The size distribution of cities: Evidence from the lab.

Rocco Rante[1] Federico Trionfetti[2] Priyam Verma[2]

[1] The Louvre Museum, France.

[2] Aix Marseille Univ, CNRS, AMSE, Marseille, France.

June 20, 2023

Abstract

In this paper, we bring fresh evidence on the city size distribution from a lab represented by the Bukhara region from 3^{rd} B.C. to the 9^{th}. This region was homogeneous in all respects (technology, amenities, climate, culture, language, religion, etc.). Yet, cities had different size. We rationalize the city size distribution of this economy with only two elements: spatial centrality and cost of traveling. We embed these two elements in a discrete choice model of location. We estimate model parameters in the data using method of moments. Further statistical tests show that while city locations and number are not distinguishable from a random draw, population is larger in spatially central location. The silk road is the only element that perturbed the otherwise homogeneous space. The silk road crossed a number of cities but only those that were stopover places for merchant's caravan have an abnormally larger population. The city size distribution passes the test of log normality and rank size relation is concave as predicted by stochastic growth theories though clearly the Bukhara region economy did not have any of the mechanisms of stochastic growth models.

J.E.L. Classification: R12, R13, F1.

Keywords: Spatial Model, Archaeological data, Centrality.

[0]Acknowledgments. This paper has received funds from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference :ANR-17-EURE-0020) and from the Excellence Initiative of Aix-Marseille University - A*MIDEX. The above acknowledgment concerns F. Trionfetti and P. Verma. These two authors also wish to thank the following students who, in the context of a training internship, have contributed to the paper: Marie Bruguet, Maha Chlouh, Hajar Chlouh, and Larua Daniala Contreras Portela. Naturally, the authors are the sole responsible for the paper.

1 Introduction

A fundamental fact about people is that they travel between places. A fundamental fact about places is that some are more central than others. These two facts give centrally located places an indisputable location advantage. Can this elementary structure made of places and traveling people explain the size distribution of cities? To answer this question we begin by asking which city size distribution should a model based on these two elements be able to replicate. Ideally, one would like to observe a city size distribution as it emerges in the absence of first and second nature characteristics. Specifically, one would like to observe a distribution emerging from an economy without location fundamentals, on a perfectly homogeneous geography, in the absence of infrastructures, with constant and identical technology in every location, and in the very long run. In this ideal lab, one would observe city size formation subject only to the two unavoidable facts on Earth: some cities are more central than others and centrality is an advantage because traveling is costly. In this paper, we use a unique data set that comes very close to this ideal situation. The data set is the result of thirteen years of archaeological exploration and contains city size and location for the universe of cities in the region of Bukhara in the 9$^{\text{th}}$ century A.D.[1] As we shall discuss in detail below, the morphological homogeneity of the region approximates very well the absence of location fundamentals. Cities have developed for twelve centuries from 3$^{\text{rd}}$ BC to 9$^{\text{th}}$ AD without sizable perturbation and in a situation of relative isolation from the rest of the world. This smooth passing of time is a very good approximation of an unperturbed environment where the determinants of city size have had the time to shape the observed distribution. The technology was constant over these twelve centuries and homogeneous across the entire area. This is a good approximation of the absence of technological change and endogenous productivity heterogeneity due, for instance to localized spillovers. Transport infrastructure where unnecessary because of the flatness of the land and the absence of natural obstacle to travel between any two points. The Silk Road was the only factor that affected some cities and not others. And yet, cities had a different sizes. These features make of the region of Bukhara in those twelve century an ideal lab to study the effect

[1]The archaeological mission began in 2009 under the direction of Rocco Rante (co-author of this paper) and continues to date. The mission is under the aegis of the Louvre Museum, in collaboration with the French Ministry of Foreign Affairs, the Archaeological Institute of Samarkand, and the Uzbek Academy of Sciences.

of centrality and travel costs on the city size distribution.

We embed centrality and travel cost in an elementary general equilibrium spatial model. In spite of its simplicity, the model explains almost perfectly the size and location of cities. The only exogenous factor, the Silk Road, helps explain what is not explained by centrality. This deterministic explanation of the city-size distribution accounts for the largest part of the variation in the data. This elementary model structure is also able to explain the evolution of aggregation of cities over twelve centuries from 3$^{\text{rd}}$ B.C. to 9$^{\text{th}}$ A.D. Lastly, we find evidence of persistence in city locations between the 9$^{\text{th}}$ and the 21$^{\text{st}}$ century. The persistence is revealed by the fact 21$^{\text{st}}$ cities are more likely to be concentrated around the location of 9$^{\text{th}}$ century cities of Bukhara than a random draw would entail.

These results provide an entirely new perspective on current theories of the city size distribution. To begin with, the city size distribution in the oasis of Bukhara passes the test of log-normality. Log-normality is the key prediction of the random growth model of Eeckhout (2004) which hinges on the presence (but not documented) local technological spillovers giving rise to growth. The economy of the Bukhara oasis clearly does not possesses such spillovers and yet the city size distribution passes the log-normality test. To further explore the relationship between centrality and log-normality we have simulated one thousand city size distributions consistent with the estimated model parameters but based on random distribution of points in space. The resulting city size distributions passes the test of log normality about 79% of the times. Thus, centrality provides an explanation for log-normality alternative to technological spillovers. Geographical centrality, however, does not necessarily give rise to a log-normal distribution. We show by simulations that any given geographical distribution of sites in the region under scrutiny may give rise to a variety of city size distributions depending on model parameters. The second new perspective concerns what is probably the most solid empirical regularity found in the data for modern economies: the concavity of the log rank log size relationship. Though concavity is ubiquitous, the only model in the literature consistent with such concavity is Duranton (2007). The data of the Bukhara region satisfies the concavity predicted by that model though clearly does not possess any of the stochastic growth processes of that model. The third new perspective concerns the explanation of the city size distribution based on location fundamentals. A location fundamental is usually is an intrinsic characteristic of a place usually referred to as a first nature characteristic (particularly fertile land, natural harbor, amenities, etc.). The city size distribution resulting only from location fundamentals would be a map from intrinsic characteristics to city size. As noted by Krugman (1996), the explanation of city size distribution based on location fundamentals is not to be excluded but, for it not to be tautological, one needs a precise measurement and an accurate definition of location fundamentals. This is precisely the new interpretation of fundamentals we propose: in our interpretation the centrality of a place is one of location fundamentals. This simple but new perspective help make sense of the fact that cities have different size even in seemingly homogeneous areas of the world such as the great plains of the U.S. or Russia. If we do not consider centrality, given that all other individual characteristics are homogeneous (say, fertility of the land, amenities, etc.) one would expect to observe all cities to have the same size. Centrality as a fundamental sheds also new light on the persistence found in Davis and Weinstein (2002). They find that that Hiroshima and Nagasaki regained their rank in the city size distribution less than two decades after the nuclear bombing. This may not be so because of their intrinsic (and unidentified) characteristics but rather because nuclear bombing did not change their centrality.

2 Related literature.

Our explanation for the city size distribution is based on the unavoidable heterogeneous centrality that space generates. This link between space and city size distribution has so far remained unexplored in the literature. The city size distribution literature focused for a long time on Zipf's law, a linear relationship between the log of rank and the log of size (measured by population) whose slope is -1. Numerous papers have investigated the extent to which such empirical relationship is found in the data; see e.g., Zipf (1949), Rosen and Resnick (1980), Krugman (1996), Eaton and Eckstein (1997), Gabaix and Ioannides (2004), Soo (2005), Rozenfeld et al. (2011), Ioannides and Skouras (2013) and in a historical perspective Davis and Weinstein (2002) and Barjamovic et al. (2019); a recent and comprehensive technical review is in Arshad et al. (2018). Results are mixed and very sensitive to the definition of 'city' but since the earliest papers found empirical support for the Zipf's law when checked on U.S. metropolitan areas the Zipf's law became the empirical regularity to be matched by theory. A theoretical breakthrough on this front comes with Gabaix (1999) who formalizes the Zipf's law through an economically micro-founded stochastic process à la Gibrat (1931) augmented with a lower bound. Such a process gives rise to a power law distribution whose exponent approaches asymptotically -1 from below as the lower bound approaches zero. The condition needed for an asymptotic slope of -1 is a 'small grain of send' (Gabaix, 1999) that prevents small cities from becoming too small.[2] Gibrat's proportional growth is the hypothesized engine of city growth and Ioannides and Overman (2003) take the original route of testing it directly using U.S. decennial censuses. This original approach provides insight on the core mechanism giving rise to the Zipf's law. They find that despite variation in growth rates as a function of city size, Gibrat's Law does hold overall. This evidence raises another concern, however: for Gibrat's law to give rise to Zipf's law, the former should be violated upward by small cities (the grain of sand mentioned above). This aspect remains unexplored in the literature. New empirical evidence is provided by Eeckhout (2004) who shows that the city size distribution for all U.S. cities (Census 2000) does not satisfy the Zipf's law and looks instead as a log-normal. He provides an economic model that gives rise to Gibrat's proportional growth (without lower bound) that, in turn, gives rise asymptotically to a log-normal distribution of city size. He also shows that Zipf's law for large cities and log-normality for all cities may be compatible results because an appropriately left-truncated log-normal distribution may give rise to Zipf's law for the remaining cities. A number of papers have explored the adherence of data to log-normality focusing almost exclusively on the U.S. see, e.g., Parr and Suzuki (1973), Levy (2009), Eeckhout (2009), and Ioannides and Skouras (2013), Lee and Li (2013), Schluter and Trede (2019) who also examine the size distribution in Germany, Schluter (2021), and Dobkins and Ioannides (2001) last section. As discussed above, however, the city size distribution explained by centrality passes the log normality test 79% of the times.

Later literature has departed from random growth to explore other mechanisms but none of them contains space. Behrens et al. (2014) develop a model that combines agglomeration economies, sorting of more talented people in larger cities, and selection of more productive firms in large markets.

[2]Technically, this is archived by assuming that the growth process is represented by a reflected Brownian motion that assign to small cities the largest of all shocks occurred to other cities whenever the shock they would otherwise receive would make them disappear.

Their model matches a number of stylized facts including Zipf's law for large cities. Davis and Dingel (2019) build a model with complementary between individual ability and learning opportunities and they are the engine that leads to heterogeneous city size. Their model is able to predict the positive correlation between skill premia and city size, the constant expenditure share on housing across cities, and the Zipf' law but the concavity remains unexplored. Other papers, e.g. Eeckhout et al. (2014) and Davis and Dingel (2020), provide explanations for the distribution of skills, occupations, and industries across cities but do not target directly the city-size distribution. This literature proposes rich economic models of city size distribution. In these models, cities grow at different rates for various economic reasons but their geographical position is totally irrelevant.

The literature on economic geography is instead all about space but for a long time has been unable to provide explanations for the city size distribution. A famous quote from Krugman (1996) goes that 'we are in the frustrating position of having a striking empirical regularity [the Zipf's law] with no good theory to account for it'. Over a decade later Duranton (2008) lamented 'the often uneasy coexistence between urban systems and the new economic geography.' Quantitative spatial models developed in recent times are in principle able to predict the size of cities but they have never been used for this purpose. As an example, Allen and Arkolakis (2014) motivate their seminal paper by the need to 'build a framework suitable to estimate the fraction of spatial inequality that is due to geography' but do not mention the need to explain the size distribution of the city. Previous papers, e.g., Redding and Sturm (2008), Brülhart et al. (2012) had developed simpler versions of quantitative spatial models able to explain important empirical facts but left the issue of city size distribution completely unexplored. Quantitative spatial models have progressed enormously since then, see e.g., Redding (2016), Redding and Rossi-Hansberg (2017), Desmet et al. (2018), Caliendo et al. (2019), Behrens and Murata (2021), Redding et al. (2022) and a thorough review in Redding (2022) but have not been applied to the study of city size distributions, neither theoretically nor empirically.

The models reviewed above relay on a number of key elements that give rise to heterogeneous city size or, in spatial models, to heterogeneity of economic activity in space. Among these elements there are local externalities (Gabaix, 1999; Eeckhout, 2004), random innovation and large number of industries (Duranton, 2007), human capital accumulation (Eaton and Eckstein, 1997), exogenous and endogenous productivity differences (Allen and Arkolakis, 2014; Desmet et al., 2018; Caliendo et al., 2019), and size-skill complementarity (Davis and Dingel, 2019) to mention only the most common of a long list reviewed and thoroughly discussed in Redding and Rossi-Hansberg (2017). None of these elements are present in the 9$^{\text{th}}$ century economy we observe. Arguably, none of them were present in the six millennia of antiquity (think of Mesopotamia), and yet cities had different sizes even then. We, therefore, focus on centrality and travel costs since they are present in every economy, modern and ancient, and are the only element present in our 9$^{\text{th}}$ century lab.

3 The Oasis of Bukhara: An ideal lab

We follow the archaeological tradition and use the term 'oasis' in the sense of a naturally delimited and homogeneously irrigated land. The oasis of Bukhara refers to the geographic area that extends over the delta of the Zerafshan in central Uzbekistan. The delta irrigates with its shallow waters a surface of land whose area measures about 5,100 square kilometers (1,969 sq mi).[3] The surface is extremely flat, the difference between the highest and lowest point in the oasis is of two hundred meters and the land slopes downwards monotonically over about seventy km from North-East to South-West. The oasis is surrounded by a desert made of clay. The time span of our study goes from the 3$^{\text{th}}$ century B.C. to the period of maximum expansions of each city, which varies by city but lies between the 9$^{\text{th}}$ and 10$^{\text{th}}$ century A.D. Bukhara is the only exception. This city was among the three largest cities until the end of the 8$^{\text{th}}$ century. Its size exploded when it became the capital of the Islamic state that extended from the Oxus river to Ferghana near the border with modern Kyrgyzstan. Today, it is the capital of the homonymous administrative region of Uzbekistan which partly overlays with the oasis. To simplify the prose we refer to the upper bound of our time span as to the 9$^{\text{th}}$ century for all cities.

Overview. The oasis characteristics make of it an ideal lab for the study of the size and geographical distribution of cities. Its geographical limits are exogenously determined by the delta of the Zerafshan, its political limits coincided with its geographical limits, and the oasis remained independent

[3] Approximately the size of the French département of Bouches du Rhône (5,087 sq km) and a bit larger than the state of Rhode Island in the United States (1,545 sq mi).

from foreign powers for the twelve centuries covered by our study.[4] There was rivalry between major cities and even a war between two cities but no city has ever had supremacy over the oasis in the time span relevant to our study. These features eliminate possible effects on city size due to political organization or supremacy. The oasis was extremely homogeneous in terms of climate, water availability, technology, natural resources, culture, religion, and language. This homogeneity is ideal for our study because allows ruling out the existence of heterogeneous location fundamentals. The economy was elementary. It produced agricultural goods and basic manufactures. Due to identical land and climate conditions the agricultural produces were the same throughout the oasis. Manufacturing production consisted of ceramic and metal object produced by use of labor, iron, clay, and water. Clay and water were essentially free goods given their vast and homogeneous availability in the oasis. Iron, was absent in the oasis and was imported from the mountain chain that extends eastbound from the North-East end of the oasis. The transport technology was elementary (horse, donkey, camels, feet) and highly costly. The typical farmer would travel by feet to the nearest market place in company of his donkey. Given the morphology of the land, roads were essentially straight lines between two points. Water was shallow, omnipresent, and bridges and canals were easily built. No reasonable location fundamentals can be imagined in this situation. High and homogeneous unit cost of travel in any direction allow ruling out irregularity of the land as a factor influencing the city size distribution. The similar size of ovens found in different cities testifies of an identical technology throughout the oasis and throughout the twelve century under scrutiny. The Silk Road came to exist and develop at the same time as human occupation of the area. As such, it does not represent a shock to a preexisting economic equilibrium. Silk merchants used the oasis cities principally as stop-over or as markets where they would exchange merchandises with other merchants. Exchange, of course, took place with locals as well. Merchants would demanded accommodation, market services (such as arranged space for trade with other merchants),

[4]The oasis was situated between two political superpowers: the successive empires that extended in the area of modern Iran and the Chinese empire. With the tacit consent of the superpower it enjoyed political independence for many centuries. In the 7th century, a political elite of foreign origin was imposed on the oasis and replaced the previous rulers. This replacement touched only the political rulers, however, leaving the economic, religious, and social structure unchanged. Table **??** in the appendix provides a timeline of major events.

food for them and for the livestock. Historic sources give us a sense of how large this additional demand was.[5] Caravans often counted dozens of people and hundreds of livestock, and more than a handful of caravans might stop in a city simultaneously.

All these features make the Oasis an ideal lab for our purposes. We can rule out local and global externalities, land irregularity, exogenous amenities, exogenous technological differences, multiple sectors and serendipidy. In no other part of the world we find such conditions met for so long time. In this world, stochastic growth models would predict cities to have all the same size. And yet, even in this world, cities did have different size. The silk road is the only external element that possibly came to influence the size of stop-over cities.

Data. In the ninth century, the oasis contained 618 sites identified by the presence of mounds, or *tepe*, resulting from the overlapping of various layers of human occupations over centuries. Sites have a specific urban structures that allow grouping them in the following categories: *manufacturing* cities, *agricultural* cities, *hamlets*, and *forts*. A manufacturing city consists of a *town* and a *business district*. The town hosts the political center, which often consists of a fortified building, and the village which hosts the population and in some cases administrative buildings. The business district hosts the commercial and manufacturing activities. Only in some cities the business district is equipped with dedicated areas for caravan stopovers. We have found evidence that the town of five manufacturing cities were endiskd by protective walls. Not surprisingly these were the cities on the border of the oasis. By the analysis of the land slopes at the border of cities it is likely that many other manufacturing cities were walled but there is no certainty, and in any case, the walls were probably of smaller magnitude. An agricultural city is made of a town structured similarly to that of the manufacturing cities, though most of the times smaller in size. Agricultural cities do not have a business district. A hamlet is an isolated settlement where neither a town nor a business district are present. It typically consisted of a few houses, or a manor and its annexes, or an isolated large farm. Forts, consist of a very small and concentrated settlement hosting only soldiers. In addition to these settlements, in the oasis there are some unstructured sites that exhibit scattered pottery or water pits. They do not show evidence of stable human

[5] See, for instance, Étienne de la Vaissière (2018).

settlement and for this reason we neglect them in our analysis. For clarity of exposition, we reserve the term 'city' for the manufacturing and agricultural sites because of their organized urban structure. In conclusion, the universe of 618 settlements consists of 53 manufacturing cities, 284 agricultural cities, 266 hamlets, and 15 forts. Figure 1 shows the oasis, the water channels, and the position of sites. Figure 2 shows the layout of the silk road and the position of manufacturing cities. The silk road touches many manufacturing cities but only eleven among these have a dedicated stop-over place for caravans. We use three measures of site size. The first is the Resi-

Figure 1: Archeological sites and the delta of Zerafshan

dential Area (RA), which is the area of the surface occupied by dwellings and public buildings if any. For manufacturing and agricultural cities this area corresponds to the town area. For hamlets and forts it corresponds to the area occupied by dwellings. The second is the Total Area (TA), which is the area of the total surface occupied by the city. This measure applies

Figure 2: Silk roads through the oasis

to manufacturing cities only and consists of RA (as defined above) plus the area of the business district. The third is Population (POP). We observe the surfaces of towns, business districts, and hamlets with precision.[6] We do not observe population but we can estimate it fairly well by assigning standard population densities used in archaeology to the areas we have directly measured. Population densities differ by type of settlements. We assign 100 inhabitants per hectare for hamlets and forts, 150 inhabitants per hectare for agricultural cities and 200 inhabitants per hectare for manufacturing cities. The reason for different population densities is that while agricultural city

[6]The method of surface identification and measurement is accurately described in Rante and Mirzaakhmedov (2019). Here, it is important to note that we observe the universe of sites and their urban structure, except for the following four cities (city code in parenthesis): Gijduvan (0002), Bukhara (0097) and Vobkent (0116) because the modern urban structure entirely covers the ancient site, and Šargh (0846) because it is completely destroyed. We know the location of these sites but we cannot measure the area covered by the sites. Data is available in the online supplementary material.

hosted population only in the town area, manufacturing cities hosted population also in alternative accommodation (soldiers in barracks, workers in the business districts, people in transit in the market area, some dwellings outside the walls, etc.). The population densities we use are standard in archaeology for those times and areas of the world and are corroborated by our archaeological excavation where we measure the size of houses, streets, buildings, etc. As robustness checks, however, we use all three measures of size.

Oasis Delimitation. For some statistical indices we use below we need to delimit the area of the oasis. The criterion is to delimit the oasis to the surface that was inhabitable in the 9^{th} century. Figure 3 shows the geographic delimitation of the oasis on a white background. The subdivision in different colors will be explained below. A cursory comparison between Figures 3 and 1 or 2 shows the the inhabitable land then was not so different from the green area appearing in satellite photographs. There are exceptions. The first one concerns the north-east strip. We see in Fig. 2 that there is a stretch of the silk road crossing what today is a quasi-desert. That area was instead fertile and green in the 9^{th} century, this is why we include it in the oasis as inhabitable land. The area to the north of the oasis appears even today as greenish and was indeed green in the 9^{th} century but was uninhabitable as it was in fact covered by swamps. This is why we exclude it from the oasis. Lastly, the inhabitable area at the south-west extreme was a bit wider than the green area appearing in satellite photographs. We then have made this area slightly larger to match what the inhabitable area was to the best of our knowledge.

Urban systems. Sites were linked by a relation of vassalage by which the agricultural cities and hamlets were under the political aegis of the nearest manufacturing city. It is interesting to note that this political interdependence was coherent with the transport and food-preservation technology of the times. Because consumption of agricultural produces had to take place near the place of production, the exchange of food against manufactures created an interdependence between manufacturing cities and other sites, with the latter being net exporters of food to the former. Indeed, it is suggestive to think that the trade structure favored or even caused the emergence of vassalage. We take this interdependence into account in our study by defining *Urban Systems* and using them as our unit of analysis. In section 4 we test the statistical validity of this associations in urban systems while here we provide the definition. An Urban System is constituted by a manufacturing city and its associated agricultural cities, hamlets, and forts (if any). The association criterion is based on least distance by which each non manufacturing site is associated to the nearest manufacturing city. The logic of such association is that farmers travel to the nearest business district to sell their produces. Fig. 3 shows the aggregation of sites in urban systems. The black dots represent the fifty-three manufacturing cities. The white dots represent all other sites. We also have created the geographical extent of the urban system and assigned to each a different color. The geographical extension is constructed first by dividing the oasis in cells of 1 arc-second by 1 arc-second and then associating each cell to the nearest manufacturing city.

Outliers. One question we have to deal with is whether to include in our calculations the two manufacturing cities situated, respectively, at the north-east and south-west extremes of the oasis (resp. site codes 104 and 95). The most famous contemporary historian, Al-Narshakhi (943), includes these two 'cities in the culture and society of the oasis' and refers to them as the oasis gateway. However, geographically they are clearly two outliers because they are extremely isolated. The nearest site to site 95 is at 10.94 kilometer and the nearest to site 104 is at 5.91 kilometers against an average of 1.40 kilometers for the other sites of the oasis. Given the transport technology of the time, the existence of these two isolated cities is probably due to the need for regularly positioned stop-over places along the silk road. The distance a caravan could travel in a day is of thirty to thirty five kilometers, which is approximately the distance that separates site 104 from its nearest manufacturing city. To check the robustness of our results we perform the empirical analysis first including and then excluding the gateway cities and their associated sites. We use the term *economic* oasis for the data-set that excludes the gateway cities and their associated sites, we use the term *archaeological* oasis for the data-set that includes all sites. Fig. 3 represents the archaeological oasis. The economic oasis obtains by suppressing the extreme south-west and extreme north-east urban systems. Summary statistics for urban systems are provided in Table 1. The second column (Obs.) help us distinguishing between the archaeological and the economic oasis. The number of observations is, respectively, 53 and 51. Additional summary statistics are provide in the appendix.

Figure 3: Urban Systems

Table 1: Summary statistics

Urban Systems

Size	Obs.	Mean	Std. Dev.	Min	Max
Residential Area (ha)	53	21.73	13.13	3.68	59.38
	51	21.85	13.36	3.68	59.38
Total Area (ha)	53	26.89	15.70	5.20	68.77
	51	26.88	15.96	5.20	68.67
Population	53	3398	2196	663	10108
	51	3401	2231	663	10108

Population is rounded up to the nearest integer.

4 Aggregation in Urban Systems.

If the aggregation based on least-distance is a good criterion we should find it reflected in the geographical distribution of sites in the oasis. In this section we test statistical the validity of such aggregation.

The need for aggregation. If we want to predict the city size distribution, the first thing to do is to define what a 'city' is. Some form of aggregation of urban areas is clearly desirable. For modern economies, for instance, we do not care to predict the size of a particular suburb or administrative subdivision of an integrated urban area; we care about predicting the size of the urban area as a single social and economic unit. Likewise, for the oasis of Bukhara it would be uninteresting to try to predict the size of any single site. In general, consider the following example where two sites, say a big site A and a small site B are located near to one another (say, New York and Newark, respectively, 8 and 0.3 millions inhabitants). Imagine we want to treat them as distinct 'cities'. Any spatial model used to predict the city size distribution would predict them to have essentially the same size because their centrality as well as other possible exogenous characteristics are essentially the same . The spatial model, by predicting the same size, would underestimate the size of A and overestimate the size of B, thus performing poorly. But this would be a bad use of the model and a bad use of the data because it fails to recognize that A and B are part of the same integrated urban area. It is the size of A+B that we want to explain and not that of A and B separately. Aggregation is therefore useful and desirable to make sense of any model and to use the data appropriately. The remaining issue is the method of aggregation. For modern economies this is not a simple task. The most popular definitions in the empirical literature for modern economies are census defined cities or metropolitan areas.[7] The difference between these two definitions is huge. According to the first definition New York is made of New York City and counts about eight millions people while according to the second definition New York is made of New York City, Long Island, the Mid and Lower Hudson Valley, Newark, Jersey City, Paterson, Elizabeth, Lakewood, Edison, Bridgeport, New Haven, Stamford, Waterbury, Norwalk, Danbury, as well as other vicinities and counts about twenty million people. Likewise, Paris *intra-muros* counts about two 2.2 million people while Grand

[7] In the U.S. Census Bureau definition, a city corresponds to the above-mentioned *place* and a metropolitan area corresponds the Metropolitan Statistical Area. Other countries use similar definitions. Other possible units of analysis are counties. This seems a suitable unit when the objective is to study the geographic distribution of economic activity without focusing on city, as in Allen and Arkolakis (2014). This measure has advantage of covering the entire population but has the weakness that a county does not correspond to a city or an urban area.

Paris counts seven million people distributed on 131 municipalities.[8] Similar differences apply, in due proportions, to smaller urban areas. The use of cities as unit of analysis has the advantage of a large coverage (about 75% of total population for the US and near to 100% for European countries) but has the drawback that the unit of analysis (the city) is defined by arbitrary administrative criteria instead of social or economic ones. Metropolitan areas are aggregation of smaller units constructed precisely with the purpose of capturing social and economic ties between the units. This is why they are a better measure than cities but have the drawback that the coverage is very small (about 30% of the population for the US). Rozenfeld et al. (2011) provide what is probably the most accurate measure of 'city' size to date. They use high-resolution remote sensing and 'build' cities by clustering urban areas according to the criterion of proximity. They define a 'city' as a maximally connected cluster of populated sites defined at high resolution. Thus, the boundaries of a city are determined by the fall in urbanization density below an arbitrary cut off threshold. We use essentially the same criterion but we benefit from one additional information that frees us from having to chose an arbitrary cut off threshold: the need for farmers to be near to the market place located in the manufacturing city. This need dictated by the transport technology allows us to determine precisely the boundaries of economic ties. We use this fact and aggregate sites into urban systems by assigning each site to the nearest manufacturing city. Our urban systems then correspond to the criterion of least distance and to the criterion of social and economic ties used to construct metropolitan statistical areas for modern economies. This aggregation needs empirical validation, however, and this is what we check in the remainder of this section.

Endogenous formation of manufacturing cities. A useful way to think about the formation of market places is to first look at archaeological evidence. Organized market places (which define a manufacturing cities) appeared only between the 4^{th} and the 6^{th} century. This fact and the absence of storage technology for food suggest a natural evolution of sites into urban systems. Let $t(\delta_{ij})$ be a function that measures the fraction of the perishable food that arrives at destination safe and sound. Of course, this measure depends on travel time δ_{ij}. Let $t'(\delta_{ij}) < 0$ and $t(\delta_{ii}) = 1$; also let $\overline{\delta}$ be such that $t(\overline{\delta}) = 0$. As the produce perishes in transit so does the income. The ex-

[8] We spare you the list.

pected utility from spending (non perished) income in market j for a resident of i is $\alpha\bar{l}t(\delta_{ij})+\mu_C\ln(\mathbb{E}[L_j^m])$ if $t(\delta_{ij})>0$ and zero if $t(\delta_{ij})=0$. Initially (say, between 3rd B.C. and 1st A.D.), when there were only a small number of sites in the oasis it is unlikely that any sub set of them were close enough so that the food possibly transported would not entirely perish. That is, initially it is likely that $\inf\{\delta_{ij}\}_{i\neq j}>\overline{\delta}$ for all i. Therefore, produce exchange, if at all, took place within every site. Given the scant population in every site there was no need for dedicated spaces or any form of organized market places. As the number of sites increased (say, between the 1st and the 4th century), accidental clusters where $\inf\{\delta_{ij}\}_{i\neq j}<\overline{\delta}$ for some i and j eventually came to exist and exploration began. The exploration process gave rise to the emergence of a single market within some or all accidental cluster simply because manufacturers, who are not tied to the land, could chose to reside in the market with better access to customers. To fix ideas, consider a process where people living in an accidental cluster chose to explore the market that gives the highest expected utility. The number of people that will travel to the market in location i within an accidental cluster that has K sites in it is

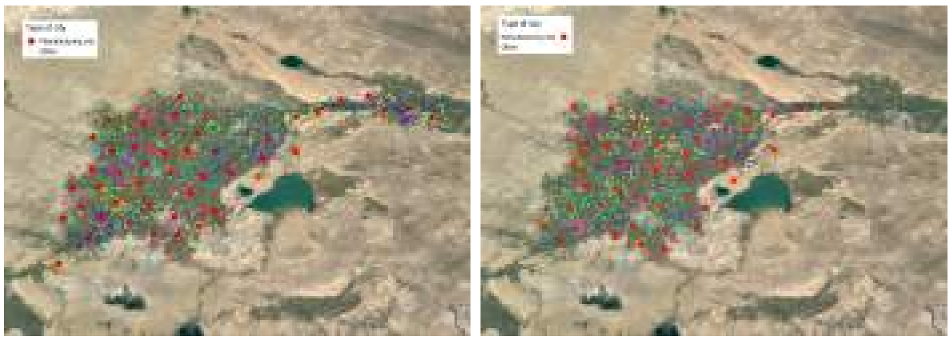
$$L_i^m=\sum_{j\neq i}^{K-1}L_jI_j \qquad \forall i, \tag{1}$$

where I_j is an indicator function such that

$$I_j=\begin{cases}1 & \text{if} \quad \alpha\bar{l}t(\delta_{ji})+\mu_C\ln(\mathbb{E}[L_i^m])=\inf_k\left\{\alpha\bar{l}t(\delta_{jk})+\mu_C\ln(\mathbb{E}[L_k^m])\right\}\\ 0 & \quad \textit{otherwise}\end{cases} \tag{2}$$

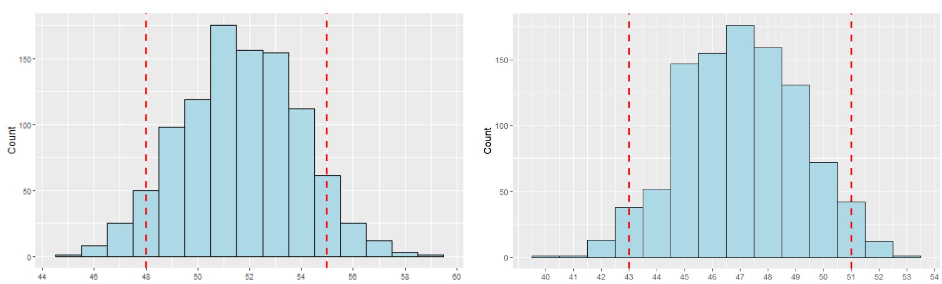
Equations (1) gives rise to heterogeneous market size even when all sites in cluster K have the same resident population or even when we solve the system (1) assuming perfect foresight $\mathbb{E}[L_k^m]=L_k^m$. Even when these conditions are met, markets will not have the same number of participants because places will not, in general, be equidistant. Heterogeneous market size makes the distribution of resident population within the accidental cluster unstable. The reason for instability is that manufacturers (who are not tied to the land) can chose where to reside. They will all chose to reside in the largest market because of largest exchange opportunities. Since manufactures do not perish we can allow for $t(\delta_{ij})=1$ for any δ_{ij} distance for manufacturers only and this would not change the results. The concentration of manufactures will wipe out all other markets since everyone needs to consume manufactures and they can only be found where manufacturers are located. At the end of the process there will be only one market place in any accidental cluster. Note that the emergence of a single market is independent from the microfoundation of the exploration decisions. As a matter of facts, one can assume a variety of exploration processes. As long as they give rise to heterogeneous market size, manufacturers will decide to reside in the largest market of the cluster and a single market will eventually emerge. Note also that a single market in a vaster area than an accidental cluster cannot emerge because perishable food would perish before reaching the market. To check that the plausibility of the formation of manufacturing cities we have just described we simulate it as follows. We generate a random allocation of 618 in the archaeological oasis. Each point may become a manufacturing city but which of them will actually become one is determined by accidental clustering of these points. For each point we compute the number of neighboring points we find within a radius of 7 km. We choose 7 km because this distance makes the consensus among archaeologists that it is the average traveling time to the market place for that historical period, that area of the world and that technology. We rank points by the number of neighbors they have. The point which has the largest number of neighbors is the first simulated manufacturing city. We the drop this point and its neighbors from the list of potential manufacturing city and we recount the number of neighbors for each of the remaining points. We identified the point with the largest number of neighbors in this reduced set of points and this will be our second simulated manufacturing city. We reiterate until we have exhausted all points. At this stage, some points will be neighbors of two manufacturing cities. We assign such points to the closest manufacturing city. After this last step we will have no left over point and no point being neighbor of more than one manufacturing cities. The set of thus created manufacturing cities and their neighbors constitute the simulated urban systems. The objective of this simulation is to verify that the number of urban system obtained from simulation is comparable to the one we observe in the data. Figure 4 shows the result of a simulation run. We repeat such simulation 100 times to obtain a simulation-based distribution of number of manufacturing cities. Fig. 5 shows the results for the archaeological and economic oasis. It is amazing how close simulations results are to the data. The mode of simulation for the archaeological oasis is 51 manufacturing cities while the data shows 53. For the economic oasis we have, respectively, 46 and 51. The data is strongly within the simulated confidence intervals.

Figure 4: Simulated number of manufacturing cities



(a) Archaeological (b) Economic

Figure 5: Simulated number of manufacturing cities



(a) Archaeological (b) Economic

Parametric test for the aggregation criterion. We use a Poisson cluster process as the statistical model appropriate to our analysis. Applied to our context, the process assumes that manufacturing cities are distributed according to a homogeneous Poisson process with intensity ρ, while all other sites are distributed according to a radially symmetric normal distribution with standard deviation σ. This process gives rise to an explicit theoretical *K-function* $K_c(r)$, where the subscript $_c$ stands for cluster process, which is

$$K_c(r) = \pi r^2 + \frac{1}{\rho}\left(1 - \exp\left(-\frac{r^2}{4\sigma^2}\right)\right). \tag{3}$$

Note that $\lim_{\sigma\to\infty} K_c(r) = \pi r^2 \; \forall r$. This is intuitive, as the variance of distribution of other sites approaches infinity, each sub-area around a manufacturing city has the same propensity to host a non manufacturing site. Hence, the distribution of these other sites approaches complete spatial randomness. We estimate the parameters of this statistical model to detect whether manufacturing cities attract other sites. We estimate the parameters by minimizing the discrepancy

$$D = \int_0^{\bar{r}} \left[(\widehat{K}_c(r))^a - (K_c(r))^a \right]^2 dr \tag{4}$$

where $\widehat{K}_c(r)$ is analogous $\widehat{K}(r)$ except that the former is computed on non-manufacturing cities only. The choice of $\bar{r}$ depends on the problem at hand. For our purposes we chose $\bar{r} = \{8, 10, 12, 14\}$ as plausible distances within which we expect fo find clustering around a manufacturing city. As per the choice of a we take the traditional $1/4$ as a benchmark. The estimated ρ and σ are 0.00034 and 0.98 and the 95% bounds are $[0.0000052, 0.00011]$ and $[7.77, 36.4]$ respectively. While ρ guides the intensity with which the manufacturing cities are formed using a Poisson process, dispersion of non-manufacturing cities around the manufacturing cities is guided by radially symmetric normal distribution with standard deviation, σ. Higher σ implies more dispersed non-manufacturing cities while smaller values indicates concentration of non-manufacturing cities around manufacturing cities. In our estimations, we find that estimated σ is smaller than what a random draw would entail (95 % CI), indicating co-concentration.

Non parametric test for the aggregation criterion. In the previous paragraph we have estimated a particular statistical model but clustering around the manufacturing cities could follow a different model unknown to us. Whatever the model, if clustering takes place, we should find non-manufacturing sites more frequently than manufacturing sites around a manufacturing site. To detect this relative concentration we use a simple distance-based index of geographic co-concentration. Since we want to measure relative concentration we use the ratio of two geographical concentration indices. To save notation we use the same capital letter to refer to a set and the number of its elements. Thus, M is the set of manufacturing cities and their number, V is the set of all other sites and their number, and N is the set and number of all sites. In our data $M = 53,\ V = 570,\ N = 618$ for the archaeological oasis while $M = 51,\ V = 560,\ N = 611$ for the economic oasis. The index of geographic concentration we use, denoted $C(r)$, is

$$C(r) = \left(\frac{\sum_{i\in M} v_i(r)}{V}\right)\left(\frac{\sum_{i\in M} n_i(r)}{N-1}\right)^{-1}. \tag{5}$$

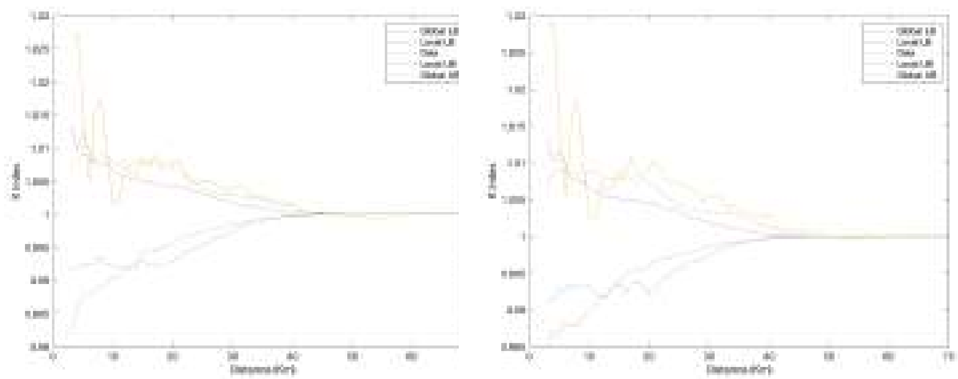
where $v_i(r)$ is the number of non manufacturing sites found within a radius r from manufacturing city i; $n_i(r)$ is instead the number of all sites found within the same radius. The maximum distance between any two sites is 131 kilometers for the archaeological oasis and and 85.7 kilometers for the economic oasis. Thus the index is computed for $r = 1, 2, ..$ to the respective maximum distance. The first term in expression (5) measures the percentage of neighbors belonging to V found within r while the second term measures that of all types. Thus, $v_i(r)/V$ is the cumulative distribution of occurrence of non-manufacturing sites from manufacturing city i in the r-disk. Analogously for $n_i(r)/(N-1)$. Dividing each of them by M gives the average cumulative distribution across i. A value $C(r) > 1$ would indicate that - on average - and within that particular radius r sites of type V are more frequently found than sites of all type. This would mean that non-manufacturing sites are more concentrated around manufacturing sites than are all sites. Vice-versa for $C(r) < 1$. Note that the values of each term for different r are necessarily not independent since V and N are constant. If $v_i(r)$ is very high for some r it necessarily has to be very low for some other r. Likewise for $n_i(r)$. This dependence across r within each term is not a concern because C_r measures relative geographic concentration. By construction, the index converges to 1 as r approaches the maximum distance. To detect the statistical significance of the values of the index obtained for the oasis we construct confidence intervals based on simulations. Since we want to establish whether manufacturing cities attract non-manufacturing sites, the simulation is constructed as follows. We create a simulated sample by drawing randomly 570 sites within the Bukhara Oasis and then randomly choosing 53 sites to get a total of 618 sites. In the economic boundary definition, we take 51 and 557 sites respectively. For every radius (r) from 1 km up to the maximum bilateral distance, we compute co-concentration indices in the sample data. We repeat the draw a hundred times and create a lower and upper local confidence bands of the indices by picking the fifth and the ninety fifth percentile for each r. We also construct a global confidence interval by keeping the percent of rejected index values across all r's equal to 5%, keeping the confidence interval range for each r fixed.

Figure 6 plot the results. The concentration of non-manufacturing sites around manufacturing cities exceeds by far what one would expect from randomness. It is very interesting to note that the peak of the distribution is found at seven kilometers for the Archaeological Oasis (Fig. 6a). For the Economic Oasis the highest and second highest local maxima are at four and seven kilometers 6b). The peaks at seven kilometers are in remarkable concordance with the standard conjecture made in archaeology that seven kilometers is the typical distance a farmer could travel in a day to go to sell produces to the market place. In conclusion, manufacturing cities tend to be surrounded by non-manufacturing sites more than by other manufacturing cities. This tells us that an urban system can be seen as an integrated set of sites we use in simulations. This result matches extremely well the prediction based on the transport technology of food: a manufacturing city attracts agricultural cities and other sites, most of them no further away than a day trip from the market place.

5 Size distribution of urban systems.

Panel A of Table 2 shows the results of the more powerful Skewness-Kurtosis and Shapiro-Wilkinson tests of log-normality. They confirm the results that the city size distribution for the Bukhara oasis passes the test of log-normality. These results are striking when compared to those of Table 2 for modern economies, where we have seen that only a few of them passed the test. Panel B of Table 2 reports the results of the KS test of log-normality. The test is performed on the three measures of size (Town Area, Total Area, Population) and the two possible units of analysis: City and Urban System. The

Figure 6: Co-concentration Index: Randomizing All Cities



(a) Archaeological (b) Economic

Notes: Figure a plots the co-concentration index for archaeological boundary which includes 53 urban systems. The index is calculated based on equation 1. The confidence intervals are created by sampling random draws of 53 cities a total of 100 times. The local confidence intervals are chosen after dropping of the observations 5% above and below for each distance. Global CI are created keeping confidence intervals for each level of distance such that we reject only 5% of the total population. Figure b is for 51 cities comprising economic oasis.

tests accept the null hypothesis of log-normality for cities (p-values larger than 0.05) with probability resp. 0.237 for town area, 0.188 for total area, 0.221 for population. The aggregation in urban systems passses the test of log normality with much higher probability. In particular, the tests run on Town Area and Population accept log-normality at 0.913 and 0.922. Panel C of Table 2 reports the results of Zipf regressions. None of the different measures of size combined with either cities or urban systems gives a slope of -1. All the slopes are significantly lower than -1 and thus, limiting the sample to larger and larger cities (catting the sample from the left) would make the slope even smaller than minus one. Furthermore, the rank-size relations for the oasis are concave *à la* Duranton (2007). This result is striking because the economy under scrutiny clearly does not have the features of Duranton (2007) economy. This contrast shows very well how interesting and useful is to use data for the ancient past. Data for modern economies matches well the predictions of Duranton (2007) but data for ancient economies does it as well. This is one reason why we revisit the theory using simpler but universal mechanisms: traveling people and geographical centrality.

6 Model

Our argument is that the geographic distribution of cities determines their size because centrality is an advantage when traveling is costly. To represent this argument we use an elementary nested multinomial logit model of location decisions in the vein of McFadden (1974). This model structure is often used in general equilibrium spatial models.

Location decisions. The indirect utility derived from being located in urban system i is

$$U_i = u_i + \epsilon_i \tag{6}$$

where u_i and ϵ_i are, respectively, the common and idiosyncratic component of utility. The latter is an *i.i.d.* random variable distributed as a double exponential (Gumbel) with mean zero, scale parameter μ_L, and variance $\mu_L^2\pi^2/6$. Let $F(x)$ and $f(x)$ denote, respectively, the cumulative and the probability distribution function. An agent will choose city i over any other city k if $u_i + \epsilon_i > u_k + \epsilon_k$ for all $k \neq i$. The probability $\mathbb{P}_i$ that an agent

Table 2: Size Distribution in the Oasis of Bukhara

B: Skwenes-Kurtosis and Shapiro-Wilkinson tests for log-normality

		Sk-Kurt			SW
Measure of 'size'	Type of unit	Sk.	Kurt	Joint	
Town Area	Urban System	0.2246	0.6155	0.4046	0.2006
Total Area	Urban System	0.3131	0.7201	0.5508	0.1624
Population	Urban System	0.6428	0.8190	0.8748	0.5597

All data for the 9^{th} century A.D. p-values for Cities near zero in all tests (not shown).

C: Kolmogorov-Smirnov test of log-normality

Measure of 'size'	Type of unit	D	p-value	№ of obs.
Town Area	Urban System	0.0769	0.913	53
Town Area	City	0.0558	0.237	342
Total Area	Urban System	0.0873	0.814	53
Total Area	City	0.0588	0.188	342
Population	Urban System	0.0757	0.922	53
Population	City	0.0567	0.221	342

All data for the 9^{th} century A.D.

A: Zipf regressions

Measure of 'size'	Type of unit	Slope	s.e.	№ of obs.
Town Area	Urban System	-1.19	.087	53
Town Area	City	-1.24	.026	342
Total Area	Urban System	-1.25	.088	53
Total Area	City	-1.15	.021	342
Population	Urban Systems	-1.25	.071	53
Population	City	-1.18	.022	342

All data for the 9^{th} century A.D.

chooses location i when there are N locations is

$$\begin{aligned}
\mathbb{P}_i &= \Pr\left[u_i + \epsilon_i = \max(u_k + \epsilon_k)\right] & (7)\\
&= \int_{-\infty}^{\infty} f(x) \prod_{k \neq i}^{N} F\left(u_i - u_k + x\right) dx & (8)\\
&= \int_{-\infty}^{\infty} \frac{1}{\mu_L} e^{-\left(\frac{x}{\mu_L}+\gamma\right)} e^{-e^{-\left(\frac{x}{\mu_L}+\gamma\right)}} \prod_{k \neq i}^{N} e^{-e^{-\left(\frac{x+u_i-u_k}{\mu_L}-\gamma\right)}} dx & (9)
\end{aligned}$$

where γ is the Euler constant. Using the change of variable $y_i = \exp(u_i/\mu_L)$, and $\delta = \exp\left[-\left(\frac{x}{\mu_L}\right) + \gamma\right]$; and then noticing that the derivative $d\delta = -(1/\mu_L)e^{-(\frac{x}{\mu_L}+\gamma)}dx \leq 0$ we obtain

$$\mathbb{P}_i = \int_0^{\infty} e^{-\delta} \prod_{k \neq j}^{N} e^{-\frac{\delta y_k}{y_i}} d\delta = \int_0^{\infty} \exp\left(\frac{-\delta \sum_{k=1}^{N} y_k}{y_i}\right) d\delta \tag{10}$$

and finally

$$\mathbb{P}_i = \frac{y_i}{\sum_{k=1}^{N} y_k} \tag{11}$$

The model has these well known properties.

$$\begin{aligned}
\lim_{\mu_L \to \infty} \mathbb{P}_i &= \frac{1}{N} & (12)\\
\lim_{\mu_L \to 0} \mathbb{P}_i &= \begin{cases} 0 \text{ if } u_i < \max\limits_{k \neq j}(u_k) \\ 1 \text{ if } u_i = \max\limits_{k \neq j}(u_k) \end{cases} & (13)
\end{aligned}$$

As μ_L (the variance) approaches infinity the idiosyncratic component of utility dominates any other aspect of the location choice and every location has the same probability of being chosen. At the other extreme, as μ_L approaches zero, the idiosyncratic component of utility has negligible impact relative to other aspects in the location decision.

Location decision is taken in a logically distinct time from the consumption and travel decision. At the time of deciding where to reside, the agent knows she will have stochastic preferences about trip destinations that will manifest themselves after the residential decision has been made. Thus, location decision is based on the expected indirect utility derived from consumption of goods plus expected indirect utility from travel from any particular location.

Expected utility from consumption. Each individual is endowed with $\bar{l}$ units of labor per unit of time. Labor is used for production and travel. Let $0 < \alpha < 1$ be the fraction of labor used for production. Then $\alpha\bar{l}$ is the real income of an individual.[9] Indirect utility from consumption equals real income plus a Gumbel distributed stochastic component with parameter μ_C analogous to the parameter μ_L introduced above. The stochastic component represents the idiosyncratic preference for a particular vendor, including oneself. The expected indirect utility from spending one's income in market i is

$$c_i = \alpha\bar{l} + \mu_C \ln\left(L_i^m\right) \tag{14}$$

where L_i^m is the number of people present in market place i.[10] L_i^m is proportional to the population of the urban system for market places that do not host silk road merchants, $L_i^m = mL_i$ with $0 < m < 1$. Exchange opportunities in market places that host silk road merchants are larger by a factor $s_i = (L_i + S_i)/L_i$, where S_i is the population of merchants stopping over. Indirect utility is increasing in L_i^m because of a wider variety of vendors (or goods) in the market.[11]

Expected utility from traveling across urban systems. There are N trip destination one can chose from if located in urban system i (including i itself). Indirect utility of a trip from i to j is $\tau_{ij} + \zeta_j$. The first addendum is the part of utility common to all agents and second addendum is the *i.i.d.* Gumbel distributed idiosyncratic destination preference with scale parameter μ_Z. The deterministic component of the indirect utility of a trip from i to j (denoted τ_{ij}) is given by the time spent in j when leaving from i. As a minimal micro-foundation of travel costs we assume that the cost of going

[9] If we assume Cobb-Douglas preferences with γ being the expenditure share on agricultural produces the real income is $\alpha\bar{l}\left(a_A\right)^{-\gamma}\left(a_M\right)^{-(1-\gamma)}$ where a_A and a_M are unit labor inputs. To save notation we abstract from this multiplicative constant in the sequel.

[10] See Small and Rosen (1981) for the derivation of expected utility from discrete choice models.

[11] Number vendors and number of goods (or variety thereof) may be used interchangeably. The same form of utility from consumption would arise in a Dixti-Stiglitz model structure because the number of variety would be proportional to resident population.

from i to j is the time lost in transit. Let, $(1-\alpha)\bar{l}$ be the time spent traveling and let d_{ij} be the distance between i and j. A return trip would then costs $2d_{ij}/s$ where $s>0$ is the speed. Then, $\tau_{ij}=(1-\alpha)\bar{l}-2d_{ij}/s$. This minimal micro-foundations can be interpreted in the spirit of the Armington model. In this interpretation, each location would produce a distinct variety of such good; then $1-\alpha$ would be the share of income spent on tradeable goods (say textile) and τ_{ij} would be the indirect utility from consumption of a particular good after paying transport services or, which is the same, after traveling to pick up the good. Nothing hinges on the assumption that all agents travel. Alternatively, one can think of a model where trips are undertaken by the local ruler, his retinue, or by other public officers (military, ambassadors, etc.). Since these trips are financed by taxes, the cost falls ultimately on individuals. In this perspective, $(1-\alpha)$ is the time equivalent of taxation raised to finance the trips. In this discrete choice model the expected utility from trips is

$$z_i \quad = \quad \mu_Z \ln\left(\sum_{j=1}^{N} \exp\left(\tau_{ij}/\mu_Z\right)\right) \tag{15}$$

Total expected utility. Expected utility associated with location i is the sum of c_i and z_i:

$$\mathbb{E}\left[u_i\right]=\bar{l}+\mu_C \ln L_i^m+\mu_Z \ln\left(\sum_{j=1}^{N} \exp\left(-\delta_{ij}/\mu_Z\right)\right) \tag{16}$$

where $\delta_{ij}=2d_{ij}/s$.

Spatial equilibrium. A spatial equilibrium is defined as a geographical distribution of the population such that the probability that a given location is chosen equals the number of individuals who actually have chosen that location. This definition implies that net migration flows are zero in equilibrium. Formally, a spatial equilibrium is the set $\{L_i^*\}$ that satisfies $L_i^*=\sum_{j=1}^{N}\mathbb{P}_i L_j^*$, for $i=1..N$. Using expressions (11) and (16) and defining $\lambda_i \equiv L_i/\overline{L}$ and $\lambda_i^m=s_i\nu\lambda_i$ we may write the spatial equilibrium as

$$\lambda_i^*=\frac{\left(\lambda_i^m\right)^{\frac{\mu_C}{\mu_L}}\left(\sum_j^N e^{\frac{-\delta_{ij}}{\mu_Z}}\right)^{\frac{\mu_Z}{\mu_L}}}{\sum_{k=1}^{N}\left(\lambda_k^m\right)^{\frac{\mu_C}{\mu_L}}\left(\sum_j^N e^{\frac{-\delta_{kj}}{\mu_Z}}\right)^{\frac{\mu_Z}{\mu_L}}}, \qquad i=1..N. \tag{17}$$

Isomorphism. It is useful at this stage to highlight similarities and differences with other spatial economic models. As mentioned in the introduction, these models include technological spillover, amenities, fixed floor space, declining cost of living, decreasing marginal returns to land. Our data belongs to a world without these elements. First, if the least technological spillover had existed, it should have materializing in an increase in productivity. Instead, there is archaeological evidence that the technology remained unchanged over the twelve centuries in our data.[12] Second, the oasis being extremely homogeneous, there is reason to believe that amenities, if any, were the same everywhere. Third, congestion due to competition over a fixed stock of floor space is unimaginable in a place where zoning was absent, space is immense, and the intermediate inputs needed for dwellings (clay and water) are essentially free goods. Fourth, congestion for floor space is a short run phenomenon which hardly persists over very long time as new dwellings are built. Fifth, given the homogeneity of the land and the low population density relative to resources, decreasing marginal returns to land may safely be ruled out. As corroborating evidence we note the land of the oasis was able to sustain the doubling of the population that occurred from the tenth to the twelfth century without any change in agricultural technology.

The only element we have in common with these models is the fact that centrality is an advantage. Even though our model is minimalist it is isomorphic to some traditional model structures. The centrality index z can be interpreted as the price index of an Armington model where each urban system produces a single differentiated and tradeable good. The centrality index is also consistent with the price index of monopolistic competition models if we assume continuous space as in Allen and Arkolakis (2014) because in such case a change in the population of a location does not affect the price index (Allen and Arkolakis (2014) footnote 5). While technological spillover and amenities may safely be ruled out, it is conceivable that the attractiveness of a location be proportional to L_i if consumers have stochastic preferences over vendors. Then, the expected indirect utility of going to market in i is proportional to L_i. Introducing this element makes the model isomorphic to all models that include local attractive or repulsive mechanisms proportional to the local population as in Helpman (1998), Allen and Arkolakis (2014),

[12]As an example, the size of oven for manufacturing production has remained the same over the entire period and over the entire oasis; which testify of a remarkably constant technology. Utensils for agricultural productions have also remained the same over space and time. This is not surprising and it is a feature of all economies in that time period.

and Redding (2016). Isomorphism is possible since the McFadden model can accommodate these agglomeration and dispersion forces in the deterministic part of utility as shown in Behrens and Murata (2021). The mean reverting Brownian Motion is different from the Gibrat law in two important ways. First, Gibrat predicts infinite average and infinite variance in infinite time. Second, at any time before infinity, the distribution resulting from Gibrat law is not log-normal.

7 Geography and city size distribution.

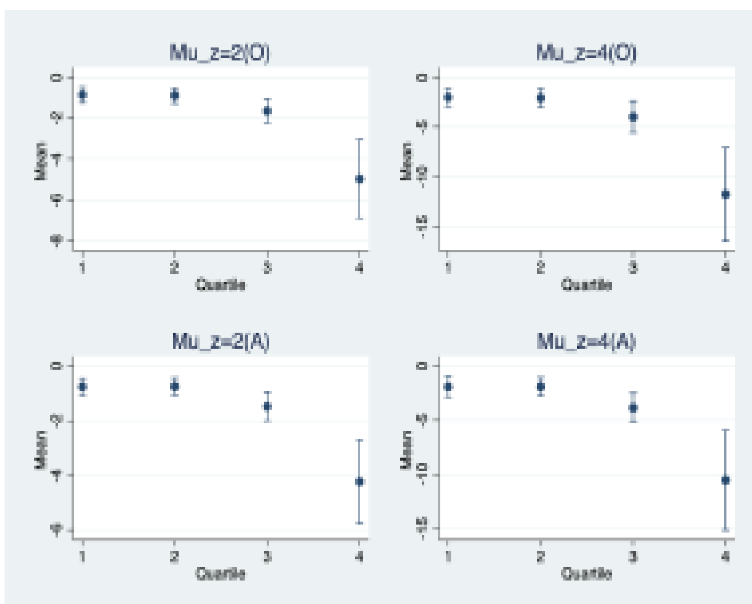
We have seen above that the city-size distribution for the Bukhara oasis passes the log-normality test. However, the deterministic part of the theoretical model in this section is not log-normal. We want to ascertain with simulations under which parametric conditions the deterministic part of the model laid out above gives rise to a city size distribution that passes the log-normality test. To this purpose we draw 1000 samples with 53 cities dropped randomly in the oasis of Bukhara (51 for the Economic Oasis). For each sample we compute the geodesic distances between these fictitious cities and we compute the model predicted population based on equation 17. We choose values of μ_z between 1 and 5. For each sample distribution we test for log-normality using Shapiro-Wilkinson test and therefore perform one thousand SW tests. In Table 3 we report the percentage of instances for which the SW test does not accept the null hypothesis of log-normality for each of the values of μ_Z considered.

Table 3: Percentage of simulated distribution passing the log-normality test

	Archaeological Oasis								
μ_z	1	1.5	2	2.5	3	3.5	4	4.5	5
%	0	0	12	57	75	79	67	55	44
	Economic Oasis								
μ_z	1	1.5	2	2.5	3	3.5	4	4.5	5
%	0	2	31	68	74	69	58	48	41

We see that the percent of distributions that pass the test of normality increases exponentially as we move from $\mu_z = 1$ peaking around $\mu_z = 3$. This simulation table and the theoretical model tell us something interesting. While the theoretical city size distribution arising from the discrete choice model is never log-normal, it can pass the log-normality test up to three quarter of times. This simulation exercises suggest that the probability of normality in the data increases with the dispersion parameter, μ_z (also migration elasticity) initially and then falls as dispersion becomes very high. If we make a comparison with the empirical estimates, we find that an estimated μ_z of 3 is around the point where our model predicts the data to have the highest probability of log normality across any geography. Hence our initial tests which do not reject log normality on Bukhara data square well with two facts from our estimation and simulation: centrality explains most of the variation in population across locations and migration elasticity (which affects centrality) across locations is around 3.

Figure 7: Duranton Concavity in Simulated Distribution



Notes: This figure plots the mean of the coefficients from the rank size regression ($log(Rank) = c_0 + c_1 log(Population) + \epsilon$) from the simulated distribution of population. We generate 1000 simulations of size 51. The regression is run by each quartile. The top left figure is for migration elasticity of 2 while the bottom right figure is for migration elasticity of 4.

We then check for the city's rank size relationship in the simulated distribution by regressing their log rank on log population by each quartile. The one stylized fact that we find in the data pertain to rank size coefficient falling as the city size increase. Figure 7 shows that the simulated data from our model also replicate this empirical fact. Even when we change the values of μ, we find concavity persist with higher μ only affecting the dispersion of the values.

8 Empirical analysis

Our objective in this section is to estimate the two key parameters in the model, which are μ_Z and μ_L. Before we estimate the model, we show an OLS of log population on a measure of centrality (with de-meaned distances) and silk route dummy. This gives us a motivating regression to a more careful estimation of the model in the next sub section. Table 4 shows that

Table 4: OLS Regression

	log Population (s.e.)
Centrality	0.69^{**}
	(0.30)
Silk Route Dummy	1.08^{***}
	(0.21)
Constant	5.5^{***}
	(0.95)
Adjusted R^2	0.32

centrality and silk route are both important explanatory variables in a linear setup and can explain close to a third of the variation without taking into account the non-linearities of the model. We also show that silk route dummy, which indicates whether the manufacturing city was a stop over point is an important explanatory variable. In our model estimations which follow next, we will account for population expansion due to silk route.

Estimation. The prediction of the theoretical model is that population distribution across regions can be represented by taking logs in equation 17 and rearranging to get

$$\ln \lambda_i^* = \frac{\mu_C}{\mu_L - \mu_C} \ln s_i + \frac{\mu_Z}{\mu_L - \mu_C} \ln \sum_{j=1}^{N} e^{\frac{-\delta_{ij}}{\mu_Z}} - \ln \sum_{k=1}^{N} s_i \lambda_i^* \left(\sum_{j=1}^{N} e^{\frac{-\delta_{kj}}{\mu_Z}} \right)^{\frac{\mu_z}{\mu_L}} \tag{18}$$

In this estimating equation, λ^* is the population share in the urban system while δ_{ij} indicates the bilateral travel cost between any two urban systems, i and j. We use geodesic distances to compute these bilateral travel costs. Parameters in this equation are μ_l, μ_z and μ_c. μ_l comes from the random term in the utility function and guides the variance of individual preferences over daily travels across locations. μ_z affects the random term determining the location choice decision for residence, while μ_c represents the i.i.d shock representing preferences over local traders. We match key moments in the data to estimate the parameters. In the moment conditions, we include variance, skewness, and range to jointly identify the parameters μ_l, μ_z and μ_c. We calibrate the parameters using method of moment estimation and the results are reported in table 5 (Model 1).

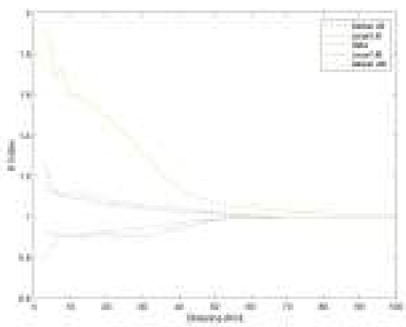
Table 5: Estimation using Method of Moments

	Parameters		
	μ_c	μ_l	μ_z
Model 1	26.99	25.69	2.93
Model 2	0.18	1.74	3.3

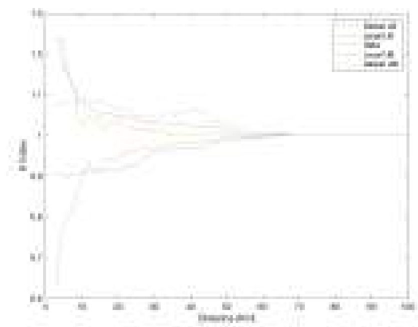
Alternatively, we use information on the size of the business district in the estimations. We regress the size of the business district for each urban unity on its population and calculate the residuals. For the cities which have positive residuals and were a silk route stopover, we replace s in the estimating equation with the positive residual values. This implies, how much extra residential space existed after controlling for population, which in turn gives a proportionally larger space for vendors to exchange a variety of goods. This in turn can make the nearby locations attractive residential place. Corresponding estimates are reported in model 2 above (5).

9 Persistence

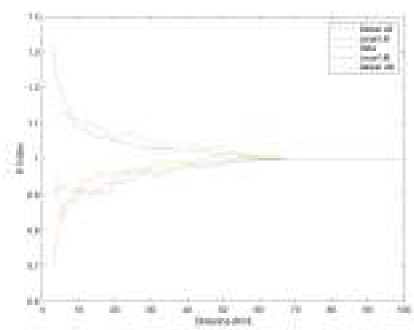
In this section, we explore to what extent the geographical location of major cities today was affected by where historically major cities of the time were located. We plot the coordinates of 38 large cities which are located in the oasis of Bukhara as of today. We test if they are more likely to be concentrated around historically 53 manufacturing cities than a random location draw would entail. For this, we draw a random sample of 38 points around 53 counterfactual distributions of historical cities. In each sample, we compute the co-concentration index which captures how likely a modern city is concentrated around a historical city than any city (historical or modern). We repeat sample draws hundred times and get the local and global CI. The local CI is computed by rejecting the top and bottom 5% of observations at each distance while the global CI is computed by rejecting 5% of observations across all distances such that the CI at each distance is of equal size. In the 9^{th} century, among the 53 manufacturing cities of Bukhara, 10 were such that the silk road passed through them as the traders would use them for halting in their journey. Our regressions in Table 7 and 8 had shown that these checkpoints were more populated than non-checkpoints. We now want to see if this historical fact mattered in the 21st century. We repeat the above sampling as we did for 53 manufacturing cities for the 10 check points to compute the co-concentration of modern cities around historical check points along the checkpoints within the Bukhara Oasis. When we look at figure 8a, we see that the co-concentration of modern cities around historical cities in Bukhara is much more than a random draw would entail. In particular, the concentration is much higher between a 5-10 kilometer radius suggesting that the modern cities located around historical cities more often, taking the geographical advantages and probably re-enforcing it through modern infrastructure. In figure 8b we find a much weaker instance of concentration around the checkpoints. For a radius up to 10 km, there seems to be co-concentration based on global CI however the co-concentration index would be contained within the stricter local CI. So we can conclude that the historical cities mattered more for the co-concentration of modern cities and their location can not be rationalized by a random draw. In figure 8c, we test whether the current Uzbek cities within the oasis of Bukhara are more likely to exist around the city location of those which existed around 1st century BC cities. We do not find them to be so. In fact we do see for short distance that it is less likely to find a modern Uzbek cities in the vicinity of an ancient



(a) Modern and ancient cities



(b) Modern cities and Silk Road



(c) Modern cities and 1st BC

Figure 8: Persistence of City locations

city relative to any other cities. It suggests that the high persistence that we see around Bukhara city locations from the 9^{th} century maybe reversed once significant time has passed. Cities which were populous in the past has some appeal in the present due to existence of large infrastructure. However with further passing of time, as those structures become obsolete, this advantage begins to fade out.

10 Conclusion

In this paper, we revisit the city size distribution debate, and expand the empirical analysis to many countries to show essentially one stylized fact which emerges from the data across almost all countries: concavity in rank size relationship. We show that Zipfs law, as predicted by random growth models, do not hold for large cities i.e the log rank and log size of cities have a -1 linear slope in the upper tails. Even when we find the slope to be -1 for a subset, we find a concave relationship for all except 2 countries. City size distribution also doesn't follow log normality, as suggested by models confirming to Gibrat's law. In both these models, geographic location of cities play no role. We provide new evidence using 9^{th} century data from the oasis of Bukhara in Central Asia to show the importance of geographic centrality in determining city size distribution using an elementary model of people who travel for work and leisure. Our simulations suggest that migration elasticity plays a central role in log normality. Our estimates for Bukhara lie in the range of elasticities which increases the likelihood for log normality. The data for Bukhara also confirms to log normality. The location of the old cities also have a persistent effect on where the modern cities are located.

References

Al-Narshakhi, A. B. M. (943). The History of Bukhara. (Transleted from Persian into English by Frye, Richard N. in 1954). Markus Wiener Publishers. Latest edition 2011.

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. The Quarterly journal of economics 129(3), 1085–1140.

Arshad, S., S. Hu, and B. N. Ashraf (2018). Zipf's law and city size distribution: A survey of the literature and future research agenda. Physica A: Statistical Mechanics and its Applications 492, 75–92.

Barjamovic, G., T. Chaney, K. Coşar, and A. Hortaçsu (2019, August). Trade, Merchants, and the Lost Cities of the Bronze Age. The Quarterly Journal of Economics 134(3), 1455–1503.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: Sorting, selection, and agglomeration. The Journal of political economy 122(3), 507–553.

Behrens, K. and Y. Murata (2021). On quantitative spatial economic models. Journal of Urban Economics 123, 103348.

Brülhart, M., C. Carrère, and F. Trionfetti (2012). How wages and employment adjust to trade liberalization: Quasi-experimental evidence from austria. Journal of international economics 86(1), 68–81.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. Econometrica 87(3), 741–835.

Davis, D. R. and J. I. Dingel (2019). A spatial knowledge economy. The American economic review 109(1), 153–170.

Davis, D. R. and J. I. Dingel (2020). The comparative advantage of cities. Journal of international economics 123, 103291.

Davis, D. R. and D. E. Weinstein (2002, December). Bones, bombs, and break points: The geography of economic activity. American Economic Review 92(5), 1269–1289.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018). The geography of development. The Journal of political economy 126(3), 903–983.

Dobkins, L. and Y. M. Ioannides (2001). Spatial interactions among u.s. cities: 1900–1990. Regional science and urban economics 31(6), 701–731.

Duranton, G. (2007). Urban evolutions: The fast, the slow, and the still. American Economic Review 97(1), 197–221.

Duranton, G. (2008). Spatial economics. In S. N. Durlauf and L. Blume (Eds.), The new Palgrave dictionary of economics (2nd ed.). Houndmills, Basingstoke: Palgram Macmillan.

Eaton, J. and Z. Eckstein (1997). Cities and growth: Theory and evidence from france and japan. Regional science and urban economics 27(4), 443–474.

Eeckhout, J. (2004). Gibrat's law for (all) cities. American Economic Review 94(5), 1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: Reply. American Economic Review 99(4), 1676–1683.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. The Journal of political economy 122(3), 554–620.

Gabaix, X. (1999). Zipf's law for cities: an explanation. Quarterly Journal of Economics 114(3), 739–767.

Gabaix, X. and M. Y. Ioannides (2004). The evolution of city size distributions. In V. Handerson and J.-F. Thisse (Eds.), Handbook of Regional and Urban Economics, Volume 4. Elsevier.

Gibrat, R. (1931). Les Inégalités économiques, applications : aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel. Paris: Libr. du "Recueil Sirey.

Helpman, E. (1998). The size of regions. In S. E. Pines, D. and I. Zilcha (Eds.), Topics in Public Economics: Theoretical and Applied Analysis. Cambridge University Press.

Ioannides, Y. and S. Skouras (2013). Us city size distribution: Robustly pareto, but only in the tail. Journal of Urban Economics 73(1), 18–29.

Ioannides, Y. M. and H. G. Overman (2003). Zipf's law for cities: an empirical examination. Regional science and urban economics 33(2), 127–137.

Krugman, P. (1996). Confronting the mystery of urban hierarchy. Journal of the Japanese and international economies 10(4), 399–418.

Lee, S. and Q. Li (2013). Uneven landscapes and city size distributions. Journal of Urban Economics 78, 19–29.

Levy, M. (2009). Gibrat's law for (all) cities: Comment. American Economic Review 99(4), 1672–1675.

McFadden, D. (1974). Conditional logit analyis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers in econometrics. Academic press.

Parr, J. B. and K. Suzuki (1973). Settlement populations and the lognormal distribution. Urban Studies 10(3), 335–352.

Rante, R. and J. Mirzaakhmedov (2019). The Bukhara oasis 1. Population, depopulation and settlement evolution. Leiden-Boston, Netherlands: Brill.

Redding, Stephen, J. (2022). Trade and geography. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), Handbook of International Economics, Volume 4, Chapter 3. Elsevier - North Holland.

Redding, S., E. Liu, and B. Kleinman (2022). Dynamic spatial general equilibrium.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. Journal of international economics 101, 148–167.

Redding, S. J. and E. Rossi-Hansberg (2017). Quantitative spatial economics. Annual review of economics 9(1), 21–58.

Redding, S. J. and D. M. Sturm (2008). The costs of remoteness: Evidence from german division and reunification. The American economic review 98(5), 1766–1797.

Rosen, K. T. and M. Resnick (1980). The size distribution of cities: An examination of the pareto law and primacy. Journal of Urban Economics 8(2), 165 – 186.

Rozenfeld, H., D. Rybski, X. Gabaix, and H. a. Makse (2011). The area and population of cities: New insights from a different perspective on cities. American Economic Review 101(5), 2205–2225.

Schluter, C. (2020;2021;). On zipf's law and the bias of zipf regressions. Empirical economics 61(2), 529–548.

Schluter, C. and M. Trede (2019). Size distributions reconsidered. Econometric reviews 38(6), 695–710.

Small, K. A. and H. S. Rosen (1981). Applied welfare economics with discrete choice models. Econometrica 49(1), 105–130.

Soo, K. T. (2005). Zipf's law for cities: a cross-country investigation. Regional science and urban economics 35(3), 239–263.

Zipf, G. (1949). Human behavior and the principle of least effort: An introduction to human ecology.

Étienne de la Vaissière (2018). Sogdian Traders: A History. Leiden, The Netherlands: Brill.