DISTINGUISHING HETEROGENEITY AND INEFFICIENCY

IN A PANEL DATA STOCHASTIC FRONTIER MODEL

Christine Amsler
Michigan State University

Peter Schmidt
Michigan State University

August 29, 2015
1. INTRODUCTION

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

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

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 \(\delta\) 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 \(a_i\) while controlling for unobservable time-invariant variables that are captured by \(b_i\).

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 \(a_i\) reflect differences in the technical efficiency of production, and as in Schmidt and Sickles (1984) a conceptual measure of inefficiency is \(u_i^* = \max_j a_j - a_i\). These \(u_i^*\) are \(\geq 0\) and one of them is \(= 0\). Differences in the value of \(b_i\), 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 \(a_i\) could be a measure of the skill of the farmer, and \(b_i\) could represent relevant but unobserved features of the production environment like soil quality or microclimate.
The models we develop may have many different applications besides the production function setting. For example, we could be estimating an earnings function in which $a_i$ could represent innate ability and $b_i$ could represent socioeconomic background. Or we could be estimating a model of lifetime (or occurrence of a particular disease, etc.) in which $a_i$ could represent lifestyle and $b_i$ could represent genetics. It would not be hard to think of many similar examples.

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:

\[ y_i = \alpha + x'_i \beta + v_i - u_i , \quad i = 1, \ldots, n , \]

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

\[ y_{it} = \alpha + x'_{it} \beta + v_{it} - u_i , \quad i = 1, \ldots, n , \quad t = 1, \ldots, T . \]

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:

\[
y_{it} = \alpha_i + x_{it}'\beta + v_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T.
\]

Schmidt and Sickles suggested the following estimates: \( \hat{\beta} = \) the usual fixed effects (within) estimate; \( \hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}; \hat{\alpha} = \max_i(\hat{\alpha}_i); \hat{\alpha}_i = \hat{\alpha} - \hat{\alpha}_i \). Consistency of \( \hat{\alpha}_i \) as an estimate of \( \alpha_i \) 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 \( \alpha_i \), and the differences in the \( \alpha_i \) 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:

\[
y_{it} = \alpha_i + x_{it}'\beta + v_{it} - u_{it},
\]

where \( \alpha_i \) 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 \( \alpha_i \) 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 $\alpha_i$ 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:

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

where $\alpha_i$ and $v_{it}$ are normal, and $u_{it}$ and $\eta_i$ 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; $\eta$ is time-invariant (persistent) inefficiency; $v$ is idiosyncratic noise; and $\alpha$ 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 ($b$) from inefficiency ($a$). We observe the basic data $y$, $x$ and $z$. We will also assume that we
observe some time-invariant variables $h_i$ that are uncorrelated with $v, a,$ and $b$ (and therefore with $v$ and $c$). The variables $h$ 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 $h$ 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 $\delta$ 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 $b$]

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

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

$$E(v_{it}|h_i) = E(a_i|h_i) = E(b_i|h_i) = 0$$

These assumptions imply that $v_{it}$ is uncorrelated with $x_i^*, a_i, b_i$ and $h_i$, and that $h_i$ is uncorrelated with $a_i$ and $b_i$.

Under these assumptions, the following moment conditions hold:

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

(MC2)  $$E h_i(y_i - x_i'\beta - z_i'\gamma) = 0$$

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

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 ($\bar{x}_{it}$) are uncorrelated with $c_i$. That would be a Breusch-Mizon-Schmidt (1989) type assumption, and if we made it the individual deviations from means would be part of $h_t$. Note also that (MC1), which requires only Assumption 1, is sufficient (given some obvious regularity conditions) to identify $\beta$; it leads to the so-called “within” estimator. However, to estimate $\gamma$ we need (MC2), in which the number of exogenous instruments $h$ is at least as large as the number of time invariant variables $z$, 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.

An important special case is the “intercept only” case in which $z_i$ contains only an intercept, so that $z_i = 1$ for all $i$, and there are no other time invariant instruments. Then $h_i = z_i = 1$, $\hat{\beta}$ equals the within estimate, and $\hat{\gamma} = \bar{y} - \bar{x}' \hat{\beta}$.

Given estimates of $\beta$ and $\gamma$, we can estimate the individual effects $c_i$. The usual estimates of the individual effects $c_i$ are given by

\begin{equation}
\hat{c}_i = \bar{y}_i - \bar{x}_i' \hat{\beta} - z_i' \hat{\gamma}
\end{equation}

In the intercept only case this is equivalent to $\hat{c}_i = (\bar{y}_i - \bar{y}) - (\bar{x}_i - \bar{x})' \hat{\beta}$. These would be, for example, the deviations from the overall mean (over $i$) of the coefficients of the individual-specific dummy variables in a fixed-effects regression calculated as OLS with individual dummy variables (“OLSDV”).

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.

A simple calculation shows that \( \hat{c}_i = c_i + \tilde{v}_i + \text{terms involving estimation error in } \hat{\beta} \text{ and } \hat{\gamma} \) that are asymptotically (as \( n \to \infty \)) negligible. Correspondingly \( \text{var}(\hat{c}_i) \equiv \sigma_{\hat{c}}^2 \equiv \sigma_c^2 + \frac{1}{T} \sigma_v^2 \).

Note that the difference between \( \hat{c}_i \) and \( c_i \) is negligible under large \( T \) asymptotics, but the difference between \( \hat{c}_i \) and \( c_i \) cannot be ignored under fixed \( T \) asymptotics.

Consistent estimation of \( \sigma_c^2 \) (\( = \sigma_a^2 + \sigma_b^2 \)) 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 \( \beta \) and \( \gamma \), define the within and between sums of squares: \( \text{SSE}_W = \sum_i \sum_t (\tilde{y}_{it} - \bar{x}_t \hat{\beta})^2 \) and \( \text{SSE}_B = \sum_i (\bar{y}_t - \bar{x}_t \hat{\beta} - z_i \hat{\gamma})^2 = \sum_i \hat{c}_i^2 \). Then \( \hat{\sigma}_v^2 = \frac{1}{n(T-1)} \text{SSE}_W \) is a consistent estimate of \( \sigma_v^2 \), \( \hat{\sigma}_{\hat{c}}^2 = \frac{1}{n} \text{SSE}_B \) 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 \( \sigma_c^2 \). 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 \( \sigma_c^2 \) 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 \( \hat{c}_i \):

\[
\hat{c}_i = L(c_i \mid \hat{c}_i) = \frac{\text{cov}(c, \hat{c})}{\text{var}(\hat{c})} \hat{c}_i = \frac{\sigma_c^2}{\sigma_{\hat{c}}^2} \hat{c}_i .
\]

(Appendix 2 discusses 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 $h$ that were uncorrelated with $v$, $a$ and $b$. In this section we will assume that we also have variables $q_1$ and $q_2$ such that:

$q_1$ is uncorrelated with $v$ and $b$ but correlated with $a$

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

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 $q_1$, or variables indicating the physical location or climate of the farm as part of $q_2$.

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

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

**Assumption 3.**

$$E(q_{1i}v_{it}) = E(q_{1i}b_i) = 0$$

Note that we do not assume that $E(q_{1i}a_i) = 0$, and indeed we want $a$ and $q_1$ to be correlated. That is what distinguishes $q_1$ from $q_2$.

It will be useful to define a little more notation. For generic random variables $y$ and $x$, with means $\mu_y$ and $\mu_x$, we define the covariance $C(y,x) = E(y - \mu_y)(x - \mu_x)' \equiv \Omega_{yx}$ and the variance $V(x) = C(x,x) \equiv \Omega_{xx}$, which of course are familiar concepts. From these we can construct the linear projection of $y$ on $x$: $L(y|x) = \Omega_{yx}\Omega_{xx}^{-1}x$. But we also define the uncentered covariance $C_y(y,x) = E(yx') \equiv \Sigma_{yx} = C(y,x) + \mu_y\mu_x'$ and the uncentered variance $V_y(x) = C_y(x,x) \equiv \Sigma_{xx}$. From these we can construct the uncentered linear projection of $y$ on $x$:

$$L_y(y|x) = \Sigma_{yx}\Sigma_{xx}^{-1}x.$$

10
The centered linear projection is the one usually discussed and used in textbooks and the literature. However, if it is known that \( \mu^y = 0 \), the uncentered linear projection has smaller mean squared error of prediction of \( y \). This point is discussed in some detail in Appendix 2.

Now we return to the context of estimating \( a_i \) on the basis of \( q_{1i} \). Let \( \Sigma_{1a} = C_\ast(q_{1i}, a_i) = E(q_{1i}a_i) \), where as above \( C_\ast \) represents the uncentered covariance, which in this case is the same as the usual covariance since \( E(a_i) = 0 \). Then we have the following additional moment conditions:

**(MC3)**

\[
E q_{1i}(\tilde{y}_i - \tilde{x}_i'\beta - z_i'\gamma) - \Sigma_{1a} = 0.
\]

Since (MC3) contains the same number of new parameters (\( \Sigma_{1a} \)) 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 \( \Sigma_{1a} \), which is

\[
(10) \quad \hat{\Sigma}_{1a} = \frac{1}{n} \sum_i q_{1i}(\tilde{y}_i - \tilde{x}_i'\beta - z_i'\gamma) = \frac{1}{n} \sum_i q_{1i} \tilde{c}_i.
\]

Now we need to recover estimates of the \( a_i \). Define the uncentered linear projection:

\[
(11) \quad \hat{a}_i = L_\ast(a_i|q_{1i}) = y'a_i q_{1i}, \quad \gamma_a = \Sigma_{11}^{-1}\Sigma_{1a}, \quad \Sigma_{11} = V_\ast(q_1) = E(q_1q_1').
\]

We observe \( q_{1i} \) so we can calculate \( \hat{\Sigma}_{11} = \frac{1}{n} \sum_i q_{1i}q_{1i}' \). Also from (10) we have an estimate of \( \Sigma_{1a} \), so we can construct a feasible version of (11), namely \( \hat{a}_i = \hat{\Sigma}_{1a}^{\ast} \hat{\Sigma}_{11}^{-1} q_{1i} \).

Note that, as discussed above and in Appendix 2, we use the uncentered linear projection because \( E(a_i) = 0 \).

As in Schmidt and Sickles (1984), our inefficiency measures are differences in the \( a_i \).

That is, our estimate of inefficiency for firm \( i \) is \( \hat{u}_i = \max_j \hat{a}_j - \hat{a}_i \). The \( \hat{u}_i \) are \( \geq 0 \) 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 \( q_{2i} \) that are uncorrelated with \( v_{it} \) and
$a_l$ but correlated with $b_l$. This allows us to estimate the $b_l$. Importantly, it also allows us to improve our estimates of the $a_i$ when $q_{1i}$ and $q_{2i}$ are correlated. Similarly, we can improve our estimates of the $a_i$ by including $h_i$ in the projection set, when $q_{1i}$ and $h_i$ 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.**

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

Note that we do not assume that $E(q_{2i}b_i) = 0$.

Under Assumption 4, the following moment conditions hold:

$$(MC4) \quad E q_{2i}(\bar{y}_i - \bar{x}_i'\beta - z_i'\gamma) - \Sigma_{2b} = 0$$

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 $\beta$ and $\gamma$ and (MC3) gives us the estimate of $\Sigma_{1a}$. Finally (MC4) implies the estimate of $\Sigma_{2b}$, which is

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

This would lead to an estimate of $b_i$ that is similar in spirit to the estimate of $a_i$ given in (11) above; namely, $\hat{b}_i = \hat{\Sigma}_{2b}^{-1}q_{2i}$. An interesting observation is that adding $q_{2i}$ and $h_i$ to the projection set allows for a better estimate of $a_i$, even though $q_{2i}$ and $h_i$ are uncorrelated with $a_i$, if $q_{2i}$ and/or $h_i$ are correlated with $q_{1i}$. We consider the linear projection of $a$ on $q^* = \begin{bmatrix} q_1 \\ q_2 \\ h \end{bmatrix}$:

$$L_a(a|q^*) = \Sigma_{aq}\Sigma^{-1}q^* = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1h} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2h} \\ \Sigma_{h1} & \Sigma_{h2} & \Sigma_{hh} \end{bmatrix}^{-1} \begin{bmatrix} q_1 \\ q_2 \\ h \end{bmatrix}$$
where $\Sigma \equiv E(q^* q^{**}) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1h} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2h} \\ \Sigma_{h1} & \Sigma_{h2} & \Sigma_{hh} \end{bmatrix}$. A technical detail is that, because we are dealing here with uncentered covariances, $h$ can be relevant (and $\Sigma$ can be nonsingular) even if $h$ contains only an intercept.

We can simplify this expression a bit if we define $q_{2h,i} = \begin{bmatrix} q_{2i} \\ h_i \end{bmatrix}$, and correspondingly

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{1,2h} \\ \Sigma_{2h,1} & \Sigma_{2h,2h} \end{bmatrix}.$$ Then

$$L_*(a | q^*) = [\Sigma_1a, 0] \begin{bmatrix} \Sigma_{11} & \Sigma_{1,2h} \\ \Sigma_{2h,1} & \Sigma_{2h,2h} \end{bmatrix}^{-1} \begin{bmatrix} q_1 \\ q_{2h} \end{bmatrix} = \Sigma_{1a} (\Sigma_{11} q_1 + \Sigma_{1,2h} q_{2h}),$$

where $\Sigma_{jk}$ represents a block of $\Sigma^{-1}$. In terms of estimates, (14) leads to

$$\hat{a}_i = \hat{a}_{1a} \left( \hat{\Sigma}_{11} q_1 + \hat{\Sigma}_{1,2h} q_{2h,i} \right).$$

The estimate in (15) is indeed better than the one in (11). For $\hat{a}$ in (11) we have, ignoring the effects of parameter estimation in $\hat{\Sigma}_1a$ and $\hat{\Sigma}_{11}$, $V_*(\hat{a}) = \Sigma_{1a} \Sigma_{11}^{-1} \Sigma_{1a}$. For $\bar{a}$ in (15) we have $V_*(\bar{a}) = \Sigma_{1a} \Sigma_{11}^{-1} \Sigma_{1a}$ where $\Sigma_{11} = (\Sigma_{11} - \Sigma_{1,2h} \Sigma_{2h,2h}^{-1} \Sigma_{2h,1})^{-1}$. Here $\Sigma_{11} - \Sigma_{1,2h} \Sigma_{2h,2h}^{-1} \Sigma_{2h,1}$ is smaller than $\Sigma_{11}$, so its inverse is bigger. Therefore $V_*(\bar{a})$ is bigger than $\text{var}(\hat{a})$. That is good because it is the explained variation. The unexplained variation (i.e. $V_*(a - \bar{a})$) is smaller.

Although we are not primarily interested in the $b_i$, we have the corresponding result that

$$L_*(b | q^*) = \Sigma_{bq} \Sigma_{11}^{-1} q^* = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1h} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2h} \\ \Sigma_{h1} & \Sigma_{h2} & \Sigma_{hh} \end{bmatrix}^{-1} \begin{bmatrix} q_1 \\ q_2 \\ h \end{bmatrix},$$

which we can express (using notation similar to that given above) as

$$L_*(b | q^*) = [\Sigma_{2b}, 0] \begin{bmatrix} \Sigma_{22} & \Sigma_{2,1h} \\ \Sigma_{1h,2} & \Sigma_{1h,1h} \end{bmatrix}^{-1} \begin{bmatrix} q_2 \\ q_{1h} \end{bmatrix} = \Sigma_{2b} (\Sigma_{22} q_2 + \Sigma_{2,1h} q_{1h}),$$

4.3 Projections onto $\hat{c}$

The methods of Sections 4.1 and 4.2 yield estimates of $a_i$ and $b_i$ but these do not add up
to either $\hat{c}_t$ or $\dot{c}_t$, whereas in some intuitive 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 $a_t$ and $b_t$ on a projection set that includes $\hat{c}_t$. To do so, we make an additional assumption, which we maintain along with Assumptions 1-4.

**ASSUMPTION 5.**

$$\sigma_{ab} = E(a_t b_t) = 0.$$  

Then we have the following simple estimates of $a_t$ and $b_t$.

$$(18) \quad \hat{a}_t = \frac{\text{cov}(a_t, \hat{c}_t)}{\text{var}(\hat{c})} \hat{c}_t = \frac{\sigma_a^2}{\sigma^2_c} \hat{c}_t, \quad \hat{b}_t = \frac{\text{cov}(b_t, \hat{c}_t)}{\text{var}(\hat{c})} \hat{c}_t = \frac{\sigma_b^2}{\sigma^2_c} \hat{c}_t.$$  

These estimates satisfy the adding up constraint that $\hat{a}_t + \hat{b}_t = \dot{c}_t$, where $\dot{c}_t$ is defined in equation (8). (It is not the case that $\hat{a}_t + \hat{b}_t = \hat{c}_t$, except in the large T case.)

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

Unfortunately the results in (18) are not feasible without estimating or specifying $\sigma_a^2$ and $\sigma_b^2$. We can estimate $\sigma_c^2$ and $\sigma^2_c$ from standard panel data methods, as discussed above, but without further assumptions we cannot estimate $\sigma_a^2$ and $\sigma_b^2$. We will return to this point in Section 4.4.

In equation (18) 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 $q^* = \begin{bmatrix} q_1 \\ q_2 \\ h \end{bmatrix}$. We ought to be able to improve on both of these estimates if we project $a$ or $b$ onto $q^{**} = \begin{bmatrix} \hat{c} \\ q_1 \\ q_2 \\ h \end{bmatrix}$. So now we will have
\[ (19A) \quad L_*(a|q^*) = \Sigma_{aq^*} V^{-1} (q^*) q^* = [\sigma_a^2, \Sigma_{1a}, 0, 0] V^{-1} q^* \]

where
\[
V = \begin{bmatrix}
\sigma_c^2 & \Sigma_{1a}' & \Sigma_{2b}' & 0 \\
\Sigma_{1a} & \Sigma_{11} & \Sigma_{12} & \Sigma_{1h} \\
\Sigma_{2b} & \Sigma_{21} & \Sigma_{22} & \Sigma_{2h} \\
0 & \Sigma_{h1} & \Sigma_{h2} & \Sigma_{hh}
\end{bmatrix}
\]
is shorthand for \( V, (q^*) \). And similarly

\[ (19B) \quad L_*(b|q^*) = \Sigma_{bq^*} V^{-1} q^* = [\sigma_b^2, 0, 0, 0] V^{-1} q^* \]

\[ (19C) \quad L_*(c|q^*) = \Sigma_{cq^*} V^{-1} q^* = [\sigma_c^2, \Sigma_{1a}', \Sigma_{2b}', 0] V^{-1} q^* \]

Let \( \tilde{b}_l, \tilde{a}_l \) and \( \tilde{c}_l \) be the estimates that we get if we evaluate (19A), (19B) and (19C) using consistent estimates of \( \sigma_a^2, \sigma_b^2, \sigma_c^2, \Sigma_{1a}, \Sigma_{2b} \) and \( V \). Then these are better (smaller mean squared error) estimates than \( b_l, a_l \) and \( c_l \). Furthermore, they satisfy the adding up constraint that \( a_l + b_l = c_l \).

We have \( \tilde{c}_l = \hat{c}_l \) in the large \( T \) case though not in the fixed \( T \) case. In the large \( T \) case,

\[ \sigma_c^2 = \sigma_c^2 \]

and therefore

\[ L_*(c|q^*) = [\sigma_c^2, \Sigma_{1a}', \Sigma_{2b}', 0] V^{-1} q^* = [1, 0, 0, 0] q^* = \hat{c}, \]

where the term \([1, 0, 0, 0]\) arises because \([\sigma_c^2, \Sigma_{1a}', \Sigma_{2b}', 0]\) is the first row of \( V \).

4.4 Estimation of \( \sigma_a^2 \) and \( \sigma_b^2 \)

As mentioned above, consistent estimation of \( \sigma_c^2 \) and \( \sigma_c^2 \) is a solved problem. As a result the estimator \( \hat{c}_l \) given in equation (8) is feasible, and so is \( \tilde{c}_l \) in (19C).

Without further assumptions it does not seem possible to obtain consistent estimates of \( \sigma_a^2 \) and \( \sigma_b^2 \) (separately). However, we can provide some bounds, as follows. We know that

\[ V, (a|q^*) = \begin{bmatrix} \sigma_a^2 & \Sigma_a' \\ \Sigma_a & \Sigma \end{bmatrix} \]

must be positive semi-definite, where \( \Sigma_a = [\Sigma_{1a}, 0] \) and as above \( \Sigma \) is the variance matrix of \( q^* \). Therefore its determinant must be greater than or equal to zero. Using a result from Searle (1982, p. 258), \( |\Sigma| \cdot (\sigma_a^2 - \Sigma_a' \Sigma^{-1} \Sigma_a) \). Therefore \( \sigma_a^2 = \]
\[ \Sigma_0' \Sigma_0^{-1} \Sigma_0 \geq 0 \] and \( \sigma_a^2 \geq \Sigma_0' \Sigma_0^{-1} \Sigma_0 = \Sigma_0' \Sigma_1 \Sigma_0' \). Similarly \( \sigma_b^2 \geq \Sigma_0' \Sigma_2 \Sigma_0' \).

We also know that \( \sigma_a^2 = \sigma_a^2 - \sigma_b^2 \) so that \( \sigma_a^2 \leq \sigma_a^2 - \Sigma_0' \Sigma_2 \Sigma_0' \) and similarly \( \sigma_b^2 \leq \sigma_a^2 - \Sigma_0' \Sigma_1 \Sigma_0' \). Therefore

\begin{equation}
\Sigma_0' \Sigma_1 \Sigma_0' \leq \sigma_a^2 \leq \sigma_a^2 - \Sigma_0' \Sigma_2 \Sigma_0' \quad \text{and} \quad \Sigma_0' \Sigma_2 \Sigma_0' \leq \sigma_b^2 \leq \sigma_b^2 - \Sigma_0' \Sigma_1 \Sigma_0' \quad (21)
\end{equation}

We would want to pick values of \( \hat{\sigma}_a^2 \) and \( \hat{\sigma}_b^2 \) that are in the allowable ranges given in (21), and that satisfy \( \hat{\sigma}_a^2 + \hat{\sigma}_b^2 = \hat{\sigma}_c^2 \). (This equality is necessary for the adding up constraint \( \bar{a}_i + \bar{b}_i = \bar{c}_i \) 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 \( \sigma_a^2 \) and \( \sigma_b^2 \) consistently. To do so we will assume parametric models for \( E(a^2|q_1) \) and for \( E(b^2|q_2) \).

We will therefore make the following assumption, which we will maintain in addition to Assumptions 1-5.

**ASSUMPTION 6.**

\[ E(a|b, q_2, v) = 0 \quad \text{and} \quad E(a^2|b, q_2, v) = \sigma_a^2 \]

\[ E(b|a, q_1, v) = 0 \quad \text{and} \quad E(b^2|a, q_1, v) = \sigma_b^2 \]

\[ E(v|a, b, q_1, q_2) = 0 \quad \text{and} \quad E(v^2|a, b, q_1, q_2) = \sigma_v^2 \]

\[ E(a^2|q_1) = \mu_a^{(2)} \cdot \exp[\lambda'_1(q_1 - \bar{q}_1)] \text{ with } \lambda_1 \neq 0 \]

\[ E(b^2|q_2) = \mu_b^{(2)} \cdot \exp[\lambda'_2(q_2 - \bar{q}_2)] \text{ with } \lambda_2 \neq 0 \]

These assumptions are much stronger than Assumptions 3, 4 and 5, because they make statements about conditional expectations, not just correlations, and because they assume parametric forms for some of the conditional expectations, and because they assume that \( E(a^2|q_1) \) and \( E(b^2|q_2) \) are not constant (there is conditional heteroskedasticity). The specific
functional forms given in the last two lines of Assumption 6 are obviously not the only ones that could have been chosen.

We can note that \( \mu_a^{(2)} \) is the expected value of \( E(a^2) \) conditional on \( q_1 = \bar{q}_1 \); loosely, conditional on \( q_1 = E(q_1) \). This is almost but not quite the same as \( \sigma_a^2 \), which is the unconditional expectation of \( a^2 \), and which (by the law of iterated expectations) equals the expectation (over the distribution of \( q_1 \)) of \( \mu_a^{(2)} \cdot \exp[\lambda'_1(q_1 - \bar{q}_1)] \). This is different from \( \mu_a^{(2)} \) because \( E(\exp[\lambda'_1(q_1 - \bar{q}_1)]) \neq 1 \) due to the nonlinearity of the exponential function. Similar statements apply to \( \mu_b^{(2)} \) and \( \sigma_b^2 \).

Clearly \( E(c_i^2|q_{1i}) = E(a_i^2 + b_i^2 + 2a_ib_i)|q_{1i} \) and so, under Assumption 6,

\[
E(c_i^2|q_{1i}) = \sigma_b^2 + \mu_a^{(2)} \cdot \exp[\lambda'_1(q_{1i} - \bar{q}_1)].
\]

Since (for large \( n \)) \( E(\hat{c}_i^2|q_{1i}) = E(c_i^2|q_{1i}) + \frac{1}{T} \sigma_v^2 \), we have (for large \( n \))

\[
E(\hat{c}_i^2|q_{1i}) = \sigma_b^2 + \frac{1}{T} \sigma_v^2 + \mu_a^{(2)} \cdot \exp[\lambda'_1(q_{1i} - \bar{q}_1)]
\]

and similarly

\[
E(\hat{c}_i^2|q_{2i}) = \sigma_a^2 + \frac{1}{T} \sigma_v^2 + \mu_b^{(2)} \cdot \exp[\lambda'_2(q_{2i} - \bar{q}_2)]
\]

Now consider nonlinear least squares based on (23). That is, we minimize the sum of squares

\[
SSE_1 = \sum_i \left( \hat{c}_i^2 - \sigma_b^2 - \frac{1}{T} \sigma_v^2 - \mu_a^{(2)} \cdot \exp[\lambda'_1(q_{1i} - \bar{q}_1)] \right)^2
\]

with respect to \( \sigma_b^2, \mu_a^{(2)} \) and \( \lambda_1 \), and with \( \sigma_v^2 \) replaced by \( \hat{\sigma}_v^2 = \frac{1}{n(T-1)} SSE_W \) (see Section 3 above). This yields an estimate of \( \sigma_b^2 \), which we will call \( \hat{\sigma}_b^2 \), plus estimates of \( \mu_a^{(2)} \) and \( \lambda_1 \) that we will ignore. Similarly, we can consider nonlinear least squares based on (24), in which case we minimize the sum of squares

\[
SSE_2 = \sum_i \left( \hat{c}_i^2 - \sigma_a^2 - \frac{1}{T} \sigma_v^2 - \mu_b^{(2)} \cdot \exp[\lambda'_2(q_{2i} - \bar{q}_2)] \right)^2
\]
to get an estimate $\hat{\sigma}_a^2$, plus estimates of $\mu_a^{(2)}$ and $\lambda_2$ that we will ignore.

Under reasonable regularity conditions, the estimates $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ are consistent. They are presumably not efficient because they ignore the information about $\sigma_a^2$ in (25) and the information about $\sigma_b^2$ in (26). 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.

4.5 Some Remarks on Adding Up

We remarked above that, except when $\hat{c}_i$ is part of the projection set, our estimates of $a_i$ and $b_i$ do not add up to our estimate of $c_i$. For a specific example, in terms of our estimates $\hat{a}_i$ and $\hat{b}_i$ defined above, we have

$$\sigma_a^2 = \Sigma'_{1a} \Sigma_{11}^{-1} \Sigma_{1a} \leq \sigma_a^2$$

and similarly $\sigma_b^2 \leq \sigma_b^2$ so that

$$\sigma_a^2 + \sigma_b^2 \leq \sigma_a^2 + \sigma_b^2 = \sigma_c^2 \leq \sigma_c^2$$

These inequalities are generally strict (correlations are not perfect) and reflect the fact that optimal predictions (fitted values) are less variable than the things they predict. So in some sense we can’t have adding up, unless we use $\hat{c}$ as part of the basis of the prediction.

The point is not necessarily that $c_i$ contains things that are not part of $a_i$ or $b_i$. For example, in the stochastic frontier setting, $a_i$ is the thing of interest, and $b_i \equiv c_i - a_i$ is everything else in $c_i$. The point is rather that $c_i$ will contain things that are not part of $\hat{a}_i$ or $\hat{b}_i$, because we will not have fully explained either $a_i$ or $b_i$. Basically these things are the part of $c_i$ that is uncorrelated with $q_i^*$. There is then a fundamental identification problem, which is the issue of how to allocate this part of $c_i$ to $a_i$ or $b_i$. As we have seen in Section 4.4, the solution to this problem requires some sort of information on $\sigma_a^2$ and $\sigma_b^2$ (or some other type of additional assumption), and that makes sense.

In other possible examples, it may be that $a_i$ and $b_i$ are not exhaustive and so we should not expect or enforce adding up. For example, in the case of the epidemiological application
given in the Introduction, we might have $a_i$ representing lifestyle, $b_i$ representing genetics, and another component (say, $d_i$) representing the quality of medical care available. We can still estimate the model of this paper if $q_1$ is correlated with $a$ but uncorrelated with $b$ and $d$, and if $q_2$ is correlated with $b$ but uncorrelated with $a$ and $d$, but we would not have adding up. Or we could extend the model to included variables $q_3$ that are correlated with $d$ but uncorrelated with $a$ and $b$, but the details of such an extension will be left to future work.

5. DISTINGUISHING HETEROGENEITY AND INEFFICIENCY – MODEL 2

We now will consider Model 2, in which the identifying assumption for $a$ is that the partial correlation of $q_1$ and $b$, given $q_2$ and $h$, equals zero. We will maintain Assumptions 1, 2 and 5 as above, and in particular we continue to assume that $a$ and $b$ are uncorrelated with $h$, but we will replace Assumptions 3 and 4 with the following assumption.

**ASSUMPTION 7.**

$$E(q_{1i}v_{it}) = 0 \text{ and } E(q_{2i}v_{it}) = 0$$

$$L_a(a|q_1, q_2, h) \text{ does not depend on } q_2 \text{ or } h$$

$$L_b(b|q_1, q_2, h) \text{ does not depend on } q_1 \text{ or } h$$

As before let $\Sigma_{1a} = E(q_{1a})$, $\Sigma_{2b} = E(q_{2b})$, $\Sigma_{1b} = E(q_{1b})$, $\Sigma_{2a} = E(q_{2a})$. Then the identifying assumptions for Model 1 were that $h$ is uncorrelated with $a$ and $b$ and that $\Sigma_{1b} = 0$ and $\Sigma_{2a} = 0$. Assumption 6 is different. Assumption 6 does not imply that $\Sigma_{1b} = 0$ and $\Sigma_{2a} = 0$; nor do $\Sigma_{1b} = 0$ 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 $a$ is ability of the farmer, $b$ is soil
quality, \( q_1 \) is education of the farmer, and \( q_2 \) 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 (\( \Sigma_{2a} \neq 0 \)), but conditional on education, ability will not be correlated with location. Also soil quality will be correlated with education (\( \Sigma_{1b} \neq 0 \)), 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 \( q_1, q_2 \) and \( h \)

As in Section 4.2, let \( \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1h} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{2h} \\ \Sigma_{h1} & \Sigma_{h2} & \Sigma_{hh} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{1,2h} \\ \Sigma_{2h,1} & \Sigma_{2h,2h} \end{bmatrix} \) and let \( \Sigma^{jk} \) represent a block of \( \Sigma^{-1} \). Then the linear projection of \( a \) on \( q^* = \begin{bmatrix} q_1 \\ q_2 \\ h \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2_h \end{bmatrix} \) is:

\[
L_*(a|q^*) = \Sigma_{aq} \Sigma^{-1} q^* = [\Sigma'_{1,2h}, \Sigma'_{2h,a}] \begin{bmatrix} \Sigma_{11} & \Sigma_{1,2h} \\ \Sigma_{2h,1} & \Sigma_{2h,2h} \end{bmatrix}^{-1} \begin{bmatrix} q_1 \\ q_2_h \end{bmatrix} \\
= [\Sigma'_{1a} \Sigma_{11} + \Sigma'_{2h,a} \Sigma_{2h,1}^{2h,1}] q_1 + [\Sigma'_{1,2a} \Sigma_{1,2h} + \Sigma'_{2h,a} \Sigma_{2h,2h}^{2h,2h}] q_{2h}
\]

Assumption 7 requires that \( L_*(a|q^*) \) should not depend on \( q_{2h} \), so we must have \( \Sigma'_{1a} \Sigma_{1,2h} + \Sigma'_{2h,a} \Sigma_{2h,2h}^{2h,2h} = 0 \), or

\[
(29) \quad \Sigma'_{2h,a} = -\Sigma_{1a} \Sigma_{1,2h} \Sigma_{2h,2h}^{-1}.
\]
Substituting this for $\Sigma'_{2h,a}$ in equation (28), we obtain

\begin{equation}
L_*(a|q^*) = \Sigma'_{1a} \left[ \Sigma_{11} - \Sigma_{12} \Sigma_{21} (\Sigma_{22}^{-1} \Sigma_{21})^{-1} \Sigma_{21} \right] q_2 = \Sigma'_{1a} \Sigma_{11}^{-1} q_1 ,
\end{equation}

using a standard result on partitioned inversion.

The same logical argument establishes that $L_*(b|q^*) = \Sigma'_{2b} \Sigma_{22}^{-1} q_2$.

Let $\xi_1 = \Sigma_{11}^{-1} \Sigma_{1a}$ and $\xi_2 = \Sigma_{22}^{-1} \Sigma_{2b}$ so that $L_*(a|q^*) = \xi_1' q_1$ and $(b|q^*) = \xi_2' q_2$. Then

\begin{equation}
L_*(c|q^*) = L_*(a|q^*) + L_*(b|q^*) = L_*(a|q_1) + L_*(b|q_2) = \xi_1' q_1 + \xi_2' q_2 .
\end{equation}

We can obtain consistent estimates of $\xi_1$ and $\xi_2$, say $\hat{\xi}_1$ and $\hat{\xi}_2$, by OLS of $c$ on $q^*$ (i.e. on $q_1$, $q_2$ and $h$). This leads us to our estimates of $a_i$ and $b_i$:

\begin{equation}
\hat{a}_i = \xi_1' q_{1i} , \quad \hat{b}_i = \xi_2' q_{2i} .
\end{equation}

### 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 $q^{**} = \begin{bmatrix} \hat{c} \\ q_1 \\ q_2 \\ h \end{bmatrix}$ and $V = V_*(q^{**}) = \begin{bmatrix} \sigma^2_{\hat{c}} & \Sigma'_{1a} & \Sigma'_{2b} & 0 \\ \Sigma_{1a} & \Sigma_{11} & \Sigma_{12} & \Sigma_{1h} \\ \Sigma_{2b} & \Sigma_{21} & \Sigma_{22} & \Sigma_{2h} \\ 0 & \Sigma_{h1} & \Sigma_{h2} & \Sigma_{hh} \end{bmatrix}$. We can calculate an improved estimate of $c$:

\begin{equation}
\tilde{c}_i = L_*(c|q^{**}_i) = \Sigma_{cq^{**}} V^{-1} q^{**}_i = [\sigma^2_{\hat{c}}, \Sigma'_{1a}, \Sigma'_{2a} + \Sigma_{2b}] 0] V^{-1} q^{**}_i ,
\end{equation}

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. Specifically, all of the elements of $V$ are variances or covariances of observable quantities, except for $\Sigma_{1a}$ and $\Sigma_{2b}$. We can evaluate these elements as $\Sigma_{1a} = \Sigma_{11} \tilde{\xi}_1$ and $\Sigma_{2b} = \Sigma_{22} \tilde{\xi}_2$.

If we have estimated or specified values of $\sigma^2_a$ and $\sigma^2_b$, then we can also calculate

\begin{align*}
\tilde{a}_i &= L_*(a|q^{**}_i) = \Sigma_{aq^{**}} V^{-1} q^{**}_i = [\sigma^2_a, \Sigma'_{1a}, \Sigma'_{2a}] 0] V^{-1} q^{**}_i ,
\end{align*}

\begin{align*}
\tilde{b}_i &= L_*(b|q^{**}_i) = \Sigma_{bq^{**}} V^{-1} q^{**}_i = [\sigma^2_b, \Sigma'_{1b}, \Sigma'_{2b}] 0] V^{-1} q^{**}_i .
\end{align*}
where these would be evaluated at the estimated values of $\Sigma_{aq\cdot\cdot\cdot}$, $\Sigma_{bq\cdot\cdot\cdot}$ and $V$. These should be better estimates than $\hat{b}_l$ and $\hat{a}_l$ because we are projecting onto more explanatory variables. They also satisfy the adding up condition that $\hat{a}_l + \hat{b}_l = \bar{c}_l$.

There remains the issue of obtaining estimates of $\sigma^2_a$ and $\sigma^2_b$. 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 again the case $V_a(q^*)$ must be positive semi-definite, and we can write

$$V_a(q^*) = \begin{bmatrix} \sigma^2_a & \Sigma_a' \\ \Sigma_a & \Sigma \end{bmatrix}$$

where $\Sigma_a = \begin{bmatrix} \Sigma_{1a} \\ \Sigma_{2a} \end{bmatrix}$. This is the same as in Section 4.4 except that now $\Sigma_{2a} \neq 0$. So the fact that this matrix is positive semi-definite implies that

$$\sigma^2_a \geq \Sigma_a'\Sigma^{-1}\Sigma_a. \quad (35)$$

Similarly, $\sigma^2_b \geq \Sigma_b'\Sigma^{-1}\Sigma_b$ where $\Sigma_b = \begin{bmatrix} \Sigma_{1b} \\ \Sigma_{2b} \end{bmatrix}$.

The same logical argument as in Section 4.4 leads to the following bounds (the analogue of equation (21) above):

$$\Sigma_a'\Sigma^{-1}\Sigma_a \leq \sigma^2_a \leq \sigma^2_c - \Sigma_b'\Sigma^{-1}\Sigma_b$$

$$\text{and} \quad \Sigma_b'\Sigma^{-1}\Sigma_b \leq \sigma^2_b \leq \sigma^2_c - \Sigma_a'\Sigma^{-1}\Sigma_a. \quad (36)$$

We will want to pick values of $\hat{\sigma}^2_a$ and $\hat{\sigma}^2_b$ that are in the allowable ranges given in (36), and that satisfy $\hat{\sigma}^2_a + \hat{\sigma}^2_b = \bar{\sigma}^2_c$ so that $\hat{a}_l + \hat{b}_l = \bar{c}_l$.

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 $\sigma^2_a$ and $\sigma^2_b$ consistently. In the present case this will require parametric models for $E(a^2|q_1,q_2)$ and for $E(b^2|q_1,q_2)$. We make the following assumption.

**ASSUMPTION 8.**
\[
E(a|q_1, q_2, h, v) = \xi'_1 q_1 \\
E(b|q_1, q_2, h, v) = \xi'_2 q_2 \\
E(ab|q_1, q_2, h, v) = 0 \\
E(v|q_1, q_2, h, a, b) = 0 \quad E(v^2|q_1, q_2, h, a, b) = \sigma^2_v \\
E(a^2|q_1, q_2, h, v) = \mu_a^{(2)} \cdot \exp[\lambda'_1(q_1 - \bar{q}_1)] \quad \text{with } \lambda_1 \neq 0 \\
E(b^2|q_1, q_2, h, v) = \mu_b^{(2)} \cdot \exp[\lambda'_2(q_2 - \bar{q}_2)] \quad \text{with } \lambda_2 \neq 0
\]

Since \( E(c_i^2|q_{1i}, q_{2i}, h_i) = E(a_i^2 + b_i^2 + 2a_i b_i)|q_{1i}, q_{2i}, h_i \), under Assumption 8,

\[
E(c_i^2|q_{1i}, q_{2i}, h_i) = \mu_a^{(2)} \cdot \exp[\lambda'_1(q_{1i} - \bar{q}_1)] + \mu_b^{(2)} \cdot \exp[\lambda'_2(q_{2i} - \bar{q}_2)]
\]

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

\[
SSE = \sum_i \left( \hat{c}_i^2 - \frac{1}{T} \hat{\sigma}_v^2 - \mu_a^{(2)} \cdot \exp[\lambda'_1(q_{1i} - \bar{q}_1)] - \mu_b^{(2)} \cdot \exp[\lambda'_2(q_{2i} - \bar{q}_2)] \right)^2
\]

where \( \hat{\sigma}_v^2 \) is a consistent estimate of \( \sigma^2_v \) (e.g. based on the within estimate of the basic model, as discussed in Section 3 above), and the term \( \frac{1}{T} \hat{\sigma}_v^2 \) arises to account for the difference between the variance of \( \hat{c}_i \) and the variance of \( c_i \).

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

\[
\hat{\sigma}_a^2 = \frac{1}{n} \sum_i \hat{\mu}_b^{(2)} \exp[\lambda'_1(q_{1i} - \bar{q}_1)].
\]

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

\section{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, as 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 $y_{it} = w_{it}' \delta + \epsilon_{it}$ where $\epsilon_{it} = v_{it} + c_i$.

We write the model for all $T$ observations for firm $i$ as $y_i = W_i' \delta + \epsilon_i$ and for all $NT$ observations we write $y = W \delta + \epsilon$. Similarly we have matrices of deviations from means $\bar{X}_i$ and $\bar{X}$. Define $H_i = e_T \otimes h_i'$, where $e_T$ is a vector of ones, and then $H$ for all $NT$ observations.

Finally, we define the instruments $G_i = [\bar{X}_i, H_i]$ for $T$ observations and $G$ for all $NT$ observations.

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

\[(A1) \quad E G_i' (y_i - W_i \delta) = 0\]

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

GMM of $\delta$ can be based on (A1). Let $\Omega = E G_i' \epsilon_i \epsilon_i' G_i$ and $\Omega_* = I_N \otimes \Omega$, and correspondingly their estimates are $\hat{\Omega}$ and $\hat{\Omega}_*$ where $\hat{\Omega} = \frac{1}{n} \sum_i G_i' \hat{\epsilon}_i \hat{\epsilon}_i' G_i$ and where $\hat{\epsilon}_i = y_i - W_i \hat{\delta}$.

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:

\[(A2) \quad \delta = (W'G_{\delta}*^{-1}G'W)^{-1} W'G_{\delta}*^{-1}G' y\]

(Continuous updating means that the initial estimate of $\delta$ leads to an estimate of $\Omega$, which leads via (A2) to a new estimate of $\delta$, which leads to a new estimate of $\Omega$, etc.)

The above procedure does not put any restrictions on the weighting matrix $\Omega$. We can impose restrictions on $\Omega$ 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]
\[
\text{Var}(v_t | G_i) = \sigma_v^2 \text{ for all } t \\
\text{Var}(c_i | G_i) = \sigma_c^2 \\
\text{Cov}(v_t, c_i | G_i) = 0 \text{ for all } t
\]

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 \( \sigma_v^2 \) and \( \sigma_c^2 \), so it is easy to get an estimate of \( \Omega \) and therefore of \( \Omega_* \). 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 \( \delta \) 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 \( \theta = \sqrt{\frac{T \sigma_v^2}{(T \sigma_v^2 + T \sigma_c^2)}} \). Now we can transform the regression equation to “whiten” the errors:

\[
\Omega^{-1/2} y_i = \Omega^{-1/2} W_i \delta + \Omega^{-1/2} \epsilon_i.
\]

This amounts to “1-\( \theta \)” 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 \( G_i \). In matrix terms for all \( NT \) observations we have

\[
\delta = [W' \hat{\Omega}_*^{-1/2} G (G' \hat{\Omega}_*^{-1} G)^{-1} G' \hat{\Omega}_*^{-1/2} W]^{-1} W' \hat{\Omega}_*^{-1/2} G (G' \hat{\Omega}_*^{-1} G)^{-1} G' \hat{\Omega}_*^{-1/2} y .
\]

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 \( L(y|x) = \Omega_{yx} \Omega_{xx}^{-1} x \) is the linear projection of \( y \) on \( x \), where (as in Section 4.1) \( \Omega \) is used to denote centered variances and covariances. It has the
property that $y - L(y|x)$ is uncorrelated with $x$. Also $\text{var}(L(y|x)) = \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$ and $\text{var}(y - L(y|x)) = \Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$.

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

For the purposes of this paper, it is important that all of these results still hold if $y$ has zero mean and therefore we use uncentered covariances and variances and the uncentered linear projection. That is, if $\Sigma_{yx} = E(yx')$, $\Sigma_{xx} = E(xx')$ and $L_*(y|x) = \Sigma_{yx}\Sigma_{xx}^{-1}x$, then “$y - L_*(y|x)$” is uncorrelated with $x” in the sense that $E[y - L_*(y|x)]x = 0$; $E(L_*(y|x))^2 = \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$; and $E[y - L_*(y|x)]^2 = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \equiv \text{MSE}(L_*(y|x))$.

We now wish to show that $\text{MSE}(L_*(y|x)) \leq \text{MSE}(L(y|x))$ when $\mu_y = 0$. First we note that, when $\mu_y = 0$, $\Sigma_{yy} = \Omega_{yy}$ and $\Sigma_{yx} = \Omega_{yx}$, but $\Sigma_{xx} = \Omega_{xx} + \mu_x\mu_x'$ is bigger than $\Omega_{xx}$. So when $\mu_y = 0$,

(A5)  \[ \text{MSE}(L_*(y|x)) = E[y - L_*(y|x)]^2 = \Omega_{yy} - \Omega_{yx}\Sigma_{xx}^{-1}\Omega_{xy}. \]

Now we need to calculate $\text{MSE}(L(y|x))$. We have

(A6)  \[ \text{MSE}(L(y|x)) = E[y - \Omega_{yx}\Omega_{xx}^{-1}x][y - \Omega_{yx}\Omega_{xx}^{-1}x]' \]
\[ = \Omega_{yy} + \Omega_{yx}\Omega_{xx}^{-1}\Sigma_{xx}\Omega_{xx}^{-1}\Omega_{xy} - 2\Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy} \]

So to show that $\text{MSE}(L_*(y|x)) \leq \text{MSE}(L(y|x))$, we need to show that

(A7)  \[ \text{MSE}(L(y|x)) - \text{MSE}(L_*(y|x)) = \Omega_{yx}[\Omega_{xx}^{-1}\Sigma_{xx}\Omega_{xx}^{-1} - 2\Omega_{xx}^{-1} + \Sigma_{xx}^{-1}]\Omega_{xy} \]

is positive semi-definite (psd).

We substitute $\Sigma_{xx} = \Omega_{xx} + \mu_x\mu_x'$ to obtain
\[ \text{(A8)} \quad \text{MSE}(L(y|x)) - \text{MSE}(L_s(y|x)) = \Omega_{yx}[\Sigma_{xx}^{-1} - \Omega_{xx}^{-1} + \Omega_{xx}^{-1} \mu_x \mu_x' \Omega_{xx}^{-1}] \Omega_{xy} \]

Now we use the matrix identity (Sherman-Morrison formula) that \( \Sigma_{xx}^{-1} = \Omega_{xx}^{-1} + (1 + \mu_x' \Omega_{xx}^{-1} \mu_x)^{-1} \Omega_{xx}^{-1} \mu_x \mu_x' \Omega_{xx}^{-1} \). With this substitution and a little arithmetic we obtain

\[ \text{(A9)} \quad \text{MSE}(L(y|x)) - \text{MSE}(L_s(y|x)) = \Omega_{yx}[\frac{\mu_x' \Omega_{xx}^{-1} \mu_x}{1 + \mu_x' \Omega_{xx}^{-1} \mu_x} - \Omega_{xx}^{-1} \mu_x \mu_x' \Omega_{xx}^{-1}] \Omega_{xy} , \]

which is indeed psd. So \( \text{MSE}(L_s(y|x)) \leq \text{MSE}(L(y|x)) \) when \( \mu_y = 0 \).
REFERENCES


