est sto m2
ml model lf myheckman (wage=educ age) (g:) (s1:) (wseen:wseen=married children educ age),
maximize technique(bhhh nr)
est sto m3
```

This particular model is more difficult to estimate without appropriate initial values. For this, we use a combination of two maximization techniques (nr and bhhh) as shown in code above.

Table 3 Heckman Estimation

	Heckman-ml	Heckman-two	QMLE
main			
education	0.98995	0.98253	0.9832
	(0.0534)	(0.0539)	(0.0544)
age	0.21313	0.21187	0.21424
	(0.0211)	(0.0221)	(0.0229)
_cons	0.48578	0.73404	0.6175
	(1.0991)	(1.2483)	(1.2871)
Selection equation			
married	0.44517	0.43086	0.42323
	(0.0668)	(0.0742)	(0.0692)
children	0.43871	0.44732	0.44254
	(0.0273)	(0.0287)	(0.0279)
education	0.05573	0.05836	0.05817
	(0.0109)	(0.0110)	(0.0110)
age	0.03651	0.03472	0.03573
	(0.0042)	(0.0042)	(0.0042)
_cons	-2.49102	-2.46737	-2.48892
	(0.1884)	(0.1926)	(0.1904)
athrho	0.87421		
	(0.1051)		
lnsigma	1.79256		1.67716
	(0.0288)		(0.0198)
Mills/g	4.22441	4.00162	4.04450
Lambda	(0.4172)	(0.6065)	(0.6168)
N	2000	2000	2000

Note: Robust Standard errors in parenthesis, except for heckman-two steps.
First Stage results available upon request.

As can be observed, the results of the QMLE are similar, although not identical, to the ones obtained using the built-in Heckman commands. Nevertheless, the QMLE standard errors are close to the ones obtained from the Heckman two-stage procedures. In this model, the main reason why even the point estimates are different from the two-step procedure is that in the two-step procedure, the parameters of the first-stage model are estimated by maximizing the objective function of the first model alone, whereas similar to the full ML approach, in the QMLE model the parameters of the first model are estimated by maximizing the likelihood of the first and second model.

4.3. OLS with endogeneity Two-Stage Least Squares

A model that is more typically used for the treatment of instrumental variables when the outcomes and endogenous variables are continuous variables is the 2SLS process. This estimation involves obtaining linear predictions for the endogenous variables using all exogenous variables, which then are used instead of the original variables in the outcome equation (two-stage predictor substitution-2SPS). Alternatively, one can obtain the predicted residuals of the first stage models, and add them to the main outcome model to control for the possible endogeneity. This strategy is also known as a control function approach or two-stage residual inclusion approach (2SRI) (Terza et al 2008; Wooldridge 2014, 2015). This model can also be estimated based on the QMLE approach, using both the standard two-stage prediction substitution approach and the two-stage residual inclusion approach (2SRI). As indicated by Wooldridge (2014), using the 2SRI approach in an MLE framework is equivalent to using a limited information maximum likelihood for IV regress.

First, let us outline the two alternative models for the estimation of equivalent 2SLS. In this case, we assume that both the first and second stage are to be estimated using MLE under the assumption of normality, instead of ordinary least squares.

```
* This one will assume the classic 2SLS prediction substitution
program myivreg2sls
args lnf xb1 g s1 xb2 s2
* First Stage Model
qui:replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`s2')))
* Second Stage model, with prediction inclusion
```

```
qui:replace `lnf'=`lnf'+ln(normalden($ML_y1,`xb1'+`g'*`xb2',exp(`s1')))
end


* This one will use 2SRI approach
program myivreg2sri
args lnf xb1 g s1 xb2 s2
* First Stage Model
qui:replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`s2')))
* Second Stage model, with predicted errors inclusion
qui:replace `lnf'=`lnf'+ln(normalden($ML_y1,`xb1'+`g'*($ML_y2-`xb2'),exp(`s1')))
end
```

Alternatively, we can also drop the assumption of normality, and specify the objective function in the program as if we were estimating the model using Ordinary Least Squares. This simplifies the model as the standard error is no longer estimated. However, for appropriate statistical inferences, the model needs to be estimated using robust option in the standard errors:

```
program myivreg2sls_2
args lnf xb1 g xb2
* First Stage Model
qui:replace `lnf'=-($ML_y2-`xb2')^2
* Second Stage model, with prediction inclusion
qui:replace `lnf'=`lnf'-($ML_y1-`xb1'-`g'*`xb2')^2
end


* This one will use 2SRI approach
program myivreg2sri_2
args lnf xb1 g   xb2
* First Stage Model
qui:replace `lnf'=-($ML_y2-`xb2')^2
* Second Stage model, with predicted errors inclusion
qui:replace `lnf'=`lnf'-($ML_y1-`xb1'-`g'*($ML_y2-`xb2'))^2
end
```

We will compare the results from the above models to the built-in packages using the example dataset from Stata:

```
webuse hsng2, clear
ivregress 2sls rent pcturban (hsngval = faminc i.region ), robust
est sto m1
ivregress liml rent pcturban (hsngval = faminc i.region ), robust
est sto m2
ivregress gmm rent pcturban (hsngval = faminc i.region ), robust
est sto m3
```

Regarding the estimation of the QMLE and QOLS models, it should be noted that for the 2SRI, the endogenous variables are directly introduced in the set of explanatory variables; however, that is not the case for the standard 2SLS. Empirically, the joint QML is difficult to estimate using the maximum

likelihood method, and thus we use a two-stage procedure. The first imposes a restriction on the link coefficient 'g' and then uses those initial values to estimate the full model.

```
constraint 1 [g]_cons=0
ml model lf myivreg2sls (rent:rent =pcturban ) (g:) (s1:) (hsngval:hsngval = pcturban  faminc
i.region) (s2:), maximize constrain(1)
matrix b=e(b)
ml model lf myivreg2sls (rent:rent =pcturban ) (g:) (s1:) (hsngval:hsngval = pcturban  faminc
i.region) (s2:), maximize   technique(nr bhhh) init(b)   robust
est sto m4
ml model lf myivreg2sri (rent:rent =pcturban hsngval)  (g:) (s1:) (hsngval:hsngval = pcturban
faminc i.region) (s2:), maximize init(b)  technique(nr bhhh) robust
est sto m5
ml model lf myivreg2sls_2 (rent:rent =pcturban ) (g:)   (hsngval:hsngval = pcturban   faminc
i.region) , maximize   technique(nr bhhh) init(b, skip)   robust
est sto m6
ml model lf myivreg2sri_2 (rent:rent =pcturban hsngval)  (g:)   (hsngval:hsngval = pcturban
faminc i.region) , maximize init(b, skip)  technique(nr bhhh) robust
est sto m7
```

The results of all estimations are presented in the table below. As indicated by Woorldridge(2014), the QMLE-2SRI approach provides results that are identical to the IV regress LIML procedure. The QMLE-2SPS, which has a setup similar to the standard 2SLS, provides point estimates that are closer to the LIML estimations, with standard errors that are also similar in magnitude (constant in equation g). In addition, using both QOLS approach we obtain the same point estimates as the standard 2SLS, with standard errors differing only due to the ML estimation procedure.

Table 4 ivregress Estimations

	ivregress 2sls	ivregress liml	ivregress gmm	QMLE-2sps	QMLE-2sri	QOLS-2sps	QOLS-2sri
main							
hsngval	0.00224	0.00267	0.00146		0.00267		0.00224
	(0.0007)	(0.0008)	(0.0004)		(0.0008)		(0.0007)
pcturban	0.08152	-0.18274	0.76155	-0.14639	-0.18273	0.08151	0.08151
	(0.4446)	(0.4942)	(0.2895)	(0.5100)	(0.5315)	(0.4230)	(0.4230)
_cons	120.7065	117.6087	112.1227	118.0348	117.6088	120.7065	120.7065
	(15.2555)	(18.8799)	(10.8023)	(17.8511)	(18.0421)	(15.6815)	(15.6813)
g				0.00261	-0.00219	0.00224	-0.00159
				(0.0008)	(0.0008)	(0.0007)	(0.0007)
s1				2.72398	2.66467		
				(0.1156)	(0.1280)		
N	50	50	50	50	50	50	50

Note: Robust Standard errors in parenthesis. First Stage results available upon request. The "g" coefficient corresponds to the Endogenous variable coefficient for the QMLE-2sps and QOLS-2sps column

4.4. Poisson with continuous endogenous variables

The last model we use to compare the performance of the QMLE approach is for the estimation of a Poisson model continuous endogenous variables, and compare the performance with built-in command are rated to the -ivpoisson-. The built-in commands in Stata have various options for the estimation of this model, including GMM with additive and multiplicative errors and a control function model using GMM. The choice of model comparison is important in the construction of the objective function, particularly because the control function estimation option of IV Poisson is only available for the model with multiplicative error specification. To appropriately compare the results, we provide estimations of two alternative specifications to the Poisson regression, a log linear model, which will be equivalent to a multiplicative error model, and a linear exponential model, which will be similar to an additive model.[1]

In this framework, we can program the three models using a two-stage control function approach as follows

```
* Poisson with 2SRI and additive error
program myivpoisson
   args lnf  xb1 g xb2 si2
  * First Stage: OLS using ML
   qui: replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`si2')))
   qui:    replace   `lnf'=`lnf'-exp(`xb1'+`g'*($ML_y2-`xb2'))+(`xb1'+`g'*($ML_y2-`xb2'))*$ML_y1-
lngamma($ML_y1+1)
end

* OLS ln(y)=xb+e or y=exp(xb)exp(e)
program myivregpoisson1
args lnf xb1 g s1 xb2 s2
   qui: replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`s2')))
   qui: replace `lnf'=`lnf'+ln(normalden(ln($ML_y1),`xb1'+`g'*($ML_y2-`xb2'),exp(`s1')))
end

*  OLS y=exp(xb)+ e    Additive error
program myivregpoisson2
args lnf xb1 g s1 xb2 s2
   qui: replace `lnf'=ln(normalden($ML_y2,`xb2',exp(`s2')))
   qui: replace `lnf'=`lnf'+ln(normalden($ML_y1,exp(`xb1'+`g'*($ML_y2-`xb2')), exp(`s1')))
end
```

Again, we implement the estimation of all these models as follows:

```
webuse website, clear
ivpoisson gmm visits ad female (time = phone frfam), add
```

---

[1] Under this specification, the two models will be y=exp(xb+e), for a multiplicative error model, and y=exp(xb)+e, for an additive error model.

```
est sto m1
ivpoisson gmm visits ad female (time = phone frfam), mult
est sto m2
ivpoisson cfunction visits ad female (time = phone frfam)
est sto m3
ml model lf myivpoisson (visits= ad female time) (g:) (time:time = phone frfam ad female) (s1:),
technique(nr bhhh) maximize robust
est sto m4
ml model lf myivregpoisson1 (visits= ad female time) (g:) (s1:) (time:time = phone frfam ad
female) (s2:),  technique(nr bhhh) maximize robust
est sto m5
** The next model is difficult to maximize unless we provide good starting values
matrix b=e(b)
matrix b[1,6]=exp(exp(b[1,6]))
ml model lf myivregpoisson2 (visits= ad female time) (g:) (s1:) (time:time = phone frfam ad
female) (s2:),  technique(nr bhhh ) maximize robust init(b)
est sto m6
```

In the results provided below, we can see that the GMM and cfunction estimations for the model in hand provide very similar results for the estimation of the impact of the endogenous variable. The additive model provides the smallest estimated standard error, nearly half as large as the other models. The results from the QMLE 2SRI model in this case appear to be quite different than the specification obtained using the IV Poisson cfunction. The reason for this difference is that the QMLE is specifying the model as if the errors were additive rather than multiplicative. Using a simpler specification (log linear model]) shows results that are closer to the IV Poisson cfunction approach, with standard errors that are somewhat smaller, whereas the specification used in the additive model provides results similar to the 2SRI shown in Column 4.

Table 5 IV poisson estimations

	ivpoisson gmm add error	ivpoisson gmm mult error	ivpoisson gmm Cfunction	QMLE-2sri	QMLE lny=xb+e	QMLE y=exp(xb)+e
main						
time	0.05893	0.06144	0.06836	0.03896	0.06794	0.03498
	(0.0108)	(0.0181)	(0.0182)	(0.0173)	(0.0169)	(0.0172)
ad	0.13734	0.13929	0.14233	0.14184	0.15729	0.13997
	(0.0102)	(0.0112)	(0.0128)	(0.0122)	(0.0120)	(0.0122)
female	-0.02477	-0.00909	0.00144	-0.04497	0.00401	-0.08436
	(0.0376)	(0.0443)	(0.0437)	(0.0399)	(0.0405)	(0.0368)
_cons	1.04151	0.99201	0.94221	1.08001	0.82229	1.15203
	(0.0386)	(0.0627)	(0.0542)	(0.0487)	(0.0505)	(0.0442)
c_time/g			0.07968	0.07745	0.0804	0.06032

			(0.0216)	(0.0203)	(0.0197)	(0.0208)
_cons						
lns1				0.80688	-0.87683	0.69126
_cons				(0.0603)	(0.0355)	(0.0781)
N	500	500	500	500	500	500

Note: Robust Standard errors in parenthesis. First Stage results available upon request. The "g" coefficient corresponds c_time coefficient from the Ivpoisson cfunction output.

4.5. Comparing QMLE to Terza (2016)

In Terza (2016), the authors propose a simplified method to address the correction of standard errors in two-stage procedures that could be applied for the appropriate estimation of endogenous models in the framework of Terza et al. (2008) and the two-stage residual imputation approach. In this paper and its companions, the author outlines the simplified correction of the standard errors when the first and second stage of the models are estimated using NLS and MLE estimations. In the previous examples thus far, we have presented the estimation of the models under the assumption that both the first and second stage of the models are estimated using MLE or both are estimated using OLS. While empirically the QMLE cannot be used when the joint ML functions are a combination of MLE and NLS objective functions, if the objective functions for both models follow the NLS objective function, the proposed estimation method is valid and identical to the correction provided in Terza (2016). For comparison, we will use the same data provided with the Terza (2016) ancillary files.

The model is based on the data from Mullahy (1997). The model estimated in this paper and replicated in Terza (2016) can be written as follows:

$$\text{Outcome: } y_1 = \exp(\beta_y y_2 + \beta_x X + \beta_u u) + v$$
$$\text{First Stage: } y_2 = \exp(\gamma_x X + \gamma_z Z) + u$$

where X is a set of exogenous variables, Z is the instrument, and y2 is the endogenous variable. This is a direct application of the 2SRI for the treatment of endogenous variables when the outcome equation is nonlinear. As indicated in Terza (2016), in the original study the model fit using a GMM estimator. Although not discussed, the equivalent specification for a GMM model would be to use a first stage that is linear in parameters, instead of the exponential model, for comparison:

$$\text{First Stage b: } y_2 = \gamma_x X + \gamma_z Z + u$$

Either model here can be estimated using MLE, assuming the normality of the errors v, u. Alternatively, we specify the objective function to minimize the sum of the squared errors. This specification is valid but requires both standard errors to be reported using the robust option in the estimation. For the benchmark model, assuming normality with a nonlinear function in both stages, the corresponding objective function can be programmed as follows:

```
program my2sri
  args lnf xb1 g s1 xb2 s2
  qui:replace `lnf'=ln(normalden($ML_y2,exp(`xb2'),exp(`s2')))
  qui:replace `lnf'=`lnf'+ln(normalden($ML_y1,exp(`xb1'+`g'*($ML_y2-exp(`xb2'))),exp(`s1')))
end
```

If we assume that the first stage of the model is linear, we can compare the results to the GMM procedure. The above program can be modified, and thus we obtain:

```
program my2srib
  args lnf xb1 g s1 xb2 s2
  qui:replace `lnf'=ln(normalden($ML_y2,(`xb2'),exp(`s2')))
  qui:replace `lnf'=`lnf'+ln(normalden($ML_y1,exp(`xb1'+`g'*($ML_y2-(`xb2'))),exp(`s1')))
end
```

Finally, if we instead want to estimate both stages of the model using NLS, we can easily modify the above programs to do so:

```
program my2sric
     args lnf xb1 g  xb2
     qui:replace `lnf'=-($ML_y2-exp(`xb2'))^2
     qui:replace `lnf'=`lnf'-($ML_y1-exp(`xb1'+`g'*($ML_y2-exp(`xb2'))))^2
end
```

Note that in this case, the objective function is the negative of the squared residual, such that the estimation would be equivalent to an NLS. For the implementation of the model, we estimate the models as follows:

```
use mullahy-birthweight-data.dta, clear
generate BIRTHWTLB=BIRTHWT/16
* Terza(2016) replication are taken directly from the paper.
* GMM estimation
gmm (BIRTHWTLB/exp({xb:CIGSPREG PARITY WHITE MALE} + {b0})-1), ///
   instruments(PARITY WHITE MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88)
 est sto m2
* As in previous cases, the model is rather difficult to estimate without good initial values
```

```
* Here we start by estimating a simplyfied model assuming g=0
constrain 1 [g]_cons=0
ml model lf my2sri (xb1:BIRTHWTLB=CIGSPREG PARITY WHITE MALE)  (g:)  (s1:)  (xb2:CIGSPREG =PARITY
WHITE MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88) (s2:) ,  maximize technique(nr bhhh)  robust
constrain(1)
matrix b=e(b)
* Benchmark model. Y1=exp(xb+c1*y2+c2*u)+e y2=exp(zb)+u
ml model lf my2sri (xb1:BIRTHWTLB=CIGSPREG PARITY WHITE MALE)  (g:)  (s1:)  (xb2:CIGSPREG =PARITY
WHITE MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88) (s2:) ,  maximize technique(bhhh nr)  init(b)
robust
est sto m3
* alternative model. GMM comparable Y1=exp(xb+c1*y2+c2*u)+e y2= zb+u
ml model lf my2srib (xb1:BIRTHWTLB=CIGSPREG PARITY WHITE MALE)  (g:)  (s1:)  (xb2:CIGSPREG
=PARITY WHITE MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88) (s2:) ,  maximize technique(bhhh nr)
init(b) robust
est sto m4
* NLS equivalent model. Both stages are estimated using NLS
ml model lf my2sric (xb1:BIRTHWTLB=CIGSPREG PARITY WHITE MALE) (g:) (xb2:CIGSPREG =PARITY WHITE
MALE EDFATHER EDMOTHER FAMINCOM CIGTAX88) ,  maximize technique(bhhh nr)  init(b, skip) robust
est sto m5
```

The results of the above estimations are shown below. In Columns 1 and 2, we report the corrected estimates from Terza (2016) and the GMM estimations from Mullahy (1997). Although these estimates are similar to each other, the GMM estimate corresponding to the endogenous variable CIGSPREG is nearly 50% larger in the 2SRI approach; all other estimates are nearly the same in terms of point estimates and standard errors.

Under the assumption of normality in Column 3, the QMLE results are close to those in Column 1. In Column 4, in which we modify the specification of the first stage to a linear model rather than an exponential one, we obtain point estimates that are closer to the GMM estimates, with standard errors for the endogenous variable that are slightly larger than the GMM estimates indicate. Finally, in Column 5, we present the estimates from the NLS estimations. Here, the point estimates are almost identical to Terza's, with standard errors that are equally close, with the differences possibly due to estimating the model using the ML estimator rather than the standard OLS estimator.

Table 6 Mullahy(1997) Model estimations

	Terza(2016)	GMM	QMLE 1st y2=exp(zb)+e	QMLE 1st y2=(zb)+e	QNLS
main					
CIGSPREG	-0.01401	-0.00989	-0.0145	-0.01092	-0.01403
	(0.0038)	(0.0029)	(0.0041)	(0.0033)	(0.0039)

PARITY	0.01666	0.01765	0.01678	0.0173	0.01667
	(0.0052)	(0.0053)	(0.0054)	(0.0053)	(0.0053)
WHITE	0.05363	0.054	0.05347	0.05486	0.05362
	(0.0127)	(0.0121)	(0.0130)	(0.0122)	(0.0130)
MALE	0.02979	0.02708	0.03089	0.0262	0.02985
	(0.0095)	(0.0092)	(0.0098)	(0.0092)	(0.0097)
_cons	1.94821	1.93898	1.94851	1.94103	1.94823
	(0.0166)	(0.0159)	(0.0168)	(0.0156)	(0.0167)
xuhat/g	0.00978		0.0103	0.00657	0.00981
	(0.0038)		(0.0041)	(0.0034)	(0.0039)
lns1			0.21148	0.21298	
			(0.0315)	(0.0315)	
N	1388	1388	1388	1388	1388

Note: Robust Standard errors in parenthesis. First Stage results available upon request. The "g/xuhat" coefficient corresponds to the coefficient of the first stage residuals as in Terza (2016).

5. Summary

In the past decade, a few papers have offered various strategies and codes to correct the estimation of standard errors in the framework of two-step models, including Hardin (2002) and Hole (2006), and more recently Terza (2016). While these strategies are potentially as simple to be implemented as Bootstrap methods are, Bootstrap methods are usually preferred in the empirical literature.

In this paper we provide an alternative strategy for the estimation of two-step models using a joint Quasi Likelihood function as described in Wooldridge (2014). While this method also requires some level of programming, it can be easily extended to allow for a large set of two-step models with minimum changes to the programs shown in text. We compare the results obtained with this strategy with the results from built-in commands in Stata, showing its reliability. We also show how the model compares to the results provided in Terza (2016). The use of this estimation method may facilitate the estimation of non-linear models with endogenous variables in the spirit of Terza et al. (2008) and Wooldridge (2014, 2015), including multinomial models with continuous endogenous explanatory variables.

6. References

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. College Station, TX: Stata Press.

Greene, W. H. (2003). *Econometric Analysis* (5 ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hardin, J. W. (2002). The robust variance estimator for two-stage models. *The Stata Journal*, 2(3), 253-266.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 153-161.

Hole, A. R. (2006). Calculating Murphy-Topel Variance Estimates in Stata: A Simplified Procedure. *The Stata Journal*, 6(4), 521-529.

Mullahy, J. (1997). Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior. *The Review of Economics and Statistics*, 586-593.

Murphy, K. M., & Topel, R. H. (1985). Estimation and Inference in Two-Step Econometric Models. *Journal of Business & Economic Statistics*, 3(4), 370-379.

Newey, W. K. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*. 36: 231-250.

Newey, W. K., & McFadden, D. L. (1994). Large Sample Estimation and Hypothesis Testing. In R. F. Engle, & D. L. McFadden, *Handbook of Econometrics* (Vol. 4, pp. 2111-2245). Amsterdam: Elsevier B.V.

Terza, J. V. (2016). Simpler Standard Errors for Two-Stage Optimization Estimators. *The Stata Journal*, 16(2), 368-385.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27, 531-543.

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

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. Journal of Econometrics, 182, 226-234.

Wooldridge, J. M. (2015). Control Function Methods in Applied Econometrics. Journal of Human

Resources, 50(2), 420-455.