INTERNATIONAL MONETARY FUND

Caucasus, Central Asia, and Mongolia

Regional Capacity Development Center

DEVELOPMENT PARTNERS: ASIAN DEVELOPMENT BANK | CHINA | EU | KOREA | POLAND | SWITZERLAND | US

MEMBER COUNTRIES: ARMENIA | AZERBAIJAN | GEORGIA | KAZAKHSTAN | KYRGYZ REPUBLIC | MONGOLIA | TAJIKISTAN | TURKMENISTAN | UZBEKISTAN

CCAMTAC - Regional Research Seminar Series

"Can LLMs Represent Households Under-Represented in Their Training Data? Inflation Expectations in Uzbekistan"

July 3, 2026

<u>Moderation:</u>

Mr. Kassymzhomart Assangaziyev, Economic Analyst, CCAMTAC

<u>Presenters:</u>

Mr. Jamshid Mamasalaev, Head of Economic Policy and Risk Analysis Division, Research Department, Central Bank of Uzbekistan

Mr. Abduazim Rasulov, Researcher, Central Bank of Uzbekistan

<u>Intervention:</u>

Mr. Nika Khinashvili, Principal Economist, Macroeconomic Research Division, National Bank of Georgia

In this research seminar, Mr. Abduazim Rasulov and Mr. Jamshid Mamasalaev of the Central Bank of Uzbekistan presented a paper examining whether large language models (LLMs), prompted to answer as demographically representative households, can reproduce inflation expectations in a market whose languages are only thinly represented in LLM training data. The study builds 150 synthetic households matched to Uzbek population characteristics and has three open-weight LLMs answer the central bank's own monthly inflation expectations survey question across roughly six years of monthly waves, comparing the results to the published survey on distributional overlap and point-estimate accuracy.

The key findings show that the synthetic households reproduce the broad shape and direction of the real survey — including its upward perception gap relative to realized inflation — and remain stable across major shocks such as the pandemic, the 2022 commodity and FX shock, and the 2024 tariff reform. At the same time, the synthetic distributions update on news faster than real households do, tend to reproduce a textbook income gradient inconsistent with the CBU survey's actual pattern, and remain highly sensitive to the specific LLM used. The authors concluded that the approach is a promising, low-cost, between-waves complement to the official survey rather than a substitute for it.

During the discussion, Dr. Nika Khinashvili, Principal Economist at the National Bank of Georgia, commended the paper's novelty and its extension of the "synthetic survey" literature to an emerging-market, non-English-language context for the first time. He raised several points for further development, including increasing the number of synthetic households toward the sample size implied by standard survey theory, addressing the narrower spread and higher persistence of the LLM-generated distributions relative to the real survey, and testing whether combining multiple LLMs or allowing synthetic households to interact with one another could improve realism. He also proposed a signed Wasserstein-distance measure to better capture changes in expectations over time, and suggested extending the method to firms and financial-market participants, whose expectations transmit into inflation more directly than household expectations do. The authors, joined partway through by Mr. Mamasalaev, agreed that expanding the household sample and broadening the analysis to firms and financial-market participants were natural next steps, while noting that the survey's own reliance on the central bank's Telegram channel for recruitment complicates any claim of full population representativeness — a point echoed by participants during the Q&A.