Phuzicists Redux

I. Divide and U. Koenker

November, 1994

Since our first report, Divide and Koenker (1991), some persistent questions have arisen about the scope and methods of our original analysis. Indeed, we have even encountered the dreaded phrase "fundamentally unsound" muttered in the hallways. In this brief note we will compare several estimates of the intransigent model of productivity proposed in our 1992 paper in an effort to resolve the doubts which linger over our earlier work. Unfortunately, as we shall see presently, some doubts are surprisingly durable. Recent innovations in the estimation of panel data models do not adequately deal with our uncertainties, indeed in some cases they only seem to accentuate them.

We begin by reviewing some attempts to estimate the productivity model using the IV techniques introduced by Hausman and Taylor (1982). As we noted in our original paper, results are extremely sensitive to the choice of instruments. Recall that the Hausman-Taylor strategy is to divide the time-varying X's and time-invariant Z's into $[X_1 X_2 Z_1 Z_2]$ with $[X_1 Z_1]$ "exogenous" i.e., independent of the individual specific effects, α_i, and $X_2 Z_2$ endogenous and therefore possibly correlated with α_i. The instruments are then

$$V = [PX_1, QX_2, QX_2, Z_1]$$

which are adequate to identify the model as long as rank $(V) \geq$ rank $[X, Z]$ which will, barring some pathological bad luck, be satisfied if there are as many columns of X_1 as there are of Z_2. Of course, if there are other valid instruments for the problem in addition to those in V we are welcome to adjoin them to V.

At great personal expense we have developed a somewhat more extensive data set consisting of 1134 observations on 300 individuals. While the average length of the available time series is quite short for each person, the maximal length is now 13 spanning nearly the full length of a career. Recall that since the data is 3 year averages $T_i = 13$ corresponds to 39 years of post Ph.d. research.

The simplest version of the productivity model includes only the first order autoregressive term, a quadratic effect in experience and the dummy variables for sex and post-1970 Ph.d. Treating y_{it-1} as endogenous and all other variables as exogenous yields[1]

$$\begin{array}{lllllll} \hat{y}_{it} & = 1.64 & + .169 e_{it} & -.0041 e_{it}^2 & -.10 y_{it-1} & -.019 s & +.045 v \\ & (.08) & (.01) & (.0002) & (.05) & (.026) & (.022) \end{array}$$

To estimate the variance components for this model we observe that for $r_{it} = \alpha_i + u_{it}$ we have, denoting $r_{i\cdot} = T_i^{-1} \sum_{t=1}^{T_i} r_{it}$,

$$E r_{i\cdot}^2 = \sigma_\alpha^2 + T_i^{-1} \sigma_u^2.$$

Thus, we can estimate the model

$$\begin{array}{lll} \hat{r}_{i\cdot}^2 & = .06 & +.016 T_i^{-1} \\ & (.01) & (.017) \end{array}$$

[1] To simplify notation we will denote log productivity as y_{it} throughout. The sex dummy is s, the Ph.d. vintage dummy is v.

and compute individual-specific $\hat{\theta}_i = \sigma_u/\sqrt{\sigma_u^2 + T_i\sigma_\alpha^2}$ and transform as in Hausman-Taylor to obtain the GLS-IV estimates

$$\begin{array}{llllll} \hat{y}_{it} & = .247 & + .167e_{it} & -.0041e_{it}^2 & -.058y_{it-1} & +.010s_i & +.75v_i \\ & (.013) & (.08) & (.0002) & (.032) & (.071) & (.04) \end{array}$$

We see that the effect of the reweighting leaves the experience effect essentially unchanged, but it rather drastically alters the intercept and the effect of the vintage effect. The latter effect is particularly disturbing in view of the fact that there is a strong *a priori* presumption that the vintage effect is zero. The induced vintage effect of the reweighting can be attributed to the negative association between the duration of the time-series, T_i, for each individual in the sample and the vintage dummy variable – individuals with post-1960 Ph.d. can't have large T_i. It is also disturbing that the autoregressive effect is essentially zero in this formulation. This may be attributable to the well-known bias of the within estimator in dynamic panel models.

An alternative strategy for estimating models of the general dynamic panel form considered here is developed by Arellano and Bond (1991) in the GMM framework. In an effort to explore this we have estimated the model in first differences

$$\begin{array}{llll} \hat{\Delta y}_{it} & = .470 & + .051\Delta y_{it-1} & -.0039(6e_{it} - 9) \\ & (.034) & (.046) & (.0003) \end{array}$$

using as instruments $(6e_{it} - 9)$ and the first 30 columns of the matrix $Z = (Z_i)_{i=1}^n$ where as in Arellano and Bond, , Z_i is a block diagonal matrix with the row vector $(\Delta y_{i1} \ldots \Delta y_{is})$ corresponding to the $(s+2)$nd observation for individual i. In our expanded sample, since some individuals have as many as $T_i = 13$ observations, the full column dimension of Z is $\max_i(T_i - 2)(T_i - 1)/2 + 1 = 92$. Since a singular value decomposition of Z indicates that only the first 1/3 of the columns of Z are linearly independent, while the remainder are nearly collinear, we have chosen to use only 29 columns corresponding to Δy and the column corresponding to the experience variable.

At first glance it appears that the model in first differences given above is qualitatively very different than the previous ones. This is not really true, however. On closer inspection we see that $\Delta e_t \equiv 3$ due to the time averaging effect and $\Delta e_t^2 = 6e_t - 9$ so we can compute the experience level with maximal productivity by dividing the intercept in the foregoing model by 6 times the "slope" coefficient which yields 20.08 a number quite close to our previous results. Obviously, the dummy effects are not estimable from this differenced version of the model.

Further exploration is obviously warranted before we can comfortably conclude that a satisfactory estimation method has been found for this problem.

References

Divide, I. and U. Koenker (1991) What is Phuzics Really Worth? preprint.

Arellano, M. and Bond S. (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.