Vol. 13. Issue.2. 2025 (April-June)
©KY PUBLICATIONS

http://www.bomsr.com
Email:editorbomsr@gmail.com

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

Econometric Models for Socioeconomic Development and Planning

Dila Ram Bhandari[1], Dr. Uma Srivastava[2]
[1]PhD scholar, Deen Dayal Upadhya Gorakhpur University, India &
Lecturer, Nepal Commerce Campus, Tribhuvan University, Kathmandu, Nepal
[2]Professor, Deen Dayal Upadhya Gorakhpur University, UP, India

DOI:10.33329/bomsr.13.2.16

ISSN:2348-0580 BOMSR BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH www.bomsr.com

Article Info
Article Received: 20/04/2025
Article Accepted: 24/05/2025
Published online: 16/06/2025

ABSTRACT

Econometric models have been widely used in analyzing and predicting socioeconomic development through mapping relationships between economic variables and social phenomenon. This paper reviews the major categories of econometric models' linear regression, panel data models, time series analysis, and structural equation modeling employed to analyze patterns of growth, distribution of income, alleviation of poverty, achievement in education, and improvements in health and other factors. Econometric models are indispensable instruments of analysis, prediction and management of socioeconomic development. It discusses model formulation, estimation procedures and the problems of multicollinearity, and observation errors that tend to be present in socio-economic data sets. By applying econometric methods, this framework allows policymakers and researchers to improve the design, implementation, and evaluation of growth strategies aimed at promoting inclusive and sustainable economic progress. Using simulated and secondary data, we demonstrate how these models quantify the impact of economic policies on growth, employment, and inequality. The study reveals that properly specified econometric models provide robust insights that can inform national and regional planning, especially in developing countries.

Keywords: growth, poverty, regression, econometric models, policymaker .

1. Introduction

Socioeconomic development and rational ways of space allocation are essential to achieve sustainable development and to enhance the quality of life. Econometric models are important tools in explaining and forecasting the behavior of socioeconomic systems. The relationship between different economic variable and social variable may be well analysed using these models and may aid the policy makers and planners to take decisions. Econometric models have a history of applications in social and economic development and planning, starting with the pioneering work of researchers and applied workers of the user's community.

Quantitative models are used extensively by governments, NGOs, and international institutions to evaluate policy effectiveness, track progress, and distribute resources. At the heart of this process is econometrics, a discipline that fuses economic theory with mathematics and statistical inference. It enables data-driven policymaking by establishing causal links between variables like GDP, employment, inflation, and public expenditure (Gujarati & Porter, 2009). In developing countries, econometric modeling is becoming increasingly acknowledged as an essential part of evidence-based planning (Wooldridge, 2016) which simplified representation of reality, created to generate hypotheses about economic behavior that can be tested. A key characteristic of an economic model is that it must be subjective in design, as there are no objective measures of economic outcomes. There are two economic models theoretical and empirical. Under the assumption that agents maximize certain objectives while adhering to model-defined constraints, theoretical models aim to derive verifiable implications about economic behavior. They offer qualitative responses to targeted inquiries, including the consequences of asymmetric information or the optimal approach for managing market failures. On the other hand, the goal of empirical models is to confirm the qualitative predictions made by theoretical models and to translate these predictions into exact numerical results. A theoretical model of an agent's consumption behavior, for instance, would typically indicate that expenditure and income are positively related. In an empirical adaptation of the theoretical model, the goal would be to determine a numerical value for the average increase in expenditure that accompanies an increase in income. Main Sectors in Socioeconomic Development Models

- **Agriculture:** Crop production, irrigation, land use, agricultural employment
- **Tourism:** Foreign Exchange and Investment, Employment Generation, GDP
- **Health:** Life expectancy, infant mortality, access to healthcare, nutrition
- **Education:** Literacy rates, school enrollment, years of schooling
- **Employment & Labor:** Unemployment rate, labor force participation, informal sector share
- **Industry & Manufacturing:** Industrial output, employment in manufacturing, investment

- **Infrastructure:** Roads, electricity access, water supply, internet connectivity
- **Finance & Banking:** Access to credit, financial inclusion, microfinance availability
- **Population & Demographics:** Population growth, migration, urbanization
- **Environment & Natural Resources:** Deforestation, water resources, climate impact
- **Governance & Institutions:** Political stability, corruption indices, quality of institutions
- **Trade & Foreign Investment:** Export/import levels, FDI inflow, trade openness
- **Technology & Innovation:** R&D spending, technology adoption, digital penetration

Tourism is a key engine of inclusive growth and is frequently included in socioeconomic models to measure its impact on income, employment, and development. In developing countries like Nepal, its role is especially important due to the country's natural beauty, heritage sites, and adventure tourism appeal. Its presence is justified by its significant contributions to national income, employment, foreign exchange earnings, and regional development. In countries with rich natural and cultural heritage, tourism serves as a major source of revenue and plays a vital role in poverty alleviation and infrastructure enhancement.

Economically, tourism boosts GDP directly through spending by international visitors and indirectly through increased demand for local goods and services. It generates employment in various sectors such as hospitality, transport, retail, and handicrafts. Moreover, it often leads to the development of supporting infrastructure like roads, airports, and communication networks, which benefit other sectors of the economy as well.

In econometric modeling, tourism is typically treated as an independent variable that can influence key development indicators such as GDP per capita, employment rates, and investment flows. Models might include variables like international tourist arrivals, tourism receipts, employment in the tourism sector, or tourism's share of GDP. For example, a regression model could be used to assess the impact of tourism receipts on GDP growth or to explore the relationship between tourist arrivals and employment levels. Tourism is also seen as a tool for regional and rural development. It encourages investment in areas that may otherwise be economically neglected, helping to reduce regional disparities. In the context of Nepal, tourism plays a crucial role in supporting local economies in mountainous and culturally rich regions. Overall, the inclusion of tourism in econometric models helps policymakers understand its potential as a driver of inclusive and sustainable economic growth.

Mathematical equations that represent a theory of economic behavior typically make up economic models. Model builders aim to incorporate enough equations to offer valuable insights into the behavior of the functioning of an economy. Econometric models have become fundamental for examining socioeconomic development, providing a systematic and quantitative framework for understanding intricate economic connections. These models facilitate the examination of patterns, hypothesis testing, and outcome prediction concerning income, education, employment, health, and poverty (Wooldridge, 2020) by combining economic theory with mathematics and statistical methods. Socioeconomic development, characterized by improvements in living standards, economic growth, and social equity, relies heavily on evidence-based policies, and econometric modeling provides the empirical foundation for crafting such policies (Gujarati & Porter, 2009).

Various econometric methods, including multiple regression analysis, time series models, panel data techniques, and simultaneous equation systems, facilitate the identification of crucial determinants of development and the assessment of intervention programs (Greene, 2018). Furthermore, the development of computational tools and the access to extensive datasets have improved the accuracy and dependability of econometric analyses, increasing their importance in tackling contemporary development issues (Stock & Watson, 2020). This paper examines how econometric models can be applied to foster sustainable and inclusive socioeconomic development, the methods used in these applications, and their importance.

Each empirical model possesses coefficients that dictate the variation of a dependent variable in response to changes in an input (how household consumption reacts to a $100 reduction in income tax) and such coefficients are typically calculated using historical data. Reality can never be perfectly captured by any economic model, however, the act of building, testing, and refining models compels economists and policymakers to sharpen their perspectives on the functioning of an economy.

2. Objective of Study

Econometric models in socioeconomic development and planning aim to quantitatively analyze relationships among key economic and social functions. This analysis helps policymakers forecast future trends, assess the effects of current and proposed policies, and devise effective sustainable development strategies. The objective of these models is to furnish empirical proof that can aid in making choices regarding employment, poverty alleviation, education, health care, business development, and income distribution.

3. Significance of Study

Evidence-Based Policy Formulation By using econometric models, planners can ground their decisions in data and statistical evidence, which diminishes guesswork and enhances the quality of public policy. The accurate forecasting of economic variables like GDP growth, inflation, investment, and unemployment basic for long-term development planning is made possible by these models. Through econometric analysis, limited resources can be allocated more effectively by pinpointing the areas or sectors that will benefit the most from interventions. This process also involves Measuring Policy Impact, Resource Allocation, Understanding Development Dynamics, and Monitoring Progress on Sustainable Development Goals (SDGs).

4. Literature Review

Econometric models have played a crucial role in enhancing comprehension of and progress toward socioeconomic development in developing countries. Numerous studies highlight that econometric modeling grants policymakers to assess poverty alleviation programs, educational reforms, health interventions, and strategies for generating employment (Todaro & Smith, 2020). Regression-based models are frequently used in developing countries to evaluate how public spending, infrastructure initiatives, and foreign direct investment affect GDP growth and income distribution (Chenery & Syrquin, 1975).

Models of time series econometrics, like ARIMA and VAR, have been widely utilized for predicting macroeconomic indicators such as inflation, exchange rates, and agricultural production, offering crucial data for national planning (Asteriou & Hall, 2015). Similarly, ARDL (Autoregressive Distributed Lag) models are widely used to examine the long-term and short-term dynamics of socioeconomic variables in African and South Asian economies (Nkoro & Uko, 2016).

As researchers seek to account for both cross-sectional and time-series variations across various regions and sectors within a country, panel data econometrics has also risen in prominence (Baltagi, 2008). Applications frequently entail assessing the effects of educational attainment on labor market outcomes or health expenditures on life expectancy. Additionally, simultaneous equation models assist in tackling endogeneity issues that are prevalent in development studies, like the case when investment and growth have reciprocal effects on one another (Wooldridge, 2020).

Even though they are useful, econometric practices in developing countries encounter problems like data scarcity, measurement errors, and structural breaks due to political instability or natural disasters (Cameron & Trivedi, 2005). Consequently, to address uncertainty and small sample sizes, researchers frequently use strong methods like instrumental variable techniques or Bayesian econometrics (Koop, 2003).

Recent developments, like the combination of machine learning with conventional econometric approaches, present promising avenues for improving the predictive accuracy and interpretability of socioeconomic models in developing contexts (Varian, 2014). Nonetheless, the effective use of these models hinges on the development of local capacity, backing from institutions, and access to trustworthy, disaggregated data.

5. Data and Methodology

To achieve effective econometric modeling for socioeconomic development, it is crucial to follow certain best practices that undertaking accuracy, relevance, and usefulness for policymaking. To prevent spurious relationships and enhance interpretability, it is crucial for models to be based on robust economic theory (Wooldridge, 2020). It is crucial to collect high-quality data which entails using reliable, timely, and disaggregated data to accurately reflect the true dynamics of development indicators (Todaro & Smith, 2020). It is essential to conduct diagnostic tests (such as for multicollinearity, heteroskedasticity, and autocorrelation) on a regular basis to confirm model assumptions and prevent biased estimates (Gujarati & Porter, 2009). The models need to incorporate country-specific factors like informal economies, governance quality, and social disparities to provide context-specific insights (Cameron & Trivedi, 2005). Finally, to uphold credibility and effectiveness in policy application, it is crucial to be transparent about methodology and to update and validate models regularly. Integrating traditional econometrics with machine learning and spatial analysis can bolster the forecasting power and relevance within a development context (Varian, 2014).

5.1 Data Sources (Econometric Models for Socioeconomic Development)

Secondary datasets obtained from the Nepal Tourism Statistics 2023, Ministry of Culture, Tourism and Civil Aviation 2000 to 2020. Key variables include GDP growth, education expenditure, employment rates, and poverty headcount ratios. Various econometric models are frequently utilized to examine and bolster socioeconomic development. Linear regression models are used to assess the effects of variables such as education, health, and infrastructure on economic growth (Wooldridge, 2020). Models of time series, including ARIMA, are employed to predict metrics such as GDP, inflation, and unemployment, thereby aiding in the formulation of future policy decisions (Asteriou & Hall, 2015). Panel data models facilitate the simultaneous analysis of cross-sectional and time-series data, providing insights into regional disparities and long-term development trends (Baltagi, 2008). When variables such as income and investment affect one another, simultaneous equation models are crucial for dealing with endogeneity problems (Gujarati & Porter, 2009). Furthermore, Logit and Probit models are employed to analyze binary outcomes like healthcare access or employment status. These models, when used properly, help policymakers. Econometric models are pivotal in socioeconomic development and planning, offering quantitative tools to evaluate policies and forecast trends. The data is analyzed by R studio.

Econometric models are widely applied in socioeconomic planning to quantify relationships between key variables such as income, employment, investment, and policy interventions. These models assist governments and international organizations in forecasting, impact evaluation, and evidence-based policymaking.

5.1.1. **Linear Regression Models in Poverty and Education Planning**

A common practice is using multiple linear regression to estimate the impact of education levels and employment on poverty rates and the relationship may be modeled as:

$$PovertyRate = \beta_0 + \beta_1 Education + \beta_2 Employment + \epsilon \qquad ....(1)$$

Where

- β_0 is the intercept,
- β_1, β_2 are coefficients,
- ϵ is the error term.

This model helps policymakers recognize which variables significantly affect poverty reduction and can enlighten education funding policies (Gujarati & Porter, 2009).

5.1.2. **Time Series Models for GDP and Inflation Forecasting**

ARIMA (Autoregressive Integrated Moving Average) models are widely used in GDP forecasting. For instance, in Nepal, ARIMA models have been employed to forecast macroeconomic indicators like GDP and inflation to align with Five-Year Plans.

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \qquad ....(2)$$

Where:

- Y_t is the value of the series at time t,
- p and q are the orders of the AR and MA parts.

These forecasts assist in allying national budgets with expected economic conditions (Box et al., 2015).

5.1.3. **Labor and Trade Policy of Simultaneous Models**

Simultaneous equation models help policymakers estimate the mutuality between labor force participation, wages, and trade openness. The models as following:

$$\begin{aligned} Wages &= \alpha_1 + \beta_1 Trade + \gamma_1 Employment + u_1 \\ Employment &= \alpha_2 + \beta_2 Wages + \gamma_2 GDP + u_2 \end{aligned} \quad \text{.........................(3)}$$

This structure reveals real-world feedback loops and is solved using Two-Stage Least Squares (2SLS) to avoid endogeneity (Wooldridge, 2016).

5.1.4. **Cobb-Douglas Production Function**

This model analyzes economic growth by combining output to capital, labor, and technology (Cobb & Douglas, 1928):

$$Y_t = A_t K_t^{\alpha} L_t^{\beta} \quad \text{..................(4)}$$

where Y_t is output, K_t is capital, L_t is labor, A_t is total factor productivity, and α, β are output elasticities. Policymakers use this to assess how investments in infrastructure ($\uparrow K$) or education ($\uparrow L$) drive growth.

5.1.5. **Vector Autoregression (VAR)**

VAR models detention interdependencies among multiple time series variables (Sims, 1980) for two variables Y_t (GDP) and X_t (education spending):

$$\begin{cases} Y_t = \alpha_1 + \beta_{11} Y_{t-1} + \beta_{12} X_{t-1} + \varepsilon_{1t} \\ X_t = \alpha_2 + \beta_{21} Y_{t-1} + \beta_{22} X_{t-1} + \varepsilon_{2t} \end{cases} \quad \text{...................(5)}$$

Authorities use VAR to forecast how shocks to education spending affect GDP and vice versa.

5.1.6. **Spatial Econometric Models**

These models description for geographic spillovers in regional planning (Anselin, 1988). A spatial lag model is:

$$Y = \rho WY + X\beta + \varepsilon \quad \text{...................(7)}$$

Where W is a spatial weight matrix, and ρ captures spillover effects used to evaluate infrastructure projects' regional impacts.

5.1.7. **Cointegration and Error Correction**

Models long-run equilibrium bonds, like GDP and energy consumption (Engle & Granger, 1987):

$$\Delta Y_t = \alpha + \beta \Delta X_t + \gamma (Y_{t-1} - \theta X_{t-1}) + \varepsilon_t \quad \text{................(8)}$$

The term $(Y_{t-1} - \theta X_{t-1})$ corrects short-term deviations from long-term trends, guiding sustainable planning.

5.1.8. **Poverty Indices**: **Foster-Greer-Thorbecke (FGT)**

Measures poverty for pointed social agendas (Foster et al., 1984):

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{z-y_i}{z}\right)^\alpha I(y_i < z) \quad \text{..................(9)}$$

Where z is the poverty line and α adjusts sensitivity to depth ($\alpha = 1$) or severity ($\alpha = 2$).

5.1.9. **Dynamic Stochastic General Equilibrium (DSGE)**

DSGE used by central banks for macroeconomic planning (Smets & Wouters, 2007). A consumption Euler equation:

$$\frac{1}{C_t} = \beta E_t\left[\frac{1}{C_{t+1}}(1 + r_{t+1})\right] \quad \text{......................(10)}$$

Policymakers simulate consequences like tax changes or monetary shocks to adjust stability. These models facilitate evidence-based policymaking, balancing theoretical rigor with practical socio-economic challenges.

5.1.10. **Panel Data Models**

Panel data regression model is used to assess policy impacts across regions over time. For analyzing longitudinal data on development indicators: (Baltagi, 2005; Wooldridge, 2010). Such models have been used by UNDP and World Bank projects across Sub-Saharan Africa and South Asia (Baltagi, 2008). The fixed effects models are applied to evaluate the effect of health infrastructure on life expectancy across multiple provinces:

$$LifeExpectancy_{it} = \beta_0 + \beta_1 HealthSpend_{it} + \mu_i + \lambda_t + \epsilon_{it} \quad \text{...................(11)}$$

Where:

- μ_i province-specific effects,
- λ_t accounts for year-specific effects.

5.1.11. **Instrumental Variables (IV)**

To address endogeneity in development variables: (Angrist & Krueger, 2001). **First stage**:

$$\mathrm{Educ}_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \nu_i \quad \text{................... (12)}$$

Second stage:

$$\mathrm{GDPpc}_i = \beta_0 + \beta_1 \widehat{\mathrm{Educ}}_i + \beta_2 X_i + \epsilon_i \quad \text{................... (13)}$$

Where Z_i is an instrument (e.g., historical school construction).

5.1.12. **Structural Equation Modeling (SEM*):*** Models' latent constructs ("institutional quality") with measurement equations (Kline, 2015).

$$\mathrm{Development}_i = \gamma_1 \mathrm{Economy}_i + \gamma_2 \mathrm{SocialWelfare}_i + \zeta_i \quad \text{...................(14)}$$

$$\mathrm{Economy}_i = \lambda_1 \mathrm{GDPpc}_i + \lambda_2 \mathrm{Unemp}_i \quad \text{................... (15)}$$

$$\mathrm{Social\ Welfare}_i = \lambda_3 \mathrm{LifeExpectancy}_i + \lambda_4 \mathrm{Education}_i \quad \text{................... (16)}$$

5.1.13. **Dynamic Panel Models (GMM):** Uses lagged variables as instruments to address autocorrelation (Arellano & Bond, 1991).

$$\mathrm{Poverty}_{it} = \alpha \mathrm{Poverty}_{i,t-1} + \beta_1 \mathrm{GDPGrowth}_{it} + \beta_2 \mathrm{SocialSpending}_{it} + \eta_i + \epsilon_{it} \quad \text{................... (17)}$$

5.1.14. **Spatial Econometrics:** Captures spillovers (neighboring regions' policies) via spatial weights matrix W (Anselin, 1988).

$$Y_i = \rho W Y_j + \beta X_i + \epsilon_i \quad \text{....................(18)}$$

5.1.15. **Cross-Sectional Regression Model)**

A basic linear regression to study determinants of socioeconomic development across regions or countries: (Barro, 1991; Temple, 1999).

$$\text{GDPpc}_i = \beta_0 + \beta_1 \text{Educ}_i + \beta_2 \text{HealthExp}_i + \beta_3 \text{Infra}_i + \epsilon_i \quad \text{..............................(19)}$$

Where

- GDPpc_i: GDP per capita (proxy for development) in region i,
- Educ_i: Education attainment (e.g., literacy rate),
- HealthExp_i: Health expenditure (% of GDP),
- Infra_i: Infrastructure index.

5. 2 Model Specification

We apply four types of econometric models:

- Multiple Linear Regression (MLR): To examine the impact of education and employment on poverty:
- $PovertyRate = \beta_0 + \beta_1 Education + \beta_2 Employment + \epsilon$
- ARIMA Time Series Model**:** For GDP forecasting:
- $Y_t = \alpha + \sum_{i=1}^{p} \phi_i \, Y_{t-i} + \sum_{j=1}^{q} \theta_j \, \epsilon_{t-j} + \epsilon_t$
- Simultaneous Equation Model**:** To capture the interaction between wages, employment, and trade:
- $\begin{aligned} Wages &= \alpha_1 + \beta_1 Trade + \gamma_1 Employment + u_1 \\ Employment &= \alpha_2 + \beta_2 Wages + \gamma_2 GDP + u_2 \end{aligned}$
- Panel Data Model**:** To assess regional disparities in life expectancy using fixed effects:

$$LifeExpectancy_{it} = \beta_0 + \beta_1 HealthSpend_{it} + \mu_i + \lambda_t + \epsilon_{it}$$

Nepal, nestled in the Himalayas, is a premier destination for adventure, culture, and nature enthusiasts. Tourism in Nepal is a vital driver of economic and social development, offering unparalleled natural and cultural experiences. However, sustainable practices are crucial to ensure its long-term benefits for both the country and its visitors. ARIMA is a powerful and flexible tool for forecasting tourist arrivals, especially when the data is non-seasonal and shows a trend. It consents tourism stakeholders to anticipate future demand, helping in better resource and policy planning. RIMA is a widely used statistical method for time series forecasting, especially effective for datasets that show evidence of non-stationarity, such as tourist arrival data. Tourist arrival data frequently display trends or seasonality, making ARIMA suitable for forecasting helps tourism plan resources, hotels manage capacity, and policymakers make informed decisions, and monthly or yearly tourist data is modeled and used to predict future trends. The published secondary data collected from

Nepal Tourism Board, Nepal Government from 2001 to 2022 and analyzed using R programming where the variables number of tourists arrival (male and female), GDP, growth rate, average earning, merchandise and length of stay.

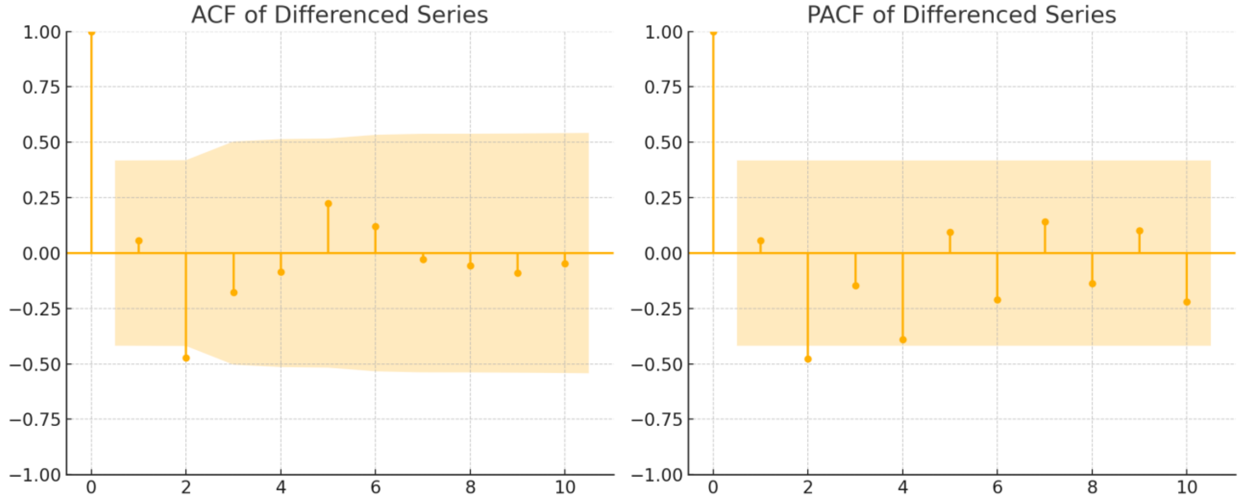

Figure 1: ACF and PACF

- The ACF plot shows significant autocorrelation at lag 1, then quickly drops, suggesting the presence of an MA (1) or short memory process.
- The PACF plot shows significance at lag 1 and 2, suggesting an AR (2) component.

This supports the chosen ARIMA (2,1,0) model, as it captures the autoregressive pattern in the differenced data.

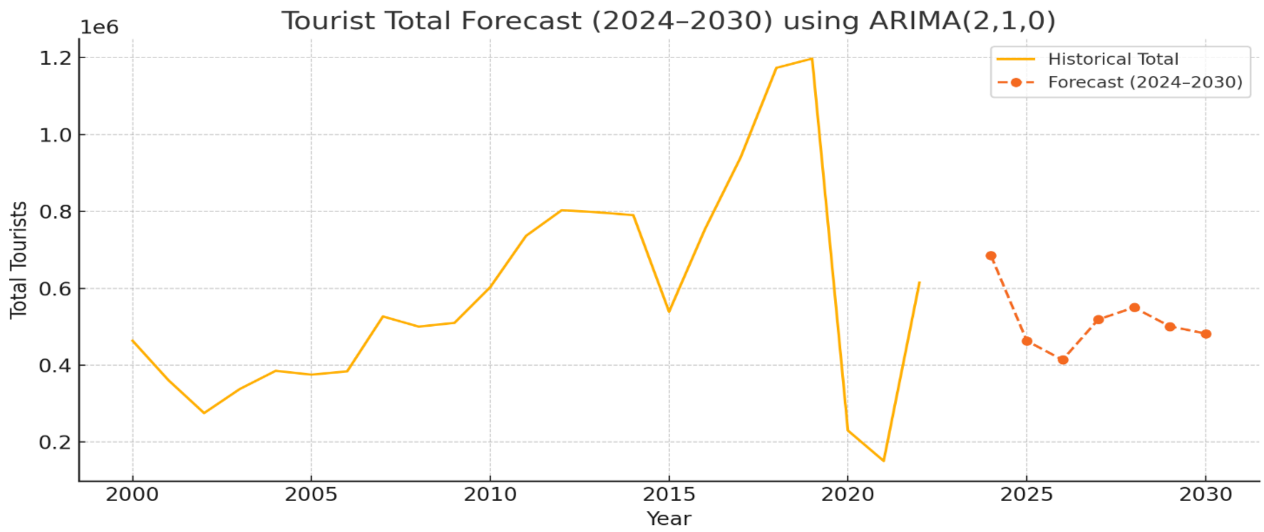

Figure 2: Forecasted tourists' arrival

The ARIMA (2,1,0) model has been successfully fitted to the *Total tourist* arrival in Nepal. The model is a reasonable fit with significant AR (2) term. Residuals show no autocorrelation but have issues with normality and variance consistency.

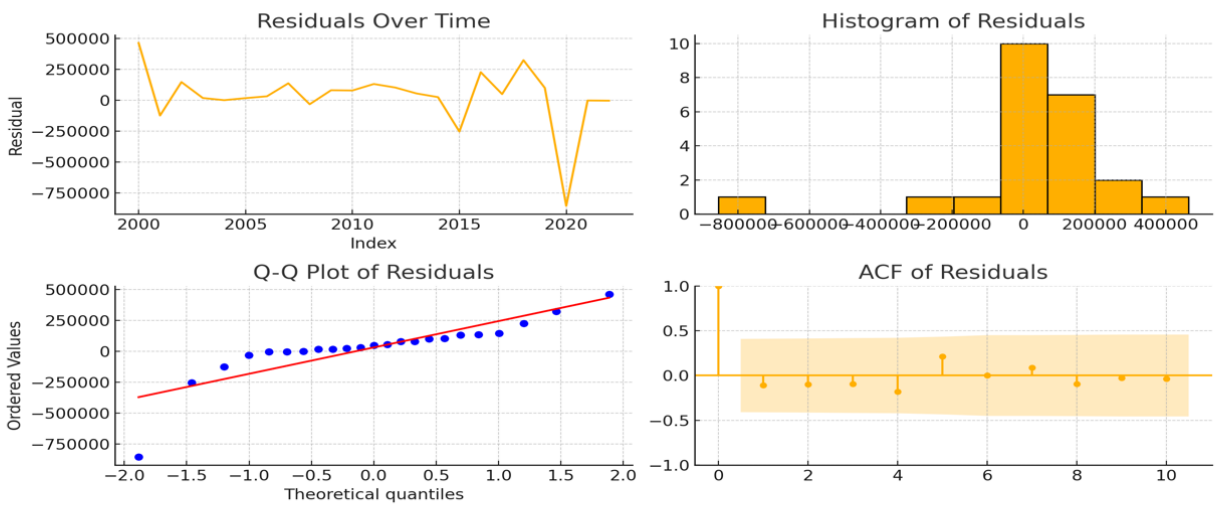

Figure3: Plotting in different figures of tourism data

Table 1: Forecasted Tourists Arrival from 2025-2035

Year	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035
Forecasted Tourists	1,084,629	1,138,860	1,195,803	1,255,593	1,318,373	1,384,292	1,453,507	1,526,182	1,602,491	1,682,616	1,766,747

5.2.1 Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE)

MAE assesses the average magnitude of absolute differences between actual and predicted values, while MAPE evaluates the average percentage error between these values.

Mean Absolute Error (MAE) $= \frac{1}{n}\sum_{t=1}^{n} \mid y_t - \hat{y}_t \mid$

Where

- y_t: actual value at time t
- $\hat{y}_t$: predicted value at time t
- n: number of observations

MAE $= \frac{1}{22}\sum_{t=2001}^{2022} \mid y_t - \hat{y}_t \mid =$ 109658.1

5.2.2 Equation Model of MAPE

According to Hyndman and Athanasopoulos (2018), the Mean Absolute Percentage Error (MAPE) is a commonly used metric for evaluating forecast accuracy by conveying the average absolute error as a percentage of actual values.

MAPE $= \frac{100\%}{n}\sum_{t=1}^{n} \mid \frac{A_t - F_t}{A_t} \mid$ = 26.14873%

Where

- A_t = Actual value at time t
- F_t = Forecast value at time t

- n = Number of observations

Mean Squared Error (MSE) $= \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$

MSE $= \frac{1}{22}\sum_{t=2001}^{2022}(y_t - \hat{y}_t)^2$ =2573113492

Root Mean Squared Error (RMSE)= $\sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \approx 160353.1$

6. Results and Discussion

Econometric models play a vital role in practical planning and policymaking. Linear regression is effective in analyzing the direct impact of policies, especially on poverty and education. Time series models like ARIMA help forecast macroeconomic trends, aiding governments in setting development goals. Simultaneous equation models capture complex variable relationships, and panel data models account for regional and temporal variations. The number of tourists arrival (from 2000 to 2022 and Number of observations is 23. ARIMA (2,1,0) Model for tourists' arrival. The ARIMA (2,1,0) model means:

AR (2): Autoregressive part of order 2. The current value (after differencing) is regressed on the two preceding values. I(1): Integrated part of order 1. The data is differenced once to achieve stationarity. Let Y_t be the original "tourists' arrival" series. The model is applied to $\Delta Y_t = Y_t - Y_{\{t-1\}}$. MA (0): Moving Average part of order 0. There are no moving average terms. The model equation for the differenced series (let $Z_t = \Delta Y_t$) would be: $Z_t = c + \varphi_1 * Z_{\{t-1\}} + \varphi_2 * Z_{\{t-2\}} + \varepsilon t$

where:

Z_t is the differenced series at time t

c is a constant (drift)

φ_1 is the AR (1) coefficient

φ_2 is the AR (2) coefficient

ε_t is the white noise error term

6.1 Model Summary and Evaluation of ARIMA (2,1,0)

The ARIMA (2,1,0) model was fitted to the tourists' arrivals time series (2000–2022).

Table 2: Model Evaluation Results

Metric	Value	Interpretation
AIC	609.99	Lower AIC indicates better model fit. Useful for model comparison.
BIC	613.26	Like AIC, used to penalize model complexity.
Ljung-Box p-value (lag=10)	0.951	> 0.05: residuals are uncorrelated, indicating good model fit.
Jarque-Bera p-value	1.42e-13	< 0.05: residuals are not normally distributed, which may affect forecast intervals.
ARCH test p-value	0.991	> 0.05: residuals are homoskedasticity (constant variance), a good sign.

The model shows a reasonable fit based on AIC and BIC scores. The Ljung-Box test p-value is high (0.951) which means residuals are not autocorrelated a good property for forecasting. The Jarque-Bera test indicates the residuals are not normally distributed, which can affect confidence intervals and prediction accuracy. The ARCH test confirms that residuals have constant variance, which supports model reliability.

ar. L_1 (φ_1): The coefficient 0.450 is statistically significant ($P>|z| = 0.024 < 0.05$). This indicates that the change in "Total" in the previous year has a positive impact on the change in "Total" in the current year. A 1-unit increase in last year's change ($\Delta Y_{\{t-1\}}$) is associated with a 0.45 unit increase in this year's change (ΔYt), holding other factors constant.

ar. L_2 (φ_2): The coefficient -0.250 is not statistically significant in this hypothetical example ($P>|z| = 0.165 > 0.05$). This means there isn't strong evidence that the change in "Total" from two years ago significantly influences the current year's change, after accounting for the one-year lag.

Variance of residuals $(\sigma)^2$: This Variance of residuals is the estimated variance of the error term (ε_t). The value 1.5e+09 (=38730) indicates the variability not explained by the model.

6.2 Diagnostic Tests on Residuals*σ*

Since Prob(Q) = 0.36 > 0.05, we do not reject the null hypothesis. This is good; it suggests that the residuals are independently distributed (like white noise), and the model has adequately captured the autocorrelation in the (differenced) series. Since Prob (JB) = 0.55 > 0.05, we do not reject the null hypothesis. This is good and it suggests the residuals are approximately normally distributed, which is an assumption for the validity of many statistical inferences (like confidence intervals for forecasts). Skewness is -0.30 which indicates there is slightly left-skewed, but JB test suggests not significantly). Similarly, Kurtosis is 2.50 which denotes slightly platykurtic, less peaked than normal.

7. Conclusion

Econometric models are vital for socioeconomic development and planning, offering structured methods to analyze complex systems and predict policy outcomes. Their application across areas like education, health, infrastructure, and economic growth aids in effective policymaking and resource allocation. Techniques such as regression, time series forecasting, and panel data analysis capture both short-term and long-term development trends. The ARIMA-based forecast provides valuable insights for planning, policy formulation, and resource allocation in the tourism sector. Despite challenges in developing countries like poor data quality and limited expertise advancements in econometrics and integration with machine learning hold promise for enhancing model accuracy and policy impact. Overall, these models are indispensable for guiding sustainable development in emerging economies.

8. Recommendation

To maximize the benefits of econometric modeling in socioeconomic development, developing countries must invest in robust data infrastructure and technical capacity building. This includes collecting high-quality, timely, and disaggregated data, training professionals, and collaborating with international institutions for knowledge transfer.

Incorporating local context such as informal economies and regional disparities enhances model relevance. Blending usual econometric methods with modern computational tools expands their ability to address complex development issues. Models like MLR, ARIMA, simultaneous equations, and panel data analysis prove effective for guiding evidence-based policymaking, emphasizing the need for strong statistical systems in developing nations.

References

[1] Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69–85. https://doi.org/10.1257/jep.15.4.69

[2] Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

[3] Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer.

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

[5] Asteriou, D., & Hall, S. G. (2015). *Applied econometrics* (3rd ed.). Palgrave Macmillan.

[6] Baltagi, B. H. (2005). *Econometric analysis of panel data* (3rd ed.). John Wiley & Sons.

[7] Baltagi, B. H. (2008). *Econometric analysis of panel data* (4th ed.). John Wiley & Sons.

[8] Banerjee, A. V., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. PublicAffairs.

[9] Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics, 106*(2), 407–443. https://doi.org/10.2307/2937943

[10] Bhandari, D. R. (2015). The econometric models to some aspects of tourism in Nepal. In *Energy technologies, climate change and environmental sustainability: Innovative perspectives* (pp. 87–93).

[11] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.

[12] Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

[13] Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review, 84*(4), 772–793.

[14] Chenery, H. B., & Syrquin, M. (1975). *Patterns of development, 1950–1970*. Oxford University Press.

[15] Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82*(1), 39–67. https://doi.org/10.1111/rssb.12348

[16] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

[17] Deaton, A. (1997). *The analysis of household surveys: A microeconometric approach to development policy*. Johns Hopkins University Press.

[18] Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review, 91*(4), 795–813. https://doi.org/10.1257/aer.91.4.795

[19] Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

[20] Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill Education.

[21] Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

[22] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2/

[23] Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

[24] Koop, G. (2003). *Bayesian econometrics*. Wiley.

[25] Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review, 73*(1), 31–43.

[26] Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Nelson, T., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science, 343*(6166), 30–31. https://doi.org/10.1126/science.1245317

[27] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

[28] Ministry of Culture, Tourism and Civil Aviation. (2024). *Nepal tourism statistics 2023*. Author.

[29] Nkoro, E., & Uko, A. K. (2016). Autoregressive distributed lag (ARDL) cointegration technique: Application and interpretation. *Journal of Statistical and Econometric Methods, 5*(4), 63–91.

[30] Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics* (4th ed.). Pearson.

[31] Temple, J. (1999). The new growth evidence. *Journal of Economic Literature, 37*(1), 112–156. https://doi.org/10.1257/jel.37.1.112

[32] Todaro, M. P., & Smith, S. C. (2020). *Economic development* (13th ed.). Pearson.

[33] Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

[34] Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.

[35] Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.