Urbanization, Long-Run Growth, and the Demographic Transition*

Jonathan J Adams†
University of Florida

This Version: October 2, 2017

Link to Most Current Version

Abstract

Advanced economies undergo three transitions during their development: 1. They transition from a rural to an urban economy. 2. They transition from low income growth to high income growth. 3. Their demographics transition from initially high fertility and mortality rates to low modern levels. The timings of these transitions are correlated in the historical development of most advanced economies. I consider a nonlinear model of endogenous long run economic and demographic change, in which child quantity-quality substitution is driven by declining child mortality. Because the model captures the interactions between all three transitions, it is able to explain three additional empirical patterns: a declining urban-rural wage gap, a declining rural-urban family size ratio, and most surprisingly, that early urbanization slows development. This third prediction distinguishes the model from other theories of long-run growth, and I document evidence for it in cross-country data.
JEL Codes: E13, J11, N10, O18, O41

*I am grateful for helpful comments and guidance from Loukas Karabarbounis, Robert Lucas, Brent Neiman, Nancy Stokey, Harald Uhlig, and participants at the University of Chicago's Capital Theory, Applied Macro, and Growth and Development Working Groups.

†Website: www.jonathanjadams.com Email: adamsjonathan@ufl.edu Post: PO Box 117140, Gainesville FL 32611

1 Introduction

Why do economies transition from milennia of near-zero income growth to modern income growth rates? Leading theories of long-run growth attempt to understand development through one of two mechanisms. A literature following Becker et al. (1990) and Galor and Weil (2000) theorize that the central mechanism is substitution of child quantity to child quality, and jointly explain the growth transition and the demographic transition. Simultaneously, a literature following Hansen and Prescott (2002) and Lucas (2004) theorize that the central mechanism is structural transformation, and jointly explain the growth transition and urbanization.

But these mechanisms are not substitutes. The incentives for quantity-quality substitution differ between urban and rural areas, and structural transformation alone cannot explain the rapid acceleration of economic growth. I propose a unifying theory which features both mechanisms and endogenously reproduces the timing and magnitude of the three transitions: growth, urbanization, and demographics. Only by considering these transitions jointly can this theory predict the following observations: a declining urban-rural wage gap, a declining rural-urban family size ratio, and that early urbanization slows development. The third prediction, that urbanization is not a panacea for growth, is a result of high preindustrial urban child mortality and is novel in this literature.

The model economy has two sectors.[1] Human capital growth drives production to shift out of the rural sector, which has diminishing returns to scale.[2] The higher returns to scale of the urban sector increases the income growth associated with any

[1]Trade is missing from this theory, which is not a trivial omission. Stokey (1996) shows that openness to trade can speed a country's human capital accumulation with capital-skill complementarity, and Stokey (2001) shows that trade accelerated England's transition. ORourke and Williamson (2005) also shows trade's large effect on the English transition, demonstrating that increased trade openness explained a much of the increase in the ratio of wages to land rents. Galor and Mountford (2008) adds trade to a unified growth model, and shows that an early transition increases demand for the human capital-intensive sector through trade, accelerating the growth and demographic transitions.

[2]This dominance is similar to the results of Ngai and Pissarides (2007) or Acemoglu and Guerrieri (2008), where the sector spending the least on a fixed factor dominates in the long run if the elasticity of substitution among sectors is greater than one.

rate of human capital growth.

Households choose how much time to work in the market, how much time to spend raising children, and how much time to spend investing in their children's human capital, in the spirit of Becker (1960). As the child mortality rate improves, the household can afford higher quantity and quality of children. Increasing the number of children increases the cost of investing a unit of human capital in each child (as in Becker and Lewis (1973)), so parents reduce fertility and spend more time on human capital investment. At high mortality levels, households have more net children as they become less costly. But as child mortality falls further, the income effect dominates the substitution effect, so households shift from child quantity to child quality.[3] As families choose fewer children and more investment per child, per capita human capital grows faster and faster. Per capita income growth rises from near-stagnation to modern levels

Urban households suffer higher child mortality than rural households, so the relative wage in urban areas is high, because households must be compensated for moving to the deadly city.[4] As human capital grows, increased knowledge reduces mortality. Declines in the difference between urban and rural mortality reduces the wage premium needed to induce households to live in an urban area, enabling further urbanization.

A large branch of the unified growth literature considers the quantity-quality trade-off to be the central mechanism behind the growth transition. The motivation for this hypothesis is generally the correlation between the growth transition and the demographic transition. Becker et al. (1990) first analyze the quantity-quality trade-off in the context of an endogenous growth model; Lucas (2002) considers in-

[3] This Giffen property of child quantity is not new. See for example Willis (1973), or (Becker, 1981, Chapter 5) for the effect of child mortality declines in particular. Soares (2005) features the Giffen property and shows that child mortality declines can contribute to escape from a Malthusian trap. This driver of quantity-quality substitution is distinct from the precautionary motive in Kalemli-Ozcan (2002)

[4] Throughout the industrial revolution, cities were unhealthy places to live, with considerably higher mortality rates than rural areas. Williamson (2002) documents this pattern for England, as does Kesztenbaum and Rosenthal (2011) for France, Hanlon and Tian (2015) for China, and Cain and Hong (2009) in the United States.

troducing land as a fixed factor, allowing for either a Malthusian or modern growth outcome. Galor and Weil (2000) model fertility increasing as workers escape their subsistence consumption constraint and work fewer hours, but who then substitute to quality as returns to education rise. Galor and Moav (2004) introduce physical capital to the framework and study inequality during the transition. Doepke (2004) consider a two sector model with a child quantity-quality decision, where education subsidies and especially child labor regulation can influence a country's transition timing. Empirical evidence supports the quantity-quality substitution during industrialization, for example in Prussia (Becker et al., 2010) and in the American South (Bleakley and Lange, 2009).

The quantity-quality decision is governed by the return to human capital, which changes over the transition period. Some authors hypothesize that this return changes due to level effects in technology or growth. For example: Galor and Moav (2002) assumes a complementarity between education and the technological growth rate, while Doepke (2004) assumes that an increase in the level of skill-intensive technology increases the return. Other hypotheses include capital-skill complementarity; Fernandez-Villaverde (2001) finds the capital-specific technological change can explain more than 50% of England's growth and demographic transitions.

I assume a different channel: declining child mortality increases the return to human capital investment, driving the quantity-quality substitution. (Galor, 2011, Chapter 4) rejects this channel on theoretical grounds.[5] Using a static model of consumption and fertility choice, he shows that declines in child mortality rates should not affect fertility and will just increase surviving children, if the household has balanced growth compatible preferences. Doepke (2005) reaches a similar conclusion.

[5]Galor also rejects the child mortality channel on empirical grounds, given that the mortality in England declined significantly during the 18th century, over a hundred years prior to the demographic transition, without an associated decline in fertility. But this is only true of the crude death rate, when the relevant measure is the child mortality rate, which Wrigley and Schofield (1983) document as not declining significantly over the same period (Figure 1). A large literature suggests child mortality improvements are central to fertility declines. For example, Eckstein et al. (1999), Kalemli-Ozcan (2002), Lagerlof (2003), or Hazan and Zoabi (2006). Some theories such as Meltzer (1992) and Kalemli-Ozcan et al. (2000) suggest that the relevant mortality improvements for growth is adult mortality. This is supported in some empirical analysis (Lorentzen et al., 2008) but not others (Acemoglu and Johnson, 2007).

The model described in Section 3 rejects this conclusion when preferences are dynastic, and households invest in each child's human capital, even with balanced growth compatibility.

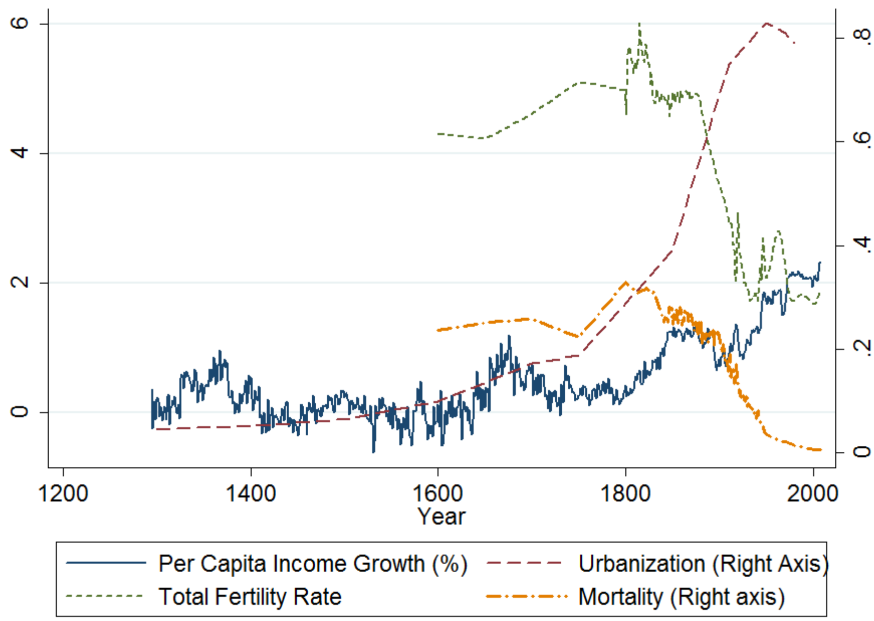

Figure 1: Transitions in England

Notes: GDP per capita is from The Maddison Project (2013) and Broadberry et al. (2010). Urbanization data are from Bairoch (1991). TFR and Mortality are from Ajus (2015) and Johansson et al. (2015) after 1800. Before 1800, they are from Wrigley and Schofield (1983).

A second set of theories focus on structural transformation as the cause of the growth transition. The motivation for this hypothesis is generally the correlation between the growth transition and urbanization. Hansen and Prescott (2002) consider an economy where only one sector uses land as an input and is perfectly substitutable with a constant returns sector. Given exogenous population and technological growth, the economy transitions from a Malthusian regime where only the land-intensive sector operates, to a modern regime where both operate. Lucas (2004) examines an endogenous growth model in which urban locations have increasing returns to scale in human capital as workers exchange ideas and learn from each other. Growth drives structural transformation out of agriculture due to the presence of a fixed factor, land.[6] Agriculture makes up the majority of employment in pre-industrial Europe (Allen, 2000) so structural transformation out of agriculture leads to urbanization if agriculture is not entirely substituted for rural non-agricultural industries. Economic growth can lead to both technological or preference-driven structural transformation, but the formal model in this paper considers technological structural transformation, motivated by evidence from Kuznets (1966), Maddison (1980), and Baumol et al. (1985), among many others.[7]

The remainder of this paper is organized as follows: Section 2 describes the empirical transitions, Section 3 describes the model environment, Section 4 defines equilibrium and characterizes several properties, Section 5 outlines the calibration procedure and simulation results, Section 6 considers the model under alternative calibrations and examines the empirical implications, and Section 7 concludes.

2 Empirical Patterns

Figure 1 plots the three transitions in England from 1295 CE. Before the industrial revolution, real income growth is consistently less than 1%. The urban share of people is less than 10%. Fertility and mortality rates are high. Then, since 1800, all of these series transition to modern values. This joint transition is an empirical regularity: among large countries with a thousand years of urbanization and income estimates, there is no evidence of a sustained transition for income growth, urbanization, fertility, or mortality before 1800.[8]

[6]A bevy of papers follow this basic approach, for example: Gollin et al. (2007) use a two-sector model of structural transformation to consider the impact of different agricultural productivity processes on countries' growth transitions. And, Michaels et al. (2012) directly relate technology-driven structural transformation to urbanization during the American transition. Strulik and Weisdorf (2008) build a two-sector model of the industrial population boom, where population growth drives productivity growth, creating an simultaneous income boom.

[7]Recent research from Herrendorf et al. (2013) and Comin et al. (2015) find that when considered together, both technology and preferences have driven structural transformation, so a more complete model of structural transformation would incorporate income effects as well.

[8]Except for Belgium and the Netherlands, which had urban shares near 30% in 1500 CE.

Moreover, these transitions occur around the same time within a country. To illustrate, I calculate the first year that each country surpasses a benchmark level for each series: (a) twenty-five years of 1% annual income growth, (b) 50% urban, (c) total fertility rate below 3, and (d) child mortality below 5%. Table 1 reports the correlation table for these transition years.[9] Countries that experience an early growth transition also tend to urbanize early, and have fertility and mortality fall early. This correlation is also observable in the current cross-section. Table 2 reports the percentage of countries surpassing the urbanization and demographic benchmarks for two income groups. Countries with 2012 GDP per capita of at least $10,000 are broadly urban with low fertility and low mortality. Countries with GDP per capita less than $1,000 tend to be rural with high fertility and high mortality.

	Income Growth	Urbanization	Fertility	Mortality
Income Growth	1			
Urbanization	0.518	1		
Fertility	0.542	0.393	1	
Mortality	0.608	0.467	0.881	1

Table 1: Correlation of Transition Years

	Urban > 50%	TFR < 3	Child Mort. < 5%
Income > $10K	93%	96%	97%
Income < $1K	7%	10%	13%

Table 2: Transitioned Percentage of Countries by Income in 2012

[9]The set of countries with one thousand years of data for income and urban population share, defined as the percentage of the population living in cities of at least 5,000 people. 18 countries are in this dataset: China, India, and 16 European countries, listed in Table 8. Historical estimates for these datasets corresponds to the modern states' current geographic area whenever possible.

2.1 Early Urbanization Predicts Later Transition

This prediction is unique in distinguishing this theory from other models of urbanization and long-run growth. Theories such as Hansen and Prescott (2002) or Lucas (2004) feature an urban sector with strictly greater returns than the rural sector. In such a model, an economy that is parameterized to choose a higher urbanization level for a given income level will grow faster.

The model presented in Section 3 also has higher urban returns, but features a trade-off: high child mortality. This reduces the household budget set, decreasing the return to human capital investment, which delays the income growth transition. Then, over the following transition, growth and urbanization are highly correlated. To accommodate this trade-off and explain the empirical pattern, theories of long-run growth and urbanization must also account for the incentives that govern the demographic transition.

Section 6 describes and tests the effect of early urbanization on transition timing. In the data, I find that early urbanization predicts a later transition to modern income growth. Across time, space, and theory, urbanization is generally associated with higher income. That's why the prediction is surprising, and distinguishes this paper from other theories of long-run growth.[10]

2.2 Urban-Rural Differences

The model also produces two other facts observed in the English transition: a declining urban-rural wage premium, and a declining rural-urban family size ratio. I focus on England, because of the quality of its long-run macroeconomic time series, and availability of historical urban and rural data on fertility, mortality, and wages. The model is calibrated to English data in Section 5.1.

[10]Several studies find related patterns and suggest alternative explanations. Acemoglu et al. (2002) show that non-European countries that were relatively urban in 1500 had lower incomes in 1995, and argue that European colonists endowed initially poorer countries with better institutions, generating faster growth. Nunn and Qian (2011) argue that the more rural Northern European countries were able to disproportionately benefit from agricultural advances, such as the introduction of the potato, than the more urban Southern European countries.

The urban-rural wage gap declines over time.[11] In the 1830's, Williamson (1987) calculates a nominal wage gap for unskilled workers of 73 %, and a real wage gap of 46%; he estimates that the majority of the gap was compensating for high urban mortality. In contrast, DCosta and Overman (2013) estimates an unconditional wage gap of 14 % in Britain from 1998-2008. Conditioning on observables such as occupation and skill further reduces the gap to 2 %, in line with estimates for other countries.[12]

The rural-urban family size ratio declines over time. Clark (2009) estimates gross fertilities for the 15th-18th century that are 27% higher on farms than in London, and 12% higher in other non-farm households than in London. Mortality differences led farm-dwelling fathers to have over twice as many surviving children than a Londoner. And other non-farm fathers had 70% more surviving children than a Londoner. By the turn of the 20th Century, (Szreter and Hardy, 2001, Table 20.6) estimates that rural fertilities were only 3-5% larger than in urban areas. In modern European countries with available data, rural crude birth rates average 98% of urban rates (United Nations Statistics Divison, 2012, Table 9). And in 2007 England, London's crude birth rate is now 30% higher than the country as a whole (Office for National Statistics, 2008, Table 6.2), although due to demographic differences its total fertility rate is lower Kulu and Washbrook (2014). This pattern is documented in many countries.[13]

3 Model

The model economy contains two production sectors: an urban sector where the only input is human capital, and a rural sector with human capital and land inputs. Land

[11] Specifically, the wage gap controlling for worker skill. Income differences between urban and rural workers may be very large if urban workers accumulate much more human capital, as in Lucas (2004). In the cross-section, Lagakos and Waugh (2013) and Young (2013) use a worker-selection model to estimate that most of the productivity gap in poor countries is due to sorting on skill.

[12] Additionally, there is cross-sectional evidence that the urban-rural productivity gap is declining in income, and nearly disappears in rich countries (Gollin et al., 2013).

[13] For example: Germany (Knodel et al., 1974, Chapter 3), Italy (Bacci, 1977), or the United States (Kiser, 1960)

is in fixed supply, but human capital grows endogenously, and is the only source of growth in the model. Households have overlapping generations, and parents decide the quantity and quality of their children.

3.1 Production

The rural production sector, denoted with the subscript R, combines human capital and land to produce output. Its production function is:

$$F_R(\tilde{h}, \tilde{l}) = \tilde{h}^{\theta}\tilde{l}^{1-\theta} \tag{1}$$

The rural firms are land intensive, such as a farm, a mine, or a logger. An individual rural firm chooses human capital $\tilde{h}$ and land $\tilde{l}$.

The urban production sector, denoted with the subscript U. It uses only human capital to linearly produce output. Its production function is

$$F_U(\tilde{h}) = \tilde{h} \tag{2}$$

Urban firms are relatively less land intensive than farms, which characterizes most of the nonagricultural sector of the economy. An urban firm might be a factory, a craftsman, or a service firm. An urban firm chooses only human capital $\tilde{h}$.

The unique final good is produced competitively by combining the output of the urban and rural sectors, with elasticity of substitution ϵ

$$F(\tilde{x_R}, \tilde{x_U}) = A(\zeta \tilde{x_U}^{\frac{\epsilon-1}{\epsilon}} + (1-\zeta)\tilde{x_R}^{\frac{\epsilon-1}{\epsilon}})^{\frac{\epsilon}{\epsilon-1}} \tag{3}$$

Final goods firms choose rural goods $\tilde{x_R}$ and urban goods $\tilde{x_U}$ as inputs.

Firms in all sectors are small and competitive, so they take prices as given. Let p_R denote the intermediate rural good's price, and p_U the intermediate urban good's price. Normalize the price of the final output good to one. Also, let r denote the rental rate of land, w_R the rural wage rate per unit of human capital, and w_U the urban wage rate per unit of human capital. Then, a rural firm solves:

$$\max_{\tilde{h},\tilde{l}} p_R\tilde{h}^{\theta}\tilde{l}^{1-\theta} - w_R\tilde{h} - r\tilde{l} \tag{4}$$

An urban firm solves:

$$\max_{\tilde{h}} p_U \tilde{h} - w_U \tilde{h} \tag{5}$$

A final goods firm solves:

$$\max_{\tilde{x_R}, \tilde{x_U}} A(\zeta \tilde{x_U}^{\frac{\epsilon-1}{\epsilon}} + (1-\zeta)\tilde{x_R}^{\frac{\epsilon-1}{\epsilon}})^{\frac{\epsilon}{\epsilon-1}} - p_R \tilde{x_R} - p_U \tilde{x_u} \tag{6}$$

3.2 Households

Agents live for two periods: in their first period of life they are children, and in the second period they are parents.[14] Generations overlap within a household: each household consists of one parent and a number of children. The parent makes all of the household's choices, choosing consumption, the number of children, and education spending. The parent must also choose whether to live in an urban or rural area, and how much time to dedicate to market work. Households do not own land; similar to Galor and Weil (2000), we suppose that an infinitesimally small fraction of the population holds all the land, and has a negligible impact on aggregate human capital and demographics.

Utility is dynastic. Parents enjoy present consumption c, their number of surviving children n, and their dynasty's discounted future utility. A parent discounting by β has utility:

$$V_t = u(c_t, n_t) + \beta V_{t+1} \tag{7}$$

where $u(c_t, n_t)$ is the period utility function, V_t is the parent's dynastic utility, and V_{t+1} is the dynastic utility of the next generation. Parents' preference for quantity of children is driven by their period utility, $u(c_t, n_t)$, because V_{t+1} is each child's future utility, not the total utility of the next generation.[15]

[14]Because adults all live to the same age, all mortality improvements are to child mortality. By construction this ignores any impact on transition dynamics from changes to adult mortality, which Lorentzen et al. (2008) suggest affects the quantity-quality trade-off, even when controlling for child mortality.

[15]This formulation is a simplification of Becker and Barro (1988), in which the discount factor is a concave function of n_t. I eliminate the dependence on n_t for tractability and parsimony. The first order condition for children is simpler and will yield a constant share of time spent working in the market with Cobb-Douglas utility. Eliminating the dependence on n_t also reduces the number of parameters to be calibrated.

The period utility function $u(c, n)$ is increasing in both arguments and must be balanced growth compatible, so that as the time cost of raising children rises, it is offset by an income effect. When necessary, I assume the functional form from Barro and Sala-i Martin (2004):

$$u(c, n) \equiv \frac{(cn^{\phi})^{\sigma}}{\sigma} \tag{8}$$

where $\phi > 0$, $\sigma < 1$ and $\phi\sigma < 1$. ϕ controls the preference for consumption relative to children, while σ controls substitutability across generations: $\frac{1}{1-\sigma}$ is the elasticity of intergenerational substitution.

Parents choose how to allocate their time to three activities: market work (τ_c), producing children (τ_n), and educating children (τ_h). They have one unit of time to allocate to these activities:

$$\tau_c + \tau_n + \tau_h = 1 \tag{9}$$

Households in sector $j \in U, R$ earn wage w_j per unit of human capital, per unit of time worked. Income is spent on consumption, so a parent with human capital h working time τ_c consumes:

$$c = w_j h \tau_c \tag{10}$$

A household choosing time τ_n produces n surviving children by:

$$n = S_j \alpha \tau_n \tag{11}$$

where parameter α is the productivity for producing children. S_j is the fraction of newborns that survive to adulthood in sector j. S_j is exogenous from the perspective of the household, but will depend on aggregate human capital, so it may vary over time. Child production is time intensive, so productivity is not improved by parental human capital.

All children are endowed with their parents' human capital, but parents can spend time educating their children to increase their human capital further. A household with human capital h choosing education time τ_h produces human capital h' for their n children linearly:

$$(h' - h)n = \xi \tau_h h \tag{12}$$

The number of surviving children n enters this equation because parents must spend time educating each of their children. All child mortality resolves before parents start to invest in their human capital.[16] The parent's ability to impart human capital is increasing in their own human capital, h, and proportional to the parameter ξ.

Combining equations (9), (10), (11) and (12) yield the combined budget constraint:

$$c + \frac{w_j(h' - h)n}{\xi} + \frac{w_j hn}{\alpha S_j} = w_j h \tag{13}$$

The household's time is used for consumption, human capital investment, or producing children. The total value in numeraire of the household's time is $w_j h$. The value of time spent in the market is what they earn and spend on consumption c. The value of time spent investing $(h' - h)n$ units of human capital is $\frac{w_j(h'-h)n}{\xi}$, and the value of time spent producing n children is $\frac{w_j hn}{\alpha S_j}$.

3.2.1 The Household's Problem

The household's problem is to choose consumption c, children n, their children's future human capital h', and location j to maximize dynastic utility. Let Λ denote the aggregate state of the economy; then the household's Bellman equation is

$$V(h; \Lambda) = \max_{c,n,h',j \in J} u(c, n) + \beta V(h'; \Lambda') \tag{14}$$

subject to the budget constraint (13), location choice set $j \in U, R$, and non-negativity constraints:

$$c \geq 0 \;\; n \geq 0 \;\; h' \geq h \tag{15}$$

Solving the household's problem yields the first order conditions:

$$u_n(c, n) = u_c(c, n)(\frac{w_j h'}{\xi} + \frac{w_j h}{\alpha S_j}) \tag{16}$$

$$u_c(c, n) w_j n = \xi \beta V'(h'; \Lambda') \tag{17}$$

[16] Tamura (2006) considers an alternative framework where some human capital investment may be lost due to child mortality risk. Reductions in child mortality increase the return to human capital even more strongly in such an environment.

and envelope condition:

$$V'(h;\Lambda) = u_c(c,n)w_j(1 + \frac{n}{\xi} - \frac{n}{\alpha S_j}) \tag{18}$$

When the preferences in (8) are applied to first order condition (16), consumption is a constant share of income:

$$\frac{c}{w_j h} = \frac{1}{1+\phi} \tag{19}$$

This also implies that $\tau_c = \frac{1}{1+\phi}$ is constant for all households. This result is due to the marginal cost of children being proportional to total income, and the homotheticity of preferences, which is required for balanced growth compatibility. As total income $w_j h$ rises, the income effect exactly offsets the substitution effect, and households spend the same amount of time $\tau_n + \tau_h$ on children, although they may reallocate their time between child quantity and human capital investment.

Different children of the same parent might choose different locations, so a household does not have a single Euler equation. Rather, the Euler equation for child k with the balanced growth preferences is:

$$(\frac{c'_k}{c})^{1-\sigma} = (\frac{n'_k}{n})^{\phi\sigma+1}\frac{w'_k}{w_j}\xi\beta(\frac{1}{n'_k} + \frac{1}{\xi} - \frac{1}{\alpha S'_k}) \tag{20}$$

Denote human capital growth by $1 + g \equiv \frac{h'}{h}$. Then the Euler equation can be rewritten using the budget constraint and consumption share in terms of fertilities, human capital growth, and wages:

$$(1+g)^{1-\sigma}(\frac{n}{n'_k})^{\phi\sigma}(\frac{w_j}{w'_k})^{\sigma} = \beta\frac{\xi}{n}(\tau_c + n'_k\frac{1+g'_k}{\xi}) \tag{21}$$

The left hand side of equation (21) is marginal utility growth across generations. On the right hand side, $\tau_c + n'_k\frac{1+g'_k}{\xi}$ is the return to human capital investment, and $\frac{\xi}{n}$ is the productivity of parental time at producing human capital for each child. As n rises, it costs more parental time to give each child a unit of human capital, so the return on parental time falls. Depending on marginal utility growth, this force creates the potential for child mortality declines to induce substitution from child quantity to child quality.

3.3 Aggregates and Laws of Motion

The state of the economy is determined by the function $\lambda(h)$, which denotes the measure of households with human capital h.

The total population in the economy N is:

$$N = \int_h \lambda(h) \tag{22}$$

The measure of households with h in sector j is denoted by $\lambda(h, j)$, and this is an equilibrium object because sector j is a choice. All households work τ_c units of time, so aggregate human capital inputs in the economy are:

$$H_j = \int_h \tau_c h \lambda(h, j) \tag{23}$$

and aggregate land is L, a fixed value. Given factor prices w_U, w_R, r, total income in the economy is:

$$Y = w_U H_U + w_R H_R + rL \tag{24}$$

Let n_j denote the fertility choice of a household in sector j. Let $h(h', j)$ denote the human capital of a household in sector j that would choose h' for their children. The distribution of households evolves by:

$$\lambda(h') = \int_j n_j \lambda(h(h', j), j) \tag{25}$$

which simply says that the number of households with h' equals the number of households that chose h' for their children, times the number of surviving children per household n_j.

Child survival $S_j(\bar{h})$ is a function of location j and average human capital, $\bar{h}$:

$$\bar{h} = \int_h \frac{h\lambda(h)}{N} \tag{26}$$

The dependence on location captures differences in child mortality across urban and rural areas. The dependence on average human capital captures the impact of the technology level on child mortality. This may come in the form of beneficial technological improvements such as clean water, food safety, and medicine.[17]

Assume the function $S_j(\bar{h})$ is increasing in $\bar{h}$ and has common limit for all j:

$$\lim_{\bar{h}\to\infty} S_j(\bar{h}) = \bar{S} \tag{27}$$

It must also be that $S_j(\bar{h}) \in [0,1]$ for all $\bar{h} > 0$. A particular form will be estimated in Section 5.

Finally, to determine the population distribution, an assumption must be made about how households allocate themselves. *Assume there is no reverse migration*: children stay where they are born unless some migration is needed from their birth location to the other location; no one leaves their birth location if net migration is flowing into it. Without this assumption, optimality conditions and constraints will only determine the allocation of aggregate human capital, but not of people, who might have differing human capital levels. In equilibrium, this assumption implies that dynasties move from rural to urban areas, and never return.[18]

4 Equilibrium

4.1 Definition

A competitive equilibrium in this economy consists of sequences for $t \geq 0$ of prices, p_R, p_U, w_R, w_U, r; aggregate allocations, Y, x_U, x_R, H_U, H_R, Z; distribution of household human capital $\lambda(h,j)$; and household allocations, $c(h,j), n(h,j)$; given initial distribution of human capital $\lambda(h)_0$ and the aggregate quantity of land L, such that:

1. The firm allocations solve (4), (5), and (6).

[17] See for example Preston (1996)'s overview, Szreter (1988)'s examination of the U.K.'s decline in particular, or Deaton (2006)'s review of Fogel (2004)'s conflicting findings. Empirically, income growth also allows for household investments in child survival, such as improved nutrition, which research such as McKeown (1976) and Fogel (2004) emphasize. But $S_j(\bar{h})$ only captures the impact of the technology level.

[18] This is not a perfect assumption. Young (2013) shows that most urban-rural migration is from the countryside to the city, but there is still a reverse flow of workers returning to rural areas.

2. The household allocations maximize (14) subject to (13) and (15).

3. Markets clear: $Y = F(x_U, x_R)$, $X_U = F_U(H_U)$, $X_R = F_R(H_R, L)$

4. The law of motion (25) holds for all human capital levels.

5. Household aggregates add up, satisfying equations (22), (23), (24), and (26), and there is no reverse migration.

4.2 Equilibrium Prices

The firms' profit maximization (equations (4), (5), and (6)) implies that equilibrium prices must relate to equilibrium factors by:

$$w_U = p_U \qquad w_R = p_R \theta (H_R)^{\theta-1} L^{1-\theta} \qquad r = p_R(1-\theta)(H_R)^{\theta} L^{-\theta} \tag{28}$$

$$p_U = A^{\frac{\epsilon-1}{\epsilon}} \zeta \left(\frac{Y}{x_U}\right)^{\frac{1}{\epsilon}} \qquad p_R = A^{\frac{\epsilon-1}{\epsilon}} (1-\zeta) \left(\frac{Y}{x_R}\right)^{\frac{1}{\epsilon}} \tag{29}$$

4.3 Equilibrium Location Choice

Households choose the location that gives them the highest utility. As usual, the household's value function is the maximum of the value of choosing each location. In most models this upper envelope is not differentiable at the point of indifference. But in this model, the value function is differentiable for indifferent households.

Proposition 1 *If households are indifferent between urban and rural locations in equilibrium, then their marginal value of human capital is equal in both locations.*

Proposition 1 is proved in Appendix A.1. Marginal value equalization implies a convenient equilibrium condition for the wage premium. Setting the envelope condition (18) equal in both locations, and substituting for consumption by equation (19) yields:

$$w_R^{\sigma} n_R^{\sigma\phi+1} \left(\frac{1}{n_R} + \frac{1}{\xi} - \frac{1}{\alpha S_R}\right) = w_U^{\sigma} n_U^{\sigma\phi+1} \left(\frac{1}{n_U} + \frac{1}{\xi} - \frac{1}{\alpha S_U}\right) \tag{30}$$

The wage premium is a compensating differential for mortality differences. If urban child survival S_U is lower than rural survival, then all else equal equation (30) will imply $w_U > w_R$. But in equilibrium all else is not equal, and urban households will change their child rearing decision n_U to partially compensate for a lower survival rate.

4.4 Equilibrium in the Limit

In this section I derive the asymptotic behavior of the economy. I show that the urban share approaches one, and the urban-rural wage, growth, and fertility gaps disappear.

The following propositions are proved in Appendix A.

Proposition 2 *If* $\lim_{t\to\infty} \bar{h} = \infty$, *then the limiting urban-rural wage premium is* $\frac{w_U}{w_R} \to 1$.

Proposition 3 *If* $\lim_{t\to\infty} \bar{h} = \infty$, $\lim_{t\to\infty} n \geq 1$ *and* $\epsilon > 1$, *then the long-run urban share converges to* 1.

Proposition 4 *If* $\lim_{t\to\infty} \bar{h} = \infty$, $\lim_{t\to\infty} n \geq 1$ *and* $\epsilon > 1$, *then the the limit of both urban and rural wages is* $\bar{w} \equiv A\zeta^{\frac{\epsilon}{\epsilon-1}}$.

Proposition 4 implies that wages, which are paid per unit of human capital, are not growing or falling in the limit. Therefore long run human capital growth $\bar{g}$ and children $\bar{n}$ are determined in the limit by the long run budget constraint and the long run steady state Euler equation:

$$\tau_c + \frac{\bar{g}\bar{n}}{\xi} + \frac{\bar{n}}{\alpha\bar{S}} = 1 \tag{31}$$

$$(1+\bar{g})^{1-\sigma} = \beta\left(\frac{\xi\tau_c}{\bar{n}} + 1 + \bar{g}\right) \tag{32}$$

5 Quantitative Analysis

Parameter values are chosen to match key features of the data, an initial condition is chosen to look like England in year 1500 C.E., and the economy's equilibrium transition path is calculated.

5.1 Calibration

Ten parameters must be calibrated: production parameters A, θ, ζ, and ϵ; preference parameters ϕ, σ, and β; and household parameters α and ξ. Initial conditions must be chosen: land L and population N_0 are normalized to one. All households are initialized with $h = 1$. The two technology functions $S_U(\bar{h})$ and $S_R(\bar{h})$ must also be characterized. Finally, assume one model period is 25 years. Calibrated values appear in Table 3, chosen to resemble in England in 1500 C.E. England is the calibrated country because England has historical data on urban and rural differences for fertility and mortality.

The rural production parameter θ is set to 0.74 so that the land share of farm income is 26%, the value for England in 1500 C.E. estimated by Clark (2010).

To calibrate the parameters $(A, \zeta, \epsilon, \alpha, \xi, \phi, \sigma, \beta)$, we target several empirical moments. First, the initial urban share is targeted to 0.064, estimated by Bairoch et al. (1988) for England in 1500. Initial human capital growth is targeted to 1.3%, the smoothed 25-year income growth at 1500 CE, in the Broadberry et al. (2010) data. Long run human capital growth $\bar{g}$ is targeted to 52%, England's 25-year real income growth rate since 1950.

Initial fertility and mortality rates are targeted to estimates from Clark (2009) for England in 1500-1800. Initial urban and rural probabilities of surviving to age 25 are $S_{U,0} = 0.59$ and $S_{R,0} = 0.68$. The ratio of urban to rural surviving children per adult is targeted to 0.77, the ratio estimated by Clark (2009). However, the estimates in levels are too high to map directly to the model, because the data is from wills and does not account for people who choose not to have children. So $n_{R,0}$ and $n_{U,0}$ are chosen to target an initial population growth rate of 8.5% per 25 years, which matches the growth rate for England from 1400-1600 estimated by Broadberry

	Parameter	Value	Interpretation
(i)	θ	0.74	Labor Share in Rural Sector
(ii)	β	0.36	Discount Factor
(iii)	σ	0.49	Utility Curvature
(iv)	ϕ	0.74	Child Preference
(v)	α	3.74	Childrearing Productivity
(vi)	ξ	3.29	Education Productivity
(vii)	A	3.68	Total Factor Productivity
(viii)	ζ	0.36	Urban Goods Weight
(ix)	ϵ	4.50	Urban-Rural Substitution Elasticity
(x)	υ	0.35	Technology Effect on Survival

Table 3: Calibrated Parameters

et al. (2010). The long run population growth is targeted to 0%, implying $\bar{n} = 1$.

Five preference and household parameters $(\phi, \sigma, \beta, \xi, \alpha)$ can be solved for jointly, given targets for human capital growth, fertility and mortality, and a target long run 5% annual rate of return on human capital investment. The five parameters are identified by five equations: the long run and initial rural budget constraints, long run and initial steady state Euler equations, and the return to human capital investment. The initial rural Euler equation is not identical to the steady steady Euler equation because there are small movements in wages and net fertilities initially, so the equilibrium value of $n_{R,0}$ and $g_{R,0}$ will not exactly match the targets.

The initial urban-rural wage premium is implied by the indifference equation (30). Chosen empirical targets imply an initial premium of $\frac{w_{U,0}}{w_{R,0}} = 1.23$. The initial urban share, normalization of $h = 1$, and market time of $\tau_c = \frac{1}{1+\phi}$ imply initial supplies of human capital $H_{R,0}$ and $H_{U,0}$. Setting the ratio of marginal products equal to the initial wage premium identifies the weighting parameter ζ in the production function, conditional on a choice of the elasticity of substitution ϵ. Targeting long run wage $\bar{w} = 1$ then implies a value for TFP A.

The child survival function $S_j(\bar{h})$ requires a functional form. This function should have four properties: $S(\bar{h}) \in (0, 1)$ for all $\bar{h} \geq 0$, $S_j(\bar{h}_0)$ matches the target for $S_{j,0}$, $S'(\bar{h}) > 0$ for all $\bar{h} \geq 0$, and $S_j(\infty) = \bar{S}$ so that in the long run, survival approaches a chosen limit. A form satisfying these properties is:

$$S_j(\bar{h}) = \bar{S} - (\bar{S} - S_{j,0}) \frac{1 + \upsilon \bar{h}_0}{1 + \upsilon \bar{h}} \tag{33}$$

This is a transformed logistic CDF, which is chosen for parsimony as it is governed by only one free parameter υ, and also for having a positive limit as $\bar{h} \to 0$. It satisfies the other desired conditions: when $\bar{h} = \bar{h}_0$, then $S_j(\bar{h}) = S_{j,0}$; $S'_j(\bar{h}) > 0$; and in the limit as $\bar{h} \to \infty$, then $S_j(\bar{h}) \to \bar{S}$.

The function is estimated on England's child mortality time series, given the targets for $S_{j,0}$ and $\bar{S}$. Appendix C describes this estimation.

The final parameter to calibrate is the elasticity of substitution ϵ. The elasticity of substitution controls the speed of urbanization as aggregate human capital grows. Figure 2 plots the transition year for urbanization and for income growth as a function of ϵ. A higher value of ϵ speeds the urbanization transition by making urban and rural sectors more substitutable: given a decline in the wage premium, more human capital will shift into the urban sector. But a higher value of ϵ also decreases growth: there are more urban households, which face lower child survival rates and spend less time investing in human capital for their children (see Section 5.2). The dashed lines are the empirical transition years. The elasticity of substitution is selected to minimize the mean squared error between the model and empirical transition years.

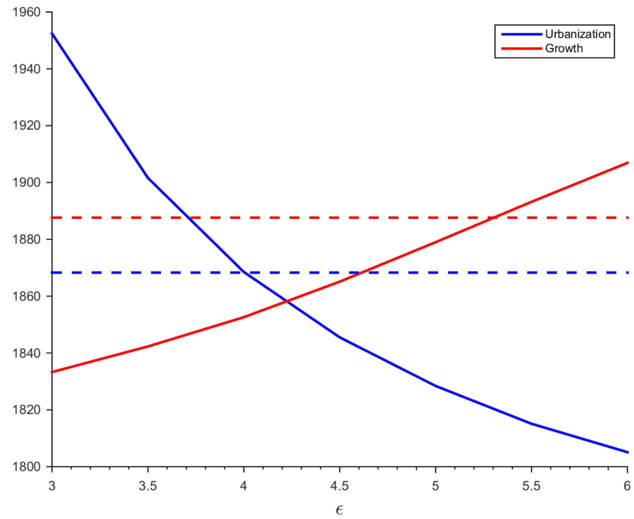

Figure 2: Elasticities of Substitution and Transition Years

Notes: Urbanization transition is when urban share > 50%. Growth transition is when annual income growth > 1%.

5.2 Results

The economy is initialized in 1500 and is run 21 periods to 2000. The economy begins with most of the population in the rural sector. As the population grows and human capital accumulates, households move to the urban sector (Figure 3). The simulated urban share surpasses 50% in year 1846, versus the empirical urban share which reached 50% around 1863. In the long run, the population fully urbanizes.

As mortality falls, surviving children become cheaper. But increasing the number of children increases the cost of investing a unit of human capital in each child. So parents reduce fertility and spend more time on human capital investment. Quantitatively, fertility falls more than one for one with the decrease in cost for unconstrained households, so surviving children fall and households substitute from quantity to quality. Income per household grows slowly at first, but eventually rises, asymptoting to the long run value (Figure 4).

To understand the dynamics of the two sectors, Figure 5 plots the Euler equation

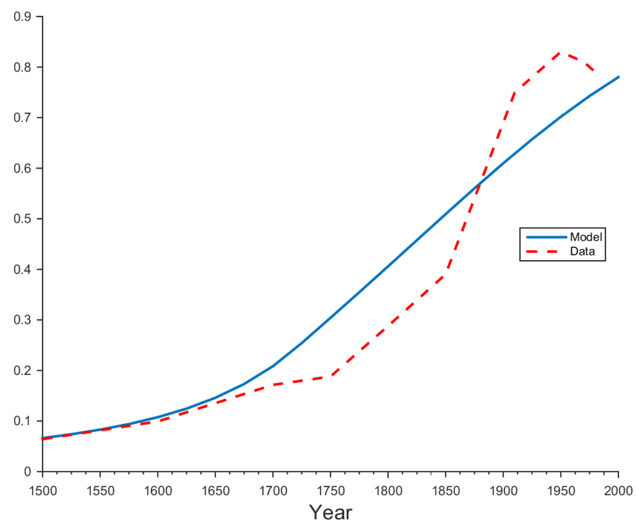

Figure 3: Simulation: Urban Share

Notes: Data from Bairoch (1991) and Bairoch et al. (1988)

in (21) in a steady state:

$$(1 + g_{ss})^{1-\sigma} = \beta \frac{\xi}{n_{ss}} (\tau_c + n_{ss} \frac{1 + g_{ss}}{\xi}) \tag{34}$$

For the steady state Euler equation, children choose the same location as their parent. $\tau_c + n_{ss}\frac{1+g_{ss}}{\xi}$ is the return on human capital, and $\frac{\xi}{n_{ss}}$ is the productivity of parental time in producing a unit of human capital for each child. With calibrated parameter values, the steady state Euler equation implies that g_{ss} is decreasing in n_{ss} for $g \in (0, \bar{g}]$.[19] In this region, households will always trade-off child quantity for quality, and never increase both. Thus an expansion in the household's budget set caused by declines in child mortality will induce substitution from quantity to quality even though quantity has become cheaper.

[19] The steady Euler Equation gives n as a decreasing function of g for $(1-\sigma)(1+g)^{-\sigma} > \beta$ which always holds when $\sigma \leq 0$, i.e. when the intergenerational elasticity of substitution is less than 1. However, the calibration gives $\sigma > 0$, so n is decreasing in g for all $g < 14.5$, which is well above the long run steady state.

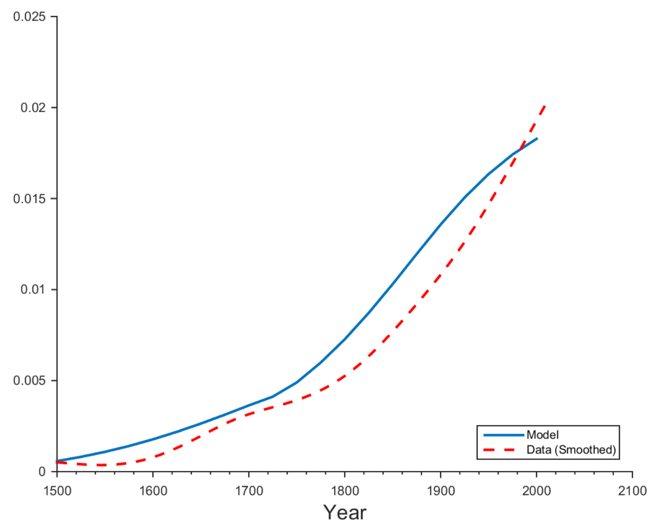

Figure 4: Simulation: Income Growth

Notes: Data from Broadberry et al. (2010) and The Maddison Project (2013), smoothed with an HP filter.

To understand this effect, Figure 5 also plots the normalized budget constraint, which divides the budget constraint (13) by total income:

$$\tau_c + \frac{gn}{\xi} + \frac{n}{\alpha S_j} = 1 \tag{35}$$

This budget constraint is plotted for three different survival levels: $\bar{S}$, $S_{R,0}$, and $S_{U,0}$. The steady state Euler equation differs slightly from the equilibrium Euler equation for initial urban or rural households, but this figure is a useful approximation for understanding the dynamics. As the rural survival rate improves, the rural budget constraint shifts towards the long run budget constraint, and the rural allocation moves down the Euler equation, shifting from quantity towards quality. The initial urban budget constraint does not intersect the Euler equation: urban households are constrained at $g = 0$ so that the non-negativity constraint 15 is satisfied. As the urban survival rate improves, the urban budget constraint shifts towards the rural budget constraint, and children increase. When the survival rate has improved

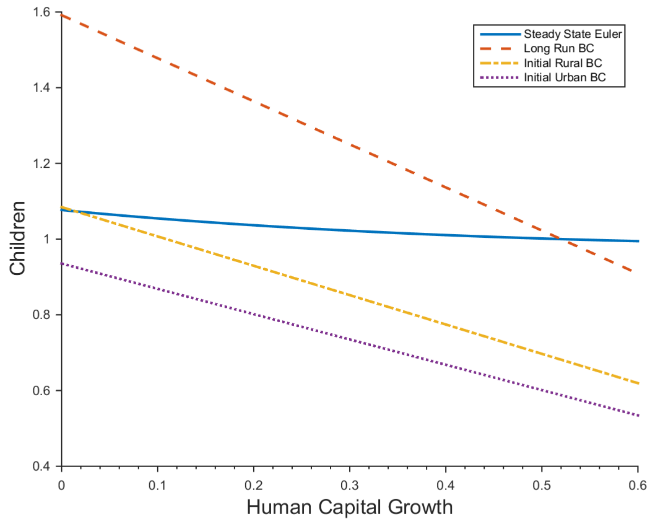

Figure 5: Quantity-Quality Substitution

sufficiently to unconstrain urban households, they follow the rural households and substitute from quantity to quality.

Figure 6 plots the ratio of urban to rural values for three quantities: wages, children, and survival, exhibiting the predictions from Section 2.2. As human capital grows, urban and rural survival rates both grow towards the same limit, so the ratio rises to one. The urban-rural wage ratio is the compensating differential for mortality differences. Williamson (1987) estimates this ratio is 1.46 in the early 1800s, versus 1.05 in the model in 1800 and 1.22 in 1500. Survival is initially lower in urban areas, so a high wage premium is necessary to make households indifferent between locations. As the survival ratio rises to one, wage ratio falls to one, and the compensating differential disappears in the limit. Urban households initially choose fewer children than rural households because they are constrained at $g = 0$ and urban children are very expensive due to their low survival rate. As the survival rate improves, the urban-rural child ratio rises as urban households have more children and rural households substitute from quantity to quality. Eventually the urban households become unconstrained and also substitute towards quality. The ratio approaches one in the long run, as the survival differential disappears.

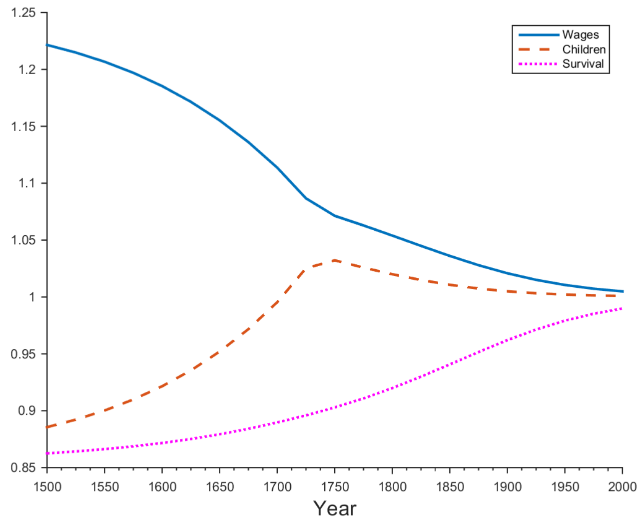

Figure 6: Simulation: Urban/Rural Ratios

While the urban-rural family size ratio increases from the initial period to the long run, fitting the empirical pattern in Section 2.2, it is not monotonic over the whole sample, which may not be true in the data. This non-monotonicity is because urban households choose higher fertilities than rural households, to compensate for high child mortalities. This fertility difference is true empirically in the modern day, but not during the 19th or early 20th centuries. To explain the fertility ratio over this period, the theory needs further urban-rural differences, such as the cost of raising children in the city, or higher urban returns to human capital (Becker, 1981, Chapter 5).

In the aggregate, fertility and mortality fall as the economy urbanizes and transitions to modern growth. Figure 7 plots births, deaths, and the net growth of each dynasty. Births are calculated before accounting for the fraction S_j that do not survive to adulthood. Total births slowly start to decline with child mortality, as fewer newborns are necessary to produce a given surviving child. In the long run, the birth rate falls to the limiting population growth rate, because child mortality disappears.

Similarly, the death rate falls to one in the long run - all adults die every period, and all children live. The difference of these series is the population growth rate which falls to zero in the long run, just as the net number of children produced by each household falls to one.

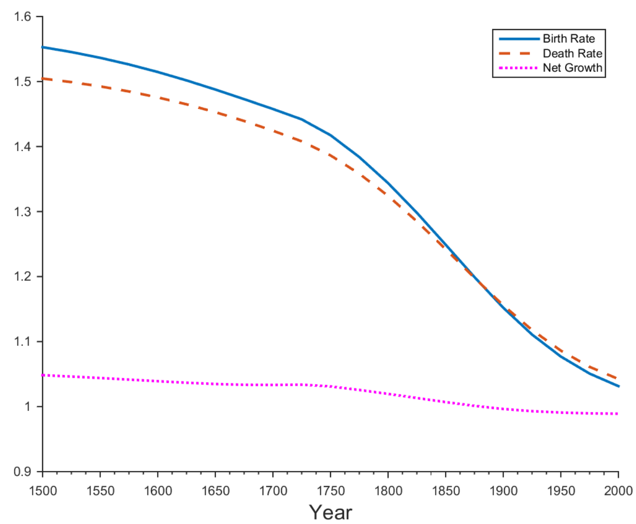

Figure 7: Simulation: Demographic Transition

6 Cross-Country Analysis

What impacts do the initial conditions have on the equilibrium dynamics? Subsection 6.1 considers the effect of changing initial calibration targets on the transition timing. Subsection 6.2 examines the empirical implications.

6.1 Model Sensitivity

The transition timing is sensitive to the initial calibration targets. In particular, three targets have large effects: the initial urban share, the initial human capital growth rate, and the initial population growth rate.

First, I vary the initial urban share target while holding constant the other targets. Varying the initial urban share chiefly operates through production parameters. In general, a change to a calibration target will not have an effect on just a subset of parameters. But the urban share's effects on calibration are relatively straightforward. Raising the initial urban share requires increasing ζ, the weight on urban goods in the final production sector, and decreasing TFP A, to keep the long run marginal productivity of human capital constant. The elasticity of substitution ϵ is kept constant, for this parameter is identified off of the transition timing. There are small changes to household parameters, which must be adjusted to keep initial population growth at the target level, but these changes are small because $s_{U,0}$ is small.

Figure 8 plots the year that the model economy surpasses 1% income growth against the initial urban share. All other calibration targets are baseline values. As the initial urban share increases, the growth transition is delayed. Because the economy is more urban, and urban parents choose lower human capital growth for their children, the economy grows more slowly for many centuries. In the long run, the economy catches up to the baseline long run growth target as urban mortality improves.

Next, I vary only the initial human capital growth target, which primarily affects household parameters. Increasing the initial growth target increases the necessary household productivity of human capital investment ξ, and decreases the productivity of child-rearing α. Intuitively, increasing ξ makes the household richer, but decreasing α raises the relative price of children quantity versus quality. Thus the initial period household chooses the same initial population growth, but a higher rate of human capital growth. Of course, other parameters must have small adjustments to maintain the long run calibration targets.

Figure 9 plots the year that the model economy surpasses 1% income growth against the initial income growth rate. Other calibration targets are unchanged from the baseline. The transition timing is very sensitive to the initial growth rate. An economy with low initial growth has poor productivity of human capital investment. This decreases the growth rate along the transition, and the economy takes longer

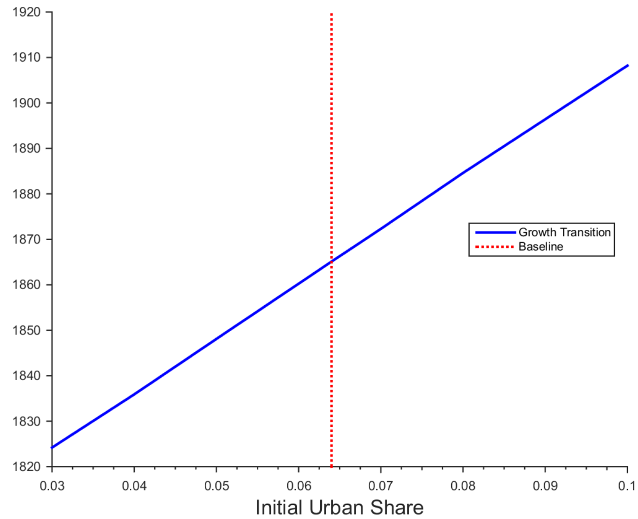

Figure 8: Transition Years and Initial Urban Share

to converge to the long run limit. Lower human capital investment has some secondary effects: urbanization is slowed, which increases income growth by shifting the population composition towards the lower mortality rural sector, but the mortality transition is also slowed for both sectors, reducing income growth.

Lastly, increasing the initial population growth target speeds the economy's transition. Higher population growth is mainly achieved by increasing the productivity of childrearing α, but with a decrease in child preference ϕ to maintain the long run population growth. Because the initial urban households are constrained at $g = 0$ due to the high child mortality, they spend all of their non-market income producing children. So an increase in α disproportionately increases initial urban children relative to rural children. It takes less time for urban households to become unconstrained, and to start substituting from child quantity to quality. The income growth transition year is plotted against the initial population growth rate, all else equal, in Figure 10. A higher population growth rate with the same household human capital growth rate speeds the income growth transition as households substitute to child quality earlier.

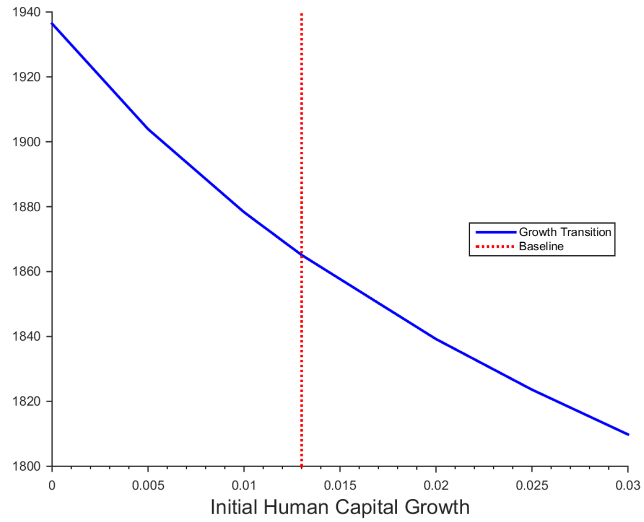

Figure 9: Transition Years and Initial Human Capital Growth

6.2 Cross-Country Empirics

The analysis in section 6.1 suggest that, all else equal, a country will have a faster growth transition if it has: 1. a lower initial urban share, 2. a higher initial income growth rate, or 3. a higher initial population growth rate.

To test these relationships in the data, I construct transition years for 29 countries[20] for which I have the relevant data in year 1500. Then I regress the transitions years T_j against country characteristics in year 1500:

$$T_j = \beta_0 + \beta_1 s_{U,0,j} + \beta_2 \Delta y_{0,j} + \beta_3 n_{0,j} + \boldsymbol{\alpha}' \mathbf{D}_j + \varepsilon_j \tag{36}$$

where $s_{U,0,j}$ is country j's initial urbanization rate, $\Delta y_{0,j}$ is their initial per capita real income growth, $n_{0,j}$ is their initial population growth, $\mathbf{D}_j$ is a vector of country characteristics for some regression specifications, and ε_j is the error term. Income and population data are from The Maddison Project (2013). For comparability, England's data is also from this source, instead of the superior data used in section

[20]Listed in column 2 of Table 8.

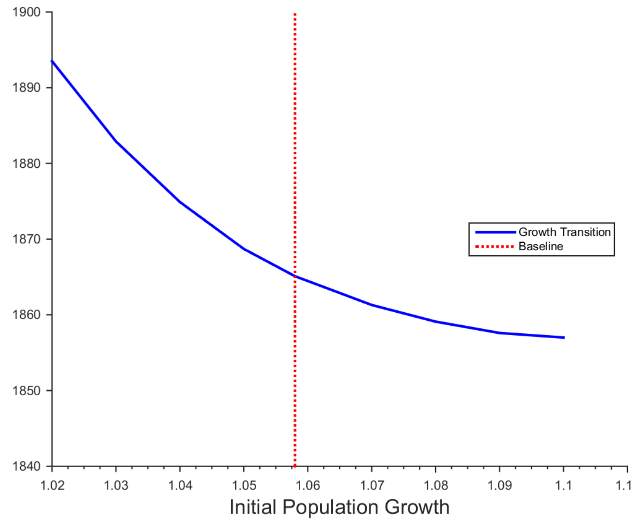

Figure 10: Transition Years and Initial Population Growth

5.1. Before 1820, income and population data are centennial, so in a given year (e.g. 1500) growth is the annualized rate over the preceding century. After 1820, income data is annual, and the transition year is defined as the first time that the 25-year moving average of income growth exceeds 1%. Finally, urbanization data is from Bairoch et al. (1988) and The Clio Infra Project (2016), interpolated over gap years.

Table 4 reports the baseline results in the first column. As predicted by the model, initial urbanization predicts a later transition, while higher income and population growth predict an earlier transition. The coefficient on initial urbanization implies that an addition 10 percentage points of urbanization should delay the growth transition by 28 years, all else equal. Both the urbanization and income growth rate coefficients are significant at the 5% level, but population growth is not, which is the case for almost every specification of these regressions.

Table 4 also reports the results of several robustness checks. The second regression uses population density as a proxy for urbanization, in case mismeasurement of the historical urbanization rates is correlated with transition timing. But population density also predicts a later transition, and the effect is significant at the 1% level.

The third regression includes a vector of geographic controls[21] considered by Ashraf and Galor (2011). The effect of urbanization is strengthened in this regression, and is significant at the 1% level. The fourth regression includes continent fixed effects, which weakens the relationship, and the fifth regression includes dummies identifying colonies, which has little effect on the estimates.

The year 1500 CE is used to initialize the baseline calibration in Section 5.1 because it is the earliest period on which urban-rural differences in mortality and fertility are available. But the empirical effects of urbanization and income growth on transition timing can be examined for other years. Table 5 reports the baseline regression for many initial years. Urbanization slows the transition for all years, although it is not always significant, particularly in 1800 CE, as countries are approaching their transition date. Income growth predicts an earlier transition in all cases except 1000 CE, when the data is especially poor.

Testing initial conditions on transition timing supports the model's predictions, yet these tests are limited by the small sample of countries with historical data before 1800 CE for all necessary variables, and by the accuracy of these historical estimates. To take advantage of more data, I next conduct a more powerful test of the relationship between early urbanization and transition timing.

In the context of the model, high initial urban shares are interpreted as reflecting high urban productivity relative to rural productivity.[22] In equilibrium, this results in a higher level of urbanization at every income level, although it may not be higher at every point in time. To illustrate, Figure 11 plots urbanization and income level for the baseline calibration, and for an alternative with China's initial urban share of 0.12. At every level of income, the alternative has higher urbanization. Why? The urban-rural wage premium is the compensating differential for the urban-rural mortality ratio. And the mortality ratio falls as the country's human capital rises. Because the urban sector is more productive relative to the rural sector in the

[21] Absolute latitude, percentage of arable land, percentage of land within 100 km. of a coast or river, percentage of land in temperate zones, and percentage of land in tropical or subtropical zones.

[22] Ashraf and Galor (2011) estimate that countries in 1500 CE with high agricultural productivity have greater population density, particularly China and India. It must be that these countries are initially urban because their urban productivity is especially high.

alternative calibration, more households must be urban for a given wage premium.

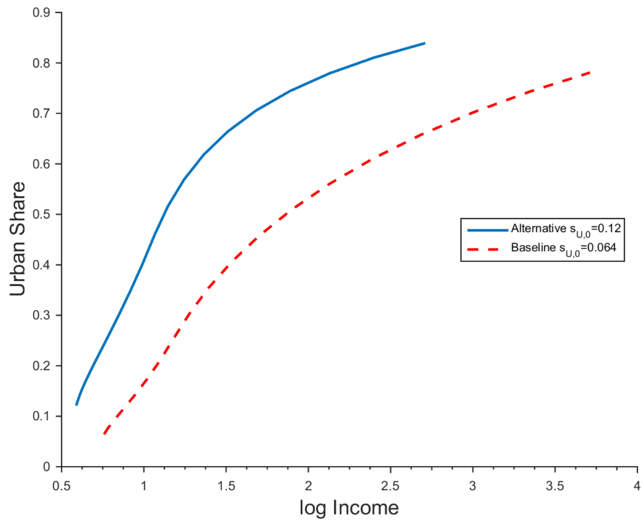

Figure 11: Urbanization and Income Levels: Model

I use a two-stage regression approach to test to see if countries with high rates of urbanization relative to income have later growth transitions, as predicted by the model. First, I run the following panel regression, for country j in year t:

$$s_{U,t,j} = \gamma \log y_{j,t} + d_j + \kappa + \varepsilon_{j,t} \tag{37}$$

This is a regression of urban share on log income with country fixed effects. Next, I regress the transition year T_j on the estimated fixed effects:

$$T_j = \psi \hat{d}_j + \varkappa + \varphi_j \tag{38}$$

Table 6 summarizes the 1st stage estimated country fixed effects. There are 57 countries[23] with urbanization data before their income growth transition, and 6,491 total year-country observations. The regressor in the second stage is an estimate and analytical standard errors will be incorrect, so standard errors are calculated by boostrapping.

[23] Listed in column 3 of Table 8.

Table 7 reports the results of the second stage regression. Countries that have a higher urbanization level conditional on their income transition much later. I estimate $\hat{\psi} = 496.5$: if a country that is 10 percent more urban conditional on income, then they will transition almost 50 years later.

Figure 12 plots countries' first-stage estimated fixed effects versus their transition year, and the second state regression line. Geographic patterns emerge. In the lower left are many Western and Central European Powers and their colonies, which were initially very rural and transitioned early. In the upper right are many Asian countries, including China and India, which were urban early in their development, but transitioned later.

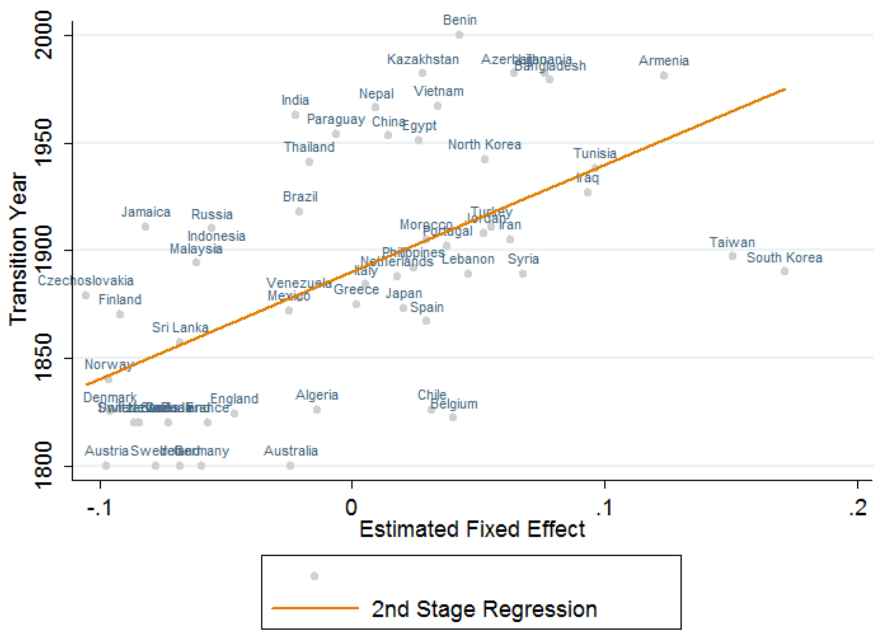

Figure 12: Estimated Country Effects and Transition Years

Both regression approaches suggest that countries relatively predisposed towards urbanization will transition to modern growth later, despite the general correlation of urbanization and income growth over time.

7 Concluding Remarks

This paper has developed a unified endogenous growth model producing three simultaneous transitions: the growth transition, urbanization, and the demographic transition. The model quantitatively reproduces the timing and magnitude of England's transitions. Because the model considers growth, urbanization, and demographics jointly, it also generates three additional empirical observations: a declining urban-rural wage gap, a declining rural/urban family size ratio, and that early urbanization delays a country's transition.

The relationship between early urbanization and transition timing is an identifying feature of the model which distinguishes it from other theories of urbanization and long run growth. I use several estimation strategies to show that the relationship between early urbanization and transition timing is robust in the historical experiences of many countries. This empirical fact raises further research questions. I have identified one plausible channel driving this relationship; could there be others? And does this channel apply to current low income countries? Future work can address these questions by applying and expanding on the theory in this paper.

References

Acemoglu, D. and V. Guerrieri (2008). Capital deepening and nonbalanced economic growth. *Journal of Political Economy 116*(3), 467–498.

Acemoglu, D. and S. Johnson (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy 115*(6), 925–985.

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly journal of economics 117*(4), 1231–1294.

Ajus, F. (2015). Gapminder total fertility rate dataset. Version 6.

Allen, R. C. (2000). Economic structure and agricultural productivity in europe, 1300–1800. *European Review of Economic History 4*(01), 1–25.

Ashraf, Q. and O. Galor (2011). Dynamics and stagnation in the malthusian epoch. *The American Economic Review*, 2003–2041.

Bacci, M. L. (1977). *A history of Italian fertility during the last two centuries.* Princeton University Press.

Bairoch, P. (1991). *Cities and economic development: from the dawn of history to the present.* University of Chicago Press.

Bairoch, P., J. Batou, and P. Chevre (1988). La population des villes europeenees de 800 a 1850: Banque de donnees et analyse sommaire des resultats. [the population of european cities. data bank and short summary of results: 800-1850]. *Centre of Internation Economic History Series* (2).

Barro, R. J. and X. Sala-i Martin (2004). *Economic Growth.* Cambridge, MA: MIT Press.

Baumol, W. J., S. A. B. Blackman, and E. N. Wolff (1985). Unbalanced growth revisited: asymptotic stagnancy and new evidence. *The American Economic Review*, 806–817.

Becker, G. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries*, pp. 209–240. Columbia University Press.

Becker, G. (1981). *A Treatise on the Family.* Harvard University Press.

Becker, G. and R. J. Barro (1988). A reformulation of the economic theory of fertility. *The Quarterly Journal of Economics 103*(1), 1–25.

Becker, G., K. Murphy, and R. Tamura (1990). Economic growth, human capital and population growth. *Journal of Political Economy 98*(5), S12–S137.

Becker, G. S. and H. G. Lewis (1973). On the interaction between the quantity and quality of children. journal of political economy. 81: 2, pp. *S279-88*.

Becker, S. O., F. Cinnirella, and L. Woessmann (2010). The trade-off between fertility and education: evidence from before the demographic transition. *Journal of Economic Growth 15*(3), 177–204.

Bleakley, H. and F. Lange (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The review of economics and statistics 91*(1), 52–65.

Broadberry, S. N., B. M. Campbell, A. D. Klein, M. Overton, and B. v. Leeuwen (2010). British economic growth: 1270-1870. *Working Paper*.

Cain, L. and S. C. Hong (2009). Survival in 19th century cities: The larger the city, the smaller your chances. *Explorations in Economic History 46*(4), 450–463.

Clark, G. (2009). Urbanization, health and income in malthusian europe. *unpublished*.

Clark, G. (2010). The macroeconomic aggregates for england, 1209–2008. *Research in Economic History 27*, 51–140.

Comin, D. A., D. Lashkari, and M. Mestieri (2015, September). Structural change with long-run income and price effects. Working Paper 21595, National Bureau of Economic Research.

DCosta, S. and H. Overman (2013). The urban wage growth premium: Evidence from british cities. In *ERSA conference papers*, Number 13. European Regional Science Association.

Deaton, A. (2006). The great escape: A review of robert fogel's the escape from hunger and premature death, 1700–2100. *Journal of Economic Literature 44*(1), 106–114.

Doepke, M. (2004). Accounting for fertility decline during the transition to growth. *Journal of Economic growth 9*(3), 347–383.

Doepke, M. (2005). Child mortality and fertility decline: Does the barro-becker model fit the facts? *Journal of Population Economics 18*(2), pp. 337–366.

Eckstein, Z., P. Mira, and K. I. Wolpin (1999). A quantitative analysis of swedish fertility dynamics: 1751–1990. *Review of Economic Dynamics 2*(1), 137–165.

Fernandez-Villaverde, J. (2001). Was malthus right? economic growth and population dynamics.

Fogel, R. W. (2004). *The escape from hunger and premature death, 1700-2100: Europe, America, and the Third World*, Volume 38. Cambridge University Press.

Galor, O. (2011). *Unified Growth Theory*. Princeton University Press.

Galor, O. and O. Moav (2002). Natural selection and the origin of economic growth. *The Quarterly Journal of Economics 117*(4), pp. 1133–1191.

Galor, O. and O. Moav (2004). From physical to human capital accumulation: Inequality and the process of development. *The Review of Economic Studies 71*(4), 1001–1026.

Galor, O. and A. Mountford (2008). Trading population for productivity: theory and evidence. *The Review of economic studies 75*(4), 1143–1179.

Galor, O. and D. N. Weil (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *The American Economic Review 90*(4), 806–828.

Gollin, D., D. Lagakos, and M. Waugh (2013). The agricultural productivity gap. *Quarterly Journal of Economics forthcoming*.

Gollin, D., S. L. Parente, and R. Rogerson (2007). The food problem and the evolution of international income levels. *Journal of Monetary Economics 54*(4), 1230–1255.

Hanlon, W. W. and Y. Tian (2015). Killer cities: Past and present. *The American Economic Review 105*(5), 570–575.

Hansen, G. D. and E. C. Prescott (2002). Malthus to solow. *The American Economic Review 92*(4), 1205–1217.

Hazan, M. and H. Zoabi (2006). Does longevity cause growth? a theoretical critique. *Journal of Economic Growth 11*(4), pp. 363–376.

Herrendorf, B., R. Rogerson, and Á. Valentinyi (2013). Two perspectives on preferences and structural transformation. *The American Economic Review 103*(7), 2752–2789.

Johansson, K., M. Lindgren, C. Johansson, and O. Rosling (2015). Gapminder child mortality dataset. Version 7.

Kalemli-Ozcan, S. (2002). Does the mortality decline promote economic growth? *Journal of Economic Growth 7*(4), pp. 411–439.

Kalemli-Ozcan, S., H. E. Ryder, and D. N. Weil (2000). Mortality decline, human capital investment, and economic growth. *Journal of Development Economics 62*(1), 1–23.

Kesztenbaum, L. and J.-L. Rosenthal (2011). The health cost of living in a city: The case of france at the end of the 19th century. *Explorations in Economic History 48*(2), 207–225.

Kiser, C. V. (1960). Differential fertility in the united states. In *Demographic and Economic Change in Developed Countries*, pp. 77–116. Columbia University Press.

Knodel, J. E. et al. (1974). *The decline of fertility in Germany, 1871-1939*, Volume 2. Princeton University Press.

Kulu, H. and E. Washbrook (2014). Residential context, migration and fertility in a modern urban society. *Advances in Life Course Research*.

Kuznets, S. (1966). *Modern economic growth: Rate, structure, and spread.* Yale University Press New Haven.

Lagakos, D. and M. E. Waugh (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review 103*(2), 948–80.

Lagerlof, N.-P. (2003). From malthus to modern growth: Can epidemics explain the three regimes? *International Economic Review 44*(2), pp. 755–777.

Lorentzen, P., J. McMillan, and R. Wacziarg (2008). Death and development. *Journal of Economic Growth 13*(2), pp. 81–124.

Lucas, R. E. (2002). *Lectures on Economic Growth.* Harvard University Press.

Lucas, R. E. (2004). Life earnings and rural-urban migration. *Journal of Political Economy 112*(S1), S29–S59.

Maddison, A. (1980). Economic growth and structural change in the advanced countries. *Western Economies in Transition. Eds.: I. Leveson and W. Wheeler. London: Croom Helm.*

McKeown, T. (1976). *The Modern Rise of Population.* Edward Arnold, London.

Meltzer, D. O. (1992). *Mortality Decline, the Demographic Transition, and Economic Growth.* Ph. D. thesis, University of Chicago, Department of Economics.

Michaels, G., F. Rauch, and S. J. Redding (2012). Urbanization and Structural Transformation. *The Quarterly Journal of Economics 127*(2), 535–586.

Ngai, L. R. and C. A. Pissarides (2007). Structural change in a multisector model of growth. *The American Economic Review 97*(1), 429–443.

Nunn, N. and N. Qian (2011). The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics 126*(2), 593–650.

Office for National Statistics (2008). *Regional Snapshot 2008.* Office for National Statistics.

ORourke, K. H. and J. G. Williamson (2005). From malthus to ohlin: Trade, industrialisation and distribution since 1500. *Journal of Economic Growth 10*(1), 5–34.

Preston, S. H. (1996). American longevity: Past, present, and future. Technical report, Center for Policy Research, Maxwell School, Syracuse University.

Soares, R. R. (2005). Mortality reductions, educational attainment, and fertility choice. *The American Economic Review 95*(3), pp. 580–601.

Stokey, N. L. (1996). Free trade, factor returns, and factor accumulation. *Journal of Economic Growth 1*(4), pp. 421–447.

Stokey, N. L. (2001). A quantitative model of the british industrial revolution, 1780–1850. In *Carnegie-Rochester conference series on public policy*, Volume 55, pp. 55–109. Elsevier.

Strulik, H. and J. Weisdorf (2008). Population, food, and knowledge: A simple unified growth theory. *Journal of Economic Growth 13*, 195–216.

Szreter, S. (1988). The importance of social intervention in britain's mortality decline c. 1850? 1914: a re-interpretation of the role of public health. *Social history of medicine 1*(1), 1–38.

Szreter, S. and A. Hardy (2001). Urban fertility and mortality patterns. In D. M. J (Ed.), *The Cambridge Urban History of Britain*. Cambridge University Press.

Tamura, R. (2006). Human capital and economic development. *Journal of Development Economics 79*(1), 26–72.

The Clio Infra Project (2016). The clio-infra database on urban settlement sizes: 1500-2000. http://www.cgeh.nl/urbanisation-hub-clio-infra-database-urban-settlement-sizes-1500-2000.

The Maddison Project (2013). The maddison project. http://www.ggdc.net/maddison/maddison-project/home.htm.

United Nations Statistics Divison (2012). *Demographic Yearbook 2013.* United Nations.

Williamson, J. G. (1987). Did english factor markets fail during the industrial revolution? *Oxford Economic Papers*, 641–678.

Williamson, J. G. (2002). *Coping with city growth during the British industrial revolution*. Cambridge University Press.

Willis, R. J. (1973). A new approach to the economic theory of fertility behavior. *Journal of Political Economy 81*(2), pp. S14–S64.

Wrigley, E. A. and R. S. Schofield (1983). English population history from family reconstitution: summary results 1600–1799. *Population Studies 37*(2), 157–184.

Young, A. (2013). Inequality, the urban-rural gap and migration. *The Quarterly Journal of Economics*, 0–25.

A Proofs

A.1 Proof of Proposition 1

In this section I prove that if households are indifferent between urban and rural locations in equilibrium, then their marginal value of human capital is equal in both locations.

Proof. Dynastic utility (7) can be expanded into the discounted sum:

$$V_t = \sum_{k=0}^{\infty} \beta^k \frac{(c_{t+k} n_{t+k}^{\phi})^{\sigma}}{\sigma} \tag{39}$$

Let $\mathcal{J}$ denote a sequence of location choices, where $\mathcal{J}(t)$ is the sector chosen in period t. Substituting for the household's consumption choice, dynastic utility becomes:

$$\begin{aligned} V_t &= \sum_{k=0}^{\infty} \beta^k \frac{(\tau_c w_{t+k,\mathcal{J}(t+k)} h_{t+k} n_{t+k}^{\phi})^{\sigma}}{\sigma} \\ &= h_t^{\sigma} \sum_{k=0}^{\infty} \beta^k \frac{(\tau_c w_{t+k,\mathcal{J}(t+k)} \frac{h_{t+k}}{h_t} n_{t+k}^{\phi})^{\sigma}}{\sigma} \end{aligned} \tag{40}$$

Normalized human capital $\frac{h_{t+k}}{h_t}$ can be expressed in terms of growth rates:

$$\frac{h_{t+k}}{h_t} = \prod_{s=t}^{t+k-1} (1+g_s) \;\; k \geq 1$$

substituting this expression into (40) gives V_t in terms of sequences of wages, locations, choices of n and g, and h_t. Lemma 5 (proved below) says that choices of n and g are independent of h_t. So given these sequences, the utility for a location sequence $\mathcal{J}$ is a function of h, proportional to current human capital to a power:

$$V_{\mathcal{J}}(h) \propto h^{\sigma} \tag{41}$$

Now consider two different location sequences $\mathcal{J}$ and $\mathcal{J}'$. Because of the proportionality in (41), it is true that:

- If a household is indifferent for some $\hat{h}$, then

$$V_{\mathcal{J}}(h) = V_{\mathcal{J}'}(h) \quad \forall h > 0 \tag{42}$$

- If a household strictly prefers $\mathcal{J}$ for some $\hat{h}$, then

$$V_{\mathcal{J}}(h) > V_{\mathcal{J}'}(h) \quad \forall h > 0 \tag{43}$$

In equilibrium, households must be indifferent between urban and rural locations for some $\hat{h}$. This follows from the equilibrium property that all households cannot strictly prefer one location, and that the household utility given a particular location decision (equation 41) is continuous in h.

Let $\mathcal{J}_U$ and $\mathcal{J}_R$ denote optimal location sequences for a household with $\hat{h}$ given a current period choice of urban or rural location respectively. The household is indifferent by definition of $\hat{h}$, so $V_{\mathcal{J}_U}(\hat{h}) = V_{\mathcal{J}_R}(\hat{h})$. Then it follows from (42) and (43) that households are indifferent between $\mathcal{J}_U$ and $\mathcal{J}_R$ for all $\forall h > 0$, and there is no other sequence of locations that any household strictly prefers.

This sequence indifference implies that for any $\mathcal{J} \in \{\mathcal{J}_U, \mathcal{J}_R\}$:

$$V_{\mathcal{J}}(h_t) = h_t^{\sigma} \sum_{k=0}^{\infty} \beta^k \frac{(\tau_c w_{t+k,\mathcal{J}(t+k)} \prod_{s=t}^{t+k-1} (1+g_s) n_{t+k}^{\phi})^{\sigma}}{\sigma} \tag{44}$$

$$\equiv h_t^{\sigma} \mathcal{V} \tag{45}$$

Thus the marginal value of human capital is equalized in both locations:

$$V'_{\mathcal{J}}(h_t) = \sigma h_t^{\sigma-1} \mathcal{V} \quad \forall \mathcal{J} \in \{\mathcal{J}_U, \mathcal{J}_R\} \tag{46}$$

■

Lemma 5 *Given a series of wages $w_{t,j}$, survival rates $S_{t,j}$, and sequence of locations $\mathcal{J}(t)$, a dynasty's choice of children n_t and human capital growth g_t is independent of its level of human capital h_t.*

The central assumption driving this result is the homotheticity of the balanced growth compatible preferences.

Proof. The combined budget constraint (13) and equilibrium choice of consumption $c = \tau_c w_j h$ imply that the budget constraint can be normalized by dividing by $w_j h$:

$$\tau_c + \frac{gn}{\xi} + \frac{n}{\alpha S_j} = 1$$

and recall that $\tau_c = \frac{1}{1+\phi}$ is constant. This normalized budget constraint and the Euler equation (21) jointly characterize the household's equilibrium behavior, and neither depends on the level of h. ■

A.2 Proof of Proposition 2

In this section I prove that if $\lim_{t\to\infty} \bar{h} = \infty$, then the limiting urban-rural wage premium is $\frac{w_U}{w_R} \to 1$.

Proof.

Suppose that $S_R = S_U$ but $w_R < w_U$. Consider the optimal rural allocations (c_R, n_R, h'_R) given w_R and S_R. A household could choose to live in the urban area and, per the combined budget constraint (13), would be able to afford the allocation $(\tilde{c_U}, n_R, h'_R)$ where $\tilde{c_U} > c_R$. Thus they would strictly prefer the urban location and this could not be an equilibrium. Similarly, if $w_R > w_U$ then an urban household could switch to a rural location and be strictly better off. The only possible equilibrium given $S_R = S_U$ must have $w_R = w_U$.

By assumption $\lim_{\bar{h}\to\infty} S_j(\bar{h}) = \bar{S}$ for all j. So in the limit, it must be that $\frac{w_U}{w_R} \to 1$. ■

A.3 Proof of Proposition 3

In this section I prove that if $\lim_{t\to\infty} \bar{h} = \infty$, $\lim_{t\to\infty} n \geq 1$ and $\epsilon > 1$, then the long-run urban share converges to 1.

Proof.

The limits for $\bar{h}$ and n imply that aggregate human capital $H = N\bar{h}$ is growing in the long run: $\lim_{t\to\infty} H = \infty$.

Use the equilibrium prices in equations (28) and (29) to express the wage premium as:

$$\frac{w_U}{w_R} = \frac{H_R^{1-\theta}\zeta x_R^{\frac{1}{\epsilon}}}{\theta L^{1-\theta}(1-\zeta)x_U^{\frac{1}{\epsilon}}}$$

Then substitute with the sectoral production functions to express the wage premium in terms of human capital inputs:

$$\frac{w_U}{w_R} = \frac{H_R^{1-\theta(1-\frac{1}{\epsilon})}\zeta}{\theta L^{(1-\theta)(1-\frac{1}{\epsilon})}(1-\zeta)H_U^{\frac{1}{\epsilon}}}$$

Aggregate human capital supplied is $\tau_c H$. The urban share of aggregate human capital is s_U. Substituting and rearranging gives:

$$\frac{w_U}{w_R}(\tau_c H)^{(\frac{1}{\epsilon}-1)(1-\theta)} = \frac{(1-s_U)^{1-\theta(1-\frac{1}{\epsilon})}\zeta}{s_U^{\frac{1}{\epsilon}}\theta L^{(1-\theta)(1-\frac{1}{\epsilon})}(1-\zeta)}$$

The agricultural labor share θ is between 0 and 1 by assumption, so if $\epsilon > 1$ then the left hand side of this equation is decreasing in H, and the right hand side is decreasing in s_U. Proposition 2 says that in the limit $w_U = w_R$, so if $H \to \infty$, the limit of the left hand side of this equation is zero. The right hand side is positive and decreasing in the urban share for $s_U \in (0,1)$, and

$$\lim_{s_U \to 1-} \frac{(1-s_U)^{1-\theta(1-\frac{1}{\epsilon})}\zeta}{s_U^{\frac{1}{\epsilon}}\theta L^{(1-\theta)(1-\frac{1}{\epsilon})}(1-\zeta)} = 0$$

So it must be that $s_U \to 1$. ■

A.4 Proof of Proposition 4

In this section I prove that if $\lim_{t\to\infty} \bar{h} = \infty$, $\lim_{t\to\infty} n \geq 1$ and $\epsilon > 1$, then the the limit of both urban and rural wages is $\bar{w} \equiv A\zeta^{\frac{\epsilon}{\epsilon-1}}$.

Proof. Use the final good production function (3) and equilibrium prices in equations (28) and (29) to express the equilibrium urban wage as:

$$w_U = A^{\frac{\epsilon-1}{\epsilon}}\zeta\left(\frac{A(\zeta x_U^{\frac{\epsilon-1}{\epsilon}} + (1-\zeta)x_R^{\frac{\epsilon-1}{\epsilon}})^{\frac{\epsilon}{\epsilon-1}}}{x_U}\right)^{\frac{1}{\epsilon}}$$

Substitute for intermediate inputs and express human capital inputs in terms of aggregate human capital and the urban share s_U:

$$w_U = A^{\frac{\epsilon-1}{\epsilon}}\zeta\left(\frac{A(\zeta(\tau_c s_U H)^{\frac{\epsilon-1}{\epsilon}} + (1-\zeta)((\tau_c(1-s_U)H)^{\theta}L^{1-\theta})^{\frac{\epsilon-1}{\epsilon}})^{\frac{\epsilon}{\epsilon-1}}}{\tau_c s_U H}\right)^{\frac{1}{\epsilon}}$$

Take the limit, given that the limits for $\bar{h}$ and n imply $H \to \infty$ and Proposition 3 implies $s_U \to 1$:

$$\lim_{t\to\infty} w_U = \lim_{t\to\infty} A^{\frac{\epsilon-1}{\epsilon}}\zeta\left(\frac{A(\zeta(\tau_c s_U H)^{\frac{\epsilon-1}{\epsilon}} + (1-\zeta)((\tau_c(1-s_U)H)^{\theta}L^{1-\theta})^{\frac{\epsilon-1}{\epsilon}})^{\frac{\epsilon}{\epsilon-1}}}{\tau_c s_U H}\right)^{\frac{1}{\epsilon}}$$

$$= A\zeta^{\frac{\epsilon}{\epsilon-1}} \equiv \bar{w}$$

■

B Regression Results

	Baseline	Proxy	Geo. Controls	Continent FE	Colony FE
Urban Share	283.3**		377.1***	151.7	282.5**
	(2.67)		(3.32)	(1.55)	(2.43)
Income Growth	-36160.9**	-61019.8***	-28379.7	-26879.2	-36329.0*
	(-2.23)	(-3.23)	(-1.28)	(-1.32)	(-1.96)
Population Growth	-5731.0	-6818.4	-4403.9	4348.0	-5717.0
	(-0.80)	(-1.03)	(-0.56)	(0.64)	(-0.78)
Log Pop. Density		14.68***			
		(3.24)			
Observations	29	29	29	29	29

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 4: Effects of 1500 CE Conditions on Growth Transition Year

	1000	1200	1300	1400	1500	1600	1700	1800
Urban Share	225.1	571.3**	556.5***	151.3	283.3**	251.0**	226.0**	69.69
	(1.03)	(2.72)	(3.08)	(1.32)	(2.67)	(2.43)	(2.38)	(0.72)
Income Growth	85115.3	-70168.6***	-76611.3***	-71015.8**	-36160.9**	-15482.4*	-21955.4**	-17511.8**
	(0.76)	(-3.02)	(-3.38)	(-2.46)	(-2.23)	(-1.93)	(-2.68)	(-2.45)
Population Growth	-43442.5	-6018.8	-3400.9	-10093.5	-5731.0	1053.8	-857.9	2152.7
	(-1.01)	(-0.58)	(-0.34)	(-0.82)	(-0.80)	(0.34)	(-0.27)	(0.76)
Constant	1865.1***	1903.9***	1892.1***	1922.2***	1865.6***	1851.1***	1863.5***	1872.9***
	(71.04)	(72.27)	(72.54)	(63.94)	(123.95)	(138.48)	(134.21)	(108.31)
Observations	18	18	18	18	29	29	32	35

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 5: Effects of Urbanization and Growth on Transition Timing: Many Initial Years

	mean	sd	min	max
Country Fixed Effects	-.010	.059	-.105	.171
Observations	6491			

Table 6: Summary of Estimated Urbanization Fixed Effects

Country Fixed Effects	496.5***
	(5.18)
Constant	1890.1***
	(298.83)
Observations	57

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 7: Impact of Estimated Urbanization Fixed Effects on Transition Timing

Notes: Standard Errors calculated by bootstrapping 500 times over 6,491 first stage observations.

C Survival Function

In this section I describe the estimation of the survival function. The one parameter version specification of the survival function is a transformed logistic cdf:

$$S_j(\bar{h}) = \bar{S} - (\bar{S} - S_{j,0})\frac{1 + \upsilon \bar{h}_0}{1 + \upsilon \bar{h}}$$

This function is able to hit both the initial target $S_{j,0}$ and the long run limit $\bar{S}$. It has all the desired properties: it is strictly increasing in $\bar{h}$, bounded by $[0, \bar{S}]$, and has finite limits as $\bar{h} \to 0$ and $\bar{h} \to \infty$.

The targets for $S_{R,0}$ and $S_{U,0}$ are from Clark (2009). I estimate the survival equation using nonlinear least squares. Child mortality data is from Johansson et al. (2015), and average income is used to approximate average human capital. Nonlinear least squares gives $\upsilon = 0.35$ when $\bar{h}_0$ is normalized to one. Figure 13 plots England's mortality data, income, and the fitted survival function given the year's income level.

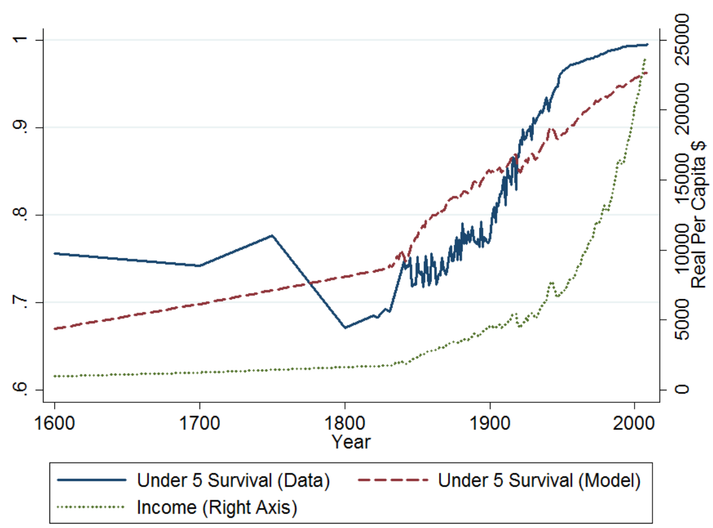

Figure 13: Empirical and Estimated Survival Rates

D Computation

In this section I describe my method of calculating the equilibrium. The strategy is to express the equilibrium allocation for each period t as a function of the rural choice of children $n_{R,t}$, and express the next period's choice $n_{R,t+1}$ as a function of period t variables. Then, an initial guess for $n_{R,0}$ is chosen, and a shooting algorithm is used to find the equilibrium value of $n_{R,0}$ and the following equilibrium allocations for all t.

First, it is useful to rewrite the location indifference condition (30) in terms of allocations instead of wages. This equation says that the right hand side of the Euler equation for urban and rural households is equal. This implies that the left hand side is also equal, so substituting with equation (21) implies:

$$w_R^{\sigma} n_R^{\sigma\phi+1}(1+g_R)^{1-\sigma} = w_U^{\sigma} n_U^{\sigma\phi+1}(1+g_U)^{1-\sigma} \tag{47}$$

Then dividing equation (30) by equation (47) yields:

$$\frac{\frac{1}{n_R}+\frac{1}{\xi}-\frac{1}{\alpha S_R}}{(1+g_R)^{1-\sigma}} = \frac{\frac{1}{n_U}+\frac{1}{\xi}-\frac{1}{\alpha S_U}}{(1+g_U)^{1-\sigma}} \tag{48}$$

Next, combine equation (48) with the normalized budget constraint (35) to yield an equation relating n_U, n_R, S_U, S_R, and parameters:

$$\frac{\frac{1}{n_R}+\frac{1}{\xi}-\frac{1}{\alpha S_R}}{(1+\xi\frac{1-\tau_c}{n_R}-\frac{\xi}{\alpha S_R})^{1-\sigma}} = \frac{\frac{1}{n_U}+\frac{1}{\xi}-\frac{1}{\alpha S_U}}{(1+\xi\frac{1-\tau_c}{n_U}-\frac{\xi}{\alpha S_U})^{1-\sigma}} \tag{49}$$

The shooting algorithm proceeds as follows. Guess a value of $n_{R,0}$. In period t, $n_{R,t}$, $S_{R,t}$, $S_{U,t}$, and the distribution of human capital Λ_t are known. In period $t=0$, $n_{R,0}$ is a guess, and $S_{R,0}$ and $S_{U,0}$ are calculated from the initial condition for Λ_0.

1. Numerically solve equation (49) for $n_{U,t}$. If the implied value of n_U is infeasible, the urban households must be constrained and their Euler equation doesn't hold, so set $n_U = (1-\tau_c)\alpha S_{U,t}$

2. Analytically solve the normalized budget constraints (35) for $g_{R,t}$ and $g_{U,t}$.

3. Calculate the wage premium $\frac{w_{U,t}}{w_{R,t}}$ from the indifference condition (30).

4. Numerically calculate the aggregate human capitals supplied $H_{R,t}$ and $H_{U,t}$ that are consistent with the wage ratio and the aggregate human capital supplied implied by Λ_t.

5. Analytically calculate the wages $w_{R,t}$ and $w_{U,t}$ implied by $H_{R,t}$ and $H_{U,t}$ using the equations for equilibrium prices (28) and (29).

6. Calculate next period's distribution of human capital Λ_{t+1} from the law of motion (25).

7. Use Λ_{t+1} to calculate next period's average human capital level and find $S_{R,t+1}$ and $S_{U,t+1}$ from equation (33).

8. Solve numerically for $n_{R,t+1}$:

(a) Express the next period's wage in location j as a function of $n_{j,t+1}$ through the Euler equation (21)

(b) Express next period's human capitals supplied $H_{R,t+1}$ and $H_{U,t+1}$ as functions of $n_{R,t+1}$ and $n_{U,t+1}$, using the equations for equilibrium prices (28) and (29).

(c) Numerically find the values of $n_{R,t+1}$ and $n_{U,t+1}$ that imply values of $H_{R,t+1}$ and $H_{U,t+1}$ that are consistent with Λ_{t+1}.

(d) If $n_{U,t+1}$ is infeasible, urban households must be constrained, so repeat steps (b) and (c) assuming $n_{U,t+1} = (1 - \tau_c)\alpha S_{U,t+1}$.

9. Return to step 1. for period $t + 1 \leq T$.

Period T approximates the long run. If the calculated long run rural children $n_{R,T}$ is within tolerance ε to the equilibrium long run value $\bar{n}$, consider the equilibrium solved. Otherwise, for $n_{R,T} > \bar{n} + \varepsilon$ revise the initial guess downwards, and for $n_{R,T} < \bar{n} + \varepsilon$ revise the initial guess upwards.

Country Name	*Transition Correlations: Table 1*	*Baseline Regressions: Table 4*	*Two-Stage Regression: Table 7*
Algeria			X
Armenia			X
Australia		X	X
Austria	X	X	X
Azerbaijan			X
Bangladesh			X
Belgium	X	X	X
Benin			X
Brazil			X
Canada		X	X
Chile			X
China	X	X	X
Czechoslovakia			X
Denmark	X	X	X
Egypt		X	X
England	X	X	X
Finland	X	X	X
France	X	X	X
Germany	X	X	X
Greece	X	X	X
India	X	X	X
Indonesia			X
Iran		X	X
Iraq		X	X
Ireland			X
Italy	X	X	X
Jamaica			X
Japan		X	X
Jordan			X
Kazakhstan			X
Lebanon			X
Lithuania			X
Malaysia			X
Mexico		X	X
Morocco		X	X
Nepal			X
Netherlands	X	X	X
New Zealand		X	X
North Korea			X
Norway	X	X	X
Paraguay			X
Philippines			X
Portugal	X	X	X
Russia	X	X	X
South Korea			X
Spain	X	X	X
Sri Lanka			X
Sweden	X	X	X
Switzerland	X	X	X
Syria			X
Taiwan			X
Thailand			X
Tunisia			X
Turkey		X	X
United States		X	X
Venezuela			X
Vietnam			X

Table 8: Countries Included in Various Estimations