YEARBOOK OF UNWE (2025), ACADEMIC PUBLISHING HOUSE UNWE
ISSN (print): 1312-5486; ISSN (online): 2534-8949; http://unwe-yearbook.org/en/
DOI: https://doi.org/10.37075/YB.2025.2.09

PANEL DATA IN ECONOMIC RESEARCH

Gema Ugalde[1], Violena Nencheva[2]
e-mail: gema.rubio@uaq.edu.mx, e-mail: violena.nencheva@uaq.mx

Abstract

This article reviews recent literature (2020 – 2025) on the use of panel data econometrics in economic research, with a focus on growth-related studies in Latin American economies. Using an exploratory-descriptive review design and searches in Google Scholar, Elsevier, Dialnet, and Redalyc, the paper synthesises core panel specifications and estimation approaches. It outlines the general linear panel model and discusses pooled OLS, fixed effects (including LSDV, within, and first-difference estimators), and random effects (error-components) models, highlighting their assumptions, strengths, and limitations. The review also summarises key specifications and diagnostic tests commonly used in applied work (e.g., Breusch–Pagan/LM, Hausman, Wooldridge, and tests for heteroscedasticity). Finally, the article briefly introduces dynamic panel models (Arellano–Bond type GMM) and panel cointegration frameworks, which are frequently, employed extensions in growth-oriented empirical research.

Keywords: panel data, fixed effects, random effects, specifications tests

JEL: C23, C33

Introduction

The use of panel data has become widespread in economics, particularly in econometrics, and it has developed into a highly productive research approach across multiple fields (Ruíz, 2010). Panel methods are frequently applied in empirical studies of development and economic growth, financial crises, foreign direct investment, unemployment, inflation, and comparative economic analysis, among other topics. Their appeal stems mainly from the desirable properties they offer for econometric estimation: panel regressions combine time-series and cross-sectional variation, enabling richer inference than models based on a single data dimension (Gujarati & Porter, 2010; Wooldridge, 2010).

The growing availability of longitudinal datasets has supported the expansion of panel-data applications. International organisations such as the International Monetary Fund (IMF), the World Bank (WB), the United Nations (UN), and

[1] Teaching Prof., PhD, Faculty of Political science, Autonomous University of Queretaro, Mexico, ORCID: 0000-0002-0727-013X

[2] Assoc. Prof., PhD, Faculty of Informatics, Autonomous University of Queretaro, Mexico, ORCID: 0000-0002-0904-7281

the Economic Commission for Latin America and the Caribbean (ECLAC) have compiled and disseminated macroeconomic indicators over long time horizons and for a large number of countries (Arellano & Bover, 1990). In parallel, governments increasingly generate microeconomic information through household and firm surveys. This growing data infrastructure facilitates more comprehensive empirical studies and, in many cases, the construction of balanced panels (Correa & Salazar, 2016).

Panel data enrich empirical analysis by capturing heterogeneity across units and over time, increasing the information available for estimation, and often reducing collinearity among explanatory variables (Gujarati & Porter, 2010). Observing the same unit repeatedly also supports the analysis of change dynamics and allows researchers to control for unobserved heterogeneity (unobserved effects), thereby addressing key limitations of traditional cross-sectional regressions (Perazzi & Merli, 2013). Moreover, panel structures typically provide greater degrees of freedom and can improve precision by reducing estimator variance (Beltrán & Castro, 2020).

At the same time, panel datasets face common data challenges, including coverage issues, missing observations, temporal biases, and measurement error. From an econometric perspective, panel settings often involve heteroscedasticity, serial correlation, and contemporaneous correlation across units, which may compromise standard inference if not adequately addressed (Rodríguez, Freire, & País, 2017). Therefore, applied research commonly relies on diagnostic and specification tests and on estimation approaches designed for panel structures, particularly fixed-effects and random-effects models (Perazzi & Merli, 2013; Gujarati & Porter, 2010).

Motivated by the widespread use of panel-data methods and their relevance for empirical economic research, this study aims to provide a literature review of panel data in economics. The remainder of the article is organised as follows: the next section presents the methodology. The subsequent section discusses the main panel-data models (including fixed- and random-effects models), key estimation techniques, specification tests, and selected empirical applications, followed by the conclusions.

Methodology

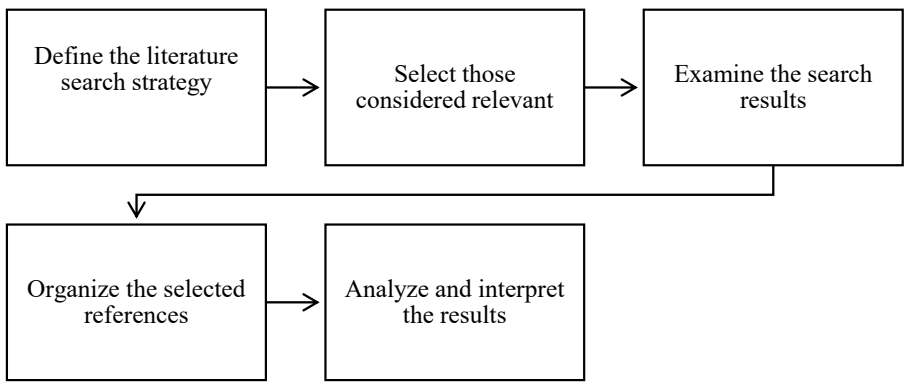
This article conducts a literature review on the use of panel data econometrics in the study of economic growth in Latin American economies, employing an exploratory-descriptive research approach and covering the period from 2020 to 2025. The focus on panel data is motivated by its well-established advantages in econometric estimation, particularly its ability to integrate time-series and cross-sectional dimensions within a unified analytical framework.

Using an exploratory-descriptive design, the study reviews the economic literature on panel data techniques, with particular attention to the general panel-data model and its main estimation approaches. The reviewed methodologies include pooled Ordinary Least Squares (OLS), Fixed Effects Models (FEM), Least Squares Dummy Variable (LSDV) estimators, Random Effects Models (REM), dynamic panel data models, and panel cointegration models. In addition, recent frontier studies in empirical economic research employing these techniques are examined to illustrate their application to growth-related analyses.

The literature review process begins with an initial exploration of the topic to identify established knowledge and existing research gaps. From a methodological perspective, the review systematises the main models, their underlying assumptions, advantages, and limitations, and organises prior studies coherently and analytically. As noted by Guirao (2015), descriptive research is particularly suitable for synthesising concepts and methodological developments within a specific field of study.

The review strategy was designed by defining appropriate keywords aligned with the research objective and by delimiting the relevant databases and catalogues. Search criteria were established with particular emphasis on time coverage, prioritising recent, methodologically relevant contributions published within a clearly defined period (Sabatés & Salas, 2020).

The review proceeds in two main stages. First, the general panel-data model, key estimation techniques, and commonly used specification tests are described. Second, selected empirical economic studies that apply panel data methods are integrated and discussed. The overall literature review process followed in this research is summarised in Figure 1.



Source: Authors' elaboration based on Sabatés and Salas (2020).

Figure 1: Literature review process

The literature search was conducted using Google Scholar, Elsevier, Dialnet, and Redalyc databases. Keywords related to panel data in economic research were employed, with economic growth as the primary inclusion criterion. From this core theme, related subtopics such as unemployment, inflation, financial inclusion, poverty, and environmental factors were identified. In terms of temporal scope, the review includes studies published over the last five years.

Results

Panel data and the general model

Panel data, also referred to as longitudinal or grouped data, combine observations across two dimensions: a cross-sectional dimension (such as individuals, households, firms, regions, or countries) and a time dimension. In this structure, the same units are observed repeatedly over time, allowing variables to vary both across entities and across periods (Gujarati, 2003). The primary objective of panel data models is to account for unobserved heterogeneity, which may otherwise bias the estimated relationship between the dependent variable (Y) and the explanatory variables (X) (Perazzi & Merli, 2013).

A standard linear panel data regression model can be expressed as follows (Wooldridge, 2010):

$$Y_{it} = \alpha_{it} + \beta_n X_{it} + u_{it} \quad (1)$$

Where:

$i = 1 \ldots N$ (individuals) cross-section

$i = 1 \ldots T$ time series

Y_{it}: Dependent variable

X_{it}: Observation at the time t for the *i-th* independent variables

β_n: Slopes of the independent variables

α_{it}: Heterogeneity of unobservable variables, may be due to individual or time effects

u_{it}: Disturbances or error term

A key assumption for consistent estimation is that the explanatory variables are exogenous with respect to the error term, formally expressed as:

$$Cov\,(X_{it}, u_{it}) = 0 \quad (2)$$

Starting from this general specification, different panel data models arise depending on how the unobserved effect α_{it} is treated. In particular, panel data approaches differ depending on whether the effects are assumed to be fixed or

random. Each specification implies distinct assumptions and estimation techniques, which are discussed in the following sections.

Estimation techniques

Panel data structures can be classified as balanced or unbalanced and as short or long panels. In a balanced panel, each cross-sectional unit has the same number of time observations, whereas in an unbalanced panel, the number of observations differs across units. A panel is considered short when the number of cross-sectional units (N) exceeds the number of time periods (T), and long when the number of time periods is larger than the number of units ($T > N$) (Wooldridge, 2010; Gujarati & Porter, 2010).

Panel linear regression models can be estimated using various techniques, depending on the panel structure and the assumptions about unobserved heterogeneity. Fixed-effects estimators include the least-squares dichotomous variables (LSDV) approach, the within-group estimator, and the first-difference estimator. Alternatively, random effects estimators may be employed. Regardless of the chosen approach, panel-data estimators must satisfy the consistency property, which depends on appropriate assumptions regarding exogeneity and the error structure (Gujarati & Porter, 2010).

Pooled Ordinary Least Squares (OLS) Regression Model or Constant Coefficients

When implementing the MCO, certain assumptions of the classical linear regression model must be met. According to Gujarati (2003), they are: 1) the model is linear in the parameters. 2) The values of the independent variables (X_i) are fixed in repeated sampling, that is, they are not stochastic. 3) The average of the residual values is equal to zero, as shown in equation 3:

$$E\ (u_i\ |X_i) = 0 \tag{3}$$

4) The assumption of homoscedasticity refers to the variance being equal. 5) There is no autocorrelation between the disturbances. 6) The covariance between and = 0. 7) The number of observations is larger than the number of parameters to be estimated and the number of explanatory variables. 8) There is variability in the values. 9) The regression model is correctly specified, that is, the functional form is correct. 10) There is no perfect multicollinearity between the explanatory variables; that is, there are no perfectly linear relationships between the explanatory variables.

The MCO model with panel data is specified in equation 4, adapted from Gujarati and Porter (2010):

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \cdots + \beta_n X_{nit} + u_{it} \qquad i = 1 \ldots N;\ t = 1 \ldots T \tag{4}$$

Where:

Y_{it}: Dependent variable

$X_{2it} \ldots X_{nit}$: Independent variables

β_1: Constant intercept term

$\beta_2 \ldots \beta_n$: Estimation parameters of independent variables

u_{it}: disturbances or error term

Using OLS, observations are grouped, and a regression is estimated. The independent variables are not stochastic and, if they are, they are not correlated with the error term. It is assumed that the independent variables are strictly exogenous, that is, they do not depend on the values of the error or disturbance term (u_{it}), and these are typically distributed, with zero mean and constant variance as in equation 5 (Gujarati and Porter, 2010).

$$u_{it} \sim N(0, \sigma^2) \tag{5}$$

The disadvantage of OLS is that it hides individual-level unobserved heterogeneity, which induces autocorrelation; thus, the error term (uit) is correlated with certain independent variables in the model, violating the assumption of no correlation between X_{it} and u_{it}. For Gujarati and Porter (2010), autocorrelation refers to the correlation between members of observations ordered in time or space. The disturbance term is related to some observation and is influenced if: $u_{it} X_{it} u_{it}$

$$Cov\ (X_{it}, u_{it}) \neq 0 \tag{6}$$

The estimated coefficients will be biased and inconsistent (Gujarati, 2003). There are two terms, autocorrelation and serial autocorrelation. The latter refers to the lagged correlation between two series; however, in the econometric literature, they are often considered synonymous (Badii et al., 2014).

Circumstances may arise, such as a specification error in the model (e.g., the omission of important variables), an incorrect functional form, or an error in the handling or transformation of the data. For grouped models, the estimators are consistent if the slope coefficients are constant across individuals and the error term is uncorrelated with the independent variables. If they are correlated, a Robust Panel Corrected Standard Error Model known as EECP (Beck and Katz, 1995).

Least Squares Model with Fixed Effects Dichotomous Variable (LSCM)

The MCVD is achieved by assigning each cross-sectional unit its own dichotomous variable (i.e., binary or qualitative) or a fixed intercept value, while accounting for heterogeneity. It involves assigning values of 0 or 1 to each individual, as appropriate (Gujarati and Porter, 2010). The general model of dichotomous variables with differential intercept is presented in equation 7:

$$Y_{it} = \alpha_1 + \alpha_2 D_{2i} + \cdots + \alpha_n D_{ni} + \beta_2 X_{2it} + \cdots + \beta_n X_{nit} + u_{it} \quad (7)$$
$$i = 1 \ldots N;\ t = 1 \ldots T$$

Where:
Y_{it}: Dependent variable
$X_{2it} \ldots X_{nit}$: Independent variables
$D_{2it} \ldots D_{ni}$: Dichotomous, binary or qualitative variables.
$\beta_2 \ldots \beta_n$: Estimation parameters of independent variables
α_1: Intercept term for individual 1
α_2: Fixed effects estimators for each individual
u_{it}:Disturbances or error term

If a dichotomous variable has categories, it should be used to avoid perfect multicollinearity; however, many explanatory variables exhibit a high degree of collinearity (Gujarati, 2003). The term multicollinearity is attributed to Frisch, who described a perfect or exact linear relationship among one or more explanatory variables in a regression model (Gujarati and Porter, 2010). The term also includes the case of perfect multicollinearity, that is:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k \quad (8)$$

In a regression with explanatory variables, i.e., kX_1, $X_2 \ldots X_k$
Where:

$X_1 = 1$ for all observations

$\lambda_1, \lambda_2 \ldots \lambda_k$ Not all are simultaneously equal to zero

Multicollinearity can occur in panel data because variables have common trends, which often happens due to a decrease or increase over time at a similar rate, for example, in macroeconomic variables such as the GDP growth rate, or microeconomic variables such as household expenditure, consumption, etc.

When estimating using fixed-effects DLS, one must be careful not to include too many dichotomous variables, whether individual, interactive, or multiplicative. Two things can happen: 1) loss of degrees of freedom and 2) inducing multicollinearity, which could make parameter estimation difficult and lead to a lack of precision or accuracy. In addition, in some situations, the model may not identify

hidden variables or variables that do not change over time, such as gender or ethnicity (Gujarati and Porter, 2010).

Care must be taken with the error term, and the classical assumption of linear regression (assumption 4 of homoscedasticity) must be modified. It can be assumed that the model exhibits heteroscedasticity, which can be corrected using tests in specialised software (Rodríguez et al., 2017). Homoscedasticity/heteroscedasticity refers to the variance. Originally, the classical assumption states that the variance of each disturbance or error term is constant (Gujarati, 2003), that is:

$$E(u_i^2) = \sigma^2 \qquad i = 1, 2, \dots, n \tag{9}$$

When heteroscedasticity exists, it is symbolised as follows:

$$E(u_i^2) = \sigma_i^2 \tag{10}$$

The conditional variance increases as X increases; therefore, the variances are not constant and heteroscedasticity are present. There are various reasons why the variances of the error or disturbance term tend to be variable, for example, the presence of atypical data may occur, whether they are very small or large observations.

It is also attributed to the model not being correctly specified or to the omission of some important variables, as well as the incorrect transformation of the data (in the first differences) or the incorrect functional form. However, the advantage is that data collection techniques are constantly improving, which tends to reduce these variations (Gujarati, 2003; Arango and Hernández, 2017).

Fixed Effects Model (FEM)

In the FEM, each cross-sectional unit has its own fixed intercept value. Fixed effects mean that the intercept can differ across individuals, but does not vary over time. The applied OLS yields fixed-effects estimators (Wooldridge, 2010). The general model is presented in equation 11:

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \cdots + \beta_n X_{nit} + u_{it} \qquad i = 1 \dots N;\ t = 1 \dots T \tag{11}$$

Where:

Y_{it}: Dependent variable

$X_{2it} \dots X_{nit}$: Independent variables

β_{1i}: The subscript expresses that the intercept may vary due to the special characteristics of each individual i

$\beta_2 \ldots \beta_n$: Slope coefficients, which do not vary over time or across individuals.

u_{it}: Disturbances or error term

Gujarati and Porter (2010) state that there are within-group fixed effects and first difference estimators.

MEF within group

In the within-group fixed effects estimator, the individual fixed effect can be eliminated, the values of the dependent and independent variables of each individual are expressed as deviations from their means, that is, the mean sample values are obtained and subtracted from the individual values of the variables, known as values corrected by the mean. They are then grouped and an OLS regression is implemented. The general model is presented in equation 12 and was adapted from Gujarati and Porter (2010):

$$y_{it} = \beta_2 x_{2it} + \cdots + \beta_n x_{nit} + u_{it}$$
$$i = 1 \ldots N;\ t = 1 \ldots T \tag{12}$$

Where:

y_{it}: Mean value of the independent variable

$x_{2it} \ldots x_{nit}$: Values corrected by the mean

u_{it}: Disturbances or error term

In this equation there is no longer an intercept term. However, among the disadvantages is that the within-group fixed effects estimator usually presents very large variances, and the disturbance or error term can also be very large, generating higher standard errors. In addition, differentiating a variable eliminates its long-term component.

MEF of first differences

Another alternative is to integrate a method of first differences, by which, for each individual, successive differences of the variables are obtained. That is, the regression of the first differences of the dependent variable (Y) on the first differences of the independent variables (X). Equation 13 is presented below with the method:

$$\Delta Y_{it} = \beta_2 \Delta X_{2it} + \cdots + \beta_n \Delta X_{nit} + (u_{it} - u_{i,t-1})$$
$$i = 1 \ldots N;\ t = 1 \ldots T \tag{13}$$

Where:

Δ: First difference operator

($u_{it} - u_{i,t-1}$): The disturbance or term of errororiginal is replaced by the subtraction or difference between the current and previous values.

If the independent variables are exogenous, i.e. they do not depend on the values of the error or disturbance term (u_{it}), then the first-difference estimator is unbiased. However, the disadvantages are the same as those of within-group fixed-effects estimators, since time-invariant independent variables (e.g., gender, ethnicity) for an individual cancel out in first differences (Wooldridge, 2010).

Random Effects Model (MEFA) or Error Components Model (ECM)

The MEFA is expressed in equation 14 and was adapted from Gujarati and Porter (2010):

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \cdots + \beta_n X_{nit} + w_{it} \quad i = 1 \ldots N;\ t = 1 \ldots T \tag{14}$$

The common intercept represents the average value of all the cross-sectional intercepts and the intercept value for an individual is expressed as:

$$\beta_{1i} = \beta_1 + \varepsilon_i \tag{15}$$

Where:

β_{1i}: Random variable with a mean value β_1

ε_i: Random term of the individual intercept with respect to the average value of all cross-sectional intercepts, with zero mean and variance. σ^2_ε.

In the MEFA the intercept values are randomly extracted from a larger population of individuals, they have a common mean for β_1, the individual differences in the intercept values of each individual are expressed in the error term, now represented by ε_i, this term is not directly observable, which is why it is known as a latent or unobservable variable (Beltrán and Castro, 2020).

$$w_{it} = \varepsilon_i + u_{it} \tag{16}$$

Where:

ε_i: Cross-sectional error, also called individual-specific error

u_{it}: Combination of cross-sectional and time series error

w_{it}: Composite error term

The composite error term is composed of ε_i and u_{it}, its assumptions are:

$$\varepsilon_i \sim N\ (0, \sigma_\varepsilon^2) \tag{17}$$

$$u_{it} \sim N(0, \sigma_u^2) \tag{18}$$

$$var\ (w_{it}) = \sigma_\varepsilon^2 + \sigma_u^2 \tag{19}$$

Equation 19 shows that the error term is homoscedastic or has equal variance. The individual error components are not correlated with each other, they are not correlated with any independent variable, nor is there autocorrelation with the cross-sectional units, nor with the time series. But it may happen that as part of it is correlated with the independent variables, which would produce an inconsistent estimation of the regression coefficients, so the most appropriate method to generate efficient estimators is the Generalized Least Squares (GLS) method (Beltrán and Castro, 2020).

Advantages and disadvantages of using fixed and random effects models

The advantages of FEMs include that the intercept may differ between individuals. Each individual or transversal unit may have special characteristics on its own, therefore, to have different intercepts, dichotomous variables can be implemented using a DVM (Baltagi, 2005; Woldridge, 2010). Its use is appropriate when the individual intercept may be correlated with one or more independent variables. However, its disadvantages are that it is not possible to add constant variables such as gender or ethnic origin, since they are collinear with the subject-specific intercept. In addition, more degrees of freedom are required when the transversal units (N) are very large.

In the FSM, the advantages are that the intercept of an individual unit is randomly drawn from a larger population with a constant mean value, there is economy of degrees of freedom, since only the mean value of the intercept and its variance are estimated. The random intercept of each cross-sectional unit is not correlated with the independent ones. Unlike the FSM, constant variables such as gender and ethnicity can be added. However, as for its disadvantages, it is observed that since it is part of the FSM, it can be correlated with the independent variables, which would produce an inconsistent estimate (Gujarati and Porter, 2010).

Once the characteristics of each model have been identified, a choice must be made between one or the other based on certain criteria, for example, identifying whether one is working with a short or long panel, with constant variables that do not change over time or with those that do change, whether dichotomous variables will be integrated, reviewing the degrees of freedom, as well as carrying out certain specification tests, such as the Hausman and Breusch-Pagan (BP) test or the Lagrange Multiplier (LM) test, to choose the most suitable model according to the data being worked with (Arango and Hernández, 2017).

First, the correlation between the individual cross-sectional error and the independent variables must be verified. The fixed effects model is appropriate when there is correlation, while, in the absence of correlation, the ideal model is the

random effects model. Table 1 shows some characteristics to take into consideration when choosing between a fixed or random effects model for the panel.

Table 1: Criteria for choosing between a Fixed Effects Model and a Random Effects Model

Items	Fixed Effects (MEF)	Random Effects (MEFA)
Check if it is a long panel where the number of periods (*T*) is greater than the number of individuals (*N*).	Optimum	
A short panel that is presented when the number of cross-sectional individuals (*N*) is larger than the periods (*T*).	Optimum	
Check if the variables are invariant over time.	Optimum	
When the cross-sectional units come randomly from a larger sample.		Optimum
Check whether time-invariant variables have been explicitly implemented in the model.		Optimum

Source: Authors elaboration based on Gujarati and Porter (2010) and Carter et al. (2011).

Dynamic panel model

This model incorporates lagged effects of the dependent variable as an explanatory variable, capturing complex temporal dynamics. They are advanced econometric tools used to analyse panel data with temporal interdependencies. These models are useful when the dependent variable at a given time is influenced by its value in previous periods, allowing to study dynamics over time. A dynamic panel model can be expressed as:

$$Y_{it} = \alpha + \lambda Y_{it-1} + \beta X_{it} + \varepsilon_{it} \quad (20)$$

Where:
Y_{it}: Value of the dependent variable for the entity in period t.
$\lambda Y_{it\text{-}1}$: Effect of the dependent variable in the previous period.
X_{it}: Set of explanatory variables.
α: Constant.

ε_{it}: Error term, composed of an unobserved individual-specific effect (u_i) and an idiosyncratic error u_{it}

The dynamic model presents temporal persistence, captures how current conditions depend on previous states, common in time series studies. Data structure combines cross-sectional (units) and temporal (periods) data. Due to endogeneity, dynamic panel models require special methods to ensure consistent estimates. Among the most commonly used are:

1. Generalized Method of Moments (GMM)

Developed by Arellano and Bond (1991), this approach uses instrumental variables to address endogeneity. It has two main variants; 1) Differential GMM: it is based on the first difference of the equations to eliminate individual effects. 2) System GMM: It introduces equations in levels and in differences to improve statistical efficiency.

2. Maximum Likelihood Method

It is suitable in scenarios with few temporal observations and when normality assumptions are valid.

3. Quadratic Variation Estimators:

Designed to capture specific dynamics, but less common due to their complexity. It has advantages such as capturing temporal dynamics: 1) it models phenomena such as inertia or delays in the effect of variables. 2) Greater realism; it better reflects the evolutionary nature of economic, social and environmental systems. 3) Control of heterogeneity; unobserved individual effects are adequately managed, reducing bias.

Limitations include: 1) Computational complexity, especially with large databases and methods such as GMM. 2) Data requirements: Requires sufficient temporal observations and variation in variables. 3) Sensitivity to specifications; estimates may be sensitive to the model or instruments used.

These are models used in economics to analyse economic growth and fiscal and monetary policies. Finance is the study of persistence in asset returns or investment decisions. And the social sciences, such as the impact of educational policies over the years. The dynamic panel model is a powerful tool to address problems where temporal dependencies are essential. Although its implementation requires technical care, its results provide a deeper and more realistic understanding of the phenomena analysed (Baltagi, 2021).

Pooled OLS model

A pooled OLS model ignores the panel structure of the data and treats it as a cross-sectional data set with repeated observations. In other words, it assumes that there is no unobserved heterogeneity across units, or that it is irrelevant to the

estimation. A pooled OLS model can be estimated using ordinary least squares on the original variables. A pooled OLS model is simple and efficient, but it can produce biased and inconsistent estimates if there is unobserved heterogeneity that is correlated with the independent variables.

The pooled OLS model is a basic method for analysing panel data, which combines observations from all entities and periods as if they were a single cross-sectional data set. This approach implicitly assumes that there are no specific individual effects (unobserved heterogeneity) across entities, which implies that any differences between units are fully captured by the variables included in the model. The general equation of the model is:

$$Y_{it} = \beta 0 + \beta 1 X_{it} + \varepsilon_{it} \tag{21}$$

Where:

Y_{it}: dependent variable for the entity over time *t*.

X_{it}: set of explanatory variables.

$\beta 0$ common intercept.

$\beta 1$ coefficients of the explanatory variables.

ε_{it}: error term.

The pooled OLS model is useful when no significant differences between units are expected or when individual effects are assumed to be irrelevant. However, this simplification has important limitations. For example, ignoring entity-specific effects can cause the model to suffer from omitted variable bias if these unobserved differences are correlated with the included explanatory variables. Furthermore, in panel data, observations within the same entity over time are often correlated, which violates the assumption of independence of errors and can lead to incorrect standard errors, affecting statistical inference.

Despite these limitations, the pooled OLS model can be a useful starting point for exploratory analyses, as long as they are complemented with specification tests, such as the Breusch-Pagan or Hausman test, to decide whether a fixed or random effects model would be more appropriate. Therefore, although simple, the pooled OLS serves as an initial framework within the econometric analyses of panel data, with applications in studies where the entities do not present substantial heterogeneity (Mendoza and Quintana, 2016).

Panel cointegration model

The panel cointegration model is an extension of time series models applied to panel data, with the aim of analysing long-term relationships between variables that are non-stationary but move together over time. In this type of model, it is assumed that the time series of each unit in the panel can be first-order integrated,

that is, they have a unit root, but their linear combinations can be stationary, suggesting the existence of a long-term equilibrium relationship between the variables. To detect these relationships, cointegration tests adapted to panel data are used, such as the Pedroni test (1999) or the Kao test (1999), which allow evaluating whether the variables in the panel are cointegrated, that is, whether they have a long-term relationship that does not dissolve over time.

The main advantage of cointegration models in panel data is their ability to account for unit-specific heterogeneities in the panel while simultaneously modelling long-term dynamics, which improves the accuracy of estimates by allowing for a more appropriate interpretation of effects that persist over time. These models are used in a variety of applications, such as the analysis of the relationship between gross domestic product (GDP) and consumption, exports and investments, or the impact of long-term economic policies. However, these approaches require careful model specification, as cointegration implies that variables have a common trend, which could hide more complex relationships if not handled appropriately.

The general equation of the panel cointegration model is based on a long-run relationship between the non-stationary variables in the panel. This equation represents a linear combination of the variables that are cointegrated, that is, those whose joint behaviour produces a stationary series. A common way of writing this model is as follows:

$$Y_{it} = \alpha_i + \beta X_{it} + u_{it} \tag{22}$$

Where:

Y_{it}: Dependent variable for the unit in time t

X_{it}: Independent variable (or set of variables) for the unit in time t

α_i: Unit-specific fixed effect or intercept term i.

β: Coefficient that measures the long-term relationship between variables,

u_{it}: Error term, which in a cointegration model can be represented as a stationary combination of the error series of the panel units.

The model is based on the premise that the series are not stationary in their own right; their combination is stationary, indicating a cointegration relationship between them. This model can be extended and adjusted to capture more complexities, such as common trends or random effects variables, depending on the specific type of cointegration (such as panel cointegration with fixed or random effects).

Furthermore, panel cointegration models require an adequate number of periods and cross-sectional units to ensure the robustness of the results, given that the detection of cointegration relationships may be sensitive to the choice of the number of observations and the presence of common trends among the panel units (Baltagi, 2008; Pedroni, 1999).

Specification Tests

When using a panel data econometric model, certain specification tests must be implemented. According to Breusch and Pagan (1980) and Montero (2011), the Breusch-Pagan (BP) test, also known as the Lagrange Multiplier (LM), poses the null hypothesis:

$$\text{H0: } Var(u_i) = 0 \tag{23}$$

With a chi-square test, when the test value is $\chi^2 p > 0.95$ H0 and is confirmed using Ordinary Least Squares (OLS), in this case, if there is correlation, a Robust Panel Corrected Standard Error Model (RPSE) is used. On the contrary, if the test value is high ($p < 0.05$) H0 is rejected and it is feasible to use a panel data model.

To determine the effect, whether fixed or random, the Hausman test is carried out, in which, H0 states that the fixed and random effects estimators do not differ. If $p < 0.05$, the null hypothesis H_0 is rejected, and the fixed effects model is preferred (Hausman and Taylor, 1981; Gujarati and Porter, 2010; Chen, Yue, and Wu, 2018).

For its part, autocorrelation indicates the dependence of the disturbances (u_{it}) (Gujarati, 2003). To verify its presence, the Wooldridge test is implemented. In this test H0 expresses that there is no autocorrelation, on the contrary, the rejection of H0 confirms its presence (Rodríguez et al. (2017). Another way to verify the existence of autocorrelation in the model is to implement the Durbin-Watson statistic (Tillman, 1995), which states that values less than 1 and close to zero support the presence of autocorrelation.

Heteroscedasticity or unequal variances refers to the fact that the variance of the disturbances or errors of the sample represented by (u_{it}) is not constant (Gujarati, 2003). In the modified Wald test for fixed effects H0 expresses that there is no heteroscedasticity while the rejection of H0 confirms that there is heteroscedasticity. Likewise, a model that presents both autocorrelation and heteroscedasticity can be corrected (Barrera et al. 2021).

Panel data in the study of economic growth in Latin American economies

Panel data methods have been widely adopted in empirical economic research, supported both by their methodological advantages and by the increasing availability of large and detailed datasets. Advances in data processing and the diffusion of specialized software such as STATA, EViews, R, and similar platforms have further facilitated the application of panel techniques in growth-related studies. As a result, panel data have become a standard econometric tool in the analysis of economic growth trajectories in Latin American economies.

By combining cross-sectional and time-series information, panel data models enable researchers to examine how structural, institutional, and policy-related factors interact over time and across countries. These models explicitly control for unobserved heterogeneity and account for both spatial and temporal dimensions, yielding more consistent and efficient estimators than those obtained from purely cross-sectional or time-series approaches (Gujarati & Porter, 2010; Wooldridge, 2010). Consequently, panel data methods are particularly well suited to the institutional diversity and structural heterogeneity characteristic of Latin America.

Recent empirical studies (2020 – 2025) have relied extensively on panel data to investigate the determinants of economic growth in the region. Most analyses adopt economic growth as the dependent variable and incorporate macroeconomic and policy-related explanatory factors. Commonly used variables include inflation, foreign direct investment (FDI), research and development (R&D) expenditure (Arévalo et al., 2024), public expenditure (Pessino et al., 2022; Angulo et al., 2023), productivity, and competitiveness (Landa & Cerezo, 2024).

A growing strand of the literature emphasizes the role of institutions and governance in shaping growth outcomes. Panel data studies highlight the complex challenge of measuring institutional quality, political instability, rule of law, and democratic processes (Husnain et al., 2024; Topolewski, 2025). In this context, traditional macroeconomic variables such as FDI, inflation, and R&D are often interpreted as indirect proxies for institutional performance (Corrales & García, 2007). For example, Delbianco and Dabús (2023), analysing low-, medium-, and high-growth regimes in Latin America, find that instability and inequality exert a negative effect on economic performance, while economic openness is not consistently associated with growth.

Several studies confirm the positive contribution of institutional quality and investment-related variables to economic growth. Husnain et al. (2024) report that institutional quality, FDI, and domestic investment significantly enhance growth, whereas inflation has a detrimental effect. Similarly, Topolewski (2025) identifies institutions as central drivers of growth and as a key mechanism for poverty reduction in Latin America. Arévalo et al. (2024) show that capital formation, education expenditure, and taxation exert statistically significant effects on growth, while R&D expenditure displays a negative association in the regional context.

Fiscal policy and public finance dynamics have also been extensively examined using panel data models. Angulo et al. (2023) find that public expenditure and international trade positively influence economic growth, while inflation, unemployment, and economic crises exert negative effects. Complementary evidence from Pessino et al. (2022) suggests that sustained public investment, par-

ticularly in physical and human capital, contributes positively to growth. In line with these findings, Rojas, Silva, and Calderón (2021) demonstrate that capital expenditures – especially infrastructure investment – stimulate growth, whereas current expenditures such as subsidies and administrative spending show limited or no positive impact, highlighting the importance of expenditure composition rather than aggregate levels.

Structural transformation and productivity improvements represent another central theme in the panel data literature. Bittencourt, De Lima, and Cerqueira (2023), using fixed-effects and GMM estimators for 14 Latin American countries, show that shifts from agricultural employment towards industry and services positively affect GDP growth when supported by adequate institutional frameworks. Likewise, improvements in productivity and competitiveness are found to be key drivers of sustained growth (Landa & Cerezo, 2024).

Human capital accumulation has emerged as a pivotal determinant of long-term growth. Using panel cointegration techniques, Martínez and Pineda (2022) document a stable long-run relationship between public investment in education and GDP per capita growth in middle-income Latin American economies. Their dynamic panel results indicate that the growth effects of education spending intensify over time, particularly in countries with lower initial educational attainment.

More recent panel data applications have expanded the analysis to include financial inclusion and environmental sustainability. Zamora and Hernández (2024), employing random effects models, find that access to credit, digital payments, and mobile banking positively influences economic growth, especially in historically underserved regions. From an environmental perspective, Gómez and Pereira (2021) use System GMM to test the Environmental Kuznets Curve hypothesis in 17 Latin American economies, confirming an inverted U-shaped relationship between economic growth and CO_2 emissions.

Additional evidence reinforces the relevance of dynamic and heterogeneous panel approaches. Bazán – Navarro et al. (2024) analyse the bidirectional long-run relationship between electricity consumption and economic growth across 31 Latin American and Caribbean countries using dynamic panel techniques, highlighting feedback effects between infrastructure development and growth. Roquez-Díaz and Escot (2018) employ heterogeneous panel cointegration and causality methods to show that the growth-trade openness relationship varies substantially across countries, reflecting differences in institutional and structural contexts. Similarly, recent panel VAR evidence for Latin America and the Caribbean indicates that shocks to economic institutions exert stronger and more persistent effects on growth than political institutional shocks.

Taken together, this body of empirical research demonstrates the analytical strength of panel data methodologies – both static and dynamic – in capturing

the multifaceted and heterogeneous drivers of economic growth in Latin American economies. By accommodating unobserved heterogeneity, endogeneity, and long-run dynamics, panel data econometrics provides a robust framework for analysing growth processes in a region shaped by structural transformation, institutional diversity, and policy volatility.

Table 2: Use of Panel Data in Economic Growth Studies – Latin American Developing Economies

Study (Year)	Countries / Period	Panel Model & Estimator	Key Variables	Main Finding
Bazán-Navarro et al. (2024)	31 LAC countries, 1980 – 2021	Dynamic panel (System GMM/ feedback)	Electricity consumption, GDP growth	Bidirectional causality: 1% ↑ EC → 0.5% ↑ Growth; Growth → ↑ EC 1.54%
Roquez-Diaz & Escot (2018)	Latin American countries	Heterogeneous panel, cointegration, Granger causality	Trade openness, GDP growth	Heterogeneous causality: varies by country
Study in panel VAR (2024)	Latin America & Caribbean	Panel VAR	Institutional quality, growth	Economic institutions → growth; reciprocal long-run dynamics
Islam & Mondal (2023)	Four lower-middle-income LA countries, 1996–2019	Dynamic panel data (likely GMM)	Remittances, financial development	Remittances boost finance sector, supporting growth
Ahamed (2021)	39 developing countries (incl. LA), 1990 – 2019	Static/dynamic panel data	Public & private investment, labor	Public investment has stronger growth impact

Source: Authors' elaboration

Conclusions

The objective of this study was to provide a literature review of the use of panel data econometrics in economic research, employing an exploratory–descriptive approach. The review confirms that panel data methods are widely used in both theoretical and empirical studies, largely due to the growing availability of longitudinal datasets produced by international and national institutions, as well as advances in econometric software capable of efficiently processing large volumes of information.

The analysis shows that panel data modelling encompasses several extensions of the general linear specification, most notably pooled OLS models, fixed-effects approaches (including LSDV, within-group, and first-difference estimators), and random-effects or error-components models. The choice among these alternatives depends on the characteristics of the phenomenon under investigation, particularly the nature of unobserved heterogeneity and whether explanatory variables vary over time. Consequently, empirical research using panel data must focus on obtaining consistent coefficient estimates while carefully aligning model assumptions with the underlying economic context.

Compared with purely cross-sectional or time-series approaches, panel data models offer clear advantages for analysing dynamic economic processes. They are particularly well suited to studying phenomena such as unemployment, labour mobility, poverty, inflation, and economic growth, enabling researchers to capture both temporal dynamics and cross-sectional heterogeneity. Panel data applications have also been instrumental in evaluating policy interventions, including the effects of labour market regulations, minimum wage policies, and production changes on employment and growth outcomes.

Despite their advantages, implementing panel data techniques requires a clearly defined research objective and careful data management. Common empirical challenges – such as missing observations, coverage limitations, temporal biases, and measurement errors – remain relevant in panel settings, even with increasingly comprehensive data sources. From an econometric standpoint, panel datasets often exhibit serial correlation, heteroscedasticity, and multicollinearity, making appropriate specification and diagnostic tests essential.

Accordingly, empirical studies must routinely employ tests such as the Breusch-Pagan or Lagrange Multiplier tests, the Hausman test, and procedures for detecting serial correlation and heteroscedasticity (e.g., Wooldridge or Durbin-Watson-type tests), as well as robust inference methods when standard assumptions are violated. Equally important is the correct specification of the functional form of the model, as misspecification may undermine the validity of empirical results.

In conclusion, panel data econometrics constitutes a powerful analytical framework for quantitative economic research. Its effectiveness, however, ultimately depends on the researcher's ability to select an appropriate model, apply rigorous diagnostic testing, and interpret results within a coherent theoretical framework, ensuring the consistency and credibility of empirical findings.

References

Ahamed, M. M. (2021). Does public investment affect economic growth in developing countries? Empirical evidence from panel data, arXiv, https://arxiv.org/abs/2105.14199

Angulo, H., Florez, W., Calderon, V., Peña, R., Barrientos, M. & Zeballos, V. (2023). Determinants of Inclusive Economic Growth in Latin America, Wseas Transactions on Business and Economics, 20, pp. 1059-1073, https://doi.org/10.37394/23207.2023.20.96

Arango, D. & Hernández, F. (2017). El impacto de especificar incorrectamente la distribución de los efectos aleatorios en las estimaciones de modelos lineales generalizados mixtos, Comunicaciones en estadística, 10(2), pp. 247-280, disponible en: http://dx.doi.org/10.15332/s2027-3355.2017.0002.04

Arellano, M. & Bover, O. (1990). La econometría de datos panel, Investigaciones económicas, 14(1), pp. 3-45, disponible en: https://www.fundacionsepi.es/investigacion/revistas/paperArchive/Ene1990/v14i1a1.pdf

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, The Review of Economic Studies, 58(2), pp. 277-297, https://doi.org/10.2307/2297968

Arévalo, J. A., Rodas, L., Ruiz, J., Moreno, L. R., Atoche, R., Arévalo, V., Gonzales, M., & Agueda, L. (2024). Effects of Public Spending on Economic Growth: An empirical approach in Latin American countries. Period 2006 – 2019, International Journal of Religion, 5(6), pp. 793-807, https://doi.org/10.61707/804ykt02

Baltagi, B. (2005). Econometric Analysis of Panel Data, 3rd ed., New York: John Wiley.

Badii, M., Guillen, O., Lugo, S. & Aguilar, C. (2014). Correlación no paramétrica y su aplicación en la investigación científica, International Journal of Good Conscience, 9(2), pp. 31-40, disponible en: http://www.spentamexico.org/v9-n2/A5.9(2)31-40.pdf

Bazán-Navarro, S., Ríos-Estepa, R., & Yepes-Cubillos, D. C. (2024). Electricity consumption and economic growth in Latin America and the Caribbean: Evidence from a dynamic panel data analysis, Energy & Environment, https://doi.org/10.1177/0958305X231188222

Barrera, P., Navarrete, J.A. & Segura, E. (2021). Análisis del emprendimiento en México a través de datos panel, En: 25 Congreso Internacional de Ciencias Administrativas, Ciudad de México: Universidad Nacional Autónoma de México.

Beck, N. & Katz, J. (1995). What to do (and not to do) with Time-Series Cross-Section Data, The American Political Science Review, 89(3), pp. 634-647, https://doi.org/10.2307/2082979

Beltrán, A. & Castro, J. (2020). Modelos de datos de panel y variables dependientes limitadas: teoría y práctica, Lima: Universidad del Pacífico.

Bittencourt, M., De Lima, R., & Cerqueira, P. A. (2023). Structural transformation and growth in Latin America: A panel data analysis, Economic Analysis and Policy, 80, pp. 101-112, https://doi.org/10.1016/j.eap.2023.02.005

Breusch, T.S. & Pagan, A.R. (1980). The Lagrange multiplier test and its application to model specification in Econometrics, The Review of Economic Studies, 47(1), pp. 239-253, https://doi.org/10.2307/2297111

Carter, R., Griffiths, W. & Lim, G. (2011). Principles of econometrics, 4th ed., New York: Wiley.

Chen, J., Yue, R. & Wu, J. (2018). Hausman-type test for individual and time effects in the panel regression model with incomplete data, Journal of Korean Statistical Society, https://doi.org/10.1016/j.jkss.2018.04.002

Corrales, J. P., & Iragorri, A. G. (2007). Crecimiento e instituciones en América Latina: Un análisis de series temporales agrupadas y cruzadas (1951 – 1999), Revista de Economía del Caribe, pp. 46-77, https://doi.org/10.14482/ecoca.01.121.005

Correa, J.C. & Salazar, C. (2016). Introducción a los modelos mixtos, Bogotá: Universidad Nacional de Colombia.

Delbianco, F. & Dabús, C. (2023). Economic growth regimes: Evidence from Latin America, Cuadernos de Economía, 42(89), pp. 129-146, https://doi.org/10.15446/cuad.econ.v42n89.94817

Guirao, S. (2015). Utilidad y tipos de revisión de la literatura, Index de Enfermería, 9(2), https://doi.org/10.4321/S1988-348X2015000200002

Gujarati, D. (2003). Econometría Básica, 4ª ed., México: McGraw-Hill.

Gujarati, D. & Porter, D. (2010). Econometría, 5ª ed., México: McGraw-Hill.

Gómez-Peña, J., & Pereira, M. (2021). Environmental degradation and economic growth in Latin America: A dynamic panel approach, Ecological Economics, 182, 106928, https://doi.org/10.1016/j.ecolecon.2021.106928

Hausman, J.A. & Taylor, W.E. (1981). Panel Data and Unobservable Individual Effects, Econometrica, 49(6), pp. 1377-1398, https://doi.org/10.2307/1911406

Husnain, M. A., Guo, P., Pan, G., & Manjang, M. (2024). Unveiling the Interplay of Institutional Quality, Foreign Direct Investment, Inflation and Domestic Investment on Economic Growth: Empirical Evidence for Latin America, International Journal of Economics and Financial Issues, 14(1), pp. 85-94, https://doi.org/10.32479/ijefi.15580

Islam, M. R., & Mondal, M. A. (2023). The effect of remittances on financial development in Latin America and the Caribbean: Evidence from dynamic panel data, arXiv, https://arxiv.org/abs/2309.08855

Landa, H. O., & Cerezo, V. (2024). Tipo de cambio real, innovación y crecimiento económico: un análisis comparativo para América Latina y Asia, Investigación Económica, 83(328), pp. 5-30, https://doi.org/10.22201/fe.01851667p.2024.328.87179

Martínez, L., & Pineda, R. (2022). Education investment and economic growth in Latin America: Evidence from panel cointegration, Latin American Economic Review, 31(1), pp. 202-223, https://doi.org/10.1186/s40503-022-00115-7

Mendoza, C. & Quintana, J. (2016). Modelos de datos de panel en análisis económico: una aplicación empírica, Revista de Econometría Aplicada, 32(1), pp. 45-67.

Montero, R. (2011). Efectos fijos o aleatorios: test de especificación. Documentos de Trabajo en Economía Aplicada, Universidad de Granada, España.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors, Oxford Bulletin of Economics and Statistics, 61(S1), pp. 653-670.

Perazzi, J. & Merli, G. (2013). Modelos de regresión de datos panel y su aplicación en la evaluación de programas sociales, Telos, 15(1), pp. 119-127, http://www.redalyc.org/articulo.oa?id=99326637008

Pessino, C., Altinok, N., y Chagalj, C. (2022). Eficiencia en la asignación del gasto público para el crecimiento en países latinoamericanos, https://doi.org/10.18235/0004310

Rodríguez, M., Freire, M. & País, C. (2017). El efecto del gasto público sanitario y educativo en la determinación del bienestar de los países de la OCDE: un modelo con datos de panel, Cuadernos de Economía, 41, pp. 104-118, https://doi.org/10.1016/j.cesjef.2017.05.001

Rojas, J. C., Silva, M., & Calderón, C. (2021). Public expenditure and economic growth: A panel data study of Latin American countries, Journal of Policy Modeling, 43(1), pp. 78-95, https://doi.org/10.1016/j.jpolmod.2020.09.001

Roquez-Díaz, D. J., & Escot, L. (2018). Trade openness and economic growth in Latin America: Evidence from heterogeneous panel data, Cogent Economics & Finance, 6(1), 1449780, https://doi.org/10.1080/23322039.2018.1449780

Tillman, J.A. (1975). The power of the Durbin Watson Test, Econometrica, 43(5/6), pp. 959-974, https://doi.org/10.2307/1911337.

Topolewski, L. (2025). An empirical analysis of the impact of institutions on economic growth: Evidence from Latin American countries, Argumenta Oeconomica, 54(1), pp. 126-136, doi:10.15611/aoe.2025.1.08

Wooldridge, J. (2010). Econometric Analysis of Cross Section and Panel Data, Cambridge: MIT Press.

Zamora, F., & Hernández, A. (2024). Financial inclusion and regional growth in Latin America: Evidence from a panel data approach, International Review of Economics & Finance, 89, pp. 53-67, https://doi.org/10.1016/j.iref.2024.02.006