COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MORATUWA

MERCon 2026
MORATUWA ENGINEERING RESEARCH CONFERENCE

The Remittance Blueprint: Data-Driven Intelligence for Sri Lanka

Dhinanjaya Fernando, Dinura Ginige, Kalana Lakshan, Chanupa Gurusinghe, Lasana Pahanga, Subavarshana Arumugam, Sandeepa Weerasekara, Sandareka Wickramanayake, Nisansa de Silva

Department of Computer Science & Engineering, University of Moratuwa, Sri Lanka

Abstract

We analyse 32 years (1994–2025) of Sri Lankan migration and remittances on a harmonised 384-month dataset from eight institutional sources, combining exploratory analysis, stationarity-corrected time-series modelling (ADF, KPSS, Johansen, VECM, IRF, FEVD) and supervised learning. Inflows are driven by *external* variables — exchange rates and global oil prices — not domestic inflation or unemployment. Ridge regression beats a SARIMA baseline by 73.8% (RMSE USD 494.8 Mn) and projects 2026 inflows of USD 9,001 Mn (± 970 Mn, 95% CI).

1994–2025	−73.8%	USD 494.8 Mn	USD 9,001 Mn
32 years · 384 monthly rows · 8 sources	RMSE vs. univariate SARIMA baseline	annualised out-of-sample RMSE (Ridge)	projected 2026 inflows (± 970, 95% CI)

1. Context, Motivation & Research Gap

- **Context:** remittances are persistently **6–9% of GDP** — the dominant foreign-exchange source and principal buffer in the **2022 sovereign default**.
- **Policy stake:** under the IMF Extended Fund Facility, remittance forecasts feed reserve targets, import cover and debt sustainability.
- **Theory:** NELM [1] casts remittances as household insurance set jointly by origin-country risk *and* destination earnings — a duality demanding a **multivariate** specification.
- **Gap:** prior Sri Lankan work [2, 3] is descriptive and annual, with no non-stationarity correction and no demographic–macro integration.

2. Contributions

- **Dataset:** a harmonised **32-year, 384-month** series from eight authoritative sources, released openly.
- **Inference: stationarity-corrected causality** (ADF/KPSS, Granger, Johansen, VECM), not spurious level regressions.
- **Breaks:** structural breaks detected at **2018, 2020** and **2022**.
- **Integration:** one pipeline linking demography, macroeconomics and temporal dynamics.
- **Forecast:** **USD 9,001 Mn** for 2026, **73.8%** better than SARIMA.

3. Data Sources & Dataset Composition

Institution	Contribution
Central Bank of Sri Lanka	remittances, policy rate, monetary aggregates
World Bank Open Data	GDP, remittance totals, poverty headcount
SLBFE	departures: gender, skill, district, contracts
ILO	labour-migration statistics
FRED / St. Louis Fed	LKR/USD exchange rate (DEXSLUS)
DCS Sri Lanka	CCPI, inflation, wages, employment
U.S. EIA	Brent crude oil prices
World Bank PPP	purchasing-power parity per capita

- **Four variable domains:** demographic (gender, skill, age, contract length, destination concentration); remittances (monthly USD Mn); domestic macro (GDP, inflation, unemployment, wages, policy rate, poverty, CCPI); external exposure (LKR/USD, PPP, destination growth, Brent).
- Built by **web extraction**, **OCR digitisation** of SLBFE reports (1994–2005) and CSV/XLSX imports; every indicator cross-validated against **an independent source**.

4. Analytical Pipeline

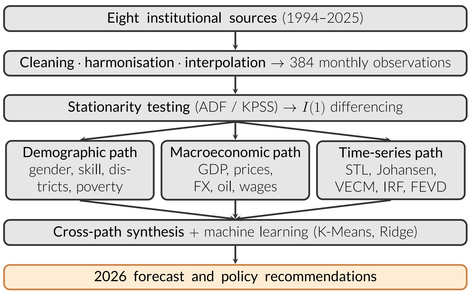


Figure 1. Three analytical paths feed one synthesis and forecasting stage.

5. Preprocessing & Stationarity Correction

- **Missingness:** columns $> 40\%$ missing dropped; short gaps median-imputed, runs interpolated, edges filled. **Crisis months retained** — clipping the $|z| > 3$ outliers of 2022 would erase the key structural break.
- **Annual → monthly interpolation.** $y(t) = y_a + (y_{a+1} - y_a)(t - t_a)/(t_{a+1} - t_a)$: injects synthetic autocorrelation, but aligns remittances with monthly regressors and matches CBSL aggregates. FX averaged within month, CCPI chain-linked, all series z-scored.
- **Unit-root correction** [4, 5]: primary series are $I(1)$, so first differences are used throughout correlation, Granger and VAR analysis.
- **Features:** remittance lags at **1, 3, 6, 12** months, a composite Gulf pull-factor, skill ratios, 12-month rolling mean and volatility.

6. Exploratory Analysis: Demographics

- **Gender inversion:** 1994 flows were **72.8% female**; the ratio inverted after 2008, peaked near **66% male**, and now sits at **60:40**. Female flows stay skill-dominant (mean 90.0%).
- **Skill polarisation** (Theil–Sen [6, 7]): skilled share -0.687/**yr** (95% CI -0.831 to -0.511), asymmetric by gender — male $+0.493$ vs. female -0.254/yr; breaks at **2018** ($F = 28.58$) and **2020** ($F = 13.31$), $p < 0.001$. Since female migrants held higher-remitting posts, the eroding female share **compresses per-capita yield**.
- **Poverty–migration hump** [8]: poverty near **65.0%** constrained early emigration; as it fell to **18.6% (2017)** emigration surged — lagged and non-linear.
- **Spatial inequality:** HHI [9] +9.00/**yr** (95% CI 6.69–12.06); by 2025 one district supplies **63.70%** of departures — SLBFE networks bypass many others.

7. Exploratory Analysis: Macro Linkages

- **Earnings, not headcount:** GDP grew +811% (USD 11.72 → 106.8 Bn). GDP–remittances $r = 0.9803$ vs. GDP–emigration $r = 0.4543$: growth reflects earning capacity, occupational upgrading and formal-channel adoption — a nuance volume-based policy would miss.
- **Gulf exposure:** emigration–oil price $r = 0.7070$ ($p < 0.001$) — an oil contraction cuts destination labour demand and inflows together.
- **Currency channel:** remittances–LKR/USD $r = 0.6645$ ($p < 0.001$), combining income and valuation effects.
- **Weak domestic signals:** despite volatile inflation (mean 9.31%), co-movement is negligible — remittances–inflation $r = -0.253$, emigration–policy rate $r = -0.026$.

8. Statistical Inference ($\alpha = 0.05$)

	Null hypothesis H_0	Test, statistic	Outcome
Demographic composition			
H1	No trend in female share of departures	Regression, $\beta = -1.358$, $p < 0.001$	**Reject**: shift to male flows
H2	No skill difference between genders	Wilcoxon [10], $W = 528.0$, $p < 0.001$	**Reject**: females outperform
H3	No link between age and contract length	Pearson, $r = 0.591$, $p < 0.001$	**Reject**: older, longer
Labour market and poverty			
H4	No poverty–emigration association	Pearson, $r = -0.359$, $p = 0.044$	**Reject**: migration hump
H5	No remittance–unemployment link	Spearman, $\rho = -0.897$, $p < 0.001$	**Reject**: inflows ↑, unemp. ↓
H6	No emigration–employment link	Spearman, $\rho = -0.118$, $p = 0.520$	**Fail to reject**
Macroeconomic drivers (joint OLS over remittances/emigration)			
H7	Inflation has no explanatory power	$F = 2.543$, $p = 0.096$	**Fail to reject**
H8	Central-bank rate has no explanatory power	$F = 7.476$, $p = 0.002$, $R^2 = 0.34$	**Reject**
H9	Domestic wages have no explanatory power	$F = 23.181$, $p < 0.001$, $R^2 = 0.61$	**Reject**
H10	Destination GDP growth has no explanatory power	$F = 0.961$, $p = 0.394$	**Fail to reject**

9. Decomposition, Cointegration & Error Correction

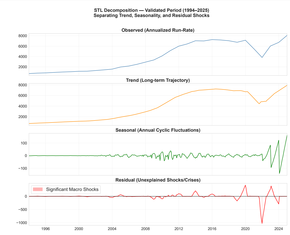


Figure 2. STL decomposition [11] of annualised remittances (period = 12, robust).

- **STL:** dominant trend ($T_f > 0.8$), negligible seasonality ($S_f < 0.1$), isolated shocks in **2020** and **2022**. ADF/KPSS confirm all key series are $I(1)$; Johansen [12] rejects $r = 0$ at 5%, justifying a VECM.

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \epsilon_t$$

Π encodes the long-run equilibrium; Γ_i the short-run dynamics.

- **Adjustment** $\alpha \approx$ **-0.15** ($p < 0.01$): **15% of disequilibrium corrected per month**, equilibrium in **6.7 months** — an **active macroeconomic stabiliser**.

10. Impulse Response & Variance Decomposition

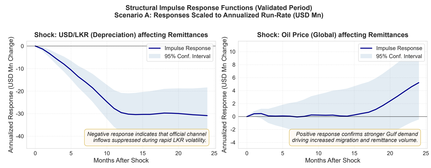


Figure 3. Structural IRFs (annualised USD Mn): LKR depreciation (left) vs. oil price shocks (right).

- Cholesky ordering by decreasing exogeneity: **oil → CCPI → exchange rate → remittances**.
- A 1-SD **LKR depreciation** *reduces* official inflows for about **three months** — a wait-and-see window where migrants *likely* divert funds to informal *Hawala/Undiyal* channels (as CBSL observed in 2022) — recovering by month six. A positive **oil price** shock lifts inflows persistently over **twelve months**.
- **FEVD:** own shocks explain **85%** of variance over months 1–6; oil and LKR/USD exceed **45%** by month 24.

11. Predictive Framework, Regimes & Forecasting

- Expanding-window walk-forward validation (`TimeSeriesSplit`) removes look-ahead bias, scored against a univariate SARIMA$(2,1,0)(1,0,0)_{12}$ baseline. $k = 4$ regimes by silhouette (**0.473**), Davies–Bouldin [13] and Calinski–Harabasz [14]: **pre-war**, **post-war commodity boom** and **2022 crisis**.

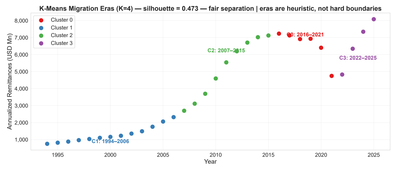
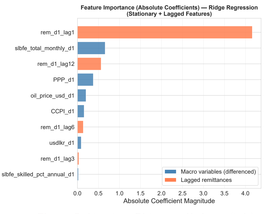


Figure 4. K-Means regimes across the 32-year window.



Model	RMSE
SARIMA$(2,1,0)(1,0,0)_{12}$	1,889
Ridge, multivariate L2	**494.8**
Δ baseline = −73.8%	

Figure 5. Absolute Ridge coefficients (left) and walk-forward accuracy (right).

- **Lag-1 and lag-12** terms dominate — lag-12 strength reflects **annual Gulf contract-renewal cycles**; **SLBFE departures**, observable **two to three months ahead**, give CBSL desks an actionable lead signal.
- Macro factors set the **long-run equilibrium** (VECM); near-term prediction is **autoregression-dominated**. **Caveat:** the gain conflates multivariate inputs with regularisation.

2026 PROJECTION **USD 9,001 Mn**
± USD 970 Mn (95% CI via RMSE), conditional on macroeconomic stability and no major disruption to Middle Eastern labour markets.

12. Policy Implications, Limitations & Conclusions

- **1. Keep the exchange rate market-clearing.** A parallel-market premium redirects diaspora savings into Hawala networks and erodes reserves, as in 2022.
- **2. Upgrade skills and diversify destinations.** SLBFE pre-departure training in construction technology, healthcare and logistics; bilateral high-wage corridors beyond the Gulf structurally reduce oil-cycle exposure, as the FEVD confirms.
- **Limitations:** interpolation masks seasonal spikes; SLBFE covers registered channels only; Ridge's linear form cannot capture abrupt Kafala-system shifts.
- **Conclusion:** remittances are **resilient but externally driven** — exchange rates and oil prices dominate the short run, outweighing domestic inflation and unemployment. Multivariate ML beats univariate SARIMA by **73.8%**: a defensible baseline for reserve planning. **Next:** destination labour-demand variables, high-frequency mobile transaction data [15], informal-channel and hybrid econometric–deep-learning models.

References

[1] O. Stark and D. E. Bloom, "The new economics of labor migration," *American Economic Review*, vol. 75, no. 2, pp. 173–178, 1985.

[2] S. S. Ramanayake and C. S. Wijetunga, "Sri Lanka's labour migration trends, remittances and economic growth," *South Asia Research*, vol. 38, no. 3_suppl, pp. 61S–81S, 2018.

[3] P. Wickramasekara, "South Asian gulf migration to the Gulf: a safety valve or a development strategy?" *Migration and Development*, vol. 5, no. 1, pp. 1–31, 2015.

[4] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.

[5] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics*, vol. 54, no. 1-3, pp. 159–178, 1992.

[6] H. Theil, "A rank-invariant method of linear and polynomial regression analysis," *Indagationes Mathematicae*, vol. 12, no. 85, p. 173, 1950.

[7] P. K. Sen, "Estimates of the regression coefficient based on Kendall's tau," *Journal of the American Statistical Association*, vol. 63, no. 324, pp. 1379–1389, 1968.

[8] W. Zelinsky, "The hypothesis of the mobility transition," *Geographical Review*, pp. 219–249, 1971.

[9] O. C. Herfindahl, *Concentration in the steel industry*. Columbia University, 1997.

[10] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics: Methodology and Distribution*, Springer, 1992, pp. 196–202.

[11] R. B. Cleveland, W. Cleveland, J. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess. 1990," *Journal of Official Statistics*, vol. 1435502, 2022.

[12] S. Johansen, "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, vol. 12, no. 2-3, pp. 231–254, 1988.

[13] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 224–227, 1979.

[14] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics-Theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.

[15] S. Lokanathan, G. E. Kreindler, N. H. N. de Silva, Y. Miyauchi, D. Dhanajaya, and R. Samarajiva, "The potential of mobile network big data as a tool in Colombo's transportation and urban planning," *Information Technologies & International Development*, vol. 12, no. 2, pp. pp–63, 2016.

Dinurang/Sri-Lanka-Emigrant-Remittance-Analysis

datasets/DS-research/Sri-Lanka-Emigrant-Remittance

ORGANIZED BY

FINANCIAL SPONSOR

TECHNICAL CO-SPONSORS