Back to the Future of Green Powered Economies

Juan Moreno Cruz M. Scott Taylor*

March 9, 2016

Abstract

The purpose of this paper is to examine how the location and productivity of available energy resources affects the spatial distribution of economic activity. More productive (power dense) resources generate far greater energy supply, stronger incentives for infrastructure investments, higher consumption per capita, and more population dense agglomerations. Using county level population data from 1086 to 1801 we investigate how England's move to the very power dense energy source represented by coal altered its economic geography. Local access to coal is responsible for over 24% of growth in all but two coal counties; and almost 40% in others. The world's first energy transition created a dramatic reshuffling of the economic landscape.

Preliminary — Please do not cite without permission.

*Moreno-Cruz, Assistant Professor, School of Economics, Georgia Institute of Technology, Atlanta, Georgia; Taylor, Professor of Economics and Canada Research Chair, Department of Economics, University of Calgary, Canada and Research Associate at the NBER. 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.

1 Introduction

In the not so distant past, solar power captured by wind flows and biomass production were the sole energy sources fueling human existence. Students of this era, particularly economic historians, have studied the role these nature given limits played in determining the economic limits to land use, and debate to this day the impact fossil fuels have had on economic activity. Students of our current era, particularly energy economists, look to a not so distant future where post-industrial green powered economies may again survive on the direct and indirect fruits of solar power. And similarly, energy economists today debate the economic effects of moving back to renewables away from fossil fuels. These two disciplines, separated by several hundred years in their respective periods of study, share a common interest in understanding how the introduction of a new energy source affects economic activity.

The purpose of this paper is to take a step back into the past to consider what our future may look like by examining how the introduction of a new energy source affects the size and location of economic activity. To do so we build a new spatial model where energy resources differ in their productivity and location, and then use this model to develop a novel empirical design to evaluate its implications. Our main theoretical result is the statement of the *density-creates-density hypothesis* which links the power or energy density of neighboring resources with the resulting geographic distribution of population and economic activity. Our main empirical result is the finding that the geographic location of coal deposits in pre-fossil fuel English counties, caused a major redistribution of population and economic activity across these counties once the economy transitioned to coal. All coal counties grew dramatically because of coal; some owe almost 40% of their growth to coal; and access to coal reshaped the entire distribution of people and production. Since our sample period runs from 1086 to 1801 many other forces are also at play in the English countryside over this tumultuous 800 years. Coal is not the driver of all changes, nor perhaps even the majority of changes, but it is a very important driver nonetheless.

To understand how access to energy resources affects the location of production we need a model of economic geography, and any model of economic geography must have three components: a reason why agglomerations exist; a reason why they are bounded in size; and a mechanism for sorting people and activity across locations. Our work is no different. The forces for agglomeration in our model arise for very standard reasons: increasing returns from specialization in production are facilitated by large markets, although the specifics are tailored to our energy application. Agglomerations are limited in size by an explicit spatial structure which ensures a rising cost of bringing marginal energy resources to the core. To

sort people across locations, we adopt a mechanism that fits both the time and place of our application - a simple Malthusian mechanism where individuals sort across competing locations on the basis of death rates and consumption per capita.

A novel element of our theory is that we differentiate energy resources accordingly to their "spatial productivity" and measure this productivity using the concept power density. The concept power density is unfamiliar to most economists but it simply measures the flow of energy a source can provide per unit area needed for its exploitation. Power is the flow of energy per unit time, and this together with the flow of labor and capital services is what drives economic activity. While our explicit measure of spatial productivity is rarely used in economic analyses, the productivity of soil, the yield of various crops, the regeneration power of forests, the energy content of fuels, and the fecundity of animal species are all reflective of this more general measure.[1] Seen in this light, measures like power density have received considerable attention in the analyses of economic historians where the interaction of an environment's spatial productivity and transport costs are held responsible for the small city sizes in pre-industrial times.[2] These arguments are however exclusively verbal, and to examine how variation in the productivity of energy resources across space affects locational choices we develop an explicit model of energy exploitation, transportation and use.

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 transmission lines that comprise the energy system.

To model an energy system we follow both energy economists and economic historians alike who 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 pre-industrial 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

[1] In Moreno-Cruz and Taylor (2013) we link the power density of both renewables and non-renewables to these more primitive determinants.

[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 Nunn and Qian (2011).

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. The combination of an energy source, its set of facilitating intermediate goods, and the pattern of production and consumption 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 we adopt a framework where increased specialization in the set of intermediate goods increases the efficacy of the energy system.

Finally we must allow for the possibility that cities compete for population. 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 1086 to 1801 through the lens of our theory.

Our theory provides four key results. First, even small differences in the power density of resources create large differences in equilibrium outcomes; because, as we show energy supply is a cubic function of power density. This reflects a scaling law arising from our spatial context and it implies resources with twice the power density, deliver eight times the energy supply. Since straw and wood are not very power dense resources, this theoretical finding supports the key role economic historians have given to energy constraints in pre-industrial economies. Second, there is a strong complementarity between the power density of energy resources and the incentives to provide improved transportation. Resources with low power density - dung, straw, wood - provide meagre incentives for transportation improvement; resources with high power densities - coal and oil - provide much larger incentives. This complementarity magnifies the workings of our scaling law creating even larger differences in the supply of more and less dense fuels. Third, the interaction of rising marginal costs of energy exploitation and increased efficiencies in the energy delivery system, produce a first rising and then falling schedule of per capita consumption as the size of markets grow. Peak consumption per capita is higher, and occurs at larger market sizes, when the energy resources in question are more power dense. Finally, if agents sort across locations trading off consumption per person and expected lifetime then locations with very energy dense

resources create very population dense agglomerations - this is the density-creates- density hypothesis.

Our model is highly abstract 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 use our theory as a lens to review the first order developments of the period and link them to implications for data. The Norman invasion, the Black Death, the end of Serfdom, the rise of London and the movement to coal all feature prominently in our narrative; but most importantly, our discussion suggests a very natural empirical strategy allowing us to move from narrative to causal argument. Our empirical design boils down to a relatively simple difference-in-difference estimation where over time and cross sectional variation in - the ability to use, and the fortune to have - coal deposits determines whether a county is treated with a new higher density energy source.

Our empirical results suggest access to coal deposits had large effects on the growth of coal counties. Since the overland transport of coal was difficult and expensive, local abundance of coal was critical to its supply price. With the exception of Newcastle coal delivered at low cost to London, most areas relied on local deposits and faced very local energy markets. In this world of very segmented energy markets, we find coal counties often owed 25% or more of their growth to the local proximity of coal. Of the 14 coal counties, only two had less than 20% of their growth over the treatment period attributed to coal; several more than 25%, and three counties almost 40%. While these impacts are large, other forces at work in the English economy were equally powerful. Textile production, booming international trade, and political developments all played a role in reshaping England over this period. To assess the importance of coal relative to these other forces, we control for access to ports and roads, for known centers of textile production, and common to country shocks. We then construct a counterfactual population distribution for English counties in 1801 assuming coal counties do not benefit from their preferential access. This construction suggests a more nuanced interpretation is in order. While having local deposits helped all of the 14 coal counties, five of the coal counties would have moved up the economic hierarchy without it; three of them were destined to remain laggards in any case; and only five of the coal counties were truly transformed.

Our work is related to contributions coming from three large and largely disjoint literatures: economic history; energy economics; and economic geography. In Moreno-Cruz and Taylor (2013) we developed a simple spatial model to show how to incorporate power density concepts into standard economic analysis. We provided methods to measure power density for both renewables and non-renewables, and examined the role it may play in a very simple economic system. Since energy was the only input and only output of the production pro-

cess, we referred it as the Only Energy Model. Here we have effectively embedded the Only Energy Model in a market economy setting. 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 UK made urbanization and further progress impossible in what he refers to as the Organic Economy period. 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 as it enters the fossil fuel driven Mineral economy period. 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 discussion of English history and the role of energy owes much to Wrigley (2010) and Allen (2009); our empirical design owes much to Nunn and Qian (2011).

Finally, our work is related to several empirical papers in economic geography where large shocks are often identified as natural experiments. Davis and Weinstein (2002) is the seminal contribution, but recent important contributions include Redding and Sturm (2008) and Bleakley and Lin (2012). Much of this work is focussed on the question of permanence. What causes permanence in the location of production and what does this tell us about competing theories of economic geography. Our work fits well in this tradition although it has some unique elements. It documents strong permanence within a given regime (both the Organic and Mineral economy periods); an extreme reshuffling across regimes; and ties a change in the locational fundamentals of energy supply to this shift. Although locational fundamentals have often been cited as important or even pivotal, their very permanence has made it difficult to assess their importance empirically. An energy transition however provides exactly this needed variation.

The rest of the paper is organized as follows. We start in section two with a descriptive sketch of the sample period using simple statistics and data representations. Following our sketch we introduce our Energy-Economy model where energy demand from firms interacts with energy supply to determine the rate of exploitation for a single renewable energy source introducing the forces behind agglomeration. We then extend our theory to allow for en-

dogenous population growth. In section 4 we use this extended model to review English population history through the lens of our theory. This discussion suggests a natural empirical framework that we then employ in section 5 to produce estimates of coal's causal impact on populations across the country. A short conclusion ends the paper. A series of sensitivity tests, further data and all proofs and lengthy calculations are relegated to an online appendix.

2 A Sketch of England's Population History 1086-1801

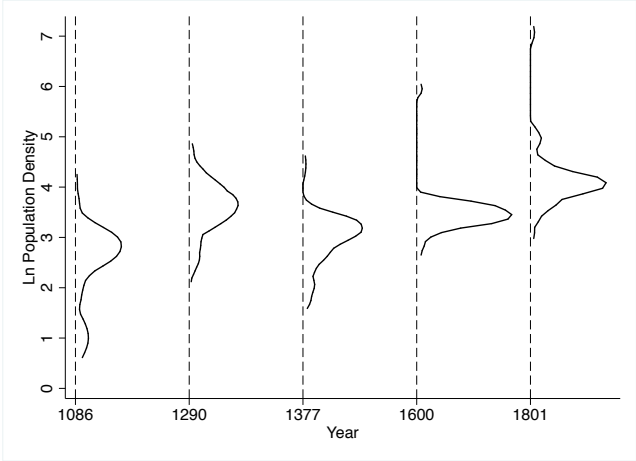
The population history of England has been studied by legions of academics with demographers, economists, historians and medievalists alike producing hundreds of contributions dissecting and debating the period from the Norman conquest to the very first years of the early modern era.[3] Demographers in particular have spent decades carefully collecting and analyzing data from this period, and here we mine this very deep vein of scholarship. Our key population data is taken from recent estimates by Broadberry et al. (2011). These authors reconstruct the population history of England by combining both official returns from census data (the Domesday census in 1086 and the first modern census in 1800) with information drawn from muster rolls, manorial accounts, parish registers, and poll tax returns[4] to develop estimates of England's population history from 1086 to 1800. Their construction builds on figures presented in the earlier work of Russell (1948), Campbell (2008) and Wrigley (2009).

The Broadberry et al. (2011) population figures are employed to construct the three panels shown in Figure 1. Panel 1.(a) presents aggregate population figures for England in the years 1086, 1290, 1377, 1600, and 1801. Panel 1.(b) exploits the across county variation in the data. It combines the county specific population figures for each of these years with areal measures of the 39 ancient English counties to generate county population density measures. This panel shows kernel density estimates for these 39 English (ancient) counties for our five snap-shot years. For ease of comparison, we present these figures in logarithms, rotate the densities by 90°, and order them chronologically. Panel 1.(c) is a rank-rank plot showing how a county's rank in population density in the first year 1086 is related to its rank in 1801. Rank is assigned so that 1 is the most dense county and 39 the least dense.

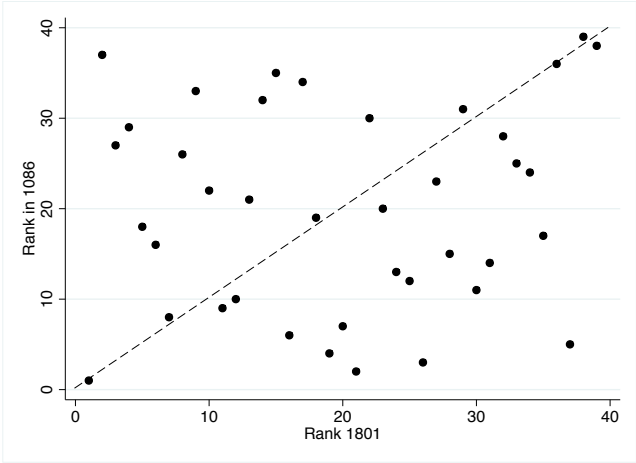
[3]Several excellent book length treatments by Sir Tony Wrigley present a fascinating picture of English demography and economics over this time period. Especially relevant are Wrigley (2010) that focusses on energy issues; Wrigley (1969) and Wrigley (1978) are also valuable resources. Allen (2009) presents 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.

[4]See the Data appendix for further details.

(a) Population Levels



(b) Kernel Distributions



(c) Population Ranks

Figure 1: Population History of England: 1086-1801

As shown in panel 1.(a) England starts with a 1086 population of approximately 1.7 million and ends with a 1801 population of over 8.5 million. This growth however represents an anything-but-constant average population growth rate of 0.3 percent per annum. In fact this period is dominated by two cycles of growth and subsequent plateau, punctuated by the worst epidemic in recorded human history. 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 1290 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. Three additional outbreaks occurred during the next fifty 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 in the 1600s.

These first three years of our sample have deeper connections that are only partially revealed by panel 1.(b). In panel 1.(b) the kernel densities appear to be quite similar for {1086, 1290, 1377}, but quite different for {1600,1800}. Given the similarities across initial periods it is tempting to ascribe them to the workings of slow and geographically neutral technological progress gradually relaxing the constraints of an Organic economy - a path of progress that was only punctuated, but not halted, by the Black Death. While this interpretation is appealing, if we recenter these initial distributions by netting out mean differences in population densities across the years we will reject the null that the 1086 and 1290 (recentered) distributions are in fact identical using a Kolmogrov-Smirnoff test. The reason for the rejection is simple: the population level and its geographic distribution in 1086 also reflects the lasting impact of the Norman conquest in 1066. Soon after William's quick victory at Hastings in the South in 1066, another claimant to the throne arose in the North (Edgar Aetheling). Over the next several years William was forced to put down a series of revolts in northern England which he did very violently and with great loss of life. This harsh retribution is still apparent years later as shown by the lower tail of the 1086 population distribution. To confirm we exclude any one of the four most Northernly counties bordering Scotland - Northumberland, Cumberland, Westmoreland or Durham - to find we can no longer reject the null hypothesis that the (recentered) distributions are the same in both 1086 and 1290 at any reasonable level of significance. For example excluding Northumberland produces a p-value of 0.147 using the Kolmogorov-Smirnov test for equality

of distributions. If we look deeper at either the raw correlations across county population densities or their rank correlations across these two periods, we again find fairly strong evidence of permanence. The raw correlation of population densities across the initial two periods, again without Northumberland, is 0.8; the rank correlation is 0.7.[5] These same results hold once we expand our sample to include 1377. Using the same K-S distributional test as previously we cannot reject the hypothesis that the (re-centered) distributions for 1290 and 1377 are in fact the same (with no exclusions); as is 1086 and 1377 when we exclude any one northern county from the test. Similarly, using the full set of counties in 1290 and 1377, we find the raw and rank correlations across are now .81 and .83.

It appears that the first three periods feature a nation recovering after the conquest with wasted counties catching up while others grew slowly towards their Malthusian limits. While these limits themselves may have been gradually relaxed by a process of slow technological progress, even the most lethal epidemic in human history appears to have had little impact on the geographic distribution of the population.

The next two observations reveal a very different England emerging. The distributions for 1600 and 1800 feature a long right tail driven by few observations and perhaps a narrowing of the distribution among all others. Our earlier supposition of ongoing and geographically neutral technological progress becomes completely untenable; again employing a K-S test on the recentered distributions we find 1377 very different from either 1600 or 1801. Although it is not apparent from Panel 1.(b), the 1600 to 1801 period did not exhibit constant population growth.[6] 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 (but with little loss of life) and almost forty years of intermittent English civil war ended with the Glorious Revolution in 1688. For almost the entire 17th century England is actively at war with either Scotland, France, or Spain. These and other yet-to-be understood forces produced relatively slow aggregate population growth for almost 100 years prior to the 1730s. Thereafter population growth rose sharply.

Finally as panel 1.(c) shows not only was growth in this latter period highly concentrated in some areas, but it produced a dramatic reshuffling of the population. Extreme permanence in the population distribution over the two periods would be reflected in observations plotted

[5]Further evidence that these wasted counties were in the process of catching up can be provided by a simple convergence regression or plot of subsequent population growth rates on levels in 1086. The differential equation governing population growth in the Isolated State case can easily be log-linearized and then manipulated to generate a convergence style regression with log population growth rates related to a common to counties constant and county specific initial log population densities. A plot of growth rates versus levels along these lines is provided in our earlier discussion paper. See Moreno-Cruz and Taylor (2012).

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

near or at the 45 degree line; but instead the plot shows a radical restructuring had occurred.

These data present three challenges to any researcher trying to understand them. First, and most simply, what drove the increase in population (and therefore density) over the period. At the start of our sample England contained 1,7 million; at the end 8.5 million. Second, why are there two quite different regimes: an early initial period where population growth appear almost balanced across counties, and what appears to be a radically different later period where population growth is highly concentrated across counties. Third, growth was not only concentrated in some counties now more than others, it was concentrated in counties that were lagging and previously backward.

3 Access to Energy and Agglomeration

As a first step in understanding how access to energy may affect the size and location of economic activity we now develop a simple spatial model. The model provides an explicit link between a region or county's access to energy resources and the resulting consumption per capita its inhabitants enjoy. It shows how more power dense environments can provide higher consumption per capita; how access to rivers or roads magnifies these advantages; and how the power density of resources affects the incentive to invest in transport improvements. These results provide us with part of the means to address our three questions.

3.1 The Energy-Economy Model

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.

3.2 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:

$$u = u(C), \quad u' > 0, \quad u'' < 0 \tag{1}$$

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.

$$Y \equiv \left[\int_0^n x_i^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1 \tag{2}$$

where each of the x_i are differentiated intermediate goods produced by a single firm and n is the mass of varieties available to consumers. Together these intermediates provide the service flow represented by the final output Y.[7]

It is important to our analysis that the mass 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 mass 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 intermediate at level x_i requires a_i units of activity. That is:

$$a_i = \alpha + \beta x_i \qquad \forall i \tag{3}$$

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:

$$a_i = f(l_i, W_i) \qquad \forall i \tag{4}$$

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

We take (4) 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.

3.3 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. 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.

As is well known, the mark up rule and zero profit condition of the monopolistic competition framework allow us to 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:

$$x_i^* = \frac{\alpha}{\beta}(\sigma - 1) \qquad a_i^* = \alpha\sigma \tag{5}$$
$$l_i^* = \alpha\sigma\partial c(w, p^W)/\partial w \qquad W_i^* = \alpha\sigma\partial c(w, p^W)/\partial p^W$$

where $c(w, p^W)$ is the unit cost function dual to (4). Since all firms are identical, total energy demanded (for production purposes) is just n times firms specific demand W_i^*.

The mass of firms, n, is endogenous and depends on the overall market size as indexed by L. To solve for the mass 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 mass of firms:

$$n = L/l_i = \frac{L}{a_i^* z c(w, p^W)/w} = \left(\frac{1-z}{z}\right)^{1-z} \left(\frac{L}{a_i^*}\right) \left(\frac{w}{p^W}\right)^{1-z} \tag{6}$$

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:

$$W^D = nW_i = L\left(\frac{1-z}{z}\right)\left(\frac{w}{p^W}\right) \tag{7}$$

The properties of demand follow directly from our assumptions. Constant returns in activity implies an increase in market size L, raises demand proportionately. Constant factor shares implies a unitary price elasticity of demand.

3.4 The Supply of Energy

Assume atomistic energy suppliers are distributed uniformly in the two dimensional space, each entitled to one squared meter of land and energy resources with power density Δ. Energy suppliers decide whether or not to take the energy they produce to the core. To move their energy to the core, suppliers must spend reduces in transportation. Assume c/Δ are the energy costs of moving resources representing one Watt of power, just one meter. A supplier located a distance r from the core will have to incur energy costs cr. If the profits from bringing their energy to the core are positive, they will do so. Otherwise, the energy they could provide will be stranded. The energy supplier's profits are given by

$$\Pi = p^W \Delta - [p^W + p^C]cr \tag{8}$$

where p^C is the price of the final output and p^W is the price of power delivered at the core.

There are three assumption reflected in (8). By construction, the energy producer supplies Δ Watts of power at p^W but buys back from itself cr watts to pay for transportation back to the core.[8] By making power requirements proportional to distance and power collected we have stayed true to the physics of transport highlighted in the Only Energy Model of Moreno-Cruz and Taylor (2013).

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

[8]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.

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 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 above zero. To ensure the energy input needed per Watt of power collected can never approach zero even in the limit, we need a production function for transport with an elasticity of substitution between energy and conventional inputs of less than one. Within this class of functions, the Leontief specification delivers the simplest results. 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, we can calculate total energy collected. First, we identify the set of marginal energy resources that provide zero profits to the owner. Denote by R^* the distance these marginal resources are from the core. At this margin, the profits provided by one Watts of power are fully expended in costly transportation to the core; that is, R^* must satisfy:

$$p^W - [p^W + p^C]\frac{c}{\Delta}R^* = 0 \text{ or } R^* = \frac{p^W}{p^W + p^C}\frac{\Delta}{c} \tag{9}$$

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 marginal units are at a distance $R^* = s\Delta/c$. Because resources are distributed uniformly across space, total energy collected will come from a circular zone with the core at its center and radius R^*. Total energy collected is the given by $W^* = \pi\Delta[R^*]^2 = \pi\Delta^3 s^2/c^2$. To find net power delivered, we need to subtract energy used in transportation. This net power supplied comes from adding up, what we will call, "energy rents." These rents are the excess of energy collection over transport; i.e. $\Delta - cr$ at all distances $r \leq R^*$ from the core. They are positive for exploitable resources, and zero for stranded resources. To add them we use a two step procedure. Along any ray from the core, there are Δ Watts of power every meter and transporting these resources to the core yields a density of $[\Delta - cr]$ net Watts of delivered power. The first step is to add up these resources along our ray over all distances less than R^*. The second step is to accumulate these quantities by sweeping across the 2π radians of our circular exploitation zone. By

doing so we obtain net power supply to the core as the sum of all energy rents:

$$W^S = \int_0^{2\pi} \int_0^{R^*} v[\Delta - c \cdot v] dv d\varphi = 2\pi \int_0^{R^*} v[\Delta - c \cdot v] dv = \frac{\pi \Delta^3 s^2}{c^2} \left[1 - \frac{2}{3} s \right] \tag{10}$$

where we note $s \equiv s(p^W/w, L)$ from (13).

3.4.1 Characteristics of Supply

Two characteristics of energy supply feature prominently in much of what follows. The first is simply that even though the potential supply of energy is infinite, it is effectively bounded by a physical limit. The reason for the upper bound on power supply can now be explained straightforwardly. As energy prices rise, more power is naturally supplied but this is sourced from more distant (read more difficult) sources. More distant sources carry with them larger transport costs and this pushes the share of energy costs s in transport costs towards their maximum of one. When s approaches one, marginal energy resources take as much energy to collect and transport as they provide - implying their contribution to net power supply is zero. At this point, it doesn't matter how valuable another unit of energy is - net supply becomes perfectly inelastic.

The implication for us is that even though a city or urban area is in some sense small in terms of the surrounding forest or farmland - our core here is after all a point in a limitless plane of energy resources - physics alone will at some point limit the exploitation zone that can fuel this city. Even a small urban center in a huge country may reach its peak draw on neighboring energy resources long before they are exhausted.

The second feature of note appears, at first blush, to be quite specific to our circular formulation. It is the fact that total power is a cubic function of power density. As we will show subsequently, it arises from the spatial context and not the particular form of our exploitation zone. The intuition is instructive. Consider the thought exercise of increasing the power density of neighboring resources while holding energy prices constant. If we do so, but leave the area of exploitation fixed; then supplied power will rise proportionately with power density; i.e. appear with power 1 in an expression like (**??**) because $W = \Delta Area$. But a higher power density also implies the marginal cost of exploitation falls. With lower marginal costs, exploitation rises and the extensive margin of exploitation moves outwards. Since area is proportional to the square of radius, total power rises with the square of power density. Adding up over both margins implies supply is a cubic of power density.

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

$$p^C = c(p_1, p_2, ...p_n) = \left[\int_0^n p_i^{1-\sigma} di\right]^{\frac{1}{1-\sigma}} \tag{11}$$

Since all firms are identical, we can substitute and rearrange to find:

$$p^C = \left[\int_0^n p_i^{1-\sigma} di\right]^{\frac{1}{1-\sigma}} = \frac{\beta\sigma}{\sigma - 1} c(w, p^W) n^{\frac{1}{1-\sigma}} \tag{12}$$

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 (6) to substitute for the number of intermediate goods to find:

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

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

$$\frac{dW^S}{d(p^W/w)} > 0 \Longleftrightarrow \sigma > 1/z \tag{14}$$

The definition of s and inspection of (13) signs the derivative.

Somewhat surprisingly, we need to impose a parameter restriction in order to ensure the supply curve for power is positively sloped. The reason is simple: cheap energy creates more energy converters; more converters raises productivity; higher productivity lowers the costs of energy exploitation, and this in turn makes energy cheap. With these positive feedbacks built into the model, we have unbounded increasing returns at the aggregate level.[9] Since unbounded increasing returns is an unattractive feature of any model, in Moreno-Cruz and Taylor (2012) we bound increasing returns with a simple parameter restriction. Bounding increasing returns ensures energy supply is always upward sloping and combining this with

[9] By using (5) we can write final output solely as a function of the mass of firms producing intermediate goods and parameters. Going further we can use (6) and (7) to write the mass of intermediates as a constant returns function of labor input and watts delivered to the core. Putting these results together we find 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: $Y = BL^aW^b$ *with* $a + b > 1$ where $a = \frac{z\sigma}{\sigma - 1}$, $b = \frac{[1-z]\sigma}{\sigma - 1}$, $B > 0$. To bound increasing returns we assume both a and b are less than one.

our energy demand produces a unique general equilibrium.[10]

But even with bounded IRS the model admits two cases: an Almost Neoclassical case where increasing returns are weak and there are no incentives for agglomeration; and an Increasing Returns case where increasing returns are stronger and incentives for agglomeration exist. While these two cases deliver similar results in a variety of settings, in the interest of space we will focus entirely on the Increasing returns case which obtains when $\sigma < 3/[1-z]$. We assume this condition holds throughout.

3.5 Consumption and Output per person

Our next two results follow most directly from supply side constraints. First, given our discussion of supply behavior it may come as no surprise that rising energy costs must eventually overwhelm the benefits of agglomeration. As a result we find:

Proposition 1 *Consumption per capita is single peaked in market size.*

Proof: see Web Appendix.

And given the importance of power density to energy supply we find:

Proposition 2 *Consumption per capita rises with power density. The peak level of Consumption per capita occurs at a larger market size when the power density of available resources is greater.*

Proof See Web Appendix.

Proposition 1 confirms what we may have already suspected: the interaction of IRS in the energy system with diminishing returns created by our spatial structure produces a singular peak in consumption per capita. Proposition 2 then tells us this peak is not only higher, but at a larger market size when resources are more power dense. These results are of course critical in that they establish a link between the power density of available resources and incentives for agglomeration.

We have thus far assumed a very diffuse transportation system: every energy producer makes their own way to the core. While this may be a reasonable representation for some farmers who bring their produce to market towns over hill and dale, in other cases they could avail themselves of nearby roads, rivers, and late in our period - canals.

[10] We prove this in the web appendix.

3.5.1 Roads, rivers and canals.

We now assume any given energy 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 $\rho < 1$. That is, while the costs of transportation by land are equal to c, a road's coefficient of friction is ρc in both directions whereas when traveling with the current a river's is also ρc but against it c/ρ. By this assumption, river transport is only useful when you are an energy producer upstream; whereas road transport reduces frictions in two directions and not one.

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. Suppliers decide how to travel to the core, that is what fraction of the trip is done by river or road and which distance is done by land. The road allows for energy producers farther from the core to profitably bring energy to it, but there are still limits to how far suppliers are willing to take their energy. Panel 2(a) shows how the exploitation zone is altered by the presence of a road. In Panel 2(a) the angle $\bar{\theta} = \arccos(\rho)$ shows the limiting angle that in combination with the limiting radius $\bar{\iota}$ determine the profitable zone. Suppliers inside this zone will bring their energy to the city, suppliers outside this zone will not. Comparing Panel 2(a) to Panel 2(b) shows the gains introduced by a road relative to the case in which there is no low cost transportation alternative. Moreover, as the transportation infrastructure becomes more competitive, that is ρ decreases, the critical angle and the limiting radius both increase. The case of a river is similar, but only half the gains are accrued because the river is one-directional.

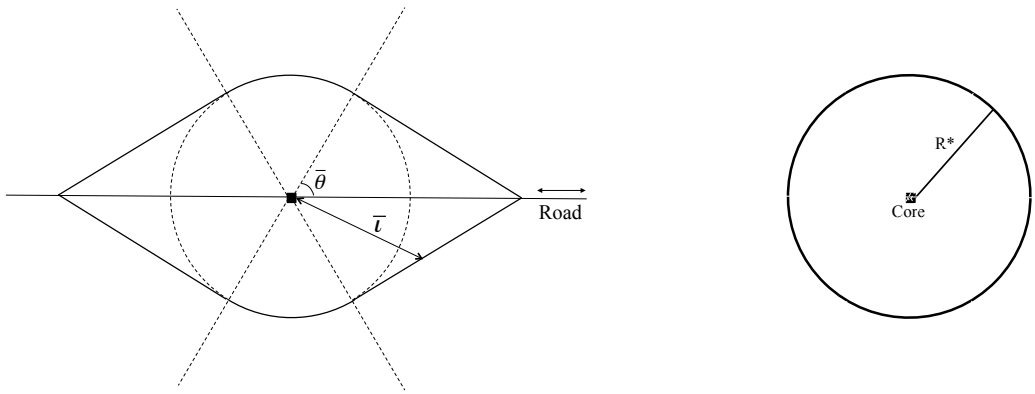
We can now calculate the energy supplied to the city when a river is available:[11]

$$W^S = \frac{\Delta^3}{c^2} s^2 \left(1 - \frac{2}{3}s\right) g(\rho) \tag{15}$$

$$g(\rho) = \pi - \bar{\theta} + \int_0^{\bar{\theta}} \left((1-\rho^2)^{1/2} \sin\theta + \rho\cos\theta\right)^{-2} d\theta \geqslant 0 \tag{16}$$

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 $g(\rho) = \pi$ since then $\bar{\theta} = 0$ and equation (15) reduces to equation (**??**).

[11] The derivation is similar to the one we show in Moreno-Cruz and Taylor (2013). Readers are directed there for more details.



(a) With Road (b) Without Road

Figure 2: Exploitation Zones

Inspection of equation (15) reveals that the role of improved transportation is identical to being granted a more dense resource base in terms of energy supplied to the core. River or roads multiply by $g(\rho)$ the power density of available resources to an extent determined by its capacity for reducing transport costs as reflected in ρ. This allows us to extend all our earlier results on power density to results concerning the existence of rivers or roads for transport. For example, it tells us 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 and raises the equilibrium energy supplied; access to a river or improved road transportation raises both consumption per capita at all population levels; and access also raises the market size at which consumption per capita peaks. Many of these results of course strengthen the forces for agglomeration in such a region.

3.5.2 Endogenous Infrastructure

While the population in England benefitted from an existing Roman road system, and extensive system of inland rivers both rivers and roads were improved by literally hundreds of Navigation and Turnpike acts rewarding and encouraging the private provision of improvements. Late in our period canal construction boomed and newly constructed canals together with road and river improvements lowered transport costs dramatically after 1700.

To see how these options may work, we assume a single agent choses how much to improve the infrastructure in order to maximize the flow of profits from energy sales.[12] The problem

[12]Since Navigation, Turnpike and Canal acts set up single corporations to both provide the service and charge tolls and tariffs we investigate the strength of these incentives by assuming a single agent can improve transportation.

for this central planer is:

$$\max_{\rho} \Pi^N = p^W W^S(\rho) - (p^W + p^C)h(\rho) \tag{17}$$

where $W^S(\rho)$ is given in (15) and $(p^W + p^C)h(\rho)$ are the flow costs of investing in infrastructure. For simplicity we have assumed one unit of the composite is used alongside each unit of energy spent in building and maintaining infrastructure. We assume these costs are an increasing function of the reduction in transportation costs so that $h'(\rho) < 0$ and $h''(\rho) > 0$. The first order condition that maximizes profits requires

$$\frac{\Delta^3 s^2}{c^2}\left(1 - \frac{2}{3}s\right) g'(\rho) = h'(\rho)$$

This equation implicitly defines ρ as a function of power density, Δ, and the energy share s. Using this implicit function we can ask how power density affects the optimal investment in infrastructure and the economic implications of this investment. The next proposition summarizes these results:

Proposition 3 *Higher power density implies greater infrastructure investment, lower energy prices and greater supply.*

Proof: See Web Appendix.

3.6 Population dynamics

To examine how the location of resources with different power density may affect observed population densities we need to add a process governing population growth and, eventually, a decision rule to sort agents across locations. It seems natural to adopt a Malthusian mechanism for population growth. It is simple, well understood, and arguably appropriate given our sample period; indeed our data comes from a time and a place where Malthus himself observed the forces he identified at work.[13] We start by examining what we refer to as an Isolated State where migration to other locations is impossible. But since villages,

[13]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).

hamlets, and market towns do not exist in isolation, we subsequently extend our analysis to allow migrants to sort across locations on the basis of expected lifetime consumption.

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. 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:

$$\begin{aligned} \eta(c) &= \eta_0 + \eta_1 c \\ \delta(c, L) &= \delta_0 - \delta_1 c + \delta_2 L \end{aligned} \tag{18}$$

Hence the population responds according to:[14]

$$\dot{L} = L[[\eta_0 - \delta_0] + [\eta_1 + \delta_1]c - \delta_2 L] \tag{19}$$

In the standard Malthusian set up $\delta_2 = 0$, and subsistence consumption is a constant independent of densities. Define the Zero Population Growth or $Z - line$ as the combination of consumption per person and population consistent with zero population growth. Setting (19) to zero the $Z - line$ can be written as:

$$c_Z(L) = \frac{\delta_0 - \eta_0}{\delta_1 + \eta_1} + \frac{\delta_2}{\delta_1 + \eta_1} L \tag{20}$$

where $c_Z(L)$ is the per capita consumption needed to maintain zero population growth in a city with population size L.

The addition of endogenous population growth does more than just tie population densities to the productivity of available energy resources, it provides another mechanism to magnify differences across regions; and since populations typically respond relatively slowly it means the knock on effects of large shocks can be literally felt for centuries. As discussed previously, positive shocks to energy supply produced a virtuous circle where new interme-

[14] 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.

diates were introduced, efficiency in the energy system rose, transport costs fell which in turn raises supply further. An example may help clarify. The introduction of a new more productive crop variety represents an increase in power density in a purely Organic economy. This raises crop yields, and raises the incentive to introduce new intermediates such as better horse collars, ploughs, or milling machines. These improvements then raise the productivity of the energy system as a whole, and generate lower cost calories (or energy services) for humans. Less costly energy services means makes it easier to reap even greater yields reinforcing the initial shock. This is the positive feedback loop making agglomeration possible.

When populations respond to shocks, a further dynamic is set in motion. Carrying forward our example, lower cost calories for humans represent increases in consumption per capita, but this then stimulates population growth. Population growth then raises the market size for horse collars, ploughs, milling machines or wagons and creates its own new intermediates. These in turn raise supply further and again reinforce the initial shock. Demographic changes tend to be quite persistent and slow moving, and hence our Malthusian mechanism may propogate shocks into the very far distant future.

4 History through the lens of theory

With our theory now fully in place it is time to reconsider England's population history through the lens of our theory. To make any headway, several simplifications are in order. For example, we assume that in addition to the forces identified in our theory exogenous (and unmodelled) technological progress drives long run economy wide changes in population levels and we focus almost exclusively on our model's steady state outcomes.

Without doing too much violence to the historical record we now divide our almost 800 year sample period into two. We refer to the first period as the Organic Economy period, since during this time the vast majority of energy used by households and industry was provided by renewable sources. We date it from 1086 to 1377. We refer to the second period as the Mineral Economy period and date it from 1377 to 1801.

Now revisit the initial two snapots snapshots {1086,1290} shown in panels 1.(a) and 1.(b). These are years untouched by the plague. Our interpretation of this data is straightforward. With little migration across counties and diffuse natural blessings across England, the geographic distribution of the population should likewise be diffuse; and the upward shift in the distribution over time reflects nothing more than the workings of slow but steady technological progress relaxing the bonds of an organic economy. To make these connections precise we first divide our 39 counties into three groups, and then place them in a Malthusian

steady state as shown in Figure 3 below. As shown one group is labelled M, and this group contains just the county of Middlesex (which contains London). A second group is the set of 24 counties without coal deposits. We label these NC for no coal. The final group is the set of 14 counties with coal, labelled C.

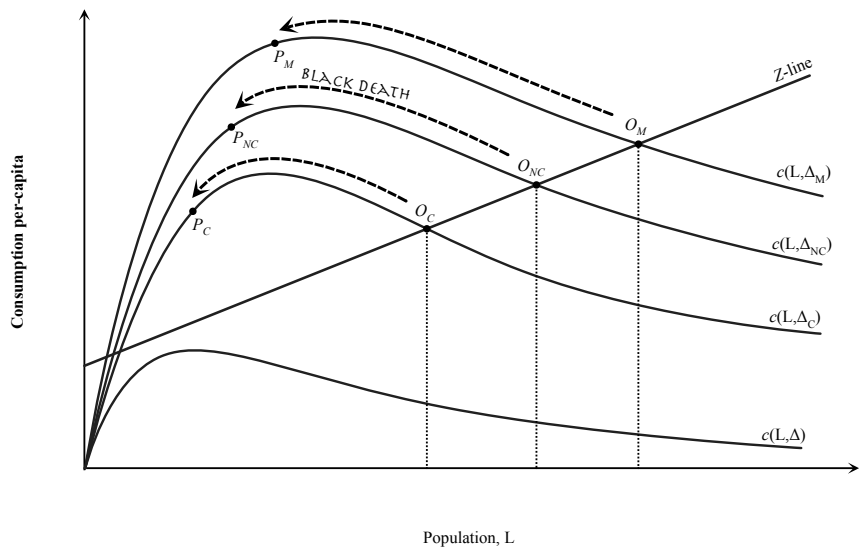


Figure 3: The Organic Economy and Black Death

The three $c(L;\Delta)$ lines drawn in the figure show how consumption per capita varies with L in regions that differ in the power density Δ of available resources. If available resources are not very power dense, then the $c\,(L;\Delta)$ curve lies everywhere below the $Z-line$ and a vacant landscape results. When resources do offer sufficient power density, then two possible steady states emerge, but only the more dense (the rightmost intersections) are stable under our Malthusian mechanism. Therefore corresponding stable steady states occur at the intersections O_M, O_{NC}, and O_C where the county specific consumption schedules intersect the zero population growth Z-line as shown in Figure 3.

In constructing Figure 3 we have made several assumptions. The first is the ordering of groups. 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. London's location on the Thames (which was 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.[15] A low transport cost alternative such as the Thames magnifies the power density of surrounding resources and should raise the density of population centers as well. In 1086 Middlesex county contains 54 people per square kilometer, but importantly this is not far in advance of the most dense NC counties (Suffolk at 31 and Norfolk at 28 persons per square kilometer). Whatever advantage the Thames and the existing Roman Road system granted to London, it was limited in an organic economy setting.[16]

The second assumption is that the relevant (i.e. stable) steady states occur on the downward sloping segments of the consumption per capita schedules. This reflects our belief, and that of significant scholarship in this area, that in some sense England was already full at this time and diminishing returns were beginning to be felt in agriculture. Our placement of the steady state past the peaks, implies further population growth would lower real wages however measured. The third assumption is zero or low labor mobility. Since the institution of serfdom and demesne farming were still very much in play, labor mobility was low. This assumption appears uncontroversial.

It is straight forward to see that the intersection of the Z-Line with the consumption per-capita curve $c(L, \Delta_M)$ occurs at a higher level of consumption and for larger population sizes than the intersection with the $c(L, \Delta_{NC})$ and $c(L, \Delta_C)$ curves. Hence, higher power density implies higher consumption per-capita and larger population sizes. This discussion leads us to the core result of our paper:

Proposition 4 ***The density-creates-density hypothesis:*** *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 has $L = 0$; (iii) if $\Delta > \Delta^{crit}$ then there are three steady states: steady state zero has zero population, $L_0 = 0$, has a vacant landscape, and is stable; steady state one has a positive population, $L_1 > 0$, is unstable, but has a relatively low population; and steady state three has a positive population, is stable, and has a larger population $L_2 > L_1 > 0$. The population level in steady state three is increasing in power density Δ.*

While our steady state configuration may capture important cross sectional aspects of the {1086,1290} distributions shown in Panel 1.(a), it says nothing about the shift in the distribution over time. Since power density is fixed over time, our theory suggests that the

[15] 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.

[16] 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. Consulting maps will show that London is the hub of the Roman system with "almost all roads leading to London".

population distributions in 1086 and 1290 should be almost identical once we allow for at most trend growth tied to technological progress. As we have already shown, this is true subject to an allowance for the wasting of northern counties by William.

The happy state of affairs in England circa 1300 soon changed. Early in the century a great famine struck Europe precipitated by repeated crop failures, and in 1348 the Black Death hits and lowers populations everywhere by almost one half. In terms of our theory this moves the system immediately to the positions we show in Figure 3 as the post plague outcomes at P_M, P_{NC}, and P_C. Real wages rise dramatically as a result of this decline, and soon after we observe the 1377 populations in our data.[17] The plague lowered the national population tremendously, but as the density for 1377 suggests the incidence of death from the plague was more severe in already dense counties. Despite this complication, our previous results show a strong similarity across these periods in the geographic distribution of population.

The Organic Period writ large features 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 power densities are constant over time. Even the most lethal epidemic in human history appears to have had little impact on the geographic distribution of the population.

4.0.1 The Trigger

Although the plague may not have altered the geographic distribution of population very much, it still plays two very important roles. Its first role is conventional: the plague provides a very large shock to the existing distribution of population. This shock may have loosened the grip that previous centers of agglomeration had on the economy allowing the post-plague period to differ from the pre-plague period. Using a major shock to explore the strength of agglomeration forces was first pioneered in seminal work by Davis and Weinstein (2002). This methodology has subsequently found wide use within the empirical literature in the new economic geography. While the plague was certainly a shock of the first order, it also plays a second and more important role in our analysis. Its second role arises from the very significant labor shortage caused by the depopulation. This shortage strengthened the hand of the laboring classes and, together with institutional changes already underway, led to the breakdown of the villeinage system over the next century and a half. Demesne land becomes smaller, more land is contracted out, and laborers gain far more freedom to move. While historians continue to debate whether the Black Death was the key driver in

[17] In our theory it is possible for real wages to fall if the population decline is severe enough to move us far past the peak of the consumption per person schedule and close to zero output in a location.

these changes or whether it was merely one of the forces at play, this debate matters little here because the facts are well established.[18] The breakdown of serfdom, the change in land practices, and the growth in mobility over this period are well documented. These institutional changes now imply laborers are far more able to migrate within and across counties in search of better opportunities. In terms of our theory, this greater freedom in the centuries following the Black Death means that England does not return to anything like its earlier 1290 or 1377 configuration of Isolated States; it returns instead to a world where labor is free to move both within and across counties.

4.1 The Mineral Economy: Within and Across County Migration

Consider the next two snapshot years {1600, 1800} in Panel 1.(b). These years show a quite different pattern emerges as we enter what historians often refer to as the the Early Modern Era. They show a radically changing and changed England. By the year 1600 two features stand out. The first is what appears to be 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 8 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.

These aggregate features are largely driven by the growth of London. London rose from 50,000 in 1520 to perhaps 200,000 in 1600. Growth the following century was also rapid. By 1700 London had over half a million inhabitants; it represented over 10% of the English population; and was by then Europe's largest city. Growth slowed closer to national averages over the next century but London was 900,000 by 1800.

To understand how flows of migration affects city size, consider Figure 4 (where we have cut the figure to concentrate on the relevant steady states). The figure also shows shows the Organic Economy configuration previously discussed plus a set of outcomes $\{M_M, M_{NC}, M_C\}$ associated with a free migration steady state.

Now consider the migration decision. Although agents may be inclined to move to the most productive region since consumption per capita is highest there, so too are death rates. To see why recall that death rates equal birth rates along the Z-line, and birth rates rise with consumption per person. This implies that all upward movements along the Z-line

[18] See the very nice article by Hatcher (1994) where he reinvigorates this debate by arguing for a pivotal role of the Black Death.

correspond to situations where agents face a higher death rate. The simplest way to interpret the Malthusian model is that everyone (man, woman, child, king or queen) faces the same instantaneous probability of death given by $\delta(c, L)$. If we accept this fiction and assume a zero discount rate for simplicity, then the expected lifetime consumption at time τ for any individual living in city i is given by:

$$e = \int_{\tau}^{\infty} c_i \exp^{-\{t-\tau\}\delta(c,L)} \, dt \tag{21}$$

$$e = c_i/\delta_i \tag{22}$$

where the last equality holds in steady state. This construct now allows us to compare expected lifetime consumption in the various locations.

It proves useful to construct a line representing all combinations of $\{c, L\}$ that yield the same expected lifetime consumption per capita:

$$c_E(e, L) = \left[\frac{e}{1 + \delta_1 e}\right] [\delta_0 + \delta_2 L] \tag{23}$$

which we will refer to as the e-line. Several properties of e-line follow directly. Holding expected lifetime consumption constant, the e-line is just a linear relationship between population size and consumption per capita. Points above the e-line 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. Finally, although it is not obvious, the e-line going through a steady state with migration is flatter than the associated Z-line going through the same steady state.

It is immediate then that points labelled $\{M_M, M_{NC}, M_C\}$ offer the same same expected lifetime consumption e. Points M_C and M_{NC} are however above the Z-line: their consumption levels exceed those necessary to generate zero population growth and, in the absence of migration, their population levels would necessarily rise. Indeed - without migration - its population would increase until the city reached the no-migration steady states at O_C and O_{NC}. An alternate possibility is that this constant excess flow of population could migrate instead to county M thereby holding their current population and its current expected lifetime consumption e. Now consider county M. Residents in this county would have the same expected lifetime consumption e, if it was to grow and reach point M_M. But of course point M_M is below the Z-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 towards its no-migration steady state at O_M. The now obvious alternative is for city M to remain at M_M and receive a constant flow of population from

counties C and NC. If the flows from city C and NC to M balance we have identified a steady state of the system with migration exhibiting expected lifetime consumption e.[19]

In theory, Figure 4 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 transport cost alternatives like rivers or roads while other cities in their county do not. Therefore, the same analysis could apply to migration between cities in the same county or in different counties.

The ability to migrate has several implications. First, real wages are higher than otherwise in the sending groups and lower than otherwise in the receiving group. We see at least the beginnings of convergence in real wages across the country. Second, average productivity in the sending groups is higher because Malthusian constraints are relaxed; it is lower than otherwise in the receiving group. Overall productivity nation wide may well be higher since migrants move to the most productive location. Third, we have drawn the figure with both the C and NC group as net suppliers of migrants, while Middlesex is a net demander. That there was large scale migration to London during this period is not at issue; the evidence for real wage convergence is however more mixed. But the concentration of population in London has an interesting implication: the migration steady state we identify by $\{M_M, M_{NC}, M_C\}$ cannot be right; in fact, as we argue below its very construction contains a contradiction.

4.1.1 The Transition

If the new configuration representing the migration steady state at M_M, M_{NC}, and M_C was right, it would have to be consistent with a London (and hence Middlesex) population of 200,000 in 1600. But 200,000 Londoners is far beyond the estimates we have of London's ability to feed and fuel its populace in an Organic economy setting.[20] Work by economic historians mapping out both London's fuel and food exploitation zones, suggests a London population of 80,000 already strained natural resources; a population well over 100,000 would have brought sharp consequences, but by the end of the 16th century London housed 200,000 people. And therein lies the rub: we are forced to conclude that the concentration of energy

[19] Since consumption per capita schedules are continuous in power density we can always construct a steady state of this type; we are not claiming this steady state is unique.

[20] See for example Galloway, et al. (1996) on fuel wood requirements and Galloway, et. al (1991) on food requirements of London.

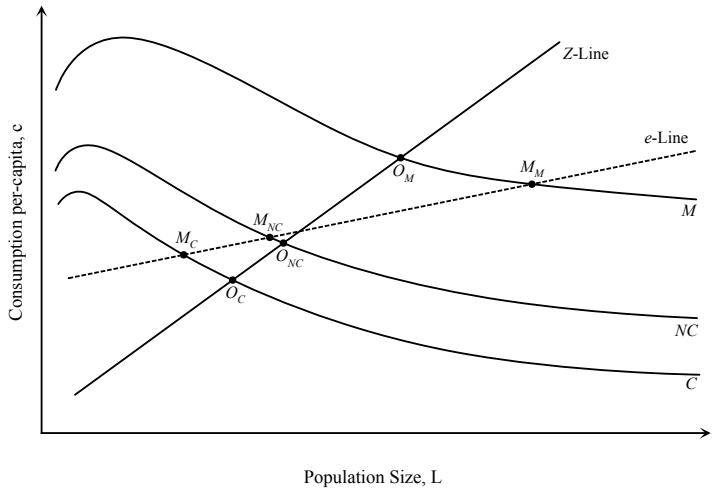


Figure 4: The Organic Economy with Migration

demand in London, created by migration, brought about its energy transition to coal in the 16th century. In short, London's growth brought about large economic incentives to solve the relatively small but numerous technical problems limiting the use of coal in domestic heating, and in other industrial processes where the fumes or chemical deposition arising from open pit burning of coal fouled the production process.

London's incipient energy crisis was solved by the transition to coal in a significant set of uses. While this view is not universally shared, the transition to a denser energy (mix) would now shift the consumption per person schedule for Middlesex upwards (not shown), raise expected lifetime consumption in the county, and perhaps even accelerate migration to the capital. London shakes off the bonds of the Organic Economy and enters the Mineral Economy period. And while the Thames is a constant advantage to London, its impact is all the greater when the energy we source is more power dense. London, already blessed by geography, is now blessed by the geographic location of vast north east coal reserves which are easily shipped coast wise and up the Thames.

4.1.2 The Treatment

What happens next is shown in Figure 5. Ignore for the moment the dashed line. As shown there is active migration from both NC and C to M given the location of the $Z - line$ also drawn. This initial situation would represent sometime in the 16th century when London has already moved significantly to coal. At this point, the same migration flows that drove the energy transition centered in London, now provide a mechanism for the diffusion of this new information regarding coal uses throughout the country. These novel ways to use coal are perhaps interesting to residents of NC counties, but not very useful since the overland

costs of transporting coal were then prohibitive. In C counties, this information is both useful and productive. Cheap coal and information from London combine to produce a much faster and relatively cheap transition to a Mineral based economy in all C counties. In terms of our theory, the consumption per person schedule for all the C counties shifts upwards and leapfrogs that of the NC counties. This knock-on effect of London's transition is shown by the dashed curve labelled C^* in the figure.

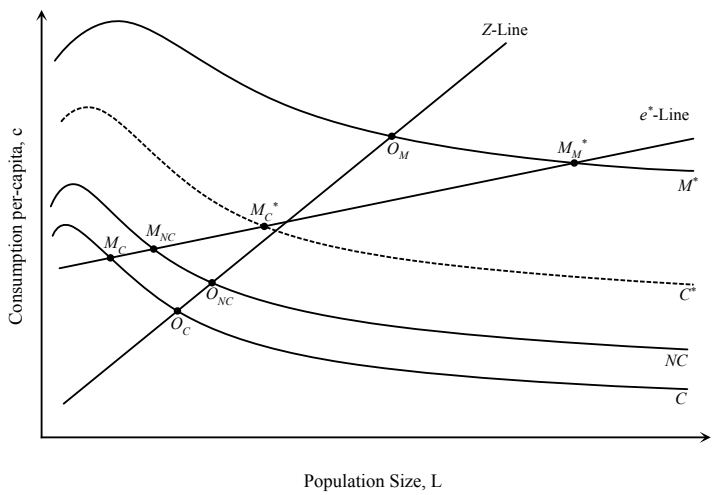


Figure 5: The Mineral Economy

The resulting change in economic opportunities in C counties then redistributes population towards the C counties by both raising their natural growth rates and lowering any migration they sent to London. The remaining NC counties do not benefit immediately from coal, and fall behind in relative terms. These two very different responses create a tremendous reordering of the rank in population densities across counties while allowing population densities themselves to approach unheard of levels. By the 18th century growth of London and Middlesex slows as it was more difficult to attract migrants from the now industrializing coal areas. And soon after we obtain the observations for 1800.

Several facts support this view. First, migration to London was a first order phenomenon. Wrigley (1967) for example constructs a simple numerical example using estimates of death and excess birth rates of the day to show how London would need to harvest a significant fraction of all births in the country to grow at its spectacular pace. His estimates led to his oft quoted claim that one in six Englishmen or women of this period spent some of their life in London. Second, we know that gross migration flows must have been much larger than net. Estimates of net flows are in the tens of thousands per year. Third, a significant portion of the migrant population - apprentices - came for the express purpose of learning new skills and transferring them back to their home counties; that is, they came to facilitate

technology transfer. Apprentices were a very important part of the London population representing somewhere between 10 and 20% of the London (i.e. up to 40,000 individuals in London of 1600).[21] Therefore, not only was their significant migration to London, it was two way migration, and a significant fraction of these migrants were just the right sort to bring new technical solutions developed in London home for productive use in their counties. Fourth, a variety of statistical evidence is at least consistent with the view that there was reordering of the rank in population density. As shown in the last panel of Figure 1, there was a dramatic reshuffling. The explanation from viewing history through the lens of our theory 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.

5 Empirical Implementation

5.1 Research Design

The perfect experiment evaluating our density-creates-density hypothesis would of course be the random assignment of energy resources with different power densities across otherwise identical counties who had little to no interaction via either trade or migration. Random assignment of resources with different power densities in a world of Isolated states would cause an initial supply shock to treated units, lower energy prices, raise consumption per capita and spur infrastructure investment. Over the longer term we would find population density was caused by differences in the power density of available resources. While the real world is not this neat, our research design strives to replicate at least part of this fiction, and control for those aspects we cannot manipulate.

Our discussion of the history suggests we adopt a simple difference in difference research design where changes in the population density in both NC and C counties is linked to the advent of information on how to use coal. The Organic economy period before 1600 is pre treatment {1086, 1290, 1377}; the Mineral economy period is post treatment {1600, 1800}. The impact of coal is identified by exploiting our sample's specific cross sectional (NC vs. C) and over time (Organic period vs. Mineral period) variation.

What is critical to our research design is that most of the technical problems surrounding the use of coal in large industrial and household uses were solved first in London; and that there exists a realistic way for these small incremental innovations to spread fairly uniformly throughout England.

[21] See Field (2010) for reference to primary sources supporting these numbers.

With regard to London's innovativeness we are also on firm ground as several authors have argued for similar effects. Allen (2010) for example argues that the incremental innovations needed in home building to take advantage of coal benefitted from local learning and experimentation in London and then spread throughout the country. What make London unique was the rate of urbanization that allowed for a testing laboratory of alternative chimney designs. As Robert Galloway writes: "Within a few years after the commencement of the seventeenth century the change from wood fuel to coal, for domestic purposes, was general and complete; (...) being used in 1612 in houses of the nobility, clergy, and gentry in London (...) as well as for the dressing of meat, washing, brewing, dyeing as for other requirements.'" (Galloway, 1882, pp 24-25)

We have argued that the constant flow of migrants into London - those seeking work, bright lights, or education via explicit apprenticeships were the perfect transmission mechanism for the many small innovations and local learning in London. That experience-based knowledge was the main way to transfer technologies is well documented (Epstein 1998, 2004). Craft guilds in London were at the center of this process by enforcing temporary property rights on the inventions of its members. They also allow for direct transfer of knowledge from master to apprentices by providing the institutions to support them when transitioning from one town to another (Epstein 2004). So for the inventions in London to be transferred and implemented in other regions there had to be an actual movement of apprentices and journeymen. In a recent paper by Klemp and coauthors identify that up to 1/5 of those trained in London ended up registering a life event (birth, marriage, death) in their parish of birth (See Table 5 and the subsequent discussion in Klemp et.al. 2013).

5.2 Timing

Our view of when the transition occurred is not universally shared. For example, while Hatcher (1993) claims that by 1600 coal was London's predominant fuel, Allen (2009) suggests a somewhat later date. Our argument for an earlier date relies on the logic of the argument presented above; the fact that coal exports to London per person were higher circa 1600 than they were in any future year; a myriad of other small pieces of information suggesting London merchants, home owners, and even cloth dyers may have shifted to coal by this time; and empirical evidence we will subsequently present that accords well with this timing. But we are not suggesting London moved entirely to coal by this time, and to some extent our difference with earlier authors is a matter of degree and not substance. We date the time information becomes freely available about coal's uses as 1600 onwards, but we examine whether our results are sensitive to this assumption. In addition, we provide

placebo tests by moving the treatment window backwards in time to investigate sensitivities.

5.3 Reverse Causation

We have excluded Middlesex for exactly this reason, and in fact rely on a causal relationship between the growing size of Middlesex and the transition to coal. In our description of the period we associate this growth to increased migration arising from differences in power density, but other, perhaps more important, forces were also at play. Most significantly the increasing concentration of international trade through London and the relative stability created by the eary Tudor and Stuart period. These were first order events that surely contributed to the rapid growth of London's population in the 16th century. In theory a similar population push may have caused the transition in the coal counties as well implying further problems with reverse causality. While this is certainly possible C counties were on average smaller and less dense than their NC counterparts pre-treatment, so if large populations drove a movement to coal (rather than cheap local access) we should find no significant treatment effect. Coal was used in almost all counties of England over much of this time period, but it was typically employed in a relatively few applications, was generally viewed as inferior, and unless local demand made firewood very dear - coal would see little use. Smithing may be the only example of an industrial demand where coal was superior to wood. But we have already seen that over the period in question, almost all of the growth in England was centered in London. This observation makes it very difficult to sustain a belief that a localized shortage of wood in coal counties drove their transitions.

While it is true that several counties surrounding London were likely influenced by London's growth and also had access to coal, they are also NC counties. Whatever coal inspired growth they achieved would only dampen the treatment effect we estimate.

5.4 Other Causes

It is possible that C counties were rapidly growing over the treatment period for other reasons - perhaps trade related - and those counties primarily benefitting from trade also happen to have coal. In an effort to mitigate this concern we will control for the existence of large and small ports and two different measures of existing road networks in our empirical work. To capture their time varying effects, we will interact these characteristics with a full set of time dummies to allow their impact to vary over time. As a consequence if C counties grew because they also contained ports booming from international trade, we will, in theory, already be accounting for this variation. Our measure of roads and ports will be constructed using pre-treatment measures (Roman roads, roads at the time of Domesday,

and historically important ports).

5.5 Data

We have collected data from several different sources. Our key population data is taken from recent estimates by Broadberry et al. (2011). We employ this data plus measures of county areas to generate a series for population density across counties and time. In addition we have exploited information on coal mines, coal output and coal trade from both volume I and II of Hatcher's classic treatise on the UK coal industry (1993). Data on international trade was collected from a variety of sources, as was information on the road and transport network in both Roman times and the date of Domesday. In some cases, recourse to historic maps was necessary to locate county towns, the extent and quality of coal fields, etc. A complete list of these resources and means to construct our data is presented in the data web appendix.

5.6 Baseline Specification

Our empirical implementation is fairly straightforward. We seek to compare the growth in population density over time in counties with and without coal deposits. While we treated all C coal counties as identical and all NC no coal counties as identical, they are in fact very different. Since counties differ in size, climate, topography and political organization we include county fixed effects in all specifications; similarly a varying climate (both natural and political), the Black Death, and technological progress call for us to include unrestricted time dummies in all specifications. In addition, we introduce variation in the treatment intensity across counties that is proportional to the size of exposed coal deposits. This measure of coal availability is taken from maps of the day and it is determined entirely by geology. Considering the extent of exposed deposits provides us with a further source of identification: within-coal-county, county variation. For comparison purposes we also provide estimates from a simple binary treatment reflecting the presence or absence of coal deposits in the county. Surprisingly, all fifteen coal fields in England were already active prior our treatment period and therefore issues of initial discovery are largely moot.

Our Baseline specification follows very closely that adopted in Nunn and Quian (2012). It is described by:

$$dens_{it} = \beta coal\_indicator_i \bullet I_t^{post} + \sum_{j=1086}^{1801} X_i^{'} I_{jt} \nu_j + \sum_{j=1086}^{1801} \gamma_j I_{jt} + \sum_{c=1}^{38} \rho_c I_{ci} + \epsilon_{it}$$

Table 1: Baseline Estimates With Percent Coal

	Dependent Variable: Ln Population Density				
	(1)	(2)	(3)	(4)	(5)
Percent Coal 1600-1801	0.255***	0.256***	0.262***	0.261***	0.245***
	(0.0615)	(0.0631)	(0.0654)	(0.0599)	(0.0543)
County Fixed-effect	Yes	Yes	Yes	Yes	Yes
Time Fixed-effect	Yes	Yes	Yes	Yes	Yes
Strong-Port	No	Yes	Yes	Yes	Yes
Weak-Port	No	No	Yes	Yes	Yes
Strong-Road	No	No	No	Yes	Yes
Weak-Road	No	No	No	No	Yes
Observations	190	190	190	190	190
R^2	0.836	0.837	0.842	0.850	0.887

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

where $dens_{it}$ is the log of population density in county i at time t; the *coal indicator* is either a binary indicator or the log of the percent coal field area in a county; I_t^{post} is an indicator variable taking the value one in 1600 and 1800; I_{jt} is an indicator variable taking a value one when $j = t$ and is zero otherwise; γ_j are the time fixed effects; and ρ_c the county fixed effects. The coefficient of interest is β, which is the estimated impact of the percent of coal area (or the binary indicator) on population density. We present the results from our baseline estimates in Table 1 below.

As we move across the table additional controls are added to the basic specification in column 1 that has only time and county fixed effects. Columns (2) first adds controls for strong (or important) ports in a county while weak (or less important) ports are added in (3). Both are interacted with time dummies to allow these to vary as growth or trade changes. Similarly columns (4) and (5) control for the density of the existing Roman road network (strong roads) and the entire set of roads available at the time of Domesday (weak roads). It is apparent that in all specifications, the treatment effect is positive and significant. Its magnitude varies only slightly across specifications and is about 0.24. This estimate implies a doubling of the area covered by coal in the county (an increase of 100%) would raise population density by approximately 24% - not a small amount.

Since the coverage of coal by counties varies quite significantly from over 30% coverage to less than 5% this estimate implies the movement to coal implied quite different things to different countries. One way to see this is to ask how much of the growth that occurred in a county from 1377 to 1801, was due to coal? To calculate this figure we first find the overall population growth for each county and then subtract for each county the contribution created by coal using its county specific measure of coal coverage area and the coefficient estimate from our full specification in column (5). The result from this exercise is the amount of growth that would have occurred in the absence of coal. Comparing this growth to the estimate of the growth caused by access to coal, give us the share of a county's growth that we can attribute to coal. We do this for every county and graph the result in Figure 6 below. The estimates range from a low of 15% of growth for Somerset to 40% for Derbyshire! For three counties the growth from coal is less than 25%, but for three others it is just under 40%. The remaining group fall somewhere in between, but in all cases these shares indicate coal played a very significant role in growth for coal counties.

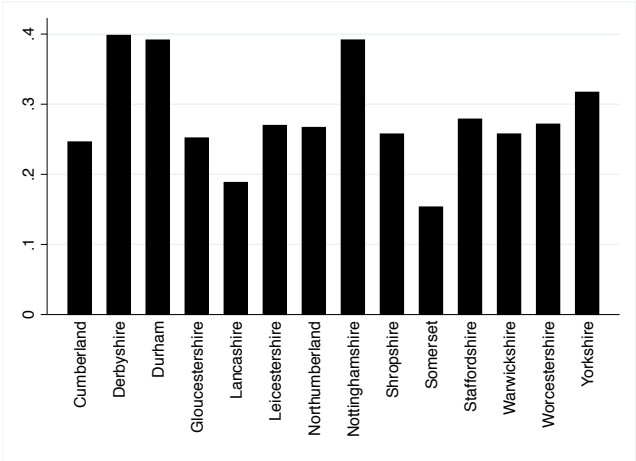


Figure 6: Share of Growth Created by Coal

Our results demonstrate that coal was important to the growth of coal counties, but was it key to the redistribution of economic activity over this period? To answer this question we present three population distributions in Figure 7 below. The top two distributions plot log population densities in 1290 and 1801 and we distinguish the 14 coal counties with dark bars. Looking just at these two panels reveals a strong correlation between having coal deposits, and moving up in the overall distribution. 11 of the 14 coal counties lie below the median in 1290; 11 of 14 lie above it in 1801; and 6 of the 10 most dense counties in 1801 are coal counties! In the last panel we have calculated a counterfactual distribution by eliminating the growth effects we attribute to coal (using table 1, column (5) estimates). This counterfactual distribution shows the role of coal deposits: all but three coal counties

now move to below median growth and all others shift down the distribution giving up their top spots; and only 2 in 10 of the most dense are now coal counties. Therefore, coal was important to all coal counties and pivotal for some.

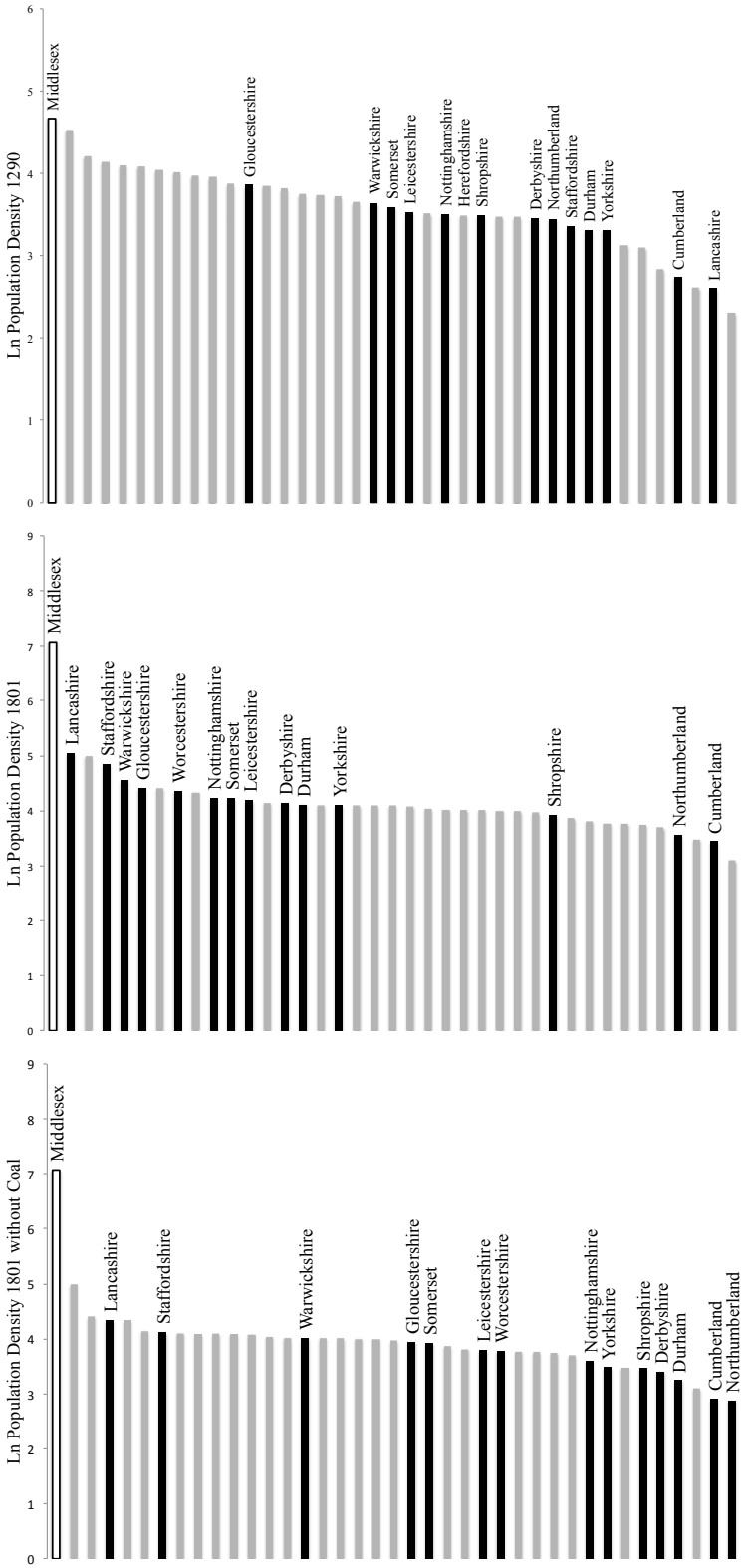


Figure 7: Coal and the Distribution of Population

5.7 Sensitivity Tests

In Table 2 below we investigate the implications of a simple binary indicator for a coal county. In this case the impact of coal is identified from just across county and across time variation. One reason we may be interested in this measure is if we think the coal coverage ratio, however constructed, is contaminated because it reflects perhaps local demand for coal. If true, this would amount to reverse causation. While we do not believe this is the case, the estimates in Table 2 are likewise consistent in their finding of a statistical significant positive impact on population density. As well, the economic significance is also large. The estimates from column (5) indicate growth on the order of 0.5 log points; indicating a coal county would all else equal be 1.65 or over one and a half times as dense than otherwise. Perhaps not surprisingly, similar results are obtained when we use the area of coal fields in the county.

Table 2: Baseline Estimates With Binary Indicator

	Dependent Variable: Ln Population Density				
	(1)	(2)	(3)	(4)	(5)
Binary Coal 1600-1801	0.511**	0.513**	0.510**	0.510**	0.498***
	(0.141)	(0.148)	(0.153)	(0.143)	(0.125)
County Fixed-effect	Yes	Yes	Yes	Yes	Yes
Time Fixed-effect	Yes	Yes	Yes	Yes	Yes
Strong-Port	No	Yes	Yes	Yes	Yes
Weak-Port	No	No	Yes	Yes	Yes
Strong-Road	No	No	No	Yes	Yes
Weak-Road	No	No	No	No	Yes
Observations	190	190	190	190	190
R^2	0.824	0.826	0.829	0.837	0.878

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Table 3 we investigate the timing of the transition. Since it is not entirely clear that our timing of events is correct, we now investigate the implications of a more flexible estimation. In this case, the post treatment dummy is replaced by a full set of time dummies. By doing we obtain an estimate for the interaction of coal county and time period by period.

Since the initial period is the base year the value of these interactions is not important but their time pattern is relevant. Table 3 contains estimations that allow for a completely flexible treatment structure. Importantly, the coefficients start relative small, fall slightly in 1377 but then start to rise sharply with over a 50% gain during this critical interval. A visual representation of these estimates together with their 95% confidence interval is shown in Figure 8 below. The figure suggests a distinctly rising pattern of impact after 1600 which is of course our treatment period.

Table 3: Flexible Estimates With Percent Coal

	Dependent Variable: Ln Population Density				
	(1)	(2)	(3)	(4)	(5)
Percent Coal 1290	0.243*	0.240*	0.237*	0.238*	0.229*
	(0.106)	(0.112)	(0.106)	(0.0955)	(0.0968)
Percent Coal 1377	0.209**	0.208**	0.211**	0.211**	0.187**
	(0.0628)	(0.0670)	(0.0731)	(0.0647)	(0.0630)
Percent Coal 1600	0.334***	0.334**	0.338**	0.338***	0.296***
	(0.0904)	(0.0927)	(0.102)	(0.0889)	(0.0725)
Percent Coal 1801	0.478***	0.476***	0.483***	0.483***	0.472***
	(0.0835)	(0.0872)	(0.0944)	(0.0789)	(0.0653)
County Fixed-effect	Yes	Yes	Yes	Yes	Yes
Time Fixed-effect	Yes	Yes	Yes	Yes	Yes
Strong-Port	No	Yes	Yes	Yes	Yes
Weak-Port	No	No	Yes	Yes	Yes
Strong-Road	No	No	No	Yes	Yes
Weak-Road	No	No	No	No	Yes
Observations	190	190	190	190	190
R^2	0.858	0.859	0.864	0.872	0.907

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

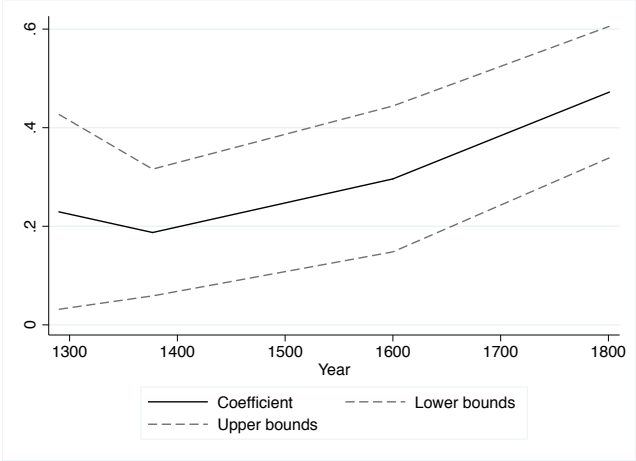


Figure 8: 95% Confidence interval flexible estimates

In Table 4 we go further and take our three year treatment window and investigate alternative placements of this window first starting in the initial 1086 period, and ending in our treatment window that starts in 1600. The initial window is just the opposite of our treatment period and hence must generate results identical to the true window after sign changes and account for constants. As we move rightwards the negative coefficients disappear by 1377 and both middle periods show no significant effect. Moving the window back into our chosen post treatment period reestablishes our results. The final output from this table is one more flexible estimation but this time within the treatment window. As shown each of the years makes a significant contribution to the treatment again suggesting our post treatment dating of 1600 was appropriate.

In Table 5 we introduce three different sensitivity analyses. First, in columns (1) and (2) we extend the treatment period to 1841. We can see the coefficient of interest is now higher for both the percent coal to 0.29 and binary coal to 0.49. In columns (3) and (4) we assume the energy transition occurred simultaneously in Surrey and Middlesex given the fuel scarcity was common across those counties. We can see that the percent coal indicator does not change but the binary coal is reduced from 0.498 to 0.370. Finally, in columns (5) and (6) we present the same specification but with a binary variable indicating whether the county was a large producer of textiles. We can see the coefficient of interest drops from 0.245 to 0.19 in the percent coal treatment and from 0.498 to 0.375 in the binary treatment.

Table 4: Treatment Estimates With Percent Coal

	Dependent Variable: Ln Population Density			
	(1)	(2)	(3)	(4)
Binary Coal 1290-1600	-0.0484			
	(0.193)			
Binary Coal 1377-1801		0.408*		
		(0.156)		
Binary Coal 1600-1801			0.498***	
			(0.125)	
Binary Coal 1600				0.287*
				(0.130)
Binary Coal 1801				0.709***
				(0.134)
County Fixed-effect	Yes	Yes	Yes	Yes
Time Fixed-effect	Yes	Yes	Yes	Yes
Strong-Port	Yes	Yes	Yes	Yes
Weak-Port	Yes	Yes	Yes	Yes
Strong-Road	Yes	Yes	Yes	Yes
Weak-Road	Yes	Yes	Yes	Yes
Observations	190	190	190	190
R^2	0.857	0.871	0.878	0.884

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Other controls

	Including 1841		Excluding Middlesex&Surrey		Including Textile dummy	
	(1)	(2)	(3)	(4)	(5)	(6)
Percent Coal 1600-1841	0.294***					
	(0.0553)					
Binary Coal 1600-1841		0.458***				
		(0.127)				
Percent Coal 1600-1801			0.243***		0.196**	
			(0.0603)		(0.0538)	
Binary Coal 1600-1801				0.370**		0.375**
				(0.134)		(0.125)
County Fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed-effect	Yes	Yes	Yes	Yes	Yes	Yes
Strong-Port	Yes	Yes	Yes	Yes	Yes	Yes
Weak-Port	Yes	Yes	Yes	Yes	Yes	Yes
Strong-Road	Yes	Yes	Yes	Yes	Yes	Yes
Weak-Road	Yes	Yes	Yes	Yes	Yes	Yes
Region-year	Yes	Yes	Yes	Yes	No	No
Textile Dummy	No	No	No	No	Yes	Yes
Observations	228	228	185	185	190	190
R^2	0.945	0.939	0.952	0.946	0.906	0.900

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Conclusions

In this paper we developed a simple spatial 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. The power density of an energy source determines the peak level of consumption per capita; roads and rivers effectively magnify the power density of available resources; and differences in power density across environments could create large differences in populations. We termed this last result the *density-creates- density* hypothesis.

To investigate this hypothesis we needed a specific mechanism to sort population across competing locations; a period when the energy system was relatively simple; and a place where data on energy sources and agglomeration was available. The natural solution was to use a Malthusian mechanism to investigate the population history of English counties over the 800 year period from the Norman invasion to the start of the Early Modern Era. The deep scholarship of this period provided to us data, many insights, and documentation of the first energy transition the world has traversed.

With help from our theory, we were able to develop a narrative account of this period consistent with many accounts of previous researchers. To move beyond this narrative we used our theory to suggest a very natural research design allowing us to investigate the causal mechanism we proposed. Our empirical results show that the movement to coal in our first energy transition was an important part of the massive reshuffling of the economic landscape. In five of the 14 coal counties, access to coal accounts for over 40% of their population growth; for all but one of this group access to coal provided over 20% of their population growth. For several counties, coal was absolutely critical; for all others it was an important driver of growth.

7 References

A Vision of Britain through Time. 2009. *1831 Census*. University of Portsmouth.

A Vision of Britain through Time. 2009. *1841 Census*. University of Portsmouth.

Ashraf, Q. and O. 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.

Beier, A.L. and R. Finlay (Eds.). 1986. *Making of the Metropolis: London, 1500-1700.* Prentice Hall Press.

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

Campbell, Bruce M.S. 2008. "Benchmarking Medieval Economic Development: England, Wales, Scotland, and Ireland, c.1290," *Economic History Review*, 61(4): 896-945.

Carus-Wilson, E. M., O. Coleman. 1963. *England's Export Trade, 1275-1547.* Oxford: Clarendon Press.

Chandler, T. 1987. *Four Thousand Years of Urban Growth: An Historical Census.* USA:The Edwin Mellen Press.

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.

Clark, P. 1972. "The Migrant in Kentish Towns, 1580-1640." In *Crisis and Order in English Towns: Essays in Urban History,* Peter Clark and Paul Slack (Eds.), 117-139. London: Routledge & Kegan Paul.

Davis, D.R. and D. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 92(5): 1269-1289.

Deane, P. and W.A. Cole. 1967. *British Economic Growth, 1688-1959: Trends and Structure* (Second Edition). Cambridge: Cambridge University Press.

Duranton, G 1998. "Labor Specialization, Transport Costs and City Size" *Journal of Regional Science*, 38(4): 553-573.

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

Epstein S.R. 1998. "Craft Guilds, Apprenticeship, and Technological Change in Preindus-

trial Europe." *The Journal of Economic History*, 58(3): 684-713.

Epstein S.R. 2004. "Property Rights to Technical Knowledge in Premodern Europe, 1300-1800" *American Economic Review, Papers and Proceedings*, 94(2): 382-387.

Feldman, R.A. 2005. *Recruitment, training and knowledge transfer in the London Dyers' Company, 1649-1826.* PhD thesis, The London School of Economics and Political Science (LSE).

Field, Jacob F. 2010. "Apprenticeship Migration to London from the North-East of England in the Seventeenth Century." *The London Journal*, 35(1): 1-21.

Finlay, R. 1981. *Population and Metropolis: The Demography of London 1580-1650.* Cambridge: Cambridge University Press.

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

Flinn, M.W. 1984. *The History of The British Coal Industry Volume 2: 1700-1830 The Industrial Revolution.* Oxford: Clarendon Press.

Floud, R. and D. McCloskey. (Eds.). 1994.*The Economic History of Britain Since 1700 Volume 1: 1700-1860.* (Second Edition). New York: Cambrdige University Press.

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

Galloway, R.L. 1882. *A History of Coal Mining in Great Britain.* London: MacMillan and Co.

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.

Grigg, D. 1980. *Population Growth and Agrarian Change: An Historical Perspective.* Cambrdige: Cambrdige University Press.

Hatcher, J. 1977. *Plague, Population, and the English Economy, 1348-1530.* Palgrave Macmillan.

Hatcher, J. 1993. *The History of British Coal Industry: Before 1700: Towards the Age of Coal,* Volume 1. Oxford: Clarendon Press.

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

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

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

Henderson, J.V. and J.F. Thisse. (Eds.). 2004. *Handbook of Regional and Urban Economics Volume 4: Cities and Geography.* The Netherlands: Elsevier: North-Holland

Hilbert, C. 1977. *London: The Biography of a City* (Revised Edition). London: Penguin Books, London.

Malthus, T. R. 1914. *An Essay on Population: Volume One.* London: J.M. Dent & Sons Ltd.

Hutchins, B.L. 1899. "Notes Towards the History of London Wages." *The Economic Journal.* 36(9): 599-605.

Leunig, T., C. Minns, and P. Wallis. 2011. "Networks in the Premodern Economy: The Market for London Apprenticeships, 1600–1749." *The Journal of Economic History*, 71(2): 413-443.

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*, 46(2): 266-279.

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

Moreno-Cruz, J. and M.S. Taylor. 2013. "A Spatial Approach to Energy Economics," *The National Bureau of Economic Research*, Working Paper No: 18908.

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.

Russell, J. C. 1948. *British Medieval Population.* Albuquerque: University of New Mexico Press.

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

Severnini, Edson R. 2013. *TThe Power of Hydroelectric Dams: Agglomeration Spillovers.* Working Paper at Carnegie Mellon University

Stone, L. 1966. "Social Mobility in England, 1500-1700." *Past & Present.* 33: 16-55.

The Edinburgh Geographical Institute, John Bartholomew & Co. *Coal and Iron* [Map]. Scale: 1:633,600.

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.

Wacher, J. 1995. *The Towns of Roman Britain* (Second Edition). London: B.T. Batsford Ltd.

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

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

Woude, A., A. Hayami, and J. Vries. (Eds.).1990. *Urbanization in History: A Process of Dynamic Interactions.* Oxford: Clarendon Press.

Wrigley, E.A. 1961. *Industrial Growth and Population Change: A Regional Study of the Coalfiled 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. 1967. "A Simple Model of London's Importance in Changing English Society and Economy 1650-1750." *Past & Present.* 37: 44-70.

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." *Economic History Review*, 62(3): 711-735.

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