Usage of an estimated coefficient as a dependent variable*

Abigail S. Hornstein**
Wesleyan University

William H. Greene***

Stern School of Business, New York University

February 28, 2012

Abstract:
Two-step estimation with large panel data sets generally involves estimating vectors of individual-specific coefficients in a first-stage. In a second-stage estimation a vector of estimated coefficients is used as the dependent variable. Potential problems of heteroskedasticity in the second stage may be mitigated by weighting all independent observations by the inverse of the variance of the dependent variable, which is obtained from the first stage estimation. This approach needs to be modified if the dependent variable in the second stage is a non-linear function of the estimated coefficient.

Key words: two-step estimation, heteroskedasticity, random parameters, GLS, OLS

JEL Classifications: C1, C3, C4, C5

---

* We thank Lawrence White and Bernard Yeung, the editor, and two anonymous referees for comments. Financial support for Hornstein was provided by the Ford Foundation/Aspen Institute and Paul Willensky fellowship at New York University's Stern School of Business. All errors remain ours.
** Wesleyan University, 238 Church Street PAC 123, Middletown, CT 06459, 860-685-3049, fax: 860-685-2301, ahornstein@wesleyan.edu; corresponding author.
*** Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012, wgreene@stern.nyu.edu.

1. Introduction

Estimation using large panel data sets often involves computing vectors of individual-specific coefficients in a first-stage regression. In a second stage regression a vector of estimated coefficients may be used as the dependent variable. Heteroskedasticity at the second stage is accommodated by weighting all observations by the inverse of the variance of the dependent "variable", where the weights are obtained from the first stage estimation. Saxonhouse (1976) was the first to show how the weighting procedure could be done effectively. This two-stage estimation may be appropriate when the variable of interest cannot be measured directly from available data. Use of the random parameters (RP) estimation methodology in the first stage may be advantageous as it simultaneously pools data across individuals while allowing for individual-specific estimates for each coefficient. In Saxonhouse's original method, the first stage equation was estimated by OLS, firm by firm. The RP methodology explicitly incorporates the possibility of firm heterogeneity while yielding more precise, efficient coefficient estimates for each firm by leveraging all information about all firms. The two-stage model can be generalized such that the first stage is $y_{it} = \gamma_i ' w_{it} + \beta_i ' x_{it} + \varepsilon_{it}$ where $\hat{\beta}_i$ is the coefficient of interest. The vector of estimated coefficients, $\underline{\hat{\beta}}$, is used to form the dependent variable in the second stage.

The model that motivates the two step approach can be written as a single estimable equation in which $\beta_i = \beta + \Delta s_i + v_i$ is substituted directly into the main equation. When the entity of interest is the coefficient itself, either the one step RP or the two step estimator will be useable. However, the RP methodology will not be sufficient when the researcher is interested in examining a function of the estimated coefficient, as we are here. Nonetheless, it may be advantageous to use an RP estimator at the first step in order to obtain more precise estimates of the coefficients of interest.

We present an empirical application showing how a function of an estimated coefficient can be used as a dependent variable, and show the advantage of an updated Saxonhouse weighted GLS procedure.

2. Dependent variable as a function of an estimated coefficient

An illustrative application of the two-stage estimation technique is found in Durnev *et al.* (2004) and Greene *et al.* (2009) who each examine the efficiency of corporate capital budgeting decisions. When a firm uses capital effectively, the marginal investment should have a net present value of zero. The marginal investment undertaken by a publicly-traded firm has a marginal *q*, which is the ratio of the change in market value to the change in book value of the firm over a specified time period. Durnev *et al.* used OLS to estimate their first stage regression while Greene *et al.* used RP methodology. Both sets of researchers then tested whether the deviation of marginal *q* from its theoretical tax-adjusted benchmark value, *h*, could be explained by firm characteristics of interest.

Using an unbalanced panel dataset of U.S. manufacturing firms from 1992-2000, Greene *et al* used the following equation to estimate marginal *q* for each firm in the dataset, with the subscript *i* denoting the firm (1…*I*); *j*, coefficient number (0…3); and *t*, time (1…T). The parameters of this model are estimated using maximum simulated likelihood. The model contains four random coefficients, $\beta_{i,0} \ldots \beta_{i,3}$, and year fixed effects.

$$\frac{\Delta V_{i,t}}{A_{i,t-1}} = \beta_{i,0} + \beta_{i,1}\frac{\Delta A_{i,t}}{A_{i,t-1}} + \beta_{i,2}\frac{V_{i,t-1}}{A_{i,t-1}} + \beta_{i,3}\frac{D_{i,t-1}}{A_{i,t-1}} + \delta_t + u_{i,t} \qquad [1]$$

such that $\beta_{i,1}$ is firm *i*'s marginal *q*.

This model can be rewritten more generically as

$$y_{i,t} = \beta_{i,0} + \beta_{i,1}X_{i,1,t} + \beta_{i,2}X_{i,2,t} + \beta_{i,3}X_{i,3,t} + \delta_t + u_{i,t} \qquad [2]$$

where $u_{i,t} \sim N[0,\sigma^2]$ or

$$y_{i,t} = \sum_{j=0}^{3}\beta_{i,j}X_{i,j,t} + \delta_t + u_{i,t} \text{ where } X_{i,0,t} = I\,. \qquad [2']$$

The four coefficients $\beta_{i,j}$ are assumed to be random coefficients due to the effects of firm heterogeneity. These coefficients are defined as:

$$\beta_{i,j} = \beta_j + v_{i,j}\ . \qquad [3]$$

This equation is expressed more generally as $\underline{\beta_j} = \underline{\beta} + \underline{v_i}$ such that $v_i \sim N[0,\Sigma]$ for all *i* = 1…*I* and for all *j* = 0… 3, and Σ is a diagonal matrix.

The Saxonhouse technique is used without modification in the second stage regression when the dependent variable is the estimated coefficient. However, when the dependent variable is a non-linear transformation of the estimated coefficient, the standard errors obtained in the first stage estimation must be adjusted to reflect the transformation of the estimated coefficient to become the dependent variable. In the second stage regressions Durnev *et al*. (2004) used as their dependent variable the deviation of the estimated marginal *q* from the theoretical benchmark value of 1.0. They used two different functions of this deviation: $\left|\hat{\beta}_i - 1.0\right|$ or $\left(\hat{\beta}_i - 1.0\right)^2$, where the second construction attached a penalty weight to those firms that had estimated marginal *q*'s very distant from 1.0. Greene et al. (2009) instead conducted separate examinations of those firms with estimated marginal *q* above and below the theoretical benchmark value.

If the second stage regression is weighted by the inverse of $\sigma_{\hat{\beta}_i}$ instead of $\sigma_{f(\hat{\beta}_i)}$, the disturbances in the equation will still be heteroskedastic because the variance of the function of β_i is not equal to the variance of $\bar{\beta}_i$. The researcher should therefore use the known information – $\hat{\beta}_i$ and $\sigma_{\hat{\beta}_i}$ – to estimate the standard error of $f\left(\hat{\beta}_i\right)$, $\sigma_{f(\hat{\beta}_i)}$. For example, the function $f\left(\hat{\beta}_i\right) = \left|\hat{\beta}_i - 1\right|$ is a continuous, non-differentiable variable but it should be treated as a variant on the truncated regression model because

$$
\begin{aligned}
|\hat{\beta}_{1,j} - 1| \quad &= \hat{\beta}_{1,j} - 1 \text{ if } \hat{\beta}_{1,j} \geq 1 \quad \text{and} \\
&= 1 - \hat{\beta}_{1,j} \text{ if } \hat{\beta}_{1,j} < 1.
\end{aligned}
$$

Consequently, the variance of $|\hat{\beta}_{1,j} - 1|$ is estimated as:

$$
\begin{aligned}
Var\,|\hat{\beta}_{1,j} - 1| \; &= \Pr(\hat{\beta}_{1,j} \geq 1) * Var(\hat{\beta}_{1,j} \mid \hat{\beta}_{1,j} \geq 1) + \Pr(\hat{\beta}_{1,j} < 1) * Var(\hat{\beta}_{1,j} \mid \hat{\beta}_{1,j} < 1) \\
&+ \Pr(\hat{\beta}_{1,j} \geq 1)\left(E(\hat{\beta}_{1,j} \mid \hat{\beta}_{1,j} \geq 1) - E(\hat{\beta}_{1,j})\right)^2 \\
&+ \Pr(\hat{\beta}_{1,j} < 1)\left(E(\hat{\beta}_{1,j}) - E(\hat{\beta}_{1,j} \mid \hat{\beta}_{1,j} < 1)\right)^2. \qquad [4]
\end{aligned}
$$

In line with the Saxonhouse methodology, when the dependent variable is $|\hat{\beta}_{1,j} - 1|$, all observations are weighted by the inverse of the standard error of $|\hat{\beta}_{1,j} - 1|$.Similarly, when $(\hat{\beta}_{1,j} - 1)^2$ is used as the dependent variable, the variance of this function is estimated as:

$$
\begin{aligned}
Var(\hat{\beta}_{1,j}-1)^2 &= Var(\hat{\beta}_{1,j}{}^2-2\hat{\beta}_{1,j}+1) \\
&= Var(\hat{\beta}_{1,j}{}^2)+4Var(\hat{\beta}_{1,j})-4Cov(\hat{\beta}_{1,j},\hat{\beta}_{1,j}{}^2) \\
&= 2(\sigma^4_{\hat{\beta}_{1,J}})+4(\sigma^2_{\hat{\beta}_{1,J}})\,. \qquad [5]
\end{aligned}
$$

Since it is assumed that $\hat{\beta}_{1,j}$ is normally distributed, then Cov($\hat{\beta}_{1,j}$, $\hat{\beta}_{1,j}{}^2$) equals zero as the first and second powers are uncorrelated. Following the Saxonhouse methodology, when the dependent variable is ($\hat{\beta}_{1,j}-1)^2$, all observations are weighted by the inverse of the standard error of ($\hat{\beta}_{1,j}-1)^2$.

On the other hand, Greene et al. (2009), used a similar two stage empirical framework with two modifications. First, the initial estimation of marginal *q* as an estimated coefficient used RP methodology, not OLS as in Durnev et al. (2004). Then, in the second stage regression the dependent variable was ($\hat{\beta}_{1,j}-1$) with separate analysis of firms with estimated $\hat{\beta}_{1,j}>1$ and $\hat{\beta}_{1,j}<1$. Thus, in this application, the variance of this function is

$$Var(\hat{\beta}_{1,j}-1)=Var(\hat{\beta}_{1,j})-Var(1)=\sigma^2_{\hat{\beta}_{1,J}} \qquad [6]$$

This example demonstrates that the Saxonhouse technique can be used without modifications when the dependent variable is a linear transformation of an estimated coefficient.

This model can be generalized for other applications whereby an estimated coefficient from one regression is used in a non-linear function as a dependent variable in a subsequent estimation. In such an instance the researcher would want to obtain the estimated variance of the non-linear function and then weight all observations by the inverse of the standard error of this non-linear function.

3. Conclusion

This estimation procedure combines the RP methodology with the Saxonhouse technique for weighted GLS. This particular combination of techniques is well suited for empirical estimation when heterogeneity should be accounted for explicitly in the empirical estimation, and when the estimated coefficients are later used to form the dependent variable in a second stage empirical estimation. The RP methodology will yield a series of estimates for each firm of the coefficient of interest along with the standard error of each estimate. The estimated standard error of these estimates is then used as the weighting matrix for the GLS estimation of the second stage regression. This combination of techniques allows more precise estimation of the coefficient of interest in the first stage and greatly decreases the standard errors of estimates obtained in the second stage regression.

REFERENCES

Durnev, Art, Randall Morck, and Bernard Yeung. 2004. "Value enhancing capital budgeting and firm-specific stock returns variation," *Journal of Finance*, *59*(1), 65-105.

Greene, William H., Abigail S. Hornstein, and Lawrence J. White. 2009. "Multinationals Do It Better: Evidence on the Efficacy of Corporations' Capital Budgeting," *Journal of Empirical Finance*, *16*(5), 703-720.

Saxonhouse, Gary R. 1976. "Estimated Parameters as Dependent Variables," *American Economic Review*, *66*(1), pp. 178-183.