In the Shadow of War: Assessing Conflict-Driven Disruptions in the Kyrgyzstan-Russia Labor Pipeline via a Gradient Boosting Approach to Nowcasting

Michelle K. Schultze

Professor Charles M. Becker, Faculty Advisor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke
University
Durham, North Carolina
2025

Acknowledgements

First and foremost, I am deeply grateful to my advisor, Professor Charles Becker, for his constant support throughout this project, from helping refine the topic to offering thoughtful, last-minute feedback. More broadly, his mentorship has shaped my academic journey, encouraging my growth as a researcher.

I would like to thank Professor Jeff DeSimone, director of the Duke Economic Analytics Laboratory, for his support with the econometric aspects of this paper and for his dedication to providing opportunities to undergraduate students. My sincere gratitude goes to Professor Kyle Bradbury for offering invaluable guidance on my machine learning and nowcasting plans. I also extend my thanks to Professor Grace Kim, our thesis seminar instructor, for her feedback on my literature review, especially during her exceptionally challenging year.

I extend my thanks to Daria Soboleva, Malika Syzdykova, Dr. Erik Zitser, Dr. Neven Valev, and Dr. Nurgul Ukueva for their help in procuring data. I am especially grateful to Daria for her detailed assistance with the minimum wage value and population data.

Lastly, to my amazing parents, Lisa and Jeff, thank you for encouraging my dreams and standing by me. To my incredible friends—thank you for your support and understanding these past few months. I'm so grateful for all the memories and excited for what's next. To Audrey—I'm so lucky to learn from you—your resilience and work ethic inspire me every day. To Charlie—thank you for your patience and for being by my side through it all.

Abstract

Kyrgyzstan, where remittances made up 30% of GDP before the Russo-Ukraine war, is central to understanding Russia–Central Asia labor migration. Wartime trends are obscured by informality—many Kyrgyz migrants work illegally in Russia—and by limited, politically manipulated Russian data. This study introduces a novel "nowcasting" approach using XGBoost and Yandex Wordstat—a Russian search query database largely overlooked in English-language research—to extend data into a usable time frame. Results show a push effect linked to war intensity, alongside a labor substitution effect: Kyrgyz migrants increasingly fill roles vacated by Russian conscripts. This shift primarily affects blue-collar and informal travelers, with remittance flows responding after a two-month lag.

JEL classification: F24; J6; R23

Keywords: Immigrant Workers; Remittances; Regional Labor Markets

1. Introduction

Three decades after the fall of the Soviet Union, millions of people still travel to Russia to work, permanently or seasonally, and send money known as remittances back to their families at home. This is the reality for the Central Asian country of Kyrgyzstan, officially called the Kyrgyz Republic, where 680,000 citizens — or more than 10% of the country's population — are estimated to live and work in Russia (Orlova, 2022), attracted by higher wages and the relative strength of the ruble. Remittances, mostly from Russia, are estimated to make up 30% of Kyrgyz GDP (World Bank, 2023)—one of the most severe remittance dependencies in the world.

Scholars have debated the macroeconomic impacts of remittance reliance for Central Asian economies, as well as the counteracting influx of Russian immigrants to Central Asia fleeing political persecution and economic turmoil. Despite the war in Ukraine entering its fourth year, very few studies have focused on the effect of the war on Central Asian migrant labor flows to Russia. One reason that this is difficult to study is informality of labor flows: measures of migration are inherently inaccurate due to both a high volume of illegal immigration and employment (Vorobeva, 2023). Second, Russian data are widely regarded as unreliable and sometimes directly fabricated, and some datasets were even removed from government websites after the onset of the war (Sonnenfeld, et al. 2022; *State,* 2024).

In an attempt to pin down Central Asian labor dynamics writ large, it is useful to focus on Kyrgyzstan as a particularly tractable case study. First, the country, though small, has published copious open data resources freely available on the internet, and it is also the beneficiary of many independent surveys focused on development outcomes. Second, the Kyrgyz language is almost completely exclusive to Kyrgyz nationals (Ethnologue, n.d.), which makes Yandex and Google search queries in the Kyrgyz language useful proxies for migrant geolocation. Third, Kyrgyzstan reflects the broader Central Asian context, sharing a post-Soviet economic legacy and Eurasian Economic Union membership in the

Eurasian Economic Union, which eases labor migration barriers (Ruziev & Majidov, 2013). Fourth, the country is fascinating for its degree of remittance reliance, making wartime effects on labor migration important to understand from a development perspective.

This topic is especially relevant due to the ongoing conflict in Ukraine. Contributing to an understanding of Russian labor dynamics, as well as ongoing migration patterns, this study directly informs policymakers and international organizations addressing migration management, remittance dependence, and poverty reduction in both Central Asia and Russia by uncovering overlooked trends and effects. More generally, this research illuminates war-related impacts on migrant labor, especially in the Post-Soviet context.

This study introduces novel strategies for nowcasting missing data points by integrating three unconventional data sources to reduce bias. The available time series for key variables—regional migrant flows (2021–23) and quarterly national data on Kyrgyz visitors to Russia (2010–16)—are too short for robust analysis. To extend these series, this study leverages search query data from Yandex Wordstat and Google Trends. XGBoost (an ensemble gradient boosting algorithm) outperforms alternative machine learning methods in terms of validation RMSE and preserving ordinality. In contrast, the quarterly Kyrgyz visitors data requires a SARIMAX model (Seasonal Autoregressive Integrated Moving Average with exogenous variables) to account for the tiny N=28 sample size and incorporate temporal dependencies. A different model is fit for each visitor type—with focus on transit, private, and business travelers.

Using these estimates, I evaluate **three hypotheses** of the effect of the War in Ukraine on Kyrgyz migration patterns:

1. **General wartime exodus:** The onset of the war in Ukraine should have produced a *decrease* in Kyrgyz migrants on average across Russia since February 2022, though exodus should occur in

two distinct “waves:” at the onset of the war, due to migrants dodging conscription and political turmoil, and after the Crocus Hall attack in April 2024, which led to discrimination and deportations for Central Asian migrants (Gozzi, 2024).

2. **Conscript replacement theory:** Migrant workers should “replace” conscripts or casualties in the Russian labor market, due to the scale of Russia’s fighting force (BBC, 2025) and a history of Central Asian migrants supporting gaps in Russian employment (Ibragimova, 2024). Using the BBC-Mediazona confirmed casualties dataset (Mediazona, 2025), I evaluate the hypothesis that regions and time periods with more reported casualties should be associated with an *increase* in migrant worker inflows or Kyrgyz visitors to Russia, ceteris paribus.
3. **Remittance trends:** Remittances should have *decreased* on average at the onset of the war due to the push factors of political turmoil and fear of being drafted. However, a positive association with lagged casualties—reflecting the labor substitution effect—should show up in increased remittances robust to the wage and inflation level.

Regressions reveal a push effect associated with severity of war, suggesting migrants are driven away from political turmoil and conscription. Crocus Hall attack effects are negative but more subtle, suggesting a longer time frame is needed. At the same time, a labor substitution effect emerges: confirmed Russian casualties are positively associated with migrant inflows, especially among assumed blue-collar and informal travelers. From regressions on remittances, the labor substitution appears to influence remittances with a two-month delay. Lastly, these regressions suggest there is a six-month incubation period for the substitution effect to manifest.

2. Literature Review

2.1 Background on Remittances, Migrant Workers, and Kyrgyzstan's development

In the face of the widespread economic devastation of the post-Soviet independence years, many households in Central Asia took advantage of higher wages in Russia by working abroad seasonally or permanently and sending money back to their families in Kyrgyzstan. A shared familiarity with the Russian language laid the groundwork for Central Asian workers to choose to travel to their northern neighbor. While some countries like Kazakhstan, Moldova, and Ukraine receive significant shares of remittances outside of Russia (Palacin & Shelburne, 2007), geographic constraints limit most of the landlocked Central Asian countries, which are far from alternative work destinations (Oshchepkov, Tilekeyev, & Gerry, 2024). In many cases, such as for Kyrgyzstan, these foreign remittances became outsized contributors to domestic consumption.

After the fall of the Soviet Union, Kyrgyzstan transitioned to a reliance on remittances. While making up less than one percent of the country's GDP until 1997, remittances rose to 31% of GDP in 2013 (3), by some estimates reaching as high as 40% of GDP (Lee, 2019). The State Migration Service of Kyrgyzstan estimates the number of remittance workers in Russia to be approximately 680,500 migrants in 2018 (Orlova, 2022), with some official estimates far higher at 1,039,000 in 2023 (Jayaprakash, 2024)—a jaw-dropping proportion of the country's population of around 7 million people (Interfax, 2023).

Many studies across global cases attribute positive net effects to Kyrgyz remittances. The average wage rate in Russia is at least 3.5 times the rate in Kyrgyzstan, representing an enormous financial inflow to Kyrgyzstan if properly harnessed (Jayaprakash, 2024). Seasonal migrants from Kyrgyzstan are motivated by "altruism and insurance" (Kumar, et al. 2018), such that they can send large paychecks home to feed their families. Some studies have uncovered a resulting positive effect on

Kyrgyz GDP both directly and through a multiplier effect (Aitymbetov, 2006). The remittance-based economy may rely on remittances as “an alternative form of liquidity” facilitating growth, and there are studies that support a strong link between remittances and development (Kumar, et al. 2018). Remittance-receiving households have a higher probability of purchasing durable goods than households not receiving remittances (Ukueva, 2010). Ukueva (2010) also argues that altruistic remittances allow countries like Kyrgyzstan with low technological development to avert poverty traps by enhancing technology transfer primarily to Bishkek, the capital, which seems to be manifesting in a widening tech sector.

However, the literature also broadly suggests that if a remittance-based economy lacks a strong financial sector, as Kyrgyzstan does, this may prevent the efficient transfer and use of remittances and reduce potential positive effects (Kumar, et al. 2018). 70% of remittances are estimated to be devoted to basic needs like food and clothes, with only 7% devoted to entrepreneurial activities, limiting long-term multiplicative benefits (Zhanaltay, 2020; Wang, Hagedorn, & Chi, 2021). Remittances are often channeled toward communal celebrations like weddings (Rubinov, 2014), as in neighboring Uzbekistan (Kakhkharov et al. 2021; Irnazarov, 2015).

Another major way that remittances undercut Kyrgyz development is vulnerability to external shocks, especially from the Russian economy. Both remittances and Kyrgyz GDP experience pro-cyclical fluctuations with Russian business cycles, with scholars finding evidence for unidirectional causality assigned to Russian GDP (Çetintaş & Baigonushova, 2018). This pro-cyclicality is mirrored in other post-Soviet countries too, like Ukraine itself in the 2000s and early 2010s (Kupets, 2012). Kyrgyz remittances are sensitive to global oil price changes, potentially channeled through the Russian economy (Zhanaltay, 2020).

2.2 Profile of Kyrgyz migrants to Russia

Prior to the war, Kyrgyz migrant laborers in Russia were increasing over time due to "simplification of procedures for labour migration between member states of the Eurasian Economic Union (EAEU), high unemployment rates and low wages in Kyrgyzstan, [and] family reunification due to earlier migration of men" (Abdukadyrova & Studenko, 2023). 40% of construction workers in Russia are migrants, and laborers also work in "retail trade, agriculture, and housing and communal services" (Dzhooshbekova, et al. 2021). Several are entrepreneurs, with 16% of migrants in 2010 living there for at least 7 years and operating their own businesses.

2.2.1. Popular Destinations

Beginning in the 1990s, the Far East and Siberia represented the largest number of Kyrgyz migrants at approximately 400,000 to 700,000 people (Dzhooshbekova, et al. 2021; Торогельдиева, 2020). This may be partly due to the vacuum left by many Russians leaving Siberia due to relaxed internal migration restrictions after the fall of the Soviet Union. Kyrgyz migrants also may travel to these areas as similar Turkic languages and ethnic appearance create a more accepting community for Kyrgyz migrants (Radio Free Europe, 2013). Proximity has also played a role, as the main rail and air transport routes went to Novosibirsk, Novokuznetsk, and Krasnoyarsk, allowing for population to dissipate from these points, only later reaching the now-popular destinations of Tyumen and Sakhalin (Кузменкин, 2021). In 2001, the Sverdlovsk region of Russia employed more than 300,000 Kyrgyz citizens in primarily "shuttle trade and services" (Dzhooshbekova, et al. 2021). In the Far East, 40 groups existed in 2006 to help Kyrgyz migrants obtain permissions and lodging in Altai, Tuva, Yakutia, and other places.

Sakha-Yakutia is a primary destination for Kyrgyz migrants, with at least 30,000 official Kyrgyz residents, including more than 8,000 in the capital, Yakutsk (Ratio Free Europe, 2019a). After a 2019 sexual assault case led to an outpouring of violence and protest against Kyrgyz migrants, the governor of the region banned the employment of foreigners in many sectors (Ratio Free Europe, 2019b). Regardless, official numbers analyzed in this study demonstrate that the region still receives a steady stream of migrants, which is acknowledged by news reports (Osmonalieva, 2024).

Moscow receives a massive portion of Kyrgyz migrants (Dogan, 2020). Many of these migrants come from rural villages in Kyrgyzstan, only 53.1% speaking Russian fluently and 67.2% of men arriving without a spouse (Тынаева, 2024; Еленин, 2018). Conditions for Kyrgyz migrants are reported by Russian media to be more egalitarian and integrated in Moscow, but crackdowns on illegal workers are more widespread.

Many Kyrgyz migrants also now live and work in the remote island of Sakhalin, some on expired credentials (Осмоналиева, 2024). Reportedly, 210,000 local Russian citizens left Sakhalin after the Fukushima power plant disaster, leading to mainly Kyrgyz, Uzbek, and Tajik immigrants filling this gap (Кожобаева, 2024). In April 2025, 30,000 Kyrgyz citizens officially worked in Sakhalin at the height of the season. Most migrants there enjoy lighter labor restrictions, try to obtain Russian citizenship, and are allowed to work for up to three years without having to leave the country to renew their migration permissions. Less discrimination and deportation raids appear to be present for these migrants, too.

2.2.2. Illegality

Prior to the war, people from Commonwealth of Independent States (CIS) countries could enter Russia without a visa, only requiring a work permit, due to the Treaty on the Eurasian Economic Union granting Kyrgyz citizens equal access to education and work, at least in principle (Корнеев, 2024). In light of supposed expanding illegal immigration (and heightened anti-immigrant sentiment), between 2012 and 2015 Russia adopted more than 50 laws enforcing new labor restrictions—both administrative and criminal (Jayaprakash, 2024), where most international migrants have 30 days to pass a series of cumbersome requirements to obtain a patent (work permit) (The New Humanitarian, 2015). While Kyrgyz citizens are exempt from these restrictions, they still require a migration card involving fingerprinting, a photo, and a medical examination, and a health insurance policy (Пакина, 2014; Полешко, 2025). The cost of these requirements, which also includes paying taxes and fees, may be simply unaffordable for migrants, forcing them to work informally in a much more vulnerable position (Jayaprakash, 2024). Migrants also may routinely pay bribes to suspend their entry bans if applicable, and they may rely on informal arrangements with "immigration officials, police, border control officials, lawyers, activists, and immigration consultancy firms." Since so many migrants live outside the eyes of the law, they become more often subjected to labor exploitation and abuse. Migrants are also frequently subjected to racial discrimination, though are harassed less by law enforcement if they gain Russian citizenship (Agadjanian, et al. 2017).

2.2.3. Women face unique problems

Around half of Kyrgyz migrants to Russia are women—in 2017, 230,029 individuals aged 14 or older recorded by official counts, with likely more in subsequent years (Romanenko & Borodinka, 2019). They typically work in lower-paying service jobs or unskilled work in factories or agriculture, earning around US$500 per month (Abdukadyrova & Studenko, 2023; Romanenko & Borodinka, 2019). Back in Kyrgyzstan, women are often praised for contributions to family income, but shamed for withdrawing from traditional gender roles (Scott, Sexmith, & Chi, 2024). Women arrive in Russia primarily with family members, with more than 50% arriving with their partner in 2017, and a further 10% bringing children (Romanenko & Borodinka, 2019). However, an increasing number arrive independently.

During COVID, women were more exposed to the virus as workers primarily in the service sector, also experiencing more layoffs, violations of their rights, and subsequent food insecurity especially if they work informally (Abdukadyrova & Studenko, 2023). Around 40% of sex workers in Russia are international migrants, with female Central Asian migrants representing a large share. In a 2006 survey of 442 women migrant laborers, 20% of retail workers were forced to provide sexual services (Romanenko & Borodinka, 2019). Heartbreakingly, this rate represented 100% of women surveyed in the entertainment industry.

2.2.4. Impact of major disruptive events

Remittances tend to decline, and migrants tend to leave or transition to illegal work, in response to major disruptive events. After the 2008 global financial crisis, Central Asian migrants to Russia plummeted by around 1 million due to more restrictive quotas by the Federal Migration Service (FMS) (Dzhooshbekova, et al. 2021). However, these numbers rebounded almost immediately after the restrictions were relaxed (Jayaprakash, 2024), with many people also working illegally to continue earning a living (Dzhooshbekova, et al. 2021). Ruble devaluation as a result of sanctions in 2014 led to the Central Asian migrant labor force declining by 22% from Tajikistan and 15.6% from Uzbekistan in 2015, and remittances also declined in value, creating similar effects.

The COVID-19 pandemic and associated lockdowns had deeply negative impacts on the Kyrgyz migrant worker population in Russia. Remittances to former Soviet countries were 1.7 times lower in April 2020 compared to the following year, April 2021 (Jayaprakash, 2024). Additionally, borders were closed and flights cancelled, leading to at least 350,000 Kyrgyz workers being stuck in Russia during the pandemic.

Different regions imposed varying lockdown protocols, and since migrant workers are only permitted to work on a per-city basis, if they lost their job, they may be unable to afford a new patent (work permit) for a different place, even if they could more easily get a job there. This led to a surge in illegal work as COVID surged (Jayaprakash, 2024). Undernutrition was prevalent, and some migrants were locked down without any job, thus unable to help those at home (Murzakulova, et al. 2020).

2.2.5. Impact of the war in Ukraine

At the onset of war, 77% of Kyrgyz migrants worked in Russia and 3% in Ukraine, so the risk of becoming "collateral" was present (Zeiner, 2023; Cabar.Asia, n.d.). Military recruiters approached Central Asian migrants in mosques, dormitories, and migration offices and claimed that upon joining, they would receive "an initial one-time payment of $2,390, followed by salaries of up to $4,160 a month" (Najibullah & Navruzshoh, 2023). Formal work status in Russia with a high salary is hard to turn down for many Central Asian migrants. Recruiters have also promised the attainment of citizenship within six months (Najibullah & Navruzshoh, 2023), while threatening deportation and five-year entry bans for those who refuse. In an effort to find additional soldiers, Russia also coerced incarcerated Central Asian migrants into volunteering for the Russian army (Najibullah, 2024). Conditions on the front are abysmal. One Kyrgyz migrant said, from the stories he has heard, that "you are stuck there until you're dead or severely disabled" (Najibullah, 2024).

Additionally, since the war started, there has been a concurrent rise in anti-immigrant sentiment in Russia that has reportedly caused many Central Asian migrants to leave Russia (Mahon, 2024). Though the Kremlin officially blamed Ukraine for the terrorist attack in March 2024 on Crocus City Hall that killed 140 people, the four suspected gunmen were from Tajikistan, thus sparking a wave of discrimination (Rickleton, 2024). Russia also introduced new restrictions on foreign nationals, especially worsening prison consequences for illegal migration or forgery of documents (Осмоналиева, 2024). Raids for checking documents are frequent, and people are frequently stopped on the street and jailed if found without the proper documents (The New Humanitarian, 2015).

However, news reports state that the appeal of the Russian labor market had soured before the crackdowns—the Crocus attack was a "watershed" moment for these pressures to come to a head (Ibragimova, 2024). There has long been labor exploitation and abuse of Kyrgyz migrants while abroad (The New Humanitarian, 2015). Reports say that rising attacks on migrants and the stagnating Russian economy are the two main reasons for Central Asian migrants going to other places instead: "[the number in] Turkey, South Korea, and Gulf states is already in the hundreds of thousands" (Rickleton, 2025; Ozat, 2023; Ibragimova, 2024).

In contrast, many Russians wishing to avoid conscription also flooded into Kyrgyzstan, producing a highly-skilled labor shock (5). Many Russian migrants arrive with a remote job or savings (Muzhykov & Turgunova, 2020). An estimated 750,000 to 1.5 million Russian refugees flooded into Central Asia, escaping authoritarian crackdowns, sanction effects, and conscription into the Russian army (Eurasianet, 2024). Most of the outlook on this phenomenon has been positive, citing the high-skill labor injection into these Central Asian economies as a huge boon to their development, if harnessed properly (Oshchepkov, Tilekeyev, & Gerry, 2024). However, this has less desirable effects on hiking real estate prices, driving some Kyrgyz households from the city center, which may act as another push factor for Kyrgyz migrants to look for work abroad (Babadjanov, 2022). The mass exodus of Russians has also contributed to labor shortages in Russia, which is reported to be offset by migrant workers (Korsunskaya, et al. 2024; Lyrchikova & Papachristou, 2025). The following analysis focuses on conscript-substitution by focusing on correlations with confirmed casualties, but wartime fluctuations apart from this (for instance, coefficients on the war and post-war dummies) should encode net dynamics.

2.3. Availability of data

This paper is necessary as an exploration of a data-poor situation, where official numbers are unreliable or unavailable. Papers studying labor migration from Post-Soviet countries to Russia tend to rely upon government data sources, including Central Bank data from Russia and the source country (Irnazarov, 2015; Kupets, 2012; Palacin & Shelburne, 2007; Çetintaş & Baigonushova, 2018; Tan Keok, 2020) or national statistical agencies such as the Kyrgyz Statistical Committee (Çetintaş & Baigonushova, 2018; Uncenta, 2024) and Federal Migration Service in Russia (Tan Keok, 2020). Scholars have also used data from international organizations, such as the Commonwealth of Independent States website (Çetintaş & Baigonushova, 2018), International Organization for Migration (IOM) and Organization for Security and Co-operation in Europe (OSCE) (Ahmed & Pawlowski, 2013), and United Nations Educational, Scientific and Cultural Organization (UNESCO), World Health Organization (WHO), and World Bank Development Indicators and KNOMAD dataset (Uncenta, 2024, *link*). These migration data tend to be yearly rather than quarterly or monthly, and they fail to disaggregate by national origin or Russian region destination. The approach used within this study to extend time series becomes necessary to address both of these deficiencies.

While these sources are consistent and complete, they sometimes underestimate true labor dynamics. One reason is due to informality, as previously described. This culminates in migration trends and their estimated effects on Kyrgyzstan being exceptionally difficult to pinpoint. Second, Russian data are widely regarded as unreliable and sometimes directly fabricated, especially for GDP measures, as a form of political propaganda (Plastun & Vorontsova, 2024; Sonnenfeld, et al. 2023; Sonnenfeld, et al. 2022; Iashchenko, 2023). Nearly 600 datasets were reported to be removed from government websites after the onset of the war, and there may be far more (*State,* 2024). To obtain Russian government datasets for this study, a VPN with Russian geolocation was used to bypass firewalls (Orphanides, 2019).

Many studies leverage national survey data as well for a more granular, individual, and qualitative look at migration dynamics. These include the Modular Population Survey in Ukraine (Kupets, 2012), Life in Kyrgyzstan Survey (Wang, Hagedorn, & Chi, 2013; Karymshakov et al. 2011), the Kyrgyz Integrated Household Survey (Ukueva, 2010; Mogilevskii, Thurlow, & Yeh, 2018), the GIZ Survey in Kyrgyzstan (Irnazarov, 2013), or authors produce their own survey and interview data (Scott, Sexmith, and Chi, 2024; Critelli, et al. 2020; Tynaliev & McLean, 2011).

One data source that was considered but not used in this paper is the Multi-Indicator Cluster Survey (MICS), which is available in Kyrgyzstan for 1995, 2006, 2014, 2018, and 2023. This range of years would, in theory, allow comparison before and after the war to triangulate impacts. However, there are no direct remittance variables, only a question about whether a child's parent is living abroad and where they currently live. This is more useful for analyzing outcomes of children or spouses left behind by migrants, rather than for the topic of this paper, which focuses on war-related regional and macro impacts over time.

Only a handful of sources have looked at migration in the post-Soviet world using unconventional data sources like Google Trends (Anastasiadou, Volgin, & Leasure, 2023). Outside of the Post-Soviet world, Google Trends is well-worn as an indicator of travel aspirations, which can be a leading indicator for migration (United Nations Global Pulse, 2014; Böhme, et al. 2020; Tjaden, Auer, & Laczko, 2019; Avramescu & Wiśniowski, 2021). Yandex is the most popular search engine in Russia (Statista 2024), but only one English-language study and nine Russian-language studies were found that used Yandex Wordstat in the social sciences literature. This may be because Google Trends has a longer historical period, Yandex Wordstat is simply a newer tool, and Yandex Wordstat is only really relevant in the Russian-speaking world.

Yandex Wordstat is an impressively rich and promising data source. While Google Trends uses relative frequencies to protect user privacy, Yandex Wordstat has no such regard for protecting user identities and reports down to absolute daily counts. This richness of the data has much greater precision for tracking search activity, especially for small quantities. As previously mentioned, there is only one piece found in the English-language literature that has used Yandex Wordstat for economic analysis (Anastasiadou, Volgin, & Leasure 2023).

In the Russian-language literature, Tsapenko and Yurievich (2022) evaluate the usefulness of using Yandex Wordstat and Google Trends for not just nowcasting but *predicting* migration patterns from Central Asia to Russia, finding a significant correlation between keywords and future migrants (Tsapenko & Yurievich, 2022). The paper also finds that typical econometric methods are insufficient for predicting migration, due to a lack of robustness to measurement error and reliance on questionable assumptions (like linearity and homoskedastic errors). They also recommend using machine learning techniques instead, which is precisely what this study fulfills in the following sections, using a similar but expanded set of keywords. They use a monthly inflows dataset beginning in 2011 that was likely taken down after the onset of the war. However, the authors recognize the superior disaggregation by migrant type for one dataset that this study *does* use for nowcasting: quarterly Kyrgyz visitors to Russia from 2010–16, extended using SARIMAX to 2024Q3.

Other Russian-language papers have used Yandex Wordstat to evaluate discrimination against migrants in the aftermath of the Crocus Hall attack and crackdowns (Четаева, 2024; Iakushkina, 2021). Beyond this, the only papers found that applied Yandex Wordstat to migration were to track internal Russian migration (Shchitova, 2020) and evaluate the potential for tourism development (Шайхеева, 2015; Sidorchukova, et al. 2024). Others focus on social capital growth (Баркова, 2024), influenza spreading (Коршунова, 2025), and semantic indexing to evaluate accessibility in particular Russian cities (Агранович, et al. 2019). The Tsapenko and Yurievich (2022) paper remains the only consequential paper for this analysis, and the fact that so few studies even in the Russian language have leveraged this database shows the untapped potential of these methods. Apart from search query data, some studies have looked at migration of high-skilled Russian labor using unusual data such as GitHub location tags and job listings (Wachs, 2022).

3. Theoretical Framework

The basic push and pull factors well-established in economic theory of international labor migration can be summarized simply as the fact that people move where expected wages and standards of living are higher (Becker, 2007). This is easily applied to the Central Asia-Russia pipeline: expected wages are historically much higher in Russia, especially since unemployment was very high in the Central Asian republics in the decades after independence. Importantly, wages should adjust as the volume of workers shifts between locations: a smaller (unskilled) labor force in Russia due to conscription would culminate in a higher wage demanded by workers in the same sectors, with companies looking to hire more cheap labor from abroad.

Importantly, the informality of this labor market implies there are lower administrative costs of migration, but high immigration barriers play a counteracting factor in the Russian case (Bosch & Farré, 2013; Ryo, 2013; Villarreal & Blanchard, 2013). Administrative, transportation, and psychological costs create friction such that the wage differential is usually not directly proportional to the volume of migration, but certainly contributes to increased flows. This is also vice versa: many migrants choose not to return home because barriers exist to their return, and they may not be able to return without revealing their illegal status.

Additionally, theory corroborates the fact that people tend to move away from disturbances—whether at home or in the host country, and especially when they are temporary or seasonal migrants (Bertoli, Moraga, & Keita, 2016). Among other things, this study estimates the elasticity of the labor supply curves according to these disturbances: when Russia wages a war that leads to disruption for migrants, how many return home, and how many stay due to the relative strength of their position in Russia compared to Kyrgyzstan? The regressions also explore labor demand elasticities: when an entire generation of young men are conscripted in Russia and a fraction of them die on the front lines, how much do hires of Central Asian migrants increase?

Collective decisionmaking also has a strong basis in the newer migration literature: migrants will often decide alongside family members whether to travel to Russia and handle risk according to whether they have a community awaiting them abroad or back home (Massey, et al. 2008; Galstyan, et al. 2024; Bilecen & Lubbers, 2021). Therefore, regions with strong Kyrgyz communities already, such as Moscow and Sakha, should have more robust migrant inflows accordingly.

4. Data

Macroeconomic indicators were taken from a variety of sources. Many indicators were drawn from theglobaleconomy.com, which aggregates and validates a wide variety of data for use by scholars (TheGlobalEconomy.com, n.d.). For both Russia and Kyrgyzstan, relevant indicators are drawn from the Organisation for Economic Co-operation and Development (OECD), the Bank for International Settlements (BIS), and Google Finance for exchange rates. For Russia, the majority of data is drawn from the Federal State Statistics Service of Russia (Rosstat) and the Central Bank of the Russian

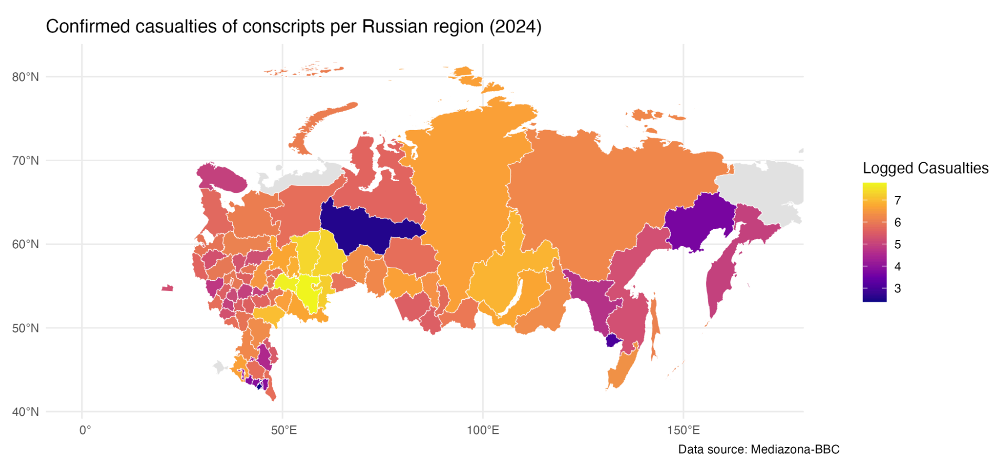
Federation, and for Kyrgyzstan, these include the National Statistical Committee of the Kyrgyz Republic and the National Bank of the Kyrgyz Republic. Beyond this database, further data is procured directly from the national government entities, such as the quarterly 2010–16 Kyrgyz visitors data drawn from the Federal Security Service of the Russian Federation (FSB), the 2021–23 Kyrgyz migration flows data drawn from Rosstat, and 2005–2025 monthly remittances from Russia to Kyrgyzstan drawn from the National Bank of the Kyrgyz Republic. Rosstat provides the most consistent set of indicators describing the Russian economy and demographics. Thus, this study relies in part on Rosstat for metrics like changes in industrial production, wage level, and so on, supplemented with private data from international organizations. Additionally, political bias should be less pronounced for the 2010–16 quarterly series due to greater political stability, thus, it is advantageous that the nowcast for quarterly visitors is trained only on data from this timeframe.

Seasonal dummies from the Muslim calendar are also included in most regressions to control for Ramadan effects or effects associated with Eid-al-Adha, called Kurban Ait in Kyrgyzstan (Advantour, n.d.), which is a well-celebrated holiday alongside Kyrgyz independence day in August (Chamberlain, 2022).

Yandex Wordstat and Google Trends data were procured for the same set of variables which paired Russian- and Kyrgyz-language queries and transformations of them. For instance, the word for work in Kyrgyz is “жумуш” (romanized: *jumush*) and in Russian is “работа” (romanized: *rabota*). The total volume of queries and the average frequency (percentage relative to maximum number of queries) for each keyword were extracted from both websites for the time period necessary. The change in volume and average frequency were also included. Lastly, the difference between the Russian- and Kyrgyz-language queries and ratio between them were also tabulated and included. Work-related search terms in Kyrgyz and Russian, as well as terms like “ticket Bishkek” or “ticket Kyrgyzstan” were also included. Money transfer words were also incorporated, including popular financial transfer companies: Zolotaya Korona and Unistream. This data will be the basis for these nowcasts: Yandex Wordstat will be used for extending the 2021–23 flows dataset due to its popularity among Russian-speaking users and availability of precise search query counts. For the 2010–16 quarterly visitors set, Yandex Trends is unavailable before 2018, so its nowcast is limited to Google Trends.

To evaluate labor substitution of migrant workers with conscripts, this study would benefit from data on the number of drafted individuals per each region. Unfortunately, conscription numbers are not consistent per region, and announced numbers are likely unreliable (Савина & Бонч-Осмоловская, 2022). There are also no official numbers announced for each region (Meduza, 2022). Luckily, a Mediazona-BBC collaboration project provides daily, region-disaggregated casualty data (Figure 1), which can reasonably proxy conscription trends with a lag. Given the lack of better alternatives, this study focuses on the more morbid version of the theory—"casualty substitution"—and uses it as a proxy for the suspected conscript substitution effect. Casualties may also reflect war intensity in a given region and period, indicating both the scale of conscription and the psychological impact of frontline deaths. These data are used to assess how migrants and remittances respond to both war intensity and the labor market vacuum left by conscripts, at region-by-year and national (quarterly and monthly) levels.

Figure 1: Map of Mediazona casualties in Russian regions in 2024



4.1. Univariate Exploratory Analysis

A summary of target variables, sample size, time frame, and nowcasting strategy is included below (Table 1). A summary statistics table for all sets of monthly, quarterly, and yearly variables is available in the online appendix. Correlation matrices demonstrate that nothing is perfectly (multi)collinear.

Table 1: Summary of Target Variables and Nowcasting Strategies

Target variable	Description	Freq.	Original Time frame	Ext'd Time frame	Method for Nowcast	Data for Nowcast
Kyrgyz inflow/ outflow/ net flows (3 series)	Number of arrivals and departures of Kyrgyz travelers, as well as the net change	Yearly, Regional	2021–23 *N=243*	2019–25 *N=486*	XGBoost (ensemble gradient boosting ML model)	Yandex Wordstat
Kyrgyz Visitors (4 types)	Number of Kyrgyz arrivals, disaggregated by type of traveler: "all," business, transit, private.	Quarterly	2010–16 *N=28*	2010–2024Q3 *N=59*	SARIMAX (due to small sample size)	Google Trends and economic covariates

The 2021–23 flow series from Rosstat benefits from regional variation and a large sample size of N=243. Maps of migrant inflows, outflows, and net flows in 2023 (Figure 3) demonstrate that there are high inflows but also high outflows in the far eastern regions of Sakha and Krasnoyarsk. The net flows graphs demonstrate that Primorsky and Kaluga exhibit the highest net departures, though most regions demonstrate positive net arrivals in 2023 despite the presence of war. Histograms of these flow variables are skewed toward zero, with very few regions breaking 2500 inflow or outflow counts, though net flows are clustered around zero (Figure 4). Average inflows across all regions (Figure 5) in the country increase very slightly over all three years, while average outflows peak in 2022 at the onset of war before decreasing somewhat, resulting in net flows also hitting a minimum in 2022. Analyzing trends for particular regions, a different urban and rural effect emerges between the popular destinations of

Moscow Oblast, Sakha (Yakutia), and Sakhalin. Sakha (Yakutia) seems to have more of an increasing trend to inflows, while inflows are decreasing in Moscow Oblast. Sakha also has a larger jump between 2022 and 2023 in outflows. Sakhalin's trend is more moderate, with an increasing trend in outflows that results in net flows falling to approximately zero in 2022 and not recovering.

Figure 3: Maps of migrant inflows, outflows, and net flows in Russian regions, 2021–23.

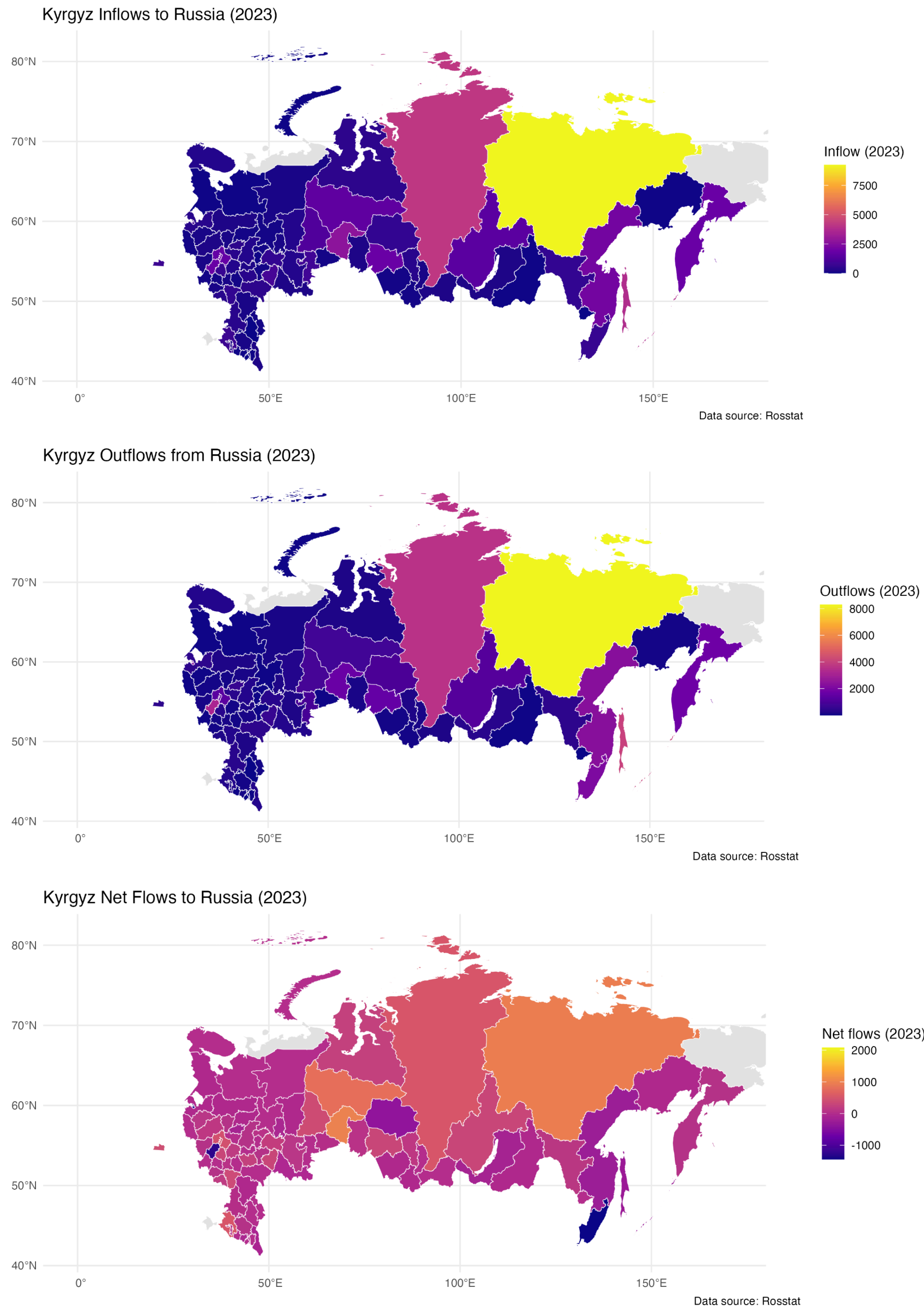
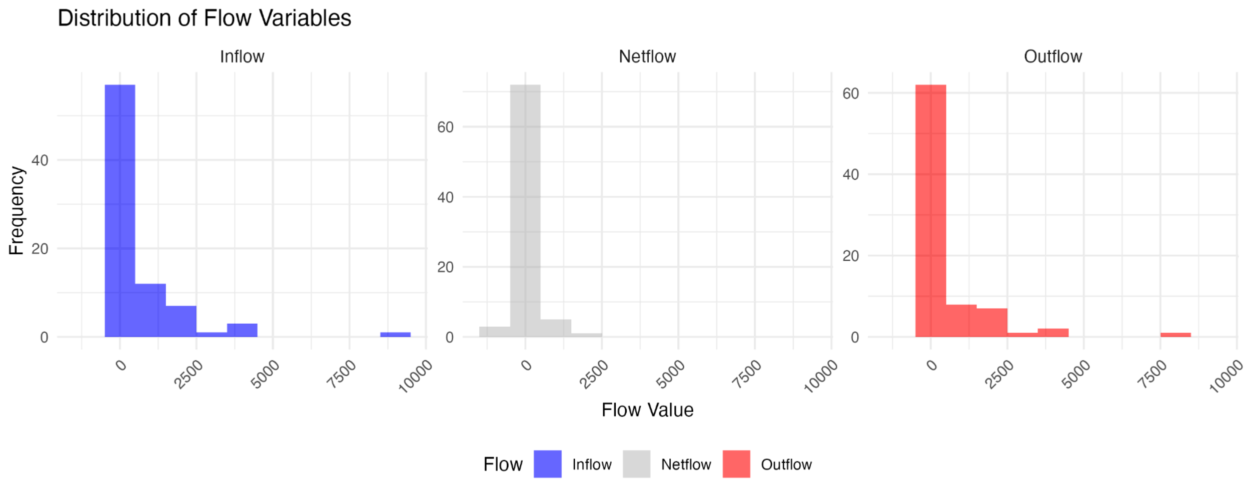


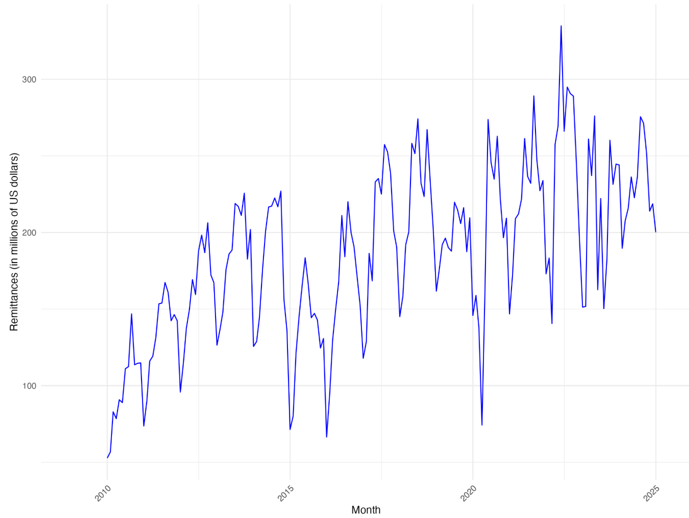
Figure 4: Histogram of inflows, outflows, and netflows of Kyrgyz migrants into each Russian region in 2023, according to data from Rosstat.



The Kyrgyz visitors by type series has a small sample size of N=28, recorded quarterly between 2010–16. Plots of logged values (Figure 8) demonstrate log-linear increasing behavior for the "all types" and business visitors series. Transit travelers and private travelers seem to have slowing growth rates in the time period of the original data, with transit travelers falling to around 1000 per year by 2016. The original data series does not provide descriptions of these types of visitors, only labels. It is fair to say that business visitors would reflect more white-collar workers (though may include some blue-collar workers), while transit and private travelers may include individuals traveling to Russia through chain migration or with the true intention of circumventing legal restrictions by overstaying beyond the allowed period.

The monthly remittances series includes primarily electric transfers, but provides a note that non-electric methods were also included beginning in 2022. A plot of monthly remittances in millions of USD (Figure 2) demonstrates distinct seasonality that breaks down after the onset of war in 2022. For regressions, the data is limited to the timeframe of 2011 to September 2024 (N=165) to maintain the availability of enough covariates for unbiased analysis.

Figure 2: Time series of remittances to the Kyrgyz Republic
source: National Bank of Kyrgyzstan



5. Methodology

5.1. Notes on Preprocessing

These results are at risk for heteroskedasticity distorting model results, since the key dependent variables are skewed toward zero with several high-leverage outliers. One way to immediately address heteroskedasticity is to apply logarithmic transformations, which reduce this skewed behavior. The problem at hand would also benefit from a logged dependent variable, since it is more relevant to interpret coefficients that describe "percentage change in y" rather than simple changes in levels (UCLA Statistical Consulting Group, n.d.). For the large collection of potential covariates to include in the regression, all applicable variables (those that do not take negative values) will also receive a log+1 transformation, since these variables are highly liable to skewed distributions (West, 2022). Where reasonable, those taking negative values are removed to prevent them from having undue influence on the regression. For rates and percentages, log transformations are not necessary because they are not liable to the same distortions.

To test for further heteroskedasticity, the Breusch-Pagan test is run after each regression (Lee, 2025). These usually all indicate heteroskedasticity, even with logged variables. A residual plot is also produced and examined, with a non-uniform distribution of residuals providing evidence of heteroskedasticity. Almost all regressions have a clustering at the center with outliers driving heteroskedastic confirmation. The full results of these tests are available in the online appendix, though positive results are common for all regressions. Thus, robust standard errors are implemented for all models, which are heteroskedasticity-robust.

It is well-known that HC2 and HC3 errors have superior finite-sample properties compared to the classical White errors (HC0) and HC1 errors (MacKinnon & White 1985, Long & Ervin 2000). HC3 errors specifically account for high-leverage points in small samples (Hayes & Cai 2007, Williams 2022). High-leverage points are certainly an issue when it comes to particular regions in Russia having a greater quantity of Kyrgyz migrant visitors each year than others, which the model has trouble correcting for in the absence of region fixed effects (which are unattractive because the sample is too small for the migration datasets). Therefore, this study uses HC3 robust standard errors for the subsequent analysis to ensure that the statistical power of the tests is as high as possible, and HC1 if the HC3 errors are not possible.

Though ordinary least squares regressions are not sensitive to scaling (beyond the coefficients themselves being liable to scaling), the machine learning algorithms used for nowcasting *are* sensitive to scaling differences, which could change the amount of weight attributed to each variable and distort the incremental learning algorithms (Vashisht, 2021). Therefore, when using these machine learning methods, basic normalization is required for the scale of the non-dummy independent variables to produce a mean of 0 and a standard deviation of 1. This allows every variable to have approximately the same distribution while encoding the sense of "extremeness" that the variable takes for that particular observation. The dummies can remain with their existing distribution of [0,1], since the value of a 0 or 1 encodes important information about presence or absence of the particular variable. The scale is fit to the training data, then applied to the validation and test sets—otherwise risking data leakage. This scaling was implemented using the usual method of the scikit-learn StandardScaler capabilities in Python (scikit-learn Developers, n.d.), and the code with summary statistics can be seen in the online appendix.

Given the wealth of covariates available, the Bayes Information Criterion (BIC), supplemented by the Akaike information criterion (AIC), performs variable selection on the regressions to optimize model fit. A stepwise process (starting with a full regression and removing predictors one-by-one that most poorly describe the response variable until an optimized model fit is reached) conserves time otherwise dedicated to estimating and presenting several regressions to determine the best fit, and also serves to select the best possible model (Yamashita, et al. 2007). The BIC more rigorously penalizes excess parameters, and is asymptotically consistent in optimizing log-likelihood of approximating the "true model" so long as the necessary predictors are available (Singh, 2024). Thus regression interpretations will defer to the BIC and supplement with the AIC (which is better for forecasting) to add depth. This is primarily a data-driven approach to evaluate the strength of particular variables in explaining the response variable, though economic theory was used to determine which covariates would be important to obtain and include in the regressions in the first place, and many variations are specified and reported in this study.

5.2. Using XGBoost to nowcast flows

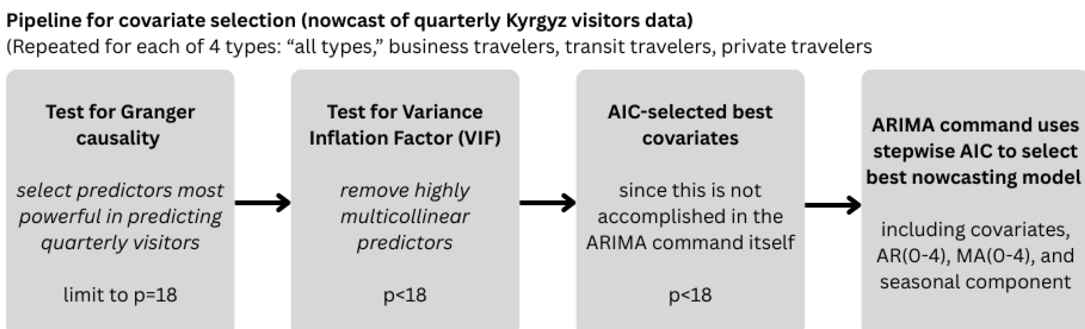
XGBoost is an ensemble gradient boosting algorithm that builds decision trees sequentially, with each tree trained to correct the errors of the previous ones (NVIDIA, n.d.). The technique has been implemented for other nowcasting studies in the economics literature. A 2020 study successfully used tree-based models like XGBoost and Random Forest to analyze how shocks affect migration patterns in agriculture-dependent African countries, using not search query data but rather weather data (Aoga, et al. 2020). The study, too, found that socioeconomic data was the most important predictor in these models for accurate estimations of migration intentions, though weather did help to improve nowcasts. Just as the analysis in this paper will demonstrate, a 2023 working paper by Boss, et al. demonstrated that combining XGBoost and Random Forest "consistently outperforms forecast horizons between 3 to 12 months," especially ARIMA models which assume stationarity that may not entirely hold in the migration setting (Boss, et al. 2023). Other machine learning models have also been shown to perform better than linear methods in this space, including Artificial Neural Networks (ANNs) which excel in modeling non-linear relationships (Gawryluk, et al. 2024), Long Short-Term Memory (LSTM) networks (Golenvaux, et al. 2020), and Dynamic Elastic Net Models (Carammia, et al. 2022). Machine learning models have also been crafted to allow for time-variant parameter estimates, even allowing for local stationarity (Ng, 2021).

In the paper, the scikit-learn pipeline (scikit-learn Developers, n.d.) is used to train and tune a series of five options: Neural Network, XGBoost (using the relevant package, (Chen & Guestrin, n.d.)), Stochastic Gradient Descent, Random Forest, and Elastic Net. Train-validation split was accomplished by training on the 2021–22 data and validating performance on the 2023 data to preserve the time dependence structure. Validation RMSE was used as a key metric to select the top models, then the plot of prediction versus actual value was used to determine if the model was actually making high-quality predictions and—importantly for the regressions to follow—at least correctly guessing the rough ordinarity of the points (for instance, that fewer migrants went to the Jewish Autonomous Oblast than Moscow in 2023). XGBoost is selected and used in subsequent analysis.

5.3. Using SARIMAX to nowcast quarterly Kyrgyz visitors

One of the most crucial tasks at hand will be to nowcast quarterly Kyrgyz migration, gaining a fine-grained sense of what is happening on the ground in the fallout of the war onset and Crocus Hall attack. Thus, we extend the original dataset (2010–16) to 2024 Q3. An overdetermination problem arises: 28 observations but around 60 covariate candidates to help fine-tune the nowcast for the next 32 quarters, beginning with 2017 Q1. This leads to OLS and thus the stepwise AIC used for model selection being unidentified due to having more predictors than samples. Therefore, the covariates must be narrowed down to a smaller number to make sure estimates are stable.

Figure 7: Flow chart for covariate selection in SARIMAX nowcast of quarterly visitors series.



Therefore, to balance avoiding multicollinearity with maintaining good predictor potential, a four-step process is required for covariate selection and model specification, outlined in a flow chart diagram (Figure 7). First, Granger causality is used to assess whether past values of covariates provide statistically significant information for predicting current logged remittances, using grangertest() from the lmtest package in R (Zeileis & Hothorn, n.d.). Only the 18 most promising covariates are carried forward, the maximum that can be fed into this pipeline without having overly multicollinearity-biased estimates that produce unreasonable estimates, balancing with the desire to feed as many covariates as possible into the pipeline for optimal selection. Next, covariates with very high variance inflation factor

(VIF) values are removed, stabilizing estimates as much as possible so the AIC can select the optimal forecasting model. Finally, the correct order of AR(1-4), MA(1-4), and seasonality are chosen with stepwise AIC via the auto.arima() command. AIC was used instead of an Elastic Net or more sophisticated regularization method due to the inclusion of confidence bands innate to the model pipeline and a desire to avoid overfit, which a more sophisticated method risks doing on such a small sample. Letting Ψ be the coefficient vector for Y_t, the BIC-selected covariates (including quarter dummies), and letting q be the current quarter the regression for visitors replicated for all four types of visitors would be as follows:

$$ln(\widehat{visitors}_q) = \mu + \Psi Y_q + \epsilon_q;\ q = 2010Q1,\ \ldots,\ 2016Q4.$$

Though there is still likely overfitting in the model results, this four-step process was determined to be the best possible pipeline to minimize the influence of noise. Without an expanded sample size, this is the best that can be reasonably done to address this small sample size and obtain as unbiased a model as possible. Augmented Dickey Fuller (ADF) tests were used to demonstrate stationarity and appropriateness of the SARIMAX approach, with business travelers and transit travelers rejecting null of a unit root at a <5% significance level, and private travelers and "all types" rejecting at the <1% level.

The approach above leads to stable nowcasted values for four types of visitors: "all types", business visitors, private travelers, and transit travelers—both with only Google Trends (called Model A) and with adding economic covariates (Model B), just like for the flows set. Model A is reliably more susceptible to noise, and Google Trends search data is rarely chosen in Model B, reflecting its sparsity and unreliability in the Russian context compared to Yandex Wordstat. Google Trends is available starting from 2005 while Yandex Wordstat extends only from 2018, which is why Google Trends is used in this context.

5.4. Methodology for analyzing inflows, outflows, and net flows with regional data

The first model is a basic difference-in-difference (DID) regression without covariates to obtain a baseline sense of war effects on the original 2021–23 dataset. The BIC/AIC was not applied for this initial specification, since all components are essential to the interpretation of the model. The variables included will be a "war presence" treatment dummy (0, 1, 1 in this very short timeframe), a "post-war" dummy to measure number of years since war onset (only 0, 1, 2 in this case). Each regression was initially run with regional fixed effects, and these results can be seen in the online appendix in the interest of space. Results were determined to be too volatile due to the small sample size—N=3 per region in the original series and N=6 in the nowcasted series. A linear time trend was considered to gain a sense of exogenous increase, but this was determined to be multicollinear with the other variables previously mentioned and thus dropped to maintain the clarity of inference for this basic case.

Next, using the approach outlined above with BIC/AIC-stepwise selection and log transformations, a set of region-by-year covariates is added to the model. Finally, an additional covariate is added to the previous specification: the Mediazona casualty data. For inflows and outflows, this analysis is repeated for the extended time series spanning 2019–24, generated using only Yandex Trends data (not economic covariates) to remove the risk of data leakage and endogeneity clouding inference. Let Φ be the coefficient vector for X_t, the BIC/AIC-selected covariates, letting α be the intercept, and let y be the current year (2021–23 for the original set and 2019–24 for the nowcasted set) for the Russian region *r*. The regression for inflows (replicated for the other dependent variables) would be as follows:

$$ln(\widehat{inflows}_{y,r}) = \alpha + \Phi X_{y,r} + \epsilon_{y,r};\ y = 2019,\ \ldots,\ 2024.$$

The full output, residual plots, and Breusch-Pagan tests can be seen in the online appendix for every regression run in this text, including this one. Throughout the model tables, Model A is the basic specification with just the war effects included, Model B has the BIC/AIC-selected economic covariates, and Model C adds casualties into the pool of potential covariates. The only difference between tables are the five different dependent variables switched out: inflows and nowcasted inflows, outflows and nowcasted outflows, and net flows (alone, because the nowcast fails to converge).

5.5. Methodology for analyzing national, quarterly Kyrgyz visitors data

To quantify war-related changes to the nowcasted Kyrgyz visitors data, the newly extended series (through 2024Q3) is regressed on quarterly dummies, Muslim calendar dummies, war effects (and post_war), and Crocus Hall attack dummies (representing the period after the attack occurred in April 2024, when many Central Asians were deported and discriminated against in Russia). The BIC stepwise algorithm approximates the true model, as before. Letting Φ be the coefficient vector for X_q, the BIC-selected covariates (including quarterly dummies), letting α be the intercept, and letting q be the current quarter, the regression for visitors (replicated for all four types of visitors) would be as follows:

$$ln(\widehat{visitors}_q) = \alpha + \Phi X_q + \epsilon_q; q = 2010Q1, \ldots, 2024Q3.$$

5.6. Methodology for analyzing remittances

The remittance data has the advantage of not only a monthly frequency, but the National Bank of Kyrgyzstan produces the dataset, meaning that it does not have the same risk of political interference by the Russian government. To get a sense of war effects, Crocus Hall effects, and casualty effects, a series of three core regressions are run with BIC-stepwise selection on these components. This approach is repeated with Russian and Kyrgyz CPI included to determine if "real" remittances went up (only Russian CPI is selected in these regressions, as it should be, according to economic logic of Russian inflation being directly related to the value of remittances being sent back). Letting Φ be the coefficient vector for X_m, the BIC-selected covariates (including monthly dummies), and letting *m* be the current month, the regression for visitors (replicated for all four types of visitors) would be as follows:

$$ln(\widehat{remittances}_m) = \alpha + \Phi X_m + \epsilon_m;\ m = Jan\ 2011, \ldots, Sep\ 2024$$

The results nodded toward two possible explanations for remittance increases—either higher wage rates or increased migration. Therefore, the same regressions were run with controlling for nominal gross wage rates in all of Russia as well as three distinct regions where Kyrgyz migrants frequently travel: Moscow Oblast, Sakha, and Sakhalin. This serves to "net out" the influence of wages, leaving the influence of additional migration via a labor substitution effect as the best explanation for positive coefficients on the casualty variables. These were also run with the bilateral exchange rate between the Kyrgyz Som and Russian Ruble as an additional control afterwards.

5.7. Summary of Limitations

The small sample size, though counteracted by the nowcasting approach, is the largest problem with this methodology as a whole. The standard errors and validations RMSEs are large in the XGBoost model, and for the quarterly visitors data. A more sophisticated model than the SARIMAX is impossible without overfitting, but limiting the nowcasting model to an OLS framework imposes an assumption that the independent variables are linear in their impacts on the response, likely not aligning with the underlying data generation process. However, this is why this study tackles the problem of estimating Kyrgyz migration from three different directions: three different data sources (Rosstat 2021–23 inflows, outflows, and net flows extended using XGBoost on Yandex Wordstat; FSB 2010–16 visitors by type extended using SARIMAX on economic covariates and Google Trends; Kyrgyz National Bank remittances 2005–2025 with no extension necessary), data frequencies (region-by-year, quarterly, and monthly), and analyzing with the same regression structure for effective comparison (using the BIC to select a model as consistent as possible with the true model). The fact that the three models align in terms of inference, especially for the labor substitution effect more so than the war or Crocus effects, can provide compelling evidence that the results are not spurious. Additionally, estimates align with trends reported in news articles, providing further evidence of validity.

Further research efforts should search for an original longer time series, either through using other unconventional data sources, or perhaps official data will be available in the future. For the poor data quality available, the XGBoost methodology does an exceptional job of nowcasting, and significant precautions are taken to account for heteroskedasticity and model misspecification. Though all findings should be taken with these limitations in mind, this study design does the best possible with the available data.

6. Results

6.1. Results for nowcasting the 2021–23 regional flows series

For both inflows and outflows, XGBoost outperforms alternatives in the Yandex Wordstat data setting (without any economic covariates). XGBoost had some high-leverage poorly guessed outliers, but overall preserved ordinality best, even when the Neural Network had a slightly better validation RMSE in many cases. Consistent with the dependent variable of migrant inflows and outflows being non-negative, it also predicted almost always non-negative points without any explicit constraint being placed on the model. Please see these model summaries for inflows and outflows (Figure 6).

For the net flows data, the Yandex data were unable to produce an accurate enough estimate for further influence due to the fact that the data fluctuated around zero and was unable to be logged before the regression, making it difficult for models to converge on an accurate estimate. Thus, predictions for net flows were not used in subsequent regressions.

Figure 5: Time series of inflows, outflows, and net flows for three location subsets: selected examples for three of the most-visited Kyrgyz regions: Sakha (Yakutia), Sakhalin, and Moscow Oblast, and an average across all Russian regions for the same period. The original data were overlaid with the nowcast made with XGBoost and Yandex Wordstat search query data, showing in-sample fit.

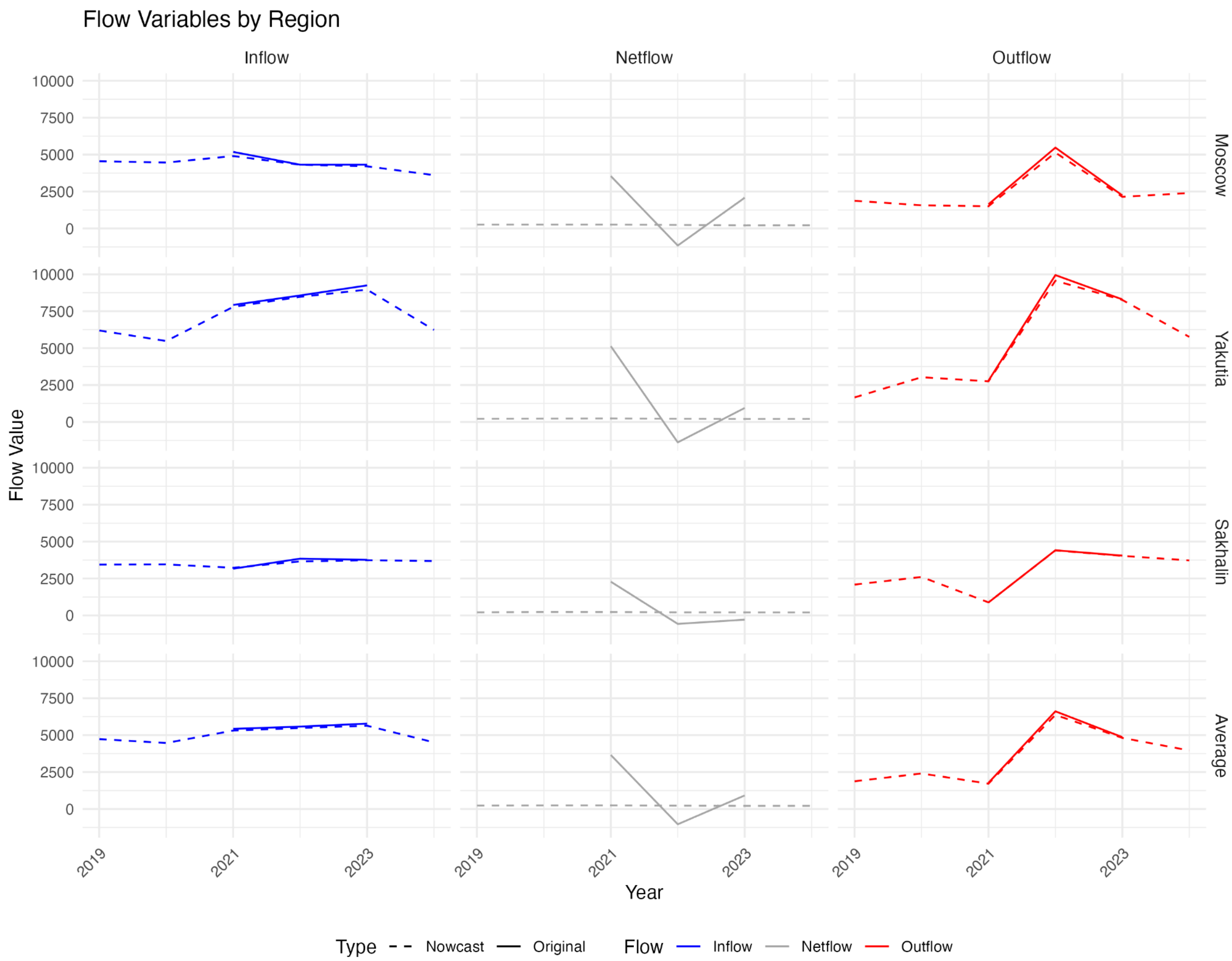
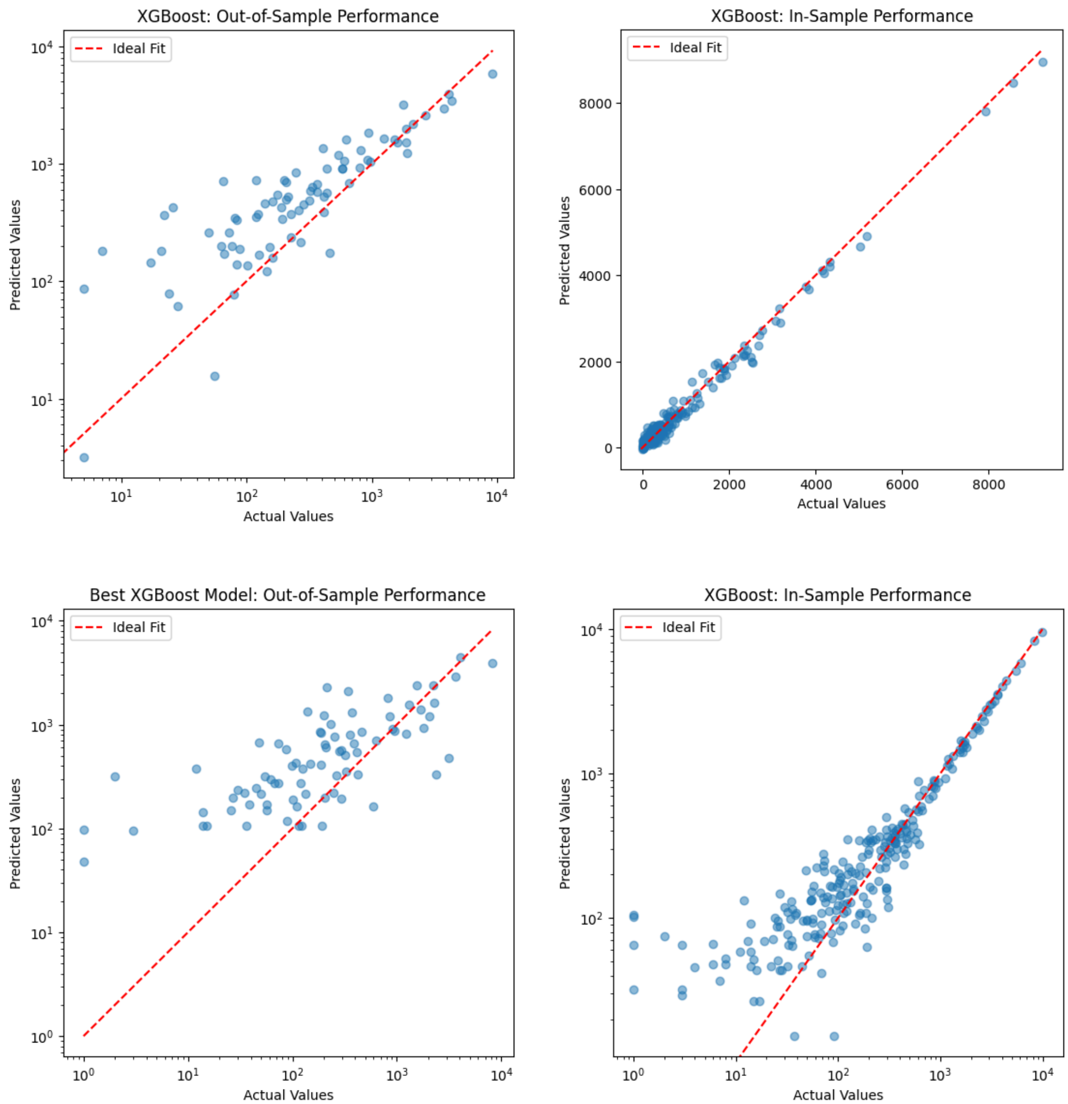


Figure 6: Accuracy of XGBoost-nowcasted inflows and outflows of Kyrgyz Migrants to Russia using only Yandex Trends data.

<u>*Out-of-sample performance on validation set (2023 data)*</u> *Model trained on 2021–22 data*	<u>*In-sample performance*</u> *Model trained on the full dataset*

Inflows

Outflows



Economic covariates are then added into the covariate pool temporarily. These estimates cannot be used for further regressions due to the risk of data leakage from the trained data making its way into the regression results on the same predictors. However, it is useful to examine the influence of these covariates on the model estimation and enhance the accuracy of the nowcast for the purpose of obtaining a reasonable point estimate. Further details are available in the online appendix. For inflows, a neural network was chosen as the best model, and XGBoost was still superior in the outflows setting. The Elastic Net model performed best with the net flows data, but the validation MSE was too poor for serious consideration. Speaking to the influence of the casualties variable in predicting migration flows, the inflows model selected casualties and lagged casualties within the top 10 most influential variables when minimizing in-sample error in the final model out of a pool of 267 possible covariates (including search queries and their transformations, region fixed effects, and socioeconomic variables). The net flows model selected casualties as the absolute most influential variable.

Bootstrapping produced confidence bands that usually overlapped zero except for extremely large estimates. This reflects that most of these models have trouble distinguishing between smaller values and do much better with predicting larger migrant counts. It is very important to recognize this limitation when interpreting the nowcasted data. Much of this has to do with the scarcity of data—both covariates and dependent variables. It is remarkable that the gradient boosting methodology can develop such a high-quality model given only three years of data, reflecting the value of the technique. These estimates are certainly imperfect, but they at the very least provide a reasonably extended time series for the purposes of this study—further regressions that take advantage of variation and ordinality of the points—rather than precise individual estimates (which are impossible in this case). The inferences using nowcasted values also typically corroborate the inferences from the original values, with some exceptions.

6.2. Results for dependent variable of inflows of migrants (Table 2)

The regressions emphasize that casualties have a much stronger relationship with inflows than the simple presence of war (called "war") or years since the war began ("post_war"). As a reminder, casualties describe soldiers conscripted from a particular region who are confirmed to have died on the front lines in that same period—recorded in the "Mediazona" variable. Casualties are selected by the BIC more than the post-war variable, providing evidence that casualties better explain wartime effects. They also absorb positive effects that were previously attributed to the war dummy. The war dummy was previously not even selected by the BIC in the absence of the Mediazona-reported casualties variable, and was estimated to have a (not significant) coefficient of 0.20 and -0.25 in the basic regression. But when including the Mediazona variable, the coefficient plummets to -1.57 for the original series and -2.23 for the nowcasted series, both significant at the <0.1% level. The casualties variable is needed to absorb the positive effect of war on migrant inflows so that the negative war effect can be registered.

Table 4: Annual inflows of Kyrgyz migrants to Russia per region
Difference-in-difference of war and post-war effects on Kyrgyzstan-to-Russia migration (2021–23, logged inflows series from Rosstat, compared to nowcasted series†) with HC3 robust standard errors.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	A	**A†**	**B**	**B†**	**C**	**C†**
war	0.20271	-0.24505	*Not selected*	*Not selected*	-1.567235***	-2.235762***
post_war	-0.19376	-0.00411	-0.650252***	-0.469852***	*Not selected*	*Not selected*
Selected covariates by BIC‡	—		log_total_pop (8.917082**) log_rural_pop (0.113944**) log_avg_nominal_gross_wages (11.078756***) lag_log_total_pop (-8.095476*) lag_log_avg_nominal_gross_wages (-8.107489**)	log_total_pop (10.065182***) log_city_pop (1.160193**) log_rural_pop (0.119949***) log_avg_nominal_gross_wages (2.425026***) lag_changeinindustrialproduction_water_utilities (-0.524917*) lag_log_total_pop (-10.510732**)	**log_mediazona** **(0.282593***)** **lag_log_mediazona** **(-0.252920***)** change_in_real_avg_wages (6.966020**) changeinindustrialproduction_all (-0.995733*) log_total_pop (10.136118**) log_rural_pop (0.095631**) log_avg_nominal_gross_wages (2.766274***) lag_log_total_pop (-9.381056**)	**log_mediazona** **(0.314573***)** **lag_log_mediazona** **(-0.147252***)** log_total_pop (11.016285**) log_rural_pop (0.108788***) log_avg_nominal_gross_wages (2.436753***) lag_changeinindustrialproduction_water_utilities (-0.534023**) lag_log_total_pop (-11.494056**) lag_log_city_pop (1.120537*)

Adj-R^2	-0.00526	0.00157	0.5762	0.5037	0.5903	0.5189

†*Extended the time series: nowcasted 2019–24 using XGBoost gradient boosting ensemble method.*
‡*Full tables with AIC-selected models and list of non-selected covariates available in the online appendix.*

Evidence for a labor substitution effect appears via a significant, positive, short-run increase to inflows in the same year as casualties are recorded from that region. “Casualties” are selected with a positive same-period coefficient (0.28 for the original series and 0.31 for the nowcast) paired with a lag with a negative coefficient (-0.25 and -0.17), both significant at the <0.1% level. This provides evidence that region-specific casualties exhibit a long-term, delayed, and positive impact on Kyrgyz migrants, consistent with the hypothesis of a labor substitution effect. At this point, the coefficient could also be simply driven by the significance of going from zero to some positive number on the onset of the war. Yet even after adjusting for HC3 errors, casualties remain extremely significant, showing that this is not just important for high-leverage points.

Each regression was run with regional fixed effects, and these results can be seen in the online appendix in the interest of space. Results were determined to be too volatile due to the small sample size—N=3 per region in the original series and N=6 in the nowcasted series. Population and industrial production controls are chosen consistently and seem to act primarily as substitutes for these fixed effects.

Total population has a positive, large coefficient, with a 1% growth in total population predicted to have between a 8.92% to 11.02% increase in migrant inflows, significant at the <0.1% level. This reflects that migrants are drawn to more populous places, of course, but this is also enhanced in rural areas (with a 1% increase in rural population associated with a significant increase in 0.095% to 0.11% migrant inflows) and city population not being as impactful. This reflects how population centers—both across Russia and in the rural far east—act as travel hubs from which migrants can disperse. Total population from the previous year carries a negative coefficient of approximately the same size—reflecting that migrants are less likely to go to places where the population has shrunk in the previous year, such as cities that have experienced mass exodus or conscription programmes, a phenomenon corroborated by news reports (Goble, 2024b). Since population estimates are typically generated from growth rates from the most recent Russian census counts, which reportedly exhibit error themselves (Goble, 2024a), this effect is at least somewhat susceptible to measurement error, but the general interpretation should still hold.

The lagged industrial production variables are selected by the BIC with the nowcasted results, and same-period industrial production is selected for the original data and combined Mediazona casualties. These all carry a significant negative coefficient, potentially because it is endogenously related to population and buoys the positive correlation with population. This is unlikely to be driven simply by high leverage points, since this is controlled for by the HC3 robust errors. The most likely explanation here is that some Russian areas have grown substantially during wartime, doing war-related manufacturing or developing close to the front lines as supply depots and bases, places which Kyrgyz migrants may be more averse to visiting or working.

Finally, average nominal gross wages in the same period are positively associated with Kyrgyz migrant inflows: the point elasticity for an average region is a 2.4–2.8% increase in migrants flowing into the region for every 1% increase in nominal wages, significant at the <0.1% level. This could be due to many explanations. First, areas with higher nominal wages will attract more Kyrgyz migrants, who according to economic theory will migrate based on higher expected wages than they can get otherwise in Kyrgyzstan. This explanation is the most probable, since the lag is not selected—only same-period wages are important. However, this could also be related to war-related labor scarcity which may drive up wages for these migrants. Additionally, there may be remaining leakages related to cost of living, even though the relevant variables (log_minwagevalue, poverty_rate) were not selected by the BIC.

6.3. Results for dependent variable of outflows of migrants (Table 3)

Observing the same regressions, but run on outflows of Kyrgyz migrants, the nowcast has more of an effect on interpretation. Sometimes the nowcast flips the sign or changes which variables are selected by the BIC. The differences could be due to either the small sample size for the original data or measurement error introduced by the nowcast. The nowcast is not perfect, but it does tend to preserve ordinality for many points when tested on the validation set and also is not trained on any economic covariates, just search query data. Thus these inferences contain validity, and they are reported together with regressions on the original data.

Table 3: Annual outflows of Kyrgyz migrants from Russia per region
Difference-in-difference of war and post-war effects on Kyrgyzstan-to-Russia migration (2021–23, logged outflows series from Rosstat, compared to nowcasted series[†]) with HC3 robust standard errors.
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

	A	**A[†]**	**B**	**B[†]**	**C**	**C[†]**
war	1.72648***	0.286289	1.75351***	*Not selected*	*Not selected*	-2.459537***
post_war	-0.46124.	0.150973	-0.86767***	*Not selected*	-0.912920***	0.512371**
Selected covariates by BIC[‡]	—		log_total_pop (0.90040***) log_rural_pop (0.12969*) log_avg_nominal_gross_wages (3.09617***)	log_city_pop (11.482873***) log_rural_pop (0.089898**) log_avg_nominal_gross_wages (2.047196***) lag_log_city_pop (-10.828279***)	**log_mediazona** **(0.332174***)** changeinindustrialproduction_electricity_utilities (-1.420790*) log_total_pop (8.681985*) log_rural_pop (0.108983*) log_avg_nominal_gross_wages (2.958660***) lag_changeinindustrialproduction_electricity_utilities (-1.205020) lag_log_total_pop (-8.019996*)	**log_mediazona** **(0.430048***)** **lag_log_mediazona** **(-0.285618***)** changeinindustrialproduction_electricity_utilities (-0.700545) log_total_pop (10.785620**) log_avg_nominal_gross_wages (2.146458***) lag_log_total_pop (-10.153473**)

Adj-R^2	0.08267	0.03767	0.5923	0.4787	0.6233	0.512

†*Extended the time series: nowcasted 2019–24 using XGBoost gradient boosting ensemble method.*
‡*Full tables with AIC-selected models and complete list of non-selected covariates available in the online appendix.*

For outflows, casualties are even more important, carrying a large and positive coefficient (0.33) significant at the <0.1% level for both. The story for the original data is a same-period increase in outflows if there are more casualties reported from that region that year—a push effect driven by fear of being drafted or simple nervousness about being in Russia during political turmoil, combined with a post-war coefficient of -0.91 also significant at the <0.1% level showing that the exodus actually tapers off as the war continues.

For the same regression on the nowcasted outflows with the longer time frame, Mediazona casualties has an even larger positive coefficient: a 1% increase in casualties is associated with a 0.43% increase in outflows in the same year, and a 1% increase in one-year lagged casualties also predicts a -0.29% decrease in outflows, both significant at the <0.1% level. This transition of a negative to a positive relationship with casualties reveals a more delayed push effect in the periods following the initial invasion. Since war presence is already controlled for, this effect is unlikely to be due simply to the significance of going from 0 to some positive number on the onset of the war, Here, the post-war coefficient predicts a 0.51% increase in migrant outflows but is less significant at the 1% level, and the war coefficient predicts a -2.46% decrease in migrant outflows, significant at the <0.1% level. This reflects that the post-war and war dummies were absorbing the Mediazona casualties association, and including the Mediazona variable provides a better fit. Together, the data suggest a mass exodus at the war's onset that tapers over time, with a delayed push effect tied to higher casualty rates in the migrant's home region.

The original data and nowcast regressions both demonstrate evidence for a push effect—delayed or not—due to the consistency of same-period casualties carrying a positive and significant coefficient in relation to outflows. Regions that have more people who fight and die on the front lines, even when controlling for population, seem to experience an increase in outflows of Kyrgyz migrants—due to dodging conscription or the push factor of more political turmoil and fear in those locations. This ends up showing up in the net flows regressions, too.

Nominal gross average wages are also positively associated with outflows in all regressions with coefficients between 2.04 and 3.10, robust to different covariate combinations and significant at the <0.1% level throughout. This likely is not because cities—which have higher wages—had more of an exodus of Russians and Kyrgyz people alike on the onset of the war, since city_pop or total_pop is also controlled for in all regressions. Though this could be spurious, this effect is likely driven by the fact that highly paying jobs or jobs in urban, high-wage-rate regions could have more turnover for these migrants, or might be places where migrants typically take seasonal construction jobs, while rural or agricultural regions have less turnover. Evidence in the remittances regression results corroborates this.

The adjusted R^2 values, between 0.47 and 0.63, are remarkable, as they express how good of a fit the model was able to achieve despite not including fixed effects and being limited to such a small covariate set. They also express that these models are parsimonious—the variables are selected because they are very important to explaining variation, and are less incidental. The covariate pool has 18 total variables, so each model does not select between 8 to 11 covariates, which is similar to the inflows regressions. Variables that were not selected include many population variables, real minimum wage, poverty rate, and more nuanced industrial production variables, such as mineral extraction and manufacturing.

6.4. Results for dependent variable of net flows of migrants (Table 4)

The net flows regressions were only completed on the original series, since none of the nowcasting models performed well on the validation set. Additionally, because of negative values being included in net flows, *the dependent variable cannot be logged*, exposing the regression to more of an influence of heteroskedasticity (and producing much larger coefficients). However, the HC3 robust standard errors help remove some of this risk.

Table 4: Annual net flows of Kyrgyz migrants to Russia per region
Difference-in-difference of war and post-war effects on Kyrgyzstan-to-Russia migration (2021–23, logged net flows series from Rosstat[†]) with HC3 robust standard errors.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	A	**B**	**C**
war	-781.741***	-639.410***	*Not selected*
post_war	156.654*	*Not selected*	*Not selected*
Selected covariates by BIC[‡]	—	log_total_pop (137.360*) log_avg_nominal_gross_wages (482.386**)	**log_mediazona (-111.634***)** log_total_pop (189.951***) log_avg_nominal_gross_wages (462.707**)
Adj-R^2	0.1386	0.2089	0.2071

[†]*The nowcasted series is not analyzed here, as all models were unable to achieve high-quality validation RMSE.*
[‡]*Full tables with AIC-selected models and complete list of non-selected covariates available at the online appendix.*

The basic model shows a -781.74 coefficient on war (significant at the <0.1% level) and a coefficient of 156.65 on post-war. In other words, this implies that the presence of war is associated with a decrease in -782 migrants (in levels) flowing into a particular region on net, but for each year the war continues, there is an associated increase in 157 migrants flowing into that region. This reflects the big picture from the inflows and outflows regressions: a mass exodus that is counteracted by inflows in the periods afterwards.

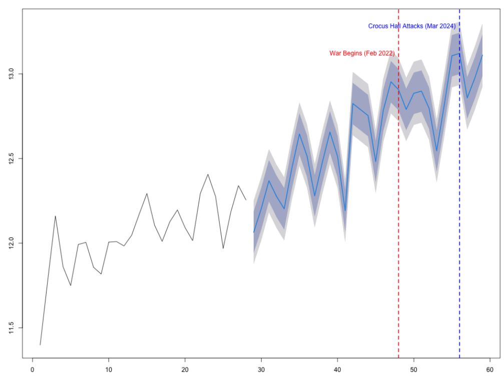
When adding the covariates, the BIC only selects total population and average nominal gross wages, both with positive coefficients significant at the 5% and 1% levels, respectively. This is unsurprising and reflects the pull factors of higher nominal wages and higher population leading to more jobs being available. However, once adding Mediazona casualties, the same variables are selected, and the coefficients become more significant, with the coefficient on population increasing slightly. A 1% increase in same-year Mediazona casualties from the same region predicts a decrease in -111.634 Kyrgyz migrants arriving in that region on net, significant at the <0.1% level. Both post-war and war are not selected at all, showing that the BIC evaluates casualties alone as best explaining net flows compared to the war dummies themselves. Additionally, the highly significant negative coefficient shows that there is very strong evidence for the push effect and exodus at war onset, and that this effect is stronger in regions with more confirmed casualties, robust to total population and nominal wage rate.

6.5. Results for dependent variable of quarterly Kyrgyz visitors

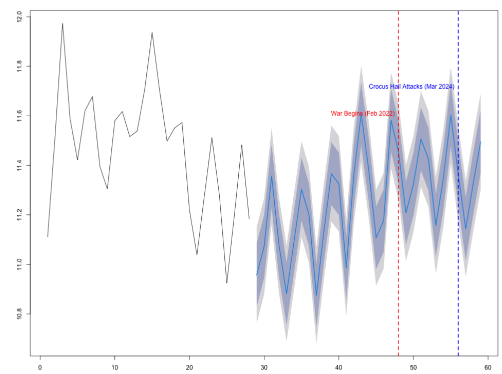
Nowcasted models were specified for four types of visitors: "all types," private travelers, business travelers, and transit travelers (Table 5). The plots (Figure 8) demonstrate changing behavior after the war onset and Crocus attacks, usually a decreasing log-linear positive trend that breaks down at these time cutoffs. Business travelers numbers decline and then rebound during COVID, which is consistent with expectations. The nowcasts were generated without implicit inclusion of war effects and trained on data from 2010–16, ending 8 years before the war began, so this decreasing behavior at the exact time of the shocks is quite remarkable. However, the BIC-selected models do not always rate these cutoffs as significant by themselves, often preferring casualties—which can be interpreted as either a measurement of war severity or a labor substitution indicator.

Figure 8: Nowcasted quarterly Kyrgyz visitors to Russia (national-level) by type, 2010–24.
Nowcast generated via SARIMAX model. Covariates including Google Trends data, socioeconomic covariates, and ARIMA order are selected based on Granger causality, non-multicollinearity (low relative VIF), and AIC. Logged dependent variable.

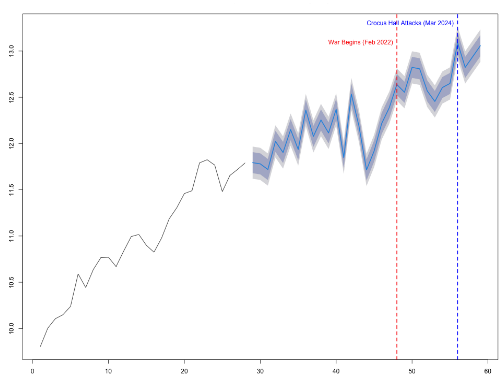
All types of visitors:



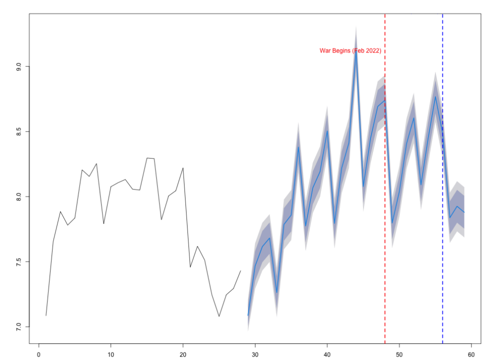
Private visitors:



Business visitors:



Transit travelers:



The private travelers nowcast model includes one Google Trends term—the logged Russian term for “[job] vacancies” which carries a negative coefficient, otherwise prioritizing lags of Kyrgyz and Russian investment, Russian consumption, and Kyrgyz GDP. These terms are broken down in the dictionary below (Table 6). The business travelers model included three Google Trends search queries: the Russian phrase “ticket to Bishkek” (negative), the Russian for “[job] vacancies” (positive), and the Kyrgyz-language term for “work” (negative). These are supplemented by lags of Russian government expenditure, consumption as a percentage of GDP in Russia, investment in Kyrgyzstan, and lagged household consumption in Russia. The model for transit travelers included two autoregressive terms (AR1 and AR2—the only one to select an AR or MA order), lagged Russian “[job] vacancies”, Kyrgyz "work," and Kyrgyz "find work," supplemented by the consumer confidence survey rating in Russia, two lags of debt service ratios for the private non-financial sector in Russia, investment percentage of GDP in Russia, and lagged investment percentage of GDP in Kyrgyzstan. It is curious to see which variables were selected for the forecast by the AIC, and there is no promise that these reflect the "true model" whatsoever, but they are variables that have strong evidence for Granger causality (ability to predict a target variable via distributed lags).

Table 5: Models for extending the time series of quarterly Kyrgyz visitors to Russia
Nowcast generated via SARIMAX model. Covariates, including AR and MA order, selected based on Granger causality, non-multicollinearity (low relative VIF), and AIC. All models include a seasonal (monthly dummy) component not displayed. Google Trends coefficients are bolded. N=59 for all variables (N=28 in the original data, with 31 additional quarters nowcasted).

	All	Private	Business	Transit
Selected model	intercept (13.1853) lag2_log_GDP_billion_currency_units_KGZ (-0.1486) Consumption_as_percent_of_GDP_RUS (-0.0564) lag3_log_GDP_billion_currency_units_KGZ (0.3901) lag1_Household_debt_to_GDP_in_percent_RUS (0.0570)	intercept (17.2092) lag4_log_Investment_billion_currency_units_KGZ (0.4734) lag2_Investment_as_percent_of_GDP_RUS (-0.0192) lag1_log_GDP_billion_currency_units_KGZ (-0.4207) **log_vakansii (-1.0563)** lag4_Investment_as_percent_of_GDP_RUS (-0.0286) lag2_Consumption_as_percent_of_GDP_RUS(0.0223)	intercept (-11.9284) Consumption_as_percent_of_GDP_RUS (-0.0825) lag3_log_Government_expenditure_billion_currency_units_RUS (2.1392) **log_bilet_bishkek (-0.0605)** **log_vakansii (1.0751)** **log_jumush (-0.0451)** lag2_log_Government_expenditure_billion_currency_units_RUS (-2.883) lag4_Investment_as_percent_of_GDP_KGZ (0.0086) lag2_log_Household_consumption_billion_currency_units_RUS (3.0464)	ar1 (-0.9984) ar2 (-0.5437) **lag_log_vakansii (1.0139)** log_Consumer_confidence_survey_RUS (0.2400) lag3_Debt_service_ratios_for_private_non_financial_sector_RUS (-0.0746) Investment_as_percent_of_GDP_RUS (-0.0401) lag4_Investment_as_percent_of_GDP_KGZ (0.0282) lag4_log_GDP_billion_currency_units_KGZ (0.9243) lag1_Debt_service_ratios_for_private_non_financial_sector_RUS (-0.0733) **log_jumush (-0.0911)** **log_jumush_izdeim (-0.1174)**

Table 6: Dictionary of search queries used for nowcasting in the study
(Logged) total counts, ratio between terms, difference between terms, and lags of each

English	Russian	Kyrgyz
salary	zarplata зарплата	ailyk айлык
(airplane) ticket to Bishkek	—	(avia)bilet Bishkek* (авиа)билет Бишкек
job search	poisk raboty поиск работы	jumush izdeim жумуш издейм
work	rabota работа	jumush жумуш
transfer money (to Bishkek)	perevod deneg перевод денег	perevod deneg v Bishkek* перевод денег в Бишкек
Unistream (Kyrgyzstan) Zolotaya Korona (Kyrgyzstan)	unistream Юнистрим zolotaya korona Золотая Корона	unistream Kyrgyzstan* Юнистрим Кыргызстан zolotaya korona Kyrgyzstan* Золотая Корона Кыргызстан
job vacancies	vakancii вакансии	—

**These pairings are not in the Kyrgyz language but operate as an indicator for Kyrgyz individuals in Russia relative to the general Russian search terms.*

Next, the war effects regressions (Table 7.1) were selected via the BIC (which is designed to reflect the closest specification to the “true model”) and evaluated alongside HC1 robust standard errors. Fascinatingly, the BIC-selected regressions on the nowcasted time series included the casualties variable for private, business, and transit travelers, but did not select any other war or Crocus effects. The nowcast may simply not be exact enough to determine these effects, but the Mediazona casualties effect must be obvious enough that it was selected in all three types—very unlikely if this were due to random chance.

Table 7: War effects on nowcasted, national-level quarterly Kyrgyz visitors to Russia

(7.1) BIC-selected regressions

Nowcast generated via SARIMAX model, incorporating Google Trends and economic covariates.
With HC3 robust standard errors
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

	All	Private	Business	Transit
covid	*Not selected*	0.1817**	-0.3421	0.5061**
war	*Not selected*			
post_war				
crocus				
post-crocus				
log_mediazona	*Not selected*	0.0422***	-0.0330*	*Not selected*
log_mediazona_lag1	*Not selected*	*Not selected*	*Not selected*	*Not selected*
log_mediazona_lag2	*Not selected*	*Not selected*	*Not selected*	0.0491**
time trend	0.0210***	-0.0107***	0.0525***	*Not selected*

(7.2) AIC-selected regressions

Nowcast generated via SARIMAX model, incorporating Google Trends and economic covariates.
With HC3 robust standard errors
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

	All	Private	Business	Transit
covid	*Not selected*	0.1817**	-0.3421	0.4123*
war	0.2736***	*Not selected*	*Not selected*	*Not selected*
post_war	0.0527***	*Not selected*	*Not selected*	*Not selected*
crocus	-0.2343***	*Not selected*	*Not selected*	-0.5759***
post-crocus	*Not selected*	*Not selected*	*Not selected*	*Not selected*
log_mediazona	*Not selected*	0.0422***	-0.0330*	*Not selected*
log_mediazona_lag1	-0.0348***	*Not selected*	*Not selected*	*Not selected*
log_mediazona_lag2	-0.0289***	*Not selected*	*Not selected*	0.0457*
time trend	0.0200***	-0.0107***	0.0524***	0.0051

The sign on the casualties coefficient is negative for just one of the regressions—business travelers (with a tiny coefficient of -.0330, significant at the 5% level, implying that a doubling or 100% increase in casualties was associated with a -3.3% decrease in business travelers)—and positive for private travelers (same period, 0.0422, significant at the <0.1% level, implying that a doubling or 100% increase in casualties was associated with a 4.2% increase in private travelers) and transit travelers (with a lag of 2 quarters, 0.0491, significant at the 1% level, implying that a doubling or 100% increase in casualties from 6 months previous was associated with a 4.9% increase in transit travelers).

This is a huge piece of evidence in support of the labor substitution effect, because the class differences between these types of travelers is consistent with what is expected for the substitution effect. Business travelers has a negative coefficient, reflecting that severity of war would likely have chilled inflows of white collar business visitors, and since conscripts are less likely to be white collar workers, business visitors should not be as susceptible to labor substitution effects for conscripted Russians. Meanwhile, private and transit travelers would include new inflows of blue collar Kyrgyz migrant laborers (with transit travelers including illegal migrants (Савина & Бонч-Осмоловская, 2022)), thus the labor substitution effect should be encoded in a positive coefficient on casualties.

It is possible that the BIC is being overly restrictive due to the tiny sample size, and so the less-restricted AIC-selected models should also be presented to corroborate these patterns (Table 7.2). For private travelers and business travelers, AIC picks exactly the same model. And for transit travelers, Crocus carries a new, large, negative coefficient, while mediazona_lag2 maintains almost exactly the same positive coefficient, like in the BIC—showing that the Crocus effect was crowded out due to the BIC restriction, but is significant and negative.

The nowcasting model for “all types” of Kyrgyz visitors was also estimated, but is by far the least sophisticated, only including two lags of Kyrgyz GDP, same-period consumption share of GDP in

Russia, and the lagged Russian household debt to GDP ratio. This is likely because the trend in the model was very linear, so including these covariates may have allowed for spurious correlations with the GDP variables to allow for overfit, especially as Kyrgyz GDP has unexpectedly increased with the onset of war, but likely not causing an increase in migrants to Russia accordingly.

The "all types" nowcast does not seem to be very trustworthy, with the BIC-selected regression including only a time trend and monthly seasonality. The AIC estimates are more interesting: a positive estimate on war and post_war, and a large negative effect of Crocus, with negative coefficients on both the first and second lag on casualties, all of which are significant at the <0.1% level. Although it makes sense that casualties, correlated with severity of war and conscription waves, would cause less visitors to go to Russia and that Crocus would lead to a huge decline in visitors, trusting this model would imply that Kyrgyz visitors as a whole—including vacationers and laborers alike—are *much* more likely to visit Russia after the start of the war. This does not seem to be true by the 2021–23 inflows data analyzed previously and many news reports (Agence France-Presse, 2023). The model, plot, and estimates are reported for the sake of comparison and transparency, but without much stake placed in these results. This phenomenon makes it clear that the disaggregation of types is important to discern migration dynamics and the labor substitution effect identified above.

The COVID control ends up being important, showing a decrease in business visitors, ceteris paribus, during the year where COVID was most relevant (2020 Q1-4), consistent with news reports (Kopytin, 2021). However, during COVID, transit and private travelers are shown to increase, robust to seasonal adjustment and time trend. This may be because the small sample size requires a SARIMAX model rather than something more sophisticated, and thus this model relies on the assumption of an underlying linear model which may simply not be correct. All economic indicators and search queries may behave differently during the onset of COVID compared to how they behaved during the period of training data, 2010–16. This is a limitation of the data and must be kept in mind. Thus, it is necessary to include the COVID dummy in the model and it is frequently selected as an important predictor, providing a control for out-of-the-ordinary behavior and increasing the precision of inference. As previously mentioned, all ADF tests passed, thus the use of this model is appropriate.

In terms of the validity of the models as a whole, it is pretty safe to assume that the behavior of many indicators should be approximately similar to in 2010–16. Exceptions include Russian data unreliability biasing up Russian GDP metrics, but Russian GDP is not included directly in any model. Additionally, war-related investment by Russia receives negative coefficients in the private and transit traveler prediction models, and therefore these estimates may be artificially depressed, but the tiny magnitude of these coefficients (less than 0.1 in absolute value, with a logged dependent variable) mean that the effect should be fairly negligible.

6.6. Results for dependent variable of remittances

Regressions on monthly remittances (Table 8) reliably estimated a negative coefficient on 8-month lag on casualties and a positive coefficient for 2-month lag on casualties, both significant at or far below the 5% level and robust to Russian CPI. This could be due to two primary explanations: either higher wages as a result of the tighter Russian labor market due to so many conscripted soldiers, or more migrants being hired and going to Russia to fill their place. Four nominal wage variables (Sakha, Sakhalin, Moscow Oblast, and overall across Russia) were added to the covariate pool to examine whether this effect was independent of the wage level. The casualty coefficients barely changed, actually increasing, showing that the increase in remittances associated with the 2-month lag is likely not due to increased wages, but rather due to increased migrants (with a 2-month lag between migration and sending money back home to Kyrgyzstan). These coefficients are also almost exactly the same when Russian and Kyrgyz CPI are added to the covariate pool. When adding the exchange rate as an additional covariate, the coefficients remain around the same magnitude and the same lags are still selected by the BIC, but only the negative 8-month lag is significant at the 10% level. There is strong evidence that the 2-month and 8-month lag are influential in explaining logged remittances, given the consistency of their coefficient estimates robust to many coefficients, plus no other lags are selected. For logged remittances not controlled by CPI but only the nominal wage levels, the BIC selected the war dummy, which carried a large and negative coefficient (-0.4346), though this was not significant.

Table 8: Regressions of logged monthly remittances from Russia to Kyrgyzstan
With HC3 robust standard errors; Jan 2011-Sep 2024 (N=165)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Basic model†	CPI-controlled model‡	Wage-controlled model§	CPI‡ and wages§	CPI‡, wages§, and exch. rate‖
Monthly dummies	*Selected (~4.75***)*	*Selected (~16.5***)*	*Not selected*	*Selected (~28.5*)*	*Selected (~5.5*)*
Time trend	0.0040***	0.0143***	*Not selected*	0.0176*	*Not selected*
COVID	-0.1355	-0.2220*	*Not selected*	*Not selected*	*Not selected*
War	*Not selected*	*Not selected*	-0.4346	*Not selected*	*Not selected*
Post_war	*Not selected*	*Not selected*	*Not selected*	*Not selected*	*Not selected*
Crocus	*Not selected*	*Not selected*	*Not selected*	*Not selected*	*Not selected*
Post_crocus	*Not selected*	*Not selected*	*Not selected*	*Not selected*	*Not selected*
Mediazona (casualties)	Lag8: -0.0172*	Lag2: 0.0281** Lag8: -0.0391***	Lag2: 0.0956* Lag8: -0.0886***	Lag2: 0.0739** Lag8: -0.0541*	Lag2: 0.0407 Lag8: -0.0505.
CPI‡	*N/A*	RUS: -0.1992***	*N/A*	KGZ: -2.1477**	RUS: -2.8431
Wages§	*N/A*	*N/A*	Mos. Obl.: 2.9932*** Sakhalin: -1.7725*** Sakha: -0.6504*	Moscow Obl: 3.0667. Sakhalin: -2.4311* Sakha: -1.6357	Moscow Obl: 3.6213* Sakhalin: -1.4441.
Exch. Rate‖	*N/A*	*N/A*	*N/A*	*N/A*	-1.0951***

†*Basic BIC-selected regression of logged monthly remittances on the following variables in the covariate pool: Month + time_trend + Ramadan_starts + Eid_al_Fitr + Eid_al_Adha + covid + war + post_war + log_mediazona + log_mediazona_lag2 + log_mediazona_lag4 + log_mediazona_lag6 + log_mediazona_lag8.*
‡*Added to the covariate pool: `log_Consumer_Price_Index_(CPI)_RUS` + `log_Consumer_Price_Index_(CPI)_KGZ`*
§*Added to the covariate pool: nominal wages for all of Russia, Moscow Oblast, Sakhalin, and Sakha (Yakutia) to represent a variety of typical destinations for Kyrgyz migrants*
‖*Added to the covariate pool: bilateral real exchange rate between the Russian ruble and the Kyrgyz som*

Remittances are less likely to be sent when inflation is higher on either the Russian or Kyrgyz side, but especially the Kyrgyz side. In the wage- and CPI-controlled regressions, a 1% increase in the Kyrgyz CPI is associated with a -2.14% decrease in logged remittances, significant at the 1% level. This may reflect that migrants have historically returned to Kyrgyzstan during times of global economic crisis (including in Kyrgyzstan), culminating in less remittances being sent back. For the Russian side, a 1% increase in the Russian CPI is associated with a -0.2% decrease in logged remittances, significant at the <0.1% level. This finding is a reflection of remittances being worth less and needed more by the migrants themselves when inflation is high.

Lastly, the fact that the wage rates for the popular destinations of Moscow Oblast, Sakhalin, and Sakha are selected and not the "all of Russia" wage rate reflects that places like these regions are more important to explaining variation in Kyrgyz remittances than the aggregate. In particular, the large coefficient on Moscow—a 1% increase in Moscow Oblast wages associated with a 3.0% increase in remittances—reflects that a huge number of Kyrgyz migrants work in Moscow and send money back. The fact that the Sakhalin and Sakha coefficients are significant but negative—a 1% increase in wages associated with a -1.6% and -2.4% decline in remittances respectively—might reflect that Kyrgyz migrants may choose to live in these rural regions due to their lower cost of living, and a rising wage may expose a rising cost of living that pushes migrants to different destinations instead. For this analysis, the small sample size prevented adding more wage variables to this regression. Thus, the wage coefficients must stand mostly as an urban or rural wage rate proxy. This is corroborated by the final regression including exchange rates, where a 1% increase in nominal wages for Moscow is associated with a 3.6% increase in remittances (significant at the 5% level), while a 1% increase in Sakhalin nominal wages is associated with a -1.44% decrease in remittances (significant at the 10% level), with

Sakha not selected due to redundancy with Sakhalin—providing evidence for this rural-urban dichotomy being a very influential component regarding wages.

7. Conclusion

From regressions on migration flows, this study uncovered a push effect associated with severity of war. Casualties have a significant, positive association with outflows. Exact push factors might include political turmoil, dodging conscription, psychological effect of more casualties being confirmed from that region. The study also found a surprising labor substitution effect of migrants for Russian conscripts: Casualties have a positive association with inflows, providing evidence that the gap in the market left by conscripts may be filled by migrant workers, though the regressions on net flows show that the push effect seems to dominate the labor substitution effect on average across regions.

From the regressions on quarterly Kyrgyz visitors, evidence of class differences in the labor substitution effect emerged. Casualties are positively associated with private and transit travelers but negatively with business travelers, consistent with the fact that the labor substitution effect is less relevant for white collar or formal workers captured by the business travelers variable, since conscripts typically leave behind blue collar jobs.

From the regressions on remittances, the labor substitution seems to affect remittance payments with a 2-month delay. Casualties from 2 months ago are positively associated with remittances, even when controlling for exchange rate, inflation, and wages. The regressions also uncover a 6-month incubation time for the labor substitution effect: casualties from 8 months ago are negatively associated with remittances, finding that the gap between the two variables is 6 months, so the labor substitution effect might take 6 months to affect migrant inflow patterns.

Lastly, the inflows regression estimated the elasticity of the migrant supply curve, finding that across regions, a 1% increase in nominal wages was associated with a 2% increase in Kyrgyz migrant inflows. The wage elasticity of supply of remittances was quantified according to fluctuations in nominal wages in Moscow, Sakha, and Sakhalin. For Moscow, a 1% increase in nominal wages was associated with a 3% or higher increase in remittances. For Sakhalin and Sakha, this wage elasticity is considerably lower or even negative, reflecting that key drivers for migration in these regions may be that employers in these rural areas can hire more cheap labor from migrants when there is a lower market wage rate. Please see a summary of key takeaways in the appendix (Table 9).

This research has significant implications both in terms of the methodology and findings. First, Google Trends largely failed as a predictor in the Russia setting, proving only to be useful for nowcasting in conjunction with economic covariates. Second, Yandex Wordstat is a far superior predictor. If the models above were estimated with monthly data rather than being constrained by the yearly aggregation, it could be a powerful predictor in even more sophisticated models. Third, gradient boosting trees outperformed other machine learning techniques, showing that migration flows do not seem to be able to be adequately described by linear models, requiring more sophisticated modeling than simple linear regressions or LASSO, and also benefiting from the huge number of predictors provided to the model. The quarterly SARIMAX models required a huge amount of complex covariate pre-processing in order to obtain stable estimates, aligning with the fact that the linearity assumption seems to be incorrect, even as stationarity holds according to the ADF tests.

In terms of findings, the push effect associated with war severity was consistent with the economic theory, but few studies focusing on any area of the world have located a similar macro-level conscript-substitution effect as what was identified in this study. Implications include that the Kyrgyz government should understand that remittance increases are likely due to more citizens going abroad rather than receiving higher wages, regardless of what official Russian numbers say. Given that remittances seem to have a dual-edge impact on Kyrgyzstan, and that so many of their citizens experience suboptimal work conditions or are conscripted while abroad, it is in Kyrgyzstan's best interest to both reduce reliance on remittances and harness them more effectively for investing in pro-development industry within its own borders. Finally, the government of Kyrgyzstan must understand that their country may not always be so lucky to be supported by financial flows from abroad, especially if Russian inflation worsens or the war's severity increases. Until Kyrgyzstan is able to invest in its people enough to create jobs and raise the standard of living for all individuals across the country, and not just in Bishkek, people will continue to risk their lives and travel thousands of miles to work in a war-afflicted country.

Abdukadyrova, L., & Studenko, O. (2023). 'the pandemic played a cruel joke on us': The vulnerabilities of Kyrgyz women migrant workers in Russia during COVID-19. *Anti-Trafficking Review*, (21), 36–51. https://doi.org/10.14197/atr.201223213

Advantour. (n.d.). *Kurman Ait, or Feast of Sacrifice in Kyrgyzstan*. Retrieved April 11, 2025, from https://www.advantour.com/kyrgyzstan/holidays/kurman-ait.htm

Agadjanian, V., Menjívar, C., & Zotova, N. (2017). Legality, Racialization, and Immigrants' Experience of Ethnoracial Harassment in Russia. *Social Problems*, *64*(4), 558–576. https://doi.org/10.1093/socpro/spw042

Agence France-Presse. (2023, November 18). *Under pressure, Central Asia migrants leaving Russia over Ukraine war*. TODAYonline. Retrieved April 11, 2025, from https://www.todayonline.com/world/under-pressure-central-asia-migrants-leaving-russia-over-ukraine-war-2307281

Aitymbetov, S. (2006). Working Paper: Emigrant Remittances: Impact on Economic Development of Kyrgyzstan. *The Economic Policy Institute*, *31*.

Anastasiadou, A., Volgin, A., & Leasure, D. R. (2023). *War and Migration: Quantifying the Russian exodus through Yandex search trends*.

Anderson, K. H., & Kroeger, A. (2011). *Remittances and children's capabilities: New evidence from Kyrgyzstan, 2005-2008* [SSRN Working Paper No. 1950540]. Social Science Research Network. https://doi.org/10.2139/ssrn.1950540

Aoga, J., Bae, J., Veljanoska, S., Nijssen, S., & Schaus, P. (2020). Impact of weather factors on migration intention using machine learning algorithms. *arXiv*. https://doi.org/10.48550/arXiv.2012.02794

Avramescu, O., & Wiśniowski, A. (2021). *Forecasting immigration using online search data*. Journal of Official Statistics, 37(4), 909–932. https://doi.org/10.2478/jos-2021-0041

Babadjanov, K. (2022, March 20). *Influx Of Russians, Return Of Labor Migrants Jacking Up Real Estate In Uzbekistan*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/uzbekistan-russians-ukraine-war-skyrocketing-prices/31761937.html

Becker, C. (2007). *Urbanization and rural-urban migration*. In G. M. Meier & J. E. Rauch (Eds.), *International Handbook of Development Economics* (pp. 1–18). Edward Elgar Publishing. Retrieved April 11, 2025, from https://www.academia.edu/68100293/Urbanization_and_rural_urban_migration

Bertoli, S., Fernández-Huertas Moraga, J., & Keita, S. (2016). The elasticity of the migrant labor supply: Evidence from temporary Filipino migrants (IZA Discussion Paper No. 10031). Institute of Labor Economics (IZA). Retrieved April 11, 2025, from https://www.iza.org/de/publications/dp/10031/the-elasticity-of-the-migrant-labor-supply-evidence-from-temporary-filipino-migrants

Bilecen, B., & Lubbers, M. J. (2021). The networked character of migration and transnationalism. *Global Networks, 21*(4), 837–852. https://doi.org/10.1111/glob.12317

Bosch, M., & Farré, L. (2013). Immigration and the informal labor market (IZA Discussion Paper No. 7843). Institute of Labor Economics (IZA). Retrieved April 11, 2025, from https://www.iza.org/en/publications/dp/7843/immigration-and-the-informal-labor-market

Boss, K., Krueger, F., Zheng, C., Heidland, T., & Groeger, A. (2023). Forecasting bilateral refugee flows with high-dimensional data and machine learning techniques. *Barcelona School of Economics Working Paper Series*, (1387). Retrieved April 11, 2025, from https://bse.eu/research/working-papers/forecasting-bilateral-refugee-flows-high-dimensional-data-and-machine

Böhme, M. H., Gröger, A., & Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 142, 102347. https://doi.org/10.1016/j.jdeveco.2019.102347

BBC Visual Journalism Team (2025, March 24). *Ukraine in maps: Tracking the war with Russia.* https://www.bbc.com/news/world-europe-60506682

Cabar.Asia. (n.d.). *Как мигрантов из Центральной Азии в россии пытаются отправить на войну*. Central Asian Bureau for Analytical Reporting. https://longreads.cabar.asia/mobilization_ru

Carammia, M., Iacus, S. M., & Wilkin, T. (2022). Forecasting asylum-related migration flows with machine learning and data at scale. *Scientific Reports, 12*, 1457. https://doi.org/10.1038/s41598-022-05241-8
Chamberlain, L. (2022, January 6). *Islamic holidays, 2010–2030 (A.H. 1431–1452)*. Infoplease. Retrieved April 11, 2025, from https://www.infoplease.com/calendars/holidays/islamic-holidays
Chen, T., & Guestrin, C. (n.d.). *XGBoost Python Package*. XGBoost Documentation. Retrieved April 11, 2025, from https://xgboost.readthedocs.io/en/stable/python/
Critelli, F. M., Lewis, L. A., Yalim, A. C., & Ibraeva, J. (2021). Labor Migration and Its Impact on Families in Kyrgyzstan: A Qualitative Study. *Journal of International Migration and Integration*, *22*(3), 907–928. https://doi.org/10.1007/s12134-020-00781-2
Dogan, Z. (2020, October 18). *Почему так много киргизов в Москве? Чем заняты и как живут на чужбине.* https://dzen.ru/a/X4xIflKE4zblP0Wg.
Dzhooshbekova, A., Chynykeeva, G., Abduvapova, A., Turdubaev, K., Elchibekov, U., Arstanov, S., Satybaldieva, C., Osmonova, S., Sulaimanov, E., Abduvapova, A., Osmonova, B., & Abdullaeva, Z. (2021). External Migration Problems of Kyrgyzstan population in the Post-soviet period. *Advances in Applied Sociology*, *11*(03), 113–129. https://doi.org/10.4236/aasoci.2021.113009
Ethnologue. (n.d.). *Kyrgyz language (KIR) – L1 & L2 speakers, status, map*. https://www.ethnologue.com/language/kir/
Eurasianet. (2024, May 3). *Relokanty: Russia's loss, Central Asia's gain*. Eurasianet. Retrieved April 4, 2025, from https://eurasianet.org/relokanty-is-russias-loss-central-asias-gain
Galstyan, M., Galstyan, N., & Hakobyan, G. (2024). From roots to routes: The role of ethnic networks in the development of migrant communities in Russia. *ResearchGate*. https://doi.org/10.13140/RG.2.2.27508.01928
Gawryluk, A., Komor, A., Kulisz, M., Zarębski, P., & Katarzyński, D. (2024). Artificial neural networks as a method for forecasting migration balance (A case study of the city of Lublin in Poland). *Sustainability, 16*(24), 11249. https://doi.org/10.3390/su162411249
Goble, P. (2024, March 31). Shortcomings of last Russian census contributing to policy mistakes now – OpEd. *Eurasia Review*. Retrieved April 11, 2025, from https://www.eurasiareview.com/31032024-shortcomings-of-last-russian-census-contributing-to-policy-mistakes-now-oped/
Goble, P. (2024, November 22). Central Asian countries sending migrant workers to Asia over Russia – Analysis. *Eurasia Review*. Retrieved April 11, 2025, from https://www.eurasiareview.com/22112024-central-asian-countries-sending-migrant-workers-to-asia-over-russia-analysis/
Golenvaux, N., Gonzalez Alvarez, P., Kiossou, H. S., & Schaus, P. (2020). *An LSTM approach to predict migration based on Google Trends*. arXiv. https://doi.org/10.48550/arXiv.2005.09902
Gozzi, L. (2024, March 27). *Moscow attack: Central Asian migrants hit by backlash in Russia.* https://www.bbc.com/news/world-europe-68665896
Hayes, A. F., & Cai, L. (2007). "Using Heteroskedasticity-Consistent Standard Error Estimators in OLS Regression: An Introduction and Software Implementation." Behavior Research Methods.
Iakushkina, O. V. (2021). Разнообразие «других»: образ московского мигранта на страницах результатов поиска Google, Yandex и DuckDuckGo. *Voprosy Kognitivnoy Lingvistiki*, *2*(046), 114–122. https://cyberleninka.ru/article/n/raznoobrazie-drugih-obraz-moskovskogo-migranta-na-stranitserezultatov-poiska-google-yandeks-i-duckduckgoo
Iashchenko, I. (2023, July 4). *Understanding Russian disinformation strategies inside and outside the country*. Aspenia Online. https://aspeniaonline.it/understanding-russian-disinformation-strategies-inside-and-outside-the-country/
Ibragimova, G. (2024, August 1). *Why are Central Asian migrant workers spurning Russia?*. Carnegie Endowment for International Peace. https://carnegieendowment.org/russia-eurasia/politika/2024/07/russia-central-asia-migrants?lang=en
Interfax. (2023, May 3). *Newsroom*. Over 1.5 mln Kyrgyz citizens work abroad. https://interfax.com/newsroom/top-stories/90223/
Irnazarov, F. (2015). Labor Migrant Households in Uzbekistan: Remittances as a Challenge or Blessing? *The Central Asia Fellowship Papers*, *11*.
Jayaprakash, R. S. (2024, April 28). *Steady, but evolving: An overview of Russia's migrant labour market*. Observer Research Foundation. https://www.orfonline.org/research/steady-but-evolving-an-overview-of-russia-s-migrant-labour-market
Karymshakov, K., Abdieva, R., Sulaimanova, B., & Sultakeev, K. (2016). Remittances Impact on Youth Labour Supply: Evidence from Kyrgyzstan. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3167324
Kopytin, Y. (2021, April 22). Kyrgyzstan – Russia flights: What changed in air traffic and what to expect? *24.kg News Agency*. Retrieved April 11, 2025, from https://24.kg/english/191198_Kyrgyzstan__Russia_flights_What_changed_in_air_traffic_and_what_to_expect/
Korsunskaya, D., Stolyarov, G., & Marrow, A. (2024, November 28). *Russia's economy struggles to find workers as Defence Sector Poaches Staff | Reuters*. Reuters. https://www.reuters.com/markets/europe/russias-labour-shortage-spreads-defence-sector-poaches-staff-2024-11-28/
Kumar, R. R., Stauvermann, P. J., Patel, A., & Prasad, S. (2018). The Effect of Remittances on Economic Growth in Kyrgyzstan and Macedonia: Accounting for Financial Development. *International Migration*, *56*(1), 95–126. https://doi.org/10.1111/imig.12372
Kupets, O. (2012). The Development and the Side Effects of Remittances in the CIS Countries: The Case of Ukraine. *CARIM-East Research Report*.
Lee, S. (2025, March 13). *A comprehensive guide to interpreting the Breusch-Pagan test*. Number Analytics. Retrieved April 11, 2025, from https://www.numberanalytics.com/blog/a-comprehensive-guide-to-interpreting-breusch-pagan-test
Long, J. S., & Ervin, L. H. (2000). "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model." The American Statistician.
Lyrchikova, A., & Papachristou, L. (2025, March 7). *Russian governor backs down on migrant labour ban after corporate outcry | Reuters*. Reuters. https://www.reuters.com/markets/europe/russian-governor-backs-down-migrant-labour-ban-after-corporate-outcry-2025-03-07/
MacKinnon, J. G., & White, H. (1985). "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." Journal of Econometrics.
Massey, D. S., Kalter, F., & Pren, K. A. (2008). Structural economic change and international migration from Mexico and Poland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 60*(48), 134–161. https://doi.org/10.1007/s11577-008-0018-3
Mediazona. (2025, March 28). *Russian losses in the war with Ukraine. Mediazona count, updated.* https://en.zona.media/article/2025/03/28/casualties_eng-trl
Meduza. (2022, October 5). A new report shows discrepancies in Russian draft statistics. *Meduza*. Retrieved April 11, 2025, from https://meduza.io/en/feature/2022/10/05/a-new-report-shows-discrepancies-in-russian-draft-statistics
Mogilevskii, R., Thurlow, J., & Yeh, A. (2018). Kyrgyzstan's Accession to the Eurasian Economic Union: Measuring Economy-Wide Impacts and Uncertainties. *SSRN Electronic Journal*, *Working Paper #44*. https://doi.org/10.2139/ssrn.3807831
Murzakulova, A., Kuznetsova, I., Mogilevskii, R., Round, J., Abdoubaetova, A., & Wolters, A. (2020). Migration and COVID-19: Challenges and Policy Responses in Kyrgyzstan. *Central Asia Program*, *247*.
Najibullah, F., & Navruzshoh, Z. (2023, April 18). *Russian Recruiters Target Central Asian Migrants At Mosques, Dorms To Join War In Ukraine*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/russia-recruits-central-asia-migrants-ukraine-war/32369341.html
The New Humanitarian. (2015, April 24). *Hope and fear: Kyrgyz migrants in Russia.* https://www.thenewhumanitarian.org/feature/2015/04/24/hope-and-fear-kyrgyz-migrants-russia

Ng, S. (2021, July 19). *Boosting high dimensional predictive regressions with time varying parameters* [Software and data]. Academic Commons. https://doi.org/10.7916/d8-pdmd-g926

NVIDIA. (n.d.). *XGBoost – What Is It and Why Does It Matter?* NVIDIA. Retrieved April 11, 2025, from https://www.nvidia.com/en-us/glossary/xgboost/

Orlova, M. (2022, October 12). *More than 800,000 Kyrgyzstanis work outside country*. https://24.kg/english/247811_More_than_800000_Kyrgyzstanis_work_outside_country/

Orphanides, K. G. (2019, August 4). VyprVPN review: A VPN with streaming prowess and security. *WIRED*. Retrieved April 11, 2025, from https://www.wired.com/story/vyprvpn-review/

Oshchepkov, A., Tilekeyev, K., & Gerry, C. (2024). *How War in Ukraine Has Shaped Migration Flows in Central Asia*. University of Central Asia.

Osmonalieva, B. (2024, November 29). *Head of Yakutia arrives in Kyrgyzstan*. https://24.kg/english/312592__Head_of_Yakutia_arrives_in_Kyrgyzstan/

Ozat, M. (2023, April 10). *Central Asian Labor Migration: Exploring New Destinations Amid Geopolitical Tensions*. Caspian Policy Center. https://www.caspianpolicy.org/research/energy-and-economy-program-eep/central-asian-labor-migration-exploring-new-destinations-amid-geopolitical-tensions

Palacin, J., & Shelburne, R. (2007). Remittances in the CIS: Their Economic Implications and a New Estimation Procedure. *United Nations Economic Commission for Europe*.

Plastun, O., & Vorontsova, A. (2024, December 25). *Illusion of stability: Never trust Russian stats*. Vox Ukraine. https://voxukraine.org/en/illusion-of-stability-never-trust-russian-stats

Radio Free Europe. (2013, August 13). *A new home in Siberia for Kyrgyz migrants*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/kyrgyz-migrants-siberia/25074085.html

Radio Free Europe. (2019, March 19). *Businesses Shuttered, Migrants Stay Home Amid Tensions Over Alleged Rape In Russian Region*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/businesses-shuttered-kyrgyz-stay-home-amid-tensions-over-alleged-rape-in-russian-region/29829961.html.

Radio Free Europe. (2019, March 28). *Russian Region Bans Employment Of Foreigners In Dozens Of Sectors*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/businesses-shuttered-kyrgyz-stay-home-amid-tensions-over-alleged-rape-in-russian-region/29829961.html.

Rickleton, C. (2024, December 18). *Fear And Economic Uncertainty Weaken Russia's Pull For Central Asian Migrants*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/central-asian-migrants-face-discrimination-russia-economic-shift-/33243452.html

Rickleton, C. (2025, January 4). *Amid Russia Pain, Central Asian Migrants Find New Destinations, Fresh Problems*. RadioFreeEurope/RadioLiberty. https://www.rferl.org/a/russia-migrants-centralasia-uzbekistan/33261194.html

Romanenko, V., & Borodkina, O. (2019). Female immigration in Russia: Social Risks and Prevention. *Human Affairs*, *29*(2), 174–187. https://doi.org/10.1515/humaff-2019-0014

Ruziev, K., & Majidov, T. (2013). Differing Effects of the Global Financial Crisis on the Central Asian Countries: Kazakhstan, the Kyrgyz Republic and Uzbekistan. *Europe-Asia Studies*, *65*(4), 682–716. https://doi.org/10.1080/09668136.2013.766044

Ryo, E. (2013). Deciding to cross: Norms and economics of unauthorized migration. *American Sociological Review, 78*(4), 574–603. https://doi.org/10.1177/0003122413487904

Scott, C. K., Sexsmith, K., & Chi, G. (2024). Gendered Migration Patterns and Social Change in Rural Kyrgyzstan. *Migration and Development*, 21632324241285086. https://doi.org/10.1177/21632324241285086

scikit-learn Developers. (n.d.). *sklearn.preprocessing.StandardScaler*. scikit-learn. Retrieved April 11, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

Shchitova, N. A., Suprunchuk, I. P., Sopnev, N. V., & Tkacheva, T. A. (2024). Геоинформационный анализ миграции молодежи на основе Big Data (на примере Ставропольского края и Республики Дагестан). *Geograficheskiy Vestnik*, *2024*(1), 99–114. https://cyberleninka.ru/article/n/geoinformatsionnyy-analiz-migratsii-molodezhi-na-osnove-big-data-na-primere-stavropolskogo-kraya-i-respubliki-dagestan

Sidorchukova, E. V., Vakulenko, E. V., Mardare, A. K., & Nekrylova, Yu. R. (2024). Analysis of the state and prospects for the development of tourism infrastructure in the Krasnodar Territory. *Vestnik Akademii znaniy*, (4), 333–340.

Singh, D. (2024, December 27). Bayesian Information Criterion: Balancing fit and model complexity. *AI-Enthusiast*. Retrieved April 11, 2025, from https://medium.com/ai-enthusiast/bayesian-information-criterion-balancing-fit-and-model-complexity-3f8973ea0336

Sonnenfeld, J. et al. (2022, August 12). *Business Retreats and Sanctions Are Crippling the Russian Economy*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4167193

Sonnenfeld, J., Roach, S. S., & Tian, S. (2023, March 6). *How the IMF naively parroted Putin's fake statistics–and botched its economic forecast for Russia*. Fortune Europe. https://fortune.com/europe/2023/03/06/imf-naively-parroted-putin-fake-statisticsand-botched-economic-forecast-russia-ukraine/

State of open data in Russia during the war. Cedar. (2024, August). https://cedarus.io/research/open-data-during-war

Statista. (2024). *Leading search engines in Russia in 4th quarter 2023, by visits share*. https://www.statista.com/statistics/1094920/leading-search-engines-by-visits-share-russia/

TheGlobalEconomy.com. (n.d.). *Global economy, world economy*. Retrieved April 11, 2025

Tjaden, J., Auer, D., & Laczko, F. (2019). Linking migration intentions with flows: Evidence and potential use. *International Migration*, 57(1), 36–57. https://doi.org/10.1111/imig.12415

Tsapenko, I., & Yurevich, M. (2022). Nowcasting migration using statistics of online queries. *Economic and Social Changes: Facts, Trends, Forecast / Экономические и Социальные Перемены: Факты, Тенденции, Прогноз*, (1 (79)). https://doi.org/10.15838/esc.2022.1.79.4

Tynaliev, U., & McLean, G. (2011). *Labor Migration from Post-Soviet Kyrgyzstan and Its Developmental Impact in the Context of National Human Resource Development*.

UCLA Statistical Consulting Group. (n.d.). *How can I interpret log transformed variables in terms of percent change in linear regression?* Retrieved April 11, 2025, from https://stats.oarc.ucla.edu/sas/faq/how-can-i-interpret-log-transformed-variables-in-terms-of-percent-change-in-linear-regression/

Ukueva, N. (2010). Migration, Remittances, and Growth. *Duke University Department of Economics*.

United Nations Global Pulse. (2014). *Analyzing social media data for global development*. https://www.unglobalpulse.org/document/analyzing-social-media-data-for-global-development/

Vashisht, R. (2021, January 6). *When to perform a feature scaling?* Atoti Community. Retrieved April 11, 2025, from https://www.atoti.io/articles/when-to-perform-a-feature-scaling/

Villarreal, A., & Blanchard, S. (2013). How job characteristics affect international migration: The role of informality in Mexico. *Demography, 50*(2), 751–775. https://doi.org/10.1007/s13524-012-0153-5

Vorobeva E. (2023). From Workers to Entrepreneurs: Central Asian Migrants in the Russian Business Market. *Central and Eastern European Migration Review* 12(2): 169-189. doi: 10.54667/ceemr.2023.21

Wang, D., Hagedorn, A., & Chi, G. (2021). Remittances and household spending strategies: Evidence from the Life in Kyrgyzstan Study, 2011–2013. *Journal of Ethnic and Migration Studies*, *47*(13), 3015–3036. https://doi.org/10.1080/1369183X.2019.1683442
West, R. M. (2022). Best practice in statistics: The use of log transformation. *Annals of Clinical Biochemistry*, *59*(3), 162–165. https://doi.org/10.1177/00045632211050531PMC+2PubMed+2bioRxiv+2
Williams, R. (2022). "Heteroskedasticity Robust Standard Errors: Some Practical Considerations." The Stata Blog.
World Bank (2023). *Personal remittances, received (% of GDP) - Kyrgyz Republic.* https://data.worldbank.org/indicator/BX.TRF.PWKR.DT.GD.ZS?locations=KG
Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A stepwise AIC method for variable selection in linear regression. *Communications in Statistics - Theory and Methods*, *36*(13–16), 2395–2403. https://doi.org/10.1080/03610920701215639Semantic Scholar+5CoLab+5ResearchGate+5
Ziener, M. (2023, April 18). *They found work in Russia. Then it became a deadly trap*. Los Angeles Times. https://www.latimes.com/world-nation/story/2023-04-18/russia-migrant-workers-drafted-military-ukraine-war
Zeileis, A., & Hothorn, T. (n.d.). *grangertest: Test for Granger Causality* [R package documentation]. RDocumentation. Retrieved April 11, 2025, from https://www.rdocumentation.org/packages/lmtest/versions/0.9-40/topics/grangertest
Агранович, В. Б., Иванова, В. С., & Колодий, Н. А. (2019). Город равных возможностей в зеркале социальных медиа. *Вестник Томского государственного университета. Философия. Социология. Политология*, (47), 248–258. https://cyberleninka.ru/article/n/gorod-ravnyh-vozmozhnostey-v-zerkale-sotsialnyh-media
Баркова, А. С. (2024). Особенности стремления к росту социального капитала личности в регионах России. *Вестник Московского университета. Серия 18. Социология и политология*, 30(2), 189–208. https://doi.org/10.24290/1029-3736-2024-30-2-189-208
Еленин, С. (2018, August 20). "КП": Как мигранты из Кыргызстана обустраивают под себя Москву. Kaktus Media. https://kaktus.media/doc/378673_kp:_kak_migranty_iz_kyrgyzstana_obystraivaut_pod_sebia_moskvy.html
Кузменкин, В. (2021, October 31). *Киргизы в Сибири: работать в Тюмень, торговать в Новосибирск, учиться в томск*. Tayga.info. https://tayga.info/172987
Осмоналиева, Б. (2024, November 14). *Нелегальных мигрантов из Кыргызстана и Таджикистана выдворили с Сахалина.* 24.kg. https://24.kg/obschestvo/311066_nelegalnyih_migrantov_izkyirgyizstana_itadjikistana_vyidvorili_ssahalina/
Савина, С., & Бонч-Осмоловская, К. (2022, 5 октября). Какие регионы отдали больше всего мужчин на войну. *Важные истории*. https://istories.media/stories/2022/10/05/kakie-regioni-otdali-bolshe-vsego-muzhchin-na-voinu/
Торогельдиева, Б. (2020). Миграция Кыргызских Граждан В Россию: Возможности И Риски. *Развитие Территорий*, (4), 67–72. https://doi.org/10.32324/2412-8945-2020-4-67-72
Тынаева, Н. (2024). *Социолог – о том, как угрозы в адрес кыргызстанок в Москве связаны с кампанией «по борьбе с проституцией» в КР.* Радио Азаттык. https://rus.azattyk.org/a/33085683.html
Кожобаева, З. (2013, April 18). *Кыргызские мигранты осваивают Сахалин.* Радио Азаттык.
Корнеев, О. (2024, August 23) *«Они могут остаться надолго» Почему мигранты из Средней Азии продолжают приезжать в Россию и чем это опасно?* Lenta.ru. https://lenta.ru/articles/2024/08/23/korneev/
Коршунова, Ю. А., Пупкова, Е., Шавырин, Д., Ломанова, А. К., Марычев, Г., Блохина, С., Кобилов, Б. Б., Мигунов, Н., Худайбердин, Д., Балабан, М. О., Строгонов, А., & Юмагузин, В. В. (2025). Использование статистики поисковых запросов в изучении интереса к теме гриппа. *Демоскоп Weekly*, (1065–1066).
Осмоналиева, Б. (2024, September 30). *Еще пять антимигрантских законов намерена принять Госдума в ближайшее время*. 24.kg. https://24.kg/vlast/306558_esche_pyat_antimigrantskih_zakonov_namerena_prinyat_gosduma_vblijayshee_vremya/
Полешко, А. (2025, March 7). *Как гражданам Киргизии устроиться на работу в России.* Migron. https://migron.ru/stati/kak-grazhdanam-kirgizii-ustroitsya-na-rabotu/
Пакина, Ю. (2024, April 22). *Прием на работу граждан Киргизии*. Миграционный Эксперт. https://all-migration.com/journal/trudoustroystvo-inostrantsa-oformlyaem-pravilno/priem-na-rabotu-grazhdan-kirgizii/
Четаева, Н. (2024, December 5). Мигранты опасны для России? (лонгрид). Калужское движение За права человека. https://kaluga-zaprava.ru/biblioteka/rf/migranty-opasny-dlja-rossii-longrid/
Шайхеева, Д. Х. (2015). Перспективы развития малого и среднего бизнеса в рамках создания единой туристической зоны в малом городе Тетюши. *Научные труды Вольного экономического общества России*, 2015. Retrieved from https://cyberleninka.ru/article/n/perspektivy-razvitiya-malogo-i-srednego-biznesa-na-territorii-novyh-subektov-rossiyskoy-federatsii.

Appendix

Table 9: Summary of Key Takeaways

Target variable	Regression method	Takeaways
Kyrgyz inflow / outflow / net flows (3 series)	After feeding in different combinations of available covariates, using Bayes Information Criterion (BIC) to select an OLS model that is most likely to be consistent with the "true model."	• Casualties have a positive association with outflows → **push effect associated with severity of war at that time and place** (political turmoil, dodging conscription, psychological effect) • Casualties have a positive association with inflows → **labor substitution effect of migrants for Russian conscripts** (gap in the market left by conscripts filled by migrant workers) • **Pull factor of expected wages:** across regions, a 1% increase in nominal wages associated with 2% increase in Kyrgyz inflows
Kyrgyz Visitors (4 types)		• Casualties positively associated with private and transit travelers but negatively with business travelers → **labor substitution effect more relevant for blue collar workers** since conscripts typically leave behind blue collar jobs
Remittances		• Casualties from 2 months ago positively associated with remittances, even when controlling for exchange rate, inflation, and wages → **labor substitution effect has 2-month delayed effect on remittance payments** • Casualties from 8 months ago negatively associated with remittances → the gap between the two is 6 months, so **labor substitution effect might take 6 months to manifest**

Online Appendix

The online appendix is accessible at the hyperlinks below, containing the following errata:

- Codebook providing more information for all variables: link
- Nowcast of inflows/outflows/net flows with and without covariates: link
- Extended regression tables for inflows/outflows/net flows: link
- GitHub repository includes a plethora of other materials including:
    - Summary statistics and covariance matrices for right-hand-side variables
    - Residual plots, Breush-Pagan test results, etc.
    - Alternative ML methods tried, including with and without covariates
    - Code for nowcasts and regressions

Additional materials are available upon request at michelleschultze25@gmail.com.