DISTINGUISHING HETEROGENEITY AND INEFFICIENCY

IN A PANEL DATA STOCHASTIC FRONTIER MODEL

Christine Amsler
Michigan State University

Peter Schmidt
Michigan State University

March 4, 2015

1. INTRODUCTION

In this paper we consider a panel data model of the form:

(1) $$y_{it} = x_{it}'\beta + z_i'\gamma + v_{it} + a_i + b_i = w_{it}'\delta + v_{it} + c_i \ , \ i = 1, \ldots, n \ , t = 1, \ldots T,$$

where $w_{it} = \begin{bmatrix} x_{it} \\ z_i \end{bmatrix}$, $\delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ and $c_i = a_i + b_i$. We view a_i and b_i as "almost fixed," in the Mundlak (1978), Chamberlain (1980) and Hausman and Taylor (1981) sense that they are random but correlated with some or all of the regressors. We will assume that z_i contains an intercept and so we can take the mean of a_i and b_i to be zero.

The existing panel data literature tells us how to estimate δ and the c_i under various assumptions. However, the aim of this paper is to estimate a_i and b_i separately. More precisely, we are interested in estimating b_i while controlling for unobservable time-invariant variables that are captured by a_i.

The specific context that we have in mind is that equation (1) represents a stochastic frontier production function model, so that y_{it} is the log output of firm *i* at time *t*. The x_{it} and z_i are measures of inputs, or observable variables to control for the production environment. (In many applications the x_{it} will be measures of inputs and there will be no z_i other than intercept, but we will opt for generality at this point.) Differences across firms in the value of b_i reflect differences in the technical efficiency of production, and as in Schmidt and Sickles (1984) a conceptual measure of inefficiency is $u_i^* = max_j \ b_j - \ b_i$. These u_i^* are ≥ 0 and one of them is $= 0$. Differences in the value of a_i , on the other hand, reflect differences in the production environment that are beyond the control of the firm and which we do not wish to include in our efficiency measures. As a specific hypothetical example, suppose that the firms are farms. Then b_i could be a measure of the skill of the farmer, and a_i could represent relevant but unobserved features of the production environment like soil quality or microclimate.

Although we will not pursue this point in the paper, our models may have many potential applications. For example, in a panel data earnings (or log wage) study, one might want to distinguish the effects of innate ability from those of socio-economic background, without necessarily having good measures of either. Or, in a longitudinal epidemiological study, one might wish to distinguish the effects of genetics from those of lifestyle. How useful our models are in answering these kinds of questions is ultimately an empirical question.

Obviously we cannot separate a_i from b_i without further assumptions. Our identification strategy will be to assume that there are some observable variables that are correlated with a_i but not with b_i, and some other variables that are correlated with b_i but not with a_i. Continuing with our agricultural example, we might assume that the education of the farmer is correlated with ability of the farmer but not with soil quality or microclimate, and we might assume that dummy variables for the physical location of the farm are correlated with soil quality or microclimate but not with the ability of the farmer.

The discussion of the previous paragraph is in terms of simple correlations, and it leads to one of the models of the paper. We also consider a second model where the identification strategy is to assume that partial autocorrelations equal zero. In terms of our agricultural example, the first model assumes that ability of the farmer is uncorrelated with physical location of the farm, whereas the second model assumes that, conditional on education of the farmer, ability of the farmer is uncorrelated with physical location of the farm.

The plan of the paper is as follows. Section 2 gives a brief review of the panel data stochastic frontier literature, to motivate the models we consider here. Section 3 lists some assumptions and gives some preliminary results from the existing panel data literature. Section 4 analyzes the model defined in terms of simple correlations. Section 5 analyzes the model defined in terms of partial correlations. Section 6 gives our concluding remarks. Some technical results are given in an Appendix.

2. A BRIEF REVIEW OF PANEL DATA STOCHASTIC FRONTIER MODELS

In this section we will give a brief review of panel data stochastic frontier models, aimed at econometricians who may not be familiar with these models. The point is to provide motivation for the models considered in this paper.

The stochastic frontier model was proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) in a cross-sectional context. The model they proposed was of the form:

(2) $$y_i = \alpha + x_i'\beta + v_i - u_i \ , \ i = 1,\ldots,n \ ,$$

where y_i is the log of output of firm i, x_i contains measures (e.g. logs) of inputs, v_i is zero-mean normal noise, and $u_i \geq 0$ is a measure of technical inefficiency. It is assumed that x, v and u are mutually independent, and u is assumed to have a specific parametric distribution, such as half-normal. This model was generalized to the panel data setting by Pitt and Lee (1981), who considered the model

(3) $$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i \ , \ i = 1,\ldots,n \ , \ t = 1,\ldots,T \ .$$

They made assumptions similar to those given above, and in particular they still assumed that v is normal and u is half-normal. The distinguishing feature of the model is that the technical inefficiency term u_i is time-invariant.

Schmidt and Sickles (1984) were the first to note that, with panel data, time-invariance of u_i can be used to avoid making distributional assumptions for u. They consider the same type of model as Pitt and Lee, as given in (3) above, with v and u viewed as random but without any specific distributional assumptions for v or u. Defining $\alpha_i = \alpha - u_i$, we then have a panel data model with individual effects:

(4) $$y_{it} = \alpha_i + x_{it}'\beta + v_{it} \; , \; i = 1,\ldots,n \; , \; t = 1,\ldots,T \; .$$

Schmidt and Sickles suggested the following estimates: $\hat{\beta}$ = the usual fixed effects (within) estimate; $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$; $\hat{\alpha} = max_i\ (\hat{\alpha}_i)$; $\hat{u}_i = \hat{\alpha} - \hat{\alpha}_i$. Consistency of $\hat{u}_i$ as an estimate of u_i requires both $T \to \infty$ (so that $\hat{\alpha}_i \to_p \alpha_i$) and $n \to \infty$ (so that $max_i(\alpha_i) \to_p \alpha$).

A serious problem with this model is that any unobservables that are time-invariant (or even very persistent) will end up in the inefficiency measure. That is, inefficiency is measured by differences in the α_i, and the differences in the α_i will capture both the technical efficiency of production (e.g. differences across farms in the skill of the farmer) and also pure heterogeneity (e.g. differences across farms in the quality of the soil) because both are likely to be at least approximately time-invariant. This point has been made forcefully by Greene in a number of articles (Greene (2004), Greene (2005a), Greene (2005b)). For example, Greene (2005a, p. 277) notes correctly that "by interpreting the firm specific term as 'inefficiency,' any unmeasured time invariant cross firm heterogeneity must be assumed away."

Greene proposes a "true fixed effects" model that contains an individual effect and an i.i.d. one-sided error:

(5) $$y_{it} = \alpha_i + x_{it}'\beta + v_{it} - u_{it} \; ,$$

where α_i is a fixed effect (parameter), the v_{it} are i.i.d. normal, and the u_{it} are i.i.d. half-normal. He interprets the α_i as measures of heterogeneity and the $u_{it} > 0$ as measures of inefficiency. However, it is arguably true that we now have the opposite problem as in the Schmidt and Sickles model, because now any time invariant (or very persistent) component of inefficiency will tend to end up in the heterogeneity measure and be left out of the inefficiency measure.

Greene also proposes a "true random effects" model in which the α_i are random (specifically, normal) and independent of the regressors and the other error components, but they are still viewed as capturing heterogeneity. This model has been generalized by Kumbhakar, Lien and Hardaker (2014) and Colombi, Kumbhakar, Martini and Vittadini (2014), who include a one-sided time invariant inefficiency term. The model (equation (1) of Colombi et al., and Model 6 of Kumbhakar et al.) is:

(6) $$y_{it} = \alpha_i + x_{it}'\beta \ + v_{it} - u_{it} - \eta_i \ ,$$

where α_i and v_{it} are normal, and u_{it} and η_i are half-normal. All four of these random components are independent of each other and of *x*, and they are i.i.d. over *i* and (where relevant) *t*. A likelihood is derived using results on the closed skew-normal family of distributions.

The interpretations of these components are as follows: u is short-run inefficiency; η is time-invariant (persistent) inefficiency; *v* is idiosyncratic noise; and α is time-invariant heterogeneity. So we distinguish time-invariant heterogeneity (soil quality) from time-invariant inefficiency (skill of the farmer) on the basis of distributional assumptions.

While this approach does successfully distinguish heterogeneity from inefficiency, it does so under very strong assumptions. In particular, the number of distributional assumptions is rather large. In this paper we will take an alternative approach, originally suggested by Chen, Schmidt and Wang (2014), who noted that "an alternative source of identification would be to identify variables that are correlated with inefficiency but not heterogeneity, or vice-versa."

3. PRELIMINARY RESULTS

The model is as given in equation (1) above. We wish to distinguish heterogeneity (*a*) from inefficiency (*b*). We observe the basic data *y*, *x* and *z*. We will also assume that we observe some time-invariant variables q_{1i} that are uncorrelated with *v*, *a*, and *b* (and therefore with *v* and *c*). The variables q_1 may include some or all of the time-invariant regressors *z* and also may include the means of some or all of the *x*'s, as in Hausman and Taylor (1981). Or, as in Amemiya and MaCurdy (1986) or Breusch, Mizon and Schmidt (1989), time specific values of *x* can be used. But q_1 may also include "outside instruments" that are not part of the basic specification.

For the benefit of readers who understand stochastic frontiers models better than the panel data literature, we will first give a brief discussion of the problem of estimating the regression coefficients δ in equation (1). This is essentially the problem of Hausman and Taylor (1981), and our discussion follows Wooldridge (2010, pp. 325-328).

Here and in the rest of the paper we assume random sampling over *i*. We make the following assumptions.

ASSUMPTION 1. [Strict exogeneity of *x* with respect to *v*, conditional on *a* and *b*]

$$\boldsymbol{E(v_{it}|x_i^*, a_i, b_i) = 0} \text{ , where } \boldsymbol{x_i^* = (x_{i1}', \ldots, x_{iT}')'}.$$

ASSUMPTION 2. [Exogeneity of q_1 with respect to *v*, *a* and *b*]

$$\boldsymbol{E(v_{it}|q_{1i}) = E(a_i|q_{1i}) = E(b_i|q_{1i}) = 0}$$

These assumptions imply that v_{it} is uncorrelated with x_i^*, a_i, b_i and q_{1i}, and that q_{1i} is uncorrelated with a_i and b_i.

Under these assumptions, the following moment conditions hold:

(MC1) $E \ \sum_t \tilde{x}_{it}(y_{it} - x_{it}'\beta) = 0$

(MC2) $E \ q_{1i}(\bar{y}_i - \bar{x}_i'\beta - z_i'\gamma) = 0$

where for any variable x_{it} , $\bar{x}_i = \frac{1}{T}\sum_t x_{it}$ and $\tilde{x}_{it} = x_{it} - \bar{x}_i$.

Note that the sum over *t* in (MC1) is necessary to make the deviations from means of *x* orthogonal to $c = a + b$. We are not necessarily assuming that the individual (single value of t) deviations from means ($\tilde{x}_{it}$) are uncorrelated with c_i. That would be a Breusch-Mizon-Schmidt type assumption, and if we made it the individual deviations from means would be part of q_{1i}. Note also that (MC1), which requires only Assumption 1, is sufficient (given some obvious regularity conditions) to identify β; it leads to the so-called "within" estimator. However, to estimate γ we need (MC2), in which the number of exogenous instruments q_1 is at least as large as the number of time invariant variables z plus the number of x's whose means are correlated with c, and where a rank condition given in Appendix 1 holds. The exogeneity of these instruments requires Assumption 2.

Some computational details about the GMM estimates based on (MC1) and (MC2) are given in Appendix 1. For our present purposes will we simply presume that these GMM estimates, which we will call $\hat{\beta}$ and $\hat{\gamma}$, are consistent.

Given estimates of β and γ, we can estimate the individual effects c_i. At this point we need to make a distinction between two different types of asymptotic analysis. Our asymptotics in this paper will always involve $n \to \infty$. However, we will distinguish asymptotic analysis as $n \to \infty$ and $T \to \infty$ (which we will call "large T" asymptotics) from asymptotic analysis as $n \to \infty$ with T fixed (which we will call "fixed T" asymptotics). Many panel data sets for stochastic frontier analysis have n much larger than T, so that the fixed T asymptotics would be the more likely to be relevant.

The usual estimates of the individual effects c_i are given by

(7) $$\hat{c}_i = y_i - \bar{x}_i'\hat{\beta} - z_i'\hat{\gamma}$$

These would be, for example, the coefficients of the individual-specific dummy variables in a fixed-effects regression calculated as OLS with individual dummies ("OLSDV"). A simple calculation shows that $\hat{c}_i = c_i + \bar{v}_i +$ terms (involving estimation error in $\hat{\beta}$ and $\hat{\gamma}$) that are asymptotically (as $n \to \infty$) negligible. Correspondingly var($\hat{c}_i$) $\equiv \sigma_{\hat{c}}^2 \cong \sigma_c^2 + \frac{1}{T}\sigma_v^2$. Note that the difference between $\hat{c}_i$ and c_i is negligible under large *T* asymptotics, but if the difference between $\hat{c}_i$ and c_i cannot be ignored under fixed *T* asymptotics.

Consistent estimation of σ_c^2 $(= \sigma_a^2 + \sigma_b^2)$ and $\sigma_{\hat{c}}^2$ is also a standard topic in the panel data literature. If $\hat{\beta}$ and $\hat{\gamma}$ are any consistent estimates of β and γ, define the within and between sums of squares: $SSE_W = \sum_i \sum_t (\tilde{y}_{it} - \tilde{x}_{it}'\hat{\beta})^2$ and $SSE_B = \sum_i (\bar{y}_i - \bar{x}_i'\hat{\beta} - z_i'\hat{\gamma})^2 = \sum_i \hat{c}_i^2$. Then $\hat{\sigma}_v^2 = \frac{1}{n(T-1)} SSE_W$ is a consistent estimate of σ_v^2, $\hat{\sigma}_{\hat{c}}^2 = \frac{1}{n} SSE_B$ is a consistent estimate of $\sigma_{\hat{c}}^2$, and $\hat{\sigma}_c^2 = \hat{\sigma}_{\hat{c}}^2 - \frac{1}{T}\hat{\sigma}_v^2$ is a consistent estimate of σ_c^2. All of these statements are true in terms of large *T* asymptotics or fixed *T* asymptotics, although the distinction between $\sigma_{\hat{c}}^2$ and σ_c^2 matters only when *T* is fixed.

Although we have an estimate of c_i, namely $\hat{c}_i$, in the fixed *T* case we should be able to do better because the variance of the error in $\hat{c}_i$ is known (equal to $\frac{1}{T}\sigma_v^2$). So we can obtain an estimator with smaller mean square error by using the linear projection of c_i on $\hat{c}_i$:

(8) $$\dot{c}_i = L(c_i|\hat{c}_i) = \frac{cov(c,\hat{c})}{var(\hat{c})}\hat{c}_i = \frac{\sigma_c^2}{\sigma_{\hat{c}}^2}\hat{c}_i \ .$$

(See Appendix 2 for a very brief discussion of linear projections.) As we would expect, this is a shrinkage of $\hat{c}_i$ toward zero.

4. DISTINGUISHING HETEROGENEITY AND INEFFICIENCY – MODEL 1

In the previous section we assumed that we observed variables q_1 that were uncorrelated with *v*, *a* and *b*. In this section we will assume that we also have variables q_2 and q_3 such that:

q_2 is uncorrelated with *v* and *a* but correlated with *b*

q_3 is uncorrelated with *v* and *b* but correlated with *a*

These variables could be inputs or functions of inputs, but mostly we have in mind other variables that would not be in the production function proper, like education of the farmer as part of q_2, or variables indicating the physical location or climate of the farm as part of q_3.

4.1 Using Variables That Are Correlated with Inefficiency but Not with Heterogeneity

For the moment we will focus on the variables q_{2i} that are uncorrelated with v_{it} and a_i but correlated with b_i. We make the following additional assumption (which we will maintain in addition to Assumptions 1 and 2).

ASSUMPTION 3.

$$\boldsymbol{E(q_{2i}v_{it}) = E(q_{2i}a_i) = 0}$$

Note that we do not assume that $E(q_{2i}b_i) = 0$, and indeed we want *b* and q_2 to be correlated. That is what distinguishes q_2 from q_1.

Let $\Sigma_{2b} = C_*(q_{2i}, b_i) = E(q_{2i}b_i)$, where "$C_*$" represents the *uncentered* covariance, which is appropriate since $E(b_i) = 0$. Then we have the following additional moment conditions:

(MC3) $$E\ q_{2i}(\bar{y}_i - \bar{x}_i'\beta - z_i'\gamma) - \Sigma_{2b} = 0\ .$$

Since (MC3) contains the same number of new parameters (Σ_{2b}) as moment conditions, it does not affect the GMM estimates $\hat{\beta}$ and $\hat{\gamma}$ from (MC1) and (MC2). It simply yields the estimate of Σ_{2b}, which is

(9) $$\hat{\Sigma}_{2b} = \frac{1}{n}\sum_i q_{2i}\left(\bar{y}_i - \bar{x}_i'\hat{\beta} - z_i'\hat{\gamma}\right) = \frac{1}{n}\sum_i q_{2i}\,\hat{c}_i\ .$$

Now we need to recover estimates of the b_i. Define the linear projection:

(10) $$b_i = \gamma_b' q_{2i} + \text{error}\ ,\ \gamma_b = \Sigma_{22}^{-1}\Sigma_{2b}\ ,\ \Sigma_{22} = V_*(q_2) = E(q_2 q_2').$$

We observe q_{2i} so we can calculate $\hat{\Sigma}_{22} = \frac{1}{n}\sum_i q_{2i}q_{2i}'$. Also from (9) we have an estimate of

Σ_{2b}, so we can construct an estimate of b_i:

(11) $$\hat{b}_i = \hat{\Sigma}'_{2b}\hat{\Sigma}_{22}^{-1}q_{2i} \ .$$

As in Schmidt and Sickles (1984), our inefficiency measures are *differences* in the b_i. That is, our estimate of inefficiency for firm *i* is $\hat{u}_i = max_j \ \hat{b}_j - \hat{b}_i$. The $\hat{u}_i$ are ≥ 0 and one of them is = 0.

4.2 Using Variables That Are Correlated with Heterogeneity but Not with Inefficiency

We now also will assume that we observe variables q_{3i} that are uncorrelated with v_{it} and b_i but correlated with a_i. This allows us to estimate the a_i, and it also allows us to improve our estimates of the b_i when q_{2i} and q_{3i} are correlated.

For this case we make the following additional assumption (which we will maintain in addition to Assumptions 1, 2 and 3).

ASSUMPTION 4.

$$\boldsymbol{E(q_{3i}v_{it}) = E(q_{3i}b_i) = 0}$$

Note that we do not assume that $E(q_{3i}a_i) = 0$.

Under Assumption 4, the following moment conditions hold:

(MC4) $$E \ q_{3i}(\bar{y}_i - \bar{x}'_i\beta - z'_i\gamma) - \Sigma_{3a} = 0$$

Under Assumptions 1, 2, 3 and 4, the moment conditions (MC1), (MC2), (MC3) and (MC4) hold. As discussed above, (MC1) and (MC2) yield the estimates of β and γ and (MC3) gives us the estimate of Σ_{2a}. Finally (MC4) implies the estimate of Σ_{3a}, which is

(12) $$\hat{\Sigma}_{3a} = \frac{1}{n}\sum_i q_{3i}\left(\bar{y}_i - \bar{x}'_i\hat{\beta} - z'_i\hat{\gamma}\right) = \frac{1}{n}\sum_i q_{3i}\hat{c}_i \ .$$

This would lead to an estimate of a_i that is similar in spirit to the estimate of b_i given in (11) above; namely, $\hat{a}_i = \hat{\Sigma}'_{3a}\hat{\Sigma}_{33}^{-1}q_{3i}$.

An interesting observation is that this added assumption allows for a better estimate of b_i, if q_2 and q_3 are correlated. We consider the linear projection of b on $q^* = \begin{bmatrix} q_2 \\ q_3 \end{bmatrix}$:

(13) $$L(b|q^*) = \Sigma_{bq^*} V_*^{-1}(q^*) q^* = [\Sigma'_{2b}, 0] \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} q_2 \\ q_3 \end{bmatrix}$$

$$= \Sigma'_{2b} \, (\Sigma^{22} q_2 + \Sigma^{23} q_3),$$

where $\Sigma = E(q^* q^{*\prime}) = \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}$ and where Σ^{jk} represents a block of Σ^{-1}. And similarly, although we are not primarily interested in the a_i, we have the result that $L(a|q^*) = \Sigma'_{3a} \, (\Sigma^{32} q_2 + \Sigma^{33} q_3)$. In terms of estimates, (13) leads to

(14) $$\breve{b}_i = \hat{\Sigma}'_{2b} \left(\hat{\Sigma}^{22} q_{2i} + \hat{\Sigma}^{23} q_{3i} \right)$$

and there is the corresponding expression $\breve{a}_i = \hat{\Sigma}'_{3a} \left(\hat{\Sigma}^{32} q_{2i} + \hat{\Sigma}^{33} q_{3i} \right)$.

The estimate in (14) is indeed better than the one in (11). For $\hat{b}$ in (11) we have var($\hat{b}$) = $\Sigma'_{2b} \Sigma_{22}^{-1} \Sigma_{2b}$. For $\breve{b}$ in (14) we have var($\breve{b}$) = $\Sigma'_{2b} \Sigma^{22} \Sigma_{2b}$ where $\Sigma^{22} = (\Sigma_{22} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32})^{-1}$. Here $\Sigma_{22} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32}$ is smaller than Σ_{22}, so its inverse is bigger. Therefore var($\breve{b}$) is bigger than var($\hat{b}$). That is good because it is the explained variation. The unexplained variation (e.g. var($b - \breve{b}$)) is smaller.

4.3 Projections onto $\hat{c}$

The methods of Sections 4.2 and 4.3 yield estimates of a_i and b_i but these do not add up to either $\hat{c}_i$ or $\dot{c}_i$, whereas in some sense we ought to respect this adding up constraint. An obvious thought is to consider the best linear predictors given by the linear projections of a_i and b_i on $\hat{c}_i$. To do so, we make an additional assumption, which we maintain along with Assumptions 1-4.

ASSUMPTION 5.

$$\boldsymbol{\sigma_{ab} = E(a_i b_i) = 0} \, .$$

Then we have the following estimates of a_i and b_i.

(15) $$\dot{a}_i = \frac{cov(a,\hat{c})}{var(\hat{c})}\hat{c}_i = \frac{\sigma_a^2}{\sigma_{\hat{c}}^2}\hat{c}_i \ , \ \dot{b}_i = \frac{cov(b,\hat{c})}{var(\hat{c})}\hat{c}_i = \frac{\sigma_b^2}{\sigma_{\hat{c}}^2}\hat{c}_i \ .$$

These estimates satisfy the adding up constraint that $\dot{a}_i + \dot{b}_i = \dot{c}_i$.

The point of Assumption 5 is to make cov$(a,\hat{c}) = \sigma_a^2$, as opposed to $\sigma_a^2 + \sigma_{ab}$, and similarly for cov$(b,\hat{c})$. This is not of fundamental importance, but we do not wish to have to estimate σ_{ab}.

The results in (15) are not feasible without estimating or specifying σ_a^2 and σ_b^2. We can estimate $\sigma_{\hat{c}}^2$ and σ_c^2 from standard panel data methods, as discussed above, but without further assumptions we cannot estimate σ_a^2 and σ_b^2. We will return to this point in Section 4.4.

In equation (15) we estimated *a* and *b* by projecting them onto $\hat{c}$, whereas in the previous section we estimated *a* and *b* by projecting them onto $q^* = \begin{bmatrix} q_2 \\ q_3 \end{bmatrix}$. We ought to be able to improve on either of these estimates if we project *a* or *b* onto $q^{**} = \begin{bmatrix} \hat{c} \\ q_2 \\ q_3 \end{bmatrix}$. So now we will have

(16A) $$L(b|q^{**}) = \Sigma_{bq^{**}}V_*^{-1}(q^{**})q^{**} = [\sigma_b^2, \Sigma_{2b}' \, , 0]V^{-1}q^{**}$$

where $V = \begin{bmatrix} \sigma_{\hat{c}}^2 & \Sigma_{2b}' & \Sigma_{3a}' \\ \Sigma_{2b} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{3a} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}$ is shorthand for $V_*(q^{**})$. And similarly

(16B) $$L(a|q^{**}) = \Sigma_{aq^{**}}V^{-1}q^{**} = [\sigma_a^2, 0, \Sigma_{3a}']V^{-1}q^{**}.$$

(16C) $$L(c|q^{**}) = \Sigma_{cq^{**}}V^{-1}q^{**} = [\sigma_c^2, \Sigma_{2b}', \Sigma_{3a}']V^{-1}q^{**}$$

Let $\vec{b}_i, \vec{a}_i$ and $\vec{c}_i$ be the estimates that we get if we evaluate (16A), (16B) and (16C) using consistent estimates of $\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_{\hat{c}}^2, \Sigma_{2b}, \Sigma_{3a}$ and V. Then these are better (smaller mean squared error) estimates than $\dot{b}_i$, $\dot{a}_i$ and $\dot{c}_i$. Furthermore, they satisfy the adding up constraint that $\vec{a}_i + \vec{b}_i = \vec{c}_i$.

We have $\vec{c}_i = \hat{c}_i$ in the large T case though not in the fixed T case. In the large T case, $\sigma_c^2 = \sigma_{\hat{c}}^2$ and therefore

(17) $$L(c|q^{**}) = \left[\sigma_{\hat{c}}^2, \Sigma_{2b}', \Sigma_{3a}'\right] V^{-1} q^{**} = [1\ ,\ 0\ ,\ 0] q^{**} = \hat{c}\ ,$$

where the term [1 , 0 , 0] arises because $\left[\sigma_{\hat{c}}^2, \Sigma_{2b}', \Sigma_{3a}'\right]$ is the first row of V.

4.4 Estimation of σ_a^2 and σ_b^2

As mentioned above, consistent estimation of σ_c^2 and $\sigma_{\hat{c}}^2$ is a solved problem. As a result the estimator $\dot{c}_i$ given in equation (8) is feasible, and so is $\vec{c}_i$ in (16C).

Without further assumptions it does not seem possible to obtain consistent estimates of σ_a^2 and σ_b^2 (separately). However, we can provide some bounds, as follows. We know that $V_* \begin{pmatrix} b \\ q_2 \end{pmatrix} = \begin{pmatrix} \sigma_b^2 & \Sigma_{b2} \\ \Sigma_{2b} & \Sigma_{22} \end{pmatrix}$ must be positive semi-definite, so its determinant must be greater than or equal to zero. Using a result from Searle (1982, p. 258), $\begin{vmatrix} \sigma_b^2 & \Sigma_{2b}' \\ \Sigma_{2b} & \Sigma_{22} \end{vmatrix} = |\Sigma_{22}| \cdot (\sigma_b^2 - \Sigma_{2b}' \Sigma_{22}^{-1} \Sigma_{2b})$. Therefore $\sigma_b^2 - \Sigma_{2b}' \Sigma_{22}^{-1} \Sigma_{2b} \geq 0$. Similarly $\sigma_a^2 - \Sigma_{3a}' \Sigma_{33}^{-1} \Sigma_{3a} \geq 0$. Therefore $\Sigma_{2b}' \Sigma_{22}^{-1} \Sigma_{2b} \leq \sigma_b^2$ and $\Sigma_{3a}' \Sigma_{33}^{-1} \Sigma_{3a} \leq \sigma_a^2$.

However, this is not the tightest bound that we can obtain. We also know that $V_* \begin{pmatrix} b \\ q^* \end{pmatrix} = \begin{pmatrix} \sigma_b^2 & \Sigma_b' \\ \Sigma_b & \Sigma \end{pmatrix}$ must be positive semi-definite, where $\Sigma_b = \begin{bmatrix} \Sigma_{2b} \\ 0 \end{bmatrix}$. This implies that $\sigma_b^2 - \Sigma_b' \Sigma^{-1} \Sigma_b \geq 0$, or $\Sigma_{2b}' \Sigma^{22} \Sigma_{2b} \leq \sigma_b^2$. This is a tighter bound than the one in the previous paragraph, because $\Sigma^{22} \geq \Sigma_{22}^{-1}$. (See the discussion at the end of Section 4.2.) By the same argument, we also have the similar result that $\Sigma_{3a}' \Sigma^{33} \Sigma_{3a} \leq \sigma_a^2$

We also know that $\sigma_b^2 = \sigma_c^2 - \sigma_a^2$ so that $\sigma_b^2 \leq \sigma_c^2 - \Sigma_{3a}' \Sigma^{33} \Sigma_{3a}$ and similarly $\sigma_a^2 \leq \sigma_c^2 - \Sigma_{2b}' \Sigma^{22} \Sigma_{2b}$. Therefore

(18) $$\Sigma_{2b}' \Sigma^{22} \Sigma_{2b} \leq \sigma_b^2 \leq \sigma_c^2 - \Sigma_{3a}' \Sigma^{33} \Sigma_{3a}$$

and $\Sigma_{3a}'\Sigma^{33}\Sigma_{3a} \leq \sigma_a^2 \leq \sigma_c^2 - \Sigma_{2b}'\Sigma^{22}\Sigma_{2b}$.

We will want to pick values of $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ that are in the allowable ranges given in (18), and that satisfy $\hat{\sigma}_a^2 + \hat{\sigma}_b^2 = \hat{\sigma}_c^2$. (This equality is necessary for the adding up constraint $\vec{a}_i + \vec{b}_i = \vec{c}_i$ to hold.) Obviously there is more than one such set of values, and subjectively choosing among them is not necessarily an attractive notion.

We now turn to the issue of finding further assumptions such that we can estimate σ_a^2 and σ_b^2 consistently. To do so we will assume parametric models for $E(b^2|q_2)$ and for $E(a^2|q_3)$. We will therefore make the following assumption, which we will maintain in addition to Assumptions 1-2. (There is no need for Assumptions 3-5 because they are implied by Assumption 6.)

ASSUMPTION 6.

$\boldsymbol{v}$ **is independent of** $\boldsymbol{a, b, q_2}$ **and** $\boldsymbol{q_3}$

$\boldsymbol{E(a|\, b, q_2) = 0}$ **and** $\boldsymbol{E(a^2|\, q_2) = \sigma_a^2}$

$\boldsymbol{E(b|\, a, q_3) = 0}$ **and** $\boldsymbol{E(b^2|\, q_3) = \sigma_b^2}$

$\boldsymbol{E(ab|q_2, q_3) = 0}$

$\boldsymbol{E(a^2|q_3) = \mu_a^{(2)} \cdot exp[\lambda_3'(q_3 - \overline{q}_3)]}$ **with** $\boldsymbol{\lambda_3 \neq 0}$

$\boldsymbol{E(b^2|q_2) = \mu_b^{(2)} \cdot exp[\lambda_2'(q_2 - \overline{q}_2)]}$ **with** $\boldsymbol{\lambda_2 \neq 0}$

These assumptions are much stronger than Assumptions 3, 4 and 5, because they make statements about independence and conditional expectations, not just correlations, and because they assume parametric forms for some of the conditional expectations, and because they assume that $E(b^2|q_2)$ and $E(a^2|q_3)$ are not constant (there is conditional heteroskedasticity). The independence assumption for v is stronger than needed but is made to simplify the conditioning assumptions and arguments that follow. The specific functional forms given in the last two lines of Assumption 6 are obviously not the only ones that could have been chosen, but they suffice to show that with suitable parametric assumptions we can estimate σ_a^2 and σ_b^2.

We can note that $\mu_a^{(2)}$ is the expected value of $E(a^2)$ conditional on $q_3 = \bar{q}_3$; loosely, conditional on $q_3 = E(q_3)$. This is almost but not quite the same as σ_a^2, which is the unconditional expectation of a^2, and which (by the law of iterated expectations) equals the expectation (over the distribution of q_3) of $\mu_a^{(2)} \cdot exp[\lambda_3'(q_3 - \bar{q}_3)]$. This is different from $\mu_a^{(2)}$ because $E(exp[\lambda_3'(q_3 - \bar{q}_3)]) \neq 1$ due to the nonlinearity of the exponential function. Similar statements apply to $\mu_b^{(2)}$ and σ_b^2.

Clearly $E(c_i^2 | q_{2i}) = E(a_i^2 + b_i^2 + 2a_i b_i)|q_{2i}$ and so, under Assumption 6,

(19A) $$E(c_i^2 | q_{2i}) = \sigma_a^2 + \mu_b^{(2)} \cdot exp[\lambda_2'(q_{2i} - \bar{q}_2)]$$

and similarly

(19B) $$E(c_i^2 | q_{3i}) = \sigma_b^2 + \mu_a^{(2)} \cdot exp[\lambda_3'(q_{3i} - \bar{q}_3)]$$

Now consider nonlinear least squares based on (19A), with our dependent variable equal to $\hat{c}_i$. That is, we minimize the sum of squares

(20A) $$SSE_2 = \sum_i \left(\hat{c}_i^2 - \sigma_a^2 - \mu_b^{(2)} \cdot exp[\lambda_2'(q_{2i} - \bar{q}_2)] \right)^2$$

with respect to $\sigma_a^2, \mu_b^{(2)}$ and λ_2. This yields an estimate of σ_a^2, which we will call $\hat{\sigma}_a^2$, plus estimates of $\mu_b^{(2)}$ and λ_2 that we will ignore. Similarly, we can consider nonlinear least squares based on (19B), in which case we minimimize the sum of squares

(20B) $$SSE_3 = \sum_i \left(\hat{c}_i^2 - \sigma_b^2 - \mu_a^{(2)} \cdot exp[\lambda_3'(q_{3i} - \bar{q}_3)] \right)^2$$

to get an estimate $\hat{\sigma}_b^2$, plus estimates of $\mu_a^{(2)}$ and λ_2 that we will ignore.

Under reasonable regularity conditions, the estimates $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ are consistent under "large T" asymptotics. However, in the "fixed T" case, the difference between $\hat{c}_i$ and c_i, which is $\bar{v}_i$ plus asymptotically negligible terms, contributes an extra additive term $\frac{1}{T}\sigma_v^2$ to $E(\hat{c}_i^2|q_{2i})$ and $E(\hat{c}_i^2|q_{3i})$. We need to correct for this term, which leads us to the estimates

(21) $$\vec{\sigma}_a^2 = \hat{\sigma}_a^2 - \frac{1}{T}\hat{\sigma}_v^2 \quad \text{and} \quad \vec{\sigma}_b^2 = \hat{\sigma}_b^2 - \frac{1}{T}\hat{\sigma}_v^2 \ .$$

These estimates are consistent in the "large T" case and also in the "fixed T" case.

These estimates are presumably not efficient because they ignore the information about σ_b^2 in (20A) and the information about σ_a^2 in (20B). However, to obtain the projections that we want (involving $\hat{c}_i$) these are nuisance parameters and a consistent estimate is all that we really need.

5. DISTINGUISHING HETEROGENEITY AND INEFFICIENCY – MODEL 2

We now will consider Model 2, in which the identifying assumption for b is that the partial correlation of q_2 and a, given q_3, equals zero. We will maintain Assumptions 1, 2 and 5 as above, but we will replace Assumptions 3 and 4 with the following assumption.

ASSUMPTION 7.

$$\boldsymbol{E(q_{2i}v_{it}) = 0} \ \textbf{ and } \ \boldsymbol{E(q_{3i}v_{it}) = 0}$$

$\boldsymbol{L(b|q_2, q_3)}$ **does not depend on** $\boldsymbol{q_3}$

$\boldsymbol{L(a|q_2, q_3)}$ **does not depend on** $\boldsymbol{q_2}$

Define $\Sigma_{2b} = E(q_2 b)$, $\Sigma_{3b} = E(q_3 b)$, $\Sigma_{2a} = E(q_2 a)$, $\Sigma_{3a} = E(q_3 a)$. Then the identifying assumption for Model 1 was that $\Sigma_{3b} = 0$ and $\Sigma_{2a} = 0$. Assumption 6 is different. Assumption 6 does not imply that $\Sigma_{3b} = 0$ and $\Sigma_{2a} = 0$; nor do $\Sigma_{3b} = 0$ and $\Sigma_{2a} = 0$ imply Assumption 6. This reflects the difference between simple and partial correlations.

To think about a case in which Model 2 is appropriate and Model 1 is not, let us return for a moment to our hypothetical agricultural example in which *b* is ability of the farmer, *a* is soil quality, q_2 is education of the farmer, and q_3 is physical location of the farm. Suppose there are two locations, and location A has better soil than location B, and also a more conveniently located school, so that education will tend to be higher in location A than in location B. Suppose people are randomly assigned to the two locations. Now suppose that education raises the ability of the farmer. So now ability will be correlated with location ($\Sigma_{3b} \neq 0$), but conditional on education, ability will not be correlated with location. Also soil quality will be correlated with education ($\Sigma_{2a} \neq 0$), but conditional on location it will not be correlated with education. So Model 2 applies.

Conversely, to think about a case in which Model 1 is appropriate, suppose that the schools are equally convenient in the two locations, and that education does not increase the ability of the farmer, but more able people like school more and so they get more education. Then ability is correlated with education and not with location, and soil quality is correlated with location and not ability. So Model 1 applies.

5.1 Projections onto q_2 and q_3

As in Section 4.2, let $\Sigma = \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}$ and let Σ^{jk} represents a block of Σ^{-1}. Then the linear projection of *b* on $q^* = \begin{bmatrix} q_2 \\ q_3 \end{bmatrix}$ is:

(22) $$L(b|q^*) = \Sigma_{bq^*}\Sigma^{-1}q^* = [\Sigma'_{2b}\,, \Sigma'_{3b}] \begin{bmatrix} \Sigma_{22} & \Sigma_{23} \\ \Sigma_{32} & \Sigma_{33} \end{bmatrix}^{-1} \begin{bmatrix} q_2 \\ q_3 \end{bmatrix}$$

$$= [\Sigma'_{2b}\Sigma^{22} + \Sigma'_{3b}\Sigma^{32}]q_2 + [\Sigma'_{2b}\Sigma^{23} + \Sigma'_{3b}\Sigma^{33}]q_3$$

Assumption 7 requires that $L(b|q^*)$ should not depend on q_3, so we must have $\Sigma'_{2b}\Sigma^{23} + \Sigma'_{3b}\Sigma^{33}$ = 0, or

(23) $$\Sigma'_{3b} = -\Sigma'_{2b}\Sigma^{23}(\Sigma^{33})^{-1}.$$

(Note that, as discussed above, $\Sigma_{3b} \neq 0$ unless q_2 and q_3 are uncorrelated.) Substituting this for Σ'_{3b} in equation (22), we obtain

(24) $$L(b|q^*) = \Sigma'_{2b}[\Sigma^{22} - \Sigma^{23}(\Sigma^{33})^{-1}\Sigma^{32}]q_2 = \Sigma'_{2b}\Sigma_{22}^{-1}q_2 \ ,$$

using a standard result on partitioned inversion.

The same logical argument establishes that $\Sigma'_{2a} = -\Sigma'_{3a}\Sigma^{32}(\Sigma^{22})^{-1}$ and that $L(a|q^*) = \Sigma'_{3a}\Sigma_{33}^{-1}q_3$.

Let $\xi_2 = \Sigma_{22}^{-1}\Sigma_{2b}$ and $\xi_3 = \Sigma_{33}^{-1}\Sigma_{3a}$ so that $L(b|q^*) = \xi'_2 q_2$ and $(a|q^*) = \xi'_3 q_3$. Then

(25) $$L(c|q^*) = L(b|q^*) + L(a|q^*) = L(b|q_2) + L(a|q_3) = \xi'_2 q_2 + \xi'_3 q_3 \ .$$

We can obtain consistent estimates of ξ_2 and ξ_3, say $\hat{\xi}_2$ and $\hat{\xi}_3$, by OLS of $\hat{c}$ on q^* (i.e. on q_2 and q_3). This leads us to our estimates of b_i and a_i:

(26) $$\hat{b}_i = \hat{\xi}'_2 q_{2i} \ , \ \hat{a}_i = \hat{\xi}'_3 q_{3i} \ .$$

Although we do not need them to calculate the estimates in (23), we can also construct estimates of the covariances $\Sigma_{2b}, \Sigma_{3b,} \Sigma_{2a}$ and Σ_{3a}. Specifically, $\hat{\Sigma}_{2b} = \hat{\Sigma}_{22}\hat{\xi}_2$, $\hat{\Sigma}_{3a} = \hat{\Sigma}_{33}\hat{\xi}_3$, $\hat{\Sigma}_{3b} = (\hat{\Sigma}^{33})^{-1}\hat{\Sigma}^{32}\hat{\Sigma}_{2b}$ and $\hat{\Sigma}_{2a} = (\hat{\Sigma}^{22})^{-1}\hat{\Sigma}^{23}\hat{\Sigma}_{3a}$.

5.2 Projections Involving $\hat{c}$

As in Section 4.4, we can also consider estimates that use the value of $\hat{c}$. So once again we define $q^{**} = \begin{bmatrix} \hat{c} \\ q_2 \\ q_3 \end{bmatrix}$ and $V = V_*(q^{**}) = \begin{bmatrix} \sigma_{\hat{c}}^2 & \Sigma'_{2b} & \Sigma'_{3a} \\ \Sigma_{2b} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{3a} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}$. We can calculate an improved estimate of *c*:

(27) $$\vec{c}_i = L(c|q_i^{**}) = \Sigma_{cq^{**}}V^{-1}q_i^{**} = [\sigma_c^2, \Sigma'_{2a} + \Sigma'_{2b}, \Sigma'_{3a} + \Sigma'_{3b}]V^{-1}q_i^{**} \ ,$$

evaluated at the estimated values of $\Sigma_{cq^{**}}$ and V. This estimate is feasible because we can estimate all of the needed variances and covariances without further assumptions.

If we have estimated or specified values of σ_a^2 and σ_b^2, then we can also calculate

(28A) $\vec{b}_i = L(b|q_i^{**}) = \Sigma_{bq^{**}} V^{-1} q_i^{**} = [\sigma_b^2, \Sigma_{2b}', \Sigma_{3b}'] V^{-1} q_i^{**}$

(28B) $\vec{a}_i = L(a|q_i^{**}) = \Sigma_{aq^{**}} V^{-1} q_i^{**} = [\sigma_a^2, \Sigma_{2a}', \Sigma_{3a}'] V^{-1} q_i^{**}$,

where these would be evaluated at the estimated values of $\Sigma_{aq^{**}}$, $\Sigma_{bq^{**}}$ and V. These should be better estimates than $\hat{b}_i$ and $\hat{a}_i$ because we are projecting onto more explanatory variables. They also have the desirable property that $\vec{a}_i + \vec{b}_i = \vec{c}_i$.

There remains the issue of obtaining estimates of σ_a^2 and σ_b^2. As in Section 4.4, we do not see how to do this without further assumptions, but we can provide bounds similar to those in equation (18) above. It is still the case that $V_* \begin{pmatrix} b \\ q_2 \end{pmatrix} = \begin{pmatrix} \sigma_b^2 & \Sigma_{b2} \\ \Sigma_{2b} & \Sigma_{22} \end{pmatrix}$ must be positive semi-definite, which implies that $\sigma_b^2 - \Sigma_{2b}' \Sigma_{22}^{-1} \Sigma_{2b} \geq 0$ and therefore $\sigma_b^2 \geq \Sigma_{2b}' \Sigma_{22}^{-1} \Sigma_{2b}$. As before, however, we can obtain a tighter bound. It is the case that $V_* \begin{pmatrix} b \\ q^* \end{pmatrix}$ must be positive semi-definite, and we can write $V_* \begin{pmatrix} b \\ q^* \end{pmatrix} = \begin{bmatrix} \sigma_b^2 & \Sigma_b' \\ \Sigma_b & \Sigma \end{bmatrix}$ where $\Sigma_b = \begin{bmatrix} \Sigma_{2b} \\ \Sigma_{3b} \end{bmatrix}$. This is the same as in Section 4.4 except that now $\Sigma_{3b} \neq 0$. So the fact that this matrix is positive semi-definite implies that

(29) $\sigma_b^2 \geq \Sigma_b' \Sigma^{-1} \Sigma_b$.

Since $\Sigma_b' \Sigma^{-1} \Sigma_b \geq \Sigma_{2b}' \Sigma_{22}^{-1} \Sigma_{2b}$, this is a tighter bound than above. And, similarly, $\sigma_a^2 \geq \Sigma_a' \Sigma^{-1} \Sigma_a$ where $\Sigma_a = \begin{bmatrix} \Sigma_{2a} \\ \Sigma_{3a} \end{bmatrix}$.

The same logical argument as in Section 4.4 leads to the following bounds (the analogue of equation (18) above):

(30) $\Sigma_b' \Sigma^{-1} \Sigma_b \leq \sigma_b^2 \leq \sigma_c^2 - \Sigma_a' \Sigma^{-1} \Sigma_a$

and $\Sigma_a' \Sigma^{-1} \Sigma_a \leq \sigma_a^2 \leq \sigma_c^2 - \Sigma_b' \Sigma^{-1} \Sigma_b$.

We will want to pick values of $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ that are in the allowable ranges given in (27), and that satisfy $\hat{\sigma}_a^2 + \hat{\sigma}_b^2 = \hat{\sigma}_c^2$ so that $\vec{a}_i + \vec{b}_i = \vec{c}_i$.

As we did for Model 1 in Section 4.4, we now turn to the issue of finding further assumptions such that we can estimate σ_a^2 and σ_b^2 consistently. In the present case this will require parametric models for $E(a^2|q_2, q_3)$ and $E(b^2|q_2, q_3)$. We make the following assumption, which we will maintain in addition to Assumptions 1 and 2. (It implies Assumption 7.)

ASSUMPTION 8.

$\boldsymbol{v}$ **is independent of** $\boldsymbol{a, b, q_2}$ **and** $\boldsymbol{q_3}$**.**

$$\boldsymbol{E(a|q_2, q_3, b) = \xi_3' q_3}$$

$$\boldsymbol{E(b|q_2, q_3, a) = \xi_2' q_2}$$

$$\boldsymbol{E(ab|q_2, q_3) = 0}$$

$$\boldsymbol{E(a^2|q_2, q_3) = \mu_a^{(2)} \cdot exp[\lambda_3'(q_3 - \bar{q}_3)]} \textbf{ with } \boldsymbol{\lambda_3 \neq 0}$$

$$\boldsymbol{E(b^2|q_2, q_3) = \mu_b^{(2)} \cdot exp[\lambda_2'(q_2 - \bar{q}_2)]} \textbf{ with } \boldsymbol{\lambda_2 \neq 0}$$

Since $E(c_i^2|q_{2i}, q_{3i}) = E[(a_i^2 + b_i^2 + 2a_i b_i)|q_{2i}, q_{3i}]$, under Assumption 8,

(31) $$E(c_i^2|q_{2i}, q_{3i}) = \mu_b^{(2)} \cdot exp[\lambda_2'(q_{2i} - \bar{q}_2)] + \mu_a^{(2)} \cdot exp[\lambda_3'(q_{3i} - \bar{q}_3)]$$

This leads naturally to a nonlinear least squares estimator in which we minimize

(32) $$SSE = \sum_i \left(\hat{c}_i^2 - \mu_b^{(2)} \cdot exp[\lambda_2'(q_{2i} - \bar{q}_2)] - \mu_a^{(2)} \cdot exp[\lambda_3'(q_{3i} - \bar{q}_3)]\right)^2$$

In the "large T" case this should yield consistent estimates of the parameters $\mu_a^{(2)}, \mu_b^{(2)}, \lambda_2$ and λ_3. In the "fixed T" case we need to incorporate the variance $\frac{1}{T}\sigma_v^2$ due to the term $\bar{v}_i$, which is not part of c_i but is part of $\hat{c}_i$. That is, the nonlinear least squares estimator would now minimize

(33) $$SSE = \sum_i \left(\hat{c}_i^2 - \frac{1}{T}\hat{\sigma}_v^2 - \mu_b^{(2)} \cdot exp[\lambda_2'(q_{2i} - \bar{q}_2)] - \mu_a^{(2)} \cdot exp[\lambda_3'(q_{3i} - \bar{q}_3)]\right)^2$$

where $\hat{\sigma}_v^2$ is a consistent estimate of σ_v^2 (e.g. based on the within estimate of the basic model, as discussed in Section 3 above).

As in Section 4.4, it is not the case that $\mu_b^{(2)} = \sigma_b^2$. However, we can construct a consistent estimate of σ_b^2 as

(34) $$\hat{\sigma}_b^2 = \frac{1}{n}\sum_i \hat{\mu}_b^{(2)}\ exp\left[\hat{\lambda}_2'(q_{2i} - \bar{q}_2)\right] = \hat{\mu}_b^{(2)} \frac{1}{n}\sum_i exp\left[\hat{\lambda}_2'(q_{2i} - \bar{q}_2)\right]$$

This calculation is a sample equivalent of the law of iterated expectations. A similar result holds for $\hat{\sigma}_a^2$.

6. CONCLUSIONS

In this paper, we have proposed methods for distinguishing two kinds of individual effects ("heterogeneity" and "inefficiency") in a panel data regression model, without making strong distributional assumptions. In one model, we do so by assuming that we observe some variables that are correlated with heterogeneity but not inefficiency, and some other variables that are correlated with inefficiency but not heterogeneity. In a second model, we assume instead that the joint linear projection of inefficiency on the two sets of observable variables depends on only one set and not on the other, and vice versa for heterogeneity. This corresponds to setting partial correlations, opposed to simple correlations, equal to zero.

As discussed in Section 2 of the paper, other papers have separated heterogeneity from inefficiency based on distributional assumptions (e.g. heterogeneity is normal and inefficiency is half normal). Like the assumptions of this paper, these are strong assumptions. An obvious question for further research is to ask how to test either or both sets of assumptions.

APPENDIX 1 – The Hausman and Taylor Estimator

We will first define some notation. The model is $y_{it} = w_{it}'\delta + \varepsilon_{it}$ where $\varepsilon_{it} = v_{it} + c_i$. We write the model for all *T* observations for firm *i* as $y_i = W_i\delta + \varepsilon_i$ and for all *NT* observations we write $y = W\delta + \varepsilon$. Similarly we have matrices of deviations from means $\tilde{X}_i$ and $\tilde{X}$. Define $Q_{1i} = e_T \otimes q_{1i}'$, where e_T is a vector of ones, and then Q_1 for all *NT* observations. Finally, we define the instruments $H_i = [\tilde{X}_i, Q_{1i}]$ for *T* observations and *H* for all *NT* observations.

Now we rewrite the moment conditions (MC1) and (MC2) as

(A1) $$E\ H_i'(y_i - W_i\delta) = 0$$

These moment conditions hold under Assumptions 1 and 2. They identify δ if there are enough of them (in obvious notation, $k_H \geq k_W$) and if the usual rank condition holds ($E\ H_i'W_i$ has full column rank).

GMM of δ can be based on (A1). Let $\Omega = E\ H_i'\varepsilon_i\varepsilon_i'H_i$ and $\Omega_* = I_N \otimes \Omega$, and correspondingly their estimates are $\hat{\Omega}$ and $\hat{\Omega}_*$ where $\hat{\Omega} = \frac{1}{n}\sum_i H_i'\hat{\varepsilon}_i\hat{\varepsilon}_i'H_i$ and where $\hat{\varepsilon}_i = y_i - W_i\hat{\delta}$. Here $\hat{\delta}$ can be a preliminary estimate, like 2SLS, or it can be part of a continuous updating GMM procedure. Then the GMM estimate based on (A1) is:

(A2) $$\tilde{\delta} = (W'H\hat{\Omega}_*^{-1}H'W)^{-1}\ W'H\hat{\Omega}_*^{-1}H'y$$

(Continuous updating means that the initial estimate of δ leads to an estimate of Ω, which leads via (A2) to a new estimate of δ, which leads to a new estimate of Ω, etc.)

The above procedure does not put any restrictions on the weighting matrix Ω. We can impose restrictions on Ω under further assumptions. The next assumption is common in the panel data literature (and was made by Hausman and Taylor).

ASSUMPTION NCH [No conditional heteroskedasticity]

$\text{Var}(v_{it}|H_i) = \sigma_v^2$ for all t

$\text{Var}(c_i|H_i) = \sigma_c^2$

$\text{Cov}(v_{it}, c_i|H_i) = 0$ for all t

Under NCH, it is well known that $\Omega = \sigma_v^2 I_T + T\sigma_c^2 E_T$ where E_T is the $T \times T$ idempotent matrix with each element equal to $1/T$. Hausman and Taylor (1981) show how to estimate σ_v^2 and σ_c^2, so it is easy to get an estimate of Ω and therefore of Ω_*. We can use that estimate in (A2). There is no asymptotic advantage to doing so, as compared to using the unrestricted weighting matrix, but the resulting estimate of δ is probably more numerically stable and has better finite sample properties than when the unrestricted weighting matrix is used.

However, as noted by Hausman and Taylor, under NCH we can do better. It is well known that (up to proportionality) $\Omega^{-1/2} = I_T - (1-\theta)E_T$, where $\theta = \sqrt{\sigma_v^2/(\sigma_v^2 + T\sigma_c^2)}$. Now we can transform the regression equation to "whiten" the errors:

(A3) $$\Omega^{-1/2} y_i = \Omega^{-1/2} W_i \delta + \Omega^{-1/2} \varepsilon_i.$$

This amounts to "1-θ" differences, e.g. the t^{th} element of $\Sigma^{-1/2} y_i$ is $y_{it} - (1-\theta)\bar{y}_i$. Then we can estimate (A3) by standard IV with instruments H_i. In matrix terms for all *NT* observations we have

(A4) $$\tilde{\delta} = [W'\hat{\Omega}_*^{-1/2} H (H'\hat{\Omega}_*^{-1} H)^{-1}\, \hat{\Omega}_*^{-1/2} W]^{-1} W'\hat{\Omega}_*^{-1/2} H (H'\hat{\Omega}_*^{-1} H)^{-1}\, \hat{\Omega}_*^{-1/2} y\ .$$

This is the Hausman and Taylor "efficient" estimator.

APPENDIX 2 – Facts about Linear Projections

Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and let *y* be scalar. Then $L(y|x) = \Sigma_{yx}\Sigma_{xx}^{-1} x$ is the linear projection of *y* on *x*. It has the property that $y - L(y|x)$ is uncorrelated with *x*. Also $\text{var}(L(y|x)) = \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$

and var$(y - L(y|x)) = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$.

We can compare this to what happens if you only use x_1. Then $L(y|x_1) = \Sigma_{y1}\Sigma_{11}^{-1}x_1$. Also var$(L(y|x_1)) = \Sigma_{y1}\Sigma_{11}^{-1}\Sigma_{1y} \leq \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$, and var$(y - L(y|x_1)) = \Sigma_{yy} - \Sigma_{y1}\Sigma_{11}^{-1}\Sigma_{1y}$ $\geq \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$. With the larger set of explanatory variables *x*, the explained variance is larger and the unexplained variance is smaller.

REFERENCES

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics,* 6, 21-37.

Amemiya, T. and T.E. MaCurdy (1986), "Instrumental Variable Estimation of an Error Component Model," *Econometrica,* 54, 869-881.

Breusch, T.S., G.E. Mizon and P. Schmidt (1989), "Efficient Estimation Using Panel Data," *Econometrica,* 57, 695-700.

Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies,* 47, 225-238.

Chen, Y.Y., P. Schmidt and H.J. Wang (2014), "Consistent Estimation of the Fixed-Effects Stochastic Frontier Model," *Journal of Econometrics,* 181, 65-76.

Colombi, R., S.C. Kumbhakar, G. Martini and G. Vittadini (2014), "Closed-Skew Normality in Stochastic Frontiers with Individual Effects and Long/Short-Run Inefficiency," *Journal of Productivity Analysis*, 42, 123-136.

Greene, W.H. (2004), "Distinguishing between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems," *Health Economics,* 13, 959-980.

Greene, W.H. (2005a), "Reconsidering Heterogeneity in Panel Data Estimates of the Stochastic Frontier Model," *Journal of Econometrics,* 126, 269-303.

Greene, W.H. (2005b), "Fixed and Random Effects in Stochastic Frontier Models," *Journal of Productivity Analysis,* 23, 7-32.

Hausman, J.A. and W.E. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica,* 49, 1377-1399.

Kumbhakar, S.C., G. Lien and J.B. Hardaker (2014), "Technical Efficiency in Competing Panel Data Models: a Study of Norwegian Grain Farming," *Journal of Productivity Analysis*, 41, 321-337.

Meeusen, W. and J. van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review,* 18, 435-444.

Mundlak, Y. (1978), "On the Pooling of Time-Series and Cross-Section Data," *Econometrica,* 46, 69-86.

Pitt, M. M. and L.F. Lee (1981), "The Measurement and Sources of Inefficiency in the

Indonesian Weaving Industry," *Journal of Development Economics*, 9, 43-64.

Schmidt, P. and R.C. Sickles (1984), "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, 2, 367-374.

Searle, S.R. (1982), *Matrix Algebra Useful for Statistics,* Wiley.

Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data,* 2nd edition, MIT Press.