An Energy-centric Theory of Agglomeration*

Juan Moreno-Cruz† M. Scott Taylor‡

December 9, 2016

Abstract

This paper sets out a simple spatial model of energy exploitation to ask how the location and productivity of energy resources affects the distribution of economic activity across geographic space. By combining elements from energy economics and economic geography we link the productivity of energy resources to the incentives for economic activity to agglomerate. We find a novel scaling law linking the productivity of energy resources to population sizes; rivers and roads effectively magnify productivity; and show how our theory's predictions concerning a single core aggregate to predictions over regional landscapes and city size distributions at the country level.

*We are very grateful to seminar participants in several countries, and for the hospitality of the Energy Institute at Haas, University of California Berkeley and the Henri Poincare Institute in Paris, France where a portion of this work was conducted. We would like to thank, without implicating, Bob Allen, Jared Carbone, Jevan Cherniwchan, Oded Galor, Gene Grossman, David Laband, Arvind Magesan, Ken McKenzie, and Jim Wilen for comments on earlier drafts or portions of this work. Excellent research assistance was provided by Nolan Derby, Fruzsina Mayer, and Rui Wan. The usual disclaimer applies.

†Assistant Professor, School of Economics, Georgia Institute of Technology, Atlanta, Georgia

‡Professor of Economics and Canada Research Chair, Department of Economics, University of Calgary, Canada; Research Associate at the NBER; Fellow of the Beijer Institute of Ecological Economics; CESifo Associate and a research professor at the Ifo Institute, Germany.

1 Introduction

The purpose of this paper is to ask how does the location and "spatial productivity" of energy resources affect the distribution of economic activity across the globe? This is a large, and as yet unanswered, research question that cannot be resolved by any one paper. Instead, we take one small, but important, step towards answering it by developing a bare bones theoretical framework to define energy's (spatial) productivity precisely; to show how differences across energy resources in their productivity have large effects on supplies; and to demonstrate how even small variation in the productivity of energy resources across landscapes can create very densely populated agglomerations of widely different sizes surrounded by large areas with dispersed production and low population density.

Even a momentary glance at the world today should give pause to any reader looking for a simple connection between today's centers of economic activity and the location of today's most important energy resources. The distribution of economic activity we observe today surely reflects a complex mix of forces. Cheap energy sources in the distant past may have produced permanent centers of economic activity that today are very distant from current energy supplies. Since the world economy has employed many different energy sources over many centuries, any serious empirical examination needs to account for the many other forces governing the location of economic activity. Three recent studies, however, have done exactly that and found strong links between the existence of spatially productive resources, population growth, and urbanization.

Nunn and Qian (2011) provides empirical evidence linking the introduction of the potato from the New World to the Old, to higher population levels and urbanization rates in the Old World. Their baseline estimates suggest the introduction of the potato, which is almost three times more spatially productive than typical staple crops, is responsible for 26% of the increase in Old World populations from 1700 to 1900. Similarly, Fernihough and O'Rourke (2014) study the introduction of coal using technology (primarily steam and smelting innovations) and link these introductions to changes in European city populations during the

Industrial Revolution.[1] They find the ability to use coal (in these uses) was responsible for over 60% of the growth in European city populations from 1750 and 1900. Moving to a contemporary setting, Severnini (2013) examines the impact of hydroelectric dam building on local communities throughout the U.S from 1920 to 1980 and finds the impact of dams on county population growth is considerable and long lasting. We take these studies, and others, as motivation for a theoretical analysis of how variation in energy constraints across space affects the geography of economic activity.[2]

The modeling choices we make are informed by four observations. The first is that energy is not physically scarce. Many sources of energy available today — solar, wind, coal and non-conventional oil and gas — represent vast, almost limitless, potential supplies. The economic costs of exploiting them however limit their use. The second is that exploiting far flung energy resources and moving energy to markets is primarily what the energy industry does. Third is a recognition that one of the most important attributes of an energy resource is its ability to deliver substantial power relative to its weight or other physical dimensions. And fourth, that energy — above all other productive factors — has a strong claim to being an essential input to human activity, production, and growth.

To investigate the potential implications of these observations for the distribution of economic activity across space, we proceed in two steps. First, we build a very simple spatial model. It features real geographic space, a fixed core that occupies no space, and energy resources that are drawn from a surrounding two dimensional plane. These assumptions eliminate any role for physical scarcity by assuming energy resources are limitless, but still costly to exploit. And since energy resources are located in geographic space, the availability of energy resources at any given location reflects spatial productivity and transport costs.

We measure spatial productivity of an energy resource by its *power density*. The power

[1] Moreno-Cruz and Taylor (2012) argue that the shift in England from biomass based fuels to coal in the latter part of the sixteenth century was at least partially responsible for the radical shift in the population distribution across the country.

[2] Economic historians have long drawn a link between the characteristics of energy resources and economic outcomes. See for example Allen (2009), Wrigley (2010), Smil (2008), Fouquet and Pearson (1998) and de Zeeuw (1978).

density of an energy resource represents its ability to provide a flow of power taking into account the area needed for its exploitation.[3] To model transport costs we go back to the fundamentals of work, force, power and resistance to understand at a very basic level how physical differences across energy resources affect their transport costs. The benefit of this return to fundamentals is that transport costs are grounded in physical laws.

Putting our assumptions together generates our Only Energy Model where the location and productivity of available energy resources determines energy supplies at our core. We show how energy supply adjusts along both intensive and extensive margins and together these produce a scaling law linking delivered energy supplies to the core with the cube of energy's spatial productivity. We also show how variation in transport costs, introduced perhaps by rivers, roads or transmission lines, effectively magnify the spatial productivity of nearby energy resources leaving our scaling law intact. Our second step is to study the economic motives behind agglomeration. To do so we embed the Only Energy Model in a general equilibrium setting where agents surrounding a single potential core choose to either agglomerate or remain dispersed in the hinterland. We create an incentive for agglomeration via conventional channels, but find several non-conventional results.

First we show how a location's "comparative advantage" (measured by the ratio of spatial productivity to transport costs) dictates whether we will see agglomeration or not; whereas a location's absolute advantage (measured by spatial productivity alone) determines how rich or populous a city may become. Minimum city size is independent of how energy rich a region may be, and low transport costs or high productivity alone do not guarantee agglomeration. Second, we show how economic activity agglomerates in places where favorable geography alone magnifies existing, and perhaps even relatively poor, resource potential because of its

[3]One benefit of using power density as a measure of spatial productivity is that it is determined by very conventional and commonly used measures of resource characteristics. For example, it reflects differences in the energy content of available resources (key to fossil fuels), recharge rates (key to many renewables), and yields (important to all staple energy crops). See Section D in our Online Appendix for a decomposition of power density into these components for both renewables and non-renewables. Power density measures the flow of energy a resource can provide in Watts per unit area needed for its exploitation and maintenance; it is typically measured in Watts per m^2.

low cost or abundant transportation options. Third, by adding up over uniform geographic spaces (landscapes), we show how our theory generates additional predictions for regional population densities and city numbers. These results at the regional level echo our scaling law but in surprising ways. For example, regional populations respond proportionately to differences in spatial productivity whereas core sizes within regions scale with the cube of spatial productivity. Finally, we combine many such landscapes or regions into a hypothetical country and show how our theory provides a simple link between the distribution of comparative advantage over geographic space and a country's distribution of city sizes. The distribution is truncated with no very small cities and it has a long right tail with few very large cities. The exact distribution flows from the interplay of nature's distribution of comparative advantage across geographic space and our model's generated scaling law relationships.

Therefore, despite its back-to-fundamentals flavor and its highly abstract presentation, the theory can provide several sharp predictions suitable for empirical testing using new GIS data (Nordhaus, 2006). And while increasing returns and transport costs play a role in the model, the scaling relationship linking energy productivity to energy availability is the key driver of most results. In this sense, we provide an energy-centric theory of agglomeration that goes some way in providing an answer to our research question.

Our work is related to previous contributions in both energy and resource economics and economic geography, but has also benefitted in perhaps less obvious ways from the contributions of economic historians. Although the Only Energy Model is constructed from first principles, it bears some resemblance to von Thunen's model of an Isolated State. In contrast to von Thunen however, transport costs and, by implication, the exploitation zone are set by appeal to physical laws governing energy use. It also bears a family resemblance to other spatial models of resource and energy use where resources and demand centers are treated as points in space (Gaudet, Moreaux and Salant (2001)); where consumers (Kolstad (1994)) or resources (Laffont and Moreaux (1986)) are distributed on line segments; where resource pools are differentiated by costs, suggestive of a spatial setting (Pindyck (1978), Swierzbinski and Mendelsohn (1989), and Chakravorty, Roumasset, and Tse (1997)); and situations where resources themselves move across space (Sanchirico and Wilen (1999)).[4] It draws on tools developed in the economic geography literature, but is more concerned with the size and location of economic cores than with the empirical questions of this literature (Head and Mayer (2004)). It is also related to the vast literature in urban and regional economics on agglomeration economies (Rosenthal and Strange (2004)) and city size distributions (Decker et al. (2007)).

It differs from all of this work in its treatment of geographic space outside of cities, its focus on energy as an essential input, its ability to provide an explicit link between unique geographic features such as rivers or coastlines and their resulting impacts on economic activity, and its ability to move from single core, to region, to country level implications. It is of course similar in some ways as well: for example, we too rely on the gains of specialization creating increasing returns and agglomeration, but we limit city size not by recourse to congestion or housing costs within the city (Helpman, 1995), but by rising energy costs created by bringing energy resources from outside.

The rest of the paper proceeds as follows. In section 2 we develop the Only Energy Model and link features of the transportation network to energy supply. In section 3 we introduce a simple general equilibrium market economy to study the incentives to agglomerate. Section 4 connects our theory to empirical work by discussing its implications for population sizes, regional densities and city size distributions. A short conclusion follows. Detailed calculations and proofs of propositions are in the Appendix. An Online Appendix contains references to data sources, some model extensions and further calculations.

[4] There is also an emerging literature in environmental and resource economics where pollution or resource flows follow a diffusion process across space (See Brock et al. 2012, Desmet et al. 2015). While these authors take space seriously, just as we do, their contributions focus on very different questions and problems.

2 The Only Energy Model

We develop a simple model of energy exploitation where energy is the only input of production. Energy is freely available everywhere on a two-dimensional plane with given density, and transporting energy in all directions has the same costs. We focus on the case of renewables since it admits a simple steady state analysis. Non-renewables are treated in our Online Appendix.[5]

2.1 The Scaling Law

We start with a definition. The area exploited in the collection of energy is related to the power obtained measured in Watts $[W]$, and to the spatial productivity of the resource, measured by its power density in Watts per meter squared $[W/m^2]$. If the flow of energy collected is W, and the available energy resource has power density Δ then the area where these resources are collected from, called the exploitation zone, EX, must equal:

$$EX = W/\Delta \tag{1}$$

where EX is measured in meters squared $[m^2]$. We assume collection is costless, but transport to the core requires the use of energy.[6] For now, assume transport costs are proportional to distance and energy collected (we will provide conditions under which this will be true subsequently). In this case, all we need to understand is unit costs. To that end, consider energy resources with power density Δ, and let c/Δ be the energy cost of moving one Watt of power, from these resources, one meter. Why we use this exact specification will become clear subsequently, but for now note that if our objective is to maximize energy deliveries to the core, then we collect energy resources until the marginal resource collected provides no

[5]Readers can think of the resource as literally renewable, or take its constant cost of collection as reflective of a exceedingly abundant non-renewable resource with no current user cost. See Section C.3 in the Online Appendix for a more formal, and conventional, treatment of non-renewables.

[6]Adding constant per unit collection/harvesting/extraction costs contributes little but notation to the analysis.

net energy. Denote by R^* the distance these marginal resources are from the core. At this margin one Watt collected is now fully expended in costly transportation to the core; that is, R^* must satisfy:

$$1 - \frac{c}{\Delta}R^* = 0 \text{ or } R^* = \frac{\Delta}{c} \tag{2}$$

The more power dense are the energy resources, the larger is the circular exploitation zone surrounding our core.[7] For example, very power dense resources (for the moment, think very dry timber) will be collected at great distances, while energy resources like straw or dung will not. If energy is an essential input then the limits imposed by (2) will in turn constrain economic activity in any core.

To find the power available for use in the core, we start by using (2) in (1) to find total power collected is simply given by: $W^* = \Delta EX = \Delta\pi[R^*]^2 = \pi\Delta^3/c^2$.

To find the power available for use in the core, we need to subtract the energy costs of transport. This net supply of power comes from adding up, what we might call, "energy rents." These rents are the excess of energy collected over transport; i.e. $\Delta - cr$ at all distances $r \leq R^*$ from the core. To add them we use a two step procedure. Along any ray from the core, there are Δ Watts of power every meter and transporting these resources to the core yields a density of $[\Delta - cr]$ net Watts of delivered power. The first step is to add up these resources along our ray over all distances less than R^*. The second step is to accumulate these quantities by sweeping across the 2π radians of our circular exploitation zone. By doing so we obtain net power supply to the core as the sum of all energy rents:

$$W^S = \int_0^{2\pi} \int_0^{R^*} v[\Delta - c \cdot v] dv d\varphi = 2\pi \int_0^{R^*} v[\Delta - c \cdot v] dv = \frac{\pi\Delta^3}{3c^2} \tag{3}$$

[7] Despite some similarities, our formulation and that of von Thunen are not the same. Whereas we associate any energy resource with a finite region of exploitation tied to its power density, the geometric transport costs of von Thunen — cleverly coined iceberg costs by Samuelson — imply an infinite exploitation zone for any and all energy resources. Without additional assumptions the iceberg assumption of von Thunen leads to the somewhat uncomfortable implication that we can move a barrel of oil (a cord of wood, a bale of hay, a pound of dung, an Ampere of electricity etc.) a billion miles and still reap some energy resources from it. See Section B in our Online Appendix for a further discussion.

Since net power is a cubic in power density, renewable resources twice as power dense deliver eight times the supply. The implication of this result is immediate: even small variations in the natural landscape affecting the productivity of energy resources can have large implications for energy supply in the core. For example, suppose the distribution of power densities Δ over potential locations $f(\Delta)$ was uniform over $[0, \bar{\Delta}]$; then 50% of the total net energy supplied is concentrated in 12% of all locations.[8] To verify that this result reflects a scaling law tied to our spatial setting rather than being an artifact of our circular region of exploitation, we now construct a setting where the exploitation zone is not circular.

To proceed we assume resources at different locations have different transport costs. This heterogeneity could arise exogenously from the nature of the resource or the terrain; but, as we will subsequently show, it will arise endogenously when agents avail themselves of nearby roads, rivers or transmission lines. For simplicity, we measure the location of a resource by its direction (in radians) relative to the core. In this more general environment we can now show:

Proposition 1 *Scaling Law: If energy resources everywhere have power density Δ but transport costs vary with the direction θ so that $c = c(\theta)$, then the exploitation zone is no longer circular and gross power collected and net power supplied remain homogenous of degree three in Δ.*

Proof. *See Appendix.* ■

The intuition for this result is easy to grasp. Suppose we increase the power density of available resources, but hold the size of the exploitation zone constant; then supplied power should rise proportionately with power density; i.e. appear with power 1 (recall the definition in (1)). This is the impact on the intensive margin of collection, but a higher

[8] To see why note the probability distribution of power W is given by: $F_W(w) = \Pr\{W < w\} = \Pr\left\{\frac{1}{3}\frac{\Delta^3}{c^2} < w\right\}$. This implies $\Pr\{\Delta < (3c^2w)^{1/3}\} = \frac{(3c^2w)^{1/3}}{\bar{\Delta}}$. The value $w_{50\%}$ for which 50% of the total net energy delivered across all locations solves $F_W(w_{50\%}) = 0.5$. Solving for $w_{50\%}$ we obtain: $w_{50\%} = (0.5^3)\bar{W} = 0.125\bar{W}$ where $\bar{W} = \frac{\bar{\Delta}^3}{3c^2}$. 50% of the net energy delivered is concentrated in 12.5% of the locations.

power density implies every meter expansion of the exploitation zone garners more resources than before. The marginal cost of collecting an extra Watt falls, and the extensive margin moves outwards. Since area scales with the square of this now expanding extensive margin, this expansion implies the set of exploitable energy resources rises with the square of power density. Adding up adjustments across both the intensive and extensive margins, means total power collected, and net power delivered, are both cubes in power density. While this logic is impeccable, it does however rely on our assumption that energy resources travel at constant costs; and at this point it is not clear exactly why, or for what energy resources, this should be true. We have also claimed that transport costs will vary with direction when agents optimize across transport options but this is far from obvious. The next subsection examine these assumptions.

2.2 Transport costs

2.2.1 Biomass, Wood, and Staple Crops

We have assumed a constant energy cost to transport resources, but said very little about why. To understand this assumption, consider the movement of resources with physical mass like biomass, wood, and staple crops. The energy costs of moving these resources amounts to the work done in overcoming friction in land transportation; and since these resources are transported continuously in steady state this work amounts to Watts of power expended for transport. To understand why costs are constant we need to use a small amount of high school physics. Recall work is equal to force times distance, $W_k = F \cdot x$. Force is in turn equal to mass times acceleration $F = M \cdot a$. In our case, the relevant acceleration is the normal force exerted by gravity since any mass moved horizontally must overcome the force of gravity g as mediated by friction in transport where μ is the coefficient of friction.[9]

[9] We are ignoring static friction encountered when the object first moves. The force that needs to be overcome to keep an object in motion is equal to the normal force times the coefficient of friction. Since the object is moving horizontally, the normal force is just gravity times the mass of the object. The coefficient of friction is a pure number greater than zero; and force is measured in Newtons.

This work is done per unit time since power is a flow. And measuring time in seconds, W_k expended in transportation is now Joules per second which represents the Watts expended in bringing resources to the core.[10]

Keeping these results in mind, revisit our unit costs of transport. If the energy resource in question has power density Δ [Watts/m ^{2}], then resources capable of providing one watt are reaped from an exploitation zone with area $1/\Delta$. If this resource — think timber or biomass — is available in a quantity d kilograms per squared meter, then these resources must weigh d/Δ kilograms. And moving these resources one meter while overcoming friction, requires a flow of power of $\mu gd/\Delta$. Therefore, when energy resources have mass (and they incur land transport costs) we have $c = \mu gd$ which is of course constant.

It is now apparent why we have assumed transport costs are linear in distance (work is proportionate to distance) and linear in power collected (work is proportionate to the mass transported).

2.2.2 Solar Power, Wind Farms, and Electricity

While constant costs may be a good assumption for some renewables, many of the most common renewable energy sources we use today - wind, solar and hydro - need to be transformed into electricity before they can be used in productive ways. And it is not obvious these resources travel at constant cost. The key observation is that line losses — that is, the power lost in electricity transmission — operate much like our constant energy costs of transport given by c. These line losses come from resistivity losses or what the industry calls *Joule heating.* These losses are the analog of the energy spent in performing work when transporting resources with mass. In our earlier discussion of energy resources with mass, we implicitly assumed resources move at constant speed (there was no acceleration of the resource, no inertial friction, and no deceleration either) which seems quite natural given our steady state analysis. The parallel assumption here is that electric power should move

[10]Expending 1 Joule of energy in 1 second means you are delivering 1 Watt of power.

at a constant current which we denote by I and measure in Amperes. But just as physical transport costs are linear in distance when objects move at constant speed, line losses are linear in distance when electricity is transmitted at constant current. Therefore, we can once again link our per unit distance transport cost, c, to fundamental determinants. Since doing so relies on concepts less familiar to most readers (Ohm's law, definitions of line resistance, etc.) we leave the details to the Appendix, and simply assert c is again constant, but now reliant on the current and the transmission line's resistivity as reflected by its material, φ, and its cross sectional area, a.[11] Therefore if solar or wind resources are geographically dispersed then their power density will determine, exactly as before, the net power that could be delivered to the core.

2.2.3 Optimization over Geography

We have thus far assumed transportation is possible from any location and, for the most part, equally costly. But rivers, roads, and canals have offered relatively cheap transport for food and fuel for centuries, while power lines, pipelines, and LNG terminals are necessary for the transport of many 21^{st} century resources. Both of these examples present a challenge to our theory since they suggest there are large and discrete differences in transport costs across space which would potentially alter the optimal transportation of energy resources, our exploitation zone, and therefore our scaling law and its implications. In this section we demonstrate how optimal use of these networks by agents implies that costs vary by direction producing an endogenously determined $c(\theta)$ schedule that satisfies Proposition 1.

Consider the decision problem of a potential energy supplier located on one meter square area containing resources that generates a flow of energy equal to Δ Watts. The supplier can take energy directly into the core or deviate to take advantage of a road nearby. Rivers and roads help to reduce the amount of work used in transportation, increasing the amount of energy delivered to the core. To capture this in our analysis we allow for the unit cost of

[11] In particular, $c = I^2(\varphi/a)$. See Section B in the Appendix for further details.

transportation by road to differ from the unit cost of transportation by land by a fraction $\rho < 1$. That is, while the cost by land is equal to c, the cost by road is ρc.[12] We assume the road is a straight line that crosses the core and expands indefinitely. The location of a supplier relative to the core is described by two terms: r, the distance from the core and θ the angle between the segment formed by the core and the supplier and the road. A potential energy supplier then has to decide on the optimal route to the core, and if the trip is worthwhile to undertake. Since the calculations involved are somewhat detailed we leave them to the Online Appendix, and just report here that optimization by agents relative to a given road/river network implicitly defines an endogenous $c(\theta)$ schedule

$$c(\theta) = c \times \begin{cases} \left((1-\rho^2)^{1/2}\sin\theta + \rho\cos\theta\right) & \theta \leq \bar{\theta} \\ 1 & \theta \geq \bar{\theta} \end{cases} \tag{4}$$

Agents located far from the road (i.e. at angles $\theta \geq \bar{\theta}$) travel straight to the core as before, while all other agents deviate to lower costs by making at least some of their trip by the lower cost option. Since optimization produces the schedule $c(\theta)$ we know from Proposition 1 that even when agent's cost minimizing path creates a far from circular exploitation zone, our scaling law still holds and variation in the productivity of energy resources again creates large differences in delivered energy. In addition, by comparing locations with and without low cost transport we find an additional result we refer to as the *magnification effect.* Our derivations are in the Online Appendix, but an important implication follows from calculating net energy supply:

$$W^S = \frac{1}{3}\frac{\pi\tilde{\Delta}^3}{c^2} \text{ where } \tilde{\Delta} \equiv \Delta(g(\rho)/\pi)^{1/3} \text{ and } (g(\rho)/\pi)^{1/3} \geq 1 \tag{5}$$

where $g(\rho) = \pi + 2(\tan(\bar{\theta}) - \bar{\theta}) \geqslant \pi$. The expression above is exactly the same form as our earlier net supply with a slight redefinition of our power density term. Therefore, we have

[12] $\rho < 1$ can alternatively represent the benefits of transmission at higher voltage, large cables, or better materials.

proven:

Proposition 2 *Magnification. Endowing our core with access to a low cost transportation option is equivalent to endowing its surrounding region with more power dense energy resources.*

Proof. *In text.* ■

This result implies that as far as energy supply is concerned, having a road or river (or transmission line) cut across the core is identical to being surrounded by more productive energy resources.[13] A moment's reflection will reveal that regardless of the number, and regardless of the efficacy of additional (straight line) transportation options we introduce, the endogenous solution for energy supply delivered to the core, will again satisfy our scaling law. This is true, since by Proposition 1 all cost minimizing paths are just different particular solutions for $c(\theta)$. Moreover, if we were to add additional roads through our core these options can only reduce costs (given optimization), and the equivalence result we report in Proposition 2 is strengthened. Every one of these additional transportation improvements is equivalent to, in energy supply terms, an additional increase in the power density of surrounding resources. With a small bit of work it can be shown that if we were to add any number of identical low cost transportation options and locate them optimally, then:

Proposition 3 *Energy supply to the core is an increasing and weakly concave function of the number of transportation options serving it.*

Proof. *See Appendix.* ■

One interpretation of this result is that locations blessed with many transport options look like they are endowed with very power dense energy resources. Another more speculative interpretation is that such a location would support a large agglomeration of economic activity. To make this connection precise we now incorporate the supply side of the Only Energy

[13]When traveling with the current a river's cost is also ρc but against it c/ρ. By this assumption, river transport is only useful when you are an energy producer upstream; whereas road transport reduces frictions in two directions and not one. Thus, we need to modify the gains function: $g_{river}(\rho) = \frac{\pi}{2} + \frac{g(\rho)}{2}$.

Model into a simple general equilibrium model where agents choose whether to agglomerate in a central core or remain dispersed and rely on home production.

3 A Simple Energy Based Model of Agglomeration

We start with a featureless plain containing energy resources with spatial productivity Δ [W/m^2]. There is a uniform density of agents across this space, and it proves useful to think of each square meter of area as representing a single farm with some number, Ω, of agents residing in it. Initially, consumption per person on each and every farm is consistent with population being constant; therefore, population density is fixed and without loss of generality, we set the number of residents on each farm to unity: $\Omega = 1$. We leave for the moment the possibility that migration or population growth may respond to the real income gains brought about by agglomeration, but return to this issue later. These residents reap the energy from their land and transform this into final products for consumption. We refer to this autarkic use of energy resources as home production.

There are many final products and for tractability we assume they are symmetric substitutes. Home production of all goods is possible, but becomes less and less efficient as the number of goods expands. It may be relatively easy for an individual to grind their own grain, bake their own bread, and brew their own beer but once they expand the set of produced goods to include leather work, shoe manufacture, and the production of candle wax, our jack of all trades becomes increasingly less efficient. This amounts to an assumption of diseconomies of scope under home production.

Into this environment we introduce the potential for agents to agglomerate at a single core labelled C. The core is located somewhere in this space, and we consider the incentives nearby agents have to agglomerate at C. Any agent who reaps energy from nearby landholdings but consumes and trades goods at C is said to agglomerate at C. We think of C as a small village or town with surrounding landholdings being locally owned. The location of C is arbitrary and fixed, but in practice this locational indeterminacy is resolved by small geographic variations creating focal points for settlements. We focus on incentives created by beneficial mutual exchange, but agglomeration for security, administrative, and insurance reasons must also have been very important in earlier times.[14] Historically, we find settlements at the mouth of rivers, on easy to defend plateaus, beside wind breaks, at valley crossings, etc. Any one of these features, even if it generates only small benefits, will make some locations for C superior to others. Propositions 2 and 3 in particular show how such variation can create local advantages in an otherwise featureless plain.

Every agent has a choice. One option is to use the Δ Watts of power generated on their land to produce at home a set of goods they value in utility. We refer to this choice as the *home production option* and index goods produced at home by z. Alternatively, the agent can use their Δ Watts of power to produce their own unique good and transport it to C to exchange for goods produced by other agents who are likewise specialized. We refer to this agglomerate and trade outcome as the *agglomerate option*.[15] Since agents produce distinct goods, we need a way to label them. A simple way to label them is to use distance from the core to identify different goods.[16] That is, good z is the good that agents at distance z from the core are especially proficient at producing; goods $z' \neq z$ are those additional goods obtained via trade in the core. It is also important to distinguish between labels for goods and labels for agents. We will refer to an agent with land holdings at distance r from the core as Agent r. This agent will consume a set of goods labelled by z, and supply one of these goods — their unique good labelled $z = r$ —to the core.

When Agent r produces their unique good and brings it to the core it sells at price $p(r)$.

[14] Major cities today grow and thrive because of the benefits of labor market pooling, knowledge spillovers and input sharing. Their current size is restricted by rising land rents and traffic congestion and not energy constraints per se. This however reflects the very success of our current energy system based on fossil fuels to deliver significant power at low cost over great distances. See our later discussion of Zipf's law since how today's distribution of city sizes may have been shaped by earlier constraints.

[15] The option of agent's consuming their own unique good rather than trading it for other goods in the core will never be optimal unless goods are perfect substitutes.

[16] Using distance as a metric for product differentiation has a very long and noble tradition in economics going back at least to the work of Hotelling, Lancaster, and Salop.

Production of this good is subject to constant returns and, by choice of units, one for one with the Watts collected at distance r. But there are transport costs to moving goods to the core just as there was in the Only Energy Model. The quantity of good r delivered to C by Agent r is $\Delta - cr$. Therefore, if Agent r chooses the agglomerate option, the income they have available to spend at the core is equal to the value of goods they deliver: $p(r)(\Delta - cr)$. Agents have love-of-variety preferences defined over the set of available goods. Let n represent the number (measure) of goods agents can potentially consume, then the utility for an agent with holdings at distance r from the core is given by:

$$u(r) = \left[\int_0^n m(z,r)^\epsilon dz \right]^{1/\epsilon} \text{ where } 0 < \epsilon < 1 \tag{6}$$

where $m(z,r)$ is the consumption of good z by agent r, and $1/(1-\epsilon) > 1$, is the elasticity of substitution between varieties.

3.1 The Home Production Option

When agents rely on home production they choose the number of goods to produce to maximize their utility. Production features increasingly strong diseconomies when production is spread across goods. To capture these diseconomies we introduce a function $\gamma(n)$ which is declining in n with $0 \leq \gamma(n) \leq 1$. We set $\gamma(0) = 1$ and assume there exists a very large, but finite, n^+ such that $\gamma(n^+) = 0$. These assumptions ensure no agent can be a jack-of-all trades without significant productivity losses. Then letting $s(z)$ be the share of each agent's energy endowment used in producing good z at home, we write the quantity of any good z produced at home as $\gamma(n)s(z)\Delta$.

It is now simple to calculate utility under home production. We start by solving for utility conditional on the number of goods produced. Since goods enter utility symmetrically, we must have $s(z) = 1/n$ when there are n goods produced. Since consumption must equal home production it follows that $m(z,r) = \gamma(n)\Delta/n$ for all z. Substituting into (6) and simplifying, we find utility for an agent producing n goods at home, u^H, is given by:

$$u^H = n^{(1-\epsilon)/\epsilon}\gamma(n)\Delta \tag{7}$$

There are two opposing forces determining an agent's optimal n. Agents prefer to diversify since utility is increasing in variety. This benefit of diversification is captured by the power function of n in (7). Working against the benefits of diversification are its costs. These costs as reflected in how γ, and hence productivity, falls and the marginal costs of diversification rise when agents produce many goods. Costs rise increasingly fast if $\gamma''(n) < 0$. Both the marginal costs and benefits of diversification are proportional to spatial productivity because of our constant-returns-within-goods assumption. Under relatively weak assumptions an optimal n^* exists and is unique.[17] At this maximum we can represent an agent's utility as $u^H(\Delta) \equiv [n^*]^{(1-\epsilon)/\epsilon}\gamma(n^*)\Delta$. We note u^H is not indexed by r because utility from home production is independent of location, and n^* is not a function of Δ because of our constant returns assumption.

3.2 The Agglomerate Option

We examine the agglomerate option in two steps. First, we solve for the utility level of a typical Agent r who agglomerates in the core and has access to a given set of n goods. To do so we construct a solution for the complete general equilibrium conditional on n. Second, we find the set of conditions under which agents producing these n goods will agglomerate.

Since preferences are identical across agents, we know that for any two goods z and z'

[17]The first order condition that maximizes (7) is $((1-\epsilon)/\epsilon)n^{-1} = -\gamma'(n)/\gamma(n)$. Any interior maximum is unique because $\gamma'(n) < 0$. One useful parameterization of $\gamma(n)$ is given by $\gamma(n) = 1-d(n)$ where $d(n) = d_0 n^{\delta}$ with $d_0 > 0$ and $\delta > 1$. $d(n)$ captures the increasingly large costs of any one individual diversifying their production plans. It is now simple to show that a rise in diversification costs, d_0 or δ, lowers the optimal n^* and lowers agent's utility; that a maximal n^+ exists; and that greater substitution across varieties leads to agents' diversifying less than otherwise.

sold in the core, consumption by any agent r must be such that

$$\frac{m(z,r)}{m(z',r)} = \left[\frac{p(z')}{p(z)}\right]^{1/(1-\epsilon)} \tag{8}$$

Budget balance for Agent r requires $\left[\int_0^n p(z)m(z,r)dz\right] = I(r)$ where income is $I(r) = p(r)(\Delta - cr)$. Solving the agent's consumption problem allows us to write the indirect utility as:

$$U(r) = I(r)/P \;\text{ where }\; P = \left[\int_0^n p(r)^{\epsilon/(1-\epsilon)}dr\right]^{(1-\epsilon)/\epsilon} \tag{9}$$

Market clearing requires demand equal to supply for every good. The aggregate supply of good z is equal to the aggregate supply of all $2\pi r$ agents living at distance $r = z$ with each agent supplying $(\Delta - cr\ )$ of the r^{th} good. Aggregate supply is then given by $X(z) = 2\pi z(\Delta - cz)$. Denote aggregate demand by all agents as $M(z)$. Then, since preferences are homothetic, market clearing requires for any two goods z and z':

$$\frac{m(z,r)}{m(z',r)} = \frac{M(z)}{M(z')} = \left[\frac{2\pi r(\Delta - cr)}{2\pi r'(\Delta - cr')}\right] \tag{10}$$

Using the solutions for relative supply from (10) in (9) for Agent r allows us to write the maximized utility of a representative agent under the agglomerate option. Denoting this by $u^A(r)$ we find maximized utility can be written as the product of three terms:

$$u^A(r) = \Delta\left[\frac{(1-(c/\Delta)r)^{\epsilon}}{(r)^{1-\epsilon}}\right]\left[\int_0^n [r'(1-(c/\Delta)r')]^{\epsilon}\, dr'\right]^{\frac{(1-\epsilon)}{\epsilon}} \tag{11}$$

The utility from agglomeration comes from three conceptually distinct sources: spatial productivity, location, and city size. The first term in (11) shows utility rises proportionally with the spatial productivity of available resources. The second term in (11) is Agent r specific and tells us that agents closer to the core have higher real incomes and utility. It is useful to note that for any given n we have $u^A(r) < u^A(r')$ if $r' < r$. This second term tells us that location matters. The third term reflects the benefit of greater choice in larger agglomerations. This city size effect is common to all agents, and it tells us that a *ceteris paribus* increase in n raises utility. Together, spatial productivity, location, and city size determine the utility of any prospective Agent r. To solve for the general equilibrium conditional on n, we note that with supplies given in (10), relative prices follow from Agent r's first order condition. Using the fact that spending must equal the value of each agent's delivered supplies, we have a complete solution to the general equilibrium.

3.2.1 The Number of Goods

We start by noting the set of goods available in the city cannot include varieties that are further than $R^* = \Delta/c$ from the core. Energy constraints put a hard limit on maximum city size since any good transported from a greater distance would have zero supply when delivered. Goods produced at points interior to this distance do have strictly positive delivered supply and agents capable of producing those goods may choose to agglomerate. Since we have already shown utility for any agent is decreasing in their distance from the core, we only need to identify the set of agents who are just indifferent between the agglomerate option and the home production option. By construction, such a marginal Agent r would have holdings such that when $r = n$ they are just indifferent between their two options. To find these marginal agents we evaluate (11) at $r = n$ to solve for their utility.

Perhaps surprisingly, we can now show using (11)evaluated at $r = n$ that utility for a marginal agent is hump shaped in n. Since this result is important we record it in a lemma.

Lemma 1 *The utility enjoyed by a marginal agent under the agglomerate option starts at zero when $n = 0$, rises to a single peak, and returns to zero when $n = \Delta/c$.*

Proof. *See Appendix.* ■

When the city (and hence n) is very small, transport costs are very close to zero. This follows from the physics of the problem since costs are linear in distance. But a marginally larger city provides a larger choice set and lower overall price index for consumption at almost identical costs. As a result, the utility of a marginal agglomerating agent initially rises with city size. But as the city grows in size, marginal agents suffer larger transport costs and face greater competition in product markets. Even though a now quite large city provides tremendous variety in consumption, energy constraints soon dominate all other considerations. The utility *for a marginal agglomerating agent* begins to fall and will eventually hit zero.

With Lemma 1 in hand, we can now examine the marginal agent's problem. We start by using a figure to sort out issues of multiple equilibria and stability, and then turn to algebra to examine the determinants of agglomeration more closely. Figure 1 plots the utility of "the marginal agent" under the two options in three different settings. The settings are differentiated by the spatial productivity of the surrounding landscape with $\Delta_L < \Delta^{cr} < \Delta_H$. In each setting, the agent can decide to remain dispersed and engage in home production; or, specialize, agglomerate, and trade. Utility under home production is a constant in this figure and is the same for any potential marginal agent. Therefore, utility for any agent under home production is proportional to Δ and we can represent the utility levels achieved in the three settings (low Δ_L, critical Δ^{cr}, and high Δ_H spatial productivity) by the height of the three horizontal lines as shown. Under the agglomerate option, utility is a single peaked function of the number of goods available in the core but shifts upward with spatial productivity. Since the number of goods available in any core also measures core size, we label the horizontal axis as agglomeration size. Correspondingly, the three hump shaped curves represent utility for a marginal agglomerating agent in settings with low Δ_L, critical Δ^{cr}, and high Δ_H spatial productivity.

It is now simple to examine the agglomerate decision. Consider the first pair of ticked curves associated with Δ_L. In this situation, utility under home production lies everywhere above that for agglomeration. This implies home production dominates agglomeration for any agglomerating agent. No agglomeration occurs. Next consider the second pair of dashed curves associated with Δ^{cr}. As shown, these two curves are just tangent at the point n^{cr}. Inspection of u^A and u^H reveals that u^A responds more than proportionately with an increase

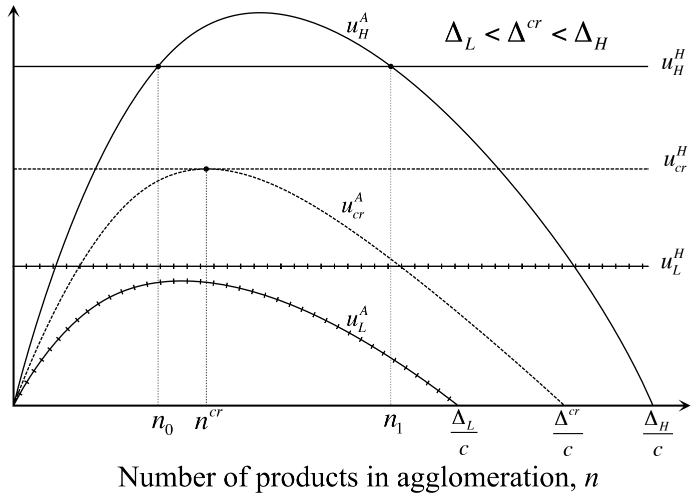


Figure 1: Agglomeration

in spatial productivity. This implies any gap between these curves must shrink as spatial productivity rises above Δ_L. By continuity there must exist a critical value as shown and labelled with cr for critical. Therefore, the agent who is at distance $r = n^{cr}$ from the core would just be willing to agglomerate. Moreover, since Agent n^{cr} is indifferent between options, we know from (11) that agents located at $r < n^{cr}$ strictly prefer the agglomerate option. Therefore, an agglomeration now forms when we raise the spatial productivity to Δ^{cr}.

Finally consider the highest productivity setting with Δ_H. There are now two intersection points: one at n_0 and the other at n_1. Consider n_0. At this point, the agent located at distance $r = n_0$ from the core would be indifferent between agglomerating or not. Again since n_0 is just a particular value for n , we know from (11) that all agents $r < n_0$ would necessarily choose to agglomerate. Therefore, n_0 represents a possible core size with n_0 as the last agent to join the core. Now consider n_1. The observations we just made apply with equal force to this new marginal agent n_1. Therefore n_1 is also a possible core size and it appears we have a situation of multiple equilibria. This is true, but only one of these equilibria is robust to small perturbations. When the core size is n_0 all agents in the range between n_1 and n_0 enjoy the utility associated with home production. In contrast if n_1 was the core size, these same agents would enjoy utility strictly greater than home production utility. Therefore, n_0 is not robust to a small perturbations in the number of agents agglomerating. In contrast, n_1 is robust and we will for the remainder only consider equilibria of this type in our analysis. Putting these results together, we have shown that if the spatial productivity of energy resources is sufficiently high, $\Delta \geq \Delta^{cr}$, agglomeration occurs. If the spatial productivity of energy resources is sufficiently low, $\Delta < \Delta^{cr}$ agents remain dispersed in home production.

While these statements are true, a small bit of algebra allows us to go further and discover an important result not apparent from the figure. The existence of an agglomeration outcome requires a solution for n^* that equates (11) evaluated at $r = n$ with (7). Since Δ cancels from both sides of this equality it is apparent that whether n^* exists or not depends only on the ratio Δ/c. Therefore, the decision to agglomerate depends on what we may call "the comparative advantage of a location" as evidenced by the ratio of spatial productivity to transport costs — Δ/c. Our previous graphical analysis now implies that there exists a critical Δ/c which determines whether agents agglomerate or not. Locations with $\Delta/c \geq [\Delta/c]^{cr}$ lead to agglomerations, those with $\Delta/c < [\Delta/c]^{cr}$ do not.

This reliance on the comparative advantage of a location seems very natural. Transport costs may for example be quite low in a desert or along the arctic tundra because elevation changes are infrequent, but of course the productivity of these locations is also extremely low. Therefore we would expect few if any settlements. In contrast, regions close to the equator are very fertile which we would associate with a large spatial productivity Δ; but these regions often feature inhospitable terrains suggesting higher transport costs as well. Again, we may find few or no settlements despite very high spatial productivity. A region with agglomeration need not have especially high productivity nor low transport costs; but it does need to exhibit a locational comparative advantage (as measured by our theory).

Proposition 4 *If the comparative advantage of a location is sufficiently high, $\Delta/c > (\Delta/c)^{cr}$, then an agglomeration arises. $(\Delta/c)^{cr}$ is determined by the productivity costs of diversifica-*

tion and agent's love of variety.

Proof. *In text.* ■

Proposition 4 links agglomeration to key parameters of the model. First, some form of increasing returns is important as reflected in the shape of $\gamma(n)$. The stronger are the returns to specialization, the more likely an agglomeration arises. Relatedly, as goods become better substitutes (the degree of substitutability across goods is captured by $\sigma = 1/[1-\epsilon]$) agglomeration is less likely since agents can do without the variety benefits a city offers. These results echo earlier work. Second, the likelihood of an agglomeration is dependent on transport costs. Low transport costs, all else equal, raise the likelihood of agglomeration. This is in stark contrast to a typical economic geography model where transport costs are incurred by final goods trade rather than input supply. Lower final goods transport costs typically make it less important to agglomerate; here lower costs raise the prices input suppliers get for energy brought to the city which makes agglomeration more likely. Third, agglomeration is dependent on a location's comparative advantage Δ/c and this measure surely varies widely within countries. This is also in contrast to typical economic geography models where the degree of increasing returns or even transport costs themselves do not vary within countries. Finally, absolute advantage as captured by Δ is irrelevant to agglomeration. More productive energy resources do not spur agglomeration unless they also come with proportionately lower transport costs. To understand these and other empirical implications of our theory we now turn to place them into the related empirical literature.

4 Empirical Implications

4.1 Lumpiness

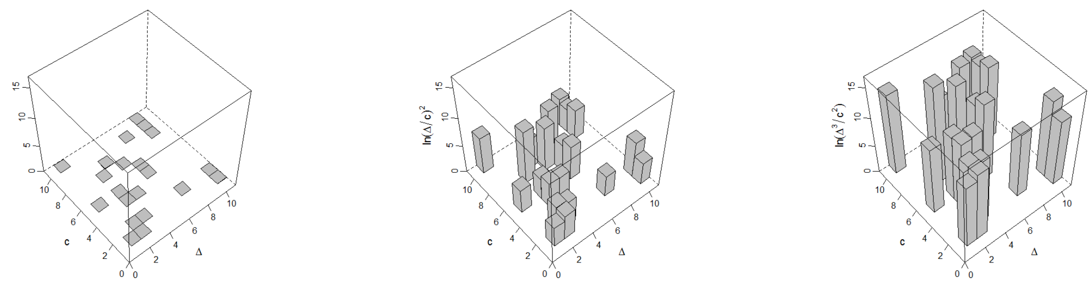
Perhaps the most central fact of economic geography is that human settlements are unevenly distributed across geographic space. For example, the G-ECON dataset built by Nordhaus and coauthors (see Nordhaus (2006)) shows that fully 85% of the world's GDP is produced within 10% of the land area, and this lumpiness of people and production is a feature of all countries. These statistics are constructed from recent data (1990s and later), but to the extent that we can trust population estimates from earlier periods, lumpiness in economic activity appears to be a feature of almost all known history.[18]

Our theory produces lumpiness across geographic space very naturally, whereas other approaches have real challenges confronting this fact. For example, most of urban economics model limits to city size via congestion costs and hence model distance and sometimes geographic space within cities, but there is no sense of geographic space across or between cities. Alternatively, models of economic geography almost always ignore geographic space in theoretical representations where locations are fixed points interchangeably referred to as cities, regions or countries. Our formulation in contrast has zero dimensional cities or cores, but real geographic space around cores and by implication — since cities cannot reap energy resources from the same area — between cities as well. As a result, lumpiness arises in three different, but related ways. First, Proposition 4 identifies a critical value of comparative advantage $(\Delta/c)^{cr}$ that divides locations into those which produce agglomeration and those that do not. If we think natural advantages vary continuously across geographic space, then smooth and continuous variation in natural conditions will result in a landscape punctuated by agglomerations.

To visualize how variation in these natural attributes can create lumpiness across geographic space, consider the three panels of Figure 2 below. To construct the panels we drew 100 Δ and 100 c from uniform distributions. We then paired the draws, calculated the ratio Δ/c, and associated them with specific squares in panel (a). If the square contains a ratio sufficiently large to produce an agglomeration it is black, and by varying the minimum ratio Proposition 4 tells us is necessary for agglomeration we can make the checkerboard

[18]One formal measure of lumpiness is the Spatial Gini Coefficient which measures inequality in the distribution of economic activity across geographic space. As with the typical Gini coefficient a measure of zero implies an equal distribution whereas 1 represents completely unequal distribution. Again using the G-ECON data, the Spatial Gini Coefficient for the world is 0.9 using GDP per unit area as the metric. Country specific measures (from Ramacharan (2009) Table 1) are, for the US 0.83, for Canada 0.91, for France 0.61, for Germany 0.57, and for the island country of Jamaica 0.32.

more black or more white.[19] A simple measure for urbanization might be the fraction of dark squares in panel (a). An increase in the cost of diversification, $\gamma(n)$, for example, would lower the necessary cut off while an increase in the absolute advantage of all locations (holding comparative advantage constant), would have no effect on the pattern of agglomeration. Such a change however makes the entire geographic space more productive and every agglomerating agent richer, and therefore shows how the geographic pattern of settlement should be independent of income levels.



(a) (b) (c)

Figure 2: Agglomeration and Lumpiness Across Space

4.2 Populations

While it is tempting to take the fraction of black squares to be a very simple measure of urbanization, this very simple depiction underestimates the extent of lumpiness across space because it ignores population differences across agglomerations. To address this issue we solve for equilibrium population sizes for each black square under two different assumptions about how populations respond to agglomeration. To start we assume population growth or migration is unaffected by the process of agglomeration, despite the fact that it raises real incomes considerably. Using this assumption we start by recalling that our initial setting

[19] We picked the threshold using the empirical distribution of draws where Δ is $U(1,2)$ and c is $U(0,1)$. The threshold is simply the mean of (Δ/c). Nothing important hinges on the choice since this is a visualization and not a simulation.

was one with a uniform density of individuals across geographic space. Recall also that for simplicity, we chose units so that *one agent* was resident on each one meter square of land, $\Omega = 1$. The first impact of agglomeration is then to concentrate these individuals in the core. Its apparent then that population in any core will reflect the equilibrium n^* as determined by tastes, technologies and comparative advantage. If we focus on differences in population arising from variation in the productivity of nearby energy resources, we can prove:

Proposition 5 *The population size of any agglomeration rises faster than the square of spatial productivity.*

Proof. *See Appendix.* ■

When spatial productivity rises, the utility benefits of agglomerating rise and this implies core expansion. This core expansion is very similar to the expansion on the extensive margin in the Only Energy Model, but here the impact of higher productivity is slightly stronger. It is stronger (greater than power 2) because expansion brings a greater variety of products to the core, and this makes agglomerating even more attractive and therefore cities larger. Not surprisingly, as goods become better substitutes this variety effect diminishes, and in the limit as ϵ approaches 1 the response of population approaches the square of spatial productivity. To visualize this extensive margin effect we have presented in the panel (b) of Figure 2,but now the height of any black square is proportional to the square of comparative advantage.[20]

Naturally, the distribution of population is much more extreme than the distribution of black and white squares. There are still rural (white squares) and urban areas (black squares), but urban areas now feature both big and small cities. Therefore, the second way our theory generates lumpiness is by showing how city size should scale with spatial productivity (for given transport costs). But even this second panel underestimates the lumpiness we could expect under our theory. Recall that an increase in spatial productivity

[20] To make these figures we have chosen to use the limiting case where ϵ approaches one. As ϵ approaches one, n becomes proportionate to Δ/c implying that city population scales with the square of Δ just as net energy supply did in the Only Energy Model.

creates both an extensive and intensive margin effect; therefore scalar increases in Δ and c, leave the extensive margin unaffected but generate more energy (and real income) for all city residents. If we assume that population growth (or migration) expands to partially or fully dissipate the real income gains brought by agglomeration, then city sizes will be affected by this intensive margin affect. To see its full potential suppose Malthusian population growth is operative (this is only one of several possible mechanisms). Recall the density of individuals across the landscape was $\Omega = 1$. If we take this existing density to represent a Malthusian steady state prior to the agglomeration option, then $u^H/\Omega = Z$ where Z is the real income per agent that sets births equal to deaths, and u^H is real income. When agents agglomerate and a core forms, this creates large income gains for agents who are not marginal. If population growth subsequently dissipates these gains, then the density of agents must adjust so that agents r within the core would have a density satisfying $\Omega^A = u^A(r)/Z$. And using (11), the total core population becomes:

$$Pop^C = \frac{2\pi\Delta}{Z}\int_0^n r'\Omega(r')dr' = \frac{2\pi\Delta}{Z}\left[\int_0^n \left[r'(1-(c/\Delta)r')\right]^{\epsilon} dr'\right]^{1/\epsilon} \tag{12}$$

and we have a further implication of our scaling law:

Proposition 6 *If population growth (or migration) dissipates the real income gains from agglomeration, then the population of any agglomeration rises faster than the cube of spatial productivity* Δ.

Proof. *See Appendix.* ■

This result can be visualized by inspecting panel (c) of Figure 2. In this panel populations are now proportional to the cube of the underlying spatial productivity, and as shown the distribution of city sizes is now wider. Therefore, the third way our theory generates lumpiness in economic activity across geographic space is by transforming variation in natural conditions into real income gains that are subsequently dissipated by further entry by migrants or new births.

4.3 Persistence and Location

Two other key facts of economic geography are that agglomerations are almost always located near discrete variations in the natural landscape such as valleys, rivers and coastlines; and that agglomerations themselves are highly persistent. Empirically, the link with water access is especially well documented. For example, Rappaport and Sachs (2003) find that in the year 2000 US coastal counties comprise 13% of the land mass, but 57% of economic income and 51% of the population. Nordhaus (2006) also reports economic activity tightly tied to water access. The key study with regard to persistence is Davis and Weinstein (2002). They show that the geographic distribution of the Japanese population across 39 prefects has been very stable for almost 8,000 years. The rank and raw correlations across these many millennia often exceed 0.8. Moreover, they also show that a major shock to built up cities, factories and populations, the Allied bombing of Japanese cities during WWII, had only a temporary and not permanent effect on the distribution of Japanese population. Their work suggests a strong degree of permanence in the geographic distribution of populations, at least for Japan. While it appears that Japan may be somewhat special because of its interior mountains and coastal nature, examples of persistence abound. For example, every one of the world's largest cities since 1000 AD would be well known to any well travelled individual in the 21st century,[21] and while substantial turnover in leading cities within countries does occur (witness the fall of Detroit, Baltimore, Cleveland and St. Louis, and the rise of Atlanta, Dallas, Houston and Miami in the U.S.) there is tremendous geographic persistence. In the US for example, 16 of the 20 most populous cities in 2010, and 16 of 20 most populous in 1850 are all located on coasts or major waterways (the Great Lakes).[22] Therefore, while the fortune of individual cities has risen and fallen, the location of economic activity has remained tightly tied to geographic advantage for literally thousands of years.

[21] If you have been to London, Istanbul, Beijing (Peking), New York and Tokyo you have been to all of the world's largest cities over this period.

[22] For example in 2010, of the top 20 largest US cities only Dallas, Phoenix, Minneapolis, and Denver are not on coasts or the great lakes.

These features would be easily explained by our theory if the benefits of an unchanging geography that lowered transport costs $c(\theta)$ led to agglomeration. While this seems a reasonable conjecture, to prove it requires us to identify the set of marginal agglomerating agents along the transport corridor and and in the hinterland. Keeping track of all these agents and considering their decisions to agglomerate or not requires significant work that we leave to the Appendix. Here we just record the basic result:

Proposition 7 *Consider a setting where agents are dispersed in home production because agglomeration is not possible given existing conditions. If we introduce into this setting a river, road, or natural transportation corridor offering a sufficiently large cost advantage , then agents with access to the corridor will agglomerate along it.*

Proof. *See Appendix.* ■

4.4 Adding up across geographic space and national populations

Finally, how do our predictions add up over space and populations to generate predictions for larger regions or countries? While a complete analysis of this problem would surely involve a discussion of strategic location decisions, the effects of population and economic growth, and perhaps even historical accident, we instead offer a simple, but useful means for understanding how the main features of our theory provides the building blocks for such an analysis. To do so, it proves useful to divide geographic space into *uniform landscapes* which are homogenous geographic regions where all locations are equally attractive.[23] Formally, a uniform landscape is defined by a common set of our four model parameters plus an area A: that is, the set $(\Delta, \Delta/c, \gamma(n), \epsilon, A)$. Apart from very small countries or islands, real countries are composed of many such landscapes. We think of uniform landscapes as being populated with many cores; and many (different) uniform landscapes comprising political units such as countries. To make this idea concrete, we depict a set of four uniform landscapes in the

[23] In ongoing and related work, Food, Fuel and the Rise of Cities (Moreno-Cruz and Taylor, 2016), we introduce a series of landscapes including those with rivers, coastlines etc. For brevity and simplicity we discuss only uniform landscapes here.

top boxes of Figure 3, and then combined bits and pieces of these landscapes in the four boxes on the second row that define a hypothetical country. The uniform landscapes in the first row differ only in their spatial productivity. Each step from right to left represents a doubling of the landscape's spatial productivity.

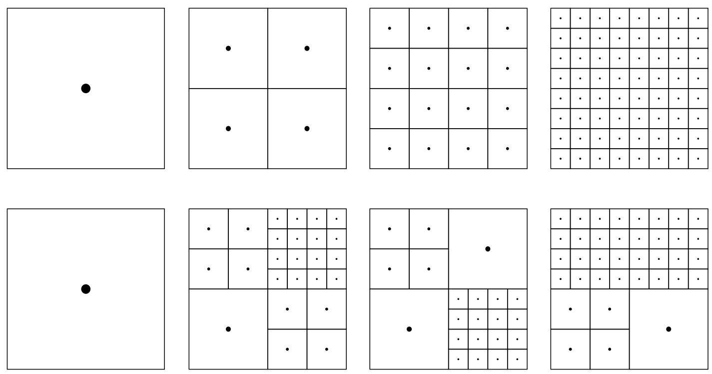


Figure 3: Cities and Regions

The fiction of uniform landscapes gives us a means to impose adding up constraints over space to see what predictions these constraints, together with our earlier theory, provide at the landscape and country level. To proceed recall that Proposition 5 tells us the exploitation zone for any successful core expands at a rate greater than the square of spatial productivity. Therefore, (using an approximate solution as ϵ approaches 1 to simplify) we can see that the first rightmost square contains 64 cores, the next contains only 16, then 4 and then 1. Core numbers fall with the square of spatial productivity. And from Proposition 6, we know that if populations in these cores are responsive to the real income gains brought about by their agglomeration, then the populations of these cores must rise at a rate greater than the cube of spatial productivity. Therefore, city populations rise at least by a factor of 8 with each step from right to left. If we now put these observations together, it implies that a uniform landscape's total population rises only proportionally with spatial productivity.[24] Regions with more productive energy resources will exhibit fewer but bigger cities. If we look across these regions differentiated solely by productivity we see a world that looks very constant

[24] We are grateful to David Stern for pointing out this implication of our work in a slightly different context.

returns; but if we compare individual cores within these regions we will see vast differences. All of these results are direct, but surprising implications of our scaling law that manifests itself differently at the core and region level.

While these are interesting (and perhaps to some readers, curious) implications of uniform landscapes, few countries are well represented by a single uniform landscape. We imagine instead that the data we obtain almost always comes from political units that look much more like the four boxes shown in the second row. And if empirical researchers cannot first identify and then condition on a landscape's unique characteristics to neatly divide this country back into uniform landscapes, we can not say anything about the likely distribution of city sizes it contains. But, is there a further implication of our scaling law for the distribution of city sizes?

It should be clear that our theory — without significant further assumptions — does not constrain the distribution of city sizes in any meaningful way. It does however present a simple means for understanding how city size distributions may be linked to our theory's primitives. To see how this occurs, consider Figure 4 below.

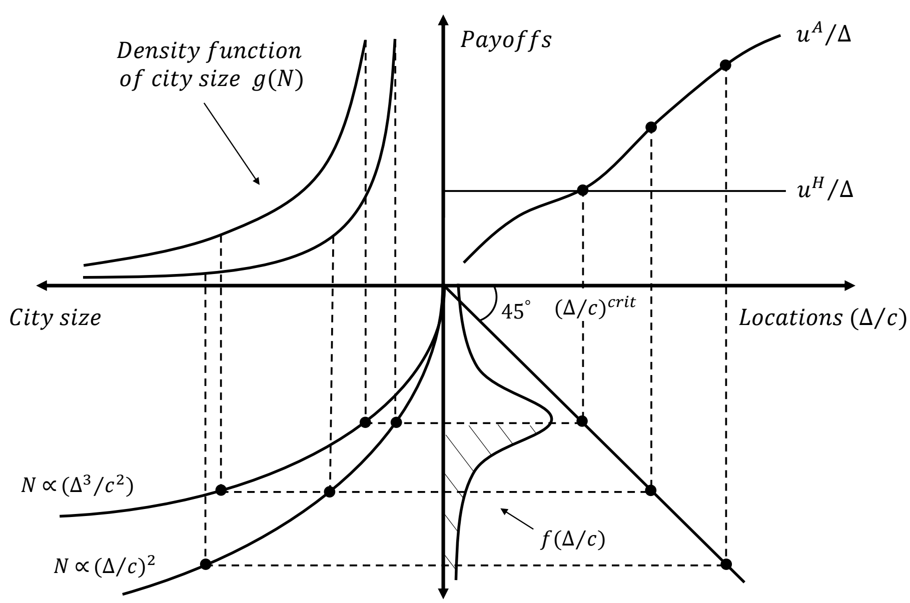


Figure 4: City Size Distribution

In Figure 4 we have constructed two potential city size distributions in quadrant II, by combing elements of our theory with an assumed distribution of comparative advantage (shown in quadrant IV). To construct the figure we assume agents everywhere have access to the same technologies and they share the same tastes, but agents are situated at locations in our hypothetical country which may differ in their comparative, and absolute, advantage as measured by Δ/c and Δ. To construct the city size distributions we start in quadrant I. Using Proposition 4, we know there exists a single critical level of comparative advantage determining whether agents agglomerate or not. If we then order our country's potential locations in terms of increasing comparative advantage along the horizontal axis, we can represent the maximal utility a marginal agglomerating agent may achieve under any particular Δ/c by the increasing curve labelled u^A/Δ. This maximal utility is increasing in Δ/c and just equals that available under home production when agglomeration becomes the chosen result. We have shown this by the intersection of the upward sloping line labelled u^A/Δ and the single horizontal line labelled u^H/Δ. u^H/Δ, given by (7), is a constant in this figure because its maximized value is only a function of taste and technology parameters, but not Δ nor c. In contrast, u^A/Δ is strictly increasing because the maximal values for utility shown in Figure 1 move up and to the right with spatial productivity. Therefore, the first quadrant shows how the agglomeration decision of many agents across a country are resolved with agents at locations where $\Delta/c \geq (\Delta/c)^{cr}$ choosing agglomeration, and agents at all other locations remaining dispersed in home production.

If we now follow the critical value $(\Delta/c)^{cr}$, downward to quadrant IV, and reflect it by use of a 45-degree line we find the portion of the distribution of comparative advantage in this country, $f(\Delta/c)$, that will be relevant to the city size distribution. We have shaded the relevant portion of this distribution; for simplicity we have made no particular assumptions about $f(\Delta/c)$ other than it is symmetric, continuous, and positive. Locations with poor values for comparative advantage become cities with probability zero; but for locations with better fundamentals, a city will form. To find how frequent we might expect cities of a certain type, we start by reading from f the relevant probability mass associated with any location's comparative advantage. But since any agent's agglomeration decision has effectively truncated f, we need to scale up the probability mass associated with a successful agglomeration by $1/[1 - F(\Delta/c)^{cr}]$. This ensures that the eventual city size distributions shown in quadrant II integrate to one. With this complication in hand, in quadrant III we map any given Δ/c into a city population size. Then taking the (correctly scaled) probability mass associated with any given Δ/c from quadrant IV, and associating it with the resulting city size in quadrant III, generates the implied city size distributions shown in quadrant II. Repeating this exercise for the remaining values traces out the complete distributions. As shown the distributions exhibit a minimum city size, a long right tail, and are monotonically declining. Any resemblance to the Pareto family of distributions is purely intentional.

To understand why we have two possible city size distributions we need to recall that both comparative and absolute advantage matter. Therefore, our distribution f cannot, without further assumption, completely determine the distribution of city sizes. Two limiting cases are instructive, and these lead to the two city size distributions shown. Suppose first that absolute advantage (spatial productivity) was the same across all locations, so that $f(\Delta/c)$ reflected only variation in transport costs. By construction then, the correlation between absolute and comparative advantage across locations is zero; and, any two cities differ only in comparative advantage. The city size distribution would be that given by the lower curve in quadrant II where population sizes are approximately proportional to the square of comparative advantage. This result is very similar to Proposition 5.

In contrast, suppose transport costs were the same across all locations and only variation in Δ was captured in f. Now locations differ in both absolute and comparative advantage but they are now perfectly positively correlated. In this case, the city size distribution is given by the uppermost curve in the figure. Populations are responding across cities due to both changes in the extensive (comparative advantage) and intensive margin (absolute advantage) as in Proposition 6. In reality we expect the correlation between comparative and absolute advantage to be less than perfect, but this just means our the two city size distributions are bounds on these other cases.

Although this construction is somewhat lengthy it is productive in showing how our theory's agglomeration decision (Proposition 4), and scaling relationships (Propositions 5 and 6), together with further assumptions, generate an equilibrium city size distribution. Several features of this construction are noteworthy. One robust feature of the construction is the truncation of the city size distribution so there are no very small cities. The extent of this truncation depends on the solution to the optimization problem in quadrant I; for example, if goods were better substitutes or if diversification less costly, then home production becomes more attractive and minimum city size grows. Another robust feature is that the scaling relationships created by the spatial structure of our model lengthens the tail of the city size distribution. Here again our scaling law shapes outcomes in somewhat surprising ways.

What is not robust about this construction is that the city size distributions as shown are falling throughout but they could at first rise steeply and then only fall with larger city sizes. Since empirical research often focusses on larger cities, this seems inconsequential. We have also implicitly assumed agents outside option is to remain dispersed in home production rather than move across cores. This seems like a natural first step since resources are often immobile, and we have a one factor model. Despite these limitations, the figure usefully summarizes how our theory, together with additional assumptions on the distribution of energy resources over space, can generate a city size distribution very much like that found in the empirical literature studying Zipf's law.

5 Conclusions

We set out a simple spatial model in order to ask how the location and productivity of energy resources affects the distribution of economic activity around the globe. Our major contribution is the introduction of a new approach to explaining the world's economic geography by focussing on the supply conditions for one, critical, and essential, input — energy. We found that adopting an explicitly spatial setting implied a scaling law linking the (spatial) productivity of energy resources to potential energy deliveries at any point. Even small differences in their spatial productivity produced large differences in available energy at any location. We then embedded this simple model of energy supply into a conventional market model where differentiated products, a return to specialized production, and a need for energy in production drove a desire to agglomerate.

Combining elements from economic geography and energy economics proved quite fruitful. We showed how the geographic pattern of agglomeration is driven by a location's comparative advantage, and when the comparative advantage of a location is sufficiently high, agents concentrate in our core, specialize, trade, and reap large income gains from doing so. This implies that geographic variation in our model's measure of comparative advantage, produces lumpiness in the geographic location of economic activity. When a region can support agglomeration, we found that the size of agglomerations scaled with the productivity of nearby energy resources. This scaling of populations then implies still greater lumpiness in the geographic location of production. And we showed how geographically unique locations, such as river and coastal locations not only provide extremely persistent motives for agglomeration, by virtue of our scaling law they also generate extremely large agglomerations.

Finally we showed how our model of a single core could add up to provide predictions at a (homogenous) regional or (many region) national level. Richer regions feature larger cities, higher population densities, and higher income agents. Although city sizes scale with the cube of energy productivity, a region's overall income or population responds only proportionately. Most real geographic space is of course comprised of many locations that differ quite markedly. And in a geographic space containing many such locations, we showed how variation in comparative advantage together with our scaling law provides a simple theory of city size distributions. The distribution is truncated with no very small cities;

it has a long right tail with few very large cities; and its exact distribution flows from the interplay of nature's distribution of comparative advantage across geographic space and our model's generated scaling law relationships.

Many of these results call out for further investigation. There is still much work to be done.

6 References

Allen, Robert C. 2009. "The British Industrial Revolution in Global Perspective." UK: Cambridge University Press

Brock, William, Gustav Engstromy and Anastasios Xepapadeas 2012. "Energy Balance Climate Models and the Spatial Structure of Optimal Mitigation Policies" Mimeo

Brock, William, Anastasios Xepapadeas and Athanasios Yannacopoulos 2014. "Optimal Control in Space and Time and the Management of Environmental Resources" Working Paper Series Athens University of Economics and Business.

Chakravorty, U., J. Roumasset, and K. Tse. 1997. "Endogenous substitution among energy resources and global warming." *Journal of Political Economy*, 195: 1201-1234.

Davis, D. and D. Weinstein. 2002. "Bombs, bones and break points: the geography of economic activity", *American Economic Review*, 92 (5): 1269-1289

Decker EH, Kerkhoff AJ, Moses ME U. 2007. "Global Patterns of City Size Distributions and Their Fundamental Drivers". *PLoS ONE* 2(9): e934. doi:10.1371/journal.pone.0000934

Desmet, Klaus, D. K. Nagy, and E. Rossi-Hansberg 2015 "The geography of development: evaluating migration restrictions and coastal flooding," National Bureau of Economic Research w21087.

Fernihough, A. and K.H. O'Rourke. 2014. "Coal and the European Industrial Revolution", National Bureau of Economic Research Working Paper No. 19802j.

Fouquet, R. and R. Pearson. 1998. "A Thousand Years of Energy Use in the United

Kingdom." *The Energy Journal*, 19(4): 1-41.

Gaudet, G., M. Moreaux, and S.W. Salant. 2001. "Intertemporal Depletion of Resource Sites by Spatially Distributed Users." *American Economic Review*, 91(4): 1149-1159.

Head, Keith and Thierry Mayer 2004. "The Empirics of Agglomeration and Trade" Chapter 59 in *Handbook of Regional and Urban Economics, Volume 4* edited by J.V. Henderson and J.F. Thisse, Elsevier B.V.

Helpman, E. 1995. "The Size of Regions." *The Foerder Institute for Economic Research* Working Paper #14-95.

Krugman, P. 1991. "Increasing Returns and Economic Geography." *The Journal of Political Economy,*, 99 (3):483?499.

Kolstad, C.D. 1994. "Hotelling Rents in Hotelling Space: Product Differentiation in Exhaustible Resource Markets." *Journal of Environmental Economics and Management*, 26: 163-180.

Laffont, Jean-Jacques and Moreaux, Michel. 1986. "Bordeaux Contre Gravier: Une Analyse par les Anticipations Rationnelles," in Gerard Gaudet and Pierre Lasserre, eds., *Ressources naturelles et theorie economique.* Quebec: Presses de l'Universite Laval, 1986, pp. 231- 53.

Moreno-Cruz, J. and M.S. Taylor. 2012. "Back to the Future of Green Powered Economies," *The National Bureau of Economic Research*, Working Paper No: 18236.

Nordhaus, W. D. 2006. "Geography and Macroeconomics: New Data and New Findings," *Proceedings of the National Academy of Sciences (US)*, 103(10): 3510-3517.

Nunn, N. and N. Qian. 2011. "The Potato's contribution to population and urbanization: evidence from a historical experiment." Quarterly Journal of Economics, 126, 593-650.

Pindyck, R.S. 1978. "The Optimal Exploration and Production of Nonrenewable Resources." J*ournal of Political Economy*, 86(5): 841-861.

Rosenthal, Stuart, S and William C. Strange 2004. "Evidence on the Nature and Sources of Agglomeration Economies" Chapter 49 in *Handbook of Regional and Urban Eco-*

nomics, Volume 4 edited by J.V. Henderson and J.F. Thisse, Elsevier B.V.

Severnini, E.R. "The Power of Hydroelectric Dams: Agglomeration Spillovers", IZA Discussion Paper No. 8082..

Sanchirico, J.N. and J.E. Wilen. 1999."Bioeconomics of Spatial Exploitation in a Patchy Environment." *Journal of Environmental Economics and Management*, 37:129-150. **Smil, V.** 2008. *Energy in Nature and Society: General Energetics of Complex Systems.* USA: MIT Press.

Swierzbinski, J.E. and Mendelsohn, R. 1989. "Exploration and Exhaustible Resources: The Microfoundations of Aggregate Models." *International Economic Review*, 30(1): 175-186.

Wrigley, E.A. 2010. *Energy and the English Industrial Revolution.* Cambridge University Press. Cambridge, UK.

Zeeuw, J.W. 1978. "Peat and the Dutch Golden Age: the Historic meaning of energy attainability", in Afdeling Agrarische Geschiedenis Landbouwhogeschool, The Netherlands.

Appendix

A Proofs to Propositions

A.1 Proof to Proposition 1:

If transportation costs vary with the direction θ, then we can identify the maximum radius in each direction θ as $R(\theta) = \frac{\Delta}{c(\theta)}$. To show that gross power is homogeneous of degree 3 we calculate the following integral:

$$W^* = \int_0^{2\pi} \int_0^{R(\theta)} r\Delta dr d\theta = \frac{\Delta^3}{2} \int_0^{2\pi} \frac{1}{c(\theta)^2} d\theta \tag{A.1}$$

which shows that for any function $c(\theta)$ gross power is a function homogeneous of degree 3 in power density. To show the same is true for power supplied, we have:

$$W^S = \int_0^{2\pi} \int_0^{R(\theta)} r(\Delta - c(\theta)r) dr d\theta = \frac{\Delta^3}{6} \int_0^{2\pi} \frac{1}{c(\theta)^2} d\theta$$

This shows power supplied W^S is homogeneous of degree three in power density, but the precise shape of the exploitation zone is determined by the form of $c(\theta)$.

A.2 Proof to Proposition 3:

Assume all roads have the same coefficient of friction ρc with $0 < \rho < 1$. Optimal deployment of the road system requires that roads are built to maximize coverage. That is, roads will split the space in equal parts. The first road, as we have assumed, would split the space in π radians, then second road would split it in $\pi/2$ radians, the third road in $\pi/4$ radians, and so on. Let n denote the number of roads. For given ρ, if $n < \bar{n}$ where $\bar{n} \equiv \frac{\pi}{2\arccos(\rho)}$ then the exploitation zones added by each road do not overlap. If $n > \bar{n}$ the exploitation zones will overlap. For $n < \bar{n}$ we have:

$$W^S = 4 \times \left[n \left[\int_0^{\bar{\theta}} \int_0^{r^*} v\left(\Delta - c(\theta)v\right) dv d\theta + \int_{\bar{\theta}}^{\pi/2n} \int_0^{r^*} v\left(\Delta - cv\right) dv d\theta \right] \right]$$

where we have exploited symmetry in the first quadrant of the cartesian space. The integral is then given by:

$$W^S = \frac{1}{3}\frac{\Delta^3}{c^2} g(\rho, n), \text{ where } g(\rho, n) = \pi + 2n(\tan(\bar{\theta}) - \bar{\theta})$$

It is easy to see now that power supplied is linear in the number of roads, n. For $n > \bar{n}$, the exploitation zones will intersect at odd multiples of $\pi/2n$. The expression for power supplied is now given by:

$$W^S = 4 \times \left[n \left[\int_0^{\frac{\pi}{2n}} \int_0^{r^*} v\left(\Delta - c(\theta)v\right) dv d\theta \right] \right]$$

$$W^S = \frac{1}{3}\frac{\Delta^3}{c^2} g(\rho, n), \text{ where } g(\rho, n) = 2n \int_0^{\frac{\pi}{2n}} \left((1-\rho^2)^{1/2} \sin\theta + \rho\cos\theta\right)^{-2} d\theta$$

To show that power supplied is increasing and concave in the number of transportation options when $n > \bar{n}$ we take derivatives with respect to n and show $\frac{\partial W^S}{\partial n} = \frac{2}{3}\frac{\Delta^3}{c^2}\left[\frac{g(\rho,n)}{2n} - \frac{\pi}{2n}\left((1-\rho^2)^{1/2}\sin\frac{\pi}{2n} + \rho\cos\frac{\pi}{2n}\right)^{-2}\right] > 0$ and $\frac{\partial^2 W^S}{\partial n^2} = \frac{-\Delta^3\pi^2\left(\sqrt{1-\rho^2}\cos\left(\frac{\pi}{2n}\right) - \rho\sin\left(\frac{\pi}{2n}\right)\right)}{3c^2n^3\left(\sqrt{1-\rho^2}\sin\left(\frac{\pi}{2n}\right) + \rho\cos\left(\frac{\pi}{2n}\right)\right)^3}$ which is negative when $\sqrt{1-\rho^2}\cos\left(\frac{\pi}{2n}\right) - \rho\sin\left(\frac{\pi}{2n}\right) > 0$ or $n > \frac{\pi}{2\arccos(\rho)} = \bar{n}$.

A.3 Proof to Lemma 1:

There are three elements to this proof. First, we need to show that u^A at $n = 0$ is zero. Second, we need to show that u^A at $n = \Delta/c$ is zero. Third, we need to show there is a single peak. We proceed in order. Define the utility in agglomeration as:

$$u^A(n,\Delta) = \Delta f(n,\Delta) g(n,\Delta), \text{ where} \tag{A.2}$$
$$f(n,\Delta) = \frac{(1-\frac{cn}{\Delta})^\epsilon}{n^{(1-\epsilon)}} \text{ and } g(n,\Delta) = \left(\left(\frac{\Delta}{c}\right)^{1+\epsilon} B_{\frac{cn}{\Delta}}(1+\epsilon,1+\epsilon)\right)^{\frac{1-\epsilon}{\epsilon}}$$

with $B_{cn/\Delta}(1+\epsilon,1+\epsilon) = \int_0^{cn/\Delta} t^\epsilon (1-t)^\epsilon dt$.

Both the denominator and the numerator approach zero as n approaches zero. Applying l'Hospital rule we find

$$\lim_{n\to 0} \frac{(1-\frac{cn}{\Delta})^\epsilon \left(\left(\frac{\Delta}{c}\right)^{1+\epsilon} B_{\frac{cn}{\Delta}}\right)^{\frac{1-\epsilon}{\epsilon}} \left(-\frac{c}{\Delta}\epsilon(1-\frac{cn}{\Delta})^{-1} + \frac{(1-\epsilon)}{\epsilon}\left(\left(\frac{\Delta}{c}\right)^{1+\epsilon} B_{\frac{cn}{\Delta}}\right)^{-1} \frac{\partial B_{cn/\Delta}}{\partial n}\right)}{n^{-\epsilon}} = 0$$

To show that $\lim_{n\to(\Delta/c)} u^A(n,\Delta) = 0$ we can simply replace $n = \Delta/c$ to find the result. The numerator is zero and the denominator is a real number.

Next, we show there is a single peak. A maximum is found at a point where $\partial u^A/\partial n = 0$. Taking derivatives and setting them equal to zero we find the maximum occurs where $-\varepsilon_{fn} = \varepsilon_{gn}$ where

$$\varepsilon_{fn} = \frac{f_n n}{f} = -(1-\epsilon) - \frac{cn\epsilon}{\Delta - cn} < 0 \text{ and } \varepsilon_{gn} = \frac{g_n n}{g} = \frac{1-\epsilon}{\epsilon} \frac{\left(\frac{cn}{\Delta}\right)^{\epsilon+1}\left(1-\frac{cn}{\Delta}\right)^\epsilon}{B_{\frac{cn}{\Delta}}(1+\epsilon,1+\epsilon)} > 0 \tag{A.3}$$

It can be shown that $\frac{\partial \varepsilon_{fn}}{\partial n} < 0$ and $\frac{\partial \varepsilon_{gn}}{\partial n} < 0$. Moreover, $\varepsilon_{fn}|_{n=0} = -(1-\epsilon)$ and $\varepsilon_{gn}|_{n=0} = (1-\epsilon)/\epsilon$. So $-\varepsilon_{fn}$ starts at a number $(1-\epsilon) > 0$ and increases monotonically in n , while ε_{gn} starts at $(1-\epsilon)/\epsilon < 1-\epsilon$ and decreases monotonically in n. The combination of continuity and the intermediate value theorem show a crossing exists and it is unique. Thus, u^A is characterized by a single peak.

A.4 Proof to Proposition 5:

Define the utility in home production as evaluated at the optimal number of goods as: $u^H(\Delta) = K\Delta$ In equilibrium we require $u^A = u^H$ where u^A is defined in equation (A.2). There are two possible equilibria, but we have shown in the text that the only logical equilibrium is the one to the right of the peak of u^A. To see how the equilibrium number of products moves with n we take total derivatives in both sides of the equilibrium equation to find:

$$(fg + \Delta f_\Delta g + \Delta f g_\Delta)d\Delta + \Delta(f_n g + f g_n)dn = K d\Delta \tag{A.4}$$

$$(1 + \varepsilon_{f\Delta} + \varepsilon_{g\Delta})\frac{d\Delta}{\Delta} + (\varepsilon_{fn} + \varepsilon_{gn})\frac{dn}{n} = \frac{d\Delta}{\Delta} \tag{A.5}$$

where

$$\varepsilon_{f\Delta} \equiv \frac{f_\Delta \Delta}{f} = \frac{cn\epsilon}{\Delta - cn} > 0 \text{ and } \varepsilon_{g\Delta} \equiv \frac{g_\Delta \Delta}{g} = \frac{(1-\epsilon)(1+\epsilon)}{\epsilon} - \frac{1-\epsilon}{\epsilon}\frac{\left(\frac{cn}{\Delta}\right)^{\epsilon+1}\left(1-\frac{cn}{\Delta}\right)^{\epsilon}}{B_{\frac{cn}{\Delta}}(1+\epsilon, 1+\epsilon)} > 0 \tag{A.6}$$

and ε_{fn} , ε_{gn} are given in equation (A.3). From equation (A.5) we find the elasticity of n with respect to Δ is given by:

$$\frac{dn}{d\Delta}\frac{\Delta}{n} = -\frac{\varepsilon_{f\Delta} + \varepsilon_{g\Delta}}{\varepsilon_{fn} + \varepsilon_{gn}} \tag{A.7}$$

Detailed observation of the definitions of the different elasticities allows us to write the following expressions $\varepsilon_{fn} = -(1-\epsilon) - \varepsilon_{f\Delta}$ and $\varepsilon_{gn} = \frac{(1-\epsilon)(1+\epsilon)}{\epsilon} - \varepsilon_{g\Delta}$. We can replace these expressions back in equation (A.7) to find

$$\varepsilon_{n\Delta} \equiv \frac{dn}{d\Delta}\frac{\Delta}{n} = \frac{(\varepsilon_{fn} + \varepsilon_{gn}) - \frac{(1-\epsilon)}{\epsilon}}{(\varepsilon_{fn} + \varepsilon_{gn})} > 1 \tag{A.8}$$

The inequality follows from recognizing that the equilibrium of interest is to the right of the hump, which implies $-\varepsilon_{fn} > \varepsilon_{gn}$ so both the denominator and the numerator are negative numbers, but the numerator is a larger number, in absolute value, than the denominator. If population increases proportionately with the area of agglomeration, the radius of the agglomeration is here determined by the marginal agent located at a distance $r = n$ and total population is $Pop = \pi n^2$. Then, by increasing Δ the area increases in proportion to r^2 and r increases more than proportionately with Δ. So population increases more than proportionately with the square of Δ. Also notice, as discussed in the main text, that as ϵ approaches 1, the change in population is proportional to the square of the change in Δ.

A.5 Proof to Proposition 6:

Using the notation introduced above, we can write equation (12) as $Pop^C = \frac{2\pi}{Z}\Delta g(n, \Delta)^{\frac{1}{1-\epsilon}}$ To find the change in population due to a change in power density Δ, we use simple hat algebra to find: $\frac{dPop^C}{d\Delta}\frac{\Delta}{Pop^C} = 1 + \frac{1}{1-\epsilon}\left[\epsilon_{gn}\epsilon_{n\Delta} + \epsilon_{g\Delta}\right] = 2 + \frac{1}{\epsilon}\frac{\epsilon_{fn}}{\epsilon_{fn}+\epsilon_{gn}}$ To show population changes more than with the cube of power density we need to show that

$$\frac{dPop^C}{d\Delta}\frac{\Delta}{Pop^C} - 3 > 0 \Rightarrow \frac{\varepsilon_{fn}}{\varepsilon_{fn} + \varepsilon gn} > \epsilon$$

Because the equilibrium is to the right of the hump, again we know $\epsilon_{fn} > -\epsilon_{gn}$ so the left hand side of the previous equation is always greater or equal to 1, it is equal to one only when $n = \Delta/c$ or when $\epsilon = 1$. In the case of $\epsilon = 1$ we find $\frac{dPop^C}{d\Delta}\frac{\Delta}{Pop^C} = 3$.

A.6 Proof to Proposition 7:

Define r_A as the distance at which agents are indifferent between bringing their energy to the core, that is $u^A = u^H$, and also indifferent between doing this by land or taking advantage of the road, that is $\theta = \bar{\theta}$. Now, define r_B as the distance of the agent that lives on the low cost alternative, $(\theta = 0)$, and choses to go to the core. Every other agent at distances r between r_A and r_B will go to the core only if they are an angle $\theta < \theta_r$ from the horizontal, where θ_r is yet to be defined. We observe that all agents along the city border have the same utility $u^A = u^H$. Given $U^A = I/P$ and P is the same for all agents, then it must be true that income I is also the same for all agents on the city border. Specifically, all agents have the same income as the agent at distance r_A and angle $\bar{\theta}$. From these observations we can identify θ_r, as the solution to the following implicit function:

$$p(r_A)(\Delta - cr_A) = p(r)[\Delta - c(\theta)r] \tag{A.9}$$

where $c(\theta)$ is given by (4). We next need to find an expression for $\frac{p(r_A)}{p(r)}$. Recall that

$$\frac{m(r_A)}{m(r)} = \frac{M(r_A)}{M(r)} = \frac{X(r_A)}{X(r)} = \frac{\int_0^{\bar{\theta}} r_A(\Delta - cr_A)d\theta}{\int_0^{\theta_r} r(\Delta - c(\theta)r)d\theta} \tag{A.10}$$

From the first order conditions of the utility maximization problem we find $\frac{m(r_A)}{m(r)} = \left(\frac{p(r)}{p(r_A)}\right)^{\frac{1}{1-\epsilon}}$ and we combine these expressions to get

$$\frac{\left(\int_0^{\theta_r} r(\Delta - c(\theta)r)d\theta\right)^{1-\epsilon}}{[\Delta - c(\theta)r]} = \frac{\left(\int_0^{\bar{\theta}} r_A(\Delta - cr_A)d\theta\right)^{1-\epsilon}}{[\Delta - cr_A]} \tag{A.11}$$

where we can find θ_r as a function of r. We next use the implicit function theorem to find whether $d\theta_r/dr < 0$. To begin, we can see the righthand side of the previous equation is independent of θ_r and r. So we can write the previous equation as

$$\left(\int_0^{\theta_r} r(\Delta - c(\theta)r)d\theta\right)^{1-\epsilon} = K(r_A) \times m(\theta, r) \tag{A.12}$$

Total derivation of the expression above yields:

$$\frac{d\theta_r}{dr} = -\frac{(1-\epsilon)\left[\int_0^{\theta_r} \Delta - 2c(\theta)d\theta\right](\Delta - c(\theta_r)r) + \left[\int_0^{\theta_r} r(\Delta - c(\theta)r)d\theta\right]c(\theta_r)}{(1-\epsilon)r(\Delta - c(\theta_r)r)^2 - \left[\int_0^{\theta_r} r(\Delta - c(\theta)r)d\theta\right]c'(\theta_r)} \tag{A.13}$$

From the definition of $c(\theta)$ we have $c'(\theta_r) < 0$, so that $\frac{d\theta_r}{dr} < 0$. Next, we need to characterize θ_r as a function of ρ. When ρ falls, the distance r_A increases. Because r_A increases, and $r_A = \rho r_B$, then the energy margin expands and θ_r increases for each r. Now that we have characterized θ_r, we can next calculate the supply of all goods that are a distance $r < r_B$. For $r < r_A$, the supply of good r brought to the core is given by:

$$X_r(r < r_A) = 4 \times \left[\int_0^{\bar{\theta}} r(\Delta - h(\theta, \rho)cr)d\theta + \int_{\bar{\theta}}^{\pi/2} r(\Delta - cr)d\theta \right] = 2\pi r[\Delta - l(\rho)cr] \quad \text{(A.14)}$$

where $l(\rho) = 1 - \frac{2}{\pi}(\bar{\theta} - (1 - \rho^2)^{1/2})$. With these results in mind, it is straight forward to show that supply increases as ρ decreases. Now, for $r_A < r < r_B$ we find

$$X_r(r_A < r < r_B) = 4 \times \left[\int_0^{\theta_r} r(\Delta - c(\theta)r)d\theta \right] \quad \text{(A.15)}$$

$$= 4r \left(\Delta\theta_r + ((1 - \rho^2)^{1/2}(\cos(\theta_r) - 1) - \rho \sin(\theta_r))cr \right) \quad \text{(A.16)}$$

$$= 4r \left(\Delta\theta_r + v(\rho, \theta_r)cr \right) \quad \text{(A.17)}$$

where $v(\rho, \theta_r) = ((1 - \rho^2)^{1/2}(\cos(\theta_r) - 1) - \rho \sin(\theta_r))$. Next we can calculate the utility from agglomeration.

The price index is given by:

$$P = \left[\int_0^n p(r)^{\frac{\epsilon}{\epsilon - 1}} dr \right]^{\frac{1-\epsilon}{\epsilon}} = \left[\int_0^{r_A(n)} p(r)^{\frac{\epsilon}{\epsilon - 1}} dr + \int_{r_A(n)}^n p(r)^{\frac{\epsilon}{\epsilon - 1}} dr \right]^{\frac{1-\epsilon}{\epsilon}} \quad \text{(A.18)}$$

$$= p(r') \left[\int_0^{r_A(n)} \left[\frac{r(\Delta - l(\rho)cr)}{r'(\Delta - l(\rho)cr')} \right]^{\epsilon} dr + \int_{r_A(n)}^n \left[\frac{\frac{2}{\pi} r \left(\Delta\theta_r + v(\rho, \theta_r)cr \right)}{r'(\Delta - l(\rho)cr')} \right]^{\epsilon} dr \right]^{\frac{1-\epsilon}{\epsilon}} \quad \text{(A.19)}$$

and the utility from agglomeration is

$$U(r', \theta) = \frac{(\Delta - l(\rho)cr')^{\epsilon}}{(r')^{1-\epsilon}} \left[\int_0^{r_A(n)} [r(\Delta - l(\rho)cr)]^{\epsilon} dr + \int_{r_A(n)}^n \left[\frac{2}{\pi} r \left(\Delta\theta_r + v(\rho, \theta_r)cr \right) \right]^{\epsilon} dr \right]^{\frac{\epsilon - 1}{\epsilon}} \quad \text{(A.20)}$$

The expression above has the exact same form as the expression we identified before in the case without reduced costs transportation options. Set r' equal to n and $r_A(n) = n/\rho$, then we recover the same expression as before where $u^A = \Delta\tilde{f} \times \tilde{g}$, where

$$\tilde{f} = \frac{(1 - l(\rho)(c/\Delta)r')^{\epsilon}}{(r')^{1-\epsilon}} \tag{A.21}$$

and

$$\tilde{g} = \left[\int_0^{r_A(n)} [r(1 - l(\rho)(c/\Delta)r)]^{\epsilon}\, dr + \int_{r_A(n)}^{n} \left[\frac{2}{\pi} r\left(\theta_r + v(\rho, \theta_r)(c/\Delta)r\right)\right]^{\epsilon} dr\right]^{\frac{1-\epsilon}{\epsilon}} \tag{A.22}$$

To complete the proof, we need to show that n increases with better transportation options. To accomplish this, we need to find how n changes with ρ. Taking total derivatives of the expression for utility agglomeration and recognizing that utility from staying home is independent of the river option, we find

$$\frac{\rho dn}{n d\rho} = -\frac{\frac{\tilde{f}_\rho \rho}{\tilde{f}} + \frac{\tilde{g}_\rho \rho}{\tilde{g}}}{\frac{\tilde{f}_n n}{\tilde{f}} + \frac{\tilde{g}_n n}{\tilde{g}}} > 0 \tag{A.23}$$

where the inequality follows from observing that

$$\frac{\tilde{f}_\rho \rho}{\tilde{f}} = -\frac{\epsilon(c/\Delta)n\rho l'(\rho)}{1 - (c/\Delta)nl(\rho)} > 0 \tag{A.24}$$

and

$$\frac{\tilde{g}_\rho \rho}{\tilde{g}} = \frac{-(1-\epsilon)(\epsilon+1)\left(\rho l'(\rho) - \frac{(\rho l'(\rho)+l(\rho))\left(1-\frac{cn\rho l(\rho)}{\Delta}\right)^{\epsilon}}{{}_2F_1\left(-\epsilon,\epsilon+1;\epsilon+2;\frac{cn\rho l(\rho)}{\Delta}\right)}\right)}{\epsilon l(\rho)} > 0 \tag{A.25}$$

As we showed before, the inequality follows from recognizing that the equilibrium of interest is to the right of the hump, which implies $\frac{\tilde{f}_n n}{\tilde{f}} < -\frac{\tilde{g}_n n}{\tilde{g}}$ so the denominator is a negative.

B Unit Costs of Electricity Transmission

Electric power measured in $Watts$ is equal to the product of voltage and current: $W = VI$ where V is voltage measured in Volts and I is current measured in Amperes. Volts times Amperes is Watts and since the power flowing from any energy resource will be measured in Watts we need to write line losses in these terms to find net power available for delivery to the core.

When electricity is transmitted over any distance, the transmission line heats up as power is lost to heat because of the line's resistance (*Joule heating*). Resistance is in turn related to the size, material and length of the line. Very simply resistance is given by $R = (\varphi/a)l$ where φ is a measure of the resistivity of the material used in the cable, a is the cross-sectional area of the cable and l is its length. Therefore, losses due to resistance are simply linear in distance much as frictional losses were linear in distance. To find the extent of losses as a function of distance, we need to transform line resistance (which is measured in *Ohms*) into *Watts*. To do so, we use another well known result from physics: *Ohm's law*. This law states that the current transmitted along a linear conductor is proportional to voltage and inversely proportional to the resistance; that is, $I = V/R$. Using this law we can now calculate the power dissipated by resistance. Denote the line losses from transmission by W^L. These losses can then be related to the voltage and amperage relevant to our line as given by $W^L = VI$. Now using Ohm's law we can substitute out voltage ($V = IR$) to find line losses W^L are proportional to both resistance and the square of current since $W^L = I^2R$. Replacing the expression for resistance we have $W^L = I^2(\varphi/a)l = cl$; where $c = I^2(\varphi/a)$.