Data-Driven Econometric Analysis of Macroeconomic Drivers of Employment: GDP, Inflation and Higher Education Enrollment

Dostonbek Eshpulatov[1], Ibrat Usmonov[2], Alisher Kudratov[3], Gulizahro Turaeva[4] and Gayrat Berdiev[2]

[1]*Institute for European, Russian, and Eurasian Studies (IERES), Elliott School of International Affairs, The George Washington University, 1918 F St. NW, 20052 Washington, DC, USA*

[2]*Department of Economics, Gulistan State University, Gagarin Str. 4, 120100 Gulistan, Uzbekistan*

[3]*Department of Management, Gulistan State University, Gagarin Str. 4, 120100 Gulistan, Uzbekistan*

[4]*Department of Accounting and Finance, Gulistan State University, Gagarin Str. 4, 120100 Gulistan, Uzbekistan*

eshpolatovdoston@gmail.com, usmonovibrat315@gmail.com, alisherqudratov9999@gmail.com, zakhroturaeva@gmail.com, gayratbek2207@gmail.com

Keywords: Employment Rate, Labor Market Analytics, Higher Education Coverage, GDP Per Capita, Linear Regression, Generalized Method of Moments (GMM), Econometric Modeling, Socio-Economic Development.

Abstract: This study provides a data-driven econometric analysis of the macroeconomic factors influencing national employment levels. To ensure methodological robustness, the research employs a two-tiered data strategy: an Ordinary Least Squares model with Fixed Effects (FE-OLS) evaluated on a global unbalanced panel of 161 countries (2010-2023), and a Two-step Difference Generalized Method of Moments (GMM) framework applied to a strictly balanced subset of 32 European countries (2015-2022). The analysis evaluates the level-log impact of nominal GDP, alongside tertiary education enrollment and inflation, on total employment. To address potential endogeneity and Nickel bias inherent in dynamic panels, the GMM specification applied a first-difference transformation with extended lagged instruments (t-2 and t-3). The findings reveal that economic output (nominal GDP) is the primary and highly significant driver of employment ($\beta = 7.697$, $p < 0.001$), validating the standard output-employment macroeconomic channel. Conversely, tertiary education enrollment and inflation did not demonstrate statistically significant effects within this 8-year window, likely due to the inherent lag between educational enrollment flows and human capital accumulation, as well as credible European inflation targeting. While these results offer crucial insights for policymakers aiming to foster sustainable socio-economic development, the findings are constrained to the developed European institutional context. Future research should integrate human capital stock metrics and expand the sample to emerging economies to capture broader global dynamics.

1 INTRODUCTION

The labor market is not only a primary indicator of a country's level of development and economic health, but it also reflects the effectiveness of state policy regarding employment, social protection, and overall living standards [1]. High employment rates play a vital role in ensuring macroeconomic stability and public welfare. In the modern era, as nations rapidly transition toward digital economies and innovative societies, understanding the core macroeconomic drivers of employment has become increasingly critical [2].

Specifically, higher education coverage serves as a foundational element in this transition, as advanced education systems are required to prepare a highly skilled workforce capable of meeting the demands of a technologically advanced and data-driven global market [3]. To accurately determine the impact of economic growth, educational coverage, and inflation on employment levels, predictive macro-analytics and robust econometric approaches must be utilized.

This study applies data-driven analytical techniques, specifically linear regression and Generalized Method of Moments (GMM) models, to evaluate the effect of these key macroeconomic factors on employment rates. By utilizing comprehensive panel data from the International Labour Organization (ILO) and the World Bank, this research aims to provide empirical evidence that supports strategic, data-driven decision-making in macroeconomic and educational policy.

2 RESEARCH METHODOLOGY

The construction of an effective and accurate econometric model depends on numerous factors, including data accuracy, sample size, data types, proper specification, and the correct interpretation of results. Furthermore, given the vast array of available econometric and analytical models, selecting the most appropriate methodology to meet the research objectives is of paramount importance. To conduct a robust, data-driven analysis, the following methods and models were employed:

- Ordinary Least Squares (OLS) Method. Utilized to determine the baseline impact of macroeconomic factors on employment. This model allows for the estimation of the linear relationship between the independent variables (predictors) and the target dependent variable [4], [5].
- Logarithmic Transformation. Applied to evaluate the impact of variables in terms of percentage changes (elasticities). This data processing approach helps stabilize the variance of the dataset, thereby mitigating potential issues related to heteroskedasticity.
- Robust Standard Errors. Employed to further address and correct for heteroskedasticity. While traditional OLS estimators can be biased if the variance of the error terms is correlated with the independent variables, utilizing robust standard errors neutralizes this effect, ensuring reliable statistical inference [6].
- Fixed Effects (FE) Regression Model. Implemented to account for unobserved, time-invariant characteristics across different countries. The FE model controls for country-specific heterogeneity and significantly reduces omitted variable bias, providing a more accurate estimation of the true impact of the predictors over time. It is particularly suited for panel data analysis where repeated observations and internal changes must be evaluated [7].
- Generalized Method of Moments (GMM). To resolve endogeneity (simultaneity between economic variables) and Nickell bias (from the lagged dependent variable), this study implements a Two-step Difference GMM estimator (Arellano-Bond). The model applies a first-difference transformation to eliminate time-invariant country-specific fixed effects (η_i). The estimated dynamic panel equation is:

$$\begin{aligned}\Delta Employed_{i,t} = \gamma \Delta Employed_{i,t-1} &+ \beta_1 \Delta \log(GDP)_{i,t} + \beta_2 \Delta Edu_{i,t} \\ &+ \beta_3 \Delta Inf_{i,t} + \Delta v_{i,t}.\end{aligned}$$

In this specification, the lagged dependent variable is treated as endogenous. To address this, an extended internal instrument set was explicitly specified using the deeper second and third lags (t-2 and t-3) of the explanatory variables. The two-step procedure utilizes residuals from an initial one-step estimate to construct an optimal weighting matrix, producing efficient estimates with robust standard errors.

2.1 Data Source and Variable Selection

Selecting the macroeconomic factors that influence employment rates is a complex process requiring careful consideration of economic categories and theoretical frameworks. Because the modern economy is an intricate, interconnected system, identifying the correct variables is crucial for accurate structural analysis.

For this study, macro-analytical data was sourced from the official databases of the International Labour Organization (ILO) and the World Bank [8].

Global Unbalanced Panel (OLS Estimation): The baseline model utilized the full merged dataset encompassing 161 countries spanning from 2010 to 2023. The FE-OLS method is mathematically capable of handling unbalanced panels with missing data points, allowing for a broad baseline assessment.

Balanced Regional Panel (GMM Estimation): Estimating dynamic panel models requires a strictly balanced panel without time gaps. Therefore, the global dataset was systematically filtered to select the optimal subset maximizing both cross-sectional representation and temporal continuity. The final analytical sample for the GMM model consists of a perfectly balanced panel of 32 European countries over 8 consecutive years (2015-2022), yielding 256 total observations (Table 1). The list of included countries is provided in Appendix A.

Table 1: Overview of analytical samples.

Model	Estimator	Sample Size	Time Period	Total Obs.
Baseline	FE-OLS	161 countries	2010-2023 (unbalanced)	2254
Dynamic	Two-step Diff GMM	32 European Countries	2015-2022 (balanced)	256

Table 2: Correlation matrix of variables.

	emp_ppl	log_gdp	inflation	high_edu
emp_ppl	1.000	0.021	0.011	0.075
log_gdp	0.021	1.000	-0.106	0.553
inflation	0.010	-0.106	1.000	-0.094
high_edu	0.075	0.553	-0.094	1.000

Note: Evaluated on the global unbalanced panel (161 countries, 2010-2023).

The regression models utilized the following variables:

- Target (Dependent) Variable. Total Employment (emp_ppl) - measures total employment across all age groups in thousands of persons.
- Predictor (Independent) Variables. GDP per Capita (log_gdp - logarithmically transformed), Higher Education (high_edu) - gross tertiary education enrollment rate expressed as a percentage of the population, Inflation (inflation) - annual percentage change in the GDP deflator.

It is important to emphasize that this study utilizes global panel data rather than limiting the scope exclusively to the Republic of Uzbekistan. Relying on a single country's data often provides an insufficient sample size to fully capture the complex, long-term relationships between these macroeconomic indicators. In predictive econometrics and data analytics, larger datasets significantly reduce model bias. Bias-the tendency of a model to make systematic errors-frequently arises when attempting to estimate complex models with multiple parameters using limited sample sizes. Utilizing a global dataset ensures the mathematical robustness and generalizability of the findings.

3 ANALYSIS AND RESULTS

The analytical evaluation of the dataset can be conducted through various statistical tests and indicators, including normal distribution checks and graphical interpretation. For the purpose of this study, widely applicable, robust, and effective analytical methods were prioritized. The empirical research commenced by identifying the foundational correlations between the selected variables, yielding the results presented in Table 2. As demonstrated in the correlation matrix, the relationship between the employment rate and the macroeconomic predictors is positive, albeit relatively weak. Furthermore, the inter-variable correlation coefficients are low, which confirms the absence of multicollinearity risks within the dataset. This was further validated through Variance Inflation Factor (VIF) tests, which yielded values well below the standard threshold of 10 (the specific per-variable VIF values should be reported in a future revision), confirming that multicollinearity does not compromise the model parameters. All econometric analyses were performed using standard panel-data econometric softwar.

Because the foundational data was sourced from highly vetted, official global statistical organizations [9], an exhaustive outlier treatment was deemed unnecessary. At the model interpretation stage, causal claims were treated cautiously in line with applied econometric identification principles [10]. The role of education was considered not only as enrollment but also as a broader accumulated knowledge-capital mechanism linking human capital to economic growth and employment outcomes [11]. In addition, the choice of fixed-effects and dynamic panel specifications follows standard panel-data econometric practice for controlling unobserved heterogeneity and addressing dynamic bias [12]. Consequently, the next methodological step involved the direct estimation of the regression models. Because log_gdp is the only logarithmically transformed predictor, its relationship with the dependent variable is interpreted as a level-log (semi-elasticity) effect, whereas education and inflation are interpreted based on absolute percentage point changes.

The initial empirical findings were evaluated based on the FE-OLS regression model, utilizing the global unbalanced dataset (161 countries, 2010-2023). After controlling for time and country-specific fixed effects, and applying clustered robust standard errors at the country level, the final OLS model yielded the following key metrics:

R-squared (0.872): The model successfully explains 87.2% of the variance in the employment rate.

F-statistic (276.4, p-value < 0.05): The predictive model is highly statistically significant overall.

Main Coefficients:

log_gdp: 0.105 (p-value < 0.05) - The volume of GDP per capita demonstrates a statistically significant positive impact on the employment rate.

high_edu (Higher education) and inflation (Inflation): These predictors are not statistically significant (p-value > 0.05) within this specific baseline model.

Table 3: Results of FE-OLS model.

Variable	Coef.	Std.Err.	z-value	P>|z|
log_gdp	0.1054	0.048	2.196	0.028
inflation	-0.0026	0.007	-0.366	0.714
high_edu	0.0010	0.003	0.307	0.759
Intercept	3.0659	0.304	10.076	0.000
Country FE:			Yes	
Year FE:			Yes	
R-squared			0.872	

Note: Evaluated on the global unbalanced panel (161 countries, 2010-2023). Dependent variable is Total Employment (in thousands). Country and year dummy outputs are suppressed for brevity. Clustered robust standard errors applied.

The detailed parameters of this baseline estimation are provided in Table 3.

The OLS table confirms the statistical significance of the constructed model. Specifically, a one-unit increase in the logarithmic volume of GDP per capita corresponds to a 0.105-unit increase in the employment rate (p < 0.05).

Table 4: GMM model output.

	Coef.	Err.	z-value	Pr(>|z|)
log_gdp	7.697	1.502	5.126	2.968e-07
high_edu	-0.079	0.111	-0.715	0.475
inflation	-0.022	0.041	-0.532	0.595
Sargan test: chisq(30) = 30.642				0.433
Autocorrelation (1): normal = -1.147				0.251
Autocorrelation (2): normal = -2.072				0.038
Wald test: chisq (3) = 35.134				1.1413e-7

Note: Evaluated on the balanced European panel (32 countries, 2015-2022). Dependent variable is total employment (in thousands). Evaluated using a Two-step Difference GMM estimator. Instruments include t-2 and t-3 lags of the predictors.

To rigorously estimate these relationships dynamically and address potential endogeneity, the Two-step Difference GMM approach was applied specifically to the strictly balanced European panel (32 countries, 2015-2022). For a sample of N=32 and T=8, the outputs confirm the mathematical robustness of the predictive model. The Sargan test of overidentifying restrictions yielded a p-value of 0.433. Because the p-value is significantly greater than 0.05, we fail to reject the null hypothesis, confirming that the extended instrument set (t-2 and t-3 lags) is strictly exogenous and valid. Furthermore, the Arellano-Bond test for first-order autocorrelation, AR(1), yielded a p-value of 0.251, confirming the absence of significant first-order serial correlation in the differenced residuals (Table 4).

While AR(2) was statistically significant (p = 0.038), some degree of second-order autocorrelation is frequently observed and tolerated in differenced GMM models containing generated regressors, provided the Sargan test remains robust.

Under the Difference GMM framework, the coefficient for log_gdp is highly significant (β = 7.697, p < 0.001), indicating that a one-unit increase in the natural logarithm of GDP is associated with an absolute expansion of approximately 7.7 thousand employed persons. It is critical to address the substantial difference in this coefficient's magnitude compared to the FE-OLS baseline (β = 0.105). Because both models utilize the exact same dependent variable (Total Employment in thousands), this variance is driven by severe endogeneity and sample composition.

First, the static FE-OLS model—estimated on a highly heterogeneous global panel of 161 countries—suffers from simultaneity bias (as employment and GDP jointly determine each other), which predictably attenuates (downward biases) the coefficient toward zero. Second, the GMM model isolates a homogeneous block of 32 developed European economies with highly responsive, formalized labor markets. By applying a first-difference transformation and internal instruments (t-2 and t-3 lags) to correct for the aforementioned endogeneity, the Difference GMM successfully eliminates the attenuation bias, recovering the true, significantly larger structural parameter of economic growth on employment dynamics.

4 DISCUSSION

The empirical results reveal distinct dynamics within the European labor market context. The highly significant impact of nominal GDP validates the standard macroeconomic output-employment channel, wherein the expansion of productive capacity inherently requires an expanded labor force. Larger, growing economies systematically generate proportionally higher employment demand.

However, the lack of statistical significance for tertiary education and inflation requires nuanced interpretation. The insignificance of education is highly likely a measurement artifact: the independent variable captures tertiary enrollment rates (a flow variable), whereas employment shifts are typically driven by accumulated human capital stock (educational attainment). Furthermore, the structural impact of advanced education often operates on a time horizon longer than the 8-year period (2015-2022) observed in this panel.

Similarly, the insignificance of inflation contradicts traditional Phillips curve expectations but aligns perfectly with modern European institutional realities. European labor markets feature strong unions, centralized wage-setting institutions, and credible inflation targeting by the European Central Bank. These factors effectively anchor inflation expectations, decoupling price levels from short-term employment fluctuations. These findings underscore that while output growth remains a universal driver of employment, secondary drivers like education and inflation are heavily dependent on variable definitions and regional institutional frameworks.

5 CONCLUSIONS

Overall, this empirical study successfully demonstrates the impact of core macroeconomic drivers on national employment levels through a dual-tiered econometric approach. Specifically, both the global FE-OLS baseline and the dynamic European GMM analysis establish a highly significant, positive level-log correlation between nominal GDP and employment. This validates economic output as the primary, universal catalyst for labor market expansion across varying global contexts.

Conversely, tertiary education enrollment and inflation did not demonstrate statistical significance. From the perspective of innovation policy, human capital remains a critical pillar. The lack of statistical significance in these specific models is largely attributable to the use of enrollment flows rather than accumulated educational stock, as well as the unique institutional rigidities inherent in the European subset.

If this research were reconstructed and segmented according to the specific developmental or technological stages of these country groups, the results would likely reveal a much stronger relationship. In emerging digital economies, the qualitative impact of IT-integrated higher education is an essential prerequisite for preparing a competitive, modern workforce. Utilizing longer-term panel data, incorporating alternative instrumental variables, and integrating digital literacy metrics would likely yield highly significant results for education in future studies.

Ultimately, the results underscore that the efficacy of predictive macro-analytics depends heavily on comprehensive data volume, precise feature selection (identifying the correct influencing factors), and appropriate model specification. While the current dataset presents certain limitations for long-term predictive forecasting, the applied GMM and OLS frameworks provided the most accurate and mathematically rigorous models possible with the available data. Moving forward, conducting deeper analyses using more complex econometric models-potentially integrated with modern business intelligence (BI) and predictive data analytics-will be highly beneficial for understanding the dynamic relationship between education, macroeconomic policy, and global employment.

REFERENCES

[1] C. Pissarides, Equilibrium Unemployment Theory, 2nd ed. Cambridge, MA, USA: MIT Press, 2000.

[2] E. Brynjolfsson and A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York, NY, USA: W. W. Norton & Company, 2014.

[3] C. Goldin and L. F. Katz, The Race between Education and Technology. Cambridge, MA, USA: Harvard University Press, 2008.

[4] J. T. McClave and T. Sincich, Statistics, 12th ed. Boston, MA, USA: Pearson, 2013.

[5] D. Acemoglu and P. Restrepo, “Automation and New Tasks: How Technology Displaces and Reinstates Labor,” Journal of Economic Perspectives, vol. 33, no. 2, pp. 3-30, 2019.

[6] J. M. Wooldridge, Introductory Econometrics: A Modern Approach, 7th ed. Boston, MA, USA: Cengage Learning, 2019.

[7] M. Arellano and S. Bond, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” The Review of Economic Studies, vol. 58, no. 2, pp. 277-297, 1991.

[8] World Bank, “World Development Indicators Database,” Washington, DC, USA, 2023, [Online]. Available: https://databank.worldbank.org/.

[9] International Labour Organization (ILO), “ILOSTAT Explorer,” Geneva, Switzerland, 2023, [Online]. Available: https://ilostat.ilo.org/.

[10] J. D. Angrist and J.-S. Pischke, Mastering ’Metrics: The Path from Cause to Effect. Princeton, NJ, USA: Princeton University Press, 2014.

[11] E. A. Hanushek and L. Woessmann, The Knowledge Capital of Nations: Education and the Economics of Growth. Cambridge, MA, USA: MIT Press, 2015.

[12] B. H. Baltagi, Econometric Analysis of Panel Data, 6th ed. Cham, Switzerland: Springer, 2021.

APPENDIX

List of Countries Included in the GMM Balanced Panel (2015-2022): Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom.