Ioana VINAȘI, PhD Candidate (corresponding author)
ioana.ianos@econ.ubbcluj.ro
Babeș-Bolyai University, Cluj-Napoca, Romania

Does Norm Selection Matter in Iterative SUR Estimation?

Abstract. *This study explores how two alternative norms perform within the Iterative Seemingly Unrelated Regression (ISUR) framework, focusing on their efficiency and accuracy across diverse scenarios – such as varying time spans and country samples. By putting these norms to the test, the research offers hands-on guidance for researchers and practitioners seeking the most effective balance between computational speed and statistical precision. The Euclidean findings reveal which norm delivers the best trade-off, serving as a practical roadmap for optimising ISUR models in real-world applications. Ultimately, this comparison not only sharpens methodological decision-making but also enhances the reliability and efficiency of empirical workflows in econometric analysis.*

Keywords: *statistical modelling, ISUR framework, comparative performance, computational optimisation, accuracy, precision.*

JEL Classification: C15, C23, C33, C52.

Received: 28 May 2025	Revised: 24 December 2025	Accepted: 4 March 2026

1. Introduction

This study fills an important methodological gap in econometric methodology by putting Seemingly Unrelated Regression (SUR) and its iterative counterpart (ISUR) under the microscope. Using Monte Carlo simulations, we explore how these estimators perform across a range of time series and cross-sectional dimensions – an area where existing literature offers limited insight. By rigorously testing their efficiency and precision, the research provides actionable guidance for applied economists aiming to streamline computation and boost model reliability. The results not only inform better estimator selection but also contribute to more robust and scalable econometric modelling for complex datasets.

The SUR model (Zellner, 1962) estimates multiple interrelated equations by leveraging error correlations, making it applicable in contexts like regional, industrial, or household analyses where shared shocks affect multiple entities. For example, demand functions for a product across households may exhibit correlated errors due to shared economic factors, such as income shocks. Each equation functions as an independent linear regression, justifying the term "seemingly unrelated," though "seemingly related" might better reflect the correlated errors (Zellner, 1962; Srivastava & Giles, 1987).

SUR enhances estimation efficiency by pooling information across equations and facilitates testing of cross-equation parameter restrictions (Zellner, 1962;

DOI: 10.24818/18423264/60.1.26.11

Greene, 2012). The model accommodates multiple equations, each with distinct dependent and explanatory variables. While Ordinary Least Squares (OLS) can yield consistent results, it is generally less efficient than SUR, which applies Feasible Generalised Least Squares (FGLS) using a variance-covariance matrix (Greene, 2012). SUR and OLS produce identical estimates only when errors are uncorrelated or explanatory variables are identical across equations (Davidson & MacKinnon, 1993).

The SUR model can be seen as either a restricted general linear model (Mardia et al., 2024) or a broader framework allowing for distinct regressors across equations. It can also extend to simultaneous equations by including endogenous variables as regressors (Hayashi, 2000).

In this paper, we evaluate the FGLS-based non-iterative model and compare convergence speeds for two iterative SUR norms. Using Python, we implemented two iterative SUR procedures tailored to these norms. The study begins with a review of the theoretical foundations of FGLS and iterative SUR, followed by practical Python implementation. No previous econometric study has compared Euclidean vs. Absolute norm convergence within ISUR. Practitioners often face slow or unstable convergence, yet lack guidance on which norm improves performance.

While SUR and ISUR are widely used, the literature does not examine whether the choice of convergence norm affects estimator accuracy, iteration counts, or computational time. This paper is, to our knowledge, the first systematic comparison of Euclidean vs. Absolute norms within ISUR estimation, using a controlled Monte Carlo environment and reproducible Python implementation. To enhance external validity, we also provide a brief empirical illustration using real-world macroeconomic indicators.

2. Literature review

Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) are common regression methods, but their efficiency declines when error terms exhibit correlation or heteroscedasticity. In such cases, the assumptions underlying these methods fail, potentially causing inefficiencies. Generalised Least Squares (GLS), introduced by Aitken (1936), addresses these issues by accounting for error structure, making it a robust tool in linear regression.

GLS minimises a weighted residual sum of squares using the inverse covariance matrix (Ω), producing unbiased and efficient estimators. When Ω is unknown, Feasible Generalised Least Squares (FGLS) uses a two-step process: OLS to estimate residuals and Ω, followed by GLS with the estimated covariance matrix (Kiviet, 1995). FGLS is asymptotically efficient, but small samples may reduce its accuracy due to variability in estimating Ω (Greene, 2003).

GLS and FGLS are particularly useful in panel data, where cross-sectional and autocorrelations complicate the estimation. Their relevance depends on the dataset structure: large cross-sections (N) emphasise cross-sectional dependence, while long time series (T) highlight autocorrelation (Baltagi, 2008). Robust methods like the

Newey-West or White-Eicker estimators adjust OLS standard errors for heteroscedasticity and autocorrelation without altering coefficients (Newey & West, 1987; White, 1980).

Iterative FGLS further refines estimates by updating Ω and re-estimating coefficients until convergence. Although this improves accuracy, it is computationally intensive and less efficient for small samples (Hansen, 2007). Recent advancements in Seemingly Unrelated Regression (SUR) models address computational challenges, robust estimators, structural changes, and high-dimensional data. For instance, Kontoghiorghes (1998) proposed efficient algorithms for large datasets, while authors (Youssef et al, 2022) demonstrated the effectiveness of robust estimators in reducing bias. Structural breaks in SUR models are addressed through statistical tests (Saraceno et al., 2023), and sparsity in high-dimensional data is tackled with the FGLasso estimator (Tan et al., 2018).

In conclusion, GLS and FGLS are powerful tools for managing heteroscedasticity and autocorrelation, particularly in large samples. Robust methods such as HAC or White-Eicker are better suited for smaller datasets. The choice of method depends on the data structure and the balance between efficiency, consistency, and computational complexity (Baltagi, 2008; Greene, 2003).

3. Model specification

In statistics, Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) can become inefficient or produce biased results when residuals in linear regression models exhibit correlation. In such conditions, the Generalised Least Squares (GLS) method, introduced by Aitken in 1936, provides a more robust estimation technique for models with correlated errors (Aitken, 1936).

In a typical linear regression framework, we observe a dataset $\{ \mathrm{y}_i, \mathrm{x}_{ij} \}_{i=1,\dots,N}^{j=2,\dots,k}$ for N statistical units (e.g., individuals, firms, or other entities), across $t = 1, \dots, T$ time periods. The response values are organised in a vector $y = (\mathrm{y}_i, \dots, \mathrm{y}_n)\mathrm{Transp}$, and the predictor values are placed in the design matrix:

$$x = (x_1^{Transp}, \dots, x_n^{Transp})^{Transp},$$

where $\mathrm{x}_i = (1, \mathrm{x}_{i2}, \dots, \mathrm{x}_{ik})$ is a vector of k predictor variables (including a constant) for the i-th unit.

The model assumes that the conditional mean of y given x is a linear function of x and that the conditional variance of the error term given x is a known nonsingular covariance matrix Ω. This is typically written as:

$$y = x\beta + \varepsilon, \mathrm{E}[\varepsilon|\mathrm{X}] = 0, \mathrm{cov}[\varepsilon|\mathrm{X}] = \Omega,$$

where $\beta \in \{\mathrm{Rk}\}$ is a vector containing the unknown regression coefficients to be estimated, ε represents the error term and Ω is the covariance matrix of the errors.

Let $\hat{\beta}$ be a possible estimate for β. Then, the residual vector for $\hat{\beta}$ will be $y - x\hat{\beta}$. The Generalised Least Squares (GLS) method estimates $\hat{\beta}$ by minimising the Mahalanobis squared length of this residual vector:

$$(\mathrm{y} - \mathrm{X}\hat{\beta})^{\mathrm{Transp}}\Omega^{-1}(\mathrm{y} - \mathrm{X}\hat{\beta}).$$

Since the objective is a quadratic form in $\hat{\beta}$, the estimator has an explicit formula:

$$\hat{\beta} = (\mathrm{X}^{\mathrm{Transp}}\Omega^{-1}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{Transp}}\Omega^{-1}\mathrm{y}.$$

This formula represents the GLS estimator, which accounts for the correlation structure of the errors, making it more efficient than the OLS when the errors are correlated or heteroscedastic.

If the error covariance Ω is unknown, a consistent estimate of it, say $\widehat{\Omega}$, can be obtained using an implementable version of GLS known as feasible least squares (FGLS) estimation. In FGLS, the modelling is done in two steps: the model is estimated by OLS or another consistent (but inefficient, according to Kiviet, 1995) estimator, and the residuals are used to construct a consistent estimator of the error covariance matrix (to do this, it is often necessary to examine the model by adding constraints additional, for example, if the errors follow a time series process, a statistician generally needs some theoretical assumptions about this process to ensure that a consistent estimator is available); using the consistent estimator of the error covariance matrix, GLS ideas can be implemented.

In panel data with N individuals and T time periods: the covariance matrix Ω accounts for both cross-sectional correlations (e.g., between individuals at the same time) and autocorrelation (e.g., correlations within an individual over time). When N is large and T is fixed, cross-sectional dependance dominates, but when T is large and N is fixed, time series properties of the errors (e.g., autocorrelation) become more important. An example equation would have the following form:

$$Y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \varepsilon_{it},$$

where $\mathrm{i} = 1, \ldots, N$, indexes individuals, $t = 1, \ldots, T$ indexes time periods and ε_{it} captures heteroscedasticity or autocorrelation.

While GLS is more efficient than OLS under the assumptions of heteroscedasticity or autocorrelation, this is not the case for FGLS. The feasible estimator is, provided that the covariance matrix of the errors is consistently estimated, asymptotically more efficient, but for small- or medium-sized samples, it may be less efficient than OLS (Hansen, 2007). Therefore, many authors prefer to use OLS and reformulate their inferences by simply adopting an alternative estimator for the variance that is robust to heteroscedasticity or serial autocorrelation. However, for large samples, FGLS is preferred over OLS in the presence of heteroscedasticity or serial correlation. (Baltagi, 2008).

A cautionary note is that the FGLS estimator is not always consistent. One case where FGLS might be inconsistent is if there are individual-specific fixed effects. (Greene, 2003)

In general, this estimator has different properties from GLS. For large (i.e., asymptotic) samples, all properties are (under appropriate conditions) shared with GLS, but for finite samples, the properties of FGLS estimators are unknown: estimators vary dramatically depending on each model, and, typically, their exact distributions cannot be derived analytically. For finite samples, FGLS can be even less efficient than OLS in some cases. Thus, although GLS can be feasible, it is not always wise to apply this method when the sample size is small (Hansen, 2007).

A common method to improve estimator accuracy in finite samples is iteration, where residuals from FGLS are used to update the error covariance estimator. This process is repeated until the estimators stabilise within a set tolerance. However, for small samples, this approach does not always improve efficiency. A practical alternative for moderate-sized samples is to use OLS while discarding the classical variance estimator: $\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$. (Greene, 2003)

In this context, the classical OLS estimator can become inconsistent due to heteroscedasticity or autocorrelation in the errors. To address these issues, a HAC (Heteroskedasticity and Autocorrelation Consistent) estimator can be used (Newey & West, 1987). For autocorrelation, the Bartlett estimator (popularised as the Newey-West estimator in their 1987 Econometrica paper) is effective. For heteroscedasticity, the Eicker-White estimator is widely applied (Eicker, 1967; White, 1980). This approach is safer and preferable, unless the sample size is large.

Determining what constitutes a "large" sample depends on the error distribution. For asymmetric error distributions, a sample size exceeding 500 observations may be required (Lumley et al., 2002).

The OLS estimator is calculated as usual by: $\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y}$, where: $\hat{\beta}$ is the vector of OLS coefficient estimates, X is the matrix of independent variables, *Y* is the vector of dependent variable values.

This estimator minimises the sum of squared residuals and is efficient under the classical assumptions of homoscedasticity and no autocorrelation in the error terms. However, in cases of heteroscedasticity or autocorrelation, robust methods such as HAC estimators are recommended. The residuals are estimated by: $\hat{\mathrm{u}} = \mathrm{Y} - \mathrm{X}\hat{\beta}$. (Baltagi, 2008)

To simplify, consider a model with heteroskedastic errors. In this case, one assumes that the variance-covariance matrix Ω of the error vector is diagonal, or equivalently, that the errors from distinct observations are uncorrelated. Each diagonal entry represents the variance of the errors for each observation and can be estimated using the fitted residuals $\hat{\mathrm{u}}_\mathrm{j}$. Thus, Ω^{OLS} can be constructed as: $\Omega^{\mathrm{OLS}} = \mathrm{diag}(\hat{\mathrm{u}}_1^2, \hat{\mathrm{u}}_2^2, \ldots, \hat{\mathrm{u}}_\mathrm{n}^2)$. Where $\hat{\mathrm{u}}_\mathrm{j}^2$ represents the squared fitted residual for each observation j. This diagonal matrix contains the estimates of the individual variances for each residual error. Once Ω^{OLS} is constructed, it can be used to adjust variance

estimates and produce heteroskedasticity-robust estimators, such as the White-Eicker robust estimation. (White, 1980; Eicker, 1967)

It is important to note that squared residuals $\hat{u}_j^2$ cannot directly be used as the entries of Ω^{OLS}; we need an estimator for the variances of the errors. To achieve this, we can use a parametric model of heteroskedasticity or a nonparametric estimator. Once this step is completed, we can proceed with the following:

The estimation of β_{FGLS1} using Ω^{OLS}can be performed through the Weighted Least Squares (WLS) method. This involves the following steps: the first step is to construct the variance-covariance matrix Ω^{OLS} based on the OLS residuals:

$\Omega^{OLS} = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_n^2)$,

where $\hat{\sigma}_j^2$ are the estimated variances of the errors. The next step is to compute the inverse of the variance-covariance matrix Ω^{OLS}:

$$\Omega^{OLS-1} = \text{diag}(\hat{\sigma}_1^{-2}, \hat{\sigma}_2^{-2}, \ldots, \hat{\sigma}_n^{-2})$$

and afterwards to estimate β_{FGLS1}, using the WLS method, the estimator β_{FGLS1} can be calculated as:

$$\hat{\beta}_{FGLS\,1} = (X^{Transp}\Omega^{OLS-1}X)^{-1}X^{Transp}\Omega^{OLS-1}Y,$$

where: X is the matrix of independent variables, Y is the vector of dependent variable observations.

This formulation allows for the estimation of the coefficients while accounting for the heteroscedasticity present in the model, thus providing more efficient estimates compared to OLS. The resulting estimator $\hat{\beta}_{FGLS\,1}$ will be consistent and asymptotically efficient under the correct specification of the error structure. (Baltagi, 2008; Greene, 2003)

The procedure can indeed be iterated to refine the estimates of the coefficients and the variance-covariance matrix. The first iteration is given by: $\hat{\beta}_{FGLS\,1}$ - starting point is the initial estimation using the Weighted Least Squares (WLS) approach:

$$\hat{\beta}_{FGLS\,1} = (X^{Transp}\Omega^{OLS-1}X)^{-1}X^{Transp}\Omega^{OLS-1}Y.$$

After obtaining the initial estimate of the coefficients, we calculated the residuals: $\hat{u}_j = Y_j - X_j\hat{\beta}_{FGLS\,1}$, where $\hat{u}_j$ are the fitted residuals for each observation j. Next, we updated the variance estimates based on the new residuals: $\hat{\sigma}_1^2 = \hat{u}_j^2$ for each j and construct the new variance-covariance matrix Ω^{OLS} : $\Omega^{OLS-1} = \text{diag}(\hat{\sigma}_1^{-2}, \hat{\sigma}_2^{-2}, \ldots, \hat{\sigma}_n^{-2})$. And second iterstion follows: $\hat{\beta}_{FGLS\,2}$ - using the updated variance-covariance matrix to obtain the next estimate of the coefficients:

$$\hat{\beta}_{FGLS\,2} = (X^{Transp}\Omega^{OLS-1}X)^{-1}X^{Transp}\Omega^{OLS-1}Y.$$

This process can be repeated iteratively, updating the variance-covariance matrix, and re-estimating β until the estimates converge or the changes in estimates fall below a specified tolerance level. This iterative approach helps in refining the estimates of the coefficients and improving the robustness of the model against heteroscedasticity. (Hansen, 2007; Greene, 2003).

This estimate Ω^{OLS} can be iterated until convergence. Under regularity conditions, any FGLS estimator (or any of its iterations, if a finite number of iterations is performed) is asymptotically distributed as: $\sqrt{\mathrm{n}}(\hat{\beta}^{\mathrm{FGLS}} - \beta) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Omega)$, where n is the sample size, $\hat{\beta}_{FGLS}$ is the FGLS estimator, β is the real parameter vector and $\mathrm{N}(0, \Omega)$ represents a multivariate normal distribution with mean 0 and variance-covariance matrix Ω.

This asymptotic distribution implies that, as the sample size increases, the distribution of the FGLS estimator approaches a normal distribution centred around the real parameter values, with variance determined by the specified covariance matrix. This property allows for valid statistical inference using the FGLS estimator, such as hypothesis testing and constructing confidence intervals, assuming that the model is correctly specified and the necessary regularity conditions hold. (Baltagi, 2008; Greene, 2003)

The iterative Seemingly Unrelated Regression (SUR) model involves using the errors obtained from the SUR model to construct the covariance matrix Ω and re-estimating the model until convergence is achieved. In our case, we considered two norms to capture the changes in the estimator. These two norms are:

Norm 1 (Euclidean norm): This norm measures the magnitude of the vector of changes in the estimates, defined as:

$$\mathrm{N1} = \sqrt{\sum(\hat{\beta}^{\mathrm{k+1}} - \hat{\beta}^{\mathrm{k}})^2}.$$

Norm 2 (Absolute norm): This norm focuses on the total absolute changes among the estimates, defined as:

$$\mathrm{N2} = \Sigma|\hat{\beta}^{\mathrm{k+1}} - \hat{\beta}^{\mathrm{k}}|.$$

Convergence is achieved when both norms satisfy N1 < ε and norm N2 < ε, where ε represents a threshold value. In this study, we consider this threshold to be 10^{-6}.

In practice, the iterative procedure concludes when the values of both norms decrease below the threshold of 10^{-6}, indicating that further updates would produce negligible changes. This criterion ensures that the estimates have stabilised sufficiently, indicating that further iterations are unlikely to produce significant changes in the parameter estimates. By setting a rigorous convergence criterion, we ensure that the estimates obtained from the SUR model are reliable and accurately reflect the underlying relationships in the data. (Greene, 2003; Baltagi, 2008)

3.1 Model Description and Hypothesis

To assess the performance of the SUR and ISUR estimators, we employed Monte Carlo simulations. As a robustness measure for each increase in sample size, we simulated 1000 different datasets. To mitigate computational issues, we utilised only 3 exogenous variables, one of which is a constant.

In the first step, we generated a random covariance matrix using the following formula: $\Omega = \mathrm{P}\,\Sigma\,\mathrm{P}^{Transp}$, where P is an orthogonal matrix and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2 \ldots, \sigma_n)$, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.

Additionally, we randomly generated a vector β for each variable and for each individual. Both Ω and β are fixed prior to the simulation cycle to facilitate the comparison of the ISUR estimators $\hat{\beta}_{\mathrm{ISUR}}$ using the N1 norm and $\hat{\beta}_{\mathrm{ISUR}}$ using the N2 norm against the previously established β.

This setup allows for a controlled environment where we can systematically evaluate how well the ISUR estimators perform relative to the real parameter values across different sample sizes and conditions. By employing both the N1 and N2 norms, we can capture different aspects of estimator performance, providing a comprehensive analysis of robustness and efficiency. (Greene, 2003; Baltagi, 2008)

Using the multivariate normal distribution function, we randomly generated the matrix X.

To obtain Y, we multiply the vector β by X and add the elements from the covariance matrix Ω for each of the n individuals, as described by the following formula:

$$\mathrm{Y}_i = \beta_{\mathrm{i0}}\mathrm{X}_{\mathrm{i0}} + \beta_{\mathrm{i1}}\mathrm{X}_{\mathrm{i1}} + \beta_{\mathrm{i2}}\mathrm{X}_{\mathrm{i2}} + \mathrm{u}_{\mathrm{i}},$$

where X_{i0} represents a vector of ones with dimensions T×1, $\mathrm{i} = 1, \ldots, \mathrm{N}$, where i is the number of individuals, X_{i1} and X_{i2} are the vectors of the two exogenous variables for individual i, generated from the multivariate normal distribution and u_{i} is the error term, which can be generated based on the specified covariance structure.

Finally, we repeat this process for each individual i to construct a panel dataset.

After constructing the panel, we executed the procedure to obtain the ISUR estimators using both N1 and N2 norms. In testing the model, we extended the panel in each dimension while keeping the other dimension constant. In the following section, we will present the simulation results for both cases: when the number of individuals N increases and when the time period T increases. (Greene, 2003; Baltagi, 2008)

To begin with, we considered β_0, β_1 and β_2 as 1, 2, and respectively, 3. These represent the real values of β used for constructing the datasets. The variables x_0, x_1 and x_2 were generated randomly based on pre-initialised means and standard deviations.

For the mean $\mathrm{E}[\mathrm{x}_1]$=1, standard deviation σ_{x1}=1, mean $\mathrm{E}[\mathrm{x}_2]$=3, standard deviation $\sigma_{\mathrm{x1}}=\sqrt{10}$, we generated random numbers from the normal distribution, which yielded the initial values for x_1 and x_2.

The residuals were generated according to the multivariate normal distribution, with the means of the residuals set to 0. The variance-covariance matrix was created using a pre-implemented function. Using these data, we constructed the dependent variable y using the equation: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, where ϵ represents the error term, generated from the specified normal distribution.

This procedure was carried out not just once, but for as many Monte Carlo replications – the final design reported in this study uses 1000 Monte Carlo datasets.

We created a function to calculate β_{OLS} and ϵ_{OLS} using regression, which ultimately returns the necessary matrices for subsequent steps. The function that computes the Seemingly Unrelated Regression (SUR) requires the variance-covariance matrix of the residuals from the previous step to return βGLS.

Additionally, we developed functions to calculate the differences between the residuals from step k and step k−1. These differences were calculated using two norms: The N1 norm, which calculates the Euclidean distance between successive estimates, provides a measure of the overall magnitude of change, while the N2 norm, which sums the absolute differences, captures the total element-wise variation between iterations.

This comprehensive approach ensures that we capture the performance of the estimators across the multiple iterations effectively. (Greene, 2003; Baltagi, 2008)

The goal of this process is to stop the calculation of ISUR when the norms fall below ε, which is initially set to 10^{-6}.

3.2 The analysis

In the first step of the Monte Carlo (MC) simulation, data is generated to compute β_{OLS} and $\varepsilon_{\mathrm{OLS}}$. The variance-covariance matrix of the OLS residuals is used to calculate the SUR estimators, β_{GLS}. We then apply an Iterative SUR (ISUR) method, which iterates until the difference between the residuals from iteration k and k−1 falls below 10^{-6}, based on the formulas for two norms. Once the difference is smaller than ε, the ISUR converges, returning the estimators, the number of iterations, the norm value, and the computation time. This process is repeated for 1000 MC simulations.

For each norm, ISUR estimators, norm values at convergence, iteration counts, and computational times are obtained. The results are concatenated across 1000 simulations, and the means and standard deviations are calculated based on changes in the dimensions T (time series) and N (number of individuals, countries, regions, etc.). To isolate the effects of each dimension, only one dimension is modified at a time, while the other remains constant.

When T is varied, N is fixed at 3, and the MC iterations are kept at 1000. T takes values of 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 and 600 sequentially. Similarly, when N is varied, T is fixed at 150, and N takes values from 2 to 9. Means and standard deviations of the estimators are calculated for each modification, ensuring clarity and comparability. Results are available upon request.

4. Results and Discussion

We will continue by presenting the results of the simulations, which focused on evaluating the performance of the estimators and testing the convergence of the two norms. To clarify the results, we calculated an average estimator for each Monte Carlo simulation: $\hat{\beta}_{k,N1}$ and $\hat{\beta}_{k,N2}$ where k = {1,2,3} represents the number of variables, and j = {1, 2,1000} represents the number of Monte Carlo simulations.

Since further dimension reduction was necessary, we computed the mean of the estimators for each increase in dimension over the 1000 Monte Carlo simulations. This resulted in three concise estimators: $\hat{\beta}_{N1}$, $\hat{\beta}_{N2}$ and β_{real} for each variable.

It is important to note that the two norms used are: Norm 1: $N1 = \sqrt{\sum(\hat{\beta}^{k+1} - \hat{\beta}^{k})^2}$ and Norm 2: $N2 = \Sigma|\hat{\beta}^{k+1} - \hat{\beta}^{k}|$. These norms will be critical in assessing the convergence and performance of the estimators throughout our analysis.

4.1 Modification of Dimension T – This refers to the number of days, weeks, months, quarters, semesters, years, etc.

The results confirm the previously stated findings regarding the extension of dimension T. Specifically, the ISUR estimator calculated using Norm 1 performs better than the one calculated using Norm 2. Based on the simulations, we can observe in Figure 1 that the difference between $\hat{\beta}_{N1}$ and β_{real} is smaller than the difference between $\hat{\beta}_{N2}$ and β_{real}.

Furthermore, it is noticeable that the standard deviations are very similar and exhibit the same decreasing trend as t increases, which aligns with the central limit theorem. A complete version of the results can be provided upon request.

Overall, modifying the dimension T is crucial for understanding the robustness and efficiency of econometric estimators under varying temporal conditions.

For this analysis, we considered a N = 3 constant values for the number of cross-section elements, and a constant number of Monte Carlo simulation MC = 1000. We want to see how the coefficients modify if we iterate the model on different values for T (from T = 50, to T = 600).

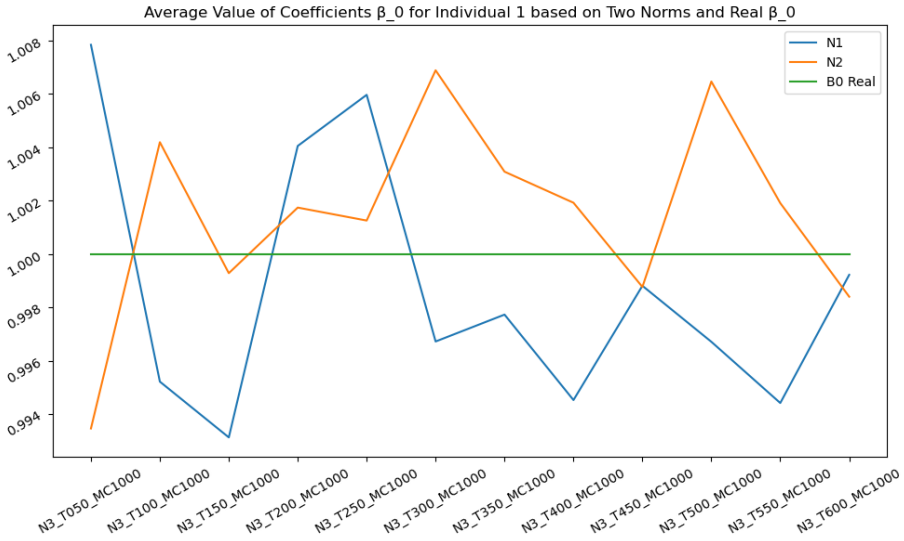

Figure 1. Average Value of Coefficients β_0 for Individual 1 based on Two Norms and Real β_0
Source: Authors' own creation.

Figure 1 illustrates the average estimated coefficients ($\hat{\beta}_0$) for individual 1, comparing the results obtained using the two norms (N1 and N2) against the actual value of β_0. The observed differences highlight the performance and accuracy of the ISUR estimators across varying sample sizes.

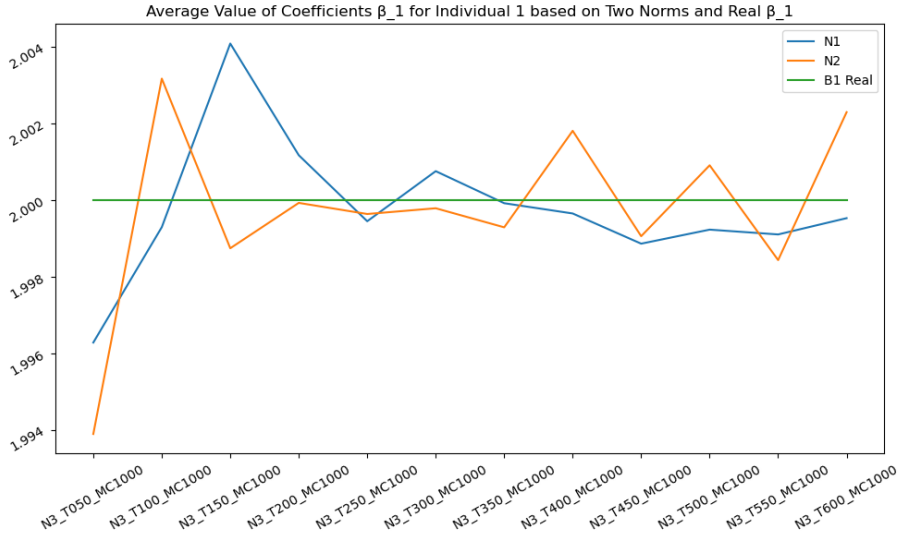

Figure 2. Average Value of Coefficients β_1 for Individual 1 According to the Two Norms and Real β_1
Source: Authors' own creation.

Figure 2 illustrates the average value of the coefficients $\hat{\beta}_1$ estimated for individual 1, comparing the results obtained using the two norms with the actual value of β_1. It provides insights into how well the estimators perform relative to the real coefficient across different sample sizes or iterations.

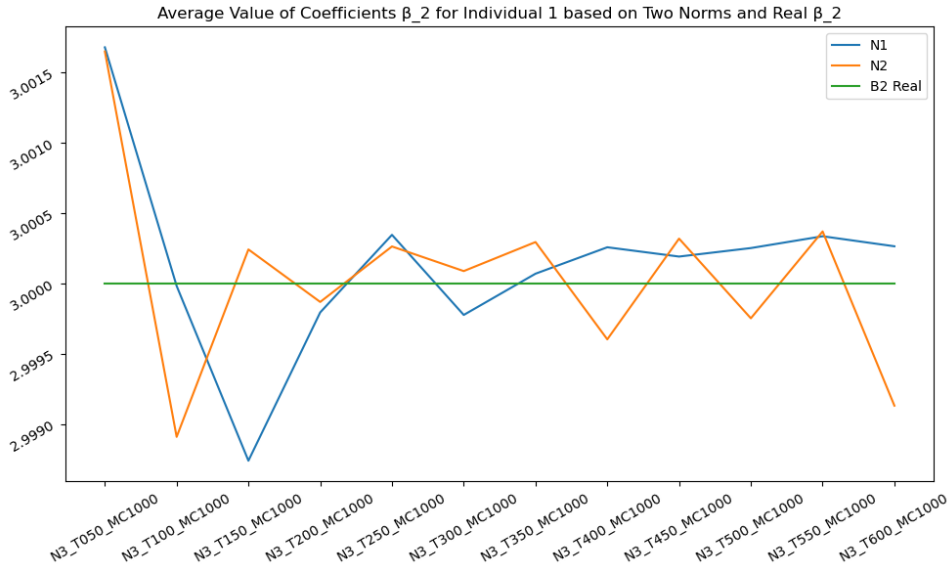

Figure 3. Average Value of Coefficients β_2 for Individual 1 based on the Two Norms and Real β_2

Source: Authors' own creation.

Figure 3 displays the average values of the estimated coefficients $\hat{\beta}_2$ for individual 1, comparing them to the actual value of β_2 across the two norms. It highlights the differences in performance between the estimators and how closely they align with the real coefficient value.

The number of iterations decreases as T increases, indicating faster convergence in longer time series. At the same time, the norms produce similar results, once again highlighting their similarity. However, if we take into account the average number of iterations when t increases, N1 is more efficient, requiring a smaller number of iterations to achieve convergence. In this case, since the number of iterations decreases as t increases, the ISUR model calculated through N1 may present a reliable alternative. In terms of the computational duration of the estimators, we can observe that N1 is more efficient, especially when t is greater than 600, by about 2 seconds. However, we should note that in the simulations, the value of N was kept at 3, which may represent a lack of robustness.

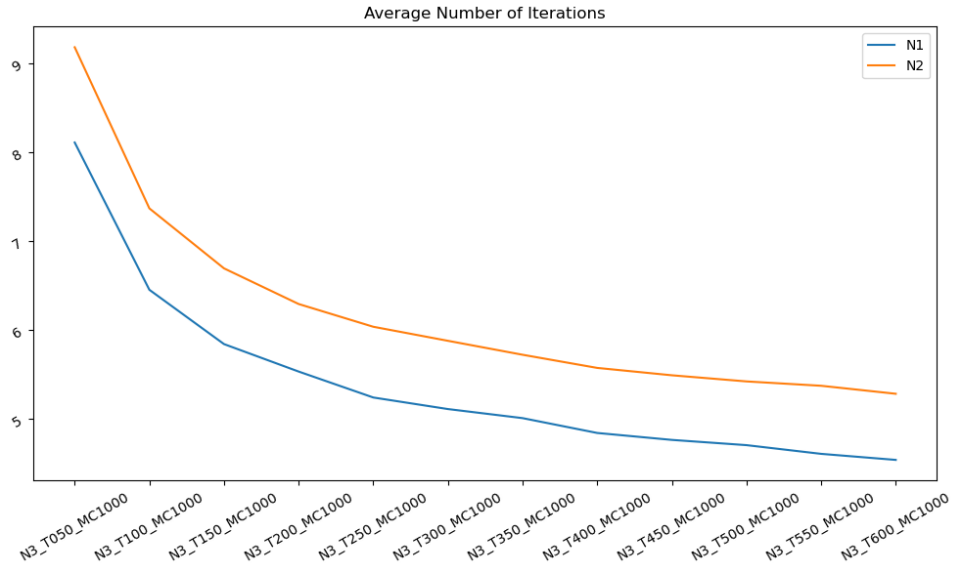

Figure 4. Average Number of Iterations Until Convergence by Norm
Source: Authors' own creation.

Here is Figure 4 representing the average number of iterations until convergence by Norm 1 and Norm 2. It illustrates the downward trend in the number of iterations as the time intervals increase, with Norm 1 requiring fewer iterations compared to Norm 2.

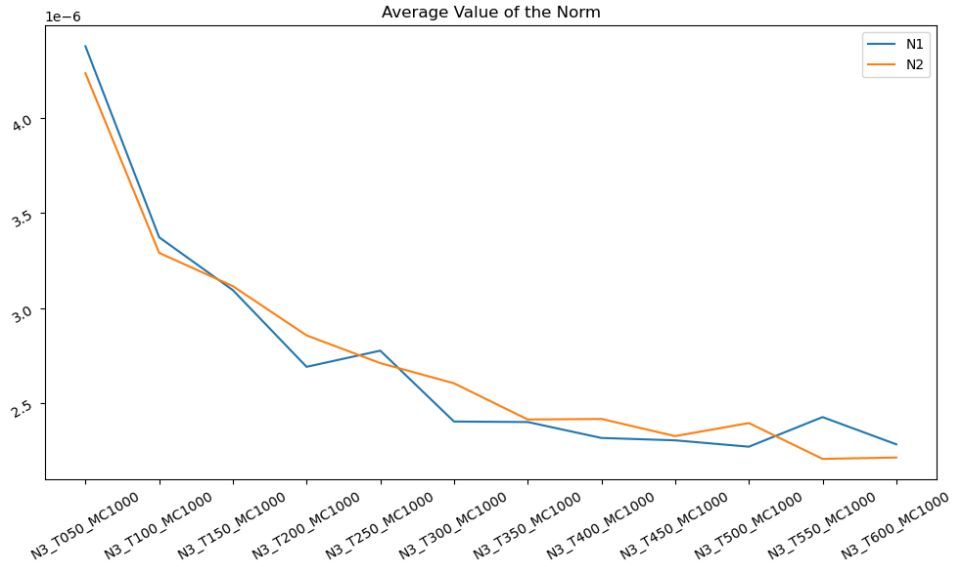

Figure 5. Average Value of the Norm at Convergence by Norm
Source: Authors' own creation.

Figure 5 shows the average value of the norm at convergence for Norm 1 (N1) and Norm 2 (N2) across different time periods (T) and Monte Carlo iterations (MC1000). The plot reveals the performance trend of both norms as the time dimension (T) increases, with the values decreasing for both norms.

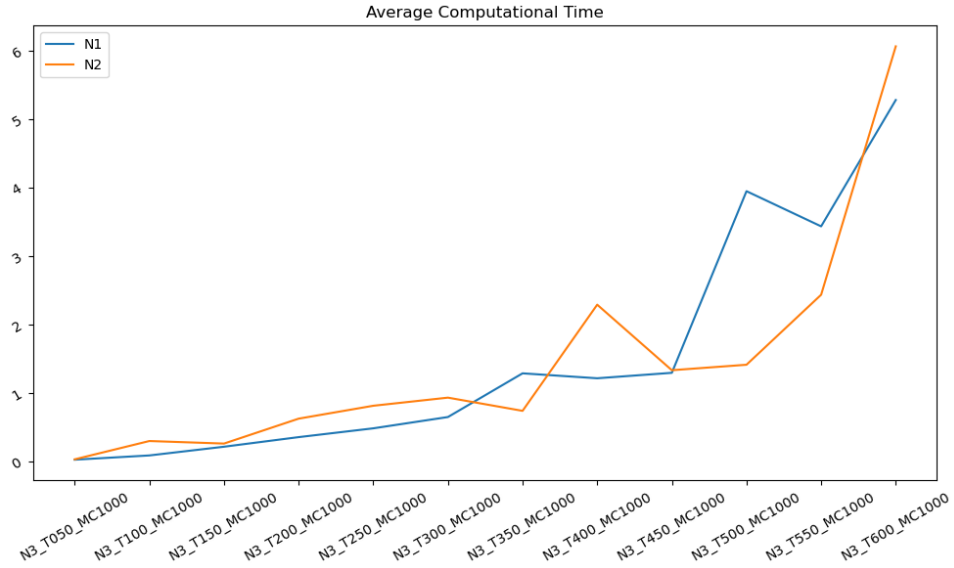

Figure 6. Average Computational Time Until Convergence by Norm
Source: Authors' own creation.

Illustrating the computational time taken for ISUR convergence as a function of dimension T (the time intervals), Figure 6 compares Norm 1 (N1) and Norm 2 (N2), where N1 consistently shows shorter computational times, particularly as T increases, confirming that N1 is more efficient in this aspect.

4.2 Modifying the Dimension N – Number of Individuals, Regions, Countries, etc.

To begin, we will present the results for the expansion of the dimension n. For testing purposes, we simulated 1000 different datasets for each n ranging from 2 to 9. To enable comparison of the results, we established three $\hat{\beta}$ coefficients, one for each variable, so that all individuals share the same coefficients: β0 = 1, β1 = 2, and β2 = 3.

The aim of modifying the dimension n is to understand how increasing the number of individuals affects the performance and reliability of the SUR and ISUR estimators, thereby providing insights into the robustness of these estimators in econometric modelling.

To this analysis, we considered a T=150 constant values for the number of time periods, and a constant number of Monte Carlo simulation MC = 1000. We want to

see how the coefficients modify if we iterate the model on different values for N (from N=2, to N=9).

The results show that the differences between the means of the estimators calculated using N1 and N2 are very small, on the order of 10^{-3}, both between them and compared to the real values. At the same time, the errors are approximately equal, which leads to the first conclusion that the estimators are performing well, and neither N1 nor N2 shows significant improvement over the other. We expected to see a decreasing trend in the errors, tending toward zero, which could represent an empirical result of the central limit theorem. However, in this case, it is not evident because increasing N does not increase T, and therefore the central limit theorem effects may not appear. Nevertheless, if we were to simulate data with a sufficiently large N, this hypothesis would likely be confirmed.

Figures 7 to 9 highlight the evolution of the two norms in terms of convergence iterations and computational time. As N increases, the number of iterations required for convergence grows. For example, at N = 2, both norms require 5 iterations, while at N = 9, this rises to 10 iterations. Computational time also increases with dimensionality, reaching 17 seconds at N = 9, compared to less than a second at smaller dimensions. The norms differ in convergence patterns: N1 typically requires fewer iterations than N2, often by more than one iteration. While their computational times are similar, N1 is slightly more efficient, with differences of less than a second between the norms.

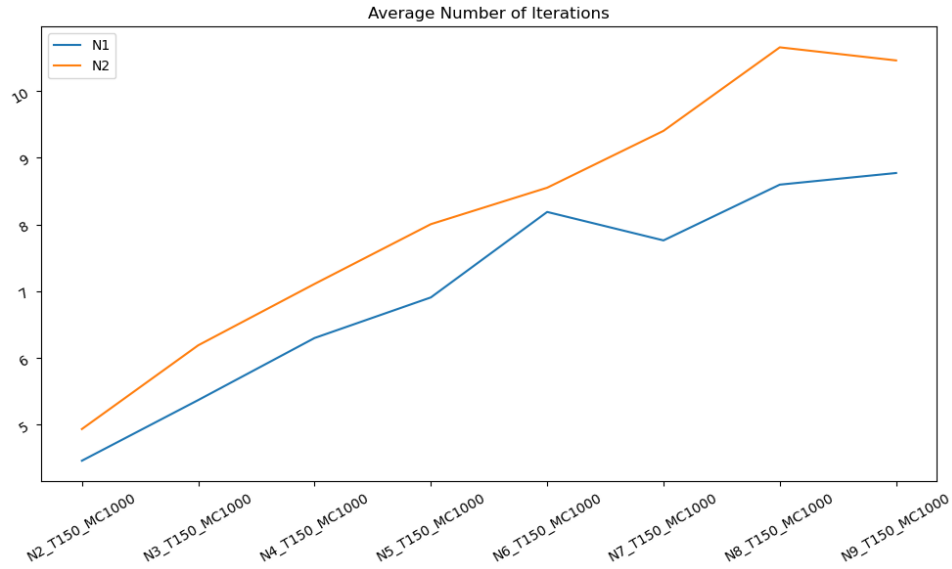

Figure 7. Average Number of Iterations Until Convergence by Norm
Source: Authors' own creation.

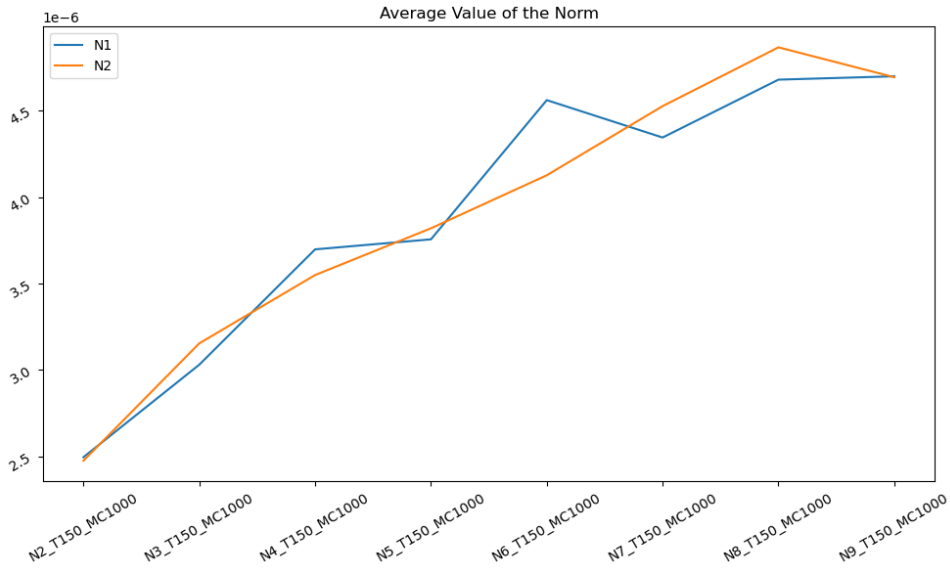

Figure 8. Average Value of the Norm at Convergence by Norm
Source: Authors' own creation.

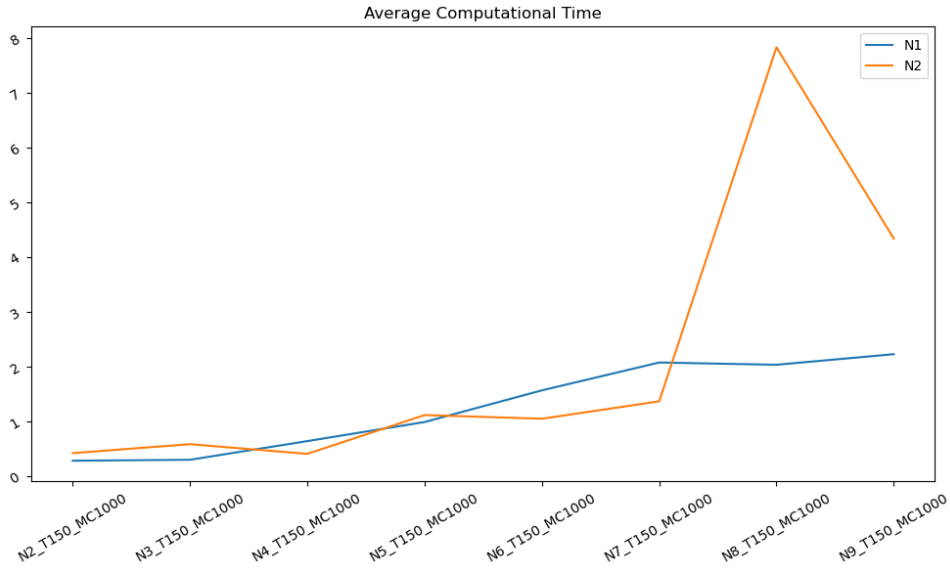

Figure 9. Average Computational Time Until Convergence by Norm
Source: Authors' own creation.

4.3 Empirical Illustration: SUR Estimation on Labor Market Indicators

To complement the Monte Carlo investigation and reinforce the practical relevance of norm selection in ISUR, we estimate a three-equation SUR model using an empirical dataset derived from World Bank indicators for 112 countries over the period 2010-2022 (see Supplementary Materials – section 2. Empirical Application). The dependent variables capture key labour-market outcomes: (i) the employment-to-population ratio (EMP_POP_i), (ii) the unemployment rate ($UNEMPL_i$), and (iii) the labour force participation rate ($LABOR\_FORCE_i$). The explanatory variables include GDP per capita (GDP_CAP_i), Internet use as a share of the population ($INTERNETUSE_i$), secondary school enrolment ($SCHOOLSEC_i$), and the share of the urban population (URBAN_i). These indicators jointly reflect structural socioeconomic conditions that tend to evolve together across countries, implying a contemporaneous correlation in their disturbances and motivating the use of a SUR specification.

Consistent with this rationale, we specify the following three-equation SUR system:

$$\mathrm{EMP\_POP_i} = \alpha_1 + \gamma_1 \mathrm{GDP\_CAP}_i + \delta_1 \mathrm{INTERNET_i} + \vartheta_1 \mathrm{SCHOOL\_SEC}_i + \omega_1 \mathrm{URBAN}_i + u_{1i},$$

$$\mathrm{UNEMPL}_i = \alpha_2 + \gamma_2 \mathrm{GDP\_CAP}_i + \delta_2 \mathrm{INTERNET_i} + \vartheta_2 \mathrm{SCHOOL\_SEC}_i + \omega_2 \mathrm{URBAN}_i + u_{2i},$$

$$\mathrm{LABOR_{FORCE}}_i = \alpha_3 + \gamma_3 \mathrm{GDP\_CAP}_i + \delta_3 \mathrm{INTERNET_i} + \vartheta_3 \mathrm{SCHOOL\_SEC}_i + \omega_3 \mathrm{URBAN}_i + u_{3i},$$

where the disturbance vector $(u_{1i'}, u_{2i'}, u_{3i})'$ is allowed to follow a non-diagonal contemporaneous covariance structure. As expected in multi-equation labour-market settings, the estimated residual correlation matrix displays sizeable off-diagonal elements, confirming the suitability of the SUR framework.

The system is estimated using both non-iterative SUR (FGLS) and Iterative SUR under the two convergence norms defined earlier: Norm 1 (Euclidean) and Norm 2 (Absolute). The Python implementation ensures that both ISUR procedures are identical in all respects except for the norm employed. For each estimation, we record (i) the number of iterations required to reach the convergence threshold of 10^{-6} (ii) computational time, and (iii) the final norm value.

Figures 10 to 15 report the empirical computational times for both convergence norms, showing that the Euclidean norm (N1) requires substantially less time to reach convergence than the Absolute norm (N2). Although the simulation study revealed greater variation in iteration counts across sample dimensions, the empirical dataset is sufficiently stable for both norms to converge in a single iteration. Even so, meaningful differences persist in the final norm magnitudes and in the

computational time required, further underscoring the superior efficiency of the Euclidean criterion in practical applications.

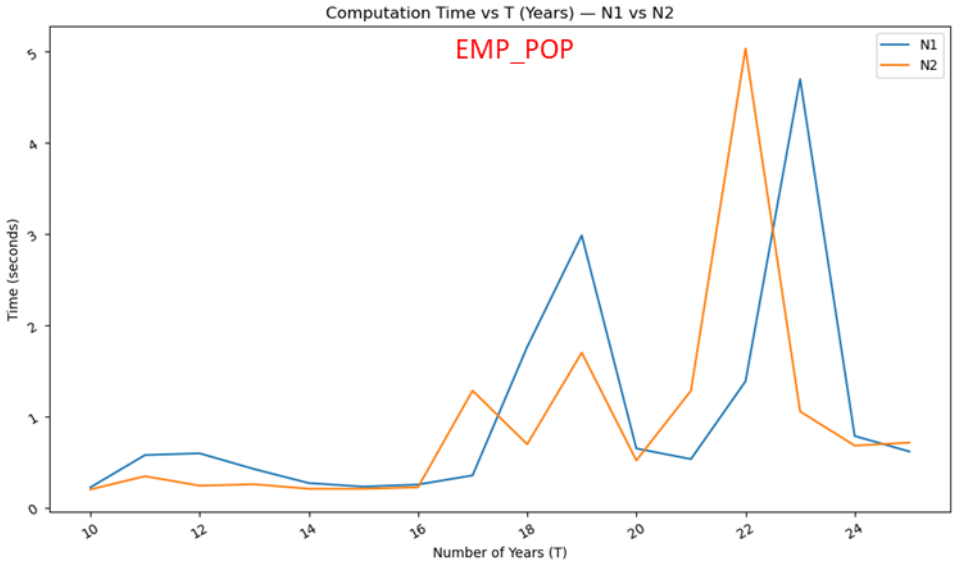

Figure 10. Computation Time by Sample Length (T): Norm 1 vs Norm 2 for EMP_POP

Source: Authors' own creation.

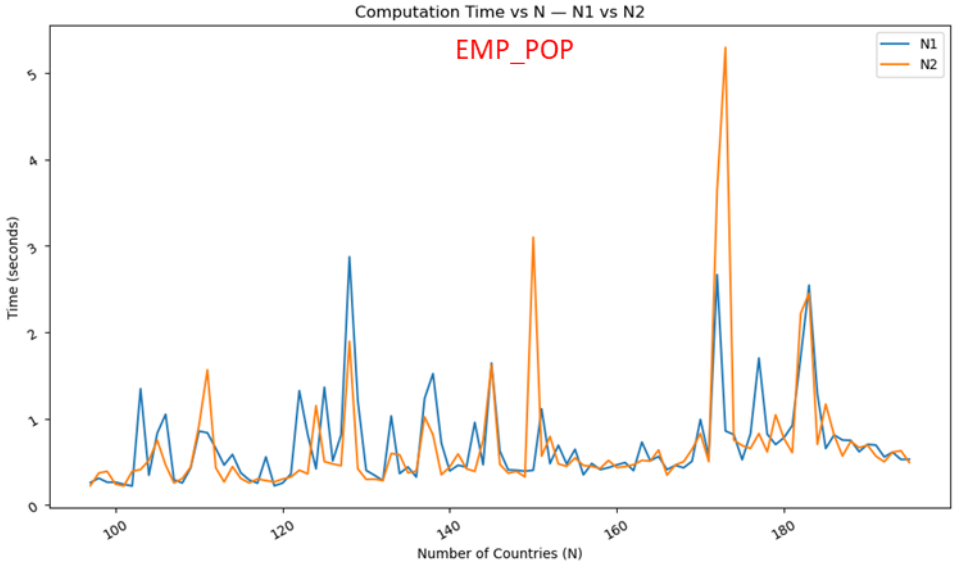

Figure 11. Computation Time by Number of Countries (N): Norm 1 vs Norm 2 for EMP_POP

Source: Authors' own creation.

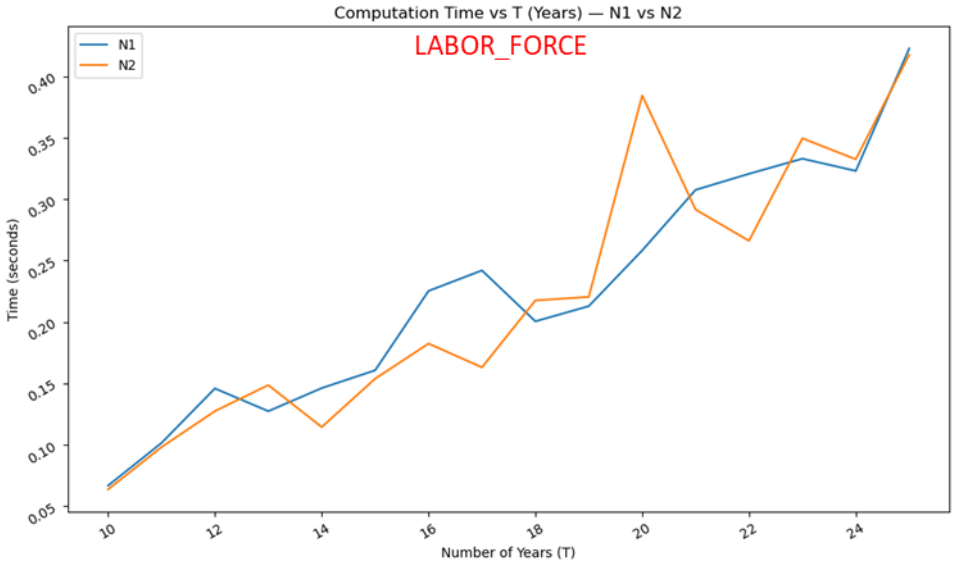

Figure 12. Computation Time by Sample Length (T): Norm 1 vs Norm 2 for
Source: Authors' own creation.

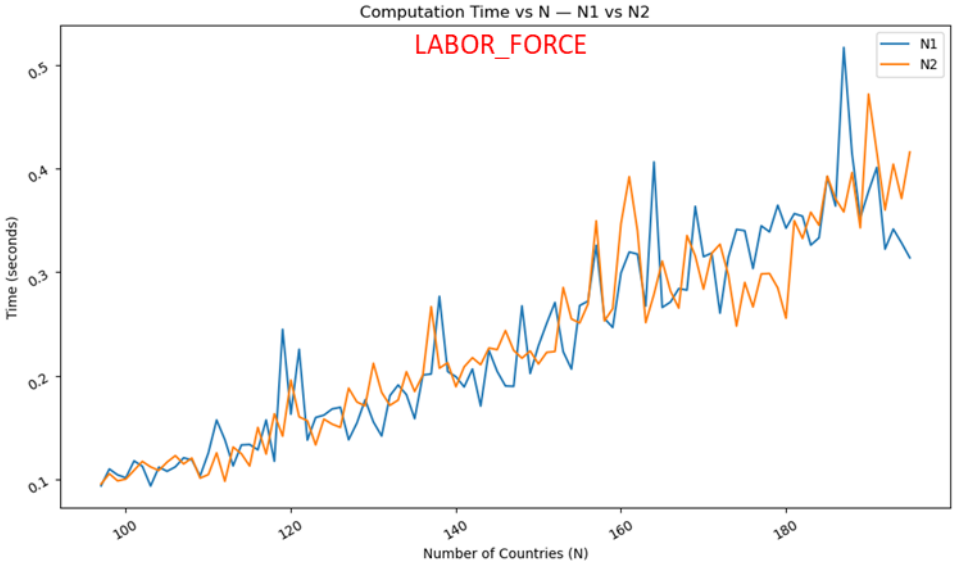

Figure 13. Computation Time by Number of Countries (N): Norm 1 vs Norm 2 for LABOR_FORCE
Source: Authors' own creation.

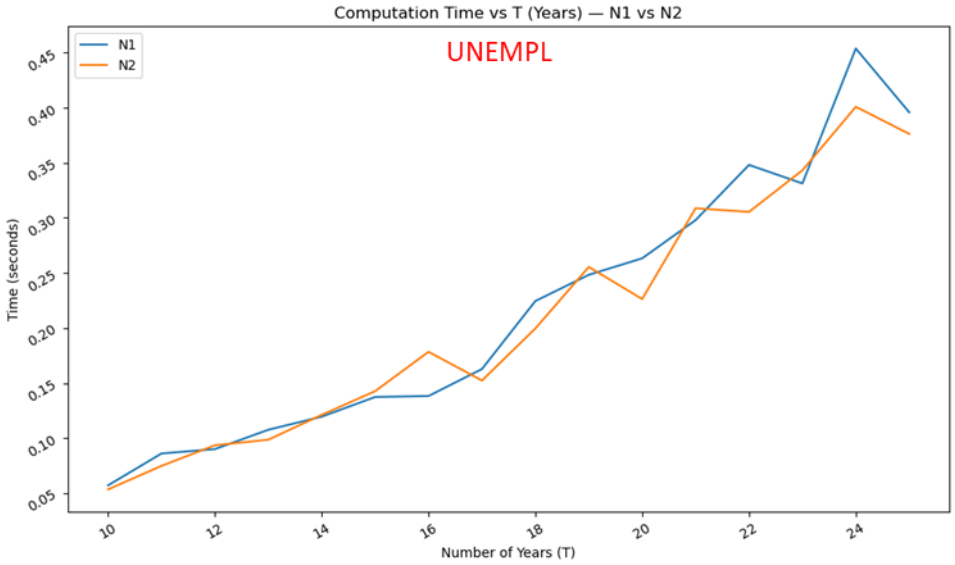

Figure 14. Computation Time by Sample Length (T): Norm 1 vs Norm 2 for UNEMPL

Source: Authors' own creation.

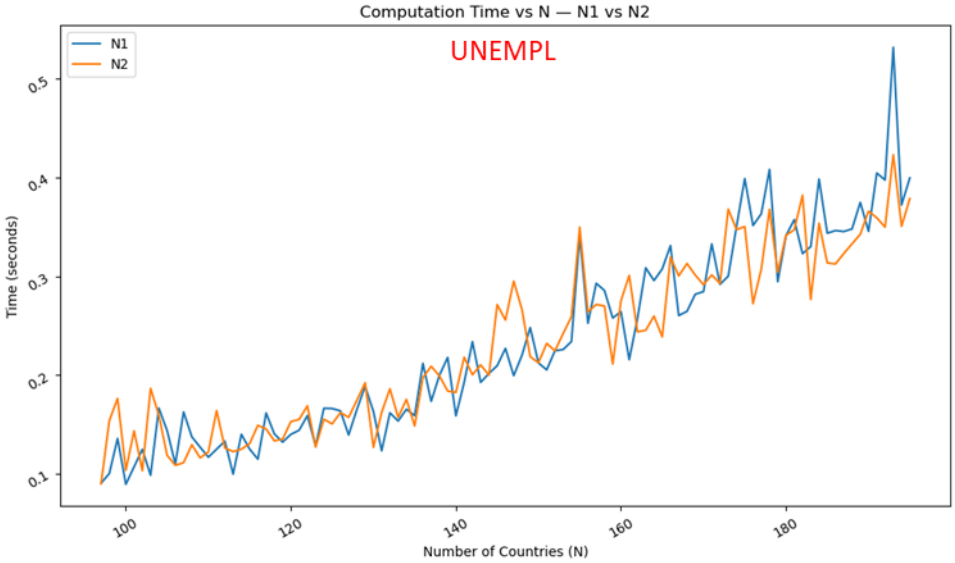

Figure 15. Computation Time by Number of Countries (N): Norm 1 vs Norm 2 for UNEMPL

Source: Authors' own creation.

Across all three equations, the empirical results are consistent with the broader patterns observed in the Monte Carlo analysis. Although both convergence norms reach the stopping criterion in a single iteration for this particular dataset, the

Euclidean norm (N1) requires noticeably less computational time and attains smaller final norm values than the Absolute norm (N2). These differences, while moderate in this relatively small empirical system, mirror the systematic efficiency advantage identified in the simulation study. Importantly, norm selection affects only the computational trajectory of the ISUR procedure: the estimated coefficients, standard errors, and overall model fit remain virtually identical across both norms and with the non-iterative SUR benchmark. Overall, this empirical illustration confirms that norm choice is an operationally meaningful modelling decision. Even in standard applications with modest dimensionality, the Euclidean norm improves computational efficiency without altering the statistical content of the estimates. As model complexity increases – through additional equations, regressors, or restrictions – the practical relevance of selecting an efficient convergence norm becomes even more pronounced, in line with the patterns revealed by the Monte Carlo experiments.

5. Conclusions

5.1 Modification of Dimension T

The values of $\hat{\beta}_{N1}$ and $\hat{\beta}_{N2}$ fluctuate around the real values β_{real}. This was expected, given that the variables also have a random component in their structure. The average deviation of coefficients β0, β1 and β2 decreases as T increases, converging toward 0 with the growth of dimension T. At low T, $\hat{\beta}_{N2}$ is more efficient compared to $\hat{\beta}_{N1}$. However, as T is modified, the differences change slightly. Ultimately, both estimators, regardless of the norm used in their calculation, can be said to be efficient, showing very similar performance. The differences $\hat{\beta}_{N1} - \beta_{\text{real}}$ and $\hat{\beta}_{N2} - \beta_{\text{real}}$ are small, on the order of 10^{-3} but they are not constant. For small T (between 50 and 200), the differences $\hat{\beta}_{N2} - \beta_{\text{real}}$ are greater than $\hat{\beta}_{N1} - \beta_{\text{real}}$. As T increases, N2 becomes more efficient. The average number of iterations for both norms decreases as the dimension T increases. From an efficiency standpoint, we can conclude that N1 (the Euclidean norm) is more efficient, as it requires fewer iterations to reach convergence. The average value of the norm decreases as T increases. Initially, with $\varepsilon = 10^{-6}$, convergence is observed for small T starting at 4×10^{-6} and as T increases, it diminishes to 2×10^{-6}. The average computational time increases as T increases; Notably, N1 (the Euclidean norm) generally requires less time on average for estimation, making it the more efficient option.

5.2 Modification of Dimension N

The values of $\hat{\beta}_{N1}$ and $\hat{\beta}_{N2}$ fluctuate around the real values βreal. This was expected, given that the variables also have a random component in their structure. The average deviation of coefficients β0, β1 and β2 decreases as N increases,

converging toward 0 with the growth of dimension N. At low N, $\hat{\beta}_{N1}$ is more efficient compared to $\hat{\beta}_{N2}$. However, as N is modified, the differences change slightly. The differences $\hat{\beta}_{N1} - \beta_{\text{real}}$ and $\hat{\beta}_{N2} - \beta_{\text{real}}$ are small, on the order of 10^{-3} but they are not constant. For small N (between 2 and 5) the differences are smaller than those for larger N. An N increases, the differences $\hat{\beta}_{N1} - \beta_{\text{real}}$ are still smaller than $\hat{\beta}_{N2} - \beta_{\text{real}}$, which indicates that $\hat{\beta}_{N1}$ is more efficient. The average number of iterations for both norms increases as the dimension N increases. From an efficiency perspective, we can conclude that N1 (the Euclidean norm) is more efficient, as it requires fewer iterations to reach convergence. The average value of the norm decreases as N increases. Initially, with $\varepsilon = 10^{-6}$, convergence is observed for small N starting at 2×10^{-6} and as N increases, it rises to 5×10^{-6}. The average computational time increases as N increases; however, it can be observed that N1 (the Euclidean norm) generally requires a shorter time on average for estimating, which allows us to classify it as the more efficient option.

We put the Iterative Seemingly Unrelated Regression (SUR) model to the test, comparing how quickly and efficiently it converges under two different norm specifications within the Feasible Generalised Least Squares (FGLS) framework. To carry out this comparison, we designed a custom Python-based algorithm capable of implementing both estimators, allowing for a head-to-head evaluation of their computational performance and precision.

This study evaluates the performance of the Iterative Seemingly Unrelated Regression (ISUR) model by comparing the convergence speed and efficiency of two norms based on Feasible Generalised Least Squares (FGLS) estimation. Using Python, we developed an algorithm to calculate the estimators and employed Monte Carlo simulations to generate data from a known variance-covariance matrix (Ω) using a multivariate normal distribution. The endogenous variable for each dimension n was constructed based on a priori values of $\hat{\beta}$.

To streamline analysis, we averaged the results of $\hat{\beta}$ across dimensions and Monte Carlo cycles. Simulations revealed that the ISUR estimator, particularly with norm N1, is highly efficient. While differences between norms are minor for smaller dimensions ($N<9$, $T<600$), increasing n and t enhances the estimators' performance, with N1 achieving faster convergence and greater computational efficiency.

This study fills a methodological gap by showing that norm selection is not innocuous: the Euclidean norm systematically reduces iteration counts and computation time without compromising estimator accuracy. To the best of my knowledge, this is the first study to examine how the choice of convergence norm affects Iterative SUR estimation. The paper provides a systematic computational comparison, complemented by an empirical application demonstrating the practical implications for real-world datasets.

This paper is the first to show systematically that Euclidean norms reduce iteration counts and computational time without loss of accuracy. These findings refine best practices for applied researchers implementing iterative GLS methods.

As the dimensionality of the model increases, the performance gap between the two norms becomes more pronounced, with Norm 1 (N1) showing a clear advantage.

Moreover, an increase in the number of cross-sectional units (N) tends to raise the number of iterations required for convergence, while a longer time span (T) generally reduces it. However, the robustness of these findings – particularly regarding iteration counts – is limited, as the simulations involving complementary dimensions were conducted using relatively small values for N and T, which may restrict the generalisability of the results.

Acknowledgements: *This work is part of "A better understanding of socio-economic systems using quantitative methods from Physics" funded by the European Union – NextgenerationEU and the Romanian Government, under National Recovery and Resilience Plan for Romania, contract no 760034/23.05.2023, cod PNRR-C9-I8-CF255/29.11.2022, through the Romanian Ministry of Research, Innovation and Digitalization, within Component 9, Investment I8".*

References

[1] Aitken, A.C. (1936), *On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh*, 56, 58-60.

[2] Baltagi, B.H. (2008), *Econometric analysis of panel data* (4th ed.). *Chichester, UK: John Wiley & Sons Ltd.*

[3] Davidson, R., MacKinnon, J.G. (1993), *Estimation and inference in econometrics*. *New York, USA: Oxford University Press.*

[4] Eicker, F. (1967), *Limit theorems for regressions with unequal and dependent errors*. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 59-82). *Berkeley, CA: University of California Press, USA.*

[5] Greene, W.H. (2012), *Econometric analysis* (7th ed.). *Upper Saddle River, NJ, USA: Pearson Education.*

[6] Hansen, C.B. (2007), *Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. Journal of Econometrics*, 140(2), 670-694. https://doi.org/10.1016/j.jeconom.2006.07.011.

[7] Hayashi, F. (2000), *Econometrics*. *Princeton, NJ, USA: Princeton University Press.*

[8] Kiviet, J.F. (1995), *On bias, inconsistency, and efficiency of various estimators in dynamic panel data models*. *Journal of Econometrics*, 68(1), 53-78, https://doi.org/10.1016/0304-4076(94)01643-E

[9] Kontoghiorghes, E.J. (2000). *Parallel algorithms for linear models: Numerical methods and estimation problems*. *New York, NY, USA: Springer.*

[10] Lumley, T., Diehr, P., Emerson, S., Chen, L. (2002), *The importance of the normality assumption in large public health data sets. Annual Review of Public Health*, 23(1), 151-169, https://doi.org/10.1146/annurev.publhealth.23.100901.140546.

[11] Mardia, K.V., Kent, J.T., Bibby, J.M. (1979), *Multivariate analysis*. *London, UK: Academic Press.*

[12] Newey, W.K., West, K.D. (1987), *A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica*, 55(3), 703-708.

[13] Parsaeian, S. (2023), *Structural breaks in seemingly unrelated regression models. Macroeconomic Dynamics*, 28(4), 1-24, https://doi.org/10.1017/S1365100523000329.

[14] Srivastava, M., Giles, D.E. (1987), *The seemingly unrelated regression equations: A survey*. In R.F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics*, 3, 2165-2225, Amsterdam, Netherlands: North-Holland.

[15] Tan, L., Chiong, K.X., Moon, H.R. (2021), *Estimation of high-dimensional seemingly unrelated regression models*. *Econometric Reviews*, 40(9), 830-851, https://doi.org/ 10.1080/07474938.2021.1889195.

[16] White, H. (1980), *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*. *Econometrica*, 48(4), 817-838.

[17] Youssef, A.H., Abonazel, M.R., Kamel, A.R. (2022), *Efficiency comparisons of robust and non-robust estimators for seemingly unrelated regressions model*. *WSEAS Transactions on Mathematics,* 21, 218-244, https://doi.org/10.37394/ 23206.2022.21.28.

[18] Zellner, A. (1962), *An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association*, 57(298), 348-368, https://doi.org/10.2307/2281644.