The Beijer Institute of Ecological Economics

DISCUSSION PAPER

Beijer Discussion Paper Series No. 258

An Energy-centric Theory of Agglomeration

Juan Moreno-Cruz and M. Scott Taylor

The Beijer Institute of Ecological Economics
The Royal Swedish Academy of Sciences
Box 50005 SE-104 05 Stockholm Sweden
Phone: +46 8 673 95 00 Fax: +46 8 15 24 64
E-mail: beijer@beijer.kva.se

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 $\mu g d/\Delta$. Therefore, when energy resources have mass (and they incur land transport costs) we have $c = \mu g d$ 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:

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

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:

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

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

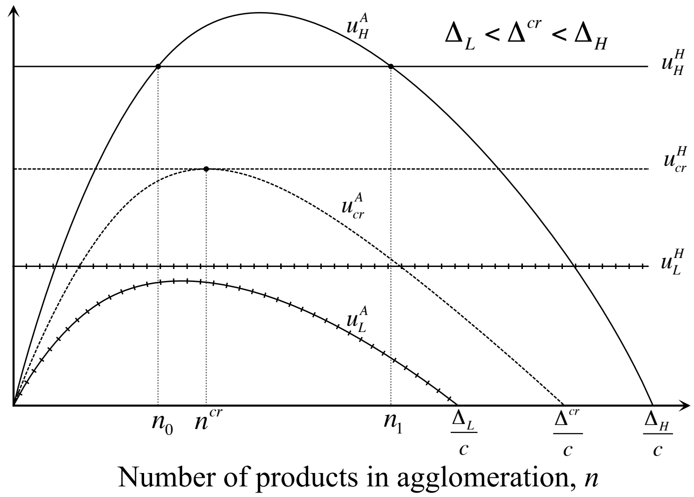


Figure 1: Agglomeration

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

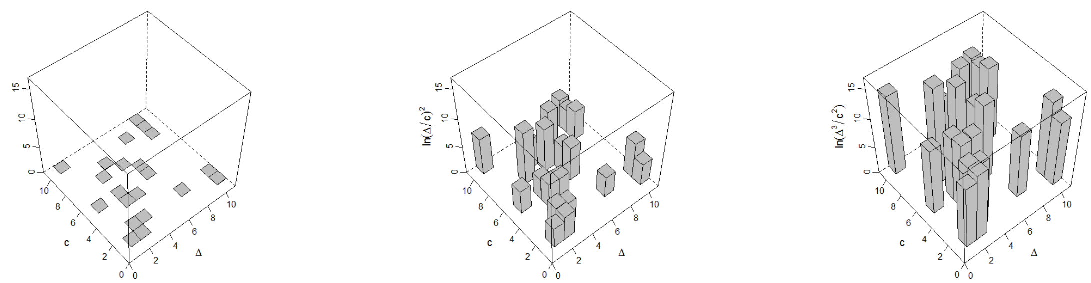
Perhaps the most central fact of economic geography is that human settlements are unevenly distributed across geographic space. For example, the G-ECON dataset built by Nordhaus and coauthors (see Nordhaus (2006)) shows that fully 85% of the world's GDP is produced within 10% of the land area, and this 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.

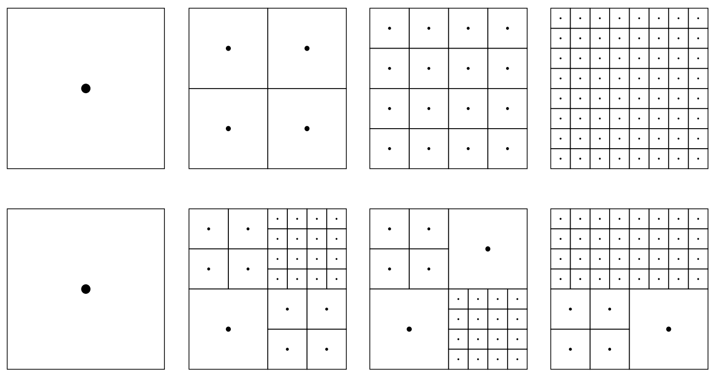


Figure 3: Cities and Regions

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.

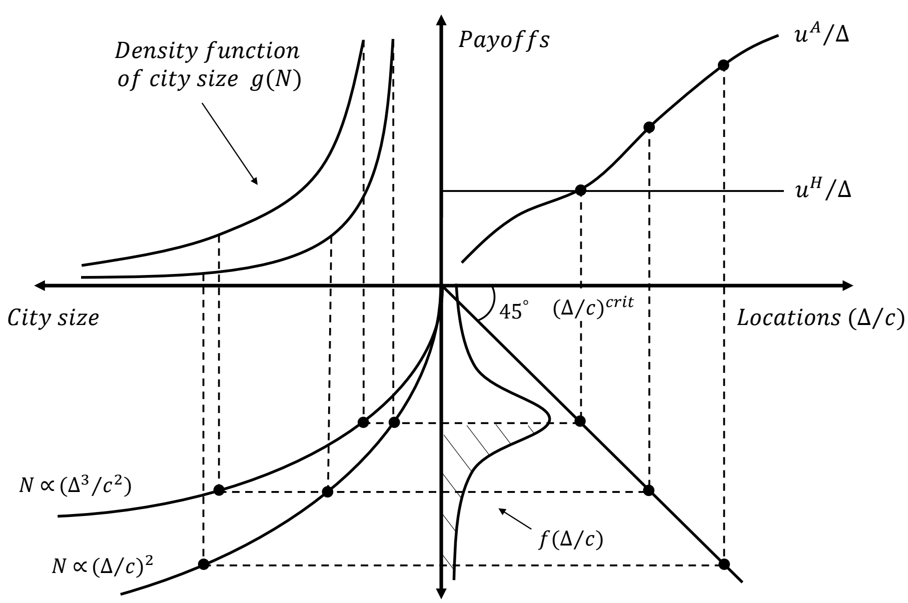


Figure 4: City Size Distribution

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
$\frac{\partial^2 W^S}{\partial n^2} = \frac{-\Delta^3 \pi^2 \left(\sqrt{1-\rho^2} \cos\left(\frac{\pi}{2n}\right) - \rho \sin\left(\frac{\pi}{2n}\right) \right)}{3c^2 n^3 \left(\sqrt{1-\rho^2} \sin\left(\frac{\pi}{2n}\right) + \rho \cos\left(\frac{\pi}{2n}\right) \right)^3}$ 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

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

(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

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

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

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

and

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

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 $c = I^2(\varphi/a)$.

Online appendix to accompany "An Energy-centric Theory of Agglomeration" by Juan Moreno-Cruz and M. Scott Taylor

A The Energy Industry

The aggregate 10 trillion dollar number is an estimate calculated as follows: annual energy consumption was 12,807.1 million tons of oil equivalent in 2013 (British Petroleum, 2015, 40). A ton of oil is equivalent to 7.33 barrels, and the average spot price of a barrel of Brent crude oil, a common benchmark for the world price of oil, was $108.56 (US) in 2013 (Energy Information Administration, 2015). Calculating annual global energy sales in 2013 using this information yields an estimate of $10.191 trillion US.

Global energy consumption in barrels of oil equivalent per second in 2013 is calculated by once again taking annual energy consumption of 12,807.1 million tons of oil equivalent, multiplying by 7.33 barrels of oil per ton and then dividing that figure by $31{,}536{,}000 = 60 \times 60 \times 24 \times 365$ seconds in 2013 to get 2976.79 barrels of oil equivalent consumed per second.

The World Trade Organization valued 2013 global fuels exports (f.o.b.) at $3.258 trillion US (2014, 62). Fuels are classified by the United Nations Statistics Division (2006) in section 3 of the Standard International Trade Classification, Rev. 4 as: coal, coke and briquettes; petroleum, petroleum products and related materials; gas, natural and manufactured; and electric current.

Data on global pipeline infrastructure listed by country is accessible through the Central Intelligence Agency's World Factbook (2013). Simple addition yields global oil and gas pipeline infrastructure totalling 3,573,235 km, with natural gas pipelines accounting for 2,855,017 km. Oil, refined products and liquid petroleum gas pipelines account for the remaining 718,218 km.

The global oil tanker fleet has a total capacity of 472.890 million deadweight tons (DWT), corresponding to a 29.1 percent share of the total capacity of the global merchant fleet. Liquefied natural gas (LNG) carriers have a combined total capacity of 44.346 million DWT, a 2.7 percent share of the deadweight tonnage of the global merchant fleet. Combining these numbers yields a total global oil and gas merchant vessel capacity of 517.236 DWT. All information on capacity and percentage share of global fleet capacity is taken from the United Nations Conference on Trade and Development (2014, 29).

Fortune Magazine's Global 500 list (2013) ranks the largest companies in the world, by revenue. Included in the list's top 10 are energy companies Royal Dutch Shell, Exxon Mobil, Sinopec Group, China National Petroleum, British Petroleum, State Grid and Total.

B von Thunen and Iceberg Costs of Transport

The transport cost assumptions adopted in von Thunen are subtly, but importantly, different from what we have assumed here. In short, iceberg costs require the energy costs of transportation to fall immediately and completely as energy is expended. At the practical level this rules out containers for fuel storage or combustion, residues left from incomplete combustion, and no mass of the vehicle carrying the load. In terms of the oat eating horse example both Von Thunen and Samuelson used, the horse cannot have any mass, the oats cannot remain resident in the horse, and there is of course no wagon to pull. It is fair to say that while the iceberg assumption is tractable it is also a knife edge assumption as we will show below. If even an epsilon of the mass of energy is wasted in moving containers, in moving engines or left in incompletely combusted particles then the transport process produces a result significantly different than von Thunen's but qualitatively the same as in our specification. Specifically it will lead to a formulation where there is a maximum zone of exploitation tied to the power density of energy.

To be precise, consider the case of renewable with mass discussed in section 2.1.1 on the main text. Assume there is some fixed cost associated with moving energy a distance dx. Then transportation costs over this increment are given by:

$$W^T(x) = \left(C(W_0) + \frac{\mu g d}{\Delta} W(x)\right) dx$$

Total energy remaining at distance $x+dx$ is given by $W(x+dx) = W(x) - \big(C(W_0) + \frac{\mu g d}{\Delta} W(x)\big)\, dx$. Rearranging terms we can rewrite this expression as

$$\frac{W(x+dx) - W(x)}{dx} = \frac{dW(x)}{dx} = -\left(C(W_0) + \frac{\mu g d}{\Delta} W(x)\right)$$

The solution to this differential equation is:

$$W(x) = \left(W_0 + \frac{\Delta}{\mu g d} C(W_0)\right) e^{-\frac{\mu g d}{\Delta} x} - \frac{\Delta}{\mu g d} C(W_0)$$

Define R as the radius for which $W(R) = 0$; that is, the energy supplied to the core by any energy source further away than R is zero. The solution for R is:

$$R = \frac{\Delta}{\mu g d} \ln\left(1 + \frac{\mu g d}{\Delta} \frac{W_0}{C(W_0)}\right)$$

The iceberg assumption occurs when $C(W_0) = 0$ since in this case R goes to infinity. The case we consider in the text arises when $C(W_0)$ is proportional to the mass of energy transported; specifically that $C(W_0) = \frac{\mu g d}{\Delta} \frac{W_0}{e-1}$ since then we obtain $R = \frac{\Delta}{\mu g d}$.

Two observations are in order. First, there exists a finite margin of exploitation for any $C(W_0) > 0$. Thus, iceberg costs represent a knife edge assumption; any value other than $C(W_0) = 0$ generates a qualitatively different result. Any and all energy sources sharing the same - infinite - margin of exploitation when $C(W_0) = 0$; they have finite and different margins of exploitation for any $C(W_0) > 0$. Second, iceberg costs have proven tractable in general equilibrium models because they allow us to model the transportation system without introducing another economic activity complicating predictions in small dimensional models. The formulation in the body of the paper does however respect this constraint. Note the only costs of transport come from moving energy (and not containers, equipment or engines even though the result is consistent with formulations with these fixed costs), the key to our result is our assumption that the mass of energy is transported even as it is used (converted) in transport.

C Extensions

The model presented in the paper is decidedly stark and abstract. In this section of the Online Appendix we present several extensions to showcase the versatility and wider applicability of the Only Energy model. In the paper we treat power density as a primitive. While this is true in principle, investments can be made to improve the quality of the resources or to reduce the costs of transporting them to the core. In the first extension we show the incentives to expand the exploitation zone and to upgrade the quality of the resources; more importantly we show that these incentives are stronger when the intrinsic quality of the resources is higher. Another important assumption in the paper is that power density is distributed uniformly across space. In our second extension we consider the case where the distribution of resources is patchy or punctiform and we also introduce the possibility of uncertain location of resources across space. We show that all the results we find using the uniform distribution hold under these alternative resource distributions. We then introduce the case of non-renewables. Our purpose here is to demonstrate how spatial productivity is germane to both renewables and non-renewables although the details are in some cases importantly different. Perhaps more importantly we demonstrate how spatially productive environments lead to a bunching of resource extractions in calendar time.

C.1 Resource Upgrading and Endogenous Exploitation Zone

C.1.1 Energy Investments

The magnification effect (Proposition 2 in the main text) tells us two quite useful things. The first is that access to low cost transport options is equivalent to raising the power density of surrounding resources. Given this equivalence, for any investment lowering the physical cost of transport there is an equivalent one raising the power density of energy resources by upgrading. Upgrading resources and investing to lower transport costs are opposite sides of the same coin.

The second is that the marginal impact of low cost transport options is greater for more power dense resources. Simply put, the energy supply consequences of say river access is far higher in a world run by, for example, coal than it would be in one run on biomass. This suggests that investments to lower transport costs will be greatest in situations where energy resources are already quite power dense.

To examine the motivation for energy investments assume the cost of building and maintaining transport that delivers an efficiency of ρ is given by $h(\rho)$. Assume $\rho > \bar{\rho}$ where $\bar{\rho}$ is the minimum physically possible value of the coefficient of friction. Since a lower value of ρ implies lower cost energy transport we assume $h(\rho)$ is a decreasing and convex function: $h'(\rho) < 0$, $h''(\rho) > 0$ with $h(1) = 0$ and $h'(1) < 0$. Then energy supplied net of infrastructure costs are $W^N = W^S(\rho) - h(\rho)$. For concreteness consider

$$W^S(\rho) = \frac{1}{3}\frac{\Delta^3}{c^2}g(\rho) \tag{C.1}$$

$$g(\rho) = \pi - 2\bar{\theta} + 2\int_0^{\bar{\theta}} \left((1-\rho^2)^{1/2}\sin\theta + \rho\cos\theta\right)^{-2} d\theta \geqslant 0 \tag{C.2}$$

as in the paper and we are improving river transportation by dredging, locks, canals, maintaining ports etc. The optimal investment problem is simply given by

$$\max_{\rho} W^N = W^S(\rho) - h(\rho) = \frac{1}{3}\frac{\Delta^3}{c^2}g(\rho) - h(\rho) \tag{C.3}$$

When the solution is interior, the first order condition that maximizes energy requires

$$\frac{1}{3}\frac{\Delta^3}{c^2}g'(\rho) = h'(\rho) \tag{C.4}$$

The left hand side of this equation is the marginal benefit from improved transport and it is again a cubic in power density showing a strong relationship between power density and the marginal benefit of further investments. The right hand side is simply marginal costs of improved transportation. Since marginal benefits are bounded and marginal costs of the first unit of investment are positive, with sufficiently low power density no investments in transport improvements will occur. In situations with higher power density an interior solution will obtain and we can write the implicit solution to (C.4) as $\tilde{\rho}(\Delta)$. Straightforward differentiation now shows[1]

$$\frac{d\rho}{d\Delta} = -\frac{\rho}{\Delta}\frac{3}{\epsilon'_{h\rho} - \epsilon'_{g\rho}} < 0$$

where $\epsilon'_{h\rho} = -\rho h''(\rho)/h'(\rho)$ and $\epsilon'_{g\rho} = -\rho g''(\rho)/g'(\rho)$ and the second order conditions sign the expression. Therefore, we have proven:

Proposition C.1 *Complementarity. There exists a critical level of power density $\Delta^c >$ 0 such that for energy resources with $\Delta < \Delta^c$ no investments in cost reducing transport occur; but for environments with $\Delta > \Delta^c$, investment is positive and more power dense energy resources call forth greater investments in cost reducing transportation investments.*

Proposition C.1 links the incentive for transport and upgrading of resources to their power density. Resources with low power density will neither be upgraded nor call forth

[1]We have already assumed $h''(\rho) > 0$; it can be shown that $g(\rho)$ is convex for all values of ρ. The second order conditions required for a maximum imply $\frac{1}{3}\frac{\pi\Delta^3}{c^2}g''(\rho) < h''(\rho)$ which simply says the costs are more convex than the benefits from investing in infrastructure.

large investments to improve transportation. Biomass for example is not a very power dense energy resource, and in a world run on biomass we should expect very local energy markets that are severely constrained by transport costs. These local markets could be expanded somewhat by simple upgrading (charcoal), but the incentives for dedicated investments should be weak. With the advent of fossil fuels new incentives arose. More power dense fuels like coal brought forth large investments in lowering transport costs and upgrading. Canals, railroads, and upgrading coal to coke could be seen as endogenous responses to its higher power density. In the petroleum era we have witnessed and continue to witness today, massive investments in pipelines, tankers and terminals, so that the set of exploitable petroleum resources now includes oil drilled in the world's most inhospitable Arctic climates and is collected from underwater wells literally miles deep. Naturally given the distances involved and the distribution of resources around the world, energy shipments now routinely cross political boundaries which has created a huge international trade in energy products where two centuries previous there was virtually none at all. Proposition 6 links this long chain of events to the changing power density of available energy sources.[2]

C.1.2 The Incentive to Upgrade

We start by providing a decomposition of power density into its component parts to understand how it may be subject to control. The decomposition links power density to directly observable and familiar characteristics such as crop yields, energy contents, and growth rates. We provide a similar decomposition for non-renewables in a subsequent section.

To start consider renewables that produce a physical harvest (timber, staple crops, fisheries, etc.) In these cases, we can write the steady state flow of energy harvested $\mathcal{F}$ [Watts] from the renewable energy resource as the product of three things: the size of the resource stock S [kg]; its current growth rate r [1/time]; and the energy content of the yield, e

[2] Notice that this analysis can be expanded to the case of electricity generation and transmission. Increasing voltages is costly but reduces line losses, therefore there is an incentive to invest for as long as the power delivered is enough to compensate for the investment in higher voltage lines.

[Joules/kg]. $\mathcal{F} = reS$, where rS represents the physical harvest per unit time, and e translates this physical flow into an energy flow in units of Watts. The physical size of the stock S can in turn be written as the product of the physical density of the resource in the environment, δ [kg/m^2] times the area containing the resource a[m^2]. Making this substitution we obtain the flow of energy as $\mathcal{F} = (re\delta)a$. Power density is just the flow of energy per unit area or $\Delta = \mathcal{F}/a = re\delta$ [W/m^2].[3] Therefore, for a renewable resource that generates a continuous physical harvest flow - like a coppiced forest, biomass, etc. - its power density is proportional to its rate of growth, or recharge rate, r, its energy content e [Joules/kg], and its physical density in the environment, δ [kg/m^2]. Since the harvest from the resource is rS we have $d = r\delta$ as the density of the harvest. Using this result we can now rewrite to find $\Delta = ed$, and $R^* = e/\mu g$. This leads to our next proposition.

Proposition C.2 *Quality versus Quantity. The extensive margin of energy collection is independent of the quantity of the resource available (measured by its physical density in the environment), but proportional to its quality (measured by its energy content).*

Proof. *In text.* ■

Proposition C.2 is quite intuitive. Recall the extensive margin is defined by zero net energy rent resources; that is, those for which transport costs completely dissipate the benefits of collection. Since transport costs are proportional to the amount collected, as is the energy contained within, it matters little whether we have an ounce of these marginal resources or a ton - marginal resources are marginal in whatever quantities we find them.

Once said, this result seems obvious, but it may explain much of the energy resource upgrading we see in the world today or in the past. Consider for example the age old collection of firewood and conversion to charcoal before transport. Charcoal has a higher energy content than wood, and therefore - via Proposition C.2 - will be collected at greater

[3] A 100kg forest growing at 10% per year generates 10 kg of firewood per year. Firewood contains 15 MJ per kg; and there are 31,536 x 10^3 seconds in a year. This piece of the forest provides 4.75 W on average for the year. If the physical density of the forest is such that it contains 50 kg of trees over each meter squared, then the power density of this forest resource is (4.75/2).

distances. The fact that the conversion of fuel wood into charcoal is incredibly inefficient in a physical sense is irrelevant. Since stranded firewood resources have a zero opportunity cost, any degree of inefficiency is acceptable if a conversion to a higher e resource is possible.

Similarly today it is common to see energy resources upgraded to make transportation more efficient (lower μ), raise energy content (e), or both. For example, the upgrading of heavy oil not only raises its energy content and lowers its transport cost by lowering viscosity, it is also very energy intensive. While we may bemoan the energy and pollution consequences of this upgrading, the opportunity cost of using stranded resources is very low and hence the logic of doing so is impeccable. Compressing or liquifying natural gas is another example where the energy content (per unit volume) is raised, transportation made easier, and yet the process is quite costly.

To illustrate more concretely the incentives to upgrade resources, consider the following simple problem (along the lines of energy investments presented above). The examples discussed above in the text suggest that the role of upgrading is increasing in energy content e, while taking concentration d as given. With this in mind, power supplied to the core is given by $W^S(e,d) = \frac{1}{3}\frac{\pi e^3 d}{(\mu g)^2}$. Assume the costs of investing in resource upgrading to deliver energy content, e, is $k(e)$ such that $k'(e) > 0$ and $k''(e) > 0$.

Resource upgrading of this form is also common among energy resources without mass that generate electricity (solar, wind, hydro). To see why recall that in this case energy rents are driven to zero at the margin of exploitation, and this occurs at a distance given by

$$l^* = \frac{\Delta}{c} = \frac{\Delta}{I^2(\rho/a)}. \tag{C.5}$$

the extensive margin now varies with power density, amperage and the details of the transmission network. Low amperage raises delivered power as do large lines and better materials. To understand how amperage matters, note that we have assumed any given location throws off power at rate Δ in Watts but said little about how this power is transmitted. In fact, this power can be transmitted in a variety of ways because current times voltage equals Watts, or $W = VI$. Therefore, for a renewable energy source x with power density Δ we have a similar decomposition with $\Delta(x) = V(x)I(x)$. Substituting this into (C.5) we see that the extensive margin for source x is given by:

$$l^*(x) = \frac{V(x)}{I(x)(\rho/a)} \tag{C.6}$$

Although voltage and amperage are not innate characteristics of energy sources, unlike energy contents and yields of renewables with mass, it is true that producing power from wind, solar, thermal generation or nuclear sources requires a set of generation technologies that have been optimized to meet their peculiar requirements. These generation technologies differ in the voltage produced. We can therefore think of otherwise homogenous providers of electricity as heterogenous in this regard, and from (C.6) we know the extensive margin relevant to any particular source x is rising in its voltage of transmission and falling in its amperage.

Putting these results together, we have a result much like we had before. Raising the power density of an energy source raises the flow of power produced but has no impact on the extensive margin if the voltage (quality) to amperage (quantity) ratio remain constant. And resources that would otherwise be stranded by their distance or low quality will either remain stranded or call forth endogenous investments to lower costs. While we prefer high voltage transmission (just as we prefer high energy content fuels) because it lowers line losses (or physical transport costs), altering voltage is of course costly just as improving roads, building canals, etc. to improve land transportation is costly. Indeed, our theory suggests that what appears to be excessively wasteful expenditures in energy used to ramp up voltages for the long distance transmission of renewables are but the mirror image of energy intensive methods to upgrade heavy oil or compress natural gas for non-renewables.

The optimization problem is given by $\max_e W^N = W^S(e,d) - k(e) = \frac{1}{3}\frac{\pi e^3 d}{(\mu g)^2} - k(e)$. The first order condition is given by $\frac{\pi e^2 d}{(\mu g)^2} = k'(e)$. Finally we can totally differentiate the first order condition to obtain: $\frac{de}{dd} = \frac{e/d}{\epsilon'_{ke}-2} > 0$, where $\epsilon'_{ke} = ek''(e)/k'(e)$ and the sign follows from the assumption that costs are more convex than the benefits from upgrading resources.[4]

C.2 Patchy, Punctiform and Probabilistic Resource Distributions

In the main text we assume resources are uniformly distributed, the space containing resource plays is connected, and there is no uncertainty regarding whether the energy resources in question are present. These are strong assumptions, but for some resources they seem innocuous. For example, crops and woodlands typically satisfy these constraints at least over fairly large areas. But for other resources they fit less well and it is unclear how our analysis would change under these assumptions. For example, the available locations for resource exploitation may be patchy (containing holes) because of land use restrictions, habitat conservation, or noise considerations. The siting decisions for wind and solar farms certainly fit this description. In other cases, most notably fossil fuels, there are often a few very significant deposits surrounded by areas with little if any resource potential (the space contains resource "plays" with widely different power densities). We will refer to this case as one where the resource distribution is *punctiform.* In some other cases it is not clear ex ante whether resources are present in any specific location although there maybe a well defined probability distribution over them (oil and gas deposits come to mind). We refer to this case as one where the distribution of resources is *probabilistic.* We will show that often very little of substance changes with alternative resource distributions although the calculations become more lengthy and the expressions less transparent.

To understand why these complications rarely matter, recall our discussion of energy rents which allowed us to define the extensive margin R^*, for a resource of given power density Δ. Let this reliance of the extensive margin on the power density of resources be written as $R^*(\Delta)$. Then since all resources within this margin provide positive energy rents it

[4]Improving transportation or upgrading resources (by increasing energy content) are equivalent problems from the perspective of the energy producer. It also follows from the equivalence between electricity transmission lines and roads that increasing transmission voltage, therefore reducing Joule losses, is equivalent to reducing ρ which reduces land friction and transportation losses.

should be apparent that they will be exploited even if the resource base is not connected nor homogenous. And if we locate all such potential resources, identify their extensive margins, and then integrate over their relevant regions this (more complicated) sum of energy rents will equal the energy supply just as before. Apart from mathematical complications, patchy and punctiform resource distributions pose no special problem. Alternatively if we assume resources are present in specific locations with given probabilities, we can again identify $R^*(\Delta)$ and integrate over this space to find what would now be expected energy supply. And if the space defined by $R^*(\Delta)$ can be divided into many resource plays with identical and independent success distributions, then a law of large numbers result could be invoked to render expected energy supply equal to ex post energy supply. At bottom the reason why these complications do not matter much is the constant returns built into transport costs by the physics of the underlying problem. Moving an object twice as far is twice the work; moving an object with twice the mass is twice the work; and if movement is output and work (energy) is the input, this production function is CRS. The CRS feature of the problem allows us to aggregate easily, define boundaries simply, and replace patchy, punctiform and probabilistic resource distributions with much simpler connected and homogenous ones in many cases.

To see exactly how to incorporate complicated resource distributions, we construct two examples.

C.2.1 Patchy and Punctiform

It may be clear from the description above that the key complication is locating the various resources in space. To make the analysis tractable and transparent, we construct discrete resource distributions. Consider a division of the space surrounding the core into concentric circles that are then divided further into wedges created by extending rays from the core. The result, shown in Figure 1, is a sequence of land parcels we will refer to as resource plays.

Let there be $n = 1, ..., N$ rays and $m = 1, ..., M$ circles, then there are $N \times M$ resource

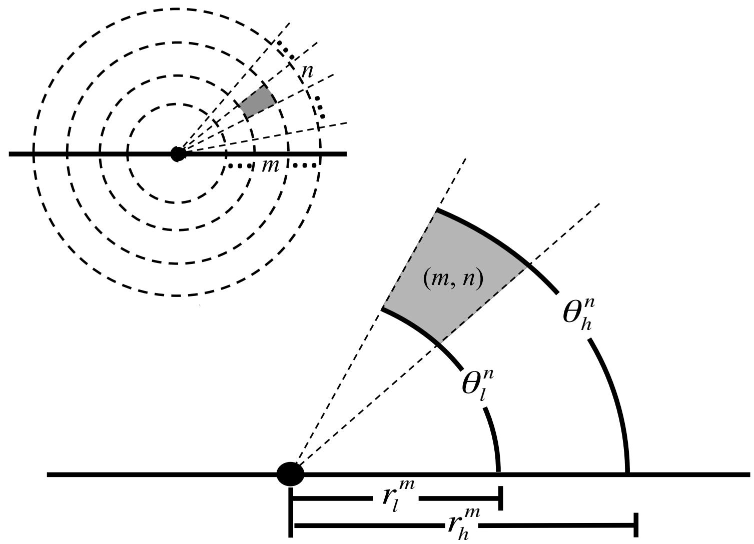


Figure 1: Uneven distribution of power density into parcels

plays each uniquely identified by the duple (m, n). Suppose each play has an associated power density Δ_{mn} with geometrical shape characterized by its width $r^m = r_h^m - r_l^m$ and the angle of the wedge $\theta^n = \theta_h^n - \theta_l^n$. Where h and l refer to both the higher and lower radius bounds defining the play; and the higher and lower angles (measured in radians) that define its location in the plane. We can write the (maximum) energy supplied by any given resource play:

$$W_{mn} = \int_{\theta_l^n}^{\theta_h^n} \int_{r_l^m}^{r_h^m} v\left(\Delta_{mn} - cv\right) dv d\varphi$$

$$W_{mn} = \frac{1}{2}\left(\theta_h^n - \theta_l^n\right)\left((r_h^m)^2 - (r_l^m)^2\right)\left(\Delta_{mn} - \frac{2}{3}c\frac{(r_h^m)^3 - (r_l^m)^3}{(r_h^m)^2 - (r_l^m)^2}\right) \tag{C.7}$$

Since an energy supplier with play (m, n) supplies energy if the play provides positive energy rents, we need to account for this complication by noting that each density Δ_{mn} has an associated energy margin $\bar{R}_{mn} = \frac{\Delta_{mn}}{c}$. This implies the actual energy supplied to the core by any resource play must be such that:

$$W^S_{mn} = \begin{cases} W_{mn} & \text{if } r^m_h \leq \bar{R}_{mn} \\ \bar{W}_{mn} & \text{if } r^m_l \leq \bar{R}_{mn} < r^m_h \\ 0 & \text{if } \bar{R}_{mn} < r^m_l < r^m_h \end{cases} \tag{C.8}$$

where $\bar{W}_{mn}$ has the same form as equation (C.7) but where r^m_h is replaced by $\bar{R}_{mn}$.

To find the aggregate energy supplied we add the n plays of each annuli m and then add all the annuli. Without further restrictions, the possibilities are very numerous. Therefore consider the case where each play within an annuli m has the same power density Δ_m. As well, order the power densities from lowest to highest such that $\Delta_0 = 0 < \Delta_1 < ... < \Delta_m < ... < \Delta_M$ so distant resources are the most power dense, and in order to eliminate potential gaps in our distribution we assume the width of each annuli is determined by the energy margins of the neighboring annuli. That is assume $r^m_l = \bar{R}_{m-1}$ and $r^m_h = \bar{R}_m$. Alternate assumptions are readily investigated. Using these assumptions we can now replace the definition of $\bar{R}_m$ back in equation (C.7) to find the energy supplied to the core:

$$W_{mn} = \frac{1}{2}\frac{(\theta^n_h - \theta^n_l)}{c^2}\left(\frac{\Delta^3_m}{3} - \Delta_m\Delta^2_{m-1} + \frac{2}{3}\Delta^3_{m-1}\right)$$

Add over all wedges in the annuli m and over all the annuli M to find

$$W^S = \frac{\pi}{3c^2}\sum_{m=1}^{M}\Delta^3_m\left(1 - 3\frac{\Delta^2_{m-1}}{\Delta^2_m} + 2\frac{\Delta^3_{m-1}}{\Delta^3_m}\right)$$

Two observations are in order. First, since the summation is over primitive determinants of the model, we could just as well replace this complicated sum with $\tilde{\Delta}$, where $\tilde{\Delta}$ is the power density of a hypothetical connected and uniformly distributed resource base yielding the same energy supply. $\tilde{\Delta} > 0$ by virtue of our ordering of power densities, and we can write it simply as:

$$\tilde{\Delta} = \left[\sum_{m=1}^{M} \Delta_m^3 \left(1 - 3\frac{\Delta_{m-1}^2}{\Delta_m^2} + 2\frac{\Delta_{m-1}^3}{\Delta_m^3}\right)\right]^{1/3}$$

Second, if we alter the power density of our hypothetical resource base, $\tilde{\Delta}$, by $\lambda > 0$ this is equivalent to uniform scaling by λ of all power densities in the heterogenous resource zone. A moment's reflection will show that energy supply is homogenous of degree three in all power densities taken together. Therefore, for many purposes we can simply write

$$W^S = \frac{\pi \tilde{\Delta}^3}{3c^2}$$

and ignore the fact that the exploitation zone in question is both patchy and punctiform.

C.2.2 Probabilistic

Here we assume the power density of the resource is uniform across space and it is given by Δ_o. This implies all the resources found inside the margin of extraction given by $R_o = \Delta_o/c$ are going to be exploited. Divide this space as we did before using N rays and M circles to identify $N \times M$ resource plays but now assume each play has a probability q of having a resource with power density Δ_o in place and a probability $1-q$ of being empty. Given Δ_o is uniform and constant, our previous assumptions imply the width of each annuli is equal to R_o/M. Thus, the geometrical shape of the parcel (m, n) is characterized by its boundaries set by $r_m^h = (m+1)\frac{R_o}{M}$ and $r_m^l = m\frac{R_o}{M}$ and the angle of the wedge $\theta^n = \theta_h^n - \theta_l^n$. In the case where parcel (m, n) is not empty, we can calculate the same double integral we calculated for the case of patchy distributions and replace the values for r_m^l and r_m^h to find

$$W_{mn} = \frac{(\theta_h^n - \theta_l^n)}{2}\left(\frac{R_o}{M}\right)^2 \Delta_o \left((m+1)^2 - m^2\right)\left(1 - \frac{2}{3}\frac{1}{M}\frac{((m+1)^3 - m^3)}{((m+1)^2 - m^2)}\right) \tag{C.9}$$

Replacing the definition for R_o and noting the wedges are of equal size given by $\theta^n = 2\pi/N$ we find:

$$W_{mn} = \frac{\pi}{NM^2}\frac{\Delta_o^3}{c^2}\left((m+1)^2 - m^2\right)\left(1 - \frac{2}{3}\frac{1}{M}\frac{((m+1)^3 - m^3)}{((m+1)^2 - m^2)}\right) \tag{C.10}$$

As we mentioned above, the power collected from parcel (m, n) is W_{mn} with probability q and it is zero with probability $1 - q$. Therefore, the expected value of energy provided by parcel (m, n) is:

$$E[W_{mn}] = q \cdot W_{mn} + (1 - q) \cdot 0 \tag{C.11}$$

We can now aggregate across parcels and use the linearity of the expected value operator to find:

$$E[W^S] = q \cdot \bar{M}\frac{\pi\Delta_o^3}{3c^2} \tag{C.12}$$

where $\bar{M}$ is a constant given by:

$$\bar{M} = \sum_{m=0}^{M}\frac{((m+1)^2 - m^2)}{M^2}\left(3 - 2\frac{1}{M}\frac{((m+1)^3 - m^3)}{((m+1)^2 - m^2)}\right) \tag{C.13}$$

In this more complicated case very little seems to change. Power density again enters as a cubic as before since now the area of *exploration* rises with the square of the extensive margin and success is proportional to this exploration zone. As well, as mentioned earlier if the number of plays were large a variety of assumptions are available on the joint distribution across the plays that would render a law of large numbers result. The simplest case being the one employed above where each play is treated as an independent and identically distributed Bernoulli random variable.

C.3 Non-Renewables: Oil, Gas and Coal

Extending our framework to non-renewables presents several challenges. First, since using non-renewable energy today precludes you from using it tomorrow the exploitation zone must change over time as the resource stock is depleted. This is true because with non-renewables, energy flows come from depleting the resource stock and not from harvesting the perpetual yield from a renewing resource. One simple and natural way to address depletion is to assume ongoing extractions hollow out the exploitation zone as the resource is extracted.[5]

Second, while we can for the most part ignore the potential impact current energy collection has on the future productivity of renewables (harvesting solar power today does not affect the likelihood of sunshine tomorrow), this is not possible with non-renewables. To see why, use the approach discussed above and assume all energy resources up to r have already been extracted. Then, the remaining non-renewable energy that could be supplied to the core is given by:

$$W^S = 2\pi \int_r^{R^*} (\Delta - cv)\, v dv = \frac{\pi\Delta^3}{3c^2} - \pi\left(\Delta r^2 - \frac{2}{3}cr^3\right) \tag{C.14}$$

where $R^* = \Delta/c$ as before. The intuition is clear. The economy loses the energy it would have been able to collect over the area already mined — this is, $\pi\Delta r^2$—net of the energy it would have expended to bring this energy to the core, $(2/3)\pi cr^3$. Previous extractions raise the cost of current extractions, and the key economic problem is to determine the rate we wish to use these resources over time. To address this problem we will assume a time separable CRRA utility function that maps delivered energy into instantaneous utility flows, and maximize the discounted sum of these flows subject to resource availability and costs.

C.3.1 A Solow-Wan Reformulation

In order to solve our intertemporal energy supply problem we start by recognizing that our spatial model with reserves differentiated by location, can be rewritten as a standard problem where there is a fixed and given resource stock exploited subject to rising marginal

[5]There is a small literature examining least cost paths for depletion in situations with multiple deposits or resources. This literature, started by Herfindahl (1967), examines when, and under what conditions, a least cost order of extraction path will be optimal. Chakravorty and Kruice (1994) contains relevant references, some discussion, and a neat result showing the typical least cost path prediction does not hold up when the resources in question are not perfect substitutes in use. This possibility is ruled out in our one energy source set up, but would be relevant in any extension with two, less than perfectly substitutable, resource types.

extraction costs. A similar reformulation was first suggested by Solow and Wan (1976) in an environment where resources were differentiated by their grade, and it proves useful to do so here.[6]

To reformulate the problem along Solow-Wan lines, we first recognize that the exploitation zone has radius $R^* = \Delta/c$, and this exploitation zone implies a corresponding limit on recoverable reserves which we denote $\bar{X}$. These recoverable reserves are simply equal to $\bar{X} = \pi\Delta^3/c^2$ which again reflects our scaling law. But if the current resource frontier is $r(t) < R^*$, then the remaining recoverable reserves at t, which we denote $X(t)$ must be equal to

$$X(t) = 2\pi \int_{r(t)}^{R^*} \Delta \iota d\iota = \bar{X} \; - \Delta\pi r(t)^2 \tag{C.15}$$

where $r(0) = 0$ since no resources have been extracted at the start of time. Cumulative extractions at t, are simply $\Delta\pi r(t)^2$.

Differentiating with respect to time, we find the needed link between remaining recoverable reserves and today's rate of extractions:

$$\dot{X} = -2\pi r(t)\Delta\dot{r}(t) = -W(t) \tag{C.16}$$

The intuition is simple. As extraction proceeds, reserves are drawn down and the resource frontier expands. The frontier expands at rate $\dot{r}(t)$ as resources with power density Δ are reaped from a ring with density $2\pi r(t)$ per unit time. The last equality in (C.16) follows because the instantaneous change in the stock must equal $W(t)$ – the flow of energy extracted at t measured in Watts. This completes the first step of the reformulation.

The second step in the reformulation is to find the associated cost function for extractions. When $W(t)$ is extracted, it is divided between deliveries to the core $W^S(t)$ and energy used in transport which we denote $W^T(t)$. We refer to these costs as extraction costs. Because at any t there is a unique $r(t)$, $W^T(t)$ must equal $r(t)[c/\Delta]W(t)$ and hence $r(t)[c/\Delta]$ represents unit

[6] Solow and Wan (1976) suggested this reformulation in a short footnote; for a more illuminating treatment see section 2 of Swierzbinski and Mendelsohn (1989).

extraction costs. While this is useful, we need to eliminate $r(t)$ and purge the problem of all spatial elements. To do so use (C.15) to substitute for $r(t)$ as a function of remaining reserves $X(t)$ and total reserves $\bar{X}$. With some simplification, we can now write the relationship between energy supplied to the core, $W^S(t)$, current extractions, $W(t)$, remaining reserves $X(t)$, and recoverable reserves $\bar{X}$, as:

$$W^S(t) = [1 - C(X(t))]W(t) \tag{C.17}$$

$$C(X) = \left(1 - \frac{X}{\bar{X}}\right)^{1/2} \tag{C.18}$$

where we now interpret $C(X(t))W(t)$ as the cost of extracting $W(t)$ units of energy from a homogenous pool of recoverable reserves $\bar{X}$, when remaining reserves equal X. $C(X(t))$ is therefore the unit extraction cost function (where we have suppressed its reliance on recoverable reserves, $\bar{X}$). With this machinery in place, we solve for the optimal extraction path.

A social planner maximizes the welfare of a representative consumer who values the energy services available for consumption in the core. By choosing service units appropriately, utility is defined over net energy supplied. The planner has a CRRA instantaneous utility function with coefficient of relative risk aversion equal to $\sigma > 0$. Social welfare is:

$$\max_{W(t)} \int_0^\infty e^{-\rho t} U(W^S(t))dt \quad where \quad U(W^S) = \frac{(W^S)^{1-\sigma} - 1}{1 - \sigma} \tag{C.19}$$

The planner maximizes (C.19) subject to the constraints (C.16) and (C.17). We write the current value Hamiltonian as

$$\mathscr{H} = U\left[(1 - C(X(t)))\, W(t)\right] - \lambda(t)W(t) \tag{C.20}$$

where $\lambda(t)$ is the co-state variable associated with the stock of resources. The optimality conditions are given by:

$$\frac{\partial \mathscr{H}}{\partial W(t)} = U'(W^S(t))(1 - C(X(t))) - \lambda(t) = 0 \tag{C.21}$$

$$\frac{\partial \mathscr{H}}{\partial X(t)} = -U'(W^S(t))C'(X)W = \rho\lambda(t) - \dot{\lambda}(t) \tag{C.22}$$

with transversality condition $\lim_{t\to\infty} e^{\rho t}\lambda(t)X(t) = 0$.

Using the definition of the utility function in (C.19) and taking the time derivative of (C.21), substituting in (C.22), and rearranging we find one differential equation linking the current rate of extractions to cumulative extractions:

$$\frac{\dot{W}(t)}{W(t)} = -\frac{\rho}{\sigma} - \frac{C'(X)}{1 - C(X)}W(t) \tag{C.23}$$

A second differential equation is provided by (C.16) while one initial condition and the transversality condition close the system.

The behavior of the dynamic system is presented in Figure 2. The $\dot{W}(t) = 0$ isocline is depicted by the solid curve in Figure 2. This curve is positive for values of $X(t) \in [0, \bar{X}]$ and has a maximum value when cumulative extraction is one quarter of total reserves; i.e. with remaining reserves $X(t) = 3\bar{X}/4$ and cumulative extraction $\bar{X} - X(t) = \bar{X}/4$. The $\dot{X}(t) = 0$ isocline is coincident with the horizontal axis. At all points above the $\dot{X}(t) = 0$ isocline, movement must be rightwards to extract all reserves, giving arrows of motion in the positive direction parallel to the horizontal axis. At points above the $\dot{W}(t) = 0$ isocline, extractions must be increasing since costs are currently too low; below the isocline just the opposite is true. This information is captured by the arrows of motion shown.

Assume we start with a new resource and hence cumulative extractions are zero. Since the arrows of motion near the origin imply all movement must be upwards and to the right, we know the system must move immediately to an initial extraction point like that shown by $W(0) = W_0$. From this initial point, the arrows of motion indicate we move upwards

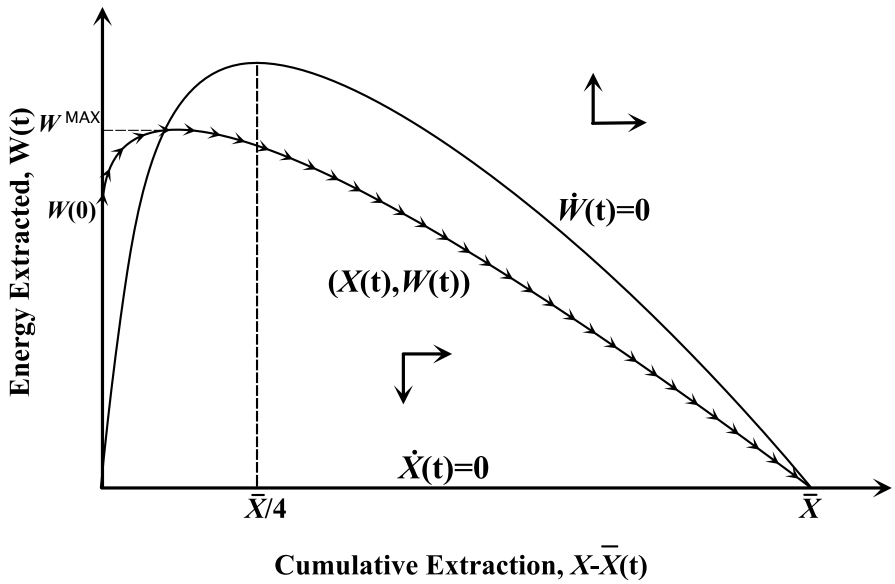


Figure 2: Optimal Extraction Rule

and to the right and cut the $\dot{W}(t) = 0$ isocline at zero slope. Once we cross this isocline, the arrows of motion tell us the extraction path must turn downwards and the transversality condition requires the path slowly approach $\bar{X}$ on the horizontal axis. Working backwards it is now apparent the transversality condition chooses the initial W_0 and this choice has to feature less extraction than that given by the peak of the $\dot{W}(t)$ isocline.

Somewhat surprisingly, extractions must at first boom and then (optimally) bust. This is true despite the uniform distribution of resources across space; despite the very conventional form of intertemporal utility and absence of demand shocks; and despite the absence of learning by doing, technological change or exploration activity. Moreover, the peak in extractions is greater the more power dense the underlying resource.

At bottom the cause is the scaling law linked to the spatial structure of the model but to understand why this is true, we need to understand the two quite different motivations governing optimal depletion. First, and not surprisingly, there are the standard Hotelling motives arising from the finiteness of the resource stock and the impatience of our planner. For example, if the costs of bringing energy to the core was zero (but there remained a finite resource stock available for use), then the shadow value of the resource *in situ* rises at the rate of time preference. Energy extracted would equal energy supplied to the core and the time profile for extractions would be given by

$$\frac{\dot{W}(t)}{W(t)} = -\frac{\rho}{\sigma} < 0 \tag{C.24}$$

where ρ is the discount rate, and σ the elasticity of marginal utility from the CRRA specification. Since optimality requires the value of marginal utilities discounted to time zero to be equalized across all periods, this is achieved by energy consumption falling at a rate proportional to time preference and the elasticity of marginal utility. This motivation follows from the finiteness of the reserves; it predicts a declining path for extractions; and, it reflects the forces identified in Hotelling's classic work (Hotelling 1931).

Second, and less familiar, are what we could call Ricardian motives. These motives come from the fact that reserves differ in their Ricardian rents: energy resources very proximate to the core have large rents and are very scarce; while very distant ones have very little rent but are abundant. Once we translate this feature of our spatial structure - via the Solow-Wan reformulation - into an implication on extraction costs, it implies that differences in Ricardian rent across reserves are now reflected in extraction costs that rise rapidly with cumulative extraction. Since any unit extracted today raises the cost of all future extractions, all else equal, it pays to shift these extraction costs into later periods. These Ricardian motives argue for a delay in extractions or what is the same, a rising path of extractions over time. Ignoring the Hotelling term given above, an extraction path that reflects only Ricardian considerations is given by

$$\frac{\dot{W}(t)}{W(t)} = -W(t)\frac{C'(X)}{1 - C(X)} > 0 \tag{C.25}$$

Equation (C.23) shows optimal extractions is the simple sum of the right hand sides of (C.24) and (C.25), and hence the interplay of these two forces produce a boom and bust path for energy production.

Descriptively, the result follows because the Ricardian motivations initially dominate

Hotelling considerations. Analytically, it follows because at the very first instant of time, energy consumption must be positive $W^S(0) > 0$, and $C'(X(0)) = -\infty$ implying $\dot{W}(t) > 0$ at least initially. And as extraction proceeds $W^S(t)$ must approach zero (the resource is finite) and $C'(X(t))$ increases. Therefore, the Ricardian forces fall over time and are eventually dominated by the Hotelling ones.

More deeply, the impact of using up the very first unit of resources on subsequent extraction costs is so costly, $C'(X(0)) = -\infty$, not because these initial energy resources have the greatest rents (which they do) but because they are so scarce in relation to the resource pool whose extraction costs are now raised. Scarcity drives the result and high rent resources are so scarce because of our scaling law. To see why, recall that energy rents fall linearly with distance, but the quantity of reserves rises with the square of distance. This implies low rent resources are abundant, and high rent resources are scarce. Increasing the power density of the energy source raises rents everywhere, but also brings in play new low rent resources at the margin of exploitation. Consequently, the motivation for pushing extractions into the future is strengthened and the peak of extractions rises.

While it is well known that the typical Hotelling's prediction can be overturned in a variety of settings, the boom and bust in extractions is a necessity in our framework and not a possibility.[7] Moreover, it follows from our scaling law which has the dual cost implications reflected in (C.18). The spatial setting provides us with a neat analytical representation for a well known empirical fact: high rent resources are scarce and low rent resources abundant - and then suggests that the logical implication of this fact is that both Ricardian and Hotelling motives drive optimal extractions. Ignoring spatial elements and their attendant impact on rents removes a key force driving optimal extractions; and taking them into account suggests an interesting parallel. Non-renewable resource extraction should be concentrated or bunched in time, just as renewables energy extractions should be bunched across space. We summarize this result in the following proposition.

[7]The first observation that extractions may boom and bust is often credited to Livernois and Uhler (1987).

Proposition C.3 *Assume intertemporal utility is of the CRRA form, then the optimal depletion path has extractions rising to a peak and then declining. Peak extractions are rising in the power density of the energy resource.*

Proof. *In text.* ■

The proof of the second part of Proposition C.3 is as follows. For any $Z = \bar{X} - X$ constant, W^{MAX} in Figure 2 is increasing in $\bar{X}$. Setting $dW/dX = 0$ we find $\frac{\partial W}{\partial \bar{X}} = 2\frac{\rho}{\sigma}Z\frac{(\bar{X}^{1/2}-Z^{1/2})-\frac{1}{2}\bar{X}^{1/2}}{(\bar{X}^{1/2}-Z^{1/2})^2}$. Therefore, $\frac{\partial W}{\partial \bar{X}} > 0$ if $\frac{\bar{X}}{4} > (\bar{X} - X)$ which always holds as the peak in the extraction always occurs to the left of the peak in the $\dot{W} = 0$ locus.

D Measuring Power Density

We start by developing the theory required for measurement. While there exist in the literature estimates of power densities for many energy sources, how these figures are constructed is unclear and rarely documented adequately. Measuring power density for some renewable resources is fairly straightforward; for example, crops dedicated to biofuels or human consumption can be turned into energy equivalents and then power flows by taking account of crop cycle, length, and area planted; similarly coppiced forests can provide stable flows of wood products for heating and cooking needs and similarly occupy well defined areas. In these cases, the renewable flows are captured by the physical quantity of fuel reaped from a resource stock.

In other cases the renewable flow does not have mass but provides either kinetic or electromagnetic energy we capture and exploit directly. In these cases the measurement is straightforward and represents the potential of these flows. For example, the power density of solar is easily estimated once we are armed with knowledge of insolation potential at a location. Similarly, wind or wave farms provide useful kinetic energy and we again find the potential power flow from the resource per unit area.

In the case of non-renewables, measurement is generally more challenging. One common method is to calculate the actual physical footprint of an energy facility's size and divide this by the current energy output. So for example, if a coal based generating station produces a constant flow of 1MW, and the mine and generating station takes up 1 km^2, then its power density is simply $[1 \times 10^6\text{W}]/[1 \times 10^3\text{m}]^2$=1W/m^2. There are several obvious problems with this method. First, the measure of power density is technology dependent. Improvements in generation technology will affect power density, and therefore power density will not be a characteristic of an energy source but rather reflect current technology in place. Second, it is difficult to know which "inputs" we should include in the measurement. For example, by including mining, crushing, and generating facilities in the calculation we make implicit decisions about which facilities to include and which to exclude. Should we also include the area taken up by transmission lines, relay stations, and other parts of the grid? In the case of oil, do we include pipelines, refineries and gas stations? If pipelines are buried and transmission lines are above ground, how do we deal with this?

How these decisions are made will materially affect the calculation. Replication of any measure produced will be almost impossible. As a result any comparison across measurement attempts will be far to reliant on the individual judgment of the researcher.

In order to resolve these issues, we present a method to measure power density for both renewables and non-renewables that is independent of technology, easy to replicate, and allows for a comparison of power densities across energy types. To do so we use restrictions from economic theory to help aggregate resources along the dimensions on which they differ, and we provide measures assuming an ideal environment where resource stocks are homogenous and where the only costs of exploitation are those arising from energy costs. Our goal is to develop measures that reflect only the physical properties of the resource and not our current, past, or future ability to reap these energy flows.

We start our discussion with the case of renewables since it is simpler to understand and relatively straightforward in practice.

D.1 Power Density for Renewables

D.1.1 Theory

Any renewable flow of energy resources produced from a renewable resource $\mathcal{F}$ [Watts] can be written as the product of the current (and steady state) physical size of the resource stock S [kg] multiplied by the energy content of the resource, e [Joules/kg] and a growth rate r [1/time]. This implies the flow is given by $\mathcal{F} = reS$.[8] The physical size of the stock can be similarly written as the product of the physical density of the resource, d [kg/m^2] times the area actually used by the resource a[m^2]. Making this substitution we obtain the flow of energy as $\mathcal{F} = (red)a$. Power density is just the flow of energy per unit area or $\Delta = \mathcal{F}/a = red$ [W/m^2].

Power density is proportional to the product of the maximal rate of regeneration, r, which measures the percentage rate of growth of the resource in an unconstrained environment. Perhaps not surprisingly, a renewable energy source that grows twice as fast has twice the power density. It is also proportional to the energy content of the fuel, e, measured in [Joules/kg], again perhaps it is not surprising that energy density matters but the specific form is of course not obvious. Finally power density also depends on a fuel's physical density, d, measured in [kg/m^2]. All else equal a fuel that produces a greater output in terms of harvest weight gives more energy.

Two special cases deserve attention. The first case applies to resources like wind or solar where there is no associated physical product. In this case we replace the stationary harvest of the resource that we used above by a measure of an average flow per unit time, and then apply energy equivalents to obtain a measure in terms of energy production per unit time. For example, average wind flow per unit area in a given location can be transformed into its kinetic energy equivalent per unit time; average solar insolation in a location is already measured in Watts per unit area terms, and measures of the energy in wave motion

[8] For example, a 100 kg forest growing at 10% per year generates 10 kg of firewood per year. Firewood contains 15 MJ per kg; and there are 31,536 x 10^3 seconds in a year. The forest provides 4.75 W on average for the year.

can likewise be measured in power density terms. In these three cases while there is no physical resource reaped, power density is simply measured by the potential energy flow these resources deliver per unit area, per unit time.

The second case arises when the harvest from a resource affects the resource stock size, and in turn the power density of the energy source. For example this is most likely to occur when resources are crops, forest land, or hydro power. In these situations it is natural to pin down the power density of the resource by assuming the energy supplier manages the resource to maximize the discounted flow of energy over time.

To investigate further it is necessary to be specific about the dynamics of resource growth. Suppose the energy source is a renewable fuel with natural growth given by $G(S)$ with $G(0) = G(K) = 0$. As is standard let G be strictly concave and let S denote the stock in physical units. For example, the energy source could be a forest, an area dedicated to biofuels, or even an area dedicated to solar or wind power. It is helpful to first take a specific example with explicit units. To that end, let $G(S) = rS[1-S/K]$ and the maximum sustainable yield harvest, H^{msy}, as one possible plan for taking from the resource. Then in perpetuity this harvest is given by:

$$H^{msy} = G(S^{msy}) = G(K/2) = rK/4 \tag{D.1}$$

where r is the intrinsic resource growth rate and K is the carrying capacity (i.e. a growth rate times a stock level). Now consider units explicitly. If K is measured in kilograms and $K =$ 100 kg, and $r = 10\%$ per unit time, then the sustainable harvest is (.1)(100)/(4)=2.5kg/unit time. If we multiply this quantity by the energy content of the fuel in [Joules/kg] denoted by e, we obtain a measure of Joules per unit time that could be harvested from the resource. Choosing to measure time in seconds, we obtain $Watts$. The final step is to divide this flow of power by the area of exploitation needed to maintain it. Since the carrying capacity is K kg, and if the fuel has a physical density, d, measured in [kg/m^2], then the total area needed for this resource flow is K/d. All this implies we can write power density for this renewable energy resource, Δ, as:

$$
\begin{aligned}
H^{msy}e &= \mathcal{F}[\text{Joules/second}] = [\text{Watts}] \\
\Delta &= \frac{(rK/4)e}{K/d} = \left[\frac{\text{Watts}}{\text{m}^2}\right] \\
\Delta &= \gamma red \qquad \gamma > 0.
\end{aligned} \tag{D.2}
$$

Power density is the simple product of three fundamental, commonly used, and potentially observable characteristics of an energy source, plus one behavioral component captured in the parameter γ. The remaining term in power density is the factor of proportionality γ which captures the intensity of harvesting. To see this note that if harvesting results in a steady state stock equal to a fraction of the carrying capacity given by κK, then $\gamma = \kappa(1-\kappa) < 1$. Harvesting zero implies $\kappa = 1$, $\gamma = 0$ and $\Delta = 0$; harvesting sufficiently high to cause extinction implies $\kappa = 0, \gamma = 0$ and again $\Delta = 0$. The example given above has $\kappa = 1/2$ and $\gamma = 1/4$. We chose this example for a particular reason: if the resource owner was interested in maximizing total energy collected over an indefinite future, then the owner would adjust their take to match that of the maximum sustainable yield. This is obvious, but more generally, if the owner discounts the value of future versus current energy flows, the optimal stationary harvest maximizing this objective would lead to $\delta = G'(S^*)$ where δ is the discount rate on future periods. This is an application of a well known result in resource economics.[9]

D.1.2 Empirics

Renewable energy has two final uses: food and fuel. One use is supplying energy to maintain bodily functions, while the other use is supplying energy for heating, light and power

[9] The optimal stock and attendant harvesting is set only by impatience and is independent of prices. In our simple example with logistic growth, power density is simply $\Delta = \gamma red$, and $\gamma = [1 - \delta/r]^2/4$ is positive as long as suppliers discount rate is not too high.

Table 1: Power Density Crops

Crop	Yield $[kg/m^2]$	Energy Content $[MJ/kg]$	Power Density W/m^2
Sugar cane	7.16	17.01	3.86
Sugar beet	4.87	9.75	1.50
Cassava	1.25	6.69	0.27
Bananas	2.05	3.73	0.24
Rice, paddy	0.43	15.5	0.21
Potatoes	1.78	3.22	0.18
Sweet Potatoes	1.25	3.60	0.14
Wheat	0.30	15.07	0.14
Barley	0.26	14.74	0.12

applications. In Table 1, we present figures on the power density of various staple crops from around the world. The figures presented for yields are estimates of "typical" yields for these crops in a system with sustainable rotation (fallow periods). As shown food crops, even staples, offer relatively small power density. Even the powerful potato offers only 0.18 W/m^2 in terms of food for fuel, but some tropical crops such as cassava (.26 W/m^2) and bananas (.24 W/m^2) provide much more. Crops that have found use as biofuels such as sugar cane (3.87 W/m^2) and sugar beet (1.50 W/m^2) have power densities one order of magnitude larger than other crops.

Table 2 presents figures on the power density of forests for six regions of the U.S. Since the productivity of forests and their composition varies so too does their power density. For example, the South Central forests have the highest growth rates (column three) whereas the North Central forests have the greatest percent of Hardwoods (column five). The power density figures are again relatively small and on the order of 0.1 to 0.15 W/m^2.

Comparing Table 1 to Table 2 we see that in general wood provides lower power density than crops. This is true even though wood density, d, and energy content, e, are much higher than the standard crop. The result follows, because a forest's rate of growth, r, is very low relative to that provided by annual crops.

Finally in Table 3 we present estimates for solar and wind energy for six regions in the

Table 2: Power Density of Wood

Forest Region	Total Acres (1,000s)	Ave. Prod. (ft^3/acre/yr)	Percent Softwood	Percent Hardwood	Average Power Dens. (W/m^2)
Northeast	79,803	57.10	25.21%	74.79%	0.10
North Central	84,215	66.54	18.72%	81.28%	0.12
Southeast	85,665	80.22	41.00%	59.00%	0.14
South Central	118,364	84.69	35.20%	64.80%	0.15
Rocky Mountain	70,969	52.00	90.29%	9.71%	0.08
Pacific Coast	75,197	81.71	89.01%	10.99%	0.12

Table 3: Power Density of Wind and Sun

Region	Solar	Wind	
	Power Density [W/m^2]	Speed [m/s]	Power Density [W/m^2]
Northeast	164	7	122
North Central	226	9	281
Southeast	217	7.5	136
South Central	219	7	115
Rocky Mountain	256	9.5	296
Pacific Coast	229	7.5	144

US. In all cases, we measure the potential provided by the resource rather than measures of our current ability to reap the resources in question. The power density of solar energy captures the yearly average amount of radiation collected by one squared meter of surface with tilt equal to the latitude of the point of measurement. Solar radiation data is reported in the first column on Table 3.

The amount of power that can be extracted from wind is a cubic function of the speed of the wind (column 2), and it is proportional to the area of the cross-section perpendicular to the velocity of the wind. Assuming one squared meter cross-section gives us the amount of power extracted at any given velocity. If we further assume one meter squared of land is used by one meter squared of cross-section then we can find the power density of wind. This is what we show in the last column of Table 3.

It is interesting to note the huge differences across these tables measuring the power density of crops or timber versus the raw energy flows in solar. Photosynthesis is, even in the best environments, a very inefficient process taking solar power from the sun and then via nature's capital equipment turning it into biomass. Estimates on this efficiency vary but a common estimate is below 1% efficiency, and this is similar to what a very naive comparison of these tables might suggest.

D.2 Power Density for Non-renewables

D.2.1 Theory

The method we presented earlier to calculate the power density of renewables cannot be applied directly because non-renewables are obviously very different. For example, the flow of energy obtained from non-renewables is proportional to the *change* in the resource stock over time, and non-renewables are not spread over vast areas but instead found in subsurface deposits of considerable thickness. To proceed we deal with these issues in turn. First, assume the resource *was* distributed uniformly on the surface of a two dimensional plane. Then the change in the resource stock would equal the physical flow of the resource, which we will denote by ϕ[kg/m^2·s], times the area exploited, a[m^2]. This gives us the kilograms mined per second. Multiplying by energy content implies the power density of a non-renewable resource distributed on the surface, could be written as:

$$W = e\phi a \tag{D.3}$$

$$\Delta = W/a = e\phi \tag{D.4}$$

This makes perfect sense: the flow of energy comes from the change in a stock; the magnitude of the resulting energy flow is determined by the stock's energy content; the change in the stock is equal to the product of the area extracted times the physical flow of the resource mined; and dividing by area exploited, we obtain the result. For future purposes it is also useful to note that the zero rent margin for this non-renewable is again equal to $R^* = e/\mu g$ since ϕ kg of the ore is being transported for each square meter mined per second; and hence

$c = \mu g \phi$. Therefore, we have a quality/quantity result similar to Proposition 2.

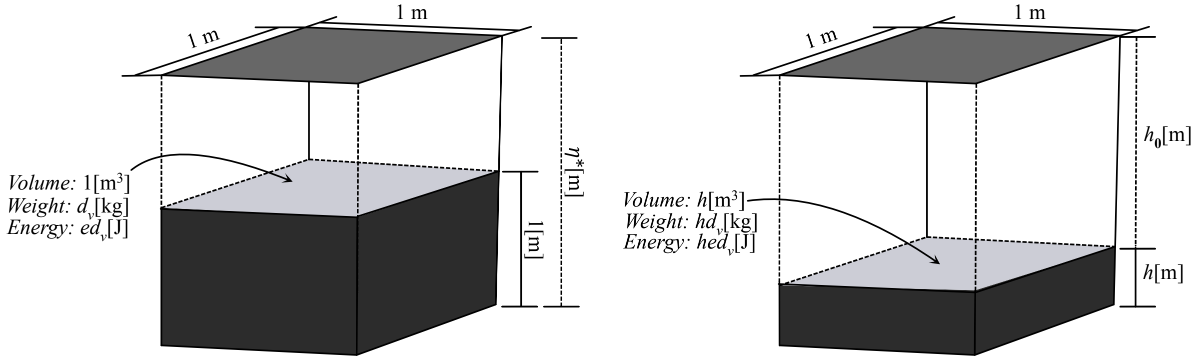
Unfortunately, this measure of power density only holds in situations where non-renewables are distributed very thinly across Earth's surface. For example, shallow deposits of the non-renewable peat may well fit this description. In the vast majority of cases however non-renewables are not thinly distributed as surface deposits. To tackle this problem we need to develop an appropriate measure of available resources in subsurface deposits. Energy resources will be brought to the surface and then transported to the core only if they provide positive energy rents; and this implies there is now an extensive margin of exploitation for any mine in terms of depth.

Consider a hypothetical resource owner with resource rights to one meter squared of surface area in our exploitation zone of homogenous resource quality. If this resource owner extracts a 1m^3 cube of energy resources with energy content e[J/kg] and volumetric density d_v[kg/m^3], then this cubic meter has mass of d_v[kg] and weight of $d_v g$[N]. The total energy contained in this cube would be simply ed_v[J] and if it was exhausted in one second the power delivered would be ed_v [Watts].

Now consider resources contained beneath this 1m^2. If resources are located at distance η[m] from the surface, then the work needed to bring them to the surface would be just $gd_v\eta$[J] since work must be done to offset gravity. Energy rents in this case are just $ed_v - gd_v\eta$. Resources where the energy cost of extraction equals their entire energy content produce zero energy rents and are located at depth $\eta^* = e/g$. The net energy available must account for the energy costs of extraction and is found by the following simple integration.

$$\Delta = \int_0^{\eta^*} [ed_v - gd_v\eta]\, d\eta = \int_0^{\eta^*} ed_v\,[1 - (g/e)\eta]\, d\eta = e^2 d_v/2g \tag{D.5}$$

where e and g are as defined before, while d_v is the volumetric density of the resource in $[kg/m^3]$.



(a) Ideal Deposit

(b) Characteristic Deposit

Figure 3: Non-renewable deposit

D.2.2 Empirics

Non-renewable resources are found in all continents. Reservoirs can be shallow or deep; they can be large or small; and fuels can be of different qualities. For example, the "quality" of coal as measured by its energy content depends on the depth of burial. Lignite coal is the coal with the lowest energy content and it is formed when Peat is buried between 200 and 1500 meters during the coalification process. The process of Lignite coalification increases the energy content from around 13 MJ/kg for Peat to 16 MJ/kg but it almost doubles its volumetric density from around 355 kg/m^3 to 700 kg/m^3. The formation of Bituminous coal occurs at greater depths (between 2500 and 6000 meters), and in the process the energy content increases to 32 MJ/kg and the volumetric density again increases to over 900 kg/m^3. Anthracite coal is formed deeper than other ranks of coal (between 6000 and 7500 meters). It is the purest form of coal (up to 96% pure carbon) with an energy content of 35 MJ/kg and volumetric density as high as 950 kg/m^3. Erosion, earthquakes and volcanic activity can expose these deposits or bury them even more. For example, the maximum depth of deposits in Argentina is 600 meters for sub-bituminous coal and minimum seam thickness of 1.8 meters, Ukraine has deposits of bituminous coal that are 1600 meters deep and have a minimum seam thickness of 0.5 meters. Australia has Lignite coal deposits at a maximum depth of 300 meters and minimum seam thickness of 3 meters. Many places in the world also feature surface deposits.

Oil and Natural Gas deposits can also be found at various depths and the quality of the resources also depends on the depth of burial, although the process of formation is quite different. The formation of petroleum occurs anywhere between 2500m and 4500m. The temperature and pressure at this depth combine to transform decayed organic matter into oil. Just as with coal, erosion and other geological forces can bury the oil deposit further or expose it. Oil deposits can be right at the surface of Earth's crust or they can be as deep as 12,000 meters; extreme temperatures below this point are likely to bake most of the crude in the deposit.

These complications mean that our theoretical construct of a continuous ideal deposit running η meters in depth from the surface is rarely obtained. To amend our calculation for any specific deposit, suppose the upper limit of the deposit is located at a distance h_0 from the surface, and suppose the thickness of the deposit is h, so the maximum depth of the deposit is $h_0 + h$; as shown in Figure 3.(b). Because the resource rights are over one meter squared the dimensions of the deposit are simply h[m^3]. In this case to measure the power density of this deposit, we use:

$$\Delta = \int_{h_0}^{\min\{h_0+h,\eta^*\}} [ed_v - gd_v\eta]\, d\eta = \int_{h_0}^{\min\{h_0+h,\eta^*\}} ed_v\,[1 - (g/e)\eta]\, d\eta$$

$$\Delta = \min\{h\,[d_v e - gd_v(h_0 + h/2)]\,, e^2 d_v/2g - h_0\,(ed_v - (gd_v h_0)/2)\} \tag{D.6}$$

We now use this formula to calculate the power density for different types of coal in Table 4 below. To give an idea of the resulting power densities we present a set of alternative estimates. We present estimates assuming a deposit's thickness, h, is either 1 or 10 meters; estimates for deposit depths, h_0, of 0 and 10,000 meters; and since coal varies in quality due to the influence of time and pressure we present estimates for a precursor (peat), low quality brown lignite coal; and higher quality bituminous coal.

As shown, higher quality coal has both higher energy content and greater volumetric

Table 4: Power Density of Different Coals

Resource	Energy Content [MJ/kg]	Volumetric Density [kg/m^3]	Power Density [GW/m^2]			
			h_0=0 m		h_0=10000m	
			h=1m	h=10m	h=1m	h=10m
Anthracite	33	1400	46.2	462	46.2	462
Bituminous	33	913	30.1	301	30	300
Lignite	16	865	13.8	138	13.8	138
Peat	14	310	4.34	43.4	4.31	43.1

density. Time and pressure increase both and this implies given our formula that higher quality coal is significantly more power dense than its precursor peat. We find the deposit depth has little effect on the energy contained in the mine, and hence comparing the power density of resources at zero depth versus those at 10,000 meters reveals only small differences. And finally, the power densities are very large. To understand the scale of the energy shock created by a movement to coal we compare these figures with similar figures for wood in Table 2.

Several observations are in order. First, a naive comparison of the figures in these tables reveals large differences in power densities which naturally suggests the introduction of coal would be a massive energy shock to an economy where wood provided most thermal energy. While we believe that this was indeed the case, a comparison of these figures requires careful attention. The power densities measured for renewables represents the flow of power we capture on our one square meter of space *ad infinitum.* The wind blows, the sun shines, crops grow and forests mature providing a potential, and constant in steady state, flow of energy we can reap. In the case of non-renewables, the energy flow depends on how fast we extract the resources contained in one square meter. A straight ahead comparison of power densities across these tables implicitly assumes we are consuming all of the non-renewable energy resources from one square meter (and all of its subsurface deposits) per second. This is incredibly fast. A better interpretation of the differences in power densities across these tables is that the potential for reaping power from non-renewables is just vastly greater.

A second observation is that any existing low cost transportation option has a bigger impact on energy supplies when the resource is more power dense. This is an implication of the magnification effect shown in Proposition 2. Therefore, existing rivers and roads would magnify the impact of coal far more than they magnified the impact of biomass resources.

Third, any comparison of wood and coal using just energy contents will severely underestimate the impact of a move to coal. Coal offers perhaps 25 MJ/kg while wood 15 MJ/kg, which means we would be willing to transport coal further than wood given similar transport technologies, but not much further. But comparisons like these miss an important point captured in the complementarity effect of Proposition C.1. Because coal is so much more power dense, the incentive for expanding the existing exploitation zone is much larger. And it was these endogenous improvements (canals, railroads, ports) in response to coal that, combined with its power density, created a huge and long lasting energy shock.

D.3 Data used in the Tables.

D.3.1 Table 1: Crops.

Data on 2011 global crop yields in Table 1 is provided by the Food and Agriculture Organization of the United Nations' (FAO) FAOSTAT Production database (2013). Global crop yield is calculated automatically by the FAO database by dividing annual crop production by area harvested. Production and harvest figures are reported to the FAO by individual countries via questionnaire or national agriculture publications. Detailed information on the collection methods for this data can be found in the Metadata section of the FAO Statistical Yearbook (2012, 357-358) as well as in the entry for agricultural production on the FAO's Methods & Standards Webpage.

Data on energy content for the crops given in Table 1 is provided by the United States Department of Agriculture, Agricultural Research Service's National Nutrient Database, Release 25. Energy content is given in kilocalories (kcal) per 100 grams, which is converted to MJ/kg by multiplying by 4.1868 kJ/kcal and then dividing by 100 to convert kJ/100g into MJ/kg.

D.3.2 Table 2: Forests.

Average power density for a forest in each of the United States regions listed in Table 2 is calculated by the formula:

(Average Productivity in Region (hg/m^2/year) $\times$ Power Density of Average Hardwood (J/m^2/year) $\times$ Percentage Hardwood in Average Forest) + (Average Productivity (kg/m^2/year) $\times$ Power Density of Average Softwood $\times$ Percentage Softwood in Average Forest)

The power density of average hardwoods and softwoods is calculated from data provided by Engineering Toolbox. To be precise, power density for individual tree species is calculated by dividing the recoverable heat value of a dry cord of wood (million BTU/cord) by the weight of a dry cord (lb/cord) given by Engineering Toolbox to get the recoverable heat value per pound of wood (million BTU/lb), which is then converted to recoverable heat value per kilogram of wood (MJ/kg), also called potential heat value per kilogram of wood, using conversion factors of 1 lb = 0.4536 kg and 1 million BTU = 1055.06 MJ. Calculations in Table 2 are based on an average dry hardwood density of 35.52 lb/ft^3 with a 14.89 MJ/Kg potential heat value, and an average dry softwood density of 27.45lb/ft^3 with a 14.87 MJ/Kg potential heat value.

Average hardwood power density is then calculated by averaging the potential heat value per kilogram of wood for five species (aspen, cottonwood, red oak, red maple and white oak) considered common in American forests by the USDA (2007, 62). An identical calculation is done for four species of softwood (hemlock, ponderosa pine, balsam fir and white pine) also considered common in American forests by the USDA (2007, 62) to get an average softwood power density. It should be noted that the variation in potential heat value per kilogram between different species of tree is always less than 1%, so average hardwood and softwood heat values do not vary much if different species of tree are chosen for the calculation.

Data on percentage of softwoods and hardwoods in an average forest is calculated using data on hardwood and softwood volumes provided by the United States Department of Agriculture (USDA) (2007, 206-208). Data on average productivity classifications for forests is also taken from the USDA (2007, 160-162). Forests are categorized by the USDA by cubic feet of wood per acre per year (cu. ft.) into one of five different classifications: 120+ cu. ft., 85-119 cu. ft., 50-84 cu. ft., 20-49 cu. ft. and 0-19 cu. ft. A simple average of the extreme values in each productivity class is taken to represent average productivity for forests within that class, while forests with a productivity exceeding 120 cu. ft. are capped at a productivity of 120 cu. ft. The productivity of an average forest in any region of the United States can then be calculated after making these assumptions. Forests on reserved land, or with an average productivity between 0-19 cubic feet per acre per year, are omitted from the stocks of total hardwood and softwood by the USDA, and hence are also omitted from the calculation of average forest productivity.

D.3.3 Table 3: Solar and Wind.

Solar energy calculations based on data from the National Renewable Energy Laboratory. The data can be downloaded from http://www.nrel.gov/gis/data_solar.html. We use the annual average direct normal irradiance for the lower 48 states and Hawaii PV 10km Resolution 1998 to 2009. The data are originally in kWh/m^2/day. We transform them to W/m^2 multiplying by 1000 to get Watts and diving by 24 to eliminate day from the calculation.

Wind energy calculations are also from NREL and can be downloaded from http://www.nrel.gov/gis/data_wind.html and it is using wind speeds at a height of 50 meters. The exact relation between speed and power is given by the following equation $W = \frac{1}{2}\rho v^3$, where ρ is the density of air and v is the speed. Increasing the height of the tower increases speed. For example increasing the height of the wind tower from 10 to 50 meters increases speed by approximately 25% which in turn would increase power density. Here we also assume wind speeds at 50 meters. To calculate power we assume air density at sea-level and temperature of 15^{o}C which is $\rho = 1.225 kg/m^3$ (Gipe P. ,2004). We assume a cross-sectional area of 1m^2 sitting on 1m^2 of land. That is, we assume the radius of the wind turbine is approximately 55% of the area it sit on.

For both solar and wind energy we use GIS to aggregate at the state level calculating the area-weighted average of the cells contained in each state. We then obtain the regional area weighted average using the same regions described above.

The six regions of the United States listed in Table 2 and Table 3 are defined by the USDA (2007, 1) as follows:

- Northeast: Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, West Virginia
- North Central: Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, Wisconsin
- Southeast: Florida, Georgia, North Carolina, South Carolina, Virginia
- South Central: Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Oklahoma, Tennessee, Texas
- Rocky Mountain: Kansas, Nebraska, North Dakota, South Dakota, Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
- Pacific Coast: Alaska, Oregon, Washington, California, Hawaii

E Endogenous Transport Costs

We solve the energy producer's problem in two stages. In the first stage transportation costs are minimized by choosing how much distance to cover by land and how much distance to cover by river. In the second stage energy rents are maximized. The cost minimization problem is given by:

$$\min_{r_1,r_2} cr_1 + \rho c r_2, \text{ subject to } r^2 = r_1^2 - r_2^2 + 2rr_2 \cos\theta \tag{E.1}$$

where r_1 is the distance travelled by land and r_2 is the distance travelled by road. The constraint follows directly from the law of triangles with r_1 being opposite to the angle θ.We can replace the constraint in the objective function to find the optimal distances travelled by land and by road:

$$r_1^* = \frac{\sin\theta}{(1-\rho^2)^{1/2}} * r \text{ and } r_2^* = \cos\theta - \frac{\rho\sin\theta}{(1-\rho^2)^{1/2}} * r. \tag{E.2}$$

If the distance r_2^* is strictly positive, the supplier deviates to the road, otherwise the supplier goes straight to the core. We can solve for the critical value of θ that separates the suppliers that go straight to the core from those who deviate to the road:

$$r_2^* > 0 \text{ if and only if } \theta \leq \cos^{-1}\rho \equiv \bar{\theta} \tag{E.3}$$

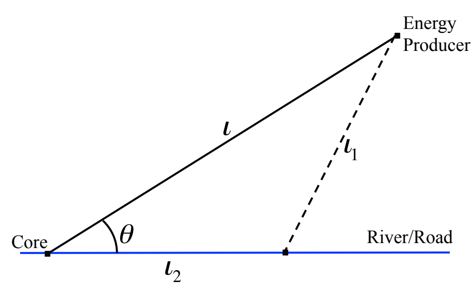
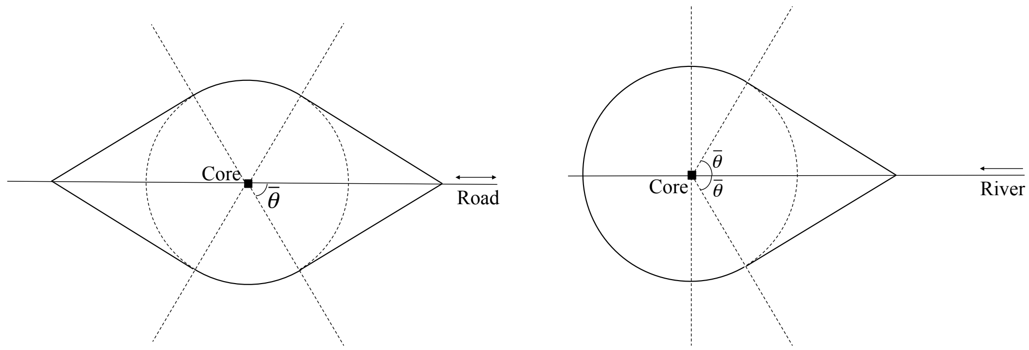


Figure 4: Transport via Road or River

Energy suppliers located at any angle $\theta < \bar{\theta}$ are "close" to the low cost alternative and choose to use it. Since $\rho = \cos(\bar{\theta})$, we know that as $\rho \to 0$ everyone deviates, since it is so cost effective. Alternatively, as $\rho \to 1$, the road offers no advantage and no one uses it.

The second part of the energy producer's problem is to decide whether or not to take its energy to the core. An energy producer situated a distance r from the core and forming an angle θ with the road will go to the core if the net energy supplied to the core is positive; i.e., if there are positive energy rents at this location. Energy supplied by this producer is given by $W^S = \Delta - c(r_1^* + \rho r_2^*)$ Replacing equations (E.2) in the previous equation makes energy rents a function we find $W^S = \Delta - c(\theta)r$ where $c(\theta)$ given by (**??**).

We depict the exploitation zone in the river and road case in the two panels of Figure 5 assuming $\rho < 1$.



(a) Area of Exploitation with a Road (b) Area of Exploitation with a River

Figure 5: Rivers and Roads

Just as before, the total energy supplied to the core is found by "adding up" all energy rents.

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

The first integral represents energy coming from suppliers who are close enough to the road to use it in transport. The second integral represents the energy coming from those who travel directly to the core. We have multiplied the integrals by 4 since we are adding up over the quarter circles of $\pi/2$ radians.

Integrating and simplifying gives us a net energy supply much like that we had before:

$$W^S = \frac{1}{3}\frac{\Delta^3}{c^2} g(\rho) \tag{E.4}$$

$$g(\rho) = \pi + 2(\tan(\bar{\theta}) - \bar{\theta}) \geqslant 0 \tag{E.5}$$

where the function $g(\rho)$ is positive and monotonic $g'(\rho) < 0$, approaches infinity as ρ goes to zero and approaches π as ρ goes to 1.

References for Online Appendix

British Petroleum. 2015. *BP Statistical Review of World Energy.* London: British Petroleum.

Central Intelligence Agency. 2013. *The World Factbook: Field Listing: Pipelines.* Available at: https://www.cia.gov/library/publications/the-world-factbook/fields/2117.html, Accessed August 7, 2015.

Chakravorty, U. and D.L. Kruice. 1994. "Heterogenous Demand and Order of Resource Extraction." *Econometrica*, 62(6): 1445-1452.

Exxon Mobil. 2010. *The Outlook for Energy: A View to 2030.*

Energy Information Administration. 2015. *Spot Prices for Crude Oil and Petroleum Products.* Available at: http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rbrte&f=m.

Engineering Toolbox. *Wood and Combustion Heat Values.* Available at: http://www.engineeringtoolbox.com/wood-combustion-heat-d_372.html, Accessed March 12, 2013.

Food and Agriculture Organization of the United Nations. 2013. *FAOSTAT Production.* United Nations. Available at: http://faostat.fao.org/site/339/default.aspx, Accessed March 12, 2013.

Food and Agriculture Organization of the United Nations. 2012. *FAO Statistical Yearbook 2012 World Food and Agriculture.* Rome: United Nations.

Food and Agriculture Organization of the United Nations. 2012. *Methods & Standards.* Available at: http://faostat3.fao.org/home/index.html#METADATA, Accessed March 12, 2013.

Fortune Magazine. 2013. *Global 500.* Available at: http://fortune.com/global500/,

Gipe P. 2004. "Wind Power: Renewable Energy for Home, Farm, and Business." Chelsea Green Publishing Company ISBN: 978-1-931498-14-2.

Herfindahl, O.C. 1967. "Depletion in Economic Theory." In *Extractive Resources and Taxation,* edited by Mason Gaffney, 63-90. Madison, Wisconsin: University of Wisconsin

Press.

Hotelling, H. "The Economics of Exhaustible Resources." 1931. *The Journal of Political Economy,* 39(2): 137-175.

Livernois JR and R Uhler. 1987."Extraction Costs and the Economics of Nonrenewable Resources" *Journal of Political Economy* 95(1): 195-203

Solow, R.M. and F.Y. Wan. 1976. "Extraction Costs in the Theory of Exhaustible Resources." *The Bell Journal of Economics*, 7(2): 359-370.

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

United Nations Conference on Trade and Development. 2014. *Review of Maritime Transport.* Geneva: United Nations.

United Nations Statistics Division. 2006. *Standard International Trade Classification, Revision 4.* Available at: http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=28, Accessed March 10, 2013.

United States Department of Agriculture. 2007. *Forest Resources of the United States.* Available at: http://www.fs.fed.us/nrs/pubs/gtr/gtr_wo78.pdf, Accessed March 12, 2013.

United States Department of Agriculture. 2012. Agricultural Research Service. *USDA National Nutrient Database for Standard Reference, Release 25.* United States Department of Agriculture. Available at: http://ndb.nal.usda.gov/ndb/search/list, Accessed March 12, 2013.

World Trade Organization. 2014. *International Trade Statistics.*

Available at: https://www.wto.org/english/res_e/statis_e/its2014_e/its2014_e.pdf.