Advanced Microeconometrics - Project 1

Group 21
Lukas Malte Kemeter (xtj271),
Felix Heuer (pfv999),
Malte Jacob Rattenborg (rnp562)

January 30, 2022

Problem 1

a:
To show that by multiplying the matrix M on D, we use the corresponding expression for M. Both M and D are NTxN matrices.

$$\begin{aligned} MD &= [I_{NT} - D(D'D)^{-1}D']D \\ &= DI_{NT} - D(D'D)^{-1}D'D \\ &= D - D \\ &= 0_{NTxN} \end{aligned}$$

The symmetry of M can be shown by the following transformation using the properties of the identity matrix:

$$\begin{aligned} M' &= [I_{NT} - D(D'D)^{-1}D']' \\ &= I_{NT} - [D(D'D)^{-1}D']' \\ &= I_{NT} - D[(D'D)^{-1}]'D' \\ &= I_{NT} - D[(D'D)']^{-1}D' \\ &= I_{NT} - D(D'D)^{-1}D' \\ &= M \end{aligned}$$

The idempotence of M can be shown by the following transformation:

$$\begin{aligned}MM &= [I_{NT} - D(D'D)^{-1}D'][I_{NT} - D(D'D)^{-1}D'] \\ &= I_{NT} - D(D'D)^{-1}D' - D(D'D)^{-1}D' + D(D'D)^{-1}D'D(D'D)^{-1}D' \\ &= I_{NT} - D(D'D)^{-1}D' - D(D'D)^{-1}D' + D(D'D)^{-1}D' \\ &= I_{NT} - D(D'D)^{-1}D' \\ &= M\end{aligned}$$

b:

Here we follow the hint given in the question to inspect the FOC of the OLS estimator, which is done in the following to show that $M\hat{u} = \hat{u}$. The given model is:

$$y = D\alpha + X\beta + u$$

In the given model the OLS regression chooses α and β to minimize the sum of square errors which can be written as:

$$\begin{aligned}u'u &= [y - D\alpha - X\beta]'[y - D\alpha - X\beta] \\ &= y'y - y'D\alpha - y'X\beta - \alpha'D'y + \alpha'D'D\alpha - \alpha'D'X\beta - \beta'X'y + \beta'X'D\alpha + \beta'X'X\beta\end{aligned}$$

To minimize this expression we take the derivatives wrt. to the parameters $\alpha \; and \; \beta$ and set these expressions to 0. Note that $X'D\hat{\alpha}$ and $D'X\hat{\beta}$ are scalars, which is the reason that taking the sum of these terms and their transposed version gives us the term multiplied by 2.

$$\begin{aligned}\frac{du'u}{d\beta} = -y'X - \hat{\alpha}'D'X - X'y + X'D\hat{\alpha} + 2X'X\hat{\beta} &= \; 0 \\ -y'X - (X'D\hat{\alpha})' - X'y + X'D\hat{\alpha} + 2X'X\hat{\beta} &= \; 0 \\ 2X'X\hat{\beta} &= 2X'y - 2X'D\hat{\alpha} \\ \hat{\beta} &= (X'X)^{-1}X'y - (X'X)^{-1}XD\hat{\alpha} \\ \frac{du'u}{d\alpha} = -y'D - D'y + 2D'D\hat{\alpha} + D'X\hat{\beta} + \hat{\beta}'X'D &= \; 0 \\ -y'D - D'y + 2D'D\hat{\alpha} + D'X\hat{\beta} + (D'X\hat{\beta})' &= \; 0 \\ 2D'D\hat{\alpha} &= 2D'y + 2D'X\hat{\beta} \\ \hat{\alpha} &= (D'D)^{-1}D'y + (D'D)^{-1}D'X\hat{\beta} \\ \hat{\alpha} &= (D'D)^{-1}D'(y + X\hat{\beta})\end{aligned}$$

We use these results to show the proposed expression in the following, by plugging in $\hat{\alpha}$ into the expression for $M\hat{u}$:

$$
\begin{aligned}
M\hat{u} &= [I_{NT} - D(D'D)^{-1}D']\hat{u} \\
&= \hat{u} - D(D'D)^{-1}D'(y - D\hat{\alpha} + X\hat{\beta}) \\
&= \hat{u} - D(D'D)^{-1}D'y + D(D'D)^{-1}D'D\hat{\alpha} - D(D'D)^{-1}D'X\hat{\beta} \\
&= \hat{u} - D(D'D)^{-1}D'y + D\hat{\alpha} - D(D'D)^{-1}D'X\hat{\beta} \\
&= \hat{u} - D(D'D)^{-1}D'y + D(D'D)^{-1}D'(y + X\hat{\beta}) - D(D'D)^{-1}D'X\hat{\beta} \\
&= \hat{u} - D(D'D)^{-1}D'y + D(D'D)^{-1}D'y + D(D'D)^{-1}D'X\hat{\beta} - D(D'D)^{-1}D'X\hat{\beta} \\
&= \hat{u}
\end{aligned}
$$

c:

To show that $\hat{\beta} = (X'MX)^{-1}X'My$ we use the expression for β derived in b). Then we substitute the solution for $\hat{\alpha}$ into $\hat{\beta}$. Recall that $D(D'D)^{-1}D' = I_{NT} - M$.

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y - (X'X)^{-1}XD\hat{\alpha} \\
&= (X'X)^{-1}X'y - (X'X)^{-1}XD(D'D)^{-1}D'(y + X\hat{\beta}) \\
&= (X'X)^{-1}X'y - (X'X)^{-1}XD[(D'D)^{-1}D'y + (D'D)^{-1}D'X\hat{\beta}] \\
&= (X'X)^{-1}X'y - (X'X)^{-1}XD(D'D)^{-1}D'y + (X'X)^{-1}XD(D'D)^{-1}D'X\hat{\beta} \\
&= (X'X)^{-1}X'y - (X'X)^{-1}X[I_{NT} - M]y + (X'X)^{-1}X[I_{NT} - M]X\hat{\beta} \\
&= (X'X)^{-1}X'My + \hat{\beta} - (X'X)^{-1}X'M\hat{\beta}
\end{aligned}
$$

$$
\begin{aligned}
(X'X)^{-1}X'M\hat{\beta} &= (X'X)^{-1}X'My \\
\hat{\beta} &= (X'MX)^{-1}X'My
\end{aligned}
$$

The computational problem with estimating the long regression was inverting the (N+K)x(N+K) matrix Z'Z. As M is a NTxNT matrix and X is a NTxK matrix, the multiplied matrices X'MX is a combined KxK matrix. In usual regressions only a few regressors are included, which means that a KxK matrix would have way fewer dimensions. Therefore, inverting a KxK matrix should not impose computational problems.

d:

To show that $My =$ and $Mx = \ddot{x}$ we calculate matrices D'D and D'y, and accordingly D'x and then use the definition of M.

We start by finding an expression for My:

$$\begin{aligned} My &= [I_{NT} - D(D'D)^{-1}D']y \\ &= y - D(D'D)^{-1}D'y \end{aligned}$$

Using the rules of matrix multiplication shows that $D'D$ can be written as:

$$\begin{bmatrix} T & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & T \end{bmatrix}_{N \times N}$$

Similarly, $D'y$ can be written as:

$$\begin{bmatrix} \sum_{t=1}^{T} y_{1t} \\ \vdots \\ \sum_{t=1}^{T} y_{Nt} \end{bmatrix}_{N \times 1}$$

Combining the two terms we derived above and using the rules of matrix multiplication we find that $(D'D)^{-1}D'y$ can be written as:

$$T^{-1} * \begin{bmatrix} \sum_{t=1}^{T} y_{1t} \\ \vdots \\ \sum_{t=1}^{T} y_{Nt} \end{bmatrix}_{N \times 1}$$

$D(D'D)^{-1}D'y$ can be written as:

$$\begin{bmatrix} 1 & \dots & 0 \\ \vdots & & \vdots \\ 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ \vdots & & \vdots \\ 0 & \dots & 1 \end{bmatrix}_{NT \times N} * \begin{bmatrix} T^{-1} * \sum_{t=1}^{T} y_{1t} \\ \vdots \\ T^{-1} * \sum_{t=1}^{T} y_{Nt} \end{bmatrix}_{N \times 1} = \begin{bmatrix} T^{-1} * \sum_{t=1}^{T} y_{1t} \\ T^{-1} * \sum_{t=1}^{T} y_{1t} \\ T^{-1} * \sum_{t=1}^{T} y_{1t} \\ \vdots \\ T^{-1} * \sum_{t=1}^{T} y_{Nt} \\ T^{-1} * \sum_{t=1}^{T} y_{Nt} \\ T^{-1} * \sum_{t=1}^{T} y_{Nt} \end{bmatrix}_{NT \times 1}$$

We then plug this term into the expression found for My above. This yields the following result, which is equal to the within-transformed y:

$$My = y_{it} - \frac{1}{T} \sum_{t=1}^{T} y_{it} =$$

We use the same procedure to find the expression for MX. We start by using the given definiton of M and calculating Mx:

$$
\begin{aligned}
MX &= [I_{NT} - D(D'D)^{-1}D']X \\
&= X - D(D'D)^{-1}D'X
\end{aligned}
$$

Note x is denoted as x_{kit} so the first indicates k, the second indicates individual i and the third indicates period t. In all other case x will be x_{it}.

By applying the rules of matrix multiplication we can see that $D'X$ can be written as

$$
\begin{bmatrix}
\sum_{t=1}^{T} x_{11t} & \dots & \sum_{t=1}^{T} x_{k1t} \\
\vdots & & \vdots \\
\sum_{t=1}^{T} x_{1Nt} & \dots & \sum_{t=1}^{T} x_{kNt}
\end{bmatrix}_{N \times k}
$$

By using the matrix $(D'D)^{-1}$ and $D'X$, we can see that $D(D'D)^{-1}D'X$ can be written as:

$$
\begin{bmatrix}
T^{-1} * \sum_{t=1}^{T} x_{11t} & \dots & T^{-1} * \sum_{t=1}^{T} x_{k1t} \\
\vdots & & \vdots \\
T^{-1} * \sum_{t=1}^{T} x_{1Nt} & \dots & T^{-1} * \sum_{t=1}^{T} x_{kNt}
\end{bmatrix}_{NT \times k}
$$

We use this matrix and plug it into the expression for Mx that we found before.

$$
MX = X - D(D'D)^{-1}D'X
$$

$$
MX = \begin{bmatrix}
x'_{11} \\
\vdots \\
x'_{1T} \\
\vdots \\
x'_{N1} \\
\vdots \\
x'_{NT}
\end{bmatrix}_{NT \times k}
-
\begin{bmatrix}
T^{-1} * \sum_{t=1}^{T} x_{11t} & \dots & T^{-1} * \sum_{t=1}^{T} x_{k1t} \\
\vdots & & \vdots \\
T^{-1} * \sum_{t=1}^{T} x_{11t} & \dots & T^{-1} * \sum_{t=1}^{T} x_{k1t} \\
\vdots & & \vdots \\
T^{-1} * \sum_{t=1}^{T} x_{1Nt} & \dots & T^{-1} * \sum_{t=1}^{T} x_{kNt} \\
\vdots & & \vdots \\
T^{-1} * \sum_{t=1}^{T} x_{1Nt} & \dots & T^{-1} * \sum_{t=1}^{T} x_{kNt}
\end{bmatrix}_{NT \times k}
$$

This expression is equal to the definition of the within-transformed x.

$$
MX = x_{it} - \bar{x}_i = \ddot{X}
$$

e:

To show that the estimated parameter $\hat{\beta}$ is the within estimator, we use the characteristics of M that we proved in a). We further use the rule: $(AB)' = B'A'$

$$\begin{aligned}
\hat{\beta} &= (X'MX)^{-1}X'My \\
&= (X'MMX)^{-1}X'MMy \\
&= (X'M'MX)^{-1}X'M'My \\
&= ((MX)'MX)^{-1}(MX)'My \\
&= (\bar{\bar{X}}'\bar{\bar{X}})^{-1}\bar{\bar{X}}'
\end{aligned}$$

This proves that the obtained parameter is equivalent to the within estimator.

f:

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}$$

The general OLS minimizes the sum of the squared error terms, which is done in the following. Note that α is assumed to be constant over time, which means the sum over T periods of α is equal to $T * \alpha$.

$$\begin{aligned}
min\sum_{t=1}^{T}(\hat{v}_{it})^2 &= \sum_{t=1}^{NT}(y_{it} - \hat{\alpha}_i + x'_{it}\hat{\beta})^2) \\
\frac{\partial\sum_{t+1}^{T}(\hat{v}_{it})^2}{\partial\hat{\alpha}_i} &- 2\sum_{t=1}^{T}(y_{it} - \hat{\alpha}_i + x'_{it}\hat{\beta})^2) = 0 \\
\sum_{T}\hat{\alpha}_i &= \sum_{T} y_{it} - \sum_{T} x'_{it}\hat{\beta} \\
T * \hat{\alpha}_i &= \sum_{T} y_{it} - \sum_{T} x'_{it}\hat{\beta} \\
\hat{\alpha}_i &= \frac{1}{T}\sum_{T} y_{it} - \frac{1}{T}\sum_{T} x'_{it}\hat{\beta} \\
\hat{\alpha}_i &= \bar{y}_i - \bar{x}'_i\hat{\beta}
\end{aligned}$$

This proves the given expression.

g:

POLS: We take the POLS regression model:

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}$$

The residuals from the pooled OLS regression, can be written as the difference in observed y_{it} and the predicted y_{it} by the individual fixed effect and $\hat{\beta}$ multiplied with the X matrix. We then show that the POLS residuals are equal to the fixed effect resiudals, which are defined as ${}_{it} = {}_{it} - \ddot{x}'_{it}\beta$. We plug in the expression found in f) $\hat{\alpha}_i = \bar{y}_i - \bar{x}'_i\hat{\beta}$.

$$
\begin{aligned}
\hat{u}_{it} &= y_{it} - \hat{\alpha}_i - x'_{it}\hat{\beta} \\
&= y_{it} - [\bar{y}_i - \bar{x}'_i] - x'_{it}\hat{\beta} \\
&= y_{it} - \bar{y}_i + \bar{x}'_i - x'_{it}\hat{\beta} \\
&= {}_{it} - \ddot{x}'_{it}\hat{\beta} \\
&= {}_{it}
\end{aligned}
$$

This shows that the error terms in the POLS model are equivalent to the error term in the FE model, if the FE model is run as a POLS regression of the transformed variables.

h:

To show the given equation we plug in the expression for $\hat{\gamma}$. Please note that $var(\gamma|Z) = 0$, because γ is a scalar and the variance of a scalar has to be 0. We then plug in the given expression that $Var(u|Z) = \sigma^2 I_{NT}$. We use the rule that $var(Av) = Avar(v)A'$.

$$
\begin{aligned}
var(\hat{\gamma}|Z) &= var((Z'Z)^{-1}Z'y|Z) \\
&= var((Z'Z)^{-1}Z'(Z\gamma + u)|Z) \\
&= var(\gamma + (Z'Z)^{-1}Z'u)|Z) \\
&= var(\gamma|Z) + var((Z'Z)^{-1}Z'u)|Z) \\
&= 0 + (Z'Z)^{-1}Z'var(u|Z)Z(Z'Z)^{-1} \\
&= (Z'Z)^{-1}Z'I_{NT}\ \sigma^2 Z(Z'Z)^{-1} \\
&= \sigma^2(Z'Z)^{-1}Z'Z(Z'Z)^{-1} \\
&= \sigma^2(Z'Z)^{-1}
\end{aligned}
$$

i:

To solve this problem, we follow the approach by (Group 18, Project 1). They have correctly pointed out how the variance-covariance matrix of $\hat{\gamma}|Z$ can be understood. (Group 18, Project 1) argue that as $\hat{\gamma}$ is the $\hat{\alpha}$ and $\hat{\beta}$ vectors stacked above each other, the variance-covariance matrix of $\hat{\gamma}|Z$ is a symmetric (N+K)x(N+K) matrix of the form:

$$var(\hat{\gamma}|Z) = \begin{bmatrix} var(\hat{\alpha}|Z) & cov(\hat{\alpha}, \hat{\beta}|Z) \\ cov(\hat{\beta}, \hat{\alpha}|Z) & var(\hat{\beta}|Z) \end{bmatrix}_{(N+K)\times(N+K)}$$

This form comes from the fact that we need to fully define the variance-covariance structure of the vector $\hat{\gamma}$, including both $\hat{\alpha}$ and $\hat{\beta}$. This means the variance-covariance matrix of $\hat{\gamma}$ has to consist of the variance-covariance matrix of $\hat{\alpha}$, $\hat{\beta}$ and the covariance matrices between these two vectors. We have N individual values of $\hat{\alpha}_i$, as they represent individual fixed effects. These N values have covariances between each other and the variances on the diagonal of the matrix. Furthermore, we have K values of $\hat{\beta}$ for each regressor, as the parameters are assumed to have a linear form and do not differ between individuals. These K coefficients have covariances between each other and the variances on the diagonal. The covariances between $\hat{\alpha}$ and $\hat{\beta}$ are the remaining part of the variance-covariance matrix characterizing the paiwise covariance between each parameter.

The solution from h) states that $var(\hat{\gamma}|Z) = \sigma^2(Z'Z)^{-1}$. We know that σ^2 is a scalar. Therefore it is only crucial to figure out how $(Z'Z)^{-1}$ should be treated as the rules of matrix multiplication point out that a scalar is multiplied elementwise to a matrix. Accordingly, to find the expression for $var(\hat{\beta}|Z)$, we only need to find the lower KxK block of $(Z'Z)^{-1}$, which is the lower KxK matrix of $var(\hat{\gamma}|Z)$, as pointed out by (Group 18, Project 1). The term for the KxK lower block of $(Z'Z)^{-1}$ is given for this question. This means to find the solution for $Var(\hat{\beta}|Z)$, we need to elementwise multiply the given expression for the KxK lower block of $(Z'Z)^{-1}$ and σ^2. We insert the definition of M, $D(D'D)^{-1}D' = I_{NT} - M$, and use the result from d), which states that $MX = \overline{\underline{X}}$. We also use the symmetry and idemponent properties of the M matrix, namely: $M = M'$ and $M = MM$. In the following, we generally follow the approach by (Group 18, Project 1), but use $D(D'D)^{-1}D' = I_{NT} - M$ instead of $M = I_{NT} - D(D'D)^{-1}D'$, which simplifies the expression in our opinion:

$$\begin{aligned}
var(\hat{\beta}|Z) &= \sigma^2(X'X - X'D(D'D)^{-1}D'X)^{-1} \\
&= \sigma^2(X'X - X'[I_{NT} - M]X)^{-1} \\
&= \sigma^2(X'X - [X' - X'M]X)^{-1} \\
&= \sigma^2(X'X - [X'X - X'MX])^{-1} \\
&= \sigma^2(X'MX)^{-1} \\
&= \sigma^2(X'MMX)^{-1} \\
&= \sigma^2(X'M'MX)^{-1} \\
&= \sigma^2((MX)'MX)^{-1} \\
&= \sigma^2(\overline{\underline{X}}'\overline{\underline{X}})^{-1}
\end{aligned}$$

j:
For the homoskedastic OLS variance estimator, we have to take into account that the estimation for error terms, $\hat{u}^2$, is based on a previously computed $\hat{\beta}$ vector: $\hat{u}_i = y_i - x_i * \hat{\beta}$ In other words, the variance for the OLS estimator can be described as the overall average of the sum of squared error terms. When calculating the variance of $\hat{\beta}$, the denominator has to be corrected by $N - K$ in order to account for included estimation errors when estimating $\hat{u}_i$. We argue that for computing $\hat{u}$, the K scalars for beta coefficients are used (which are itself estimated values). Thus the variance needs to be adjusted in order to account for the potential stochastic variation in σ^2 (β estimates were used to compute u_i) and K degrees of freedom need to be deducted in the variance formula.
By the same logic we estimate the variance in the panel case. We have N*T observations, while we still have to subtract K degrees of freedom to account for the estimated K parameters. Additionally, we estimate average error terms for each person in the panel case, which accounts to N averages. These calculated averages from the sample serve as estimates for the expected value in the the population. These estimated averages are stochastic variables by themselves, which is the reason we have to account for the additional variance they add to the estimation of $\hat{\beta}$. Nevertheless, as found in g), when the FE model is estimated by running a POLS on the transformed variables we receive the same error term as in the POLS case. Furthermore in h) we showed that the variance of the estimate of the POLS parameter γ only takes into account variable Z. This means that the POLS estimation of the transformed variables does not account for this additional stochastic variation when estimation the N fixed effects. This is the reason we subtract N degrees of freedom from the denominator in the homoskedastic panel case to account for this additional stochastic uncertainty. The resulting corrected standard errors should be estimated by:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{NT - N - K} * \sum_i \sum_t \ddot{\epsilon}_{it}^2$$

The standard error correction matters, because otherwise the estimated standard errors would be underestimated. Overall the general OLS variance estimator, which would be the estimator if we apply the POLS on the transformed variables, as shown in (g), does not account for the estimated averages and therefore is too small. The degrees of freedom correction increases the estimated variance accordingly.

Problem 2

(1)

To formulate a linear regression model, we apply a log transformation according to:

$$F = A * K^{\alpha} * L^{\beta} \;\Rightarrow\; ln(F) = ln(A) + \alpha ln(K) + \beta ln(L)$$

$$ln(F) \equiv y_{it} = c_{it} + \gamma * x'_{it} + \epsilon_{it}$$

where $\gamma = [\alpha \ \beta]$ contains the coefficients for the two input regressors, $x_{it} = [x1_{it} \ x2_{it}] = [ln(K) \ ln(L)]$, ϵ_{it} being the error term, y_{it} the log output level and c_i being firm fixed effects. Note that c_i is denoted as time variant above but will be assumed to be time invariant as the general specification of c_{it} is not estimable. Based on Table 1, the following section discusses the choice between different estimators for the given setting.

Table 1: Assumptions

	Consistency	Efficiency	Estimator
POLS	full rank of $X'X$ contemp. exog. $E[c_i x_{it}] = 0$ $E[\epsilon_{it} x_{it}] = 0$	$\mathrm{E}[\mathrm{c}_i^2 \vert x_i] = \sigma_c^2 = 0$	$(\mathrm{X'X})^{-1} X'y$
FE	full rank of $\ddot{\mathrm{X}}'\ddot{\mathrm{X}}$ strict exog. $E[\epsilon_{it} \vert c_i, x_{i1}, .., x_{iT}] = 0$	$E[\epsilon_i \epsilon_i' \vert x_i c_i] = \sigma_\epsilon^2 I_T$	$(\ddot{\mathrm{X}}'\ddot{\mathrm{X}})^{-1} \ \ddot{\mathrm{X}}'\ddot{\mathrm{y}}$
FD	full rank of $\triangle X' \triangle X$ sequential exog. $E[\epsilon_{it} \vert x_{it+1}, x_{it}, x_{it-1}]$	$E[u_i u_i' \vert x_i c_i] = \sigma_u^2 I_{T-1}$ with $u_i = \triangle \epsilon_i$	$(\triangle X' \triangle X)^{-1} \triangle X' \triangle y$
RE	full rank of $\tilde{X}'\tilde{X}$ strict exog. $E[\epsilon_{it} x_i] = 0$ $E[x_{it} c_i] = 0$	$\mathrm{E}[\epsilon_i \epsilon_i' \vert x_i, c_i] = \sigma_\epsilon^2 I_T$ $\mathrm{E}[\mathrm{c}_i^2 \vert x_i] = \sigma_c^2$	$(\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{y}$

Full rank of $X'X$ implies that the matrix is invertable.
ϵ correspond to the error terms of a linear panel-data model.

Pooled Ordinary Least Squares (POLS) vs. Fixed Effects (FE):

When using POLS, consistent estimation is likely to be violated due to a correlation of the regrossors and the fixed effects. If e.g. agricultural firms in the warm south are more productive or are more attractive to employees, southern firms might be relatively more labor intensive compared to their northern peers. Location would then be a time-constant fixed effect correlated with the regressors, violating the consistency of OLS. Given the rejection of consistency, efficiency of POLS does not have to be considered. In contrast to POLS, the FE estimator allows for correlation between fixed effects and regressors ($E[c_i x_i] \neq 0$). Instead, by assuming time constant fixed effects, within transformation (de-meaning) can be used to eliminate individual fixed effects and thus restores exogeneity and consistent estimation.

Random Effects (RE) vs. Fixed Effects (FE) :

If fixed effects are both constant over time and random, a RE model would be efficient and consistent under the assumption of strict exogeneity. In general, POLS, FE, and FD are all consistent in the RE setting, however, if the fixed effects are assumed to be random and with a non zero variation, the RE model is preferable due to efficiency gains. This is because it includes both between and within variation by not requiring data transformations and still capturing variations in c_i. In the given example, the RE model should be used in case total factor productivity (TFP) varies randomly between firms and is uncorrelated with the level of capital and labor. The later appears to be unlikely given sectoral or regional factors potentially affecting labor and capital input. The Hausmann test is a way to test whether the RE or FE estimator should be preferred. The null hypotheses for the Hausmann test is $H_0 : E[x_{it}c_i] = 0$, which is the consistency assumption of the RE model. The test statistic uses the efficiency assumption of the RE estimate to simplify the expression. This means we test the consistency and efficiency assumptions of the RE against the FE model. The test statistic is given by:

$$H = (\hat{\gamma}_{RE} - \hat{\gamma}_{FE})'[V(\hat{\gamma}_{FE}) - V(\hat{\gamma}_{RE})]^{-1}(\hat{\gamma}_{RE} - \hat{\gamma}_{FE}) \sim \chi^2(2)$$

By running this test we receive a test statistic of 131.3159 with a p-value of 0. This leads us to the conclusion that the null hypothesis can be rejected. Rejecting the Hausmann test means the assumptions of RE do not hold, which would lead to an inconsistent estimate under RE. Hence, we prefer the FE model over the RE model.

First Differences (FD) vs. Fixed Effects (FE)

Another approach allowing for correlation between fixed effect and regressor is the FD estimator which also requires time-invariant fixed effects. The first-difference transformation is used to eliminate the fixed effects. As sequential exogeneity is implied by strict exogeneity and unlikely to be empirically applicable, we discuss strict exogeneity as the assumption for consistency of the FD model. To decide between the FE and FD estimator we consider the efficiency assumptions, as both are consistent in the same setting. Efficiency under FD is ensured if individual error terms are IID and serially correlated over time, meaning that error terms are following a random walk. E.g. time-varying shocks to the economy might happen repeatedly only to a subset of firms, which would lad under a unit root to different levels of subsequent ϵ_t's across firms. Intuitively this is the case if e.g. an innovative firm is more likely to be innovative in the future, serial correlation of individual error terms would be implied. This appears to be intuitive in the given example. To decide between the FE and the FD model, we conduct the test for serially correlated error terms. This test only yields valid results in case of strict exogeneity. The test is done by regressing the FD error terms on the lagged FD error terms:

$$\Delta\hat{\epsilon_{it}} = \rho\Delta\hat{\epsilon_{it-1}} + error_{it} \qquad , \quad t = 3, ..., T, i = 1, ..., N$$

We define two different null hypotheses for this test. As shown in the exercise class the efficiency assumption of FE, that error terms are uncorrelated over time, implies the null hypothesis is: $H_0 : \rho = -0.5,$. The efficiency of the FD model relies on error terms following a unit root, which leads to the null hypothesis: $H_0 : \rho = 0,$. Running the regression tests the assumption $H_0 : \rho = 0,$, which yields an estimate of $\rho - 0.2015$ and a t-value of -14.332, as shown in Table 3. We conclude that the null hypothesis of a unit root can be rejected. Next we conduct a WALD-test to test the assumption of no serial correlation of error terms: $H_0 : R\gamma = r,\ R = (1),\ r = -0.5$. The WALD test statistic is then given by: $W = (R\hat{\gamma}_{FD} - r)'(R\hat{V}_{FD}R')^{-1}(R\hat{\gamma}_{FD} - r)$, with $\hat{V}_{FD}$ being the variance-covariance matrix of the first difference estimator. We receive a WALD-statistic of 451.1252 and a p-value of 0. The null hypothesis of no serial correlation can also be rejected. This leads us to the conclusion that the efficiency assumption of neither the FD nor the FE model seem to hold. We rely on our intuitive argument, that the efficiency assumptions of the FD model are more realistic to hold. The corresponding results of the discussed estimation methods are given in Table 2. Column (1) and (2) compare the POLS results for the IID standard errors and panel robust standard errors. It is observable that the t-values decrease strongly when introducing robust standard errors. Furthermore, the prediction error of the model might be larger for high performing firms as other factors might play a role for their success, such as market power due to a lack of competition. Therefore, we include panel robust standard errors in all further models, as we are not confident to assume homoskedasticity. As discussed above, we conclude the FD estimation to be the most likely to derive valid results. Accordingly, the effect of log capital and log labour input on log sales amount to 0.5535 and 0.061 respectively.

(2)

In order to test for constant returns to scale, we test the null hypothesis of $\alpha + \beta = 1$ by using a Wald test. The underlying assumption of the Wald test is $\hat{\gamma}_{FD} \sim N(\gamma, var(\gamma))$, where $\hat{\gamma}_{FD}$ corresponds to our FD parameter estimations of β and α and their variance (V). This implies consistency of our parameter estimations, such that $\hat{\beta}$ and $\hat{\alpha}$ converge to the true parameter values. As we are argued that our preferred specification is the first differences model, consistency requires strict exogeneity and the full rank condition to be valid. This assumption requires the model to be correctly specified, which rules out omitted variables and reverse causality[1]. Furthermore, the Wald test requires a normally distributed parameter estimator which is fulfilled by the asymptotic characteristics of the estimator, which is asymptotically normal distributed due to its average properties. As the FD estimator does not report a constant, we can define the R matrix to be one for the first and second regressor. The WALD test would then be given by the null hypothesis: $H_0 : R\gamma = r,\ R = (1, 1),\ r = 1$ The WALD test statistic is then given by:

$$W = (R\hat{\gamma}_{FD} - r)'(R\hat{V}_{FD}R')^{-1}(R\hat{\gamma}_{FD} - r)$$

[1]Specific assumptions for the Wald test to be valid are discussed in Section (3)

with $\hat{V}_{FD}$ being the variance-covariance matrix of the FD estimator and $\hat{\gamma}_{FD}$ the FD coefficients. We estimate a WALD test statistic of 287 for the FD set-up. The p-value is given by zero. Table 4 reports the WALD test statistics for all considered models. Under any reasonable decision rule, we would reject the null hypothesis that there are constant returns to scale.

(3)

In 2.2 we argued that consistent coefficient estimates are required for valid interpretation of the WALD test. Consistency of the FD model is based on strict exogeneity, which assumes that regressors are uncorrelated to the error term in every period. Table 5 provides a test for this crucial assumption, by regressing the outcome of interest on the inputs as well as their leads. Under strict exogeneity we would predict that the leads of the inputs should not affect the outcome. We see that labor in t+1 has a significant effect at the 1%level and that capital in t+1 is significant at the 10% level. This is counterintuitive and suggests that strict exogeneity for employment cannot be assumed and is questionable for capital. Violation of the identification assumption results from predetermined regressors. Hence, the inputs in period t are correlated with the error term in period t-1. This might occur if not only inputs affect sales, but also if sales affect the level of chosen inputs. Intuitively, a positive correlation between sales and improved financial ability which in response enables an increased usage of capital and labor input could be an explanation. Such a relationship would suggest a correlation between regressors and the error term in previous periods. As our data includes yearly observations, it might be that regressors in t are also correlated with the error term in t, as economic shocks can materialize in changed input in the same year. The econometric consequence is that the regressors are a function of past and present estimation errors such that the exogeneity assumption $E[\epsilon_{it}|x_{it}] = 0 \ \forall t$ underlying the FD estimation is violated. Consequently, we obtain inconsistent estimates of β which violates the assumptions of the WALD test. Thus, our conclusion of rejecting constant returns to scale in (2) would not be valid. Instrumental variables (IV) are our main suggestion to restore exogeneity. By predicting our endogeneous regressors with valid IVs, we could obtain consistent estimation and restore validity of results in (2). If contemporaneous exogeneity holds for predetermined regressors, one can restore consistency using lagged regressors as instruments. This would in practice mean to use $(x_{it-1}, x_{it-2}..., x_{i1})$ as instruments for the regressor Δx_{it}. The underlying assumption for this approach to yield consistent estimates is $E[\epsilon_{it}|x_{i1}, x_{i2}, ..., x_{it}, c_i] = 0$. Nevertheless, as argued before we do not assume contemporaneous exogeneity to hold. To ensure that the use of lagged regressors as IVs is valid, we would propose to use monthly data for this approach. By using monthly data, it is unlikely that economic shocks in one month would change inputs in the same month due to the time it requires to hire individuals or acquire capital. This would fulfill contemporaneous exogeneity and lagged regressors could be used as instruments to ensure a consistent estimation.

Table 2: Regression Results

Y = log sales	POLS	POLS	RE	FE	FD
Constant	0.00	0.00	0.00		
	(0.00)	(0.00)	(0.00)		
LEMP	0.678***	0.678***	0.726***	0.702***	0.554***
	(69.72)	(18.82)	(21.61)	(50.54)	(20.48)
LCAP	0.304***	0.304***	0.186***	0.146***	0.061***
	(34.90)	(9.57)	(7.19)	(12.21)	(2.79)
R^2	0.91	0.91	0.65	0.50	0.17
Robust standard errors	-	✓	✓	✓	✓

t values in parenthesis and $p < 0.1^*$, $p < 0.05^{**}$, $p < 0.01^{***}$.

Table 3: Test for serially correlated error terms

Y = FD residuals	
e(it-1)	-0.2015***
	(-14.33)
R^2	0.04

t values in parenthesis and
$p < 0.1^*$, $p < 0.05^{**}$, $p < 0.01^{***}$.

Table 4: WALD Test Results

	POLS	RE	FE	FD
WALD statistic	19.62	98.76	178.39	287.41
$\Pr(\alpha + \beta = 1)$	$9.4452e^{-06}$	0	0	0

Table 5: Test for strict exogeneity

Y = log sales	FD
LEMP	0.555^{***}
	(31.87)
LCAP	0.063^{***}
	(3.54)
LEMP(t+1)	0.031^{***}
	(3.25)
LCAP(t+1)	-0.015^{*}
	(-1.74)
R^2	0.17

t values in parenthesis and
$p < 0.1^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$.