Policy Research Working Paper 10534

Agglomeration Economies and Transport Connectivity Revisited

A Regional Perspective Based on Evidence from the Caucasus and Central Asian Countries

Atsushi Iimi

WORLD BANK GROUP
Transport Global Practice
July 2023

Policy Research Working Paper 10534

Abstract

Transport connectivity is an important determinant of agglomeration economies and urbanization. However, measuring its impacts is a complex task when causality is considered. An important empirical challenge comes from potential endogeneity of infrastructure placement. To deal with the endogeneity problem, first, the paper constructs detailed georeferenced connectivity measurements based on micro shipping data collected over 10 years. Then, the system generalized method of moments regression is applied. Using unique data from the Caucasus and Central Asian countries, the paper estimates the impact of transport connectivity on agglomeration economies. It finds that agglomeration economies are significant and persistent in the region. Thus, the existing firm clusters are likely to continue growing. However, a constraint is also found. Large cities exhibit congestion diseconomies. Finally, the paper shows that the improvement of transport connectivity, especially local market accessibility, has a significant effect on agglomeration. By contrast, no clear evidence to support the impact of improved regional connectivity on agglomeration is observed yet. To take full advantage of agglomeration economies at the regional level, further efforts may be needed, for instance, toward increasing efficiency in transportation and logistics, improving the freight load, and/or reducing the time and costs of border crossing, which add to overall transport costs and times.

This paper is a product of the Transport Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The author may be contacted at aiimi@worldbank.org.

Agglomeration Economies and Transport Connectivity Revisited: A Regional Perspective Based on Evidence from the Caucasus and Central Asian Countries

Atsushi Iimi¶

Eastern and Southern Africa, Transport
The World Bank

Key words: Transport connectivity, Firm agglomeration, Dynamic panel data regression.

JEL classification: R32, R41, C23, N75, N74.

¶ Corresponding author.

I. Introduction

1. The new economic geography literature suggests that agglomeration economies are one of the most important determinants of firm productivity, thus, affecting how urban areas are formed and how an economy grows (e.g., Krugman, 1991; Fujita et al., 1999). Supporting evidence can be found all over the world. Mare and Graham (2013) estimate the elasticity of agglomeration in New Zealand. The impacts of agglomeration economies may be heterogenous across countries and industries. In Europe, it is shown that the agglomeration patterns are determined by not only physical proximity but also cultural similarity (Procher, 2011). In the U.S. market, Korean assembly manufacturers flock together with upstream firms, but consumer goods producers are spatially more fragmented (Lee et al., 2012). In the high-tech industry, localization economies may be limited to very close proximity (De Silva and McComb, 2012). Giuliano et al. (2019) provide a good literature review on agglomeration economies and urban form.

2. Among others, transport accessibility is an important determinant of firm location. Firms are likely to be located where access to the road network is good (e.g., Boudier-Bensebaa, 2005). Foreign direct investment is also found to be dependent on proximity to major transport infrastructure, such as highways and hub ports (Belderbos and Carree, 2002; Cieślik and Ryan, 2004; Deichmann et al., 2005; Milner et al., 2006). As a result, the literature generally supports the view that improvement in transport connectivity can facilitate efficient firm activities, stimulating economic growth (e.g., Datta, 2012; Donaldson, 2018; Duranton and Venables, 2018; Ferraz and Coutinho, 2019; Banerjee et al., 2020; IMF 2020; Li et al., 2020).

3. Despite the relatively richness of the literature, the potential impacts of transport connectivity remain more complex than we believe, especially when long-term causality is considered (e.g., Jiang et al., 2017; Valila, 2020). One of the most important empirical challenges is potential endogeneity between economic outcomes and infrastructure placement (e.g., Datta, 2012; Banerjee et al., 2020). While infrastructure investment may increase firm productivity, public infrastructure is often placed where economic productivity is inherently high. Therefore, regardless of its real impact, access to good infrastructure is positively correlated with firm productivity and growth. This infrastructure endogeneity problem is

particularly difficult to address in the transport sector because there are few available time series data in the sector.

4. The current paper's contribution to the literature is at least twofold: First, the paper applies a dynamic panel regression model to obtain an unbiased estimate of the transport connectivity impact on firm agglomeration. In recent years, the instrumental variable (IV) method has been used to deal with the endogeneity issue (e.g., Datta, 2012; Jedwab and Maradi, 2012; Donaldson, 2018; Banerjee et al., 2020). It is shown that quasi-random (or quasi-exogenous) factors, such as geographic conditions, historical events and "unintended" benefits from the network nature of infrastructure, can be good instruments because they are less relevant to the underlying economic performance but may have influenced infrastructure development in subsequent years. Notably, however, the selection of instruments is still dependent on the judgement of researchers and subject to post statistical tests. The current paper applies a dynamic panel data regression model where valid instruments can be explored more systematically.

5. Second, the current paper develops a way of generating detailed spatial data to measure transport connectivity based on micro shipping data. A challenge to apply the above-mentioned dynamic panel data approach is that detailed panel data are needed for a relatively long period of time. A practical norm of the length of time, T, may be five to ten rounds (Labra and Torrecillas, 2018). It means that it takes 5-10 years if annual data are used. This is significantly costly, although some earlier studies collected a series of household surveys. For instance, Khandker et al. (2009) use panel data collected over a period of 7 years. Dercon et al. (2009) analyze five rounds of household surveys carried out for 10 years. However, this approach is costly and often impossible. In the transport sector, there are few data that are regularly collected.

6. In the literature, in addition, potential benefits from transport connectivity are often defined in a dichotomous fashion, depending on proximity (e.g., a distance of 2 km) to road infrastructure (e.g., Dercon et al., 2009; Khandker et al., 2009; IMF, 2020). This binary definition of connectivity is often not realistic particularly when panel data with large T is considered. The potential impacts could be distributed more widely, depending on where people

live and where they go. It is noteworthy that transport infrastructure forms a network. Thus, transport connectivity should be measured in a continuous manner and at a more granular level.

7. To overcome these data challenges, the paper proposes to construct a spatial connectivity dataset using detailed shipment data. Our spatial data allows for estimating transport costs or travel time at each individual road level. Dorosh et al. (2009) use similar geospatial data to identify the impact of market accessibility on crop production in Africa. Their cross-sectional approach remains subject to the potential risk of omitted variable bias and uncontrolled infrastructure endogeneity. This paper takes advantage of stop-by-stop freight shipment data collected over the period 2010-2020.[1] Detailed historical changes in transport connectivity in the Caucasus and Central Asia region are measured. Then, they are regressed on firm location data.

8. The remaining sections are organized as follows: Section II provides a brief regional and country context and elaborates on how our transport measurements are constructed for the sample region. Section III discusses our empirical strategy and summary statistics. Section IV presents the main estimation results. Section V discusses robustness and heterogeneity of the results. Then, Section VI concludes.

II. Regional and Country Contexts and Transport Connectivity Measurements

9. Recent global crises, such as the Coronavirus Disease 2019 (COVID-19) and the war in Ukraine, have reminded us of the importance of maintaining efficient and reliable transport and logistics networks to ensure economic growth.[2] Because of the stagnation and faster-than-expected recovery in merchandise trade and global outputs, the world shipping capacity remains constrained by shortages in logistical equipment, containers, service operators, such as truck drivers and port operators, and cargo vessels (UNCTAD 2021). According to the Drewry database, the average global maritime freight rate reached over US$10,000 per 40-foot container

---

[1] See ADB (2020) for more details. We would like to express our special thanks to the Central Asia Regional Economic Cooperation (CAREC) member countries and the Asian Development Bank (ADB) Corridor Performance Measurement and Monitoring (CPMM) team for sharing relevant raw shipping data.

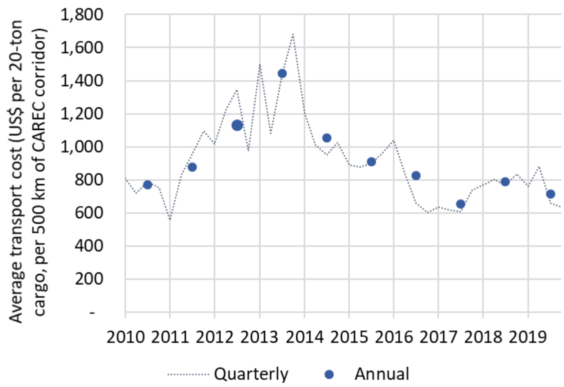
[2] The following analysis examines the period from 2010 to 2020. The impact of the COVID crisis may not be reflected sufficiently in our dataset.

in 2021, five times higher than the level prior to the COVID crisis. Road transport costs also increased substantially. The harmonized consumer price index for transport in Europe jumped from 106.7 in 2020 to 127.6 in 2022 (Eurostat). No doubt transport connectivity is now more important than ever.

10. Regional integration and transport connectivity are of particular importance for the Caucasus and Central Asia region, which is landlocked and located at a strategic place between Europe and East Asia, and between the Russian Federation and South Asia (e.g., Incaltarau et al., 2022). After the collapse of the Soviet Union in 1991, the region experienced long economic stagnation and transition as well as regional disintegration. In recent years, the region has started recovering with relatively robust growth (Pomfret, 2010, 2021). As pointed out by Cheong and Turakulov (2022), trade facilitation and regional reintegration are of vital importance to sustain growth in the region. Since the early 2010s, a number of large transport investments have been made to reestablish the regional connectivity. For instance, Kazakhstan embarked upon a massive road and rail investment program, Nurly Zhol, to connect major cities, logistics centers and free-trade zones to the regional market, i.e., border crossings, including ports on the Caspian Sea dry ports. The country spent about US$37 billion over five years: 2013-2017 (UNECE, 2019).

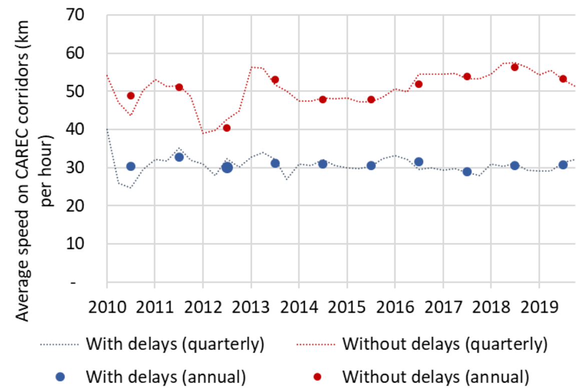
11. Due to such efforts, the regional connectivity seems to have been improved. According to the ADB CAREC Corridor Performance Measurement and Monitoring (CPMM) program, average transportation costs on the regional corridors were halved from US$1,400 per 20-ton cargo per 500 km in 2014 to US$700 in 2019 (**Figure 1**). Average speed also increased gradually. Note that the recorded travel times include not only driving time but also various delays and roadside stops, such as loading and offloading, and security checks. With such delays excluded (i.e., data “without delays”), the average speed had a slight increasing trend since 2015 (**Figure 2**).

Figure 1. Average transport costs on corridors



Source: ADB (2020).

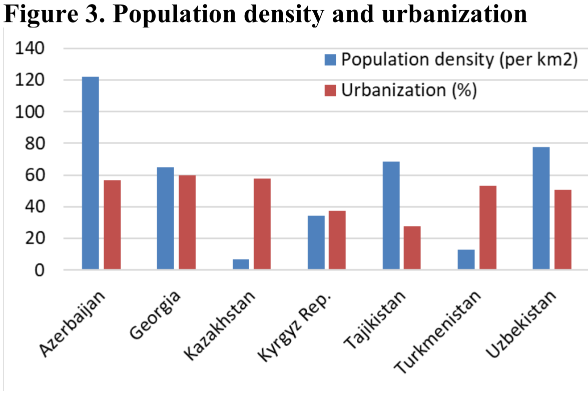
Figure 2. Average travel speed on corridors



Source: ADB (2020).

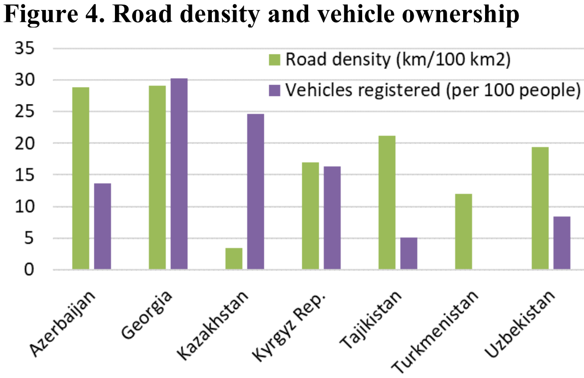
12. Transport connectivity differs substantially across countries. The current paper considers seven countries: Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan.[3] Although traditionally available transport data are thin, they already indicate general challenges in the transport sector. Population density is generally low except for Azerbaijan (**Figure 3**). Limited urbanization also implies a challenge to connect people in remote areas. The urbanization process is particularly slow in Kyrgyz Republic and Tajikistan. Road density is very low in Kazakhstan and Turkmenistan. On the other hand, Azerbaijan and Georgia are considered to be relatively well connected (**Figure 4**). On the fleet side, vehicle ownership is particularly high in Kazakhstan and Georgia.

Figure 3. Population density and urbanization



Source: WDI.

Figure 4. Road density and vehicle ownership



Sources: WDI and WHO.

---

[3] In the empirical analysis, only five countries are analyzed: Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan and Uzbekistan due to data availability.

13. Unfortunately, however, these national-level transport and socioeconomic indicators cannot present actual transport connectivity on the ground. By construction, these indicators barely change over time even if some roads are substantially improved. For instance, road density does not change unless a new road is constructed. In reality, many road projects aim at rehabilitating or maintaining existing roads, not developing new ones. In addition, the indicators aggregated at the national level ignore potential spatial heterogeneity in a country. Transport connectivity often differs significantly across subnational areas.

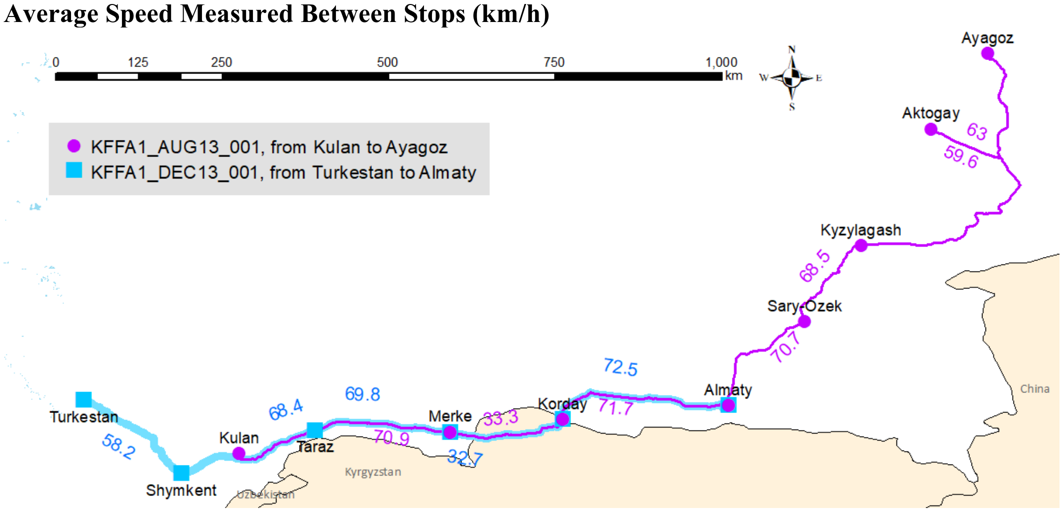
14. To capture historical changes in transport connectivity more precisely, spatial techniques are used to construct detailed georeferenced data based on stop-by-stop freight shipping data. The CMPP collected by the ADB database takes a time/cost-distance (TCD) approach (UNESCAP).[4] The data is focused on regional long-haul shipments originated from and/or destined for the CAREC member countries over the period of 2010-20. It comprises detailed micro shipping cost and time, which are tracked movements of a truck or cargo from door to door along transport corridors, including borders and other inspection points.

15. The original database contains over 187,000 observations, covering both road and rail shipments. Excluding data with missing information (e.g., origin or destination), the paper uses 93,786 road shipment data, which could be georeferenced. A total of 14,342 shipping carriers transported these cargos among 565 locations.[5] Note that as shown in **Figure 5**, each observation is recorded by a shipping carrier at every major stop point (normally, city or border point). Thus, even on the same road segment, we observe different transport costs and times. For instance, some carriers may have driven faster than others. The figure shows average speeds recorded by two Kazakh shipping carriers in 2013. Origins, destinations and stopped points can be different across carriers. Even if they are different (e.g., between Shymkent and Merke), the speed (or time) data are available at the road segment level. Similarly, the transport costs differ across carriers and the cost data are also disaggregated at the road level.

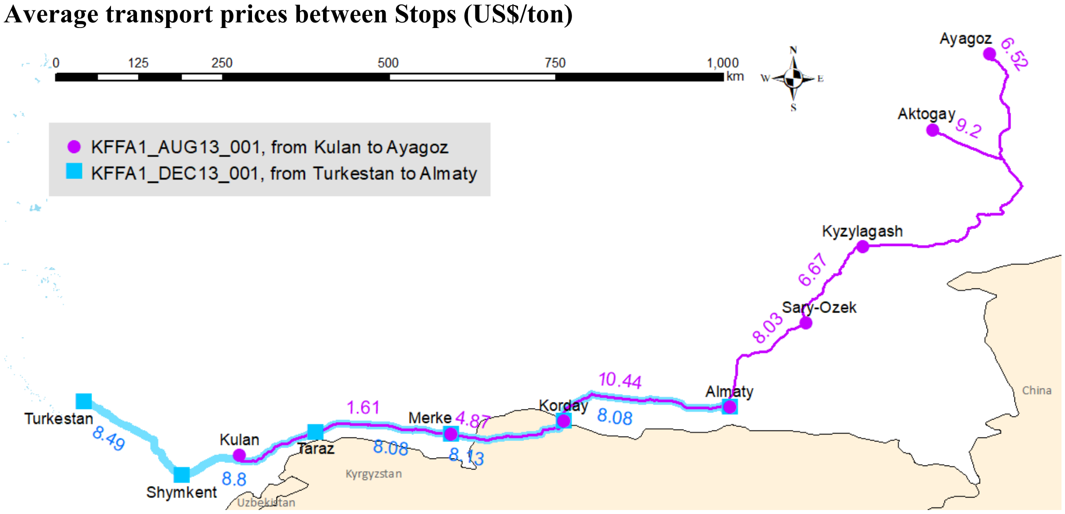
[4] See ADB (2014) for detailed data collection methods.
[5] While 527 locations are identified as origins, 516 locations are georeferenced as destinations.

16. To aggregate the data at the road segment level, annual average costs and times are calculated when more than one carrier passed a particular road segment. For other road segments where CPMM data are not available, the normal speed and road user costs are assumed.[6]

Figure 5. Examples of Shipping Data:
Average Speed Measured Between Stops (km/h)



Average transport prices between Stops (US$/ton)



17. The developed spatial database allows to estimate transport costs and time between any pair of two locations by using spatial software. The best route to move from one location to another is identified by minimizing accumulated transportation costs or travel time. Different

[6] Road user costs are assumed to be 6.5, 7.2 and 9.2 U.S. cents per ton-km for regional highways, national primary and secondary roads, and other tertiary roads, respectively. Similarly, average speed is assumed to be 49.5, 41.3 and 27.8 km per hour based on road class.

types of transport connectivity are examined by using different definitions of destination cities. Two thresholds are considered: 1 million and 100,000 in population. The former aims to capture the proximity to mega markets in the region. There are eight cities in the five countries to be analyzed (Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan and Uzbekistan). The latter threshold is used to measure accessibility to smaller national or local markets. Forty-one towns and populated areas are found in the seven countries.

18. In the region, there exists a substantial difference in connectivity. Even within a country, some areas are better connected than others. In general, the connectivity looks good along a regional corridor connecting Almaty, Bishkek, Tashkent, Dushanbe and Ashgabat, and a national corridor connecting Almaty, Astana, Pavlodar and Oskemen in Kazakhstan. The measured connectivity looks different, depending on the destination. **Figures 6** and **7** show the connectivity to the nearest city with more than 1 million and 100,000 populations, respectively. The connectivity also varies depending on whether it is measured by travel time or transport costs (**Figures 8** and **9**). The distributions are broadly similar. Transport costs and speed are correlated, but there are certain differences at a granular level.

Figure 6. Travel time to the nearest city with >1 million populations, 2019

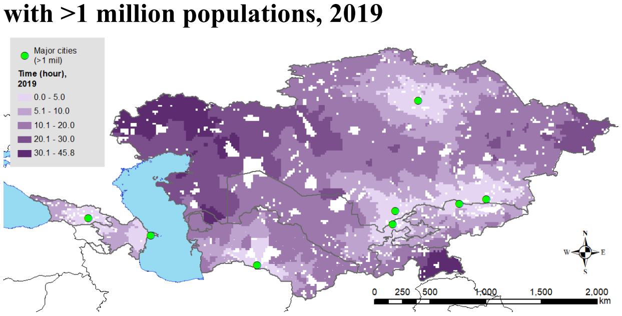


Figure 7. Travel time to the nearest city with >100,000 populations, 2019

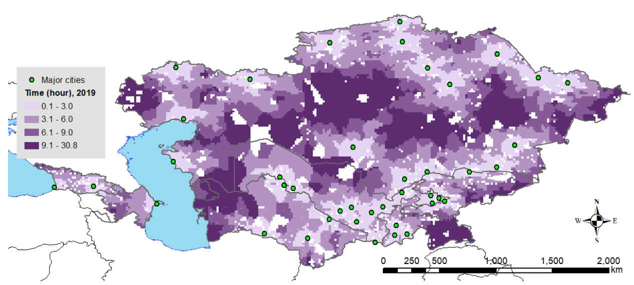


Figure 8. Transport costs to the nearest city with >1 million populations, 2019

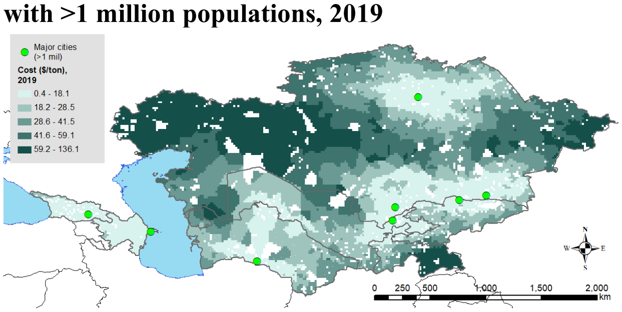
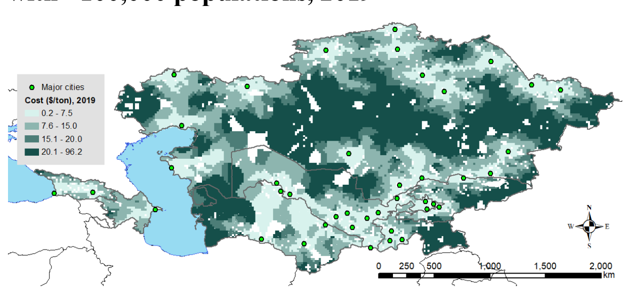
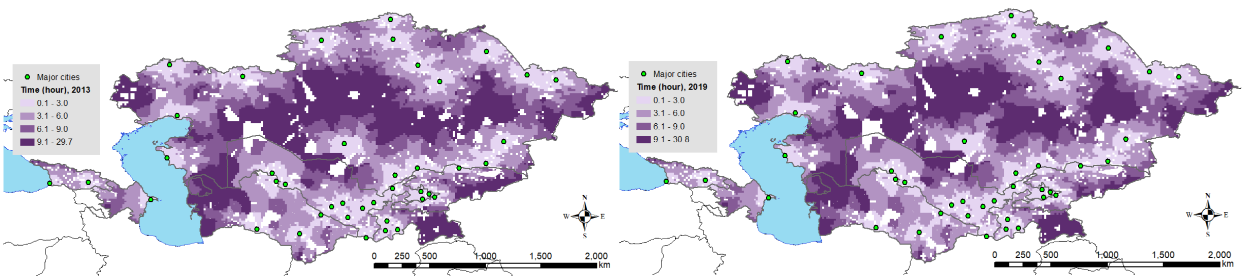


Figure 9. Transport costs to the nearest city with >100,000 populations, 2019



19. More importantly, these connectivity measurements change over time. For instance, when the 2013 and 2019 data are compared, the average trip time to the nearest large city declined in some areas (**Figure 10**). In other places, transport costs seem to have increased. The following analysis takes advantage of these changes over time to identify the unbiased economic impacts of transport connectivity.

Figure 10. Trip time to the nearest city with >100,000 populations
(2013) **(2019)**



III. Empirical Models and Data

20. To examine the dynamics of growth and transport connectivity, the paper focuses on a particular economic outcome, agglomeration economies. The economic geography literature suggests that agglomeration economies occur when a number of firms are located in the same place (e.g., Krugman, 1991; Fujita et al. 1999). They can share common intermediate inputs and produce complementary goods and services. Transportation connectivity is one of the important determinants of their investment location (Holl, 2004; Procher, 2011; Lee et al. 2012; Mare and Graham, 2013).

21. The Caucasus and Central Asian countries are of particular interest from the agglomeration point of view because they have experienced significant structural changes after the collapse of the Soviet Union. Notably, some economies shrank, and others recovered rapidly. The following analysis is focused on five countries: Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan and Uzbekistan where the firm registry data are available. They seem to have had different growth dynamics (**Figure 11**). In Kazakhstan, for instance, the number of active enterprises registered in

the national firm census database declined in 2015 and then increased more recently. In Uzbekistan, the number of enterprises and organizations increased particularly after 2019. In Kyrgyzstan, the growth rate has been gradually reducing (**Figure 12**).

22. Note that the definition and coverage of firm registry data differ across countries. While Georgia reports the number of registered enterprises, Kyrgyzstan statistics only cover active businesses for which detailed financial and operational data are available. Uzbekistan statistical data aggregate not only economic entities but also other nonprofitable organizations (**Table 1**). From the empirical point of view, ideally, the same definition should have been used. These differences have to be taken into account in the analysis.

Figure 11. Number of registered enterprises

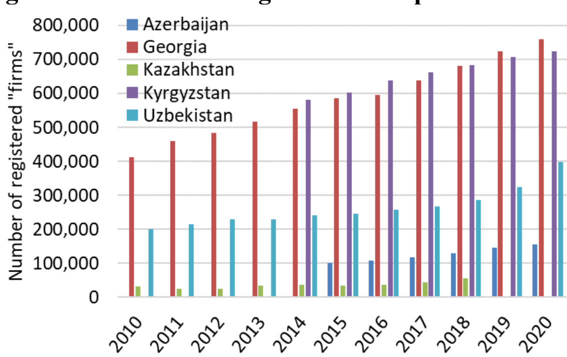


Figure 12. Growth rates of number of enterprises

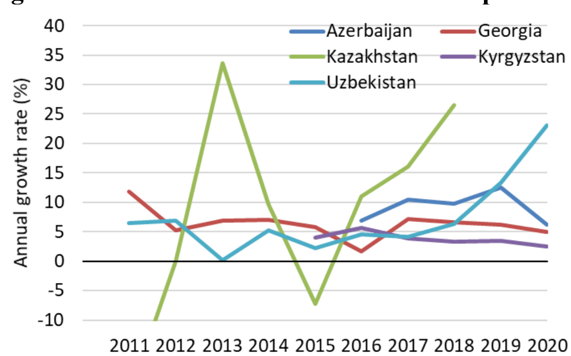


Table 1. Definition of available data on the number of firms registered

Country	Description of data	Administrative level	Data period
Azerbaijan	Number of registered statistical units (excluding individual entrepreneurs)	73 districts	2015-20
Georgia	Number of registered entities in the business sector	12 regions	2010-20
Kazakhstan	Number of registered enterprises	163 districts	2010-18
Kyrgyzstan	Number of operating businesses	9 regions	2014-20
Uzbekistan	Number of active enterprises and organizations	14 regions	2010-20

23. Our hypothesis is that firm agglomeration is facilitated by improved transport connectivity. Even within one country, there is considerable variation in the number of registered enterprises across districts or regions. For the period of 2015-18, for instance, significant increases were observed in Western Georgia, Karaganda Region of Kazakhstan and Kyrgyzstan (**Figure 13**). Some of these areas are consistent with the places where transport connectivity was improved

(**Figure 14**). However, they are not perfectly matched. There is high correlation between the number of enterprises and travel time required to access the nearest large city (**Figure 15**).

Figure 13. Changes in the number of enterprises

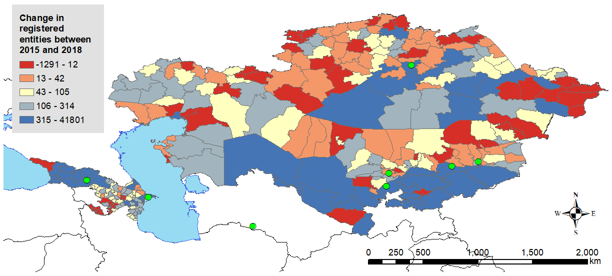


Figure 14. Changes in travel time to the large city

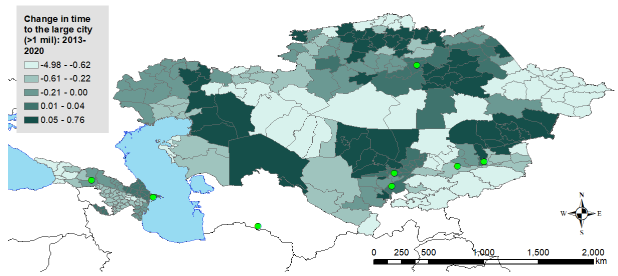
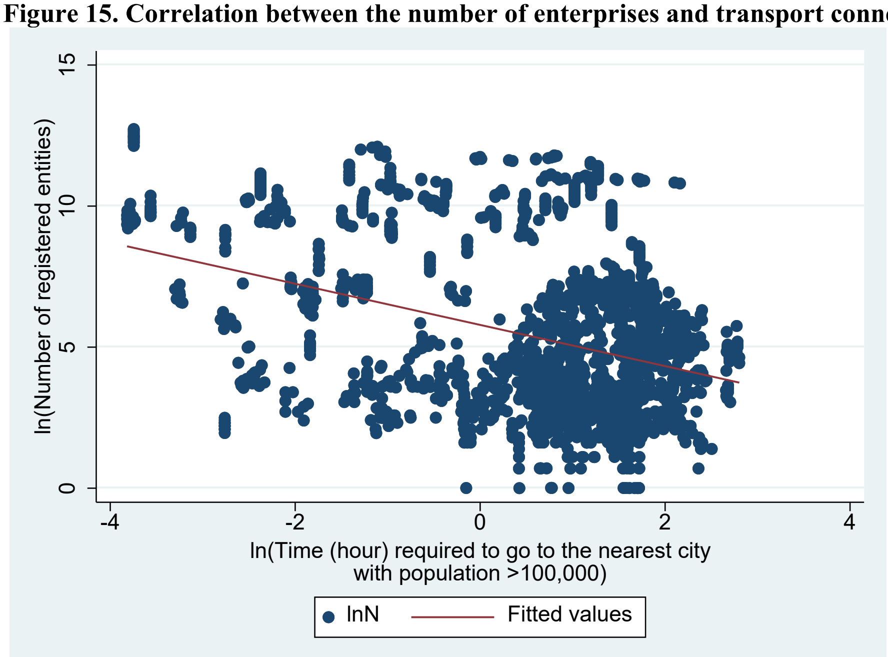


Figure 15. Correlation between the number of enterprises and transport connectivity



24. Causality remains open to argument, which is the main question of the paper. The endogeneity of infrastructure placement is of particular concern. There are many potential factors that are unobservable for researchers but may affect both firm location and infrastructure investment. To deal with this problem, the following analysis uses the dynamic panel data regression. The fixed-effects regression model can mitigate the endogeneity bias under certain assumptions (e.g., parallel trend assumption) but cannot eliminate it in general (e.g., Nickell,

1981).[7] In the region, IMF (2020) applies the fixed-effects method to estimate the impact of transport infrastructure under Nurly Zhol on firm revenues and profits in Kazakhstan. However, the fixed-effects model may lead to biased estimates if there are any time-variant individual effects that are correlated with the residuals.

25. The generalized method-of-moments (GMM) estimator is more efficient than the fixed-effects estimator. In this paper, the system GMM using instruments from both levels and differences (e.g., Roodman, 2009). The traditional difference GMM approach (Arellano and Bond, 1991) where only differences lagged differences are used as instruments tends to poorly perform particularly when the time series persistence is high in a dependent variable. The system GMM can generate more instruments under the assumption that their changes are uncollated with the individual fixed effects and better control for the endogeneity and autocorrelation issues simultaneously. The reliability and validity of instruments will be examined by ex post statistical tests, such as the Hansen overidentifying restriction tests.

26. Assuming that agglomeration economies are influenced by transport connectivity and other factors, the following dynamic panel model is considered:

$$\ln N_{it} = \sum_p \alpha_p \ln N_{it-p} + \sum_k \beta_{v_k} \ln x_{ikt} + u_i + u_t + \varepsilon_{it} \qquad (1)$$

where N_{it} is the number of firms registered at location i at year t. Our unit of analysis is region or district. The level of spatial disaggregation is different across countries (see **Table 1** above). All time-invariant characteristics that are specific to location i, such as local culture, language and politics, are assumed to be captured by the location-specific fixed-effects, u_i, and removed by the first difference transformation of the dynamic panel model. The systematic differences in definition and coverage of firm data across countries are also expected to be controlled by u_i.

27. To deal with the persistence of agglomeration economies, the lagged dependent variables are included on the right-hand side. As shown above, our dependent variable exhibits strong

[7] As will be shown below, the dynamic panel model estimator is more suitable to our data where there is a strong first-order serial correlation in the first-difference residuals.

time-series persistence. Thus, the coefficients α's are expected to be significant. In our data, it was found that the first three lags of $\ln N$ significantly impact the current value of $\ln N$, while the older lags do not. Since the lagged dependent variable, $\ln N_{it-p}$, is generally correlated with the individual-specific fixed-effects u_i, the ordinary least squares (OLS) estimator will be biased by construction.

28. The time dummy variables, u_t, are also included in the model. Note that since our model includes the three lags of the dependent variable as regressors, the first three periods, i.e., t=2010, 2011 and 2012, are used as a baseline.

29. Transport connectivity is included in X. Two types of connectivity are considered: (i) regional connectivity and (ii) local market connectivity. Two thresholds on population are used. For the former, the nearest city with a population of more than 1 million is considered, such as Almaty and Tashkent. For the latter, relatively small local markets with a population of more than 100,000 are considered. In both case, two measurements are constructed based on transportation costs and travel time. Thus, we have four transport variables: regional connectivity measured by time (HR_{REGION}), local connectivity by time (HR_{LOCAL}), regional connectivity by transport cost (TC_{REGION}), and local connectivity by cost (TC_{LOCAL}).

30. In addition, population density at location i, $POPD_{it}$, is also included in X. This is expected to control three potential issues. First, it can help to control for the size effect in our data. The size of administrative areas varies substantially. Some districts, such as metropolitan areas, are highly populated compared with other rural areas. It may be normal that more firms exist where more people live, regardless of agglomeration economies.

31. Second, the local population can capture the impact of broader external economies, i.e., urbanization economies. In theory, urban form is an endogenous process (e.g., Palivos and Wang, 1996; Konishi, 1996). Highly populated areas are often developed as consumer or amenity cities and employment centers (e.g., Glaeser et al., 2001; Craig et al., 2016). If this is the case, the coefficient of $POPD$ should be estimated to be significantly positive.

32. Finally, there is a possibility to observe congestion diseconomies. Urban growth may eventually be deterred by high traffic congestion and housing costs (e.g., Henderson, 1974; Sedgley and Elmslie, 2001; Muroishi and Yakita, 2021). If this is the case, the coefficient could be negative. In any case, population density is an important part of economic dynamics and the population variable needs to be treated as an endogenous variable.

33. The GMM approach not only eliminates any time-invariant location-specific characteristics u_i but also addresses the potential endogeneity of all these endogenous variables. Both twice-lagged differences and levels of the endogenous variables, i.e., the lagged dependent variables, the transport variables and population density, are used as instruments. Our panel data are unbalanced. The length of time T differs across countries, as shown in **Table 1**. The longest length is 11 years for Georgia and Uzbekistan. Azerbaijan data only covers 6 years. The unconditional GMM generates a total number of instruments of about 200. This may look high enough to cause overfitting bias, i.e., instrument proliferation (e.g., Roodman, 2009). The empirical robustness will be examined by reducing the instrument count in the following sections.

34. The summary statistics are shown in **Table 2**. The number of groups varies across countries, from 9 regions in Kyrgyzstan to 163 districts in Kazakhstan. In total, our sample data comprises 255 administrative areas in the five countries. At each location, on average 7,637 enterprises are located. The average travel time to go to the nearest largest city with more than 1 million inhabitants is 11 hours. The travel time to a smaller city with 100,000 inhabitants is much shorter at 3.4 hours but still has a wide variation from nearly zero to 16 hours.

35. The transport cost to bring goods to the nearest city also varies significantly. It costs on average US$25 per ton if it is transported to a large city. The average transport cost to a nearby small city is much lower at US$8 per ton. Note that all cost data are converted and normalized to 2010 constant dollars using the U.S. consumer price index. The population size differs across districts and regions, from a few thousands to over 3 million with an average of 291,000 inhabitants. This translates into a wide range of population density from 0.3 to 7,800 persons per km^2.

Table 2. Summary statistics

Variable	Abb.	Obs	Mean	Std. Dev.	Min	Max
Number of "enterprises" registered at location *i* at time *t* [1]	*N*	1374	7637	27500	1	331224
Travel time required to go to the nearest large regional city with more than 1 million inhabitants (hour)	HR_{REGION}	1374	11.15	9.88	0.02	44.97
Travel time required to go to the nearest local city with more than 100,000 inhabitants (hour)	HR_{LOCAL}	1374	3.40	2.89	0.02	16.18
Transport costs to convey goods to the nearest large regional city with more than 1 million inhabitants ($/ton)	TC_{REGION}	1374	25.20	21.88	0.02	116.50
Transport costs to convey goods to the nearest local city with more than 100,000 inhabitants ($/ton)	TC_{LOCAL}	1374	8.29	7.53	0.02	48.27
Population density at location *i* at time *t* (population per km^2)	*POPD*	1374	197.39	909.73	0.28	7840.24
Time dummy variables:						
t=2013		1374	0.13	0.34	0	1
t=2014		1374	0.13	0.34	0	1
t=2015		1374	0.13	0.34	0	1
t=2016		1374	0.13	0.34	0	1
t=2017		1374	0.14	0.35	0	1
t=2018		1374	0.18	0.39	0	1
t=2019		1374	0.07	0.26	0	1
t=2020		1374	0.07	0.26	0	1

[1] The definition of enterprises differs across countries.

IV. Main Estimation Results

36. First of all, the OLS and fixed-effects models are performed to check how our agglomeration equation behaves. The estimation results are broadly consistent with prior expectations (**Table 3**). The agglomeration variable is found to have high persistence. The first three lagged agglomeration variables are all significant in the OLS models. As expected, firm agglomeration is positive related to transport connectivity. Both coefficients of regional and local connectivity are negative. Note that our transport variables represent the level of deterrence of transportation. Thus, the negative coefficients mean that the number of firms increases with transport connectivity.

37. Certainly, however, there are a number of unobservables omitted from our model. The fixed-effects models can control for time-invariant unobservables. The results turned out to be

less conclusive, though.[8] While the coefficient of HR_{LOCAL} is significantly negative, TC_{REGION} has a positive coefficient. The fixed-effects results may also be biased because of the uncontrolled endogeneity issue.

Table 3. OLS and fixed effects estimation

	OLS			OLS			Fixed effects			Fixed effects		
	Coef.	Std.Err.		Coef.	Std.Err.		Coef.	Std.Err.		Coef.	Std.Err.	
lnN_{t-1}	0.692	(0.072)	***	0.696	(0.072)	***	0.383	(0.062)	***	0.376	(0.061)	***
lnN_{t-2}	0.155	(0.060)	***	0.153	(0.060)	***						
lnN_{t-3}	0.133	(0.040)	***	0.131	(0.040)	***						
lnHR_{REGION}	-0.011	(0.005)	**				0.151	(0.122)				
lnHR_{LOCAL}	-0.032	(0.005)	***				-0.191	(0.111)	*			
lnTC_{REGION}				-0.004	(0.005)					0.206	(0.057)	***
lnTC_{LOCAL}				-0.027	(0.005)	***				0.038	(0.050)	
ln*POPD*	-0.034	(0.005)	***	-0.031	(0.005)	***	0.105	(0.287)		-0.146	(0.272)	
t=2012							0.184	(0.032)	***	0.166	(0.031)	***
t=2013							0.298	(0.031)	***	0.244	(0.031)	***
t=2014	0.054	(0.029)	*	0.052	(0.028)	*	0.369	(0.028)	***	0.324	(0.027)	***
t=2015	-0.062	(0.029)	**	-0.064	(0.029)	**	0.333	(0.030)	***	0.287	(0.029)	***
t=2016	-0.052	(0.027)	*	-0.053	(0.027)	**	0.431	(0.034)	***	0.383	(0.034)	***
t=2017	0.076	(0.025)	***	0.076	(0.025)	***	0.562	(0.037)	***	0.516	(0.036)	***
t=2018	0.312	(0.034)	***	0.309	(0.035)	***	0.845	(0.050)	***	0.800	(0.047)	***
t=2019	0.105	(0.034)	***	0.093	(0.034)	***	0.625	(0.051)	***	0.591	(0.048)	***
t=2020	0.090	(0.028)	***	0.076	(0.028)	***	0.641	(0.051)	***	0.626	(0.048)	***
constant	0.327	(0.032)	***	0.329	(0.035)	***	2.463	(0.824)	***	2.715	(0.757)	***
Obs.	1,374			1,374			1,893			1,893		
R-squared	0.9925			0.9924			0.9247			0.7214		
F stat.	16771.57			16728.29			138.56			124.32		
No. of group dummy variables							261			261		
Hausman test stat.							652.67	***		688.06	***	

Note: The dependent variable is ln*N*. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

38. To manage the potential endogeneity issue, the system GMM is adopted. The results are shown in **Table 4**. First of all, it is found that the selected instrumental variables are valid according to the Hansen overidentifying restriction test. The test statistics are estimated at 220.2 to 253.3, depending on the specification. For the serial correction, as expected, the first-order autoregressive test AR(1) detects significant correlation in the residuals of the differenced

[8] In the fixed-effects models, the first lag of the dependent variable is included on the right hand side. It is found that only the first lag has a significant impact on the current value.

equation. The test statistics are estimated at about -5.7, which is significant to indicate the presence of autocorrelations. On the other hand, the differenced residuals do not have significant second-order autoregressive AR(2) process, which also supports the validity of our GMM approach.

39. As usually observed in the literature (e.g., Procher, 2011; Mare and Graham, 2013; Giuliano et al., 2019), agglomeration economies are strong and significant in our estimation. The first lag of the dependent variables has a significant coefficient of about 0.67. Thus, firms are very likely to be co-located where other firms already exist. Moreover, the second and third lags are also significant. It implies that the agglomeration process is highly endogenous and persistent over time. Therefore, in the sample region, the current firm clusters are likely to continue growing further, attracting more new enterprises. On the other hand, the lagging areas where fewer firms are located may further lose their attractiveness. This looks consistent with the existing argument that the Eastern European and Central Asian countries have experienced dynamic transition processes until recently (Pomfret, 2010, 2021; Incaltarau et al., 2022).

40. Regarding transport connectivity, it is found that efficient accessibility to local markets is particularly important for firm agglomeration. With the time-based connectivity measurement used, the coefficient of $\ln HR_{LOCAL}$ is estimated at -0.092, which is statistically significant, meaning that firms flock together where local market accessibility is good in terms of travel time. On the other hand, the effect of regional connectivity measured by HR_{REGION} is consistently insignificant. In our data, there is no evidence that improved regional connectivity would facilitate agglomeration of firms. This may be contradictory to prior expectations of policy makers, who may have envisaged deepening regional development and integration through massive investment in regional transport corridors (e.g., ADB 2014; Cheong and Turakulov, 2022). Notably, however, it normally takes a long time for the impacts of infrastructure investment to materialize. Emerging evidence is that the reginal connectivity has surely improved (ADB 2020). More time may be needed to observe the likely real impact at the regional level.

41. When market accessibility is measured by transport costs, the results look less stable. As in the case where the time-based variable is used, the local connectivity TC_{LOCAL} has a significantly negative coefficient of -0.076. Thus, again, firm agglomerations are likely to be fostered by local market access improvement. The impact of TC_{REGION} turned out to be positive, implying that firms are located where it costs more to transport goods from or to the regional mega market, which is counterintuitive. When both time and cost measurements are included in the equation, the only time-based local market accessibility, HR_{LOCAL}, has a significant coefficient, which is estimated at -0.168. The statistical significance of the travel-time based connectivity variables disappears. All the indications are that efficient and rapid access to local markets is essential to foster firm clusters and stimulate agglomeration economies.

Table 4. System GMM estimation with all possible instruments

	GMM			GMM			GMM		
	Coef.	Std.Err.		Coef.	Std.Err.		Coef.	Std.Err.	
lnN_{t-1}	0.660	(0.063)	***	0.676	(0.064)	***	0.666	(0.064)	***
lnN_{t-2}	0.174	(0.050)	***	0.170	(0.051)	***	0.174	(0.051)	***
lnN_{t-3}	0.147	(0.047)	***	0.146	(0.049)	***	0.146	(0.048)	***
lnHR_{REGION}	0.011	(0.018)					0.030	(0.057)	
lnHR_{LOCAL}	-0.092	(0.015)	***				-0.168	(0.056)	***
lnTC_{REGION}				0.041	(0.020)	**	-0.026	(0.055)	
lnTC_{LOCAL}				-0.076	(0.018)	***	0.095	(0.059)	
ln*POPD*	-0.080	(0.014)	***	-0.081	(0.014)	***	-0.078	(0.012)	***
t=2014	0.062	(0.030)	**	0.058	(0.031)	*	0.061	(0.030)	**
t=2015	-0.054	(0.032)	*	-0.060	(0.033)	*	-0.055	(0.033)	*
t=2016	-0.048	(0.026)	*	-0.052	(0.026)	**	-0.049	(0.026)	*
t=2017	0.083	(0.027)	***	0.079	(0.027)	***	0.077	(0.026)	***
t=2018	0.370	(0.038)	***	0.359	(0.038)	***	0.358	(0.038)	***
t=2019	0.257	(0.053)	***	0.234	(0.053)	***	0.233	(0.053)	***
t=2020	0.242	(0.044)	***	0.216	(0.044)	***	0.223	(0.045)	***
constant	0.393	(0.061)	***	0.310	(0.091)	***	0.306	(0.068)	***
Obs.	1,374			1,374			1,374		
Wald stat	489840			477473			554138		
No. of groups	255			255			255		
No. of instruments	214			214			312		
Arellano-Bond test:									
AR(1)	-5.76	***		-5.78	***		-5.77	***	
AR(2)	-0.12			-0.12			-0.04		
Hansen overidentifying restriction test:									
Chi2	223.20			222.42			253.37		
p-value	0.125			0.133			0.965		

Note: The dependent variable is ln*N*. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

V. Discussion

42. One may wonder if too many instruments are created in our system GMM (i.e., the problem of instrument proliferation). Especially in the specification including both time- and cost-based transport variables (i.e., the last column model in **Table 4**), the number of constructed instruments reaches 312. When a large number of instruments are created as in the system GMM setting, the Hansen test statistics tend to be weak because the endogenous variables are almost perfectly fit at the first stage (e.g., Roodman, 2009). The robustness of the results is examined through reducing the number of instruments. With different sets of lags, the system GMM estimation is performed again (**Table 5**). It is confirmed that the estimation results are robust when the instrumenting variables are limited to the second to at least fifth lags. If the number of instruments is reduced further (e.g., the number of lags is equal to 3), the statistical validity of instruments would be lost. The Hansen overidentifying restriction test statistic is estimated at 224.59 and rejects the null hypothesis. The p-value on the Hansen test is 0.02.

Table 5. System GMM estimation with different lags of instruments

	lags of instruments = 3			lags of instruments = 5			lags of instruments = 7		
	Coef.	Std.Err.		Coef.	Std.Err.		Coef.	Std.Err.	
lnN_{t-1}	0.667	(0.064)	***	0.667	(0.064)	***	0.666	(0.064)	***
lnN_{t-2}	0.174	(0.051)	***	0.174	(0.051)	***	0.174	(0.050)	***
lnN_{t-3}	0.146	(0.049)	***	0.145	(0.048)	***	0.146	(0.048)	***
lnHR_{REGION}	0.038	(0.064)		0.040	(0.060)		0.032	(0.057)	
lnHR_{LOCAL}	-0.184	(0.059)	***	-0.172	(0.057)	***	-0.165	(0.056)	***
lnTC_{REGION}	-0.031	(0.063)		-0.029	(0.058)		-0.027	(0.056)	
lnTC_{LOCAL}	0.111	(0.063)	*	0.098	(0.060)		0.092	(0.058)	
ln*POPD*	-0.076	(0.012)	***	-0.077	(0.012)	***	-0.078	(0.012)	***
t=2014	-0.060	(0.030)	**	-0.060	(0.030)	**	-0.061	(0.030)	**
t=2015	-0.116	(0.020)	***	-0.116	(0.020)	***	-0.116	(0.020)	***
t=2016	-0.110	(0.021)	***	-0.110	(0.021)	***	-0.110	(0.021)	***
t=2017	0.015	(0.020)		0.016	(0.020)		0.016	(0.020)	
t=2018	0.296	(0.028)	***	0.298	(0.029)	***	0.298	(0.029)	***
t=2019	0.170	(0.040)	***	0.172	(0.040)	***	0.173	(0.039)	***
t=2020	0.160	(0.032)	***	0.161	(0.032)	***	0.162	(0.031)	***
constant	0.344	(0.080)	***	0.350	(0.074)	***	0.369	(0.073)	***
Obs.	1,374			1,374			1,374		
Wald stat	549658			552955			552778		
No. of groups	255			255			255		

No. of instruments	199	261	298
Arellano-Bond test:			
AR(1)	-5.80 ***	-5.78 ***	-5.76 ***
AR(2)	-0.03	-0.04	-0.03
Hansen overidentifying restriction test:			
Chi2	224.59	248.09	250.43
p-value	0.02	0.433	0.912

Note: The dependent variable is ln*N*. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

43. One of the unexpected results from the above GMM estimation may be that population density ($POPD$) has a significantly negative coefficient, which is about -0.08 regardless of specification. It means that holding everything else constant, firms are more likely to be located where population density is lower. This may be counterintuitive because this coefficient is normally expected to be positive. That is, the more people, the more firms. However, our finding is opposite, indicating congestion diseconomies, rather than urbanization economies. It can be interpreted to mean that the populated areas in the region have not evolved yet as sustainable producer cities but as consumer or amenity cities (e.g., Craig et al., 2016; Giuliano et al., 2019). Generating jobs for inclusive growth may be a consistent challenge in the region (ADB, 2019).

44. To investigate the congestion effect further, the conventional structural test is applied. The sample data is divided into two subsamples according to population density. The median is 12.67. The estimated Wald test statistic is 134.06, which is significant (**Table 6**). Thus, the structure of the estimated equation differs between these two subsamples. Note that in this structural test model, only the second lags are used as instruments to avoid the problem of instrument proliferation.

45. As expected, congestion diseconomies are only observed in the first subsample where population density is relatively high. The coefficient of $\ln POPD$ is estimated at -0.051, which is statistically significant. On the other hand, the coefficient is insignificant for the second subgroup with lower population density. In addition, it is also found that the transport connectivity impacts on agglomeration only where population density is high. HR_{LOCAL} has a significant coefficient of -0.062 for the first subsample. The coefficient for the second group is also negative but not significant. The result also makes sense and reminds the importance of transport connectivity to mitigate congestion, while fostering agglomeration.

46. In sum, it can be concluded that large cities in the Caucasus and Central Asia region exhibit congestion diseconomies. In general, firms dislike to be located in populated areas. Despite the congestion effect, however, it is still important to improve local market accessibility in order to bring in more new firms.

Table 6. Separated GMM estimation by population density

	POPD > 12.67			*POPD* < 12.67		
	Coef.	Std.Err.		Coef.	Std.Err.	
$\ln N_{t-1}$	0.683	(0.087)	***	0.655	(0.059)	***
$\ln N_{t-2}$	0.177	(0.038)	***	0.176	(0.060)	***
$\ln N_{t-3}$	0.122	(0.070)	*	0.152	(0.050)	***
$\ln HR_{REGION}$	-0.009	(0.019)		0.035	(0.027)	
$\ln HR_{LOCAL}$	-0.062	(0.012)	***	-0.021	(0.027)	
ln*POPD*	-0.051	(0.020)	**	-0.071	(0.053)	
t=2014	-0.007	(0.032)		0.083	(0.036)	**
t=2015	-0.057	(0.035)	*	-0.061	(0.039)	
t=2016	-0.043	(0.030)		-0.058	(0.033)	*
t=2017	0.080	(0.034)	**	0.073	(0.032)	**
t=2018	0.148	(0.040)	***	0.494	(0.045)	***
t=2019	0.118	(0.053)	**	0.138	(0.079)	*
t=2020	0.100	(0.044)	**	0.195	(0.102)	*
constant	0.400	(0.093)	***	0.208	(0.088)	**
Obs.	1,374					
Wald stat	.					
No. of groups	255					
No. of instruments	212					
Arellano-Bond test:						
AR(1)	-5.70	***				
AR(2)	0.23					
Hansen overidentifying restriction test:						
Chi2	163.36					
p-value	0.861					
Structural test (Ho: All coef. are the same):						
Chi2	134.06	***				

Note: The dependent variable is ln*N*. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

47. Another interesting structural test may be carried out to investigate potential heterogeneity across countries. Recall that our data comprises five countries. Obviously, there may be systematic differences in agglomeration process among the countries. From the growth point of

view, it is of particular interest to examine the potential differences between rapidly growing economies and other countries. To this end, the sample data is divided into two groups: The first comprises Azerbaijan, Kazakhstan and Uzbekistan where the numbers of enterprises have grown relatively significantly in recent years (see **Figure 12** above). The second group is composed of Georgia and Kyrgyzstan.

48. The estimated results are shown in **Table 7**. It is found that the transport connectivity impact is only significant in the growing economies: Improved transport connectivity facilitates firms' agglomeration. The estimated effect is significant at -0.123. In other countries where growth has been relatively modest, the estimated transport impact is insignificant. This implies that that in these low-growth countries, firms may not effectively be connected with each other. The local transport network may remain to be improved to fully exploit agglomeration economies. By contrast, the growing economies seem to take advantage of agglomeration economies through enhancing firms' accessibility to local markets. The evidence is consistent with the view that urbanization is the engine of growth (e.g., Bertinelli and Black, 2004).

Table 7. Separated GMM estimation by country group

	Growing economies			Other countries		
	Coef.	Std.Err.		Coef.	Std.Err.	
$\ln N_{t-1}$	0.622	(0.066)	***	1.679	(0.092)	***
$\ln N_{t-2}$	0.172	(0.049)	***	-0.609	(0.108)	***
$\ln N_{t-3}$	0.169	(0.047)	***	-0.071	(0.015)	***
$\ln HR_{REGION}$	0.010	(0.023)		-0.002	(0.002)	
$\ln HR_{LOCAL}$	-0.123	(0.022)	***	-0.001	(0.003)	
$\ln POPD$	-0.079	(0.018)	***	-0.002	(0.002)	
t=2014	0.074	(0.032)	**	-0.008	(0.010)	
t=2015	-0.048	(0.035)		-0.026	(0.010)	***
t=2016	-0.043	(0.028)		-0.048	(0.007)	***
t=2017	0.086	(0.029)	***	0.010	(0.010)	
t=2018	0.424	(0.041)	***	-0.015	(0.003)	***
t=2019	0.344	(0.072)	***	-0.013	(0.002)	***
t=2020	0.329	(0.061)	***	-0.021	(0.004)	***
constant	0.475	(0.079)	***	0.046	(0.012)	***
Obs.	1,374					
Wald stat	.					
No. of groups	255					
No. of instruments	209					
Arellano-Bond test:						
AR(1)	-5.71	***				
AR(2)	0.02					
Hansen overidentifying restriction test:						
Chi2	133.18					
p-value	0.997					
Structural test (Ho: All coef. are the same):						
Chi2	816.7	***				

Note: The dependent variable is ln*N*. Robust standard errors are shown in parentheses. *, ** and *** indicate statistical significance at the 10, 5 and 1 percent level, respectively.

VI. Conclusion

49. In this paper, the impact of transport connectivity on agglomeration economies was reexamined with unique data from the Caucasus and Central Asian countries. Transport connectivity is considered to be an important policy variable to facilitate agglomeration economies, which are one of the most important determinants of firm productivity, urban formation and economic growth. However, the potential impacts of transport connectivity remain more complex than we believe, especially when long-term causality is considered. One of the most important empirical challenges is potential endogeneity between economic outcomes and

infrastructure placement. In addition, there are few timeseries data measuring transport accessibility.

50. The paper particularly contributed to two issues. First, the paper applied spatial data and techniques to generate detailed georeferenced connectivity data in the region, using micro shipping data over 10 years. As shown, this is a useful approach to measure historical changes in transport accessibility at the highly disaggregated level. The same approach can be used everywhere as long as shipping data are available. Second, using such granular connectivity data, a dynamic panel regression model was applied to estimate the unbiased impact of transport connectivity on firm agglomeration. This is an alternative approach to the IV technique which is used in recent empirical studies investigating various impacts of transport infrastructure (e.g., Datta, 2012, Donaldson, 2018).

51. For the Caucasus and Central Asian countries, regional integration and transport connectivity are of vital importance. The region experienced significant economic transition during the 1990s. Since the early 2010s, a number of large transport investments have been made to reestablish regional connectivity. The paper cast light on such a dynamic period of time in the region and examined how seemingly improved transport connectivity affected the spatial distribution of firms, i.e., agglomeration economies.

52. It is found that the system GMM regression could control for the endogeneity problem to generate consistent estimation results. The systematically generated lagged differences and levels of the endogenous variables were found to be valid instruments. As expected, it is shown that agglomeration economies are significant in the region. Firms tend to be located where other firms are also located. The agglomeration process was also found to be highly persistent. Thus, the existing firm clusters are likely to continue growing, attracting more new enterprises. The evidence is consistent with the economic geography literature. However, it was also found that populated areas in the region suffer from congestion diseconomies. Holding everything else constant, firms are less likely to choose populated areas. This indicates that large cities are faced with a certain growth constraint: While firms tend to flock together, city growth is still not self-sustained to not only consume firms' outputs but also supply quality labor.

53. To facilitate agglomeration economies, the paper confirmed that local transport connectivity is important. This is a robust result regardless of specification. The potential endogeneity problem of infrastructure investment was controlled by the GMM approach. It means that it is important for firms to secure efficient or quick access to local markets. Transport costs may matter but look less important than travel time. On the other hand, the expected effect of regional connectivity is not statistically significant. There is no evidence yet that improved regional connectivity could facilitate firms' agglomeration. Perhaps more time may be needed to observe the likely real impact of connectivity. To take advantage of agglomeration economies at the regional level, further efforts may be needed, for instance, toward increasing efficiency in transportation and logistics and/or reducing the time and costs of border crossing, which add to overall transport costs and times. The estimated impact of connectivity is found to be heterogeneous. It is significant only in relatively populated areas, and in growing economies. Thus, local market connectivity is a key policy instrument to manage agglomeration economies and stimulate urban growth.

References

ADB. 2014. Central Asia Regional Economic Cooperation Corridor Performance Measurement and Monitoring: A Forward-Looking Retrospective. Asian Development Bank.

ADB. 2019. Good Jobs for Inclusive Growth in Central Asia and the South Caucasus: Regional Report. Asian Development Bank.

ADB. 2020. CAREC Corridor Performance Measurement and Monitoring Annual Report 2019. Asian Development Bank.

Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, Vol. 58, pp. 277-297.

Banerjee, Abhijit., Esther Duflo, and Nancy Qian. 2020. On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, Vol. 145, Article 102442.

Belderbos, René, and Martin Carree. 2002. The location of Japanese investments in China: Agglomeration effects, Keiretsu, and firm heterogeneity. *Journal of The Japanese and International Economies*, Vol. 16(3), pp. 194-211.

Bertinelli, Luisito, and Duncan Black. 2004. Urbanization and growth. *Journal of Urban Economics*, Vol. 56(1), pp. 80-96.

Boudier-Bensebaa, Fabienne. 2005. Agglomeration economies and location choice: Foreign direct investment in Hungary. *Economics of Transition*, Vol. 13(4), pp. 605-628.

Cheong, Inkyo, and Valijon Turakulov. 2022. How Central Asia to escape from trade isolation?: Policy targeted scenarios by CGE modelling. *The World Economy*, Vol. 45(8), pp. 2622-2648.

Cieślik, Andrzej, and Michael Ryan. 2004. Explaining Japanese direct investment flows into an enlarged Europe: A comparison of gravity and economic potential approaches. *Journal of the Japanese and International Economies*, Vol. 8(1), pp. 12-37.

Craig, Steven, Janet Kohlhas, and Adam Perdue. 2016. Empirical polycentricity: The complex relationship between employment centers. *Journal of Regional Science*, Vol. 56(1)m oo, 25-52.

Datta, Saugato. 2012. The impact of improved highways on Indian firms. *Journal of Development Economics*, Vol. 99(1), pp. 46-57.

Deichmann, Uwe, Kai Kaiser, Somik Lall, and Zmarak Shalizi. 2005. Agglomeration, transport, and regional development in Indonesia. Policy Research Working Paper No. 3477. Washington DC: The World Bank.

Dercon, Stefan, Daniel Gilligan, John Hoddinott and Tassew Woldehanna. 2009. The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages, *American Journal of Agricultural Economics*, Vol. 91(4), pp. 1007-1021.

De Silva, Dakshina, and Robert McComb. 2012. Geographic concentration and high tech firm survival. *Regional Science and Urban Economics*, Vol. 42(2), pp. 691-701.

Donaldson, Dave. 2018. Railroads and the Raj: The economic impact of transportation infrastructure. *American Economic Review*, Vol. 108(4-5), pp. 899-934.

Dorosh, Wang, You, and Schmidt. 2012. Road connectivity, population, and crop production in Sub-Saharan Africa. *Agricultural Economics*, Vol. 43, pp. 89-103.

Duranton, Gilles, and Anthony Venables. 2018. Place-based policies for development. World Bank Policy Research Working Paper 8410.

Ferraz, Joao Carlos, and Luciano Coutinho. 2019. Investment policies, development finance and economic transformation: Lessons from BNDES. *Structural Change and Economic Dynamics*, Vol. 48, pp. 86-102.

Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. The Spatial Economy. MIT Press.

Giuliano, Genevieve, Sanggyun Kang, and Quan Yuan. 2019. Agglomeration economies and evolving urban form. *The Annals of Regional Science*, Vol. 63, pp. 377-398.

Glaeser, Edward, Jed Kolko, and Albert Saiz. 2001. Consumer city. *Journal of Economic Geography*, Vol. 1(1), pp. 27-50.

Henderson, J.V. 1974. The sizes and types of cities. *The American Economic Review*, Vol. 64(4), pp. 640-656.

IMF. 2020. Republic of Kazakhstan: Selected Issues. International Monetary Fund Country Report No. 20/38.

Incaltarau, Cristian, Ilkhom Sharipov, Gabriela Carmen Pascariu and Teodor Lucian Moga. 2022. Growth and convergence in Eastern Partnership and Central Asian countries since the dissolution of the USSR – Embarking on different development paths? *Development Policy Review*, Vol. 40(1), e12547.

Jedwab, R. and A. Moradi. 2012. "Colonial Investments and Long-Term Development in Africa: Evidence from Ghanaian Railroads," unpublished paper, George Washington University; STICERD, London School of Economics; and University of Sussex.

Jiang, Xiushan, Xiang He, Lei Zhang, Huanhuan Qin, and Fengru Shao. 2017. Multimodal transportation infrastructure investment and regional economic development: A structural equation modeling empirical analysis in China from 1986 to 2011. *Transport Policy*, Vol. 2017, pp. 43-52.

Khandker, Shahidur, Zaid Bakht, and Gayatri Koolwal. 2009. The poverty impact of rural roads: Evidence from Bangladesh. *Economic Development and Cultural Change*, Vol. 57(4), pp. 685-722.

Konishi, Hideo. 1996. Voting with ballots and feet: Existence of equilibrium in a local public good economy. *Journal of Economic Theory*, Vol. 68(2), pp. 480-509.

Krugman, Paul. 1991. Increasing returns and economic geography. *Journal of Political Economy*, Vol. 99(3), pp. 483-499.

Labra, Romilio, Celia Torrecillas. 2018. Estimating dynamic panel data: A practical approach to perform long panels. Revista Colombiana de Estadistica, Vol. 41(1), pp. 31-52.

Lee, Ki-Dong, Seok-Joon Hwang, and Min-hwan Lee. 2012. Agglomeration economies and location choice of Korean manufactures within the United States. *Applied Economics*, Vol. 44, pp. 189-200.

Li, Xiaolong, Zongfa Wu, and Xingchen Zhao. 2020. Economic effect and its disparity of high speed rail in China: A study of mechanism based on synthesis control method. *Transport Policy*, Vol. 99, pp. 262-274.

Mare, David, and Daniel Graham. 2013. Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics*, Vol. 75(C), pp. 44-56.

Milner, Chris, Geoff Reed, and Pawin Talerngsri. 2006. Vertical linkages and agglomeration effects in Japanese FDI in Thailand. *Journal of The Japanese and International Economies*, Vol. 20(2), pp. 193-208.

Muroishi, Madoka, and Akira Yakita. 2021. Agglomeration economies, congestion diseconomies, and fertility dynamics in a two-region economy. *Letters in Spatial and Resource Sciences*, Vol. 14, pp. 51-63.

Nickell, Stephen. 1981. Biases in dynamic models with fixed effects. *Econometrica*, Vol. 49(6), pp. 1417-1426.

Palivos, Theodore, and Ping Wang. 1996. Spatial agglomeration and endogenous growth, *Regional Science and Urban Economics*, Vol. 26(6), pp. 645-669.

Pomfret, Richard, 2010. Constructing market-based economies in Central Asia: A natural experiment? *European Journal of Comparative Economics*, Vol. 7(2), pp. 449-467.

Pomfret, Richard, 2021. Central Asian economies: Thirty years after dissolution of the Soviet Union. *Comparative Economic Studies*, Vol. 63(4), pp. 537-556.

Procher, Vivien. 2011. Agglomeration effects and the location of FDI: Evidence from French first-time movers. *Annals of Regional Science*, Vol. 46, pp. 295-312.

Roodman, David. 2009. Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, Vol. 71(1), 135-158.

Sedgley, Norman, and Bruce Elmslie. 2001. Agglomeration and congestion in the economics of ideas and technological change. *The American Journal of Economics and Sociology*, Vol. 60(1), pp. 101-121.

UNCTAD. 2021. "Review of Maritime Transport 2021." United Nations Conference on Trade and Development.

UNECE. 2019. "Logistics and transport competitiveness in Kazakhstan." United Nations Economic Commission for Europe.

Valila, Timo. 2020. Infrastructure and growth: A survey of macro-econometric research. *Structural Change and Economic Dynamics*, Vol. 53, pp. 39-49.