NBER WORKING PAPER SERIES

BACK TO THE FUTURE OF GREEN POWERED ECONOMIES

Juan Moreno Cruz
M. Scott Taylor

Working Paper 18236
http://www.nber.org/papers/w18236

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2012

We are very grateful to seminar audiences at several universities, and would like to thank, without implicating, Bob Allen, Herb Emery, Oded Galor, Gene Grossman, Arvind Magesan and Jim Wilen for comments. The usual disclaimer applies. Excellent research assistance was provided by Jevan Cherniwchan, Nolan Derby and Rui Wan.

Back to the Future of Green Powered Economies
Juan Moreno Cruz and M. Scott Taylor
NBER Working Paper No. 18236
July 2012
JEL No. N0,Q4,Q5

ABSTRACT

The purpose of this paper is to introduce the concept of power density [Watts/m²] into economics. By introducing an explicit spatial structure into a simple general equilibrium model we are able to show how the power density of available energy resources determines the extent of energy exploitation, the density of urban agglomerations, and the peak level of income per capita. Using a simple Malthusian model to sort population across geographic space we demonstrate how the density of available energy supplies creates density in energy demands by agglomerating economic activity. We label this result the density-creates-density hypothesis and evaluate it using data from pre and post fossil-fuel England from 1086 to 1801.

Juan Moreno Cruz
School of Economics
Georgia Institute of Technology
221Bobby Dodd Way
Atlanta, Georgia 30332
juan.moreno-cruz@econ.gatech.edu

M. Scott Taylor
IEE Canada Research Chair
Department of Economics
The University of Calgary
2500 University Drive, N.W.
Calgary, AB T2N 1N4
CANADA
and NBER
mstaylor@ucalgary.ca

1 Introduction

In the not so distant past - and perhaps not so distant future - economies will rely, once again, on renewable energy. In the not so distant past, solar power captured by wind flows and biomass production were the sole energy sources fueling human existence. The direct and indirect fruits of solar power provided wood for heating, crops for eating, and fodder for draft and game animals. Students of this era, particularly economic historians, debate the role these nature given limits played in determining the economic limits to land use (i.e. diminishing returns) identified by the giants of classical economics - Ricardo, Malthus and Adam Smith. Students of our current era, particularly energy economists, look to a not so distant future where post-industrial green powered economies may survive on the direct and indirect fruits of solar power captured by wind turbines, wave motion, and silicon panels. And similarly, many energy economists today debate the role green power can play in powering a modern developed economy in the future.

These two disciplines are separated by several hundred years in their respective periods of study but share a common interest in understanding how the characteristics of renewable energy sources determine the limits of an organic - or green powered - economy.[1] The purpose of this paper is to take a step back into the past to consider what our future may look like if all energy is supplied by renewables. We do so by developing a simple general equilibrium model that incorporates elements of an energy system thought to be important by both economic historians and energy economists, and then ask how the physical limits of renewable energy sources determine the economic limits of a green powered economy.

We develop a model with four key features. First, we allow energy sources to differ in their power density. Power density measures the flow of energy a source can provide per unit area needed for its exploitation and maintenance. In the case of renewables power density comes directly from the area requirement for a resource (say timber or biofuels or solar power), together with its ability to provide a flow of energy measured in Watts.

It is somewhat surprising that a key characteristic of renewable energy, ignored in most economic analyses, is the vast areas needed to provide power at anything approaching a commercial scale. This areal constraint is well known in the case of biofuels since the land tied up in their production is key to their cost, but area considerations are also important in hydro power which displaces millions of people around the globe, and important to both solar and wind power when they are asked to provide significant energy flows. The density of available energy resources has however received considerable attention in the analyses of

[1]In a series of contributions Sir Tony Wrigley refers to economies based wholly on the direct and indirect flows of solar power as "organic economies". Although this usage is not universally accepted we will use it interchangeably with its, somewhat misleading, alternative "green powered economies".

economic historians, because the interplay of energy density and transport costs is commonly held to be responsible for the small city sizes in pre-industrial times.[2] To examine these arguments we develop a spatial model of energy exploitation, transportation and use that allows us to define and introduce the power density concept into economics.

Second, energy economists and economic historians alike agree that every fuel source has unique attributes that often requires - and always leads to - the introduction of a set of facilitating goods that aid in its collection, transport, and conversion. For example, scythes, carts, and horse collars were all important facilitating goods in organic economies just as today solar collectors, wind turbines, and DC converters are important for today's renewables. Since these goods are "intermediate" in the supply chain somewhere between the exploitation and collection, and the final use of energy, we will refer to them as intermediate goods. The role of these intermediate goods is to facilitate the conversion of raw energy inputs into final energy services at a lower cost than otherwise. Our model gives a key role to these energy converters since they are the backbone of any real world energy system.

Third, the combination of the density of an energy source, the set of facilitating intermediate goods, and the pattern of production and consumption in an economy defines, what we will call, an "energy system." While referring to a set of interrelated demands and supplies as a system is standard fare in economics, energy system experts and energy historians have a more inclusive definition in mind. Their energy system is one where complementarities between and across system components link consumption, transport and exploitation in a mutually reinforcing way creating elements of increasing returns and path dependence. To capture and hopefully understand the potential role played by these system wide complementarities thought to be important in past, present, and presumably future energy systems, we adopt a framework where increased specialization in the set of intermediate goods increases the efficacy of the energy system. By design the extent of these system wide complementarities is bounded and, by construction, easily manipulatable.

Fourth, our spatial setting allows us to model the costly exploitation and transport of energy explicitly. We assume transport is costly because it requires energy, and calculate the work needed to move real resources across space. Transport requires both energy and other inputs and we assume both inputs are essential. Since energy has to be collected, refined, transported, etc. it is difficult to imagine a world where energy could be delivered at zero energy cost to consumers; it is also impossible to think of energy being collected, moved or distributed without complementary inputs such as wheel barrows, wagons, pipelines or

[2] The link between city size and power density is very clearly made in many contributions of Vaclav Smil (see for example Smil (2006, 2008)); it is also a recurrent theme in the work of Wrigely (see Wrigley (2010) for one example) although Wrigley does not use the term power density. See also our discussion of Nunn and Qian (2011) in section 6.

transmission lines that comprise the energy system.

Within this context we develop several results. In our simplest model with only one energy source and a social planner who maximizes delivered power, we show how the concept of power density affects energy supply. We find energy supply is a cubic function of power density. This finding not surprisingly, has large implications. For example, doubling the density of an energy source implies eight times the energy supply. Alternatively in a world with two energy sources, an energy source twice as dense as its counterpart would be responsible for 8/9 of the energy supplied. These results reflect the power of the cubic, but the cubic arises from a scaling law once we take into account the costs of transporting energy in a world where geography matters. The critical role we find for power density continues throughout our analysis and corroborates the key role economic historians have given to energy constraints in an organic economy.

Once we move to a market economy demand considerations - such as market size - also play an important role. Larger markets support the entry of more specialized inputs and create a more efficient energy system. In some cases we find energy prices fall with market size. The reason is simple: while a larger demand for energy pushes up prices, it also leads to improvements in the energy transportation system which, at least for some period, can swamp the impact of sourcing energy from higher marginal cost sources. As a result both the price of a Watt of power and the price of energy services delivered to consumers can fall with market size before eventually rising.

It may then come as no surprise that utility and consumption per capita can at first rise and then fall with market size. Indeed we find that more dense energy sources provide greater peak consumption per capita and they do so at larger population levels. This feature of the consumption per person profile creates an obvious energy-density driven incentive for the agglomeration of activity. In this sense density in energy supply can create its own density in energy demand - a result we term the *density-creates-density hypothesis*.

Our model is highly abstract, largely static, and of very small dimension. It fits no real world economy past nor present, but does carry with it several potentially testable hypotheses. To move towards an empirical evaluation we develop three additional results. We first show how to measure power density by decomposing it into three common, observable, and fundamental features of energy resources: their renewability; their energy content; and their dispersal in the environment. This makes the measuring of power density straightforward in theory, although clearly still difficult in practice.

Second, since any real world landscape is cut by roads and rivers we extend our analysis to examine how these low transport cost options affect our results. Surprisingly we are able to show an interesting equivalence. A city on a river or road enjoys an energy supply identical to that of a land-locked and isolated city with more dense resources at hand. Higher power density of available resources is equivalent to having improved transportation options.

Finally, population centers are not isolated states and hence we must allow for the possibility that cities compete for both population and energy resources. To do so, we couple our market model with a simple model of Malthusian population dynamics where consumption per capita and population density jointly determine birth and death rates. The Malthusian application gives us a simple mechanism to sort population across space and investigate the density-creates-density hypothesis. We then evaluate its predictions by reviewing the population history of England from 1100 to 1800 through the lens of our theory.

Our work is related to contributions coming from three large and largely disjoint literatures: economic history; energy economics; and economic geography. Our simplest model with transport costs for energy collection generates results much like those in Von Thunen's Isolated State (1826). We likewise solve for a circular area of exploitation whose margins are determined by a zero rent condition, but our work differs since we generate transport costs by taking explicit account of the density of the energy resources, the frictions in transport, and the physical forces required to transport objects. As a result we derive a primitive per meter transport cost - similar to that assumed by Von Thunen and popularized by Samuelson as iceberg transportation costs - but now explicitly linked to fundamentals.

We also differ when we extend the simplest model to a market economy where energy demanders and suppliers interact and determine the features of the energy system. Here we find that the increasing returns generated by complementarities in the energy system combines with transport costs to generate a reason for economic activity to agglomerate. Therefore, in contrast to Von Thunen's work there is a reason for a village or city to exist. Not surprisingly our Energy-Economy model is related to earlier contributions in economic geography. Like Henderson (1974, 1980) the interplay of increasing returns and transport costs generates an optimal city size. We differ however by linking the density of economic agglomeration to the density of the energy sources supplying the city. And similar to Fujita et al (2000) we generate increasing returns at the economy level from increased specialization in economic activity.

Work in economic history is also related. Perhaps most importantly, Wrigley (2010) argues the low density of available energy sources in the U.K made urbanization and further progress impossible. He argues for a view of the Industrial Revolution where positive feedbacks between the density of fossil fuels, the resulting urbanization, and eventual technological progress drive the transformation of the UK economy. Similarly, Smil (2008) argues that the low density of biomass based fuels kept villages small and the market size for any innovation too low to foster growth. Recent work by Nunn and Qian (2011) is also directly on point as they link the introduction of a new high density energy source (the lowly potato) to population growth and urbanization in the Old World. Important work specific to the period we study is Wrigley (1969, 1998), Allen (2009), and Clark (2009).

Our work is far more abstract than these data-heavy accounts by historians. It offers to this literature a formal model of a green or organic economy that links the limits of urbanization to primitives of the energy source, other economic attributes, and rivers and road characteristics. It shows how migration, differential mortality, and the availability of energy sources jointly determine city size and economic well being.

Finally, there is a large and growing literature on the economics of renewable energy and a healthy debate over its likely role in any future energy system.[3] Contributions vary from specific industry studies examining the costs of renewable energy to more macro approaches evaluating the potential limits to growth. None of this work has, to our knowledge, recognized that we have literally millennia of green economy experience under our belt already, and as a result our examination of past green economies should provide guidance for the future.

The rest of the paper is organized as follows. In section two we introduce the concept of power density and discuss how energy use over time has evolved in this context. This section answers why it may be valuable to incorporate the concept of power density into economic models. In the next section we show how to develop a simple model linking power density to energy supply. This section develops definitions and establishes production possibilities within our Only Energy model. Section four introduces an Energy-Economy model where energy demand from firms interacts with energy supply to determine the rate of exploitation for a single renewable energy source. Sections 5 and 6 contain our Malthusian application and a discussion of data and hypothesis testing. A short conclusion ends the paper. All proofs and lengthy calculations are relegated to the appendix.

2 Why Power Density?

The density of an energy source can be measured in a variety of ways. One method is to rank sources according to their energy content measured in Joules/kg. Under this metric, the energy contained in a kg of crop residue is less than that in a kg of wood, which is in turn less than that of coal, and less than that of oil. Another method is to recognize that some energy resources take up large areas for their maintenance and exploitation (crop residue, wood, game animals) in comparison to what might be called punctiform resources where energy resources are highly concentrated geographically (coal, fast moving rivers and

[3]See for example the divergent views in Bryce (2010) and Droege (2009).

uranium). To capture both the energy content of sources and their geographic requirements, some authors have employed the metric power density.

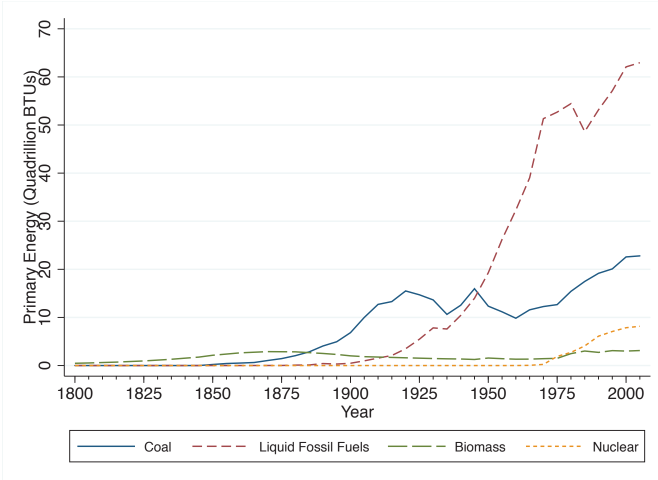
The density of energy resources has long been a key concept in the analyses of economic historians, because, as we will show, humans have always turned to increasingly dense fuels during our industrialization. Straw and dung - to fuel wood and charcoal and coal - to oil and natural gas and nuclear - all represent transitions to denser and denser fuels where density refers to the flow of energy these resources provide (measured in $Watts$ of power) per m^2 needed for their maintenance or exploitation.[4]

Density is important to capture because the concentration and agglomeration of economic activity first in villages, then cities and now mega-cities has meant the density of energy consumption has grown tremendously. Energy demand in a major city like New York, Tokyo or London requires a flow of power in the range of 100 $Watts/m^2$ (Smil, 2008). Transporting and concentrating energy sources to meet this demand is a critical and major function of any developed economy.

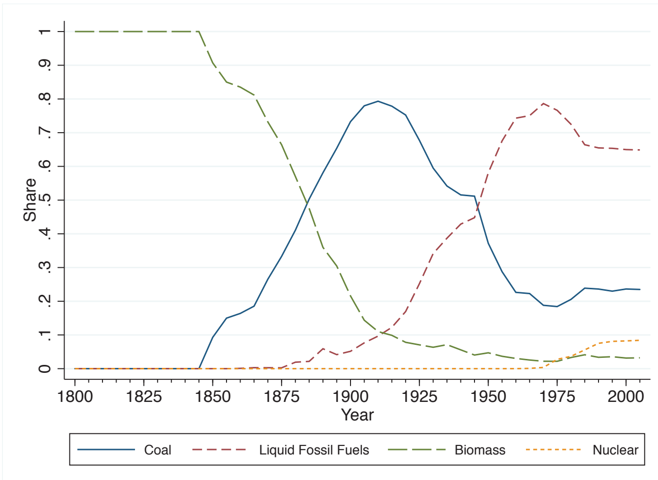
To appreciate this relationship, consider the three panels shown in Figure 1. All panels employ a historical time series for U.S. primary energy consumption from 1800 to 2005. The first panel shows how the composition of US energy consumption has changed over time. It is interesting to note that new energy sources are added but old energy sources do not disappear. For example, the primary energy generated by biomass in the U.S. was approximately the same in 2005 as it was one hundred years earlier. The second panel shows the same data but now in terms of shares. The share data suggests an almost smooth repeating pattern of energy substitution across the last two hundred years, one that many have taken as evidence of the inevitability of our move into renewables. Finally, in the third panel this same data (excluding nuclear power) is used to construct the average energy content and average power density of U.S. primary energy consumption indexing both to 100 in 2000.[5] This figure casts the data in a different light. The history of primary energy consumption in the U.S. has been one of a dramatic rise in the energy content and power density of consumption. As shown the average energy content of consumption has risen

[4] For example, the newest wind turbine can in a very windy location capture 65 $megajoules/m^2$ and this is, what an engineer would call, its power density. Similarly, dry wood as an energy content of 15 $megajoules/kg$; a healthy forest delivers 20 kg/m^2 in perpetuity, and hence dry wood has a power density of 300 $megajoules/m^2$

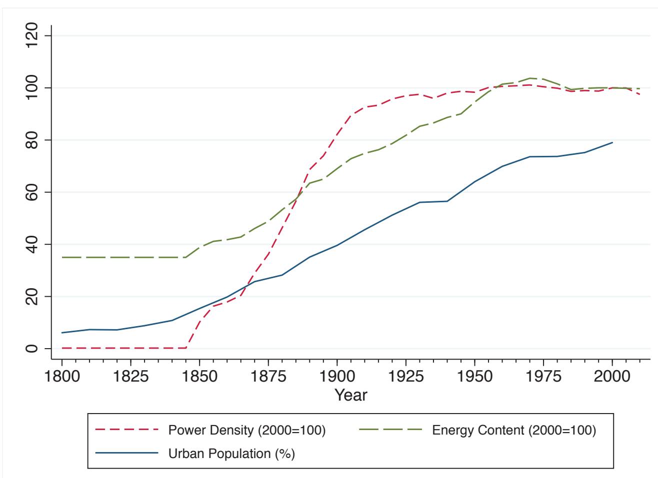
[5] The three panels of Figure 1 are constructed using data from five sources: (i) energy consumption data from the US Energy Information Administration Annual Report of 2009, (ii) data on energy content from the World Energy Council's Survey of Energy Resources (2010), (iii) data on power density from Smil (2010), (iv) urban and rural population data for period 1800-1990 from the US Census Bureau's 1990 Census of Population and Housing (1990) and (v) urban and rural population data for the year 2000 from the U.S. Census Bureau's American Fact Finder. The urban and rural population data was only available for 10 year intervals; the data presented in the figure were linearly interpolated to fill in the missing 5-year intervals.



(a) U.S. Primary Energy Consumption



(b) U.S. Energy Use in Shares



(c) Weighted Average of U.S. Energy Content and Power

Figure 1: U.S. Energy Transitions

more than 2.5 fold, while the average power density of consumption has risen almost 100 fold over the period. Also graphed in the third panel is the share of the U.S. population living in urban centers of more than 2500 people. The close connection between these series is intriguing. A major goal of this paper is to understand how they may be related and its major question is to ask whether the density of available energy sources causes density in economic activity.

Given the last panel in the figure it is perhaps not surprising that economic and energy historians have repeatedly highlighted the role power density plays in limiting city size and economic progress. A few quotes should suffice to convey the importance they ascribe to this measure. For example, energy historian Vaclav Smil states in a discussion on the limits of city size:

"...pre-industrial cities needed at least ten and up to thirty Watts per square meter of their built up area. This means that if they relied entirely on wood, they needed nearby areas of between 50-150 times their size in order to have sustainable phytomass" (Smil, 2006, p. 73).

Smil goes on to note how the cost of transporting energy from such large areas is why city size was so limited in pre-industrial times.

But power density is not only thought relevant to city size in early human history, economic historian Sir Tony Wrigley has in a number of articles and one book length treatment argued that overcoming the limitations of green power were key to the industrial revolution in the U.K.

"As long as supplies of both mechanical and heat energy were conditioned by the annual quantum of insolation and the efficiency of plant photosynthesis in capturing incoming solar radiation, it was idle to expect a radical improvement in the material conditions of the bulk of mankind" (Wrigley, 2010, p. 17).

And finally, several present day energy analysts have pointed to the low power density of renewables to argue against their usefulness in any future energy system. For example, energy expert and provocateur Robert Bryce argues:

"The essential problem with renewables is that they fail the first test of the Four Imperatives: power density...The weak power density of renewables has become so apparent that the Nature Conservancy,...., recently coined the term energy sprawl'" (Bryce, 2010, p. 84).

It is of course difficult, if not impossible, to examine these arguments further without a formal model that allows us to understand the role power density may play in an economic system.

3 The Only Energy Model

We start by introducing a simple model of energy exploitation where energy is the only input and only output of production. We do so to introduce our assumptions on geography and costly transport, to define power and energy density, and to derive several preliminary results useful to our subsequent analysis.

We assume consumption and production activities are located at an economic core while potential energy sources are distributed in the surrounding space. The economy's core contains all of its production and consumption units but is zero dimensional. The exploitation zones where energy sources can be found are two dimensional planes allowing us to employ definitions of area, distance, and density. Distance is meant to capture any and all costs incurred when incremental amounts of energy are exploited. For some energy sources, distance is more than a metaphor. Timber resources for example may only be available some distance from the core; and for some renewable energies - solar or wind power, or biofuels - measures of area are important to consider. For other energy sources, like fossil fuels, area considerations are less important, but here increasing distance can reflect the difficulties firms have in accessing incremental resources due to well depth, weather conditions, and non-standard geological formations.

Introducing geography into an energy model allows us to assess the arguments made regarding power density and the roles it may play in a coupled economy-energy system. It also allows us to distinguish our work from much of traditional thinking in energy and resource economics where questions concerning the optimal depletion of a given resource base have received much attention. Instead, we treat energy resources as essentially limitless but with increments coming at a rising marginal cost.

The area exploited to find and collect energy is related to the power demands of the core measured in watts, [W], and the power density, Δ, measured in [W/m^2] of the particular energy source exploited. If the flow of needed power is W, then the area exploited must equal:

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

where EX is measured in m^2.

3.1 The Geography of Exploitation

If the energy resources are distributed uniformly, then minimum cost search implies energy will be collected from a circular area with the economy's core at its center. To calculate the average distance resources travel to the core, we need to relate the distance a resource travels with the density of resources at that distance. Suppose we exploit a circular area defined by a radius R. The density of resources at any given radii r less than R is simply given by the circumference at this distance, $2\pi r$, divided by the sum of all radii up to and including the limiting radius of R. Since the total area exploited is πR^2, the density function for energy resources must be given by $f(r) = 2r/R^2$. The average distance any resource is transported is then simply the distance r times the density of resources at that distance. We will refer to this as the average carry distance, ACD, and it is given by:

$$ACD = \int_0^R r[2r/R^2]dr = \frac{2}{3}R \tag{2}$$

While we assume energy resources are evenly distributed across (a two-dimensional) space, it should be clear that other options are easily dealt with. For example, suppose the core was located with resource availability in mind (not an unlikely scenario), then resources may be dense near the core and less dense further away. Alternatively, we could allow for the equally likely possibility that a new energy source is located further from the core than was the older established source (think of the potential wind resources on the Great Plains or solar power farms in Arizona!). In all these cases, an appropriate adjustment of the density $f(r)$ is warranted and will generate average carry distances less than or greater than that given in (2).

Since Area isπR^2 and EX represents the area of exploitation, we have $ACD = (2/3)(EX/\pi)^{1/2} = (2/3\sqrt{\pi})(W/\Delta)^{1/2}$ using (1). ACD is measured in meters; to calculate the work done in transporting energy resources we need to find what we will call "Total Carry", measured in kg-meters. This is found by multiplying the average carry distance by the total weight of resources transported. Total energy resources transported is equal to the area under exploitation times the density of fuel over this area. Therefore, Total Carry is given by:

$$\begin{aligned}
\tilde{TC} &= ACD \times EX \times d = [\mathrm{m}][\mathrm{m}^2][\mathrm{kg/m}^2]{=}[\mathrm{kg} \times \mathrm{m}] \\
\tilde{TC} &= \left(\frac{2}{3\sqrt{\pi}}\right)\left(\frac{W}{\Delta}\right)^{1/2} EX.d \\
\tilde{TC} &= \left(\frac{2d}{3\sqrt{\pi}}\right)\left(\frac{W}{\Delta}\right)^{3/2}
\end{aligned} \tag{3}$$

where again use has been made of (1), and d is the (uniform) physical density of this resource in [kg/m^2]. Total carry is the number of kilogram-meters covered in transporting energy to the core when the power requirements are W and the energy source has density Δ.

Not surprisingly, the greater are the power requirements, the greater are total carrying costs. Less obvious is that marginal (carrying) costs are strictly increasing in power. As power demand rises, larger and larger areas for exploitation must come at the fringe of already sourced areas.

3.2 The Physics of Transport

Production possibilities are determined by the amount of power that can be supplied to the core. This section sets out maximum supplied power as a function of energy resource characteristics. A first step in doing so is to recognize that power collected does not equal power available for use in the core. The collection and transportation of resources, like everything else, requires power. The net power supplied to the core is equal to total power collected in the exploitation zone, W, minus the power needed for transportation. In obvious notation, $W^S = W - W^T$.

To calculate the power used in transport, a little high school physics is required. Recall Work is equal to force, f, times distance, x, or work is $W_k = f.x$. Force is in turn equal to mass, M, times acceleration g; as any mass moved horizontally must overcome the force of gravity as mediated by friction in transport.[6] All this implies:

$$\begin{aligned} W_k &= \mu[M \times g]x = [\text{Newton.meters}] = [\text{Joules}] \qquad (4) \\ W_k &= \mu g[M \times x] \\ W_k &= \mu g[\text{Total Carry}] \end{aligned}$$

where μ is the coefficient of friction, g is the acceleration of gravity at 9.81m/s^2, and W_k is then measured in Joules. This work is done per unit time since power is flow as is the flow of labor services and the flow of useful output. If we measure time in seconds, then the flow of work, W_k, measured in Joules per second is now power requirements measured in Watts.[7] So the total cost of delivering the flow of power W to the core is given by $TC(W)$. Using (3) and (4) we can write this total cost as:

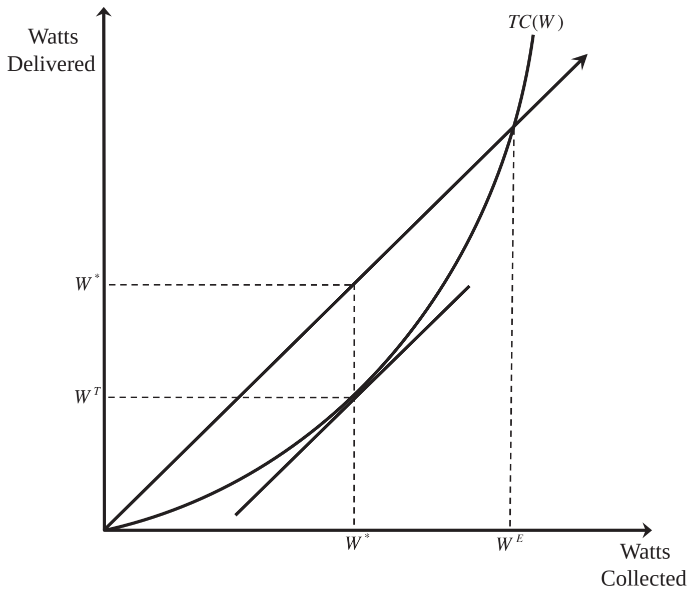
$$W^T = \mu g \widetilde{TC}(W) = TC(W) \qquad (5)$$

Note $TC(W)$ is a strictly convex function where $TC(0) = TC'(0) = 0$. The calculation of supplied power W^S is shown in Figure 2. The vertical height of $TC(W)$ at any W gives the

[6] 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.

[7] Expending a Joule of energy per second means you are delivering power of one Watt.

Figure 2: Optimal Energy Exploitation



power used in transporting to the core the resources represented by their associated power on the horizontal axis. By construction, the gap between the 45 degree line and $TC(W)$ represents supplied power. Therefore extending our exploitation to collect power W^E results in zero power supplied since all of it would be exhausted in transport. In contrast, W^* maximizes supplied power because the vertical distance between W^* and $TC(W^*)$ represents maximum of $W^S = W^* - TC(W^*)$.

One important feature of the figure is that it shows the energy system is *productive*. A *productive* energy system is one that can provide positive supplied power to the economy: that is, there exists extraction levels for which the power needed for delivery to the core do not exhaust power collected. The system here is always productive regardless of the costs of transport. Even if we push the coefficient of friction towards infinity making transport very costly in terms of energy, the slope of the total carry curve at zero Watts, is still zero. In contrast the slope of the 45 degree line remains one even at zero Watts, and hence for any finite coefficient of friction, there will always exist an opportunity to deliver positive power to the core.[8]

It is straightforward to show that the maximum supplied power is increasing in power density Δ as this shifts $TC(W)$ outwards, and increasing, for the same reason, when the transportation system becomes more efficient in the physical sense (μ falls). Associated

[8]This does not imply that all landscapes have sufficient resources to support life or a productive economy. We discuss these issues in later sections.

with any W^* is an area of exploitation EX^*. If we focus on efficient (in terms of maximized supplied power) outcomes then we have a simple model of how the density of an energy resource affects power available for use in the core.

When supplied power is maximized, total power is given by:

$$W^* = \frac{\pi\Delta^3}{[\mu g d]^2} \tag{6}$$

The optimal solution is a cubic function of power density. The intuition is illuminating. Suppose we increase power density but leave the area of exploitation fixed; then supplied power should rise proportionately with power density; i.e. appear with power 1 in an expression like (6). But a higher power density also implies the marginal cost of exploitation falls. With lower marginal costs, exploitation rises and the radius of search increases. Since area is proportional to the square of radius, total power rises with the square of power density when the extensive margin of exploitation changes. Putting this logic together it is apparent why power density is so powerful in determining outcomes.[9]

To find the portion of power used in transport and the remainder that is supplied to the core, evaluate total carry at W^* and subtract to find:

$$W^S = \left(\frac{1}{3}\right) W^*, \quad W^T = \left(\frac{2}{3}\right) W^* \tag{7}$$

Two thirds of the energy collected is spent in transport leaving only one third to supply the power needs of the core.

The area of exploitation is also a function of power density since at the optimum:

$$EX^* = W^*/\Delta = \pi\frac{\Delta^2}{[\mu g d]^2} = \pi[R^*]^2 \tag{8}$$

$$R^* = \frac{\Delta}{\mu g d} \tag{9}$$

Having a more dense energy source means the energy costs of exploiting far away regions is less: dense energy sources feed further exploitation. The area of exploitation for a fuel twice as dense as another is four times larger. This feature seems to ring true. Villages in the Middle Ages rarely sourced fuel wood from more than a few miles away.[10] Coal

[9]This result is not a function of our assumed circular exploitation zone and is in fact an example of a scaling law. Expanding margins in two dimensions always leads to a squared term for power density. Section 5.2 demonstrates this for non-circular zones that arise when the landscape is cut by rivers, roads or canals.

[10]Towns on rivers would be an exception since their transport costs are far less. See our discussion of rivers in section 5.2.

however has been transported hundreds of miles on barges, wagons and trains for much of the last two centuries. Today oil is sourced from some of the most inhospitable climates in the far reaches of the planet and from wells drilled literally miles deep below the ocean surface. Equation (8) and (9) tells us this is no surprise - dense energy sources beget large exploitation zones.

To see exactly why this is true, think about the costs of moving one Watt of power just one meter. Since friction is a constant returns process this is all we really need to understand. One Watt of power, takes up $1/\Delta$ square meters of area which in turn implies the energy resources it represents must weigh d/Δ kilograms. Moving this mass one meter, and overcoming friction, requires a flow of power of $\mu gd/\Delta$. Therefore, $\mu gd/\Delta$ is the number of Watts needed to transport one Watt worth of an energy source with power density Δ, one meter. It is now apparent that (9) identifies those marginal energy resources which are located R^* meters from the core and provide zero net energy supply: that is, $1 - R^* \frac{\mu gd}{\Delta} = 0$. The more dense are the energy resources, the lower are the transport costs of obtaining a marginal watt of power, and not surprisingly the energy exploitation zone must grow.

Our discussion of marginal energy resources suggests an alternative method of finding net supply. Net power supplied must come from adding up, what we could call, "energy rents". These rents, $1 - \frac{\mu gd}{\Delta} r$, are collected at distances $r \leq R^*$ from the core. To add them up consider this two step procedure. Along any ray from the core, there are Δ watts of power every meter and transporting these resources at distance r from the core yields a density of $\Delta \left[1 - \frac{\mu gd}{\Delta} r\right]$ net watts of power at distance r from the core. Add these resources along our ray over all distances less than R^*, and then 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:

$$\begin{aligned} W^S &= \int_0^{2\pi} \int_0^{R^*} v\Delta[1 - \frac{\mu gd}{\Delta} \cdot v] dv d\varphi \qquad (10) \\ W^S &= \frac{\pi \Delta^3}{3\left[\mu gd\right]^2} \end{aligned}$$

It is sometimes useful to find power supply using this method, once the zero rent margin R^* has been identified. It is also useful in highlighting our formulations similarity to Von Thunen (1836). In Von Thunen's work crops of various types generate rent gradients falling with distance to the city center; and the market economy's demands for wood, cereals, or meat sets relative prices and land use. Aggregating across energy sources is in fact easy, if we assume their collection costs are independent. In this case, two independent energy sources with power densities Δ_1 and Δ_2 would provide net power in total of:

$$W^S = W_1^S + W_2^S = \frac{\pi \bar{\Delta}^3}{3\left[\mu g d\right]^2}$$

where $\bar{\Delta} = \sqrt[3]{\Delta_1^3 + \Delta_2^3}$. It is as if the economy had only one energy source with density $\bar{\Delta}$ rather than two. We will exploit this result and aggregate all energy resources into one. The simplicity this affords allows us to tie a set of fundamentals comprising $\frac{\mu g d}{\Delta}$ to unit distance - or iceberg - transport costs, and thereby provide micro foundations for Von Thunen's classic work.[11]

3.3 What determines power density?

Thus far we have taken power density as a given fixed constant. This may be a reasonable representation of reality for solar and wind and water resources where the flow is beyond human manipulation. It may however be less accurate in situations where energy suppliers could vary their take of an energy resource (timber, crops, livestock) and thereby affect their future productivity. If what we take today affects future productivity of energy resources (a not unlikely result), then our energy maximizing planner may want to take this into account and choose a path that maximizes some measure of aggregate energy flows over time. While this may seem rather esoteric at present, it proves useful to consider this possibility now.

To understand how this additional margin of adjustment may affect our results, we need to examine the concept of power density more closely. Recall power density is the rate at which an energy resource can deliver a flow of energy services per unit time divided by the area needed to find, collect, and maintain the resource. It is found by multiplying the quantity of a fuel we can collect per unit time by the fuel's energy content per unit fuel, and then dividing by the area under exploitation. This calculation will produce Δ measured in the units [Watts/m^2]. Watts are a measure of energy flow per unit time, and meters squared are an indicator of area. An example is instructive especially for the renewable resource case as all green energies are at bottom governed by renewable resource dynamics.

Suppose a potential fuel is renewable with natural growth given by $G(S)$ with $G(0) = G(K) = 0$, and G being strictly concave. For example, this 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. Take logistic growth as our example (i.e. $G(S) = rS[1 - S/K]$), and the maximum sustainable yield harvest, H^{msy}, as one possible

[11] See Samuelson (1983) for a modern exposition of Von Thunen.

plan for taking from the resource. Then in perpetuity this harvest is given by:

$$H^{msy} = G(S^{msy}) = G(K/2) = rK/4 \tag{11}$$

where r is the intrinsic resource growth rate and K is the carrying capacity. Now consider units explicitly. If K is measured in kilos 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, Δ as:

$$\Delta = \frac{(rK/4)e}{K/d} = \left[\frac{\text{Watts}}{\text{m}^2}\right]$$
$$\Delta = \gamma red \qquad \gamma > 0.$$

Power density is the simple product of three fundamental, commonly used, and potentially observable characteristics of an energy source, and one behavioral component captured in the parameter γ. Power density is proportional to the product of an energy source's 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 not surprising that energy content 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.

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 $\gamma = \kappa(1-\kappa) < 1$. The example given above has $\kappa = 1/2$ and $\gamma = 1/4$. We chose this example for a particular reason: if the planner was interested in maximizing total energy collected over an indefinite future, then its clear the planner would adjust their take to match that of the maximum sustainable yield. This is obvious, but more generally, if the planner discounted 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.[12] In a stationary environment (like ours), the intensity of harvesting is set to equalize the return from the resource to the energy supplier's discount rate. Leaving a marginal unit of the resource in situ raises future harvests by $G'(S^*)$ in perpetuity; taking it out today rather than tomorrow earns a return of δ. And therefore, in our example our perfectly patient planner who only cares about total energy flows choses the maximum sustainable yield since $0 = G'(S^*)$.[13]

The point of this exercise is that although power density may in some cases be subject to manipulation, in a stationary environment it is fully determined by parameters of the system. It is simply linked to four potentially observable magnitudes: the rate of resource regeneration, the energy content of the fuel provided, its physical density, and the discount rate applied to future rents by energy suppliers. This result is in fact general and not a function of our logistic example, and the constancy of power density will also carry over when we introduce a market economy with price determination.

3.4 Summary

The only energy model, although simple, tells us three things that carry forward with only small amendment throughout our analysis. First, differences in power density across energy resources are likely to create large differences in energy supply. Given the cubic formulation even relatively small differences in power density matter greatly to outcomes. Second, we are likely to search far and wide to exploit dense energy resources. The very density of the energy resource we seek, fuels our efforts at obtaining more. Finally, even though we have assumed the potential supply of energy resources is limitless, the net supply of energy resources is what matters. And the net energy supply will always be constrained by the physical necessity of using energy in the costly exploitation and transportation of energy.

4 The Energy-Economy Model

Thus far we have said nothing about how a market economy would interact with the technological possibilities described. While there are many ways to embed our energy model in a market economy setting, three large pieces of the puzzle are missing from the analysis thus far. First, what is the demand for power in the core and how does this demand set prices and determine exploitation zones? Second, how are the costs of transportation determined

[12] See for example Neher (1990, p.28) for a discussion.

[13] 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.

by the cost of other factors used in transportation? And finally, how can we incorporate elements of our only energy model into a tractable general equilibrium framework that allows for the positive feedbacks thought to be important in real world energy systems?

4.1 Tastes, Technology, Endowments

There are L identical consumers each endowed with one unit of labor. Their utility is defined over the consumption, C, of a final output good Y which provides energy services. Consumers' income comes from providing labor services and from reaping resource rents. Since nothing important hinges on consumer numbers we will model a representative consumer with labor endowment L. Utility of our consumer is strictly increasing and strictly concave in consumption:

$$u = u(C), \quad u' > 0, \quad u'' < 0 \tag{12}$$

Final output Y, is produced *inter alia* by a set of goods intermediate in the supply chain between energy exploitation and final consumption. Naturally we will refer to them as intermediate goods. We take them to represent a set of capital goods and consumer durables tailored to a specific energy source. These intermediate goods are of course the backbone of any real world energy system since they are the means by which energy services are delivered to firms and consumers in the economy.

While any one firm or individual may employ only a subset of these intermediates to provide energy services, aggregate output in the economy would be a function of all available intermediates. One simple, tractable, and common way to capture the reliance of final output on the set of intermediates is to adopt a constant elasticity Dixit-Stiglitz specification for final goods production.

$$Y \equiv \left[\sum_{i=1}^{n} x_i^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1 \tag{13}$$

where each of the x_i are differentiated intermediate goods produced by a single firm. Together these intermediates provide the service flow represented by the final output Y.[14]

It is important to our analysis that the set of available intermediates is endogenously determined. To do so we follow standard practice and assume $\sigma > 1$ to allow for their production by monopolistically competitive firms. The number of firms is then determined by a free entry condition. Each of these firms incur both fixed and variable costs to produce any intermediate. Let a_i be the activity level of firm i, then the production of a typical

[14] The service interpretation is well known and well used in the international trade literature. See for example, Markusen (1990).

intermediate at level x_i requires a_i units of activity. That is:

$$a_i = \alpha + \beta x_i \qquad \forall i \tag{14}$$

where α represents the set up cost to produce a typical intermediate while β represents its constant marginal cost of production in terms of activity.

Activity is the employment of labor in conjunction with power. Specifically, the activity of each firm, a_i, requires both a flow of energy and labor given by the following constant returns function:

$$a_i = f(l_i, W_i) \qquad \forall i \tag{15}$$

We take (15) to be a Cobb-Douglas function with a labor share in the value of output given by z. Constant returns is innocuous given increasing returns are already assumed in the mapping from activity to intermediate output; the constant shares assumption is useful in identifying specific regions of the parameter space to characterize results.

Combining these assumptions we see that final energy services to consumers (and producers) are provided by a set of differentiated processes that use both conventional inputs (labor) and energy. In the abstract, these processes are energy converters that provide the services of heat, light and power; and any energy converter that transforms energy into services with the help of conventional inputs is broadly consistent with our formulation (a toaster, flashlight, and laser printer all fit the bill). In the concrete, we have ignored the particularities of these converters by assuming all intermediate goods are symmetric substitutes; and importantly, we have adopted a formulation where the greater is the number of ways energy services can be delivered, the higher is the system's overall productivity in translating energy and inputs into services.

4.2 The Demand for Energy

The demand for energy comes from two sources. Energy is used in the production of intermediates, as well, energy is used directly in transportation much as it was in the only energy model. It proves useful to treat energy used in transportation as an input use problem, and we leave it to our discussion of energy supply in the next section.

To determine energy demand arising from production, note that the minimum cost production of Y implies demands for each x_i intermediate good is of the form:

$$x_i = A p_i^{-\sigma} \qquad \text{for some } A > 0. \tag{16}$$

Individual producers, taking A as given, adopt a fixed mark-up rule over marginal costs making intermediate prices, p_i

$$p_i = \left(\frac{\sigma}{\sigma - 1}\right) \beta c(w, p^W) \tag{17}$$

where $c(w, p^W)$ is the unit cost function dual to (15), w represents the wage rate of labor and p^W the price of a Watt of power. Free entry implies that profits to intermediate goods producers will be eliminated in equilibrium and hence this implies:

$$p_i x_i = c(w, p^W)[\alpha + \beta x_i]$$

As is well known, the mark up rule and zero profit condition neatly solve for the level of output and activity per firm. Shephard's Lemma then returns the implied factor demands for labor and energy per firm. Together these are simply:

$$\begin{aligned} x_i^* &= \frac{\alpha}{\beta}(\sigma - 1) \qquad a_i^* = \alpha\sigma \\ l_i^* &= \alpha\sigma\partial c(w, p^W)/\partial w \qquad W_i^* = \alpha\sigma\partial c(w, p^W)/\partial p^W \end{aligned} \tag{18}$$

Since all firms are identical, total energy demanded (for production purposes) is just n times firms specific demand W_i^*.

The number of firms, n, is endogenous and depends on the overall market size as indexed by L. To solve for the number of firms use the firm specific demands for labor and note full employment requires $\sum_{i=1}^{n} l_i^* = L$. By substituting for l_i^* we can solve for the number of firms:

$$n = L/l_i = \frac{L}{a_i^* z c(w, p^W)/w} = \left(\frac{1-z}{z}\right)^{1-z} \left(\frac{L}{a_i^*}\right) \left(\frac{w}{p^W}\right)^{1-z} \tag{19}$$

because $wl_i = za_i^* c(w, p^W)$. We can now add up across firms to find aggregate energy demand, W^D (Watts demanded) as:

$$W^D = nW_i = L\left(\frac{1-z}{z}\right)\left(\frac{w}{p^W}\right) \tag{20}$$

The properties of demand follow directly from our assumptions. Constant returns implies an increase in what we will refer to as market size, L, raises demand proportionately. Constant factor shares implies a unitary price elasticity of demand; energy demand would however respond negatively to a price increase under any suitable generalization.

4.3 The Aggregate Production Function

An important feature of our framework is that its aggregate behavior depends on characteristics of the energy system. In some cases, the aggregate behavior is very neoclassical; in others, its behavior reflects increasing returns coming from positive feedbacks in the energy system. Not surprisingly, it is useful to understand some of the forces in play even before we attempt to solve for the general equilibrium. To do so recall that final output is given by (13). By substituting from (18) we can write final output solely as a function of the number of firms producing intermediate goods and parameters. To go further use (19) and (20) to write the number of intermediates as a constant returns function of labor input and watts delivered to the core. Putting these results together and setting power supply to demand, we can solve for the economy's aggregate production function relating final energy service output to the economy's (inelastically supplied) endowment of labor and the (endogenously determined) supply of power:

$$Y = BL^a W^b \ \textit{with}\ a + b > 1, \qquad (21)$$
$$a = \frac{z\sigma}{\sigma - 1},\ b = \frac{[1-z]\sigma}{\sigma - 1},\ B > 0,$$

While energy is an endogenous variable, this doesn't stop us from characterizing the returns to scale when both inputs are varied. It is apparent from (21) that the aggregate production function is homogenous of degree $\sigma/[\sigma - 1] > 1$: therefore, the aggregate production function *always* exhibits increasing returns to its two inputs - labor and (delivered) power. Thus far we have assumed that σ is greater than one in order to support monopolistically competitive firms producing unique intermediates; but, without any further restrictions the production function's degree of increasing returns is unbounded. Very simply, as σ approaches one from above, the degree of increasing returns approaches infinity. Unbounded increasing returns is of course an undesirable feature in any economic model since it makes a mockery of resource constraints.

Typically we bound increasing returns to eliminate undesirable or unbelievable outcomes. One natural way to bound increasing returns is to assume diminishing marginal returns to each input. That is, by assuming $a < 1$ and $b < 1$. Diminishing returns to each input individually seems hardly novel or controversial but it is helpful and this is what we will do here.[15] One implication of assuming diminishing returns to each input is that the sum of their exponents $\sigma/[\sigma - 1]$ must be less than two. Therefore, diminishing returns provides a simple bounding condition on the aggregate production function's degree of increasing

[15] In all of that follows we assume each input exhibits diminishing marginal productivity: $a < 1$, and $b < 1$.

returns that is independent of factor shares: $\sigma > 2$.

While we have bounded increasing returns we have not eliminated their role. In much of what follows we focus on two cases. We will refer to one of these as the Almost Neoclassical case; the other as the Increasing Returns case. To understand why we obtain two cases, we first write the endogenous supply of power as a function of market size, $W = W(L)$. Then differentiate 21 with respect to market size and rearrange slightly to obtain:

$$\frac{dY}{Y} = \left[\frac{z\sigma}{\sigma - 1} + \frac{[1-z]\sigma}{\sigma - 1}\left[\frac{dW}{dL}\frac{L}{W}\right]\right]\frac{dL}{L} \tag{22}$$

This derivation highlights the link between the energy system's ability to provide additional power in response to an increase in market size, with the degree of aggregate returns the economy as a whole will exhibit. Our discussion above on the aggregate production function's degree of increasing returns concerned the typical hypothetical when both inputs are varied in proportion; in contrast, 22 characterizes the equilibrium degree of increasing returns the economy exhibits when its one primary factor, L, is varied and the economy's endogenous power supply adjusts accordingly. As shown this equilibrium behavior depends on the energy system's ability to provide additional power, dW/dL, since this response determines the magnitude of the term in large brackets.

The two possibilities are now fairly clear. In one case, the model generates almost neoclassical results despite the existence of positive feedbacks and increasing returns. This case is associated with a weak response of power to a change in market size and occurs when power responds less than proportionately with market size. The model behaves very differently from a neoclassical benchmark when power responds more than proportionately with market size. In this second case, the economy's behavior will look anything but neoclassical.

It is an interesting feature of the model that the same conditions which make increasing returns in the aggregate production function stronger (i.e. a relatively low value for σ that renders $\sigma/[\sigma - 1]$ large) also ensure the power supply response to a change in market size is large. And a relatively high value of σ that renders $\sigma/[\sigma - 1]$ small, also ensures the power supply response to a change in market size is small. Therefore, our two cases (by and large) can be described as the Almost Neoclassical case with large σ and weak increasing returns; or, the Increasing Returns case with small σ and strong increasing returns.

We now turn to an examination of supply.

4.4 The Supply of Energy

We assume a sole energy supplier chooses the area of exploitation to maximize flow profits taking prices as parametric.[16] The energy supplier's optimization problem is to choose total power to maximize profits:

$$\underset{W}{\text{Max}}\ \Pi = p^W W - (p^W + p^C)TC(W) \tag{23}$$

where W is total power embodied in the energy resources exploited, p^C is the price of the final output, and $TC(W)$ is total carry as defined before.

There are three assumptions reflected in (23). The first concerns how we have dealt with the issue of net power. By construction, the firm chooses total power to maximize profits, but supplies to the market only that power net of use in transportation. It in effect sells W watts of power at p^W but buys back from itself $TC(W)$ watts to pay for transportation back to the core.[17] By making power requirements proportional to total carry we have stayed true to the physics of transport highlighted in the Only Energy Model.

The second assumption is that transportation uses both energy and the final output good as inputs. The assumption provides a link between the efficiency of providing energy services and the cost of energy exploitation. We mentioned previously how lower energy prices created entry into intermediate good production and raised productivity; now we see that this higher productivity (created, for example, by an energy price decline) will imply lower exploitation costs and greater energy supply. This assumption closes the positive feedback loop in the energy system.

The final, and most important, assumption is the limited substitutability between energy and final output in transport. Every unit of energy used in transportation must be matched with a unit of final output to provide transportation. This assumption reflects a critical modeling decision. It ensures the energy costs of energy exploitation are bounded strictly above zero. Therefore, even if final output becomes very cheap relative to the price of power, the energy input needed per Watt of power collected can never approach zero even in the limit. This ensures energy is an essential input in energy collection.

The Leontief specification we adopt delivers simple results, but the critical assumption is a production function for transport with an elasticity of substitution between energy and conventional inputs of less than one. Since the energy costs of transport come from the

[16] Assuming a single producer is innocuous, since the natural alternative of assuming atomistic energy supplies produces identical results. Assuming the supplier maximizes flow profits is innocuous when energy density is not affected by the degree of energy exploitation.

[17] Electric transmission pays a transport cost via transmission losses; natural gas pipelines run their turbines on natural gas; diesel fuel runs diesel fuel tanker trucks, etc.

forces needed to overcome friction, one way to think about this assumption is that it respects the physical reality that friction cannot be eliminated - even in the limit - regardless of the scale of conventional inputs applied in transportation.

With these assumptions in place, the firm's maximization problem has a simple solution:

$$\left(\frac{p^W}{p^W + p^C}\right) = TC'(W) \tag{24}$$

Let s be equal to the share of energy costs in extracting and transporting resources to the core, then $s = p^W/[p^W + p^C]$ and the solution to our firm's problem becomes $s = TC'(W)$. In our Only Energy Model we had $s = 1$ and hence $1 = TC'(W^*)$ determined the region of energy exploitation.

Since $TC'' > 0$, we can invert (24) to write total power as $W^* = \psi(s)$ and power supplied as $W^S \equiv \psi(s) - TC(\psi(s)) = g(\psi(s))$. It is now simple to show:

$$\frac{dW^S}{ds} = [1 - s]\psi' > 0 \tag{25}$$

where use has been made of (24). Not surprisingly, the supply of power to the core is increasing in the relative price of energy to a unit of final output.

Since energy demand is a function of relative factor prices, we need to put supply in similar terms. To do so recall that final good production is a constant returns activity, and hence active production requires its price equal unit cost. This implies

$$p^C = c(p_1, p_2, ...p_n) = \left[\sum_{i=1}^{n} p_i^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{26}$$

Since all firms are identical, and inputs are priced via (17), we can substitute and rearrange to find:

$$p^C = \left[\sum_{i=1}^{n} p_i^{1-\sigma}\right]^{\frac{1}{1-\sigma}} = \frac{\beta\sigma}{\sigma - 1} c(w, p^W) n^{\frac{1}{1-\sigma}} \tag{27}$$

The unit cost of energy services reflects not only the cost of labor and energy, but also the set of intermediate goods used in producing the service flow. An increase in the set of intermediate goods tailored to a energy source raises overall productivity and therefore lowers the costs of energy services delivered.

To simplify further use the linear homogeneity of $c(w, p^W)$ and (19) to substitute for the number of intermediate goods to find:

$$\frac{p^W}{p^C} = \frac{s}{1-s} = z\left(\frac{1-z}{z}\right)^{\frac{\sigma(1-z)}{\sigma-1}} \left(\frac{\sigma-1}{\beta\sigma}\right)\left(\frac{L}{a^*}\right)^{\frac{1}{\sigma-1}} \left(\frac{p^W}{w}\right)^{\frac{z\sigma-1}{\sigma-1}} \tag{28}$$

And hence we can now link energy supply to relative factor prices to find:

$$\begin{gathered}\frac{dW^S}{d(p^W/w)} = \psi'[1-s]\frac{ds}{d(p^W/w)} > 0 \\ \Longleftrightarrow \sigma > 1/z\end{gathered} \tag{29}$$

Recall $\psi' > 0$. The definition of s and inspection of (28) signs the derivative.

Somewhat surprisingly, we need to impose a parameter restriction in order to ensure the supply curve for power is positively sloped. A higher price for power raises the incentive for further collection and this operates to ensure supply is positively sloped. This direct and familiar force is however potentially offset by changes in the productivity of transport created when energy price changes alter the number of intermediate goods. When the price of power rises, intermediate goods producers face higher energy costs and exit the industry. This lowers productivity, raises the cost of transporting energy, and naturally works against a positive supply response.

4.5 General Equilibrium

To solve the model we need to equate the supply and demand for energy which solve for W^* and $(p^W/w)^*$; with these two determined the rest of the system follows directly.[18] This is Proposition 1:

Proposition 1 *A unique interior general equilibrium exists.*

Proof: See Appendix.

The result follows for three reasons. First, our assumption of diminishing returns to each input bounds the strength of increasing returns in the system and in particular this ensures the energy supply curve is upward sloping (i.e. it ensures $\sigma > 1/z$). Second, given Cobb-Douglas preferences there is no choke price for power. Third, the supply of power is positive even at very low relative prices because our energy system is always productive.

One very important feature of power supply, not apparent from the analysis above, is that the supply of power is bounded above by a physical limit; as a result, the power

[18]The market for the composite clears via Walras Law. To check, note our representative consumer's aggregate income is $I = wL + \Pi$.

supply curve becomes increasingly inelastic as prices rise. To understand why recall $W^S \equiv \psi(s) - TC(\psi(s)) = g(\psi(s))$. Now substituting for the particulars of the model we can write total power $W^* = \psi(s)$, and supplied power $W^S = g(\psi(s))$ in explicit form as follows:

$$W^S = \left[1 - \frac{2}{3}s\right] W^* \quad W^* = \frac{\pi s^2 \Delta^3}{[\mu g d]^2} \quad R^* = \left[\frac{s\Delta}{\mu g d}\right] \tag{30}$$

where we note $s \equiv s(p^W/w, L)$ from (28).

In many ways the economy's supply of power is very similar to what we had in the Only Energy Model. Total power, W^*, is again a cubic function of power density Δ; net power supplied to the core, W^S, is a simple fraction of total power; and evaluating supply at $s = 1$ returns our earlier solution. The area of exploitation is again quadratic in power density but now also reflects the relative price of energy. With high energy prices, power is sourced from more distant (read more difficult) sources. Very high power prices also push the share of energy costs towards their maximum of one, and at this point power supply becomes perfectly inelastic. This should not be surprising because when s approaches one, the model becomes more and more like the Only Energy Model.

The reason for the upper bound on power supply is straightforward. When s approaches one, energy is in effect the only economically relevant input. The exploitation zone expands but as s rises it approaches the point where marginal resources take as much energy to collect and transport as they provide - implying their contribution to net power supply is zero. At this point, net supply becomes perfectly inelastic. To see this note that as s approaches unity, R^* approaches $\Delta/\mu g d$ and this implies that one watt of marginal resources delivered to the core has an energy cost of collection and delivery of $R^* \mu g d/\Delta = 1 watt$. No net power is provided. The supply curve is perfectly inelastic.

4.5.1 Energy Prices

There are several "energy prices" in the model. There is the relative factor price of power, p^W/w; the relative price of energy services to the wage, p^C/w; and the share of energy prices in extraction and transport costs $s = p^W/[p^W + p^C]$ which reflects, and is increasing in, the relative price of power to energy services, p^W/p^C. There are also two cases to consider: Almost Neoclassical and Increasing Returns; and at least two important primitives we may want to link to energy prices: market size, L, and energy density, Δ.

Of these many possibilities, two results - that are independent of any case driven assumptions - determine many others. The first is that the share of energy costs in exploitation, s, is always falling in power density. It should come as no surprise that an increase in power density always lowers power prices however written. Higher power density lowers the relative factor price of power; the price of energy services to the wage; and lowers the energy cost share as well. These results are straightforward and rely on no new assumptions. They reflect the pure supply shock nature of a change in Δ, since an increase in density shifts power supply outwards while driving down relative factor prices in a diagram like figure 1.

The second key result is that the share of energy costs in exploitation is always rising in market size. An increase in market size creates a demand shock that raises the demand for power at the same time as increasing the supply of final output. Consequently, their relative price adjusts upward and s rises. This result is also independent of other case driven assumptions, and it carries with it several important implications. For example, while it says larger markets always create larger exploitation zones and increasing power supply, we have already seen there is a limit to this process. Specifically, dW/dL must eventually approach zero in (22) so that final output responds less than one for one with market size. Therefore, regardless of the potential for increasing returns in the economy - as shown by our hypothetical exercise using (21) - and regardless of whether the economy's output responds more than one for one with market size over some range - as seems plausible given (22) - the economy must eventually suffer from diminishing returns to increases in market size. It also implies there is a maximum size to exploitation zones given when $s = 1$.

This result follows for two main reasons. The first is that we have bounded increasing returns by requiring each input suffer diminishing returns. In doing so, we rule out the possibility that increases in labor alone can raise output per person forever. The second reason is that power becomes more expensive with larger markets, and this drives up the costs of exploitation. Higher energy prices drive potential revenues up, but when the share of energy cost in exploitation approach one further increases in prices drive costs up by the same amount.

While the economy must eventually suffer from the limited availability of energy, over some range, energy prices (measured in a variety of ways), may fall with market size. This possibility comes about from efficiency gains in the energy system. Larger markets support the entry of additional firms producing intermediates that raises the overall efficiency of the energy system. How large these gains are determines the aggregate degree of increasing returns in the model. Not surprisingly then our results here are case dependent. Two definitions ease the notation: the Increasing Returns case satisfies $\sigma < 3/[1-z]$; the Almost Neoclassical case satisfies $\sigma \geq 3/[1-z]$. With these definitions in hand, we can report

Proposition 2 *In the Almost Neoclassical case, an increase in market size raises the relative price of energy services to the wage, and raises the relative price of power.*

The parameter restriction in this proposition is exactly that which ensures the bracketed term in 22 is always less than one. More descriptively given the measure of potential increasing returns is given by $\sigma/[\sigma-1]$, it says when increasing returns are weak the increasingly costly energy exploitation always offsets any efficiency gains a larger market may provide for the energy system. As a result, the economy's results mimic those of a typical neoclassical set up with one fixed factor but constant returns overall.

Results are quite different when increasing returns are stronger.

Proposition 3 *In the Increasing Returns case an increase in market size at first lowers the relative price of energy services to the wage but eventually leads to an increase.*

The parameter restriction defining the Increasing returns case is exactly that which ensures the bracketed term in (22) is greater than one for small values of s, but less than one for high values. Under slightly stronger conditions ($\sigma < 3$) it is also true that the relative factor price of energy also falls and then rises with market size. Therefore, both the raw price of power and the price of energy services will fall with market size. Since the cost of energy services depends on both raw power prices and the number of intermediates in the energy system, it is always true that the relative price of raw power turns upward first only later to be joined by an increasing price for energy services.

4.6 Consumption and Output per person

Given the last two propositions it should not be surprising that the relationship between market size and output and consumption per person are determined by similar forces. Proving this requires some further algebra that we relegate to an Appendix.

Proposition 4 *In the Almost Neoclassical case, output per capita and consumption per capita fall with market size; in the Increasing Returns case, output per capita and consumption per capita are single peaked in market size.*

Proof: see Appendix.

When increasing returns are weak, the model operates much like a competitive neoclassical formulation with one fixed factor and one variable factor. Output per person and consumption per person fall with population size, and this represents the Almost Neoclassical features of our model.

When increasing returns are stronger, then things are vastly more interesting. Both consumption and output per person reach an interior maximum at a given market size and then fall thereafter. The single inflection point reflects the weighing of the model's increasing returns properties against the rising marginal costs of additional energy exploitation. Since both processes are monotonic - increasing returns always weaken and marginal costs are always rising - the peak is unique.

Two other results are relevant.

Proposition 5 *Output per person and Consumption per person rise with power density.*

The entire schedules in both cases shift upwards when the energy source has greater power density. In addition,

Proposition 6 *In the Increasing Returns case, peak levels of output and consumption per capita occur at larger market sizes when power density is greater.*

Proof See Appendix.

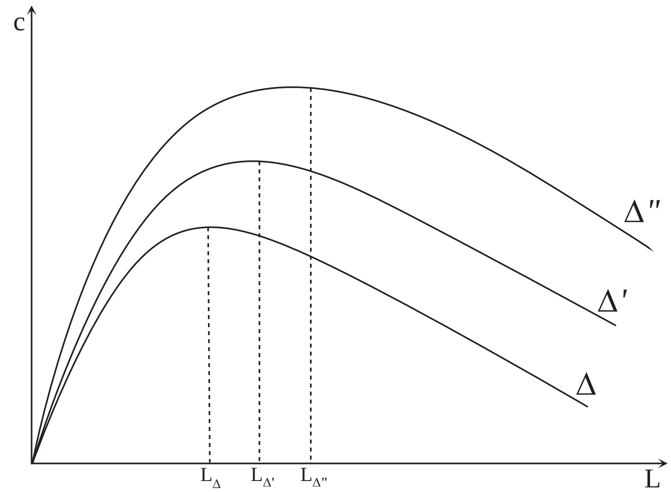
More dense energy sources provide greater peak utility and support larger populations at these peaks. These features are shown in the two panels of Figure 3 where we have assumed power densities $\Delta'' > \Delta' > \Delta$. The top panel shows the Increasing Returns case; the bottom panel the Almost Neoclassical case.

The Energy-Economy model adds several new results and amends some from the Only Energy model. Most importantly, it demonstrates how the interaction of rising marginal costs of obtaining power with an ever more efficient energy system determines the pattern of energy prices, consumption and output per capita. When increasing returns are weak, the model generates results that we have called Almost Neoclassical. Increases in market size drive up energy prices, lower consumption and output per person, and offer no incentives for economic agglomeration. Indeed incomes and consumption are maximized when any population center is of vanishingly small size.

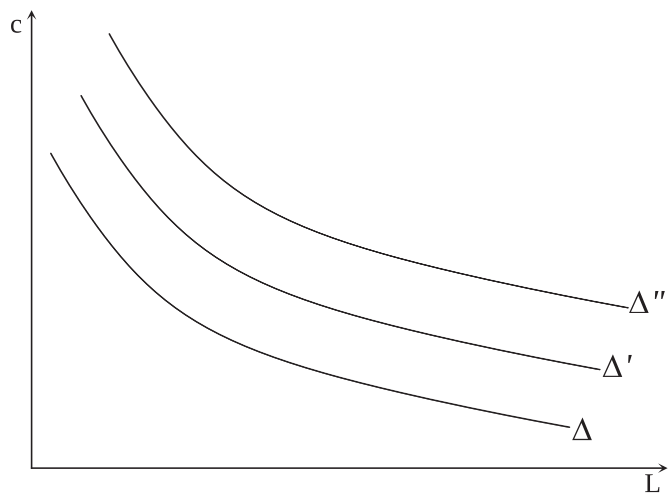
When increasing returns are strong, as in the Increasing Returns case, energy prices initially fall with increases in market size despite the obvious pressure this puts on supplies. Energy prices decline because the larger market for power allows for the introduction of new methods and means for harnessing energy to productive ends. These new means create aggregate productivity gains that fuel positive feedbacks in the energy system. Moreover the strength of these impacts - declining energy prices, higher incomes per capita, and strong incentives to agglomerate - are all stronger when the energy source is more dense.

4.6.1 Roads, rivers and canals.

Rivers, roads and canals were all important features of the energy transportation system in 1800; just as power lines, oil pipelines, and LNG terminals are important features today. What these features have in common is that they represent low friction and presumably low



(a) Increasing Returns Case



(b) Almost Neoclassical Case

Figure 3: Energy Density and Consumption Per Capita

cost methods of transporting energy to markets. In this section we examine the impact of having a low friction alternative by examining the impact of having a river or a road cross a city. Analytically, we treat roads as rivers that lower transport costs in both directions.

The analysis proceeds in two steps. First, given the lower transport costs along a river or road energy suppliers may now decide to take a longer route to the core if it offers lower costs. This decision problem affects the shape of our exploitation zone. Second, since a river or road lowers transport costs the overall size of the exploitation zone will also change.

To proceed consider the decision problem of a potential energy supplier located on one meter squared of land with energy produced with uniform power density equal to Δ Watts per meter squared. The supplier can take energy directly into the city or deviate to take advantage of a road or river nearby. Rivers and roads help to reduce the amount of work used in transportation, increasing the amount of energy delivered to the city. To capture this in our analysis we allow for the coefficient of friction of the river or road to differ from the coefficient of friction of land by a fraction ρ. That is, while the coefficient of friction of land is equal to μ, a road's coefficient of friction is $\rho\mu$ in both directions whereas when traveling with the current a river's is also $\rho\mu$ but against it μ/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.

We assume the river or road is a straight line that crosses the core of the city and expands indefinitely. The location of a supplier relative to the city is described by two terms: ι, the distance from the city and θ the angle between the segment formed by the city and the supplier and the river.

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. Note that since every unit of energy used in transportation is matched with a unit of final output, minimizing energy use also minimizes total transport costs. In the second stage profits are maximized. The cost minimization problem is given by:

$$\min_{\iota_1,\iota_2} \frac{\mu g d}{\Delta}\Delta\iota_1 + \frac{\rho\mu g d}{\Delta}\Delta\iota_2 \tag{31}$$

subject to

$$\iota_1^2 = \iota_2^2 + \iota^2 - 2\iota\iota_2\cos\theta \tag{32}$$

where ι_1 is the distance travelled by land and ι_2 is the distance travelled by river or road.

The constraint follows directly from the law of triangles with ι_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 river:

$$\iota_1^* = \frac{\iota\sin\theta}{(1-\rho^2)^{1/2}} \tag{33}$$

and

$$\iota_2^* = \iota\cos\theta - \frac{\rho\iota\sin\theta}{(1-\rho^2)^{1/2}}. \tag{34}$$

If the distance ι_2^* is strictly positive, the supplier deviates to the river, otherwise the supplier goes straight to the city. We can solve for the critical value of θ that separates the suppliers that go straight to the city from those who deviate to the river:

$$\iota_2^* > 0 \text{ if and only if } \theta \leq \cos^{-1}\rho \equiv \bar{\theta} \tag{35}$$

The second part of the energy producer's problem is to decide whether or not to take its energy to the city. An energy producer situated a distance ι away from the city and forming an angle θ with the river will go to the city if its profits are greater than zero. Profits are given by

$$\Pi = p^W \Delta - (p^C + p^W)\frac{\mu g d}{\Delta}(\iota_1^* + \rho\iota_2^*)\Delta \tag{36}$$

Replacing equations (33) and (34) in the previous equation makes the profits a function of the distance to the city ι. There is a critical distance ι^* at which suppliers become indifferent between bringing their energy production to the city or leaving it on the ground. This distance is determined by setting $\Pi = 0$. The critical distance is given by:

$$\iota^* = \frac{\Delta}{\mu g d}\frac{p^W}{p^W + p^C}\begin{cases} \left((1-\rho^2)^{1/2}\sin\theta + \rho\cos\theta\right)^{-1} & \text{if } \theta \leq \bar{\theta} \\ 1 & \text{if } \theta \geq \bar{\theta} \end{cases} \tag{37}$$

which determines the area of exploitation.

Total energy supplied to the city can then be written as:

$$W^S = 2 \times \left[\int_0^{\bar{\theta}}\int_0^{\iota^*} v\left(\Delta - \frac{\mu g d}{\Delta}\left(\iota_1^* + \rho\iota_2^*\right)\Delta\right)dvd\theta + \int_{\bar{\theta}}^{\pi}\int_0^{\iota^*} v\left(\Delta - \frac{\mu g d}{\Delta}v\Delta\right)dvd\theta\right]$$

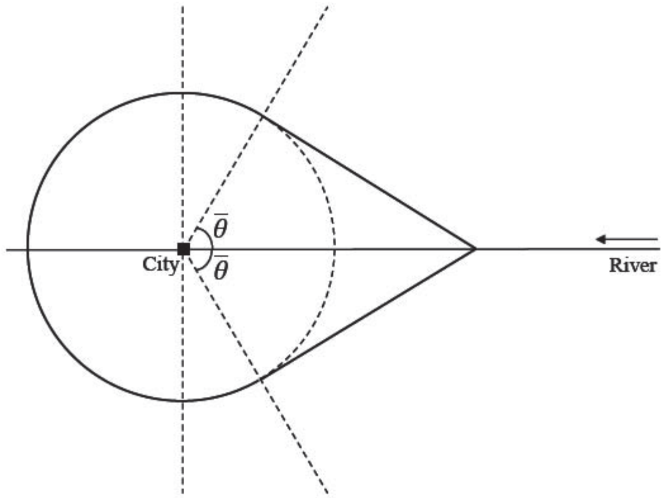
The integral is multiplied by 2 since we are only adding up over the half circle of π radians. The first integral represents the supply coming from suppliers who are close enough to the river to use it in transport. The second integral represents the supply coming from those who travel directly to the city core and should be familiar from our discussion of (10).

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

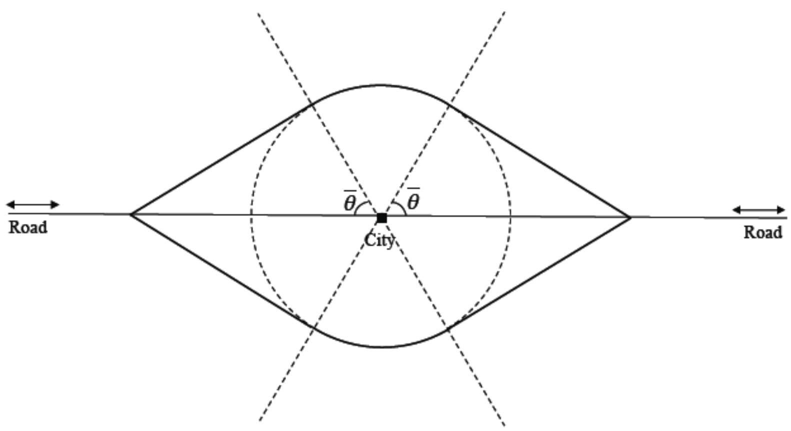
$$W^S = \frac{\Delta^3}{(\mu g d)^2}s^2\left(1 - \frac{2}{3}s\right)(\pi + g(\rho)) \tag{38}$$

$$g(\rho) = \int_0^{\bar{\theta}} \left((1-\rho^2)^{1/2}\sin\theta + \rho\cos\theta\right)^{-2} - \bar{\theta}d\theta \geqslant 0 \tag{39}$$

where the function $g(\rho)$ is positive and monotonic and approaches infinity as ρ goes to zero. Setting $\rho = 1$ means the river or road offers no advantage in terms of transportation. This implies $g(\rho) = 0$ since then $\bar{\theta} = 0$ and equation (38) reduces to equation (30). We depict the exploitation zone in the river and road case in the two panels of Figure 4 assuming $\rho < 1$.



(a) Area of Extraction with a River



(b) Area of Exploitation with a Road

Figure 4: Exploitation and Transport Costs

It is now possible to redefine terms slightly to rewrite supply as:

$$W^S = W^* \left(1 - \frac{2}{3}s\right)$$

$$W^* = \frac{\pi s^2 \tilde{\Delta}^3}{(\mu g d)^2} \quad where \quad \tilde{\Delta} \equiv \Delta m(\rho)$$

$$m(\rho) = ((\pi + g(\rho))/\pi)^{1/3} > 1$$

which is exactly the same form as our earlier supply curve! The role of improved river transport is identical to being granted a more dense resource base in terms of energy supplied to the core. River location multiplies by $m(\rho)$ the power density of available resources to an extent determined by its capacity for reducing transport costs as reflected in ρ. An important feature of this result is that the impact of the improved transport, in terms of energy supply, is that it is increasing in the density of available resources; therefore, improved transport is most powerful for very dense resources, and least powerful for low density resources.

These calculations can be easily expanded to accommodate the case of a road for which the direction of flow is not important. In this case energy suppliers can take advantage of the reduced coefficient of friction in either direction away from the city. Using similar steps, this leads to the supply function being given by:

$$W^S = W^* \left(1 - \frac{2}{3}s\right), where\ W^* = \frac{\pi s^2 \hat{\Delta}^3}{(\mu g d)^2}$$
$$\hat{\Delta} \equiv \Delta \left((\pi + 2g(\rho))/\pi\right)^{1/3}$$

which is of course very similar to the flow of the river case. Summarizing:

Proposition 7 *Consider an improvement in river or road transportation that changes $\rho = 1$ to some $\rho < 1$, then this improvement: (1) has the same effect on a city's energy supply as does an increase in the power density of its surrounding resources; and (2) increases energy supply most in situations where resources have high power density.*

This is a very important result because it extends all our earlier results on power density to results concerning the existence of rivers or roads for transport. For example, we know that a unique general equilibrium still exists; the existence of a river or road (modeled as a movement from $\rho = 1$ to some $\rho < 1$) shifts the supply curve for energy outwards lowering the relative price of energy while raising the equilibrium energy supplied, final output and the number of intermediate goods in the energy system; access to a river or improved road transportation raises both consumption and output per capita at all population levels while it raises the density of individuals at which output and consumption per capita peak.

4.6.2 Summary

The way is now clear for us to introduce a sorting mechanism that allocates population across space to investigate the role energy density may play in determining the location and geography of economic activity. For example, we could follow Henderson (1975, 1973) and introduce a process of city formation by developers; alternatively, we could follow more recent work by Fujita et al. (1996) by introducing competing regions and trade to examine the motives for mobile labor to agglomerate in one region.

While many mechanisms for sorting the population across locations could be fruitfully added to our Energy-Economy model, this paper's purpose is to ask whether there is a causal relationship between the density in available energy supply and the resulting density in economic activity. Any empirical study of this relationship seems difficult, if not outright heroic, in any contemporary economy with multiple energy sources (fossil fuels, renewables, and nuclear), with energy sources not tied to geographic locations (nuclear), and with agglomeration patterns that have both driven, and been driven by, the availability of dense energy sources (does anyone think the suburbs would exist without the automobile, fossil fuels, and nearby refineries?).

Sorting out all of these complications to identify the role of energy density would be difficult if not impossible. To eliminate some of these complexities, we have chosen to look back to a simpler period in our history when all energy was in fact green energy, and all economies were, in fact, organic economies. This simpler setting offers to us the possibility of identifying the role of energy density in shaping the geography of economic activity.

5 The Malthusian Application

To allow market size to respond endogenously to the conditions set by a landscape's energy resources we start by examining an Isolated State whose market size is determined solely by population growth. We adopt a Malthusian mechanism for population growth since it is simple, well understood, and arguably appropriate given our empirical application covers years well before any demographic transition; indeed our data comes from a time (1086-1801) and a place (England) where Malthus himself observed the forces he identified at work.[19] But villages, hamlets, and cities do not exist in isolation. They are geographically proximate competitors for population. Therefore we also extend the results from our analysis of the Isolated State to allow for migration both within and across regions.

[19] There is an ongoing debate over the usefulness of the Malthusian model as a means to understand population growth in England over this time period. While there are several issues at dispute, one basic problem is that almost any population history can be rationalized by the Malthusian model if we allow for shifts in the preventive check over time. Since these shifts are usually motivated by forces outside the model, they are clearly difficult to evaluate empirically. Consequently, while Clark (2007) presents an unabashed Malthusian view of the world prior to 1800, Allen (2008) disputes much of this evidence and argues the application is simplistic and often problematic. A more favorable recent evaluation, using cross country data to the year 1500, is presented in Ashraf and Galor (2011).

5.1 The Isolated State

We adopt a very standard Malthusian specification where population growth responds to increases in consumption per capita, but amend it to allow for a higher death rate arising from crowding in cities. There is a baseline birth rate, η_0, and baseline death rate, δ_0, with $\eta_0 - \delta_0 < 0$ representing the net rate of population reduction when consumption per capita is zero. This baseline population growth is then adjusted by assuming births rise proportionately with consumption per capita, $\eta_1 c$ and deaths fall proportionately with consumption per capita $\delta_1 c$, where $c = C/L$. During the Malthusian era, death rates in cities were known to be much higher given their poor sanitation and crowding. We incorporate this into our population dynamics by assuming death rates due to crowding rise proportionately with population density, $\delta_2 L$. As a result, the birth and death rate functions are given by:

$$\begin{aligned} \eta(c) &= \eta_0 + \eta_1 c \\ \delta(c, L) &= \delta_0 - \delta_1 c + \delta_2 L \end{aligned} \tag{40}$$

Hence the population responds according to:[20]

$$\overset{\bullet}{L} = L[[\eta_0 - \delta_0] + [\eta_1 + \delta_1]c - \delta_2 L] \tag{41}$$

In the standard Malthusian set up $\delta_2 = 0$, and subsistence consumption is a constant independent of city size. Once we allow city size to affect death rates, consumption must rise with city size to hold population growth at zero. Define the Malthusian line as the combination of consumption per person and population consistent with zero population growth. We define the Malthusian $M - line$ as:

$$c_M(L) = \frac{\delta_0 - \eta_0}{\delta_1 + \eta_1} + \frac{\delta_2}{\delta_1 + \eta_1} L \tag{42}$$

where $c_M(L)$ is the per capita consumption needed to keep the population constant in a city with population size L.

Since we have already characterized how consumption per person changes with market size and power density it is now apparent that:

Proposition 8 *In the Increasing Returns case, there exists a critical level of power density, $\Delta^{crit} > 0$; if power density is below this critical value, $\Delta < \Delta^{crit}$ then the only steady state*

[20] An implicit assumption in this analysis is that everyone lives in some form of agglomeration, be it beside a manor, village, town, or city. Adding a fixed purely rural population is easy to accomplish but adds little to our results.

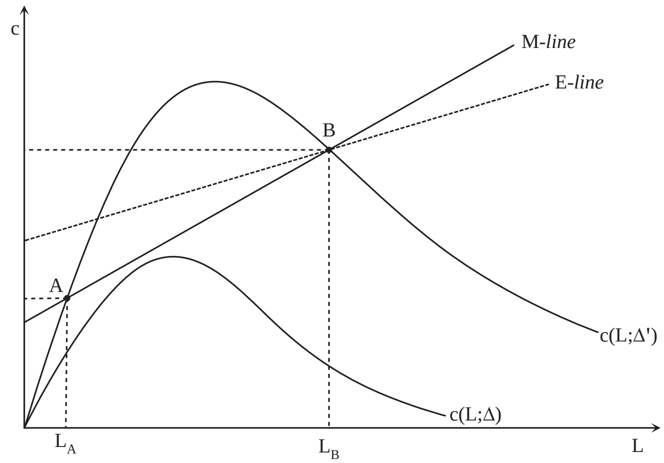


Figure 5: The Isolated State

has $L = 0$ and the landscape is vacant; (iii) if $\Delta > \Delta^{crit}$ then there are three steady states: steady state zero has zero population $L_0 = 0$ and a vacant landscape; steady state one has a positive population $L_1 > 0$ but relatively low population density; steady state three has a positive and larger population $L_2 > L_1 > 0$.

This result establishes a link between the density of available energy resources, and the density of economic activity in the region. It is shown graphically in Figure 5. Let $c(L; \Delta_i)$ define the curve giving consumption per capita in city i as a function of labor L and its power density Δ_i. If available resources are not dense, then no city can exist and a vacant landscape must result. This is the case shown in the figure by the $c\,(L; \Delta)$ curve which lies everywhere below the $M - line$. When resources do offer sufficient power density, $\Delta' > \Delta$, then two possible city sizes emerge: small cities (point A) and large cities (point B).

It is obvious that small cities offer residents lower consumption per capita, but less obvious that they offer lower death rates and longer lifetimes as well. To see why note that when all individuals are identical, we can interpret $\delta(c, L)$ as the instantaneous death rate for any citizen in a city giving the pair $\{c, L\}$. Since birth rates equal death rates along the $M - line$ and birth rates rise with c, it is now clear that death rates at B exceed those at A. We can in fact go further to calculate expected lifetime consumption at any steady state. To do so, assume a zero discount rate for simplicity, and then the expected lifetime consumption for any individual living in city i is given by:

$$e_i = \int_{\tau}^{\infty} c_i \exp^{-\{t-\tau\}\delta(y,L)}\, dt \tag{43}$$

$$e_i = c_i/\delta_i \tag{44}$$

where the last equality holds in steady state. We can compare the expected lifetime consumption any individual could expect in any city by simply looking at e_i across cities.

It proves useful to construct a line representing all combinations of $\{c, L\}$ yielding a given e. To find the combinations of consumption per capita and population giving an expected lifetime consumption per person equal to e, use the death rate $\delta(y, L)$ from (40) in (44) and rearrange to find the following relationship which we refer to as the $E - line$:

$$c_E(e, L) = \left[\frac{e}{1 + \delta_1 e}\right] [\delta_0 + \delta_2 L] \tag{45}$$

One such $E - line$ is shown in Figure 5. Several properties of $E - lines$ follow directly. Holding expected lifetime consumption, e, constant, the $E - line$ is just a linear relationship between population size and consumption per person. Points above the $E - line$ associated with any given e, correspond to combinations of consumption and population sizes yielding higher expected lifetime consumption than e; points below correspond to lower expected lifetime consumption than e. Any given $E - line$ is drawn conditional on e; raising e shifts the $E - line$ upwards. Higher expected lifetime consumption implies both a higher vertical intercept, and a steeper slope, for the $E - line$. The $E - line$, evaluated at $e = e^*$ must pass through the steady state yielding e^*. Finally, although it is not obvious, the $E - line$ going through any steady state is flatter than the associated $M - line$ going through the same steady state.[21]

Using this construction reconsider the two possible interior steady states shown in Figure 5. Since the small city steady state at A falls below the $E - line$ going through the large city steady state at B, it must provide lower expected lifetime consumption per capita than the large city steady state. In this sense the large city steady state welfare dominates the small city steady state. In addition, under the Malthusian adjustment mechanism, the steady states at zero and at B are stable. Whereas the steady state at A is not. This implies that

[21] Some algebra will show this, or some logic chopping can explain it. Movements along the E line must reflect equal proportionate changes in consumption and death rates; changes in consumption cause less than proportionate changes in the birth rate since the birth function is affine; this implies movements along an E line away from a steady state generates a situation where deaths exceed births and hence this movement must be below the corresponding M line. The M Line is flatter than the E line through any steady state.

small cities when they exist are fragile constructs. A small negative perturbation leads to their disappearance; a small positive perturbation starts the process of transformation into large cities. Small cities are therefore both temporary and not very productive phenomena; large cities are far more productive and far more permanent features of the landscape.

These results should be contrasted with the case when increasing returns are weak, since this case produces almost standard Neoclassical results.

Proposition 9 *In the Almost Neoclassical Case: (i) there exists one interior steady state with $L > 0$ and this interior steady state is stable under the Malthusian adjustment mechanism; (ii) population density at the interior steady state is increasing in power density; and (iii) there exists a steady state at $L = 0$ but this steady state is not stable.*

City size is uniquely determined by power density in the Almost Neoclassical case; whereas cities could be large or small in the Increasing Returns case. In the Almost Neoclassical case any agglomeration or village smaller than the (unique) steady state will grow to become a city; in the Increasing Returns case, relatively small agglomerations die out creating city death, or they may grow large and stabilize. In the Almost Neoclassical case every landscape can support a city; whereas in the Increasing Returns case there is a minimum power density required. In short, the Increasing Returns case paints a far richer view of city formation and growth where multiplicity of outcomes, instability, and thresholds determine, in conjunction with the natural environment, the geographic distribution of population and economic activity.

5.2 Within and Across County Migration

Thus far we have assumed our cities are truly Isolated States: their own population growth determines market size. Since there is no inter-city migration, and no simultaneous small or large city pair feeding and sustaining each other, we have only a very partial understanding of how the distribution of power densities across space may influence the distribution of cities and their size. The ability of the population to migrate changed greatly over our 800 year sample period, and therefore it is important to understand how migration interacts with the distribution of power densities to determine city sizes. For example, prior to the Black Death serfdom limited the mobility of the population but the resulting wage pressures after the Black Death enhanced the power of serfs. These forces led to the Peasant revolt in 1381, the elimination of serfdom by c.1500, and a corresponding increase in mobility.[22]

[22]See Wrigley (2010, Chapter 5) for an extensive discussion of population growth and migration from the mid sixteenth to the mid nineteenth century.

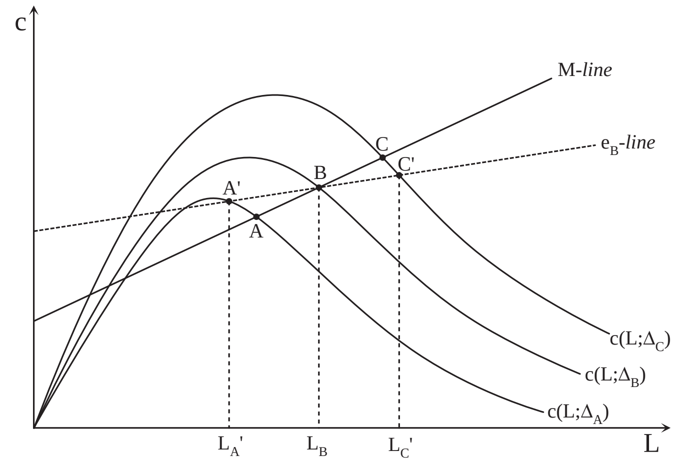


Figure 6: Migration in the three county example

We introduce migration by constructing a simple three point distribution linking power density to city sizes and migration. This example highlights many of the features of the mapping from power density to city size, and has the benefit of being transparent.

5.2.1 A 3 County Example

To understand how flows of migration affects city size, consider Figure 6. We have drawn three Isolated States which differ in the power density of their neighboring resources. By assumption, C has the highest power density and A the lowest; in addition we have assumed we are in the Increasing Returns case.[23] Also shown in Figure 6 is the Malthusian or $M - line$ giving those combinations of consumption per person and population generating zero population growth. The intersection of this line with our three consumption per capita curves generates the associated stable steady states labelled A, B and C.

Now consider the migration decision. Start with city B, which we will refer to as the middle city. At B, there is an associated consumption per person and population size; and using (43) we can calculate the expected lifetime consumption per capita any resident of B would obtain. Denote this e_B, and using now familiar methods we find the following relationship

$$c_E(e_B, L) = \left[\frac{e_B}{1 + \delta_1 e_B}\right] [\delta_0 + \delta_2 L] \tag{46}$$

which we will refer to as the e_B-line since it is the $E - line$ evaluated at steady state B which gives e_B. By construction the e_B-line must pass through point B, and it is flatter than the

[23]The interested reader is invited to follow through our analysis with the Almost Neoclassical case.

$M - line$ as shown.

It is immediate then that points labelled A' and C' in Figure 6 offer the same expected lifetime consumption per capita as at B. Point A' is however above the $M - line$: its consumption level exceeds that necessary to generate zero population growth and, in the absence of migration, its population level would necessarily rise from $L_{A'}$. Indeed - without migration - its population would increase until the city reached the no-migration steady state at A. An alternate possibility is that this constant excess flow of population could migrate instead to city C thereby holding its current population at $L_{A'}$ and its current expected lifetime consumption equal to that offered at B. Now consider city C. Residents in this city would have the same expected lifetime consumption as those at B, if it was to grow and reach point C'. But of course point C' is below the $M - line$, and hence consumption per capita at this level falls short of that necessary to generate zero population growth. Without further adjustments, its population would decline from the level $L_{C'}$ towards its no-migration steady state at C. The now obvious alternative is for city C to remain at C' and receive a constant flow of population from city A. If the flows from city A to C balance we have identified a steady state of the system with migration. For future reference we will label steady state expected lifetime consumption in a migration steady state when there are three cities (labelled A, B and C) as e^*_{ABC}.

In theory, Figure 6 shows how a given distribution of power densities across previously Isolated States maps into a steady state distribution of city sizes with active migration. We have assumed here that cities differ in the power density of their available resources as they would if they came from different counties. This interpretation is however overly restrictive. Cities within the same county may differ in power densities if some have access to low friction alternatives like rivers or roads while other cities in their county do not. Therefore, our three cities could be in one county, two counties, or three counties.

The migration steady state has several important features that generalize to the many region case. First, migrants flow from small cities to big cities. Big cities offer higher consumption per capita but higher death rates. Migration does not equalize consumption or real wages across the population centers, but it does equalize expected lifetime consumption. Second, migration widens the distribution of city sizes. The biggest city got bigger and the smallest city got smaller. Third, for any city or region sending migrants we find a rise in productivity as consumption (and output) per person rises. In addition overall productivity may also rise - migrants move to a region with higher average productivity and lessen the Malthusian pressures keeping productivity low in their sending regions. Fourth, migration is associated with an increase in the birth rate in sending regions but depresses it in the large city receiving it.

5.2.2 Existence

Our analysis thus far assumes the existence of a middle city B with zero migration and an expected lifetime consumption of $e^*_B = e^*_{ABC}$. Extending the analysis to solve for the steady state consumption per person is not difficult since e plays the role of a price that clears the market for migrants. To see this imagine removing city B entirely from our analysis, and then draw in the associated $E-lines$ for city A and city C evaluated at their zero migration steady states in Figure 6. In this case, since we are considering the no migration steady states, denote these the $e^*_A - line$ and the $e^*_C - line$ respectively.

Consider first the $e^*_C - line$ through point C. At this level of expected lifetime consumption, deaths in city C match births; but if this level of consumption was maintained in A, then its clear that births would far exceed deaths. A would "like" to supply migrants to C but there is no demand for them. Therefore, the expected lifetime consumption represented by the $E - line$ through C is too high to represent a steady state with migration. Now consider the $e^*_A - line$ through A. At this level of expected lifetime consumption, deaths in city A match births but if this level of expected consumption was to be maintained in C then it is clear that deaths would exceed births in C. C has a demand for migrants but there is no supply coming from A. Therefore, the expected lifetime consumption represented by the $e^*_A - line$ through A is too low to represent a steady state with migration. Recall that as we increase expected lifetime consumption the $E - line$ shifts up and becomes steeper. This raises the supply from A and decreases the demand from C in a continuous manner. Since we had excess demand evaluated at e^*_A and excess supply at e^*_C, there will exist a $e^*_C > e^*_{AC} > e^*_A$ where they balance, and this defines our steady state.

To generalize this argument some new notation is helpful. Recall $c(L; \Delta_i)$ defines the curve giving consumption per capita in city i as a function of labor L and its power density Δ_i. Now consider the intersection of this curve with a given $E - line$. Assuming an intersection exists, let $L(e, \Delta_i)$ denote the (largest) strictly positive value of the population at the intersection between an $E - line$ evaluated at e and the consumption per capita curve associated with Δ_i. In our analysis above we implicitly assumed the intersection at $L(e^*_C, \Delta_A) > 0$ existed; that is, there was a strictly positive population level in city A which could support the expected lifetime consumption in the no migration steady state of city C. It should also be apparent that every consumption per capita curve has a maximum expected lifetime consumption it can support which is given by the tangency of an $E - line$ and the consumption curve $c(L; \Delta_i)$. Denote the maximum expected lifetime consumption city i with density Δ_i could sustain as $e^{\max}_{\Delta_i}$. By definition then the population level $L(e^{\max}_{\Delta_i}, \Delta_i)$ is the smallest positive population city i could sustain in any migration steady state. Then we know that if $e^*_C < e^{\max}_{\Delta_A}$, then $L(e^*_C, \Delta_A) > 0$ and city A can survive in a migration steady state. Finally, we note that $e_{\Delta_i}^{\max}$ is strictly increasing in Δ_i; and putting this all together we obtain:

Proposition 10 *Consider a finite number of Isolated States with surrounding resources having power density Δ_i with the labels $i = 1, 2....N$ chosen so that power density is strictly increasing with N. Then if $e_N^* < e_{\Delta_1}^{\max}$ a steady state with active migration across all cities exists.*

Proof: see Appendix.

5.2.3 City Death

The construction above rules out the possibility of any city being destroyed by migration. It suggests however how such a result can occur. Consider a three city case using the notation and assumptions from the previous proposition. Suppose a migration steady state occurs because $e_3^* < e_{\Delta_1}^{\max}$. Since we know that expected lifetime consumption is strictly rising in a city's power density we can increase the power density of city 3 enough to violate this condition. After we do so, it is now possible for the steady state with city 1 to disappear. This will necessarily be true when we find $e_{23}^* > e_{\Delta_1}^{\max}$; that is, when a migration steady state between city 2 and city 3 produces a steady state expected lifetime consumption greater than the maximum supportable in city 1. When this occurs, city 1 can not compete and it loses its population.

In a world with many cities and varying availability of energy resources, there will always exist a middle consumption city like our previous construction of B. All regions with power densities below this level will export people; while all those above it will import people. Of course there is no need for the number of cities that receive or send people to match, in fact one great city could receive almost all of the population while cities in every other region send people to it. With this in mind, we now turn to our empirical examination of England in the middle ages.

6 A Population History of England

The population history of England from 1000 AD to 1800 AD has been studied by legions of academics with demographers, historians and economists producing literally hundreds of contributions dissecting and debating this period at great length.[24] Our goal here is not to

[24]Several excellent book length treatments by Sir Tony Wrigley present a fascinating picture of English demography and economics over this time period. Especially relevant are Wrigley (2010) that focusses on

dispute their findings but present for discussion a view of this history through the lens of our theory. To do so we need to be specific about how we relate data to our theory; and specific about what our theory says about the data.

6.1 Data

We have collected data from several different sources to present a largely descriptive account of the population history of this period. Population data is available from recent estimates by Broadberry et al. (2011) for all 39 ancient English counties for five time periods: 1086, 1290, 1377, 1600 and 1800. We employ this data plus measures of county areas to generate a series for population density across counties and time.

6.2 Units, Resources and Time periods

To link the theory to data several assumptions need to be made. To start we need to associate a geographic unit with a region where resources are homogenous in terms of their power density. Our data is at the county level and hence, while it is not ideal, we take counties as our unit of analysis. There are 39 ancient counties (which excludes Wales and Northern Ireland) in our data and together they constitute England. By assumption, all locations in a county are surrounded by resources of the same power density, and therefore all agglomerations of economic activity (villages, boroughs or cities) within a county have access to resources with the same power density. However, by virtue of Proposition 7 even locations surrounded by resources of the same power density will differ in terms of the resources they can access. As a result, a city or village that can exploit a low friction alternative such as a river or road can be thought of as one surrounded by more power dense resources. The importance of this result is that it introduces a source of within county heterogeneity so that any one county could support agglomerations of different sizes and there can be within county migration as described previously. Naturally, we expect counties to differ in the power density of resources and this may fuel across county migration as well.

For the most part we assume the availability of resources is fixed over time. For example, the power density of renewable resources is fixed over time since it reflects natural blessings coming from solar insolation, rainfall, and soil conditions together with existing rivers. Improvements in our ability to capture these flows surely occurred as it did in many other facets of economic life over this period. Therefore our working assumption is that the

energy issues; Wrigley (1969) and Wrigley (1978) are also valuable resources. Allen (2009) presents an engaging and rigorous examination of the forces leading to the Industrial Revolution focussing on energy sources, induced innovation and the role of international trade.

distribution of renewable resources is fixed, but we allow their productivity to grow over time at a common economy wide rate. For example, changes in the availability or productivity of crops could have raised power densities; just as a changing climate during the Little Ice Age could have lowered them. We sweep all such changes into a residual called technological progress by implicitly assuming their impact is common across counties. A clearly heroic but necessary assumption.

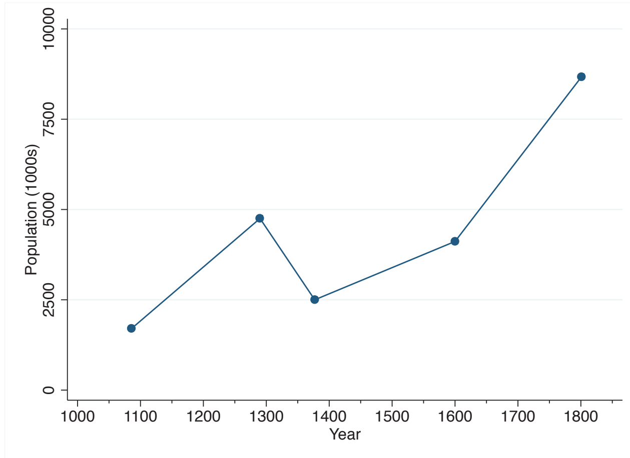
In similar fashion we treat coal as an energy source whose geographic location is exogenous, but whose exploitation is dependent upon economic conditions determined within the model and exogenous factors such as technological progress outside of the model. For example, we take the existence of coal deposits at Newcastle and the slowly growing penetration of coal in the energy supply as exogenous, but explain the coastal transport of Newcastle coal to London coal within the model. Over most of this period coal was used as a source of domestic and industrial heating. Only late in the period was coal used as a source of power for pumping, stamping, etc. in an expanding list of industries. Given the size of coal reserves in place we treat coal as inexhaustible during this period, and ignore the impact of depletion on costs.[25] There are three key differences between coal and renewables: coal has a much greater power density; coal is a punctiform resource that appears in only select locations; and, finally, an ability to use coal as an energy source changed drastically over the period. Coal was about 10% of England's energy consumption in 1600; 50% in 1700; and 75% in 1800.[26]

6.3 A Sketch of 700 Years

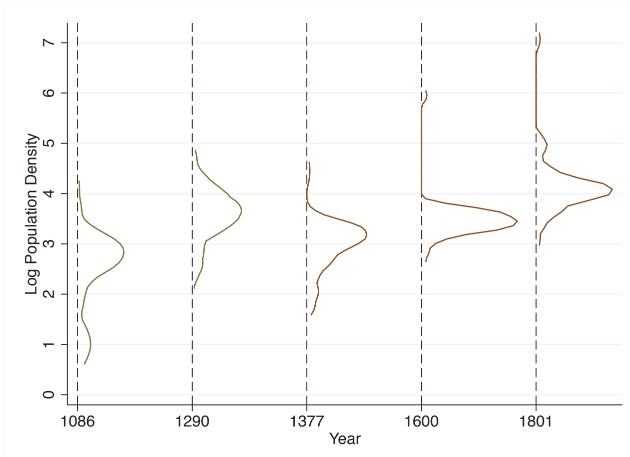
The population history over our entire sample period is shown in Panel A of Figure 7 (ignore for the moment panel B). Panel A presents five data points representing population estimates for the years 1086, 1290, 1377, 1600, and 1801. As shown England starts with a 1086 population of approximately 1.7 million and ends with a 1801 population of a little under 9 million; this growth however represents an anything-but-constant annual population growth rate of 0.23 percent per annum. In fact this period is dominated by two cycles of growth and subsequent plateau, although this is not readily apparent from the figure. The earliest period from 1086 to 1290, although affected by the Norman conquest and the resulting political upheavals it created, was a relatively prosperous time for the population. Population growth during the thirteenth century was robust and the population of close to 4.5 million citizens in

[25] There is little in our analysis which needs amendment if energy comes from a non-renewable like coal. Depletion raises the marginal cost of obtaining the resources, and this rises over time at a rate related to the size of reserves, their power density, and demand. Given the size of coal reserves, assuming away rising depletion costs seems innocuous. Empirical evidence on this point is provided in Chapter 4 of Allen (2009).

[26] See Wrigley (2010, p. 37, Table 2.1).



(a) Population



(b) Population Density

Figure 7: Population History of England: 1086-1801

1290 may have represented the country's peak population given its almost exclusive reliance on low density renewable energy sources at this time. While there are records of coal use and shipping in the 13th century, the amounts involved are relatively small and the uses made of coal limited. For example, there is no mention of coal use in the Domesday book of 1086, although there were small shipments of coal into London during this period. We view the 1086 to 1290 period as one where England's economy was organic, labor mobility was low, and the Malthusian forces we have highlighted determined population growth. The 1290 population figure may have represented something close to a Malthusian steady state for England.

The Black Death struck England in the spring of 1348 and swept across the country with great intensity for the next two years. Although records are far from exact, the population plunged with those infected suffering a 70% mortality rate. Periodic outbreaks occurred during the next several hundred years, but this first outbreak created the greatest and most widespread loss of life. Not surprisingly, estimates for the aggregate population in 1377 are well below the 1290 peak. During the next three centuries the population recovered slowly only surpassing its pre-plague levels by perhaps 1600. Although it is not apparent from Panel A of the figure, the 1600 to 1800 period did not exhibit constant population growth.[27] The period started with a minor resurgence of the plague in 1625 and a major one in 1665; the great fire of London followed in 1666 and the Restoration in 1688. For almost the entire century England is actively at war with either Scotland, France, or Spain. These and other yet-to-be understood forces produced relatively slow population growth until the 1730s.[28] Thereafter population growth rose sharply. Most importantly, this period witnessed the spectacular growth of London. London rose from perhaps 350,000 in 1650 to become Europe's largest city by 1700 with over half a million inhabitants.

By the end of our sample period in 1801, England was a key colonial player, the world's leading power, a prolific trading nation, the world's largest coal producer, and home to Europe's largest most cosmopolitan city of almost one million people - London.[29] Coal use rose significantly over this early modern period starting with the introduction of a significant coal trade to London in the fourteenth century. At this time, most use was for domestic heating although coal also found uses in a range of industries from lime and salt production to smelting. As new mines opened in western England, Scotland, and Wales, the economic incentives for adopting coal use grew and technical barriers were overcome in a whole host of industries where coal replaced charcoal or wood. During this period coal moved from being exclusively a source of domestic and industrial heat to a provider of mechanical power via the Newcomen and Watt steam engines.

6.4 History through the Lens of Theory

Our theory makes several predictions linking the distribution of population over space to the distribution of energy resources. To make any headway at all in mapping one to the other, much simplification is required. For example, we make no attempt to discuss our model's predictions out of steady state; we assume exogenous (and unmodelled) technological progress drives long run economy wide changes in population levels; and we ignore completely the important role international trade may have played in these events. In doing so we hope to limit the discussion to elements which are novel to our theory, as there are already well

[27]See for example the discussion in Wrigley (1969) Chapter 3 and especially Figure 3.3.

[28]For example Wrigley (1969, p.96) reports London deaths in in 1603, 1625 and 1665 of 43,000, 63,000 and 97,000 respectively. In plague years one sixth to one quarter of a city's population may die.

[29]See Table 3.2 p. 61 Wrigley (2010) for estimates of London's population from 1520-1800.

known and excellent book length treatments highlighting important features of this period.

Without doing too much violence to the historical record we divide the 700 years from 1100 to 1800 into two periods reflecting the drastic changes moving England from an organic economy in 1086 to what Wrigley would call a mineral based economy by the early 1800s. We refer to the first period as the Organic Economy period, since during this time period the vast majority of energy used by households and industry was provided by renewable sources. It starts with the Norman conquest in 1066 and ends with the onset of the Black Death in the mid-14th century. The second period starts after the Black Death and runs until the first official census in 1801. The Black Death provides a useful starting point for the second period as it wipes the slate clean by lowering population levels.

We refer to this second period as the Early Modern Era since it precedes the most tumultuous and rapid changes of the Industrial revolution that transformed England into a true mineral based economy post 1800. We chose to stop our examination in 1801 since our primary focus has been on the organic economy. The nineteenth century saw revolutionary changes in energy use created by steam engines, railroads, electricity etc. and any examination of this period would require a wholesale investigation of the Industrial Revolution which is well beyond the scope of this paper. The Early Modern Era offers us something more manageable: an economy in transition from almost complete reliance on renewables at its beginning to one with significant dependence on coal at its end.

6.4.1 The Organic Economy Period

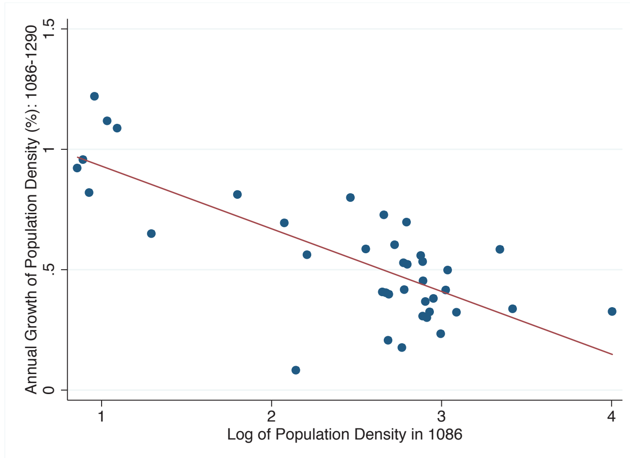
To start our discussion we consider the second panel of Figure 7. The figure plots kernel density estimates of the population densities in the 39 English counties across our five snapshot years. We present these figures in logarithms given the two orders of magnitude difference across counties in their population densities. We smooth the raw data using a kernel density estimator. Turning the figure 90° shows a series of density functions stacked over time; keeping it vertical shows us how these densities are related over time.

Consider first just the initial years 1086 and 1290 which represent the organic economy period. Our interpretation of this data is straightforward. With little migration across countries and diffuse natural blessings across England, the geographic distribution of the population should likewise be diffuse. The population level and its geographic distribution in 1086 however reflected the impact of the Norman conquest, and the losses of life and dislocation that occurred when William the Conqueror put down a series of revolts in northern England. This is apparent in the lower tail of the 1086 population distribution. Several northern counties were laid to waste and hence the geographic distribution of population across counties in 1086 represents a disturbed initial condition for a system moving towards what we have assumed to be a Malthusian steady state in 1290. We identify this steady state with the no-migration outcome of our model where a given geographic distribution of power densities across England together with slow technological progress determine populations.

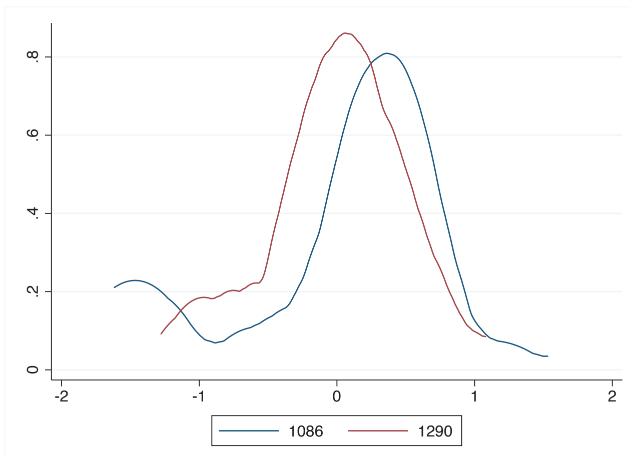
Outliers on the lower end of the 1086 distribution are exactly those of the counties William harried in the tumultuous period where Norman control over the North of England was in jeopardy. These counties are (in order of least dense to most) Cumberland, Westmoreland, Lancashire, Northumberland, Durham, Yorkshire and Cheshire. These counties recovered quickly and a simple plot of subsequent population growth rates on levels in 1086 is shown in Panel A of Figure 8. The figure provides at least suggestive evidence that these lagging counties were in fact catching up from this initial destruction during the next two hundred years. Since the unit of analysis here is a county and counties can differ quite substantially in the power density of available resources, the lack of a tight fit in this convergence style regression is to be expected. With little migration, each county should be approaching its own steady state conditional on the power density of neighboring resources and we have not conditioned on the determinants of power density in each locale.

One other feature of Figure 7 is that the overall distribution of population densities shifts upwards during the 1086 to 1290 period. We think this overall change in population reflects the workings of slow but steady technological progress relaxing the bonds of an organic economy. Evidence for this interpretation is shown in panel B of Figure 8 below where we compare the population distributions in 1086 and 1290 after recentering all observations by subtracting their period specific means. Once we correct for mean differences in population density across periods, the two distributions are quite similar (at least visually). The major difference is again the counties wasted by William's northern campaign. We can buttress this visual inspection with a Kolomogorov-Smirnov test. This test will however reject the null that the 1086 and 1290 (recentered) distributions are in fact identical, unless we exclude at least one wasted county. If for example, we exclude the two most Northerly counties bordering Scotland - Northumberland and Cumberland - we cannot reject the null hypothesis that the (recentered) distributions of population across the remaining counties is the same in both 1086 and 1290 at the 10% level of significance. With this smaller set of 37 counties, we do however reject at even the 1% level the hypothesis that the 1086 distribution and either the 1600 or 1801 distributions are the same.

There is as well a strong correlation across these periods on a county by county basis. For example, the correlation between population densities across counties (in levels or logs) across the two years is above .7 and the Spearman's rank coefficient is always above .65. These last two results suggest a rather strong permanence in the population distribution over this time period.



(a) Population Density and Population Density Growth



(b) Population Density

Figure 8: Population History of England: 1086-1801

Overall we take these results as evidence in favor of our theory's view of the period. The organic period featured a nation recovering after the conquest with wasted counties catching up while others grew slowly towards their limits. There was little change in the geographic distribution of population which is as it should be when migration is limited and key energy resources are constant over time. A further examination of this period could evaluate our theory more closely. We could for example relate the rank order of counties in terms of population density to what our theory sets out as the determinants of power density using data on solar insolation, precipitation, soil conditions and access to rivers and roads. At present we do not have sufficient data for such a test, although it is worth noting that some county specific attributes are obviously known. For example, even in 1086 Middlesex county - which contains London - is the most dense county in England. This is of course long before London became a hub for international trade, a recipient of coal shipments or even the capital of England. At this time, London's location on the Thames (which was cleverly chosen by the Romans) did however give it a far wider exploitation zone than inland cities with similar natural environments. The Thames is a tidal river meaning ships can both enter and leave with the current, while London's location is far enough inland to protect from attack and storms but close enough to benefit from tidal flows and a deep channel. During much, if not all, of this time period London was sourcing wood resources from the Thames valley far up river and coast-wise as well.[30] According to Proposition 7, a low friction alternative such as the Thames magnifies the power density of surrounding resources and should raise the density of population centers as well.[31]

In 1086 Middlesex county is the densest part of England with 54 people per square kilometer, but importantly this is not far in advance of Suffolk at 31 and Norfolk at 28 persons per square kilometer. Whatever advantage the Thames granted to London it was limited in an organic economy setting.

6.4.2 The Early Modern Era

Now consider the next three years in panel B of Figure 7. These years show a quite different pattern emerges as we enter the early modern era. The period starts with a population distribution in 1377 not unlike that of 1290. As we have already shown the plague lowered the national population tremendously, but as the density shows the incidence of death from the plague was not uniform. Not surprisingly death rates were greatest in the most dense counties (with the notable exception of Middlesex) and as a result the upper tail of the distribution for 1377 is somewhat compressed. The next two snap shots for 1600 and 1801 present a radically changing and changed England. By the year 1600 two features stand out. The first is a narrowing of the population distribution across a majority of the counties. The second is the run away growth of Middlesex county. By 1600 Middlesex county had a density almost 9 times greater than its next competitor (Surrey); while the same calculation in 1290 shows the ratio of Middlesex to the next competitor (Suffolk) is less than 2. In 1801, a similar feature appears, but now Middlesex is only 7 times as dense as its closest competitor. Density is however now an astounding 1178 people per square kilometer. In total the distributions shift upwards over time; develop a long right tail; and, have their mass shift left. All three features are readily explained by our theory.

We have already shown migration across cities both within and across counties tends to

[30] See, for example, Figure 1, p. 459 from Galloway et al. (1996). The figure, although for a somewhat later period, shows London's zone extends up the Thames valley past Henley, and includes almost all of Middlesex, and parts of Surrey, Berkshire, Oxfordshire, Buckinghamshire, Essex, Kent, and Hertfordshire.

[31] See Hilbert (1977) for a discussion of early London. Since *Londinium* was in fact a Roman invention, it was also the beneficiary of an ancient roman road system.

make big cities bigger and smaller ones smaller. If one big city in particular had ready access to a new higher density energy source, then consumption per capita would rise in the city, its population density would rise, and migrants would flock to it. The city would grow despite a rise in mortality. The remaining cities would lose population, and some villages or even cities would cease to exist. These impacts tend to tighten the population distribution in the sending counties, while lengthening the far right tail containing the receiving county. These results come about when one city has differential access to a dense power source together with the ability of the population to migrate.

While the Thames is a constant advantage to London and operative during both periods, the impact of the Thames would be much greater in the early modern era. To understand why imagine river and coastal shipping allowed you to deliver a net 100 kg of fuel rather than 50 kg of fuel; with straw or crop residue each kg brings at best 10 MJ; so river access brings you 500 MJ more energy than otherwise. If the fuel source is coal with 40 MJ per kg, then river access brings you 2000 MJ more energy than otherwise. Since low friction alternatives lower the energy cost per kg moved, while dense fuel sources deliver more joules per kg, it is easy to see how the benefit of a river is increasing in the power density of the fuel. If we couple this fact with the increased mobility during this period, our theory predicts just such a run a way pattern as the data depicts.

6.5 Some nagging questions

Our discussion has shown how it might be useful to think of our 700 year history in two periods distinguished by their energy sources. The geographic distribution of the population in the first period was relatively stable; in the second period major changes took place. In both periods one county emerged as an outlier, although the extent to which Middlesex dominated the data changed over time. While these results are consistent with our theory there are alternative explanations, some related empirical evidence in favor of our approach, and at least one substantive question that needs an answer.

6.5.1 Supply or Demand Shocks and Reverse Causality

Some combination of technological progress and the rise of London because of international trade could also explain much of the data presented. For example, we have already invoked technological progress to explain rising densities over the entire period although some of these increases may have come from income gains from trade. Since London was a major port, international trade could also be the source of London's growth, and the coal trade along the east coast could be *in response* to London's growth rather than the cause of it. Allen (2010) for example presents a convincing argument along these lines as part of his discussion of the industrial revolution. At some level the differences between these interpretations is small. Our explanation for the rise of London is primarily a supply side story. Coal deposits within London's exploitation zone led to its growth via migration flows from the rest of England. This growth would as well raise the attraction of London as a destination for imports and an outlet for exports giving rise to greater trade. Allen's interpretation is a demand side story. Rising trade drove London's growth which was facilitated by the proximity of coal and the east coast coal trade. Since London being a port on the Thames is important to both the shift in supply or shift in demand behind these interpretations it will be difficult to disentangle whether demand or supply considerations were key.

To determine whether proximity to dense fuels is sufficient to create a dominant city the size of London is difficult and the relevant data less convincing. To evaluate this possibility we should note that access to the new high density energy source – coal – affected far more than just London. In many, if not all, counties the population redistributed itself in ways consistent with our theory. To see this we present in Figure 9 a plot showing the rank of counties in terms of population densities in 1290 versus the same ranking in 1800. The ranking is inverse: the least dense county is ranked 1 and the most dense county ranked 38. If the rank order of counties did not change, then the counties most densely populated in 1290 would also be those most densely populated in 1801: all points would lie on a 45 degree line. If there was a strong relationship over time, the data should show a strongly positive relationship. As shown, there is very little - if any - relationship across these periods.

The implication is clear and immediate - the distribution of population densities created by the advent of coal and the beginnings of a mineral based economy in the early modern era had little to do with the densities created by the forces of the organic economy determining outcomes in 1290. A list of major cities and their population circa 1086 or 1290 with a comparison in 1800 will show the same result, as will a comparison of county densities. With the almost singular exception of London, the population redistributed itself over this period from traditional centers of population to new coal rich areas. Centers of population such as Norwich, York and Chester were eclipsed by Birmingham, Manchester, Leeds and Sheffield.[32]

6.5.2 Why isn't Newcastle - London?

If proximity to dense energy resources creates great cities, then why didn't Newcastle grow into London? Or alternatively, why didn't London fade into insignificance given its lack of coal reserves? On the face of it, this appears to be a glaring contradiction in our

[32] Chapter 3 of Wrigley (2010) contains a detailed account of these changes.

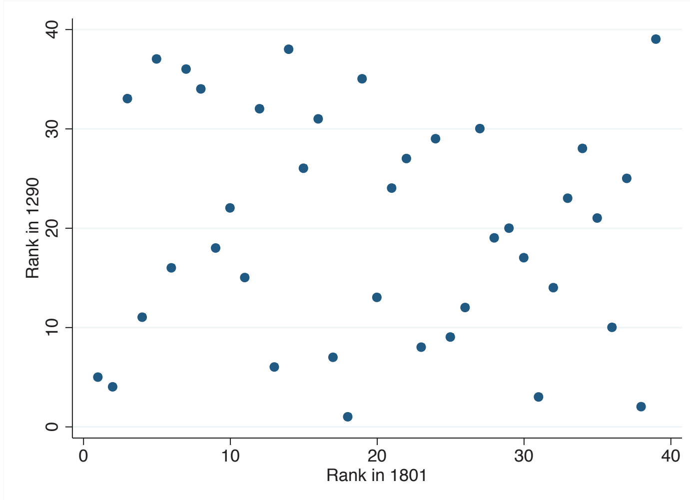


Figure 9: Population Rank: 1290 vs. 1801

analysis. The contradiction is however more apparent than real, and its resolution highlights an important feature of the theory that we have yet to discuss.

Part of the problem is that we think of London as far from the coal reserves of Newcastle. We associate distance on a map with difficulty to transport. In fact, the costs of water transport are only a small fraction of land transport implying that reserves not on major waterways were essentially trapped near their mine mouths. Estimates of overland transport costs of coal not surprisingly vary widely, but all estimates indicate an almost prohibitive cost of overland transport for any lengthy distance. For example, Turnbull (1987, 547) reports the price of coal from the pithead doubles in 10 miles overland or 200 miles via (coastwise) water. Using this metric, London is only the equivalent of 15 miles overland from Newcastle. Massachele (1993, 273) however reports less onerous costs using data on wheat shipments in the 14th century. He finds overland-to-river-to-coastwise transport costs in the ratio of 8:4:1. Using this metric London is less than 40 miles from Newcastle. In either case, transport costs are significant and especially so in relation to the value of the product. This is perhaps why the population centers in the Northwest of England grew so quickly given their proximity to coal fields, but surely Newcastle was still closer to the mines than London and so our original question remains.

The answer to this question relies on a feature of the model we have chosen to ignore in our analysis thus far. Throughout we have assumed that the exploitation zones for distinct agglomerations did not intersect. This allowed us to generalize the analysis of an Isolated State to a many city and many region country transparently. This is also the natural starting assumption since the maximum size of an exploitation zone is given by that in the

Only Energy model (since this occurs when $s = 1$) and we have already shown how this zone is a quadratic in power density. When energy resources have low power density, intersections are not likely; when they are very dense intersections are far more likely.

In the appendix we examine what happens when two cities compete for resources because their exploitation zones intersect. Their intersection defines a contested zone where the transport of resources to either center is profitable, and we assume the division of the region into exclusive zones depends on relative profitability. As we show, the city with a higher relative price for energy as measured by s commands most of the contested zone and will obtain most of the resources. We should also recall that s is strictly increasing in the population size of the city. Putting this logic together tells us that when a more populous city's exploitation zone intersects with a smaller city's, it will draw most of the contested energy resources to itself. In this case, the model predicts a form of inter-regional trade where resources physically close to one center will be delivered to a larger but more distant center. Since London was already England's largest city before the early modern era by virtue of the density of its surrounding resources and the benefits of the Thames, the initial lead given to it in the organic economy period made it - and not Newcastle - the largest beneficiary of the coal reserves on the Tyne.

This feature of the theory is of course very relevant in explaining why cities and countries today not endowed with oil, coal or natural gas can still thrive and grow. It suggests that sometime in their history their initial natural advantages gave them a size sufficient to become attractive energy markets. And it was their initial success in the world of organic economies that allowed them to source the far denser and far flung energy resources we see in the world today.

6.5.3 Food or Fuel Constraints?

Throughout we have aggregated all energy sources into one in order to highlight the role of power density. In doing so we have ignored the distinction between energy constraints coming from the costly collection of fuel for heating or industrial purposes (timber, straw, dung) and energy constraints coming from the costly collection of fuel for human growth and sustenance (crops, meat, dairy). While all of these fuels are the direct or indirect fruits of solar power, distinguishing between them may be important in some cases. It is unlikely that either the food or fuel constraints always bites first; the tightness of these constraints will likely be place and time specific. Nevertheless, it is important to recognize that our theory is equally applicable to situations where food as fuel constrains city size. A recent and very interesting paper by Nathan Nunn and Nancy Qian presents evidence directly on point.

Table 1: Crops and Power Density

Crops	Yield per acre (kg)	Energy Value (MJ)	Acres to deliver 42 MJ/day	Energy Content (MJ/kg)	Power Density (W/m^2)
Wheat	650	8,900	1.7	13.69	.07
Barley	820	11,400	1.4	13.90	.08
Oats	690	9,300	1.6	13.47	.07
Potatoes	10,900	31,900	0.5	2.92	.24

Nunn and Qian (2011) provide evidence linking the introduction of the potato from the New World to higher population levels and urbanization rates in the Old World. While measures of agricultural productivity may seem far afield from power density, it is important to recognize that any estimate of a crop yield in terms of caloric payback per acre is, in fact, an estimate of power density.[33] For example, Nunn and Qian motivate their work by providing estimates in terms of the number of acres of farmland required to provide 42 megajoules of energy per day for an annual crop cycle of one year (42 megajoules represents a caloric budget of 10,000 calories per day for a family of two adults and three children). Since joules per unit time must be proportional to Watts while acres are obviously a measure of area, the figures from their Table 1 can be manipulated to obtain power densities of various crops.[34]

In Table 1 the first three columns represent figures taken from Nunn and Qian's Table 1; the last two columns follow from simple manipulations of the first three. Energy content, e, is column two divided by one; power density, Δ, comes from multiplying column two by 1,000,000 to obtain joules, dividing this number by the number of seconds in a day to obtain Watts, and then dividing by the number of squared meters required to produce this energy flow. The result is a measure of the power density for each crop shown in the last column.

The Table shows two things. First, the power density of potatoes is at least three times greater than the other staples! Second, given the figures in Table 1 it is interesting to speculate why we tend to overlook the lowly potato. As the table shows the power density of the potato doesn't come from its enormous potential energy; in fact, its energy content, measured in $Joules/kg$ or e in our formulation is fully one third of the other staples. Instead

[33] Care has to be taken to make sure we are calculating the net yield after accounting for other inputs. A rise in yields due to manuring for example does not reflect an increase in the power density of the crop. In the potato case, cropping requires more labor inputs than for most cereals and we may want to account for these added energy costs to get a net energy yield figure. The table below does not contain this adjustment, but it is simple to make in our framework since it corresponds to an adjustment for net energy supply. See Nunn and Qian (2011) for a discussion of potential adjustments.

[34] A megajoule is a 1million joules; an acre is equivalent to 4046 sq. meters; and a day contains 24x60x60 seconds. A Watt is a joule/second. Using these equivalences we can easily calculate the power densities shown in the last column.

it is the ability of potatoes to provide extremely dense resources per unit area, or d in our formulation. This information is given in column one which shows the potato providing thirteen to sixteen as much as other staples. And this calculation brings out an important point: any analysis which focusses on energy content (just one determinant of power density) rather than power density itself, is likely to overlook how important the potato (or any other energy source) can be in economic development.[35]

Nunn and Qian use these estimates of the caloric payback per acre for various crops (wheat, barely, oats and potatoes) to argue that the introduction of the potato could have large effects on incomes, and via this channel populations and urbanization rates post 1700 in the Old World. Their baseline estimates show a very significant role for the potato: a 1% increase the amount of land suitable for potatoes in a country raises the urban share of its population by .36%; and overall, the introduction of the potato is responsible for 26% of the increase in the Old World population from 1700 to 1900. Combining these results with our theory surely goes some way in explaining the findings of Nunn and Qian since higher population levels and greater urbanization all follow directly from access to a higher density energy source.

While these are important results, they have only limited relevance to our specific case. Nunn and Qian find significant impacts on populations starting only in 1750 which is long after we observe large shifts in the population distribution across English counties. Earlier improvements in English agriculture could however play a role in raising densities and urbanization over our period.

7 Conclusions

This paper set out a simple general equilibrium model of an organic or green powered economy to understand both the past and potential future role renewable energy may play in an economy. We developed a model where energy sources differed in their power density, where the collection and exploitation of energy was costly, and where improvements in the energy system were complementary to energy exploitation. Within this context we showed how the concept of power density could play a major role in determining the size and density of urban agglomerations. An energy source twice as dense delivers 8 times the energy supply; power density determines the peak level of consumption per capita and its supporting population; roads and rivers effectively magnify the power density of available resources; and power den-

[35]Just as ignoring the speed with which a crop grows or the fallow period would likewise be unwise. These two features are captured in the reliance of power density on a crop's rate of growth r, and the intensity of cropping γ.

sity itself was shown to be the product of three fundamental characteristics of energy sources (renewability, dispersal, and energy content) plus a behavioral component reflecting energy suppliers' impatience.

We showed how the strength of complementarities within the energy system determined whether the model produced outcomes reflecting neoclassical features (the Almost Neoclassical case) or features of strong increasing returns (the Increasing Returns case). Since the Almost Neoclassical case produced familiar results we focussed throughout on the Increasing Returns case. Here we found that increases in market size produced falling energy prices and rising output and consumption per capita, but given the constraints imposed by the energy costs of collection and transport, energy costs eventually rose producing a single peaked profile for both output and consumption per capita, and a single dipped U-shaped profile for both energy and energy service prices. More dense energy sources produce greater peak consumption, support larger populations at these peaks, and present an obvious power density driven incentive for the agglomeration of economic activity. We labelled this tendency for power dense resource environments to create population dense centers of economic activity, the density-creates-density hypothesis.

To investigate this hypothesis we needed a specific mechanism to sort population across competing environments; a period when the energy system was relatively simple; and a place where data on energy sources and agglomerations was available. Our solution was to add a simple Malthusian mechanism to our Energy-Economy model and evaluate its performance by linking the distribution of power densities across geographic space with resulting population densities in English counties from 1086 to 1800. England provides a wealth of detail, large cross county variation in resources, and a well documented shift from a purely organic economy in 1086 to an economy where coal played a large role in 1800.

Our Malthusian model showed how natural environments with sufficient power density produced both small and large cities. Small cities are however fragile and not very productive constructs that either die out or grow to large cities. Large cities are both more permanent and more productive features of the landscape. In environments with low power density, no agglomeration can arise.

When we add migration in search of higher expected lifetime consumption we find migrants flow from low power density locations to high power density locations; migration tends to widen the distribution of city sizes and can produce city death; large cities offer higher per capita consumption but lower life expectancy than small cities; and the introduction of a new dense energy source in one location raises incomes everywhere but redistributes population towards the new dense source. Using these insights we reviewed the population history of English counties from the Norman conquest to the start of the nineteenth century. It is impossible to over estimate the depth and breadth of scholarship dedicated to understanding the population and economic history of this time and place, and our brief sketch of this 700 years of history must surely do great violence to many important facts and features of this period. Our hope is that our analysis throws new light on old questions, and provides provocative new questions to answer. A direct implication of our theory is that the massive redistribution of the English population to the virtually trapped, but power dense, resources in the coal producing regions is perfectly explicable; and that the rise of London to dominance by 1600 was due to its early, easy, and cheap access to the new high density energy source - coal. Whether our theory's implications for migration, productivity, and fertility patterns stand up to closer scrutiny is as yet unknown, but we are hopeful because excellent work establishing a causal link from power densities to population sizes has already be done.[36]

With regard to the future, our look at the past has been illuminating. While our energy system today is extremely complex and features an incredible transport of energy from producers to consumers that suggests locations are irrelevant, it would be unwise to underestimate the forces identified here. The power density and punctiform nature of fossil fuels has created both the ability, and the necessity, for the energy system we see today. Any significant movement back to an organic economy in our future will create new geographic winners and losers, require radical changes in our energy transportation system, and will create strong pressures for the redistribution of economic activity towards those regions with relatively dense, but still green, renewable resources.

[36]See our earlier discussion of Nunn and Qian (2011).

8 References

Ashraf, Q. and Oded Galor. 2011. "Dynamics and Stagnation in the Mathusian Epoch." *American Economic Review*, 101(5):1-41

Allen, Robert C. 2001. "The Great Divergence in European Wages and Prices from the Middle Ages to the First World War." *Explorations in Economic History*, 38: 411-447.

Allen, Robert C. 2008. "A Review of Gregory Clark's A Farewell to Alms: A Brief Economic History of the World." Journal of Economic Literature, 46(4): 946–73.

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

Allen, Robert C. 2011. *Global Economic History: A Very Short Introduction.* Oxford: Oxford University Press.

Bairoch, P. 1988. *Cities and Economic Development: From the Dawn of History to the Present.* Translated by Christopher Braider. Chicago: The University of Chicago Press.

Broadberry, S., B. Campbell, and B. van Leeuwen. 2011. *English Medieval Population: Reconciling Time Series and Cross Sectional Evidence.* LSE: Discussion paper.

Bryce, R. 2010. *Power Hungry: The Myth of "Green" Energy and the Real Fuels of the Future.* New York: Public Affairs.

Clark, G. 2005. "The Condition of the Working Class in England, 1209-2004." *Journal of Political Economy*, 113(6): 1307-1340

Clark, G., and D. Jacks. 2007. "Coal and the Industrial Revolution, 1700-1869." *European Review of Economic History*, 11(1), 39-72.

Clark, G. 2007. *A Farewell to Alms: A brief Economic History of the World.* Princeton, N.J.: Princeton University Press.

Davis, D. R. and Weinstein, D. (2002). "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 92(5):12691289.

Deane, P. and W. A. Cole. 1967. *British Economic Growth, 1688-1959: Trends and Structure.* 2nd ed. Cambridge: Cambridge University Press.

Droege, P., ed. 2009. *100% Renewable: Energy Autonomy in Action.* London: Earthscan.

Dyer, Alan. 1995. *Decline and Growth in English towns (1400-1640).* UK: Cambridge University Press.

Fouquet, R. 2008. *Heat Power and Light.* Edward Elgar

Fujita, M., P. Krugman, and A. J. Venables. 2000. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge: The MIT Press.

Galloway, J.A., D. Keene, and M. Murphy. 1996. "Fuelling the City: Production and distribution of firewood and fuel in London's Region, 1290-1400." *The Economic History*

Review 49 (3), 447-472.

Galor, O. and O. Moav. 2002. "Natural Selection and the Origin of Economic Growth." *The Quarterly Journal of Economics*, CXVII (4), 1133-1191.

Hatcher, J. 1994. "England in the Aftermath of the Black Death." *The Past and Present Society.* 144:3-35.

Henderson, V. 1974. "The Sizes and Types of Cities." *American Economic Revie*, 64: 540-656.

Henderson, V. 1980. "Community Development: The Effects of Growth and Uncertainty" *American Economic Review.* 70: 894-910.

Hilbert, C. 1977. *London: The Biography of a City.* revised edition, London: Penguin Books London.

Markusen, J. 1990. "Micro-foundations of external economies." *Canadian Journal of Economics*, 23(3): 495-508.

Massachele, J. 1993. "Transport Costs in Medieval England." The Economic History Review, Vol. 46, No. 2, 266-279.

Mitchell, B. R. and P. Deane. 1962. *Abstract of British Historical Statistics.* Cambridge: Cambridge University Press.

Morillo, S. 1996. *The Battle of Hastings: Sources and Interpretations.* UK: The Boydell Press.

Nef, J.U. 1966. *The Rise of the British Coal Industry.* Volumes One & Two. London: Frank Cass & Co. Ltd.

Neher, P.A. 1990, *Natural resource economics: conservation and exploitation.* Cambridge University Press.

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.

Rauch, J.E. 1993. "Does history matter only when it matters little? The case of city-industry location." *Quarterly Journal of Economics*, 108:843-867.

Samuelson, P.A. 1983. "Thunen at Two Hundred." *Journal of Economic Literature*, 21(4): 1468-1488.

Schofield, J. 1984. *The Building of London: From the Conquest to the Great Fire.* London: British Museum Publication Ltd.

Smil, V. 2006. *Energy: A Beginner's Guide.* Oxford, UK: Oneworld Publications.

Smil, V. 2008. *Energy in Nature and Society: General Energetics of Complex Systems.* MIT Press.

Smil, V. 2010. *Energy Myths and Realities: Bringing Science to the Energy Policy Debate.* Washington, D.C.: The AEI Press.

Turnbull, G. 1987. "Canals, Coal and Regional Growth during the Industrial Revolution", *The Economic History Review*, 40(4): 537-560.

Von Thunen, J.H. 1826. "Der Isolierte Staat." In *Beziehung auf landwrirtschaft und Nationalekonomie.* Hamburg.

Walker, D. 1995**.** *The Normans in Britain.* Historical Association Studies, Oxford: Basil Blackwell.

World Energy Council. 2010. *Survey of Energy Resources.*

Wrigley, E.A. *Industrial Growth and Population Change: A Regional Study of the Coal-filed Areas of North-west Europe in the Later Nineteenth Century*. Cambridge: Cambridge University Press.

Wrigley, E.A. 1962. "The Supply of Raw Materials in the Industrial Revolution." *The Economic History Review*, 15: 1-16.

Wrigley, E.A. 1969. *Population and History*. New York: McGraw-Hill (World University Library).

Wrigley, E.A. and R. S. Schofield. 1981. *The Population History of England, 1541-1871: A reconstruction*. London: Edward Arnold.

Wrigley, E.A. 1987. *People, Cities and Wealth: The Transformation of Traditional Society.* Oxford: Basil Blackwell.

Wrigley, E.A., R. S. Davies, J. E. Oeppen, and R. S. Schofield. 1997. *English population history from family reconstitution 1580-1837.* Cambridge: Cambridge University Press.

Wrigley, E.A. 2003**.** "Energy Constraints and Pre-industrial Economics." In *Economia e energia secc. XIII-XVIII S.* Cavaciocchi (ed), 156-172. Florence, Le Monnier.

Wrigley, E.A. 2009. "Rickman revisited: the population growth rates of English counties in the early modern period." *The Economic History Review*. 62(3): 711-735

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

A Appendix: Proofs

Proposition 1

From equation (20) we know the demand for energy is decreasing in p^W/w, it approaches infinity when p^W/w approaches zero and it approaches zero when p^W/w approaches infinity. From equations (25) and (29) we know energy supply is increasing in p^W/w as long as $\sigma > 1/z$. Thus, to prove existence it is enough to show the supply function is zero when p^W/w is zero. From inspection of (28) we know that p^W/p^C is zero when p^W/w is zero, which implies $s = 0$. From (24) this implies $TC'(W)$ is zero which by equation (5) implies energy supply is also zero.

Proposition 2

We first show that an increase in market size raises the relative price of power to the wage, that is $d(p^W/w)/dL > 0$. Total differentiating the demand and supply equations and rearranging we find

$$\widehat{W}_S = \frac{6(1-s)}{3-2s}\widehat{s} + 3\widehat{\Delta} \tag{A.1}$$

$$\widehat{W}_D = \widehat{L} - \widehat{\left(\frac{p^W}{w}\right)} \tag{A.2}$$

where a variable with a hat on top means $\hat{X} = dX/X$. By the definition of s and equation (28) we find

$$\widehat{s} = (1-s)\widehat{\left(\frac{p^W}{p^C}\right)} \tag{A.3}$$

and

$$\widehat{\frac{p^W}{p^C}} = \left(\frac{z\sigma - 1}{\sigma - 1}\right)\widehat{\left(\frac{p^W}{w}\right)} + \left(\frac{1}{\sigma - 1}\right)\widehat{L} \tag{A.4}$$

Equating (A.1) and (A.2), setting $\widehat{\Delta} = 0$ and rearranging terms we find

$$\frac{\widehat{\left(\frac{p^W}{w}\right)}}{\widehat{L}} = \frac{1 - \frac{6(1-s)^2}{3-2s}\frac{1}{\sigma-1}}{\frac{6(1-s)^2}{3-2s}\frac{z\sigma-1}{\sigma-1} + 1} \tag{A.5}$$

Therefore, the relative price of power to the wage raises with market size if

$$\sigma - 1 > \frac{6(1-s)^2}{3-2s} \tag{A.6}$$

The righthand side of the inequality is bound between 0 and 2 (because s can only take values between 0 and 1). When $\sigma \geq 3/[1-z]$ the lefthand side of the inequality is larger than 3 so that the inequality always holds.

Next, we show that an increase in market size raises the relative price of energy services to the wage. To do so, first take equation (28) and multiply both sides by w/p^W and find:

$$\frac{\widehat{p^C}}{w} = \frac{\sigma(1-z)}{\sigma-1}\widehat{\left(\frac{p^W}{w}\right)} - \left(\frac{1}{\sigma-1}\right)\widehat{L} \tag{A.7}$$

Replacing (A.5) in (A.7) we find

$$\frac{\widehat{\left(\frac{p^C}{w}\right)}}{\widehat{L}} = \frac{\frac{\sigma(1-z)-1}{\sigma-1} - \frac{6(1-s)^2}{3-2s}\frac{1}{\sigma-1}}{\frac{6(1-s)^2}{3-2s}\frac{z\sigma-1}{\sigma-1}+1} \tag{A.8}$$

Therefore, the relative price of manufactured goods increases with the size of the economy if and only if

$$\sigma(1-z)-1 > \frac{6(1-s)^2}{3-2s} \tag{A.9}$$

As before, the righthand side in the inequality (A.9) is bounded between 0 and 2. When $\sigma > 3/(1-z)$ we can see the lefthand side is always greater than 2 which ensures p^C/w increases monotonically with the size of the economy.

Proposition 3

The result follows directly from the proof of proposition 2. When $\sigma < 3/(1-z)$ the lefthand side of (A.9) is less than 2. Thus, for s close to zero the condition is violated and an increase in market size lowers the relative price of energy services to the wage. However, as the size of the market increases, s also increases. To see this replace equations (A.4) and (A.5) back in equation (A.3) to show:

$$\frac{\widehat{s}}{\widehat{L}} = \frac{\frac{z\sigma}{\sigma-1}}{\frac{6(1-s)^2}{3-2s}\frac{z\sigma-1}{\sigma-1}+1} > 0 \tag{A.10}$$

Hence, as the size of the market increases, s approaches 1 and the condition in (A.9) eventually holds.

Proposition 4

Because energy services are used in transportation, consumption is defined as the residual of output that remains after transporting the energy to the city: $C = Y(L, W^S) - TC(W^*, \Delta)$

where $Y(L, W^S)$ is given in (21), $TC(W, \Delta)$ in given in (5) and $W^S = W^* - TC(W^*)$. We are interested in characterizing consumption per capita (C/L) as a function of L and Δ. Total differentiation of consumption per-capita C/L yields:

$$\frac{d(C/L)}{dL} = \frac{dC}{dL}\frac{1}{L} - \frac{C}{L^2} \tag{A.11}$$

To be able to sign the derivative we need to characterize each element in individually.

To begin we need to find $\frac{dC}{dL}$. Totally differentiating consumption and making use of the relation between energy supplied, energy extracts and energy used in transportation we find

$$dC = Y_L dL + \left(Y_{W^S} - \frac{TC_{W^*}}{1 - TC_{W^*}}\right) dW^S - \frac{TC_\Delta}{1 - TC_{W^*}} d\Delta \tag{A.12}$$

where from (24) we have $TC_{W^*} = \frac{p^W}{p^C + p^W} > 0$ and from equation (5) we have $TC_\Delta < 0$.

Next, we want to find relations that allow us to simplify equation (A.12) into something more manageable. First, we can solve for p^W/w using equation (20) and replace the result in (28) to find the ratio of prices

$$\frac{p^W}{p^C} = \frac{\sigma - 1}{\sigma\beta}(1 - z)\left(\frac{1}{\alpha\sigma}\right)^{\frac{1}{\sigma-1}} \frac{(L^z (W^S)^{1-z})^{\frac{\sigma}{\sigma-1}}}{W^S} \tag{A.13}$$

Comparing this equation with (21) allows us to write $\frac{p^W}{p^C} = (1-z)\frac{Y}{W^S} = TC_{W^*}/(1 - TC_{W^*})$ where the last inequality follows from (24). We also find the marginal productivity of energy in final output is $Y_{W^S} = \frac{\sigma(1-z)}{\sigma-1}\frac{Y}{W^S}$. Replacing these results in (A.12) we find

$$dC = Y_L dL + \left(\frac{1-z}{\sigma-1}\frac{Y}{W^S}\right) dW^S - \frac{TC_\Delta}{1 - TC_{W^*}} d\Delta \tag{A.14}$$

Now that we have simplified equation (A.12) we can set $d\Delta = 0$ in (A.14) to find

$$\frac{dC}{dL} = \frac{1-z}{\sigma-1}\frac{Y}{W^S}\frac{dW^S}{dL} + Y_L \tag{A.15}$$

Replacing (A.15) and noticing from (21) that $Y_L = \frac{z\sigma}{\sigma-1}\frac{Y}{L}$ we can rewrite equation (A) as:

$$\frac{d(C/L)}{dL} = \left(\frac{1-z}{\sigma-1}\frac{\hat{W}^S}{\hat{L}} + \frac{z\sigma}{\sigma-1}\right)\frac{Y}{L^2} - \frac{C}{L^2} \tag{A.16}$$

Replacing equation (A.10) in equation (A.1) we find

$$\frac{\hat{W}^S}{\hat{L}} = \frac{6(1-s)^2 z\sigma}{6(1-s)^2(z\sigma-1)+(3-2s)(\sigma-1)} > 0 \tag{A.17}$$

so that

$$\frac{d(C/L)}{dL} = \left(\frac{1-z}{\sigma-1}\frac{6(1-s)^2 z\sigma}{6(1-s)^2(z\sigma-1)+(3-2s)(\sigma-1)} + \frac{z\sigma}{\sigma-1}\right)\frac{Y}{L^2} - \frac{C}{L^2} \tag{A.18}$$

Now we are ready to show that consumption per capita peaks when $\sigma < 3/(1-z)$. To show that consumption per capita is single peaked, $\frac{d(C/L)}{dL}$ must be positive for small values of L, negative for large values of L and the sign of $\frac{d(C/L)}{dL}$ can change only once. First we find the critical value of L where the slope is zero by setting (A.18) equal to zero:

$$\underbrace{\left(\frac{1-z}{\sigma-1}\frac{6(1-s)^2 z\sigma}{6(1-s)^2(z\sigma-1)+(3-2s)(\sigma-1)} + \frac{z\sigma}{\sigma-1}\right)}_{LHS} = \underbrace{\frac{C}{Y}}_{RHS} \tag{A.19}$$

Taking the derivative with respect to L and replacing equation (A.10) we can show that $dLHS/dL < 0$. From the definition of s we know that as the size of the market increases, the price of energy services to the wage also increase and s approaches 1. This implies that the LHS approaches $\frac{z\sigma}{\sigma-1} < 1$ where the last inequality follows from the assumption of decreasing returns to labor. We have characterized now the LHS as monotonically decreasing in L and approaching $\frac{z\sigma}{\sigma-1} < 1$ for large L.

Next, we show that the RHS is monotonically increasing in L. To see this we can express $C/Y = 1 - TC(W^*, \Delta)/Y$. From equations (A.15) and (A.17) we know $dC/dL > 0$, so it must be true that Y increases faster than TC when the size of the marker increases. This implies that when L increases, TC/Y decreases eventually reaching 0 when L approaches infinity. This implies that the RHS increases as a function of L and reaches 1 for very large L.

Given that LHS is monotonically decreasing in L and RHS is monotonically increasing with L, then LHS and RHS intersect if and only if $LHS|_{L=0} > RHS|_{L=0}$. To calculate $LHS|_{L=0}$ simply set $s = 0$ to find $LHS|_{L=0} = \frac{1-z}{\sigma-1}\frac{6z\sigma}{6(z\sigma-1)+3(\sigma-1)} + \frac{z\sigma}{\sigma-1}$. To calculate $RHS|_{L=0}$ we first need to recall that $TC(W^*, \Delta) = \frac{2}{3}\frac{\pi\Delta^3}{(\mu g d)^2}s^3$ which from (30) is equal to $TC(W^*, \Delta) = \frac{2}{3}\frac{sW^S}{1-2s/3}$. Now, using $p^W/p^C = (1-z)Y/W^S$ and from the definition $p^W/p^C = s/1-s$ we find $C/Y = 1 - \frac{2}{3}\frac{(1-s)(1-z)}{1-2/3s}$. Hence when L approaches zero, $RHS|_{L=0} = \frac{1+2z}{3} < 1$. Putting these two result together, there will be an intersection, and this intersection will be unique when $\sigma < 3/(1-z)$. This is the Increasing Returns Case. If $\sigma > 3/(1-z)$ there is no intersection and C/L is monotonically decreasing with the size of the market. This is the Almost Neoclassical case.

The proof for production per capita is similar. Production is equal to $Y = \frac{\alpha}{\beta}(\sigma - 1)\left(\frac{1}{\alpha\sigma}L^z(W^S)^{1-z}\right)^{\sigma/(\sigma-1)}$. The rate of change of production per capita can be written as $\widehat{\frac{Y}{L}} = \frac{1-(1-z)\sigma}{\sigma-1}\widehat{L} + \frac{(1-z)\sigma}{\sigma-1}\widehat{W^*}$. Replacing equation (A.17) we find that

$$\frac{\widehat{\frac{Y}{L}}}{\widehat{L}} = \frac{1-(1-z)\sigma}{\sigma-1} + \frac{(1-z)\sigma}{\sigma-1}\frac{6(1-s)^2 z\sigma}{6(1-s)^2(z\sigma-1)+(3-2s)(\sigma-1)} \tag{A.20}$$

We know s approaches zero as L approaches zero, so production per capita first raises with the size of the market when the previous equation is positive, that is when $\sigma < 3/(1-z)$. For large L production per capita always falls because as s approaches 1, the expression $\widehat{\frac{Y}{L}}/\widehat{L}$ approaches $\frac{1-(1-z)\sigma}{\sigma-1}$ which is negative, assuming decreasing returns to labor in equilibrium. If $\sigma > 3/(1-z)$, then production per capita falls monotonically with the size of the market.

Proposition 5

Equating (A.1) and (A.2), setting $\widehat{L} = 0$ and rearranging terms we find

$$\frac{\widehat{\left(\frac{p^W}{w}\right)}}{\widehat{\Delta}} = -\frac{3}{\frac{6(1-s)^2}{3-2s}\frac{z\sigma-1}{\sigma-1}+1} < 0 \tag{A.21}$$

Replacing (A.21) in (A.2) we find $dW^D/d\Delta > 0$. Rearranging the expression in (A.14) and, given $W^S = W^D$ in equilibrium, we find $dC/d\Delta > 0$.

Proposition 6

The key result to prove here is that the peak of consumption per capita increases with Δ. That is, both the C/L and L are larger at the peak when Δ increases. The proof follows from equations (A.18) and (A.19). Given the characteristics of LHS and RHS in equation (A.19), if LHS increases with Δ and RHS decreases with Δ for all L, then it is necessarily true that L increases with Δ (See Figure 10).

We start calculating the result of the LHS:

$$\frac{dLHS}{d\Delta} = -\frac{12z\sigma(1-z)(2-s)(1-s)}{(3(\sigma+2z\sigma-3)+2(7-\sigma-6z\sigma)s+6(z\sigma-1)s^2)^2}\frac{ds}{d\Delta} > 0 \tag{A.22}$$

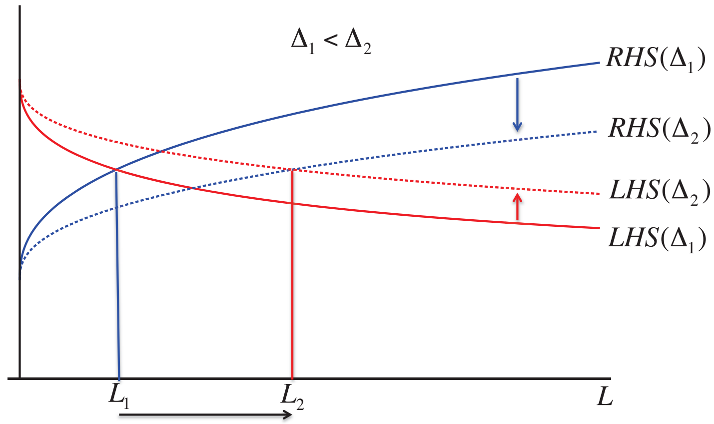


Figure 10: Consumption per-capita

where $ds/d\Delta < 0$ directly from equations (A.1)-(A.5). Similarly, for the RHS we have

$$\frac{dRHS}{d\Delta} = \frac{2(1-z)}{(3-2s)^2}\frac{ds}{d\Delta} < 0 \tag{A.23}$$

Then peak value of L increases with Δ.

To show that the peak of consumption per capita increases with Δ we simply need to show that C/L is an increasing function of Δ for all L.

$$\frac{d(C/L)}{d\Delta} = \frac{dC}{d\Delta}\frac{1}{L} - \frac{C}{L^2}\frac{dL}{d\Delta} \tag{A.24}$$

were the second term is equal to zero, so that $\frac{d(C/L)}{d\Delta} = \frac{dC}{d\Delta}\frac{1}{L} > 0$ directly from Proposition 5.

Proposition 7, 8 and 9

In the text.

Proposition 10

The slope of the E-Line is positive which implies its point of tangency with the $c(e^{max}_{\Delta_i}; L_i)$ will occur to the left to the maximum for $c(e; L_i)$. So it follows from Proposition 6 that $\frac{\partial c(e;L_i)}{\partial L}$ increases with Δ. Hence a higher Δ_i requires the corresponding E-line that is tangent to

$c(e; L_i)$ to be higher. As e increases, so it does $\frac{\partial E-line}{\partial L}$. This implies an increase in $\frac{\partial c(e^{max}_{\Delta_i}; L_i)}{\partial L}$ at the new point of tangency which defines an increasing function $e^{max}_{\Delta_i}(\Delta_i)$. That is, the point of tangency for any city will be higher the higher is its power density. Thus, if the migration equilibrium expected lifetime consumption is such that $e^*_N < e^{max}_{\Delta_1}$ and by definition $e^*_N \leq e^{max}_{\Delta_N}$ then there will be an intersection between the $e^* - Line$ and all the $c(e; L_i)$ curves, for any $e^{max}_{\Delta_1} \leq e^* \leq e^{max}_{\Delta_N}$

B Why Newcastle didn't become London

Cities within regions are proximate and may compete for energy resources which would in turn affect their exploitation zones, energy supply, etc. To examine how this interaction may play out we consider the case of only two cities that are close enough to have their exploitation zones overlap in an area we call the contested zone. Label these two cities A and B, with exploitation zones of radii R_A and R_B respectively. We assume the distance between the two cities is small enough to create a contested zone ($R_{AB} < R_A + R_B$) but large enough to exclude the city cores from this zone ($R_{AB} > R_A$ and $R_{AB} > R_B$).[37] Energy producers on the zero rent margin for city A will necessarily find it more profitable to go to city B; and likewise producers on the zero rent margin for city B will find it more profitable to go to city A. In between these clear cut cases there will exist a set of energy producers that are indifferent between bringing their energy to city A or city B. Consider energy producer I that holds Δ units of power and is located inside the intersection of the two exploitation zones. The distance from energy producer I to city core A is R_{AI} and the distance from energy producer I to city core B is R_{BI}.

Profits for energy producer I are given by

$$\Pi_{AI} = p^W_A \Delta - (p^W_A + p^C_A)(\mu g d) R_{AI} \tag{B.1}$$

$$\Pi_{BI} = p^W_B \Delta - (p^W_B + p^C_B)(\mu g d) R_{BI} \tag{B.2}$$

The producer has Δ watts of power to sell, but pays costs proportional to the energy cost of delivering them to either core. The energy costs of transport are simply $(\mu g d/\Delta)$ for each watt of power moved one meter. Therefore energy costs to core A are $(\mu g d/\Delta)\Delta R_{AI} = (\mu g d) R_{AI}$ and similarly for city B.

The energy producer will be indifferent between the two cities if the amount of con-

[37] Other cases are clearly possible to investigate but harder to deal with analytically. For example, if the city cores are very close together the indifference curve takes on a different shape, but our simulation shows the results we focus on below are very similar.

sumption goods obtained from each city are equalized. That is, profits measured in terms of consumption goods must be equal for the indifferent energy supplier. That is, the indifferent supplier must be found where:

$$\frac{\Pi_{AI}}{p_A^C} = \frac{\Pi_{BI}}{p_B^C} \tag{B.3}$$

In addition, we also know that energy suppliers on the margins of each city's exploitation zone earn zero profits at current prices. Therefore the zero rent producers must satisfy:

$$\Pi_A = p_A^W \Delta - (p_A^W + p_A^C)(\mu g d) R_A = 0 \tag{B.4}$$

$$\Pi_B = p_B^W \Delta - (p_B^W + p_B^C)(\mu g d) R_B = 0 \tag{B.5}$$

By using (B.1), (B.2), (B.4), and (B.5) the indifference condition (B.3) can be written as

$$(R_A - R_{AI}) = \frac{(1 - s_A)}{(1 - s_B)} (R_B - R_{BI}) \tag{B.6}$$

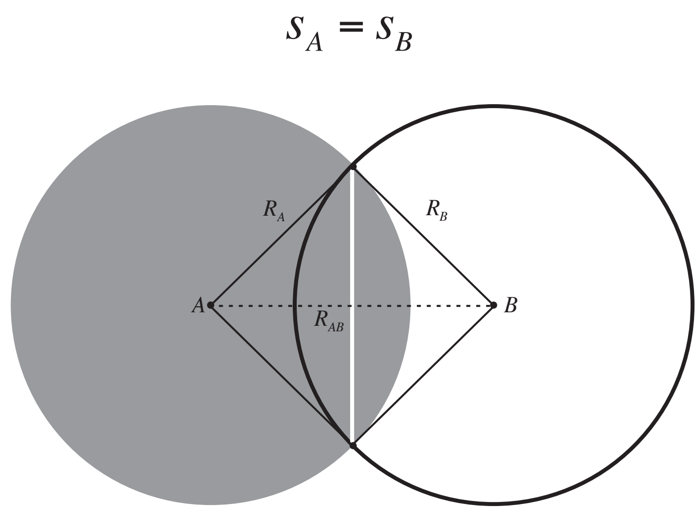
This condition defines a curve dividing the contested zone according to whether $s_A > s_B$. Note if $s_A > s_B$ then $R_A - R_{AI} < R_B - R_{BI}$ and city A gets the biggest share of the contested zone. Not surprisingly, the city with the highest relative price for energy gets the lion's share of the contested zone.

To fully characterize this curve we need to understand how the position of the indifferent energy producer changes inside the contested zone. We have assumed thus far that our indifferent producer is on the straight line joining the two city centers, but this is just one point. To solve for the entire line consider the angle formed between the line segment AB and the line segment AI. Call this angle ϕ and use the law of triangles to find the following condition:

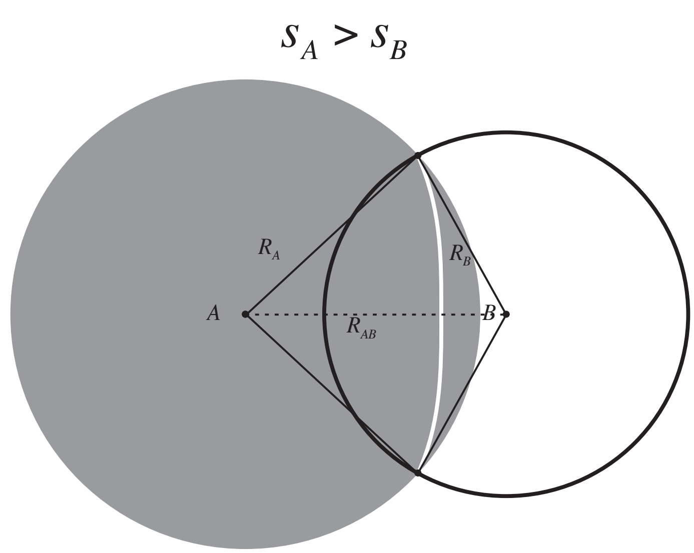
$$R_{BI}^2 = R_{AI}^2 + R_{AB}^2 - 2R_{AI}R_{AB}\cos(\phi) \tag{B.7}$$

Using equations (B.6) and (B.7) we can now find an expression for $R_{AI}^*(\phi)$ that determines the position of the indifferent energy producer for any given ϕ.

Figure 11 shows how this curve is determined by the relative size of the exploitation zones. In each figure the curve splitting the contested zone is marked by a red dotted line. The area to the left of this curve belongs to city A, while city B takes the remainder. In Figure 13(a) the two regions are identical, hence the two cities split the contested zone equally. In Figure 13(b) Region A is larger than Region B and captures most of the contested zone. Holding s_B constant, as s_A approaches 1 all of the contested area belongs to city A and as s_A approaches zero, all of the contested area belongs to city B.



(a) Two regions are equal



(b) Region A is larger than Region B

Figure 11: Regions of Exploitation

In the case of symmetric regions shown by the first figure the solution to the curve is somewhat simpler. When regions are symmetric $R_{AI} = R_{BI}$ and the indifferent energy producer is situated a distance R^*_{AI} away from the city A forming an angle ϕ with the segment AB:

$$R^*_{AI} = \frac{R_{AB}}{2\cos(\phi)} \tag{B.8}$$

which is what one would expect given that the set that describes this "indifference curve" is a straight line that divides the contested area in two.

With this characterization in hand we can solve for the energy supplied to each city which is given by the following integral:

$$W^S = 2 \times \left[\int_0^{\phi_{cr}} \int_0^{R^*_{AI}} v(\Delta - \mu gd \cdot v) dv d\varphi + \int_{\phi_{cr}}^{\pi} \int_0^{R_A} v(\Delta - \mu gd \cdot v) dv d\varphi\right] \tag{B.9}$$

where ϕ_{cr} is the angle that determines the intersection between the two areas of exploitation and it is the solution to

$$R_B^2 = R_A^2 + R_{AB}^2 - 2R_A R_{AB}\cos(\phi_{cr}) \tag{B.10}$$

which is given by:

$$\phi_{cr} = \arccos\left(\frac{R_A^2 - R_B^2 + R_{AB}^2}{2R_A R_{AB}}\right) = \arccos\left(\frac{R_{AB}}{2R_A}\right) \tag{B.11}$$

where the second inequality follows from the assumption of symmetry.

The solution to this integral gives a supply curve again quite similar to our earlier results. Specifically:

$$W^S = (\pi - \phi_{cr})\frac{\Delta^3}{(\mu gd)^2} s_A^2 \left(1 - \frac{2}{3}s_A\right) + \Psi \tag{B.12}$$

$$\Psi = \frac{R_{AB}^2 \Delta}{4}\left(\tan(\phi_{cr}) - \frac{1}{3}\mu gd R_{AB}(\operatorname{arctanh}(\tan(\phi_{cr})) + \frac{1}{2}\sec(\phi_{cr})\tan(\phi_{cr}))\right) > 0 \tag{B.13}$$

Several results follow from these calculations. First, any contested zone lowers energy supply from what it would have been in its absence. The explicit formulation given above takes out a pie slice from the exploitation zone with radial circumference of ϕ_{cr}, and then adds back in the net energy delivered from the triangle of the exploitation zone (within this pie slice) that the city actually gets to keep $\Psi > 0$. Second, energy supply is smaller the closer are the cities together and of course larger the more dense is the energy source to begin with. Putting these results together with those from earlier sections, it is clear that cities with overlapping exploitation zones can only exist if power density is sufficiently high (recall density must exceed a certain minimum to support a positive population); and the distance between cities who could overlap in exploitation zones needs to be greater in poor less power dense regions, but could be quite small in power dense rich regions. Third, even if a city is surrounded by dense resources a distant city with a large exploitation zone may end up benefiting from those resources since they will be transported to it rather than the closest city. A large exploitation zone results from, all else equal, a large population since s is increasing in population size. Therefore, London and Newcastle's exploitation zones did intersect; and coal was carried to London rather than Newcastle because its value was higher there.