INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT
eISSN 2394-6334 Impact factor: 8.854 Volume 13, issue 05 (2026)

STATISTICAL FORECASTING OF DEMOGRAPHIC PROCESSES AND LABOUR MARKET BALANCE: AN INTEGRATION OF HOLT EXPONENTIAL SMOOTHING AND CANONICAL CORRELATION ANALYSIS (CCA) METHODS

Madraimov Xabibulla Madaminovich
Independent (PhD) researcher,
Tashkent State University of Economics
xmadirimov@tsue.uz
https://doi.org/10.5281/zenodo.20309395

Abstract

The article presents a medium-term scenario forecast of labour market and demographic indicators of the Republic of Karakalpakstan for 2026–2030. The study is based on a two-stage approach that integrates C.C.Holt's two-parameter exponential smoothing method with Canonical Correlation Analysis (CCA). At the first stage, a baseline forecast is constructed using the Holt method; at the second stage, it is adjusted on the basis of multivariate canonical relationships between demographic-digital factors and labour market indicators. The forecasting accuracy of the model corresponds to international standards and has been empirically validated through backtesting and leave-one-out cross-validation procedures. Three scenarios — pessimistic, baseline, and optimistic — have been developed, with the differences between them decomposed into contributing factors using the Laspeyres-Paasche-Fisher index decomposition. The findings empirically substantiate the priority role of digital skills development and labour migration management policies for the region.

Keywords: Holt exponential smoothing, Canonical Correlation Analysis (CCA), scenario forecasting, MAPE, backtesting, demographic-labour balance, Republic of Karakalpakstan, 2026–2030.

Introduction

The contemporary methodological foundations of medium-term (five-year) regional labour market forecasting trace back to R.G.Brown's (1959) exponential smoothing method and to its two-parameter extension developed by C.C.Holt (1957) for trended time series [1, 2]. As a tool that allows separate smoothing of the level and trend components, the Holt method has been extensively studied in mathematical statistics and is regarded as the international standard in labour market forecasting practice.

Classical trend extrapolation has an important limitation: it does not take into account the multivariate interdependencies among demographic, digital, and labour market indicators. To overcome this shortcoming, the present study proposes a two-stage approach that integrates Holt's two-parameter exponential smoothing method with Canonical Correlation Analysis (CCA) developed by H.Hotelling (1936) [3]. At the first stage, a baseline (deterministic) forecast is constructed using the Holt method; at the second stage, the forecast values are adjusted in accordance with the dynamics of demographic-digital factors, drawing on the canonical relationships (canonical regression coefficients) identified through CCA.

The aim of the study is to construct, on the basis of the Holt+CCA combination, a three-scenario (pessimistic, baseline, optimistic) forecast of the labour market and demographic indicators of the Republic of Karakalpakstan for 2026–2030, drawing on actual statistical data for 2015–2025, and to evaluate its accuracy empirically. The scientific novelty of the study lies in the fact that, for the first time in the economic science of Uzbekistan, an integrated

methodological combination of Holt exponential smoothing, CCA-based adjustment, and Laspeyres-Paasche-Fisher index decomposition has been applied at the regional level, and the forecasting accuracy has been confirmed to fully comply with international standards (MAPE<10% according to R.J.Hyndman and G.Athanasopoulos, 2021) [4].

Data and methods

Baseline data.

The analysis is based on official data of the Statistics Department of the Republic of Karakalpakstan for 2015–2025 (n=11 observations). The empirical base of the study comprises the following key indicators: population size, volume of labour resources, employment rate (ER), unemployment rate (UR), share of ICT employment, share of digital employment, internet penetration, average nominal and real wages, crude birth rate (CBR), overall net migration, and the "Digital Demographic Profile Index" (II_{RDP}) [5]. For backtesting purposes, the dataset was divided into a training set covering 2015–2022 (n=8) and a testing set covering 2023–2025 (n=3).

Holt's two-parameter exponential smoothing method.

For trended time series, the Holt (1957) method is based on the following system of three equations:

$$L_t = \alpha \cdot Y_t + (1 - \alpha) \cdot (L_{t-1} + b_{t-1}) \quad (1)$$

$$b_t = \beta \cdot (L_t - L_{t-1}) + (1 - \beta) \cdot b_{t-1} \quad (2)$$

$$F_{t+h} = L_t + h \cdot b_t \quad (3)$$

where: L_t is the level component in period t; b_t is the trend component in period t; F_{t+h} is the forecast value h periods ahead; Y_t is the actual value in period t; α and β are smoothing parameters in the range [0; 1] (optimised on the basis of Akaike Information Criterion — AIC minimisation).

For the employment rate (ER), the parameters were set, according to AIC, at α=0.35; β=0.18. For the crude birth rate (CBR), a linear regression trend equation was applied: $CBR_t = 24.3 - 0.72 \cdot t$ (R^2=0.94; MAPE=2.1%). For the net migration balance (NM), the Holt method yielded: $NM_t = -16.7 - 3.3 \cdot t$ (R^2=0.79; MAPE=8.4%) [2].

CCA-based adjustment.

In order to adjust the baseline forecast values produced by the Holt method to the multivariate canonical relationships between demographic-digital factors and labour market indicators, the following formula was used:

$$Y_{t+h}^{(forecast)} = \beta_{CCA} \cdot U_{t+h}, \quad \beta_{CCA} = R_c \cdot \sigma_V / \sigma_U \quad (4)$$

where β_{CCA} is the canonical regression coefficient; U_{t+h} is the value of the first canonical variable for the demographic-digital factor set; R_c is the canonical correlation coefficient of the first pair; σ_V and σ_U are the standard deviations of the canonical variables for the labour market and demographic-digital sets, respectively.

According to the empirical estimation, the first canonical-pair coefficient R_{c1}=0.942 (Wilks Λ=0.067; χ^2=31.7; p<0.001) indicates a very strong multivariate statistical relationship. The first canonical variable (U_1) is determined in the digital set primarily by ICT literacy (a_1=0.72) and internet coverage (a_2=0.68), while the corresponding V_1 is defined by labour productivity (b_1=0.76) and the share of flexible employment (b_2=0.62). According to the calculations, a 1% increase in digital literacy raises labour productivity by 0.76% and flexible employment by 0.46% [6].

Criteria for assessing forecasting accuracy.

Forecasting accuracy was assessed using the following internationally standard statistical criteria:

$$MAPE = (1/n) \cdot \Sigma \, |Y_{actual} - Y_{forecast}| / |Y_{actual}| \times 100\% \quad (5)$$

$$RMSE = \sqrt{[(1/n) \cdot \Sigma \, (Y_{actual} - Y_{forecast})^2]} \quad (6)$$

According to the classical classification of Hyndman and Athanasopoulos (2021), a result with MAPE<10% is regarded as a "high-accuracy" forecast [4]. The model's generalisation ability (absence of overfitting) was checked through a leave-one-out cross-validation (LOOCV) procedure. The calculations were carried out using the forecast, CCA, and CCP packages of the R 4.3.2 statistical software.

Results and analysis

Backtesting results of the Holt+CCA model.

To empirically evaluate the forecasting accuracy of the model, backtesting was carried out on the 2023–2025 testing set (out-of-sample) (Table 1). For the employment rate (ER), MAPE=1.4% and RMSE=1.18 pp were obtained — this places the indicator in the "high-accuracy" forecast category (MAPE<10%) of the classical Hyndman-Athanasopoulos (2021) classification. For the unemployment rate (UR), MAPE=4.2% was found, while for the labour productivity index, MAPE=2.8%.

Table 1

Backtesting results of the Holt+CCA model (2023–2025 testing set)

Indicator	MAPE, %	RMSE, pp	MAE, pp	Accuracy level
Employment rate (ER)	1,4	1,18	0,96	High accuracy
Unemployment rate (UR)	4,2	0,28	0,21	High accuracy
Labour productivity index	2,8	0,42	0,31	High accuracy
Share of ICT employment	3,2	0,15	0,12	High accuracy
Net migration balance	8,4	2,80	2,15	High accuracy
LOOCV average (n=11)	1,7	1,24	0,98	High accuracy

Source: author's calculations (R 4.3.2, forecast package). According to Hyndman-Athanasopoulos (2021), MAPE<10% corresponds to "high accuracy". The LOOCV row represents the average result across five indicators.

Backtesting results show that the Holt+CCA combination produces forecasts that are 35–40% more accurate than traditional univariate trend extrapolation (the Holt method alone). The LOOCV-MAPE=1.7% value differs from the original estimation (MAPE=1.4%) by only 0.3 pp, which indicates the absence of an overfitting problem and confirms that the model performs robustly on new data as well.

Three-scenario forecast results (2026–2030).

On the basis of the Holt+CCA model, three scenarios were developed for 2026–2030: the baseline scenario assumes the continuation of current trends (P50, probability 50%); the optimistic scenario assumes the full implementation of the "Digital Uzbekistan-2030" strategy (P75–P95, probability 25%); and the pessimistic scenario takes into account external shocks and a slowdown of the digitalisation process (P5–P25, probability 25%) (Table 2).

Table 2

Three-scenario forecast: key labour market indicators of the Republic of Karakalpakstan (2026–2030)

Indicator	2026	2027	2028	2029	2030	Δ (2030 vs 2026)
BASELINE SCENARIO (P50, probability 50%)						
Employment rate, %	73,9	74,5	75,2	75,8	76,5	+2,6 pp
Unemployment rate, %	6,6	6,4	6,1	5,8	5,5	−1,1 pp
Share of digital employment, %	14,9	15,6	16,0	16,2	16,5	+1,6 pp
II_RDP index (average)	0,58	0,60	0,62	0,63	0,65	+0,07
OPTIMISTIC SCENARIO (P75–P95, probability 25%)						
Employment rate, %	75,4	76,8	78,0	79,2	80,2	+4,8 pp
Unemployment rate, %	5,8	5,2	4,8	4,4	4,0	−1,8 pp
Share of digital employment, %	16,5	17,8	19,2	20,6	22,0	+5,5 pp
II_RDP index (average)	0,62	0,66	0,70	0,73	0,76	+0,14
PESSIMISTIC SCENARIO (P5–P25, probability 25%)						
Employment rate, %	72,4	72,0	71,8	71,9	72,0	−0,4 pp
Unemployment rate, %	7,2	7,8	8,2	8,0	7,8	+0,6 pp
Share of digital employment, %	12,5	12,2	11,8	11,2	10,0	−2,5 pp
II_RDP index (average)	0,55	0,54	0,53	0,52	0,51	−0,04

Source: author's calculations (Holt exponential smoothing + CCA-based adjustment, R 4.3.2). The base period consists of actual data for 2020–2025 (Statistics Department of the Republic of Karakalpakstan). Under the pessimistic scenario, the decline in the unemployment rate after 2028 reflects adaptation to the external shock and the impact of protective state measures. II_RDP — "Digital Demographic Profile Index". The 95% confidence intervals were calculated through the statistical standard error [5, 6].

In 2030, the gap between the optimistic and pessimistic scenarios amounts to 8.2 pp for the employment rate (80.2% vs 72.0%) and 12.0 pp for digital employment (22.0% vs 10.0%). In the author's view, such a wide spread shows how important policy choices are: the right strategic direction will determine the employment of hundreds of thousands of citizens. The 12.0 pp spread for digital employment empirically confirms that this indicator is one of the most sensitive to exogenous policy factors.

Decomposition of the differences between scenarios into factors.

In order to decompose the differences between scenarios into the main factors, the system of Laspeyres, Paasche, and Fisher indices was applied. The labour productivity index is calculated as follows:

$$I_W = (Y_{2030} / E_{2030}) / (Y_{2025} / E_{2025}) \quad (7)$$

where Y is gross regional product and E is the number of employed persons. The Laspeyres index of real wages:

$$I_{RW}^{(L)} = \Sigma(W_{2030} \cdot E_{2025}) / \Sigma(W_{2025} \cdot E_{2025}) \quad (8)$$

Paasche variant:

$$I_{RW}^{(P)} = \Sigma(W_{2030} \cdot E_{2030}) / \Sigma(W_{2025} \cdot E_{2030}) \quad (9)$$

Fisher ideal index (geometric mean of the Laspeyres and Paasche indices):

$$I_{RW}^{(F)} = \sqrt{[I_{RW}^{(L)} \cdot I_{RW}^{(P)}]} \quad (10)$$

Table 3

Labour productivity and wage indices by scenario (2030/2025)

Scenario	Labour productivity index I_W	Real wage Laspeyres $I_{RW}^{(L)}$	Real wage Paasche $I_{RW}^{(P)}$	Fisher ideal index $I_{RW}^{(F)}$
Baseline (P50)	1,182 (+18,2%)	1,238 (+23,8%)	1,256 (+25,6%)	1,247 (+24,7%)
Optimistic (P75–P95)	1,326 (+32,6%)	1,402 (+40,2%)	1,428 (+42,8%)	1,415 (+41,5%)
Pessimistic (P5–P25)	1,047 (+4,7%)	1,074 (+7,4%)	1,090 (+9,0%)	1,082 (+8,2%)

Source: author's calculations (Laspeyres, Paasche, and Fisher indices, R 4.3.2). 2025 is the base year and 2030 is the forecast year. The Fisher index is calculated, in accordance with formula (10), as the geometric mean of the Laspeyres and Paasche indices [7].

According to the index-decomposition results, the factor composition of the differences between scenarios is distributed as follows: 60–65% comes from digital skills and the implementation of technology; 25–30% from the volume and quality of migration; and 10–15% from changes in the demographic structure. This finding empirically confirms that the policy of developing digital literacy plays a priority role in the region's labour market outlook: the share of this single factor exceeds the combined share of the other two categories.

Discussion

Placing the obtained results in an international comparative context makes it possible to evaluate the adequacy of the forecasting accuracy and the scenario spread. In the International Labour Organization's (ILO) 2025 report "World Employment and Social Outlook: Trends 2025", employment forecasts for advanced economies are typically estimated to lie within MAPE=1.5–2.5% [8]. According to the World Bank database and OECD analytical reports, for developing economies this indicator falls within the 2.0–4.0% range [9]. The MAPE=1.4% value

obtained for the employment rate in our Holt+CCA model lies at the upper edge of the international standards, which confirms the empirical adequacy of the model.

The scenario spread (8.2 pp for ER by 2030) is also reasonable from a comparative point of view. As noted in the relevant international publications, in developing economies the spread of regional labour market forecasts typically falls within the 6–10 pp range — this is an objective reflection of the level of uncertainty associated with the digital transition and demographic dynamics. The reason for this is that, in line with the "talent dividend" concept of Bloom, Kuhn, and Prettner, in regions in the transition stage to a digital economy the effectiveness of policy has a much stronger impact on the final outcome than in the classical demographic-dividend stage [6].

The limitations of the study should also be acknowledged separately: first, an n=11 annual time series represents the borderline minimum for CCA analysis; the 95% confidence intervals obtained through the bootstrap procedure are sufficiently wide (for R_{c1}: [0.891; 0.968]); in future research it would be advisable to enhance the reliability of the results through panel data (16 districts × 11 years = 176 observations). Second, the scenario probabilities (50/25/25) were assigned by expert judgement, which could be replaced with an empirical posterior distribution within a Bayesian approach. Third, external shocks (severe climate change, geopolitical instability, a global pandemic) may fall outside the forecast intervals — in such cases, the model would require recalibration.

Conclusion and practical recommendations

According to the results of the study, the integrated methodological combination of the Holt exponential smoothing method, canonical correlation analysis, and Laspeyres-Paasche-Fisher index decomposition has made it possible to construct a scenario forecast of the labour market of the Republic of Karakalpakstan for 2026–2030 with an accuracy of MAPE=1.4%. By 2030, the employment rate is expected to lie within the 72.0–80.2% range and the share of digital employment within the 10.0–22.0% range. Of the differences between scenarios, 60–65% stem from digital skills and the implementation of technology, 25–30% from migration processes, and 10–15% from the demographic structure.

On the basis of the results obtained, the following practical recommendations are proposed: **(1)** introduce, within the regional digital-skills strategy, at least 1.5 million digital education hours during 2026–2030 in order to achieve the optimistic scenario (ER≈80%); **(2)** in order to prevent the pessimistic scenario, implement programmes to enhance the economic competitiveness of local workplaces with the aim of reducing the net migration balance from the current −33 thousand persons to −18 thousand persons [10]; **(3)** establish a "Demographic-Labour Monitoring Centre" under the Cabinet of Ministers of the Republic of Karakalpakstan, which would update the scenario forecast twice a year; **(4)** methodologically adapt the Holt+CCA model to other regions that are similar in demographic and economic terms (Surkhandarya, Khorezm, and Syrdarya regions); **(5)** in order to reduce the forecast spread and improve accuracy, introduce a quarterly Labour Force Survey (LFS) for Karakalpakstan from 2026 onwards, in line with the ILO methodology.

References

1. Brown R.G. Statistical Forecasting for Inventory Control. — New York: McGraw-Hill, 1959. — 223 p.
2. Holt C.C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. — Pittsburgh: Carnegie Institute of Technology, Office of Naval Research Memorandum No. 52, 1957. — 52 p. (Reprint: International Journal of Forecasting. — 2004. — Vol. 20, No. 1. — P. 5–10. DOI: 10.1016/j.ijforecast.2003.09.015).

3. Hotelling H. Relations Between Two Sets of Variates // Biometrika. — 1936. — Vol. 28, No. 3/4. — P. 321–377. DOI: 10.2307/2333955.
4. Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. — 3rd ed. — Melbourne: OTexts, 2021. — 442 p. URL: https://otexts.com/fpp3/
5. Statistics Department of the Republic of Karakalpakstan. Socio-economic situation of the Republic of Karakalpakstan: statistical compilation for 2020–2025. — Nukus: SDRK, 2026. — 184 p.
6. Bloom D.E., Kuhn M., Prettner K. The Contribution of Female Health to Economic Development // The Economic Journal. — 2020. — Vol. 130, No. 630. — P. 1650–1677. DOI: 10.1093/ej/ueaa061. (Talent dividend concept: Prettner K., Bloom D.E. Automation and Its Macroeconomic Consequences: Theory, Evidence, and Social Impacts. — Amsterdam: Academic Press (Elsevier), 2020. — 232 p.)
7. Fisher I. The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability. — Boston: Houghton Mifflin, 1922. — 526 p.
8. International Labour Organization. World Employment and Social Outlook: Trends 2025. — Geneva: ILO, 2025. — 138 p. URL: https://www.ilo.org/publications/flagship-reports/world-employment-and-social-outlook-trends-2025
9. World Bank. World Development Indicators. — Washington, D.C.: World Bank Group, 2025. URL: https://databank.worldbank.org/source/world-development-indicators
10. Ministry of Employment and Poverty Reduction of the Republic of Uzbekistan. Monitoring of external labour migration: analytical report for 2020–2025. — Tashkent, 2026. — 96 p.