TA Notes : Econometric Theory II

Paúl Corcuera

January, 2021

This notebook is indented to different topics that will be useful for ECON627 at VSE.

Contents

1 Linear Algebra

1.1 Basic Definitions

Definition 1.1 (Vector Space) *A vector space over a field F is a set V together with two operations:*

1. *Addition:* $+ : V \times V \to V$
2. *Scalar Multiplication:* $* : F \times V \to V$.

We refer to the elements $v \in V$ as vectors, and the elements of $\alpha \in F$ as scalars.

Definition 1.2 (Linear Transformation) *A linear transformation between two vector spaces V and W is a mapping $T : V \times W$ such that :*

1. $T(v_1 + v_2) = T(v_1) + T(v_2)$
2. $T(\alpha v) = \alpha T(v)$ *for any scalar* $\alpha \in F$.

We can represent linear transformation as matrices. Let

$$A = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ a_{21} & \dots & a_{2k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix}.$$

Then A maps from $\mathbb{R}^k$ to $\mathbb{R}^n$. The number of columns of a matrix encode the number of basis vectors of the input space. More specifically, the columns of matrix A show where the **standard** basis vectors of $\mathbb{R}^k$ end up after the transformation.

To see this, consider a transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. The linear transformation, and the standard basis vectors are given by

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \hat{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \hat{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

You can confirm that the columns of A correspond to the output vectors of the linear transformation onto the basis vectors.

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \cdot \hat{i} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \mathrm{col}_1(A)$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \cdot \hat{j} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \mathrm{col}_2(A)$$

Furthermore, linearity implies that

$$A \cdot \begin{bmatrix} 5 \\ 3 \end{bmatrix} = A(5 \cdot \hat{i} + 3 \cdot \hat{i}) = 5A \cdot \hat{i} + 3A \cdot \hat{j} = \begin{bmatrix} 5 \cdot 1 + 3 \cdot 1 \\ 5 \cdot 1 + 3 \cdot 2 \end{bmatrix}$$

We can conclude that, in general, multiplication AB consist on "projecting" the rows of A onto every column of B.

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, B = \begin{bmatrix} b_1 & \dots & b_k \end{bmatrix}$$

$$AB = \begin{bmatrix} a_1 \cdot b_1 & \dots & a_1 \cdot b_k \\ \vdots & \ddots & \vdots \\ a_n \cdot b_1 & \dots & a_n \cdot b_k \end{bmatrix}$$

The operations (dot products) inside the matrix AB must be well defined, and this requires that the rows vectors of A have the same size of the column vectors of B.

1.2 Dot Products & Projections

Economists often study settings where units possess two or more group memberships, some of which can change over time. A prominent example comes from **?** (henceforth AKM) who proposed a panel model of log wage determination that is additive in worker and firm fixed effects.

This so-called "two-way" fixed effects or "AKM" model takes the form:

$$y_{gt} = \alpha_g + \psi_{j(g,t)} + w'_{gt}\delta + \varepsilon_{gt} \qquad (g = 1, \dots, N,\ t = 1, \dots, T_g \geq 2) \tag{1}$$

where the function $j(\cdot,\cdot) : \{1, \dots, N\} \times \{1, \dots, \max_i T_g\} \to \{1, \dots, J\}$ assigns a worker g and year t observation to one of J firms. Here α_g is a person effect, $\psi_{j(g,t)}$ is a firm effect, w_{gt} is a time-varying covariate, and ε_{gt} is a mean-independent time-varying error.

We can rewrite the original AKM model as:

$$y_i = x'_i\beta + \varepsilon_i \qquad i = 1, ..., n \tag{2}$$

where i indexes a particular person-year observation (g, t), x_i is a vector that collects all the worker, firm dummies as well as the time-varying covariates w_{gt} so that $\beta = (\alpha, \psi, \delta)'$ is a $k \times 1$ vector that collects all the worker, firm fixed effects along with δ.

Interest in AKM models often centers on understanding how much of the variability in log wages is attributable to firms. It is common to summarize the firm contribution to wage inequality using the following two parameters:

$$\sigma^2_\psi = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\left(\psi_{j(g,t)} - \bar{\psi}\right)^2 \qquad \text{and } \sigma_{\alpha,\psi} = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\left(\psi_{j(g,t)} - \bar{\psi}\right)\alpha_g \tag{3}$$

where $\bar{\psi} = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\psi_{j(g,t)}$. The variance component σ^2_ψ measures the contribution of firm wage variability to inequality, while the covariance component $\sigma_{\alpha,\psi}$ measures the additional contribution of systematic sorting of high wage workers to high wage firms.

The function `leave_out_KSS` provides unbiased estimates of σ^2_ψ and $\sigma_{\alpha,\psi}$ as well as an estimate of $\sigma^2_\alpha = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\left(\alpha_g - \bar{\alpha}\right)^2$ using the leave-out bias correction approach proposed by KSS.

1.3 The KSS Correction

We now provide some general intuition about the KSS leave-out methodology. A more formal discussion can be found in KSS.

1.3.1 The Plug-in Estimator

Suppose the researcher is interested in the variance of firm effects, σ^2_ψ. To simplify the exposition, we normalize the firm effects so that their firm-size weighted mean is equal to zero, i.e. $\bar{\psi} = 0$ and re-write σ^2_ψ as

$$\sigma^2_\psi = \sum_{j=1}^{J} s_j \psi_j^2 \tag{4}$$

where s_j gives the employment share of firm j, $s_j = \dfrac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T_g}\mathbf{1}\{j(g,t)=j\}$.

It is customary to report "plug-in" estimates of a given variance component using the corresponding OLS estimate. For instance, the plug-in estimate of the variance of firm effects σ^2_ψ is given by

$$\tilde{\sigma}^2_\psi = \sum_{j=1}^{J} s_j \hat{\psi}_j^2 \tag{5}$$

where $\hat{\psi}_j$ is the OLS estimate obtained after estimating equation (1) via OLS.

1.3.2 The Bias in the Plug-in Estimator

The estimated firm effect, $\hat{\psi}_j$, represents a noisy estimate of the true firm effect, ψ_j. The presence of noise in $\hat{\psi}_j$ is not an issue when one is interested in ψ_j as the OLS estimator $\hat{\psi}_j$ is assumed to be unbiased in this context, i.e. $E[\hat{\psi}_j] = \psi_j$.

However, estimation error in $\hat{\psi}_j$ is going to lead to biases when estimating ψ_j^2 using its "plug-in" analogue $\hat{\psi}_j^2$ since:

$$E[\hat{\psi}_j^2] = E[(\hat{\psi}_j - \psi_j + \psi_j)^2] = \psi^2 + \underbrace{\mathbb{V}[\hat{\psi}_j]}_{\text{Bias}} \tag{6}$$

where $\mathbb{V}[\hat{\psi}_j]$ is the squared standard error of $\hat{\psi}_j$. Intuitively, when we take the square of $\hat{\psi}_j$ we are not only squaring its signal, ψ_j, but also the estimation error in each $\hat{\psi}_j$. The latter is going to introduce a bias when estimating ψ_j^2.

The same logic applies when analyzing the bias of the plug-in estimator of the variance of firm effects since

$$E[\tilde{\sigma}^2_\psi] = \sigma^2_\psi + \underbrace{\sum_{j=1}^{J} s_j \mathbb{V}[\hat{\psi}_j]}_{\text{Bias}} \tag{7}$$

1.3.3 The Problem with "Standard" Standard Errors in High Dimensional Models

The formula above shows that the bias of the plug-in estimator of the variance of firm effects is

$$E[\tilde{\sigma}^2_\psi] - \sigma^2_\psi = \sum_{j=1}^{J} s_j \mathbb{V}[\hat{\psi}_j]. \tag{8}$$

Therefore, all that is required for a bias correction is an estimate of the (squared) standard error of each firm effect, $\mathbb{V}[\hat{\psi}_j]$. Similarly, if our interest is on the variance of person effects, then we would need the standard error on each of the person effects, $\mathbb{V}[\hat{\alpha}_i]$. If we are interested in the covariance of worker and firm effects, then we would need the covariances in sampling errors between each $\hat{\alpha}_i$ and $\hat{\psi}_{j(i,t)}$.

The discussion above highlights that an estimate of the sampling variability of the OLS coefficient vector $\hat{\beta} = (\hat{\alpha}, \hat{\psi}, \hat{\delta})$ is required in order to derive an unbiased estimate of the variance components of the AKM model displayed in equation (2).

Recall that the sampling variability in $\hat{\beta}$, assuming independence across observations, is given by

$$\mathbb{V}[\hat{\beta}] = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} \sigma_i^2 x_i x_i' \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}. \tag{9}$$

where $\sigma_i^2 = Var(\varepsilon_i)$.

One might be tempted to provide an estimate of $\mathbb{V}[\hat{\beta}]$ using heteroskedasticity *consistent* ("HC" or robust), standard errors. Standard **?** HC standard-errors are calculated using a plug-in estimate of σ_i^2 based on

$$\tilde{\sigma}_i^2 = (y_i - x_i'\hat{\beta})^2. \tag{10}$$

so that the HC based estimate of $\mathbb{V}[\hat{\beta}]$ is given by

$$\tilde{\mathbb{V}}[\hat{\beta}] = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} \tilde{\sigma}_i^2 x_i x_i' \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}. \tag{11}$$

However, HC standard errors based on $\tilde{\sigma}_i^2$ are downward biased (**?**). From an asymptotic perspective, HC standard errors are inconsistent in any high dimensional model where the number of parameters grows in proportion to the sample size (**?**). Such "many regressor" asymptotics are natural in the worker-firm fixed effects model as we often have fewer than 5 worker moves on average per firm.

1.3.4 The Leave Out Correction

KSS provide a heteroskedasticity-*unbiased* (HU) estimate of the standard error of any coefficient obtained from a linear regression model.

The KSS HU standard error estimate is based on a leave-out estimate of σ_i^2:

$$\hat{\sigma}_i^2 = y_i(y_i - x_i'\hat{\beta}_{-i}). \tag{12}$$

where $\hat{\beta}_{-i}$ is the OLS estimate of β from equation (2) after leaving observation i out.

KSS then replaces σ_i^2 in $\mathbb{V}[\hat{\beta}]$ with its unbiased estimate $\hat{\sigma}_i^2$ to derive a HU estimate of $\mathbb{V}[\hat{\beta}]$:

$$\hat{\mathbb{V}}[\hat{\beta}] = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} \hat{\sigma}_i^2 x_i x_i' \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}. \tag{13}$$

Going back to the example of the variance of the firm effects, we can extract from $\hat{\mathbb{V}}[\hat{\beta}]$ the corresponding squared standard error of each firm effect, $\hat{\mathbb{V}}[\hat{\psi}_j]$. We can then use the latter to bias correct the corresponding estimate of the variance of the firm effects as follows:

$$\hat{\sigma}_\psi^2 = \tilde{\sigma}_\psi^2 - \sum_{j=1}^{J} s_j \mathbb{V}[\hat{\psi}_j]. \tag{14}$$

The MATLAB function `leave_out_KSS` is going to print the bias corrected variance of firm effects, $\hat{\sigma}_\psi^2$, as well as the bias-corrected covariance of worker and firm effects and variance of person effects. `leave_out_KSS` also provide the correct standard errors — based on $\hat{\mathbb{V}}[\hat{\beta}]$ as opposed to $\tilde{\mathbb{V}}[\hat{\beta}]$ — when regressing the firm effects on some observable characteristics, see Section 9.

2 Computing the KSS Correction

We now demonstrate how one can implement the KSS correction in two-way models using MATLAB and the function `leave_out_KSS`. We continue to work with a simple example based on an AKM model. In what follows, we use the words "workers" and "firms" when describing the procedure for simplicity but the the function `leave_out_KSS` can be applied to any two-way fixed effects model (e.g. patients and doctors, students and teachers, strata and treatment arms).

2.1 Setup

We begin with some auxiliary lines of code that define the relevant paths, call the CMG package package developed by Yiannis Koutis and set-up the parallel environment within MATLAB.

2.2 Importing the Data

The Github Repo contains a matched employer-employee testing data where we observe the identity of the worker, the identity of the firm employing a given worker, the year in which the match is observed (either 1999 or 2001) and the associated log wage.

Important!: the original data must be sorted by individual identifiers (id). For instance, one can see that the testing data is sorted by individual identifiers (and year, using `xtset id year` in Stata)

2.3 Calling the Main Function

The function `leave_out_KSS` relies on three mandatory inputs: `(y,id,firmid)`. We can obtain an unbiased variance decomposition of the associated AKM model by simply calling

3 Interpreting the Output

The code starts by printing its two key inputs: the algorithm used to compute the statistical leverages (exact vs. JLA) — we explain this distinction in Section 5 — and the level at which the leave-out correction is carried (observation vs. match) — we explain this in more detail in Section 7.

The output printed by `leave_out_KSS` is composed by three sections.

Section 1: Here we provide info on leave-out connected set. This is the largest connected set of firms that remains connected after removing any worker from the associated graph, see Lemma 1 and the Computational Appendix of KSS for details. The code provides some summary statistics (e.g. # of movers, # of firms, mean and variance of the outcome, etc) for the leave-out connected set.

Section 2: After printing the summary statistics, the code computes the statistical leverages of the design, denoted as P_{ii}. Computation of $\{P_{ii}\}_{i=1}^{n}$ represents the main computational bottleneck of the routine.

Section 3: The code then enters its third, and final, stage where the main results are printed. The code starts by reporting the — biased — estimates of the variance components that result from the "plug-in" approach of treating OLS estimates as measured without error. Finally, the code prints the bias corrected variance of firm effects and the covariance of worker and firm effects.

4 What Does the Code Save?

`leave_out_KSS` saves three scalars: the variance of firm effects (`sigma2_psi` in [4]), the covariance of worker and firm effects (`sigma_psi_alpha`) and the variance of person effects (`sigma2_alpha`).

`leave_out_KSS` also saves on disk one .csv file. This .csv contains information on the leave-out connected set. This file has 4 columns. First column reports the outcome variable, second and third columns the worker and the firm identifiers (as originally inputted by the user). The fourth column reports the statistical leverages of the regression design. If the code is reporting a leave-out correction at the match-level, the .csv will be collapsed at the match level. By default, the .csv file is going to be saved in the main directory under the name `leave_out_estimates`. The user can specify an alternative path using the option `filename` when calling `leave_out_KSS`.

5 Scaling to Large Datasets

`leave_out_KSS` can be used on extremely large datasets. The code uses a variant of the random projection method, denoted as the Johnson– Lindenstrauss Approximation (JLA henceforth) algorithm in KSS for its connection to the work of **?**, see also **?**. We now discuss briefly the main computational bottleneck of the procedure and the JLA algorithm.

5.1 Computational Bottleneck

Recall from the discussion in Section 1 that the KSS leave-out bias correction procedure relies on leave-out estimates of σ_i^2

$$\hat{\sigma}_i^2 = y_i(y_i - x_i'\hat{\beta}_{-i}) \tag{15}$$

where $\hat{\beta}_{-i}$ is the OLS estimate of β from the AKM model in equation (2) after leaving observation i out.

Clearly, re-estimating $\hat{\beta}_{-i}$ by leaving a particular observation i for n times, is infeasible computationally. Fortunately, one can re-write $\hat{\sigma}_i^2$ as

$$\hat{\sigma}_i^2 = y_i \frac{(y_i - x_i'\hat{\beta})}{1 - P_{ii}} \tag{16}$$

where P_{ii} measures the influence or leverage of observation i, i.e. $P_{ii} = x_i'S_{xx}^{-1}x_i$. The expression above highlights that all that is needed for computation of $\hat{\sigma}_i^2$ are the n statistical leverages $\{P_{ii}\}_{i=1}^n$. However, exact computation of P_{ii} may remain prohibitive when n is in order of tens of millions or higher.

5.2 Approximating the Statistical Leverages

The JLA algorithm introduced by KSS provides a stochastic approximation to $\{P_{ii}\}_{i=1}^n$ using the random projection ideas developed by Johnson and Lindenstrauss (1984). We defer the reader to the Computational Appendix of KSS for further details.

The number of simulations underlying the JLA algorithm is governed by the input `simulations_JLA` (which is denoted as p in the computational appendix). Intuitively, more simulations imply a higher accuracy – but higher computation time — when estimating $\{P_{ii}, B_{ii}\}_{i=1}^n$.

Note: The user might want to pre-specify a random number generator seed to ensure replicability when calling the function `leave_out_KSS`.

We now demonstrate the performance of the code on a large dataset

We can see from the output that the leave-out connected set has almost 14 million person-year observations. The code is able to complete in around 4 minutes (on a 2020 Macbook pro with 6 cores and 16GB of RAM).

The computational appendix in KSS shows that the JLA algorithm can cut computation time by a factor of 100 while introducing an approximation error of roughly 10^{-4}.

The current code uses an improved estimator of both P_{ii} and $M_{ii} = 1 - P_{ii}$ which are both guaranteed to lie in $[0, 1]$. These improved estimators are then combined to derive an asymptotically unbiased JLA estimator of a given variance component provided that $\frac{n}{p^4} = o(1)$, see this document.

One can check the stability of the estimates for different values of `simulations_JLA`. For instance, if we double `simulations_JLA` from 50 to 100 and run the code again on the same data:

We obtain virtually the same variance components as when `simulations_JLA`=50 while significantly increasing the computational time! If the user does not specify a value for `simulations_JLA`, the code defaults to `simulations_JLA`=200.

We conclude this section by noting that the user can also calculate an exact version of $\{P_{ii}\}_{i=1}^{n}$. This can be done by setting the option `type_of_algorithm` to `exact`.

Warning: Calling the option `exact` in large datasets can be very time consuming! We now load again the original, smaller, testing data and then compare the exact and JLA based estimates of the variance components

The variance components estimated via JLA are extremely close to the `exact` estimates but only take a fraction of the time to compute. If the input data has more than 10,000 obs, the code defaults to using the JLA algorithm unless the user specifies type_of_algorithm to "exact".

6 Adding Controls

We have demonstrated the functioning of `leave_out_KSS` using a simple AKM model with no controls ($w_{gt} = 0$). It is easy to add a matrix of controls to the routine. Suppose for instance that we want to add year fixed effects to the original AKM model. This can be done as follows

When controls are specified, the code proceeds by partialling them out. That is, it first estimates by OLS the AKM model in the leave-out connected set

$$y_{gt} = \alpha_g + \psi_{j(g,t)} + w_{gt}'\delta + \varepsilon_{gt} \tag{17}$$

from which we obtain $\hat{\delta}$. We then work with a residualized model where the outcome variable is now defined as $y_{gt}^{new} = y_{gt} - w_{gt}'\hat{\delta}$ and project this residualized outcome on worker and firm indicators and report the associated (bias-corrected) variance components.

7 Leaving out a Person-Year Observation vs. Leaving Out a Match

By default, the code reports leave-out corrections for the variance of firm effects and the covariance of firm and worker effects that are robust to unrestricted heteroskedasticity and serial correlation of the error term within a given match (unique combination of worker and firm identifier), see Remark 3 of KSS. Intuitively, leaving out matches is analogous to "clustering" the standard error estimates on match. We discuss the interpretation of the leave-out corrected variance of person effects when leaving a match out in Section 8.

The user can specify the function to run the KSS correction when leaving only an observation out using the option `leave_out_level`. When leaving a person-year observation out, the resulting

KSS variance components are robust to unrestricted heteroskedasticity but not serial correlation within match. Below we demonstrate how to compute KSS adjusted variance components when leaving a single (person-year) observation out.

When $T = 2$ (i.e the underlying matched employer-employee data spans only two years), as in this example, it turns out that the KSS adjusted variance of firm effects and covariance of firm and worker effects is robust to any arbitrary correlation between ε_{g2} and ε_{g1}.

8 Variance of Person Effects when Leaving Out a Match

By leaving a match-out, we can bias correct the variance of firm effects and covariance of worker and firm effects while allowing for unrestricted hetoreskedasticity and serial correlation of the error term ε_{gt} within each worker-firm match.

However, the person effects, α_g, of "stayers" — workers that never leave a particular firm — are not leave-match-out estimable. This implies that we cannot compute an unbiased estimate of $\Omega_g = Var(\varepsilon_{g1}, ..., \varepsilon_{gT_g})$ for stayers. An estimate of Ω_g for both stayers and movers is required in order to provide a bias correction for the variance of person effects, see Section 1 and Remark 3 in KSS.

The current implementation of the code estimates Ω_g for stayers by leaving only a single observation out – i.e., by assuming Ω_g is diagonal. This approach yields an upper bound estimate on the variance of person effects (computed across both stayers and movers).

There are several alternatives that the user can explore:

1. Estimate a variance decomposition in a sample of movers only: For movers, it is possible to estimate a leave-out bias corrected variance of person effects that is robust to both unrestricted heteroskedasticity and serial correlation in the error term of the AKM model within a given match. Therefore, one can provide an unbiased variance decomposition of all the three components of the two-way fixed effects model by simply feeding to the function `leave_out_KSS` a movers-only sample.

2. Drop adjacent wage observations for stayers: Under the assumption that the errors are serially independent after m periods, it suffices to keep every *mth* stayer observation and apply the leave person-year out estimator. For example, if $m = 2$ and we have a balanced panel with $T = 5$, we can restore independence of the errors in the stayer sample by keeping any of the following pairs of stayer time periods: (1,4), (2,5), (1,5). One can choose from the available pairs randomly for each stayer with equal probability.

3. Drop interior wage observations for stayers: To minimize concerns regarding serial correlation, the user can drop all but the first and last wage observation of each stayer. Note that dropping stayer wage observations reduces their weight in the variance components. Future versions of the code will allow the variance components to be defined in terms of weights other than the number of micro-observations.

9 Regressing Firm Fixed Effects on Observables

It is common in empirical applications to regress the fixed effects estimated from the two-way model on some observables characteristics. Using the AKM model again as our leading example, suppose we are interested in the linear projection of the firm effects ψ_{gt} on some observables Z_{gt}:

$$\psi_{j(g,t)} = Z'_{gt}\gamma + e_{gt} \tag{18}$$

Typical practice is to estimate γ using a simple regression where the estimated firm effects, $\hat{\psi}_{j(g,t)}$, are regressed on Z_{gt}

$$\hat{\gamma} = \left(\sum_{g,t} Z_{gt}Z'_{gt}.\right)^{-1} Z_{gt}\hat{\psi}_{gt} \tag{19}$$

KSS show that inference on $\hat{\gamma}$ needs to be adjusted because the estimated firm fixed effects $\{\hat{\psi}_j\}_{j=1}^{J}$ are correlated with one another.

To see this, suppose that we have a simple AKM model with only two time periods, set $w_{gt} = 0$, and take first differences $\Delta y_g \equiv y_{g2} - y_{g1}$ to eliminate the worker fixed effects so that the AKM model becomes

$$\Delta y_g = \Delta f'_g \psi + \varepsilon_g \tag{20}$$

where $\Delta f_g = f_{g,2} - f_{g,1}$ and $f_{gt} = \{\mathbf{1}_{j(g,t)=1}, .., \mathbf{1}_{j(g,t)=J}\}$ is the vector containing the firm dummies. In this model,

$$\hat{\psi} = \psi + \underbrace{\sum_{g=1}^{N} (\Delta f_g \Delta f'_g)^{-1} \Delta f_g \varepsilon_g}_{\text{Correlated Noise}} \tag{21}$$

Note how the dependence in the vector of estimated firm fixed effects, $\hat{\psi}$, is induced by the regressor design $\sum_{g=1}^{N}(\Delta f_g \Delta f'_g)^{-1}$. As shown in Table 3 of KSS, ignoring this correlation can easily lead standard errors to be underestimated by an order of magnitude in practice.

The package provides the HU standard errors on $\hat{\gamma}$ using the function `lincom_KSS` which is designed to emulate the Stata function lincom and therefore works as a post-estimation command. We demonstrate the functioning of `lincom_KSS` with an example.

In this example, we are interested in testing whether the difference in person-year weighted mean firm effects between region 1 and region 2 is statistically different from zero. This amounts to running a regression where the dependent variable is the vector of estimated firm effects and the set of observables, Z_{gt} , here is represented by a constant and a dummy for whether the firm of worker g in year t belongs to region 2.

The resulting coefficient (and standard error) can be computed by calling the function `leave_out_KSS` specifying that we want to run the `lincom` option and using the region dummy as Z_{gt} (the constant is automatically added by the code).

We can see from the output above (make sure to scroll until the end) that the difference in person-year weighted mean firm effects between the two regions is equal to 0.26. The traditional HC or "robust" standard errors on this coefficient is around 0.05 while the HU standard error derived in KSS is roughly twice as large (0.09).