Department of Economics

ON WEIGHTED ESTIMATION IN LINEAR REGRESSION IN THE PRESENCE OF PARAMETER UNCERTAINTY

Judith A. Clarke

Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2

September, 2007

Abstract

We examine the problem of estimating the linear regression model's coefficients when there are uncertain linear restrictions about these parameters. Theorems are provided that generalize results obtained by Magnus and Durbin (1999) and Danilov and Magnus (2004).

Author Contact:

Judith A. Clarke, Tel.: +1 250 721 8542; fax.: +1 250 721 6214

Email address: jaclarke@uvic.ca

1. Introduction

Consider estimating the coefficients of the model

$$y = X\beta + Z\gamma + u, \qquad u \sim N(0, \sigma^2 I_{n \times n}) \tag{1}$$

where $y(n \times 1)$ is a vector of observations, $X(n \times k)$ and $Z(n \times m)$ are nonrandom regressor matrices, $\beta(k \times 1)$ and $\gamma(m \times 1)$ are parameter vectors. With $W \equiv [X : Z]$ (of full column-rank $K = (k+m)$) and $\varphi \equiv [\beta' \ \gamma']'$, model (1) is

$$y = W\varphi + u . \tag{2}$$

Suppose, in addition, J linear uncertain beliefs exist about φ expressed as $R\varphi = r$; $r(J \times 1)$ and $R(J \times K)$ are nonstochastic with R of full row-rank, $J < K$. This framework extends that of Magnus and Durbin (1999) and Danilov and Magnus (2004) who suppose that X contains required variables, whereas Z consists of doubtful regressors, included to perhaps provide a "better" estimator of β; i.e., their prior beliefs regard exclusion restrictions. See also Magnus (1999, 2002), Danilov (2005) and Zou et al. (2007). Given the uncertainty about γ, these authors consider the following estimator of β

$$b_{w,1} = \lambda b_u + (1-\lambda)b_r \tag{3}$$

where b_u is the unrestricted least squares (OLS) estimator, b_r is the restricted least squares (RLS) estimator obtained under $\gamma = 0$, $\lambda \equiv \lambda(g_u, \hat{u})$ is a random weight function, $\hat{u}$ is the OLS residual vector and g_u is the OLS estimator of γ; the dependence of λ on $\hat{u}$ is usually via an error variance estimator. Using an F-test to choose between b_u and b_r results when $\lambda = I_{[0,c]}(F)$, where F is the F-test statistic, c is the critical value from an F distribution with m and n-K degrees of freedom, and $I_{[0,c]}(F) = 1$ when $F \in [0,c]$, 0 otherwise. Then, $b_{w,1}$ is the traditional pretest estimator (e.g., Judge and Bock, 1978; Giles and Giles 1993), inadmissible and never preferred, using risk under quadratic loss or mean squared error, to either of its component estimators. Indeed, the risk of the pretest estimator can be greater than either of its component estimators, an unattractive feature for a strategy adopted to improve knowledge on β.

As the aim is to obtain a preferred estimator of β, as opposed to undertaking a test about whether $\gamma = 0$, it makes sense to allow λ to be a continuous function (with $0 \leq \lambda$

≤ 1). This gives rise to many possible combination estimators, including Magnus and Durbin's (1999) weighted-average least squares (WALS) estimator, the shrinkage estimators of Judge and Bock (1978, pp.240-42), which extend the James and Stein (1961) estimators, and the James-Stein type estimator of Kim and White (2001).

The goal is to optimally mix the component estimators based on a chosen criterion; e.g., mean squared error (MSE) or risk under quadratic loss. First glance suggests that this task will depend on the model's features: β, γ, X, Z and σ^2. Indeed, this has led researchers to assume orthonormal regressors or to explore risk of the prediction vector (E(y|X,Z)) rather than the coefficient vector. No longer is this needed with Magnus and Durbin's elegant "Equivalence Theorem". This theorem shows that determining λ to minimize the MSE of $b_{w,1}$ reduces to ascertaining λ such that the MSE of $\lambda\hat{\theta}$ is minimized where $\hat{\theta} = (Z'MZ)^{1/2} g_u \sim N(\theta, \sigma^2 I_{m\times m})$ with M = $I_{n\times n} - X(X'X)^{-1}X'$. We need only determine λ such that $\lambda\hat{\theta}$ is a preferred estimator of θ - the mean vector of an m-variate normal distribution; a task that is independent of specific regression details.

Mixing just b_r and b_u is likely restrictive – *all* of the regressors in Z are either in or out. Researchers may examine partially restricted models that contain *some* of Z's columns. There are 2^m models to choose between with m auxiliary regressors; let $\mathscr{M}_i$ be the model that imposes that none, one, some or all of the elements of γ are zero, b_i the subsequent LS estimator of β, g_i the restricted estimator of γ and $\hat{u}_i$ the associated residual vector, i=1,...,2^m; i.e., $\hat{u}_i = y - Xb_i - Zg_i = y - Wp_i$. The combination estimator that weights all possible 2^m estimators of β is examined by Danilov and Magnus (2004):

$$b_{w,2} = \sum_{i=1}^{2^m}\lambda_i b_i.$$
(4)

with weights that satisfy $\lambda_i \geq 0$, $\sum_{i=1}^{2^m}\lambda_i = 1$ and $\lambda_i \equiv \lambda_i(\hat{\theta}, \hat{u}_i)$. Note: $b_i = b_u$ (= b_r) when all (none) of Z's columns are regressors and $b_{w,2}$ collapses to $b_{w,1}$ when only b_u and b_r are combined. Danilov and Magnus extend the equivalence theorem to cover the weighted estimator defined by expression (4).

These significantly useful findings are shown for estimating β when there exists uncertainty about which auxiliary regressors to include in the specification. In this note, we generalize these equivalence theorems to the estimation of φ, so covering estimation of γ in addition to β, when the prior beliefs relate to any linear combination of the components of φ rather than simply exclusion restrictions. The optimal (in terms of MSE) combination estimator is determined solely by ascertaining the optimal estimator of the mean of a normal random variate with unknown variance, which has nothing to do with the regression model's structure nor the specific form of the prior linear beliefs.

2. Setup

Our focus is on estimating the full coefficient vector φ with uncertain beliefs $R\varphi=r$. Allowing for the unrestricted model, the fully restricted model and all possible partially restricted models that incorporate some of the J restrictions, there are 2^J models to consider; let $\mathcal{M}_i$ be the i'th model ($i=1,\ldots,2^J$). Let A_i be a $J\times a_i$ selection matrix of rank $a_i \geq 0$ (i.e., $A_i' = [I_{a_i \times a_i} \ 0]$ or a column-permutation thereof) such that model $\mathcal{M}_i$ corresponds to model (2) subject to $A_i'R\varphi = A_i'r$. The matrix R is of full row rank, which implies that $A_i'R$ also has full row rank. As before, we denote p_i as the LS estimator of φ associated with model $\mathcal{M}_i$ with residual vector $\hat{u}_i = y - Wp_i$. Define

$S = W'W,$ $\quad\quad\quad\quad\quad\quad\quad\quad p_u = S^{-1}W'y,$

$p_r = p_u - S^{-1}R'[RS^{-1}R']^{-1}(Rp_u\text{-}r),$ $\quad p_i = p_u - S^{-1}R'A_i[A_i'RS^{-1}R'A_i]^{-1}A_i'(Rp_u\text{-}r),$

$\theta = [RS^{-1}R']^{-1/2}(R\varphi\text{-}r),$ $\quad\quad\quad\quad \hat{\theta} = \theta + [RS^{-1}R']^{-1/2}RS^{-1}W'u,$

$P_i = [RS^{-1}R']^{1/2}A_i[A_i'RS^{-1}R'A_i]^{-1}A_i'[RS^{-1}R']^{1/2},$

$M_W = I_{n\times n} - WS^{-1}W',$ $\quad\quad\quad\quad Q = S^{-1}R'[RS^{-1}R']^{-1/2},$

$\hat{u} = M_W u,$ $\quad\quad\quad\quad\quad\quad\quad\quad \hat{u}_i = \hat{u} + WQP_i\hat{\theta},$

$\Omega = S^{-1} - QQ',$ $\quad\quad\quad\quad\quad\quad \Omega_i = S^{-1} - QP_iQ',$

$B_i = I_{J\times J} - P_i,$ $\quad\quad\quad\quad\quad\quad H_i = WQP_i.$

The OLS estimator of φ is $p_u = p_i$ with $A_i = P_i = 0_{J \times J}$ and the RLS estimator, obtained from imposing all J restrictions, is $p_r = p_i$ with $A_i = P_i = I_{J \times J}$. Consider the combination estimator:

$$p_{w,1} = \lambda p_u + (1-\lambda) p_r, \tag{5}$$

where the weight λ satisfies $0 \leq \lambda \leq 1$ with $\lambda \equiv \lambda(\hat{\theta}, \hat{u})$. We also consider a more general weighted estimator:

$$p_{w,2} = \sum_{i=1}^{2^J} \lambda_i p_i. \tag{6}$$

where the weights satisfy $\lambda_i \geq 0$, $\sum_{i=1}^{2^J} \lambda_i = 1$ and $\lambda_i \equiv \lambda_i(\hat{\theta}, \hat{u}_i)$. The estimators $p_{w,1}$ and $p_{w,2}$ generalize, respectively, the weighted estimators of Magnus and Durbin (1999) and Danilov and Magnus (2004) to all coefficients when there is uncertainty about linear combinations of these parameters.

3. Generalized equivalence theorems

As $p_r = p_u - Q\hat{\theta}$ and $p_i = p_r + Q B_i \hat{\theta}$ we have

$$
\begin{bmatrix} p_u \\ \hat{\theta} \\ p_r \\ p_i \\ \hat{u} \\ \hat{u}_i \end{bmatrix} \sim N \left(\begin{bmatrix} \varphi \\ \theta \\ \varphi - Q\theta \\ \varphi - QP_i\theta \\ 0 \\ H_i\theta \end{bmatrix}, \sigma^2 \begin{bmatrix} S^{-1} & Q & \Omega & \Omega_i & 0 & QH_i' \\ Q' & I_{J \times J} & 0 & B_iQ' & 0 & H_i' \\ \Omega & 0 & \Omega & \Omega & 0 & 0 \\ \Omega_i & QB_i & \Omega & \Omega_i & 0 & 0 \\ 0 & 0 & 0 & 0 & M_W & M_W \\ H_iQ' & H_i & 0 & 0 & M_W & M_W + H_i'H_i \end{bmatrix} \right). \tag{7}
$$

Using (7) enables us to establish the following theorems.

Theorem 1. Denote $p_{w,1} = \lambda p_u + (1-\lambda) p_r$, where $\lambda \equiv \lambda(\hat{\theta}, \hat{u})$, $u \sim N(0, \sigma^2 I_{n \times n})$ and $\hat{\theta} \sim N(\theta, \sigma^2 I_{J \times J})$. Then, the MSE of $p_{w,1}$ is

$$MSE(p_{w,1}) = \sigma^2 \Omega + Q[MSE(\tilde{\theta}_1)]Q' \tag{8}$$

where $\tilde{\theta}_1 = \lambda \hat{\theta}$.

Proof. As $p_u = p_r + Q\hat{\theta}$, $p_{w,1} = p_r + Q\tilde{\theta}_1$ and using that p_r is independent of $\hat{\theta}$ and $\hat{u}$, we have $E(p_{w,1} | \hat{\theta}, \hat{u}) = E(p_r) + Q\tilde{\theta}_1$. So

5

$$E(p_{w,1}) = \varphi + Q[E(\tilde{\theta}_1 - \theta)]. \tag{9}$$

In addition, $\text{var}(p_{w,1} \mid \hat{\theta}, \hat{u}) = \text{var}(p_r) = \sigma^2 \Omega$, which implies

$$\text{var}(p_{w,1}) = \sigma^2 \Omega + Q[\text{var}(\tilde{\theta}_1)]Q'. \tag{10}$$

Combining expressions (9) and (10) completes the proof.

Theorem 2. Denote $p_{w,2} = \displaystyle\sum_{i=1}^{2^J} \lambda_i p_i$ where $\lambda \equiv \lambda(\hat{\theta}, \hat{u}_i)$, $u \sim N(0, \sigma^2 I_{n \times n})$ and

$\hat{\theta} \sim N(\theta, \sigma^2 I_{J \times J})$. Then, the MSE of $p_{w,2}$ is

$$\text{MSE}(p_{w,2}) = \sigma^2 \Omega + Q[\text{MSE}(\tilde{\theta}_2)]Q' \tag{11}$$

where $\tilde{\theta}_2 = B\hat{\theta}$ with $B = \displaystyle\sum_{i=1}^{2^J} \lambda_i B_i$.

Proof. Let $B = \displaystyle\sum_{i=1}^{2^J} \lambda_i B_i$. As $p_i = p_r + QB_i\hat{\theta}$, $p_w = p_r + QB\hat{\theta}$ then, since p_r is

independent of $\hat{\theta}$ and $\hat{u}_i$, the conditional mean is $E(p_{w,2} \mid \hat{\theta}, \hat{u}_i) = E(p_r) + Q\tilde{\theta}_2$

with $\tilde{\theta}_2 = B\hat{\theta}$. The unconditional mean is then

$$E(p_{w,2}) = \varphi + Q[E(\tilde{\theta}_2 - \theta)]. \tag{12}$$

In addition, $\text{var}(p_{w,2} \mid \hat{\theta}, \hat{u}_i) = \text{var}(p_r) = \sigma^2 \Omega$, which implies

$$\text{var}(p_{w,2}) = \sigma^2 \Omega + Q[\text{var}(\tilde{\theta}_2)]Q'. \tag{13}$$

The result then follows by combining (12) and (13).

Theorems 1 and 2 show that the MSE properties of the weighted estimators of φ crucially depend on the MSE properties of the weighted estimator of θ, the mean

vector of a normal random variate. Underlying θ is the compatibility of the prior information with the coefficient vector: $R\varphi-r$. This implies the irrelevance of the specific regression data in determining how best to weigh the various estimators of φ.

Special cases of Theorem 1 and 2 are, respectively, Theorem 2 of Magnus and Durbin (1999) and Theorem 1 of Danilov and Magnus (2004). For both, let J=m,

$R(m \times K) = \begin{bmatrix} 0_{m\times k} & I_{m\times m} \end{bmatrix}$, $r = 0$, $p_u = [(b_r - \hat{\theta}g_u)' \quad g_u']'$, $p_r = [b_r' \quad g_r']'$,

$p_i = [b_i' \quad g_i']'$,

$b_r = (X'X)^{-1}X'y$, $g_u = (Z'MZ)^{-1}Z'My$, $g_r = 0$, $M = I_{n\times n} - X(X'X)^{-1}X'$,

$\hat{\theta} = (Z'MZ)^{1/2}g_u$, $\theta = (Z'MZ)^{1/2}\gamma$, $Q' = [-(Z'MZ)^{-1/2}Z'X(X'X)^{-1} \quad (Z'MZ)^{-1/2}] =$

$[Q_1' \; Q_2]$, $\Omega = \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & 0 \end{bmatrix}$

and $P_i = (Z'MZ)^{-1/2}A_i[A_i'(Z'MZ)^{-1}A_i]^{-1}A_i'(Z'MZ)^{-1/2}$. Define $b_{w,1} = \lambda b_u + (1-\lambda)b_r$, $g_{w,1} = \lambda g_u + (1-\lambda)g_r$, $b_{w,2} = \sum_{i=1}^{2^m} \lambda_i b_i$ and $g_{w,2} = \sum_{i=1}^{2^m} \lambda_i g_i$. Using our results with j=1,2:

$$MSE(b_{w,j}) = \sigma^2(X'X)^{-1} + Q_1 MSE(\tilde{\theta}_j)Q_1'$$

(14)

    (14)

& $MSE(g_{w,j}) = Q_2 MSE(\tilde{\theta}_j)Q_2$.

    (15)

Expression (14) corresponds to that presented by Magnus and Durbin (1999) for j=1 (i.e., $b_{w,1}$) and to that reported by Danilov and Magnus (2004) for j=2 (i.e., $b_{w,2}$). Collapsing the results in this fashion highlights that the same weight function that provides an optimal estimator of θ gives an optimal estimator of both γ and β, not just β as stressed by Magnus, Durbin and Danilov.

3. Concluding Remarks

We have considered combining estimators of the coefficient vector of a linear regression model in the presence of uncertainty about linear restrictions. This setup

generalizes that of Magnus and Durbin (1999) and Danilov and Magnus (2004), and also related work by Magnus (1999, 2002), Danilov (2005) and Zou et al. (2007), all of whom limit attention to estimating some of the coefficients when others are not of interest. We generalize their results to our framework: the optimal MSE weighted estimator is determined by optimally estimating the mean vector of an uncorrelated, homoskedastic distribution. This finding implies, for instance, that the extensive literature on how best to estimate the mean of a normal distribution carries over to the linear regression model with uncertain parameter beliefs.

References

Danilov, D., 2005. Estimation of the mean of a univariate normal distribution when the variance is not known. Econometrics Journal 8, 277-291.

Danilov, D., Magnus, J.R., 2004. On the harm that ignoring pretesting can cause. Journal of Econometrics 122, 27-46.

Giles, J.A., Giles, D.E.A., 1993. Pre-test estimation and testing in econometrics: Recent developments. Journal of Economic Surveys 7, 145-197.

James, W., Stein, C., 1961. Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 361-379.

Judge, G.G., Bock, M.E., 1978. The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics. North-Holland, Amsterdam.

Kim, T.H., White, H., 2001. James-Stein-type estimators in large samples with application to the least absolute deviations estimator. Journal of the American Statistical Association 96, 697-705.

Magnus, J.R., 1999. The traditional pretest estimator. Theory of Probability and its Applications 44, 293-308.

Magnus, J.R., 2002. Estimation of the mean of a univariate normal distribution with known variance. Econometrics Journal 5, 225-236.

Magnus, J.R., Durbin, J., 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. Econometrica 67, 639-643.

Zou, G., Wan, A.T.K., Wu, X., Chen, T., 2007. Estimation of regression coefficients of interest when other regression coefficients are of no interest: The case of non-normal errors. Statistics & Probability Letters/j.spl.2006.11.019.