


RESEARCH PAPER

Urbanization, long-run growth, and the demographic transition

Jonathan J. Adams* 

Department of Economics, University of Florida, PO Box 117140, Gainesville, FL 32611, USA

*Corresponding author. E-mail: adamsjonathan@ufl.edu

(Received 11 May 2018; revised 1 December 2020; accepted 2 December 2020;
first published online 3 March 2021)

Abstract

Advanced economies undergo three transitions during their development: (1) transition from a rural to an urban economy, (2) transition from low-income growth to high-income growth, (3) transition from 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.

Key words: Demographic transition; fertility; growth; mortality; structural change; urbanization

JEL Codes: E13; J11; N10; O18; O41

1. Introduction

Why do economies transition from millennia 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. Literature following Becker *et al.* (1990) and Galor and Weil (2000) theorizes that the central mechanism is the substitution of child quantity to child quality, and jointly explains the growth transition and the demographic transition. Simultaneously, literature following Hansen and Prescott (2002) and Lucas (2004) theorizes that the central mechanism is structural transformation and jointly explains 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.

© Université catholique de Louvain 2021. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 association of early urbanization with lower modern incomes is novel to applied theories of very long-run growth and demographics,¹ and is closely related to the empirical literature's concept of the "reversal of fortune", whereby high preindustrial income is associated with slower growth. Indeed, Acemoglu *et al.* (2002) specifically use urbanization as a proxy for early income and show it is negatively correlated with modern income levels for former colonies. Their explanation for this relationship is the colonial transmission of institutions. This contrasts with the present paper in two ways. First, I am concerned with the relationship between growth and urbanization directly, not just as a proxy for income. Second, the explanation in this paper is that high child mortality in early urban centers disincentivizes human capital growth. This is an independently important effect, which I demonstrate in section 2.2, showing that early urbanization is associated with delayed growth, even when controlling for the alternative explanations of colonial history and geographic factors [e.g., Diamond (1998), Acemoglu *et al.* (2005) or Nunn and Qian (2011)]. In the long run, urbanization is beneficial for growth; a large literature supports this, and it is true in this paper's model as well. However, the factors that cause preindustrial civilizations to be more urbanized are associated with delayed transitions to modern growth.

The model economy has two sectors.² Human capital growth drives production to shift out of the rural sector, which has diminishing returns to scale.³ The higher returns to scale of the urban sector increase the income growth associated with any 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 a 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

¹Low-income economies leapfrogging higher income economies are a feature of many well-known models [e.g., Brezis *et al.* (1993) or Brezis and Krugman (1997)], but not a prediction of existing theories of very long-run growth that can simultaneously explain why countries transition from preindustrial to modern growth.

²Trade 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. O'Rourke 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.

³This 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 >1 .

income effect dominates the substitution effect, so households shift from child quantity to child quality.⁴ 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.⁵ As human capital grows, increased knowledge reduces mortality. Declines in the difference between urban and rural mortality reduce 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 introducing 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)], in the American South [Bleakley and Lange (2009)], and across the developing world in the 20th century [Chatterjee and Vogl (2018)].

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) assume 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. This joins a growing literature arguing that child mortality improvements are central to the transition to modern growth. The exact mechanism—child mortality's effect on the return to human capital—differs from other papers in this literature. For example, Kalemli-Ozcan (2002) and Kalemli-Ozcan (2008) show that reductions in child mortality induce substitution from child quantity to quality by reducing the precautionary motive to have many children. Bhattacharya and Chakraborty (2017) find that mortality improvements can speed the adoption of modern contraception, which is

⁴This Giffen property of child quantity is not entirely 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.

⁵Throughout 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 Keszenbaum and Rosenthal (2011) for France, Hanlon and Tian (2015) for China, and Cain and Hong (2009) in the United States.

complementary to substituting towards child quality. Ehrlich and Lui (1991) argue that reductions in mortality raise the incentive for parents to invest in their children's human capital, as longevity improvements raise the value of future old-age support from their offspring. Other papers suggesting that child mortality improvements drive fertility declines include Eckstein *et al.* (1999), Lagerlof (2003), Hazan and Zoabi (2006), and Bhattacharya and Chakraborty (2012).⁶

The hypothesis that child mortality is fundamental to the growth transition is not without controversy. [Galor (2011), Chapter 4] rejects this channel on theoretical grounds. 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) and Strulik (2017) reach a similar conclusion. However, the model described in section 3 differs from this conclusion when preferences are dynastic, and households invest in each child's human capital, even with balanced growth compatibility. 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. This is true of the crude death rate, but the relevant measure is the child mortality rate, which Wrigley and Schofield (1983) document as not declining significantly over the same period (Figure 1).

The 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 the land as an input and is perfectly substitutable with a constant returns sector. Given the 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.⁷ Agriculture makes up the majority of employment in preindustrial 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.⁸

⁶Some 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)].

⁷A 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 a simultaneous income boom.

⁸Recent 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.

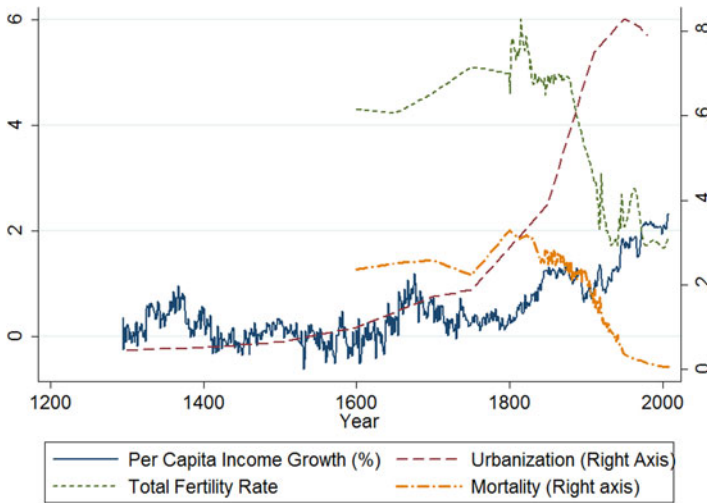


Figure 1. Transitions in England.

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).

The intersection of these two broad growth literatures—quantity-quality substitution and structural transformation—is limited. The present paper argues that the intersection is important for understanding long-run transitions and the interaction between the two forces generate effects that cannot be observed when considered independently. Few papers populate this intersection, but in an important related paper, Baudin and Stetler (2018) also consider a growth model with urban and rural differences in demographic decisions; they use the framework to show that migration costs can slow an economy’s transition and increase urban-rural inequality.

The remainder of this paper is organized as follows: section 2 describes the empirical patterns, 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 $<1\%$. The urban share of people is $<10\%$. 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 1800.⁹

⁹Except for Belgium and the Netherlands, which had urban shares nearly 30% in 1500 CE.

Table 1. Correlation of transition years

| | 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 2. Transitioned percentage of countries by income in 2012

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

Moreover, these transitions occur around the same time within a country. This is well-known, but to illustrate, I calculate the first year that each country surpasses a benchmark level for each series: (a) 1% annual income growth, (b) 50% urban, (c) total fertility rate below 3, and (d) under-five child mortality below 5%. [Table 1](#) reports the correlation table for these transition years.¹⁰ 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 <\$ 1, 000 tend to be rural with high fertility and high mortality.

2.1 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.

The urban-rural wage gap declines over time.¹¹ Williamson (1987) calculates a nominal wage gap in the 1830s for unskilled workers of 73% and a real wage gap of

¹⁰The set of countries includes those with 1,000 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. In total, 18 countries are in this dataset: China, India, and 16 European countries, listed in [Table 5](#). Historical estimates for these datasets correspond to the modern states' current geographic area whenever possible. The annual income growth rate is a 50-year moving average, so that recent years with high frequency data correspond to the early Maddison estimates with low frequency data points.

¹¹Specifically, 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.

46%; he estimates that the majority of the gap was compensating for high urban mortality. In contrast, D'Costa and Overman (2013) estimate an unconditional wage gap of 14% in Britain from 1998 to 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 is 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 Division (2012), Table 9]. This pattern is documented in many countries.¹³

2.2 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 to scale 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 to scale, but an additional factor in the urban sector inhibits growth: high child mortality. If an early economy is relatively urbanized all else equal, its high child mortality 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.

In this section, I estimate the relationship between early urbanization and transition timing using cross-country data to document how country characteristics including the preindustrial urbanization rate affect the timing of a country's transition to modern growth.

I construct growth transition years for 43 countries¹⁴ for which I have the relevant data in the year 1500. The transition years use the same definitions as in Table 1: the first year that a 50-year moving average of income growth exceeds 1%. Then I regress the transition years T_j against country characteristics in the year 1500:

$$T_j = \beta_0 + \beta_1 s_{U,0,j} + \beta_2 \Delta y_{0,j} + \beta_3 n_{0,j} + \alpha' D_j + \varepsilon_j, \quad (1)$$

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, D_j is a vector of country characteristics for some regression specifications, and ε_j is the error term. One

¹²Additionally, there is cross-sectional evidence that the urban-rural productivity gap is declining in income, and nearly disappears in rich countries [Gollin *et al.* (2014)].

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

¹⁴Listed in column 2 of Table 5.

conclusion of the sensitivity analysis in section 6.1 is that initial income and population growth rates must be controlled for in these regressions, for they are associated with other factors that affect the transition timing, such as the productivity of human capital investment and preference for children.

Income and population data are from The Maddison Project (2013).¹⁵ For comparability, England's data is also from this source, instead of the superior data used in section 5.1's model calibration. 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 and population data are annual for most countries. Finally, urbanization data is from Bairoch *et al.* (1988) and The Clio Infra Project (2016), interpolated over gap years.¹⁶

Table 3 reports the baseline results in Column (1). 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 additional 10 percentage points of urbanization should delay the growth transition by 25 years, all else equal. Both the urbanization and income growth rate coefficients are significant at the 5% level or lower, but population growth is not, which is the case for almost every specification of these regressions.

Table 3 also reports the results of several robustness checks. Column (2) reports the results with no controls, which gives a weaker relationship. Column (3) 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. Column (4) includes a vector of geographic controls¹⁷ considered by Ashraf and Galor (2011). The effect of urbanization is strengthened in this regression and is significant at the 1% level. Column (5) includes continent fixed effects, which weakens the relationship, although this may be because continents are correlated with colonial status.

To demonstrate that the effect of urbanization on transition timing is independent of the colonial institution channel documented by Acemoglu *et al.* (2002), I next run regressions with dummies for colonial history. Specifically, I include a dummy to indicate whether countries were colonized, as well as a dummy to indicate whether countries were colonizers before the industrial revolution (Tables 4 and 5).¹⁸ The

¹⁵For 14 of the countries in the baseline 1500 CE sample, I have country-specific estimates of all controls except for income. For these countries, I instead use Maddison's regional income estimates when controlling for initial income growth. This applies to the 7 Eastern European countries, and 7 of the African countries (but not Egypt).

¹⁶Urbanization is defined as the share of the population living in a city with at least 5,000 people. The Clio Infra Project (2016) dataset is used for all countries in the 1500-1800 CE year samples, except for China, for which the urbanization definition is inconsistent. Bairoch *et al.* (1988) is used for China, as well as for all samples before the year 1500 CE.

¹⁷Absolute 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.

¹⁸Table 4 considers alternative specifications without the colonizer dummy, as well as alternative colonial classifications. In the baseline, I define as colonizers only those countries with large scale colonies before the Industrial Revolution: England, France, Portugal, Spain, Belgium, and the Netherlands. This excludes countries that acquired large colonies after their growth transitions such as Germany and the United States, as well as countries with only limited colonial holdings such as Sweden. Belgium is included as it was a part of the Netherlands during its growth transition. Countries classified as colonies comprise all of the non-European countries listed in the second column of Table 5, except for Iran, Iraq, Japan, and

Table 3. Effects of 1500 CE conditions on growth transition year

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---------------------|--------------|--------|--------------|-----------|-------------|-------------|-----------|-------------|
| Urban share | 248.5** | 178.0* | | 299.6*** | 165.0* | 296.7** | 304.6** | 245.4** |
| | (2.55) | (1.90) | | (3.07) | (1.79) | (2.51) | (2.67) | (2.32) |
| Income growth | -44,807.9*** | | -66,116.1*** | -25,589.9 | -39,578.7** | -38,997.8** | -26,533.9 | -43,803.3** |
| | (-3.05) | | (-4.10) | (-1.49) | (-2.13) | (-2.04) | (-1.46) | (-2.32) |
| Population growth | -606.6 | | -4,578.5 | -3,322.3 | 4,140.0 | 755.1 | -1,609.1 | 5,493.2 |
| | (-0.10) | | (-0.86) | (-0.49) | (0.71) | (0.11) | (-0.24) | (0.95) |
| Log Pop. density | | | 14.47*** | | | | | |
| | | | (3.62) | | | | | |
| Geographic controls | | | | X | | | X | |
| Continent dummies | | | | | X | | | X |
| Colonial dummies | | | | | | X | X | X |
| Observations | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 43 |
| R ² | 0.298 | 0.081 | 0.387 | 0.479 | 0.501 | 0.312 | 0.545 | 0.562 |

t statistics in parentheses.

*p < 0.1, **p < 0.05, ***p < 0.01.

Table 4. Alternative colonial classifications

| | Baseline classification | | | | | Expanded colonizers | | |
|---------------------|-------------------------|-----------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Urban share | 247.0** (2.40) | 243.5** (2.46) | 246.1** (2.43) | 288.7** (2.49) | 254.9** (2.63) | 306.4*** (2.78) | 274.4** (2.59) | 277.4** (2.69) |
| Income growth | -45,277.0** (-2.57) | -33,239.9* (-1.94) | -45,947.4** (-2.48) | -42,645.5** (-2.23) | -36,289.6** (-2.07) | -32,558.8* (-1.81) | -42,458.2** (-2.24) | -33,661.6* (-1.86) |
| Population growth | -615.9 (-0.10) | -2,753.5 (-0.42) | -670.8 (-0.10) | 289.6 (0.04) | -3,285.9 (-0.50) | -2,423.7 (-0.37) | 327.1 (0.05) | -2,486.6 (-0.37) |
| Geographic controls | | X | | | X | X | | X |
| Colonizer dummies | | | | X | | X | X | X |
| Observations | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 43 |
| R^2 | 0.298 | 0.529 | 0.298 | 0.310 | 0.529 | 0.543 | 0.315 | 0.536 |

All of these regressions include dummies for countries that were colonized but vary who receives a dummy for being a colonizer. The “Expanded Colonizers” add Turkey, Germany, Russia, and the United States to the baseline set. This classification adds no additional colonies in the set of observations over the inclusion of just Turkey, so regressions with no colonizer dummy for this classification are omitted.

t statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5. Countries included in various estimations

| Country name | Transition correlations: Table 1 | Baseline regressions: Table 3 | Two-stage regression: Table 10 |
|----------------|--|---|--|
| Albania | | X | X |
| Algeria | | X | X |
| Angola | | | X |
| Armenia | | | X |
| Australia | | X | X |
| Austria | X | X | X |
| Azerbaijan | | | X |
| Bangladesh | | | X |
| Belgium | X | X | X |
| Benin | | | X |
| Brazil | | | X |
| Bulgaria | | X | X |
| Canada | | X | X |
| Cape Verde | | | X |
| Chile | | | X |
| China | X | X | X |
| Costa Rica | | | X |
| Cuba | | | X |
| Czechoslovakia | | X | X |
| Denmark | X | X | X |
| Egypt | | X | X |
| El Salvador | | | X |
| England | X | X | X |
| Estonia | | | X |
| Finland | X | X | X |
| France | X | X | X |
| Germany | X | X | X |
| Greece | X | X | X |
| Honduras | | | X |
| Hungary | | X | |
| India | X | X | X |
| Indonesia | | | X |
| Iran | | X | X |

(Continued)

Table 5. (Continued.)

| Country name | Transition correlations: Table 1 | Baseline regressions: Table 3 | Two-stage regression: Table 10 |
|--------------|--|---|--|
| Iraq | | X | X |
| Ireland | | | X |
| Italy | X | X | X |
| Jamaica | | | X |
| Japan | | X | X |
| Jordan | | | X |
| Kazakhstan | | | X |
| Laos | | | X |
| Latvia | | | X |
| Lebanon | | | X |
| Lithuania | | | X |
| Libya | | X | |
| Madagascar | | X | |
| Malaysia | | | X |
| Mexico | | X | X |
| Morocco | | X | X |
| Mozambique | | X | |
| Nepal | | | X |
| Netherlands | X | X | X |
| New Zealand | | X | X |
| Nicaragua | | | X |
| North Korea | | | X |
| Norway | X | X | X |
| Paraguay | | | X |
| Philippines | | | X |
| Poland | | X | |
| Portugal | X | X | X |
| Russia | X | X | X |
| Somalia | | X | |
| South Africa | | X | |
| South Korea | | | X |
| Spain | X | X | X |
| Sri Lanka | | | X |

(Continued)

Table 5. (Continued.)

| Country name | Transition correlations: Table 1 | Baseline regressions: Table 3 | Two-stage regression: Table 10 |
|---------------|----------------------------------|-------------------------------|--------------------------------|
| Sudan | | | X |
| Sweden | X | X | X |
| Switzerland | X | X | X |
| Syria | | X | X |
| Taiwan | | | X |
| Thailand | | | X |
| Tunisia | | X | X |
| Turkey | | X | X |
| Uganda | | | X |
| United States | | X | X |
| Uruguay | | | X |
| Venezuela | | | X |
| Vietnam | | | X |
| Yugoslavia | | X | |

regression in Column (6) of Table 3 includes these colonial fixed effects and estimates a larger effect than in the baseline that is significant at the 5% level. Column (7) includes both colonial fixed effects and geographic controls, demonstrating that the urbanization channel appears robust, even when controlling for both colonial and geographic explanations of reversals of fortune. Finally, Column (8) includes colonial fixed effects as well as continent fixed effects and gives a statistically significant coefficient unlike in Column (5) when the colonial status was not accounted for.

The year 1500 CE is used to initialize the baseline calibration in section 5.1 because it is the earliest period for which the rich Clio Infra dataset gives urbanization estimates. But the empirical effects of urbanization and income growth on transition timing can be examined for other years. Table 6 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, or in 1000 CE when the sample is small and the data especially poor.

I also consider alternative measurements of the growth transition timing. The baseline is the first year that the annual growth trend exceeds 1%. Table 7 reports the baseline regression for other thresholds in Columns (2)–(4). Above 1.5%, the relationship is not robust, suggesting that when countries become sufficiently developed, modern factors may overwhelm the early effects. Column (5) reports the regression where transition timing is defined as the year a country passes an income threshold, rather than a growth threshold: US\$5,000 in 2008. This effect is also statistically significant at the 5% level. Finally, Columns (6)–(7) report the effects of

Turkey. Iraq is excluded because British colonial rule was barely a decade. However, it is included in alternative specifications where Turkey is considered a colonizer.

Table 6. Effects of urbanization and growth on transition timing: many initial years

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|-------------------|-----------------------|-------------------------|-------------------------|-------------------------|-------------------------|------------------------|-------------------------|------------------------|
| | 1000 | 1200 | 1300 | 1400 | 1500 | 1600 | 1700 | 1800 |
| Urban share | 196.1 (1.06) | 479.7** (2.70) | 463.0** (2.98) | 149.4 (1.49) | 248.5** (2.55) | 211.5** (2.35) | 182.1** (2.31) | 41.25 (0.53) |
| Income growth | 75,750.4 (0.80) | -63,562.3*** (-3.27) | -68,919.9*** (-3.58) | -69,323.2*** (-2.89) | -44,807.9*** (-3.05) | -17,095.4** (-2.30) | -23,501.9*** (-3.18) | -15,826.0** (-2.60) |
| Population growth | -36,317.6 (-1.13) | -1,639.8 (-0.19) | 1,782.0 (0.21) | -9,173.8 (-0.98) | -606.6 (-0.10) | 629.7 (0.23) | -143.4 (-0.05) | 1,212.1 (0.50) |
| Constant | 1,868.9*** (93.35) | 1,893.5*** (90.58) | 1,881.9*** (86.10) | 1,918.9*** (88.66) | 1,871.4*** (148.22) | 1,865.0*** (174.99) | 1,873.6*** (175.85) | 1,879.9*** (137.15) |
| Observations | 22 | 22 | 22 | 22 | 43 | 43 | 46 | 46 |
| R^2 | 0.159 | 0.485 | 0.514 | 0.354 | 0.298 | 0.163 | 0.240 | 0.200 |

t statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7. Effects of urbanization and growth on transition timing: alternate measures

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|-------------------|--------------|-------------|-----------|----------|-----------|--------------|----------------|
| | 1% | 1.2% | 1.4% | 1.6% | 5000 USD | 3 TFR | 5% Child Mort. |
| Urban share | 248.5** | 260.5** | 235.3** | 118.9 | 115.6** | 108.1* | 102.4** |
| | (2.55) | (2.70) | (2.36) | (1.23) | (2.08) | (1.87) | (2.04) |
| Income growth | -44,807.9*** | -31,513.2** | -25,288.7 | -9,323.5 | -13,591.8 | -37,716.8*** | -16,710.2** |
| | (-3.05) | (-2.12) | (-1.64) | (-0.62) | (-1.54) | (-4.11) | (-2.10) |
| Population growth | -606.6 | -5,817.3 | -3,810.7 | -8,907.7 | -4,563.9 | -9,540.9** | -4,670.8 |
| | (-0.10) | (-0.92) | (-0.58) | (-1.40) | (-1.17) | (-2.47) | (-1.39) |
| Observations | 43 | 42 | 42 | 41 | 33 | 39 | 38 |
| R^2 | 0.298 | 0.316 | 0.228 | 0.165 | 0.300 | 0.553 | 0.348 |

t statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

early urbanization on the demographic transitions: fertility and child mortality. As predicted, early urbanization delays the demographic transitions just as it delays the growth transition, although the effect on total fertility is only significant at the 10% level. The regressions in Columns (5)–(7) have fewer observations because several countries in the baseline sample have not yet met the relevant thresholds and so are excluded from the regressions.

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. The land 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:

$$F_R(\tilde{h}, \tilde{l}) = \tilde{h}^{\theta-1} \tilde{l}^{1-\theta}. \tag{2}$$

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 , uses only human capital to linearly produce output. Its production function is

$$F_U(\tilde{h}) = \tilde{h}. \tag{3}$$

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 the elasticity of substitution ε and weighting parameter ζ

$$F(\tilde{x}_R, \tilde{x}_U) = A(\zeta \tilde{x}_U^{(\varepsilon-1)/\varepsilon} + (1 - \zeta) \tilde{x}_R^{(\varepsilon-1)/\varepsilon})^{\varepsilon/(\varepsilon-1)}. \tag{4}$$

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. All sectors feature free entry of firms. 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. 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:

$$\max_{\tilde{h}, \tilde{l}} p_R \tilde{h}^{\theta-1} \tilde{l}^{1-\theta} - w_R \tilde{h} - r \tilde{l}. \tag{5}$$

An urban firm solves:

$$\max_{\tilde{h}} p_U \tilde{h} - w_U \tilde{h}. \tag{6}$$

A final goods firm solves:

$$\max_{\tilde{x}_R, \tilde{x}_U} A(\zeta \tilde{x}_U^{(\varepsilon-1)/\varepsilon} + (1 - \zeta) \tilde{x}_R^{(\varepsilon-1)/\varepsilon})^{\varepsilon/(\varepsilon-1)} - p_R \tilde{x}_R - p_U \tilde{x}_U. \tag{7}$$

3.2 Households

Individuals live for two periods: in their first period of life, they are children and in the second period they are parents.¹⁹ 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; rather, I suppose that an infinitesimally small fraction of the population holds all the land and has a negligible impact on aggregate human capital and demographics. This is a useful simplification to avoid keeping track of the distribution of land ownership in addition to the other state variables of the model.²⁰

Utility is dynastic, formulated as in Razin and Ben-Zion (1975). Parents enjoy present consumption c , their number of surviving children n , which for tractability is not restricted to integers. Parents also care about their dynasty’s future wellbeing, represented by discounting the average utility of each child. A parent discounting by β has utility:

$$V_t = u(c_t, n_t) + \beta(s_{t+1}^U V_{U,t+1} + (1 - s_{t+1}^U)V_{R,t+1}). \tag{8}$$

where $u(c_t, n_t)$ is the period utility function, V_t is the parent’s dynastic utility and $s_{t+1}^U V_{U,t+1} + (1 - s_{t+1}^U)V_{R,t+1}$ is the average dynastic utility of the next generation. In this expression, s_{t+1}^U is the share of the household’s children who will choose to live in the urban location, $V_{U,t+1}$ is the dynastic utility of children who will choose the urban location, $1 - s_{t+1}^U$ is the share who will choose to live in the rural location and $V_{R,t+1}$ is the dynastic utility of children who will choose the rural location. Parents’ preference for quantity of children is driven by their period utility, $u(c_t, n_t)$, because V_{t+1} is the average utility of the children, not the total utility of the next generation.²¹

The period utility function $u(c, n)$ increases 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):

$$u(c, n) \equiv \frac{(cn^\phi)^\sigma}{\sigma}, \tag{9}$$

¹⁹Because 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 trade-off affects the quantity-quality, even when controlling for child mortality.

²⁰An alternative approach to avoid tracking land wealth would be to follow Galor and Weil (2000) and let workers earn their average product instead of their marginal product.

²¹The Razin and Ben-Zion (1975) structure contrasts to Becker and Barro (1988), in which the discount factor is a concave function of n_t . I choose the Razin and Ben-Zion (1975) structure 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.

where $\phi > 0$, $\sigma < 1$ and $\phi\sigma < 1$. ϕ controls the preference for consumption relative to children, while σ controls substitutability across generations: $1/(1 - \sigma)$ 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:

$$\tau_c + \tau_n + \tau_h = 1. \tag{10}$$

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:

$$c = w_j h \tau_c. \tag{11}$$

A household choosing time τ_n produces n surviving children by:

$$n = S_j \alpha \tau_n, \tag{12}$$

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.

Parents produce education to increase the human capital of their children. The education produced per child k is denoted by d_k . Households may choose to endow children going to different locations with different education levels (although they will not in equilibrium). Therefore $d_U s^U n$ is the total education for children headed to urban locations, while $d_R(1 - s^U)n$ is the total education for children headed to rural locations. All child mortality resolves before parents start to invest in their human capital so the number of surviving children n affects the allocation of education, rather than the gross number of children.²² Total education produced is linear in the time spent educating τ_h , and the productivity of parental time in producing education is proportional to parental human capital h . With the productivity parameter ξ , education is given by

$$d_U s^U n + d_R(1 - s^U)n = \xi \tau_h h. \tag{13}$$

A child's future human capital is increasing in the education it receives. Children are also endowed with their parents' human capital during the childrearing process. This captures the distinction that only some human capital accumulation is an economic decision (education) while other accumulation occurs naturally. Human capital accumulation is assumed to be linear, so for a child who will choose location k , its future human capital h'_k is given by

$$h'_k = d_k + h. \tag{14}$$

²²Tamura (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.

The endowment of parental human capital ensures that human capital growth is non-negative, even if households are constrained, in which there is zero education and thus zero human capital growth. Without this lower bound on human capital accumulation, there is a potential for inescapable poverty traps and other equilibria.

Combining equations (10), (11), (12), and (13) yield the combined budget constraint:

$$c + \frac{w_j n}{\xi} (d_U s^U + d_R (1 - s^U)) + \frac{w_j h n}{\alpha S_j} = w_j h. \tag{15}$$

The household’s time is used for consumption, education, or producing children. The total value in the 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 total cost of time spent investing in human capital is $((w_j n)/\xi)(d_U s^U + d_R (1 - s^U))$, and the total cost of time spent producing n children is $w_j h n / \alpha S_j$. Production of these goods is linear, so the marginal cost of producing an additional unit of human capital per child is $w_j n / \xi$ while the marginal cost of producing an additional child is $(w_j / \xi)(d_U s^U + d_R (1 - s^U)) + (w_j h / \alpha S_j)$.²³ Crucially, child mortality and education affect the marginal cost of additional children, but not the marginal cost of human capital. As child mortality improves and education rises relative to the human capital stock, the household is incentivized to substitute from child quantity towards child quality.

3.2.1 The households’ problem

The household’s problem differs depending on whether the parent was born in a rural or urban location. The rural-born households’ problem is to choose location j , consumption c , children n , and the education d_k and future human capital h'_k of their children who will choose location k , to maximize dynastic utility. Let Λ denote the aggregate state of the economy, and s^U the share of children choosing the urban location; then the household’s Bellman equation is

$$V^R(h; \Lambda) = \max_{c, n, d_U, d_R, h'_U, h'_R, j \in J} u(c, n) + \beta (s^U V(h'_U; \Lambda') + (1 - s^U) V(h'_R; \Lambda')), \tag{16}$$

subject to the human capital accumulation equations (14), budget constraint (15), location choice set $j \in U, R$, and non-negativity constraints:

$$c \geq 0 \quad n \geq 0 \quad d_U \geq 0 \quad d_R \geq 0. \tag{17}$$

Urban-born households face a simpler problem: they do not choose location. I assume there is no reverse migration from urban to rural locations in order to pin down the distribution of households.²⁴ The Bellman equation of an urban-born

²³Why does h enter the marginal cost of children but not the marginal cost of education? Parents give up $w_j h$ income per unit time spent on either activity, but h also increases their productivity at producing education, exactly offsetting so that h does not affect the marginal cost of education.

²⁴Without 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. Of course,

household is

$$V^U(h; \Lambda) = \max_{c,n,d_U,h'_U} u(c, n) + \beta V(h'_U; \Lambda'), \tag{18}$$

subject to constraints (14), (15), and (17). In this case, the urban share of the households' children is necessarily $s^U = 1$, given the assumption of no reverse migration.

Solving the households' problems yields the first-order conditions:

$$u_n(c, n) = u_c(c, n) \left(\frac{w_j h'}{\xi} + \frac{w_j h}{\alpha S_j} \right), \tag{19}$$

$$u_c(c, n) w_j n \geq \xi \beta V'(h'_U; \Lambda'), \tag{20}$$

$$u_c(c, n) w_j n \geq \xi \beta V'(h'_R; \Lambda'), \tag{21}$$

and envelope condition:

$$\frac{\partial}{\partial h} V^j(h; \Lambda) = u_c(c, n) w_j \left(1 + \frac{n}{\xi} - \frac{n}{\alpha S_j} \right). \tag{22}$$

The first-order conditions (20) and (21) hold with equality when the household is unconstrained in its choice of d_U and d_R , respectively.

When the preferences in (9) are applied to first-order condition (19), consumption is a constant share of income:

$$\frac{c}{w_j h} = \frac{1}{1 + \phi}. \tag{23}$$

This also implies that $\tau_c = (1/(1 + \phi))$ 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.

The household has a Euler equation for children choosing each location. Combining the first-order conditions (20) and (21) with the envelope condition (22) gives the Euler equation for children choosing location k :

$$\left(\frac{c'_k}{c} \right)^{1-\sigma} \geq \left(\frac{n'_k}{n} \right)^{\phi\sigma+1} \frac{w'_k}{w_j} \xi \beta \left(\frac{1}{n'_k} + \frac{1}{\xi} - \frac{1}{\alpha S'_k} \right). \tag{24}$$

This Euler equation reveals how child mortality affects the incentive to invest in child quality. When the future survival rate S'_k is higher, it increases the budget set in the next period by making children less costly, so that households can consume more with the

assuming that dynasties move from rural to urban areas and never return 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. Baudin and Stelter (2018) consider the implications when migration costs exist and the migration rate is no longer undetermined.

same level of human capital. Thus, an increase to S'_k increases the return on education, which appears on the right-hand side of the Euler equation.

Denote human capital growth by $1 + g_k \equiv (h'_k/h)$. Then the Euler equation can be rewritten using the budget constraint and consumption share in terms of fertilities, human capital growth, and wages:

$$(1 + g_k)^{1-\sigma} \left(\frac{n}{n'_k}\right)^{\phi\sigma} \left(\frac{w_j}{w'_k}\right)^\sigma \geq \beta \frac{\xi}{n} \left(\tau_c + n'_k \frac{1 + g'_k}{\xi}\right). \tag{25}$$

The left-hand side of equation (25) is marginal utility growth across generations. On the right-hand side, $\tau_c + n'_k((1 + g'_k)/\xi)$ is the return to human capital investment, and ξ/n is the productivity of parental time at producing human capital for each child. The Euler equation holds with equality when households are unconstrained. The right-hand side is the marginal benefit of investing more parental time into education. This benefit decreases in n because when a household has more children, it requires more time to invest each child with a unit of education.

3.3 Aggregates and the distribution of human capital

The state of the economy is determined by the function $\lambda(h)$, which denotes the measure of households with human capital h . Households are not ex ante heterogeneous; all heterogeneity in this model is captured by the distribution $\lambda(h)$.

Human capital is distributed heterogeneously because dynasties live different amounts of time in different locations and households in different locations make different choices about child quantity versus quality. This heterogeneity in investment rates across locations is the central mechanism through which urbanization and income growth interact. As often is the case when savings rates are correlated with wealth, households in this model do not aggregate to a representative household, because urban and rural households invest in human capital at different rates and their levels of human capital are potentially correlated with their location. This nontrivial heterogeneity is necessary to study the interaction between the different transitions but adds complexity over other theories of long-run growth with urban and rural locations that admit representative households, such as Laitner (2000), Hansen and Prescott (2002), or Tamura (2002).

In particular, the distribution of human capital $\lambda(h)$ is necessary to track in order to characterize the distribution of people across locations. Market clearing and optimization pin down the allocations of aggregate human capital to urban and rural sectors, but in order to map aggregate human capital stocks into shares of the population living in each location, $\lambda(h)$ must be known.

The total population in the economy N is:

$$N = \int_0^\infty \lambda(h)dh. \tag{26}$$

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:

$$H_j = \int_0^\infty \tau_c h \lambda(h, j) dh, \tag{27}$$

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

$$Y = w_U H_U + w_R H_R + rL. \tag{28}$$

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:

$$\lambda(h') = \sum_{j \in \{U, R\}} n_j \lambda(h(h', j), j), \tag{29}$$

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} :

$$\bar{h} = \int_0^\infty \frac{h \lambda(h)}{N} dh. \tag{30}$$

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.²⁵

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

$$\lim_{\bar{h} \rightarrow \infty} S_j(\bar{h}) = \bar{S}. \tag{31}$$

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.

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 (5), (6), and (7).

²⁵See 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.

- (2) The household allocations maximize (16) and (18) subject to (14), (15), and (17).
- (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 (29) holds for all human capital levels.
- (5) Household aggregates add up, satisfying equations (26), (27), (28), and (30)

4.2 Equilibrium prices

The firms’ profit maximization [equations (5), (6), and (7)] implies that equilibrium prices must relate to equilibrium factors by:

$$w_U = p_U \quad w_R = p_R \theta (H_R)^{\theta-1} L^{1-\theta} \quad r = p_R (1 - \theta) (H_R)^\theta L^{-\theta} \tag{32}$$

$$p_U = A^{(\epsilon-1)/\epsilon} \zeta \left(\frac{Y}{x_U} \right)^{1/\epsilon} \quad p_R = A^{(\epsilon-1)/\epsilon} (1 - \zeta) \left(\frac{Y}{x_R} \right)^{1/\epsilon} . \tag{33}$$

The urban sector has linear production, so the free entry condition ensures that $w_U = p_U$ holds in equilibrium.

4.3 Equilibrium location choice

Rural-born 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 (22) equal in both locations, and substituting for consumption by equation (23) yields:

$$w_R^\sigma n_R^{\sigma\phi+1} \left(\frac{1}{n_R} + \frac{1}{\xi} - \frac{1}{\alpha S_R} \right) = w_U^\sigma n_U^{\sigma\phi+1} \left(\frac{1}{n_U} + \frac{1}{\xi} - \frac{1}{\alpha S_U} \right) . \tag{34}$$

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 (34) 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.

An implication of Proposition 1 is that all children in the same rural household receive the same education, i.e., $d_U = d_R$. This can be seen from the education first-order conditions (20) and (21); the Proposition implies that if one holds with equality, then the other must hold with equality given that children are indifferent between urban and rural locations. Therefore, for a given household, $h'_U = h'_R$. This does not imply that urban and rural households make the same choices; it merely implies that a rural parent allocates the same education to each of its children.

It is necessary to check that the premise that households are indifferent between locations holds in equilibrium. Urban-born households are not allowed to choose their location in order to pin down the joint distribution of population and human capital in equilibrium. If equilibrium features migration from rural to urban areas, then both types of households are indifferent between locations, and Proposition 1 holds. When households are indifferent between locations, labor demand determines the allocation of human capital across sectors. This is how dynasties with the same initial human capital may end up in different locations.

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

Proposition 2 implies that the urban/rural wage gap disappears in the limit. This is because the wage gap is a compensating differential for child mortality differences, which also disappear. This does not imply the urban and rural incomes are equalized in the long run; these wages are paid per unit of human capital, not per worker. Rather, if the urban sector has more human capital per worker, then urban incomes will be higher.

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

Proposition 3 implies that the urban and rural sectors produce substitutes (i.e., $\varepsilon > 1$) then the share of the population employed in the rural sector goes to zero in the long run. This is a standard result as in Ngai and Pissarides (2007) or Acemoglu and Guerrieri (2008). In the knife-edge case, if the final good was aggregated with a Cobb-Douglas production function ($\varepsilon = 1$), then both sectors could have non-zero shares in the long run.

Proposition 4: If $\lim_{t \rightarrow \infty} \bar{h} = \infty$, $\lim_{t \rightarrow \infty} n \geq 1$ and $\varepsilon > 1$, then the limit of both urban and rural wages is $\bar{w} \equiv \alpha \zeta^{\varepsilon(\varepsilon-1)}$.

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:

$$\tau_c + \frac{\bar{g}\bar{n}}{\xi} + \frac{\bar{n}}{\alpha\bar{s}} = 1, \tag{35}$$

$$(1 + \bar{g})^{1-\sigma} = \beta \left(\frac{\xi\tau_c}{\bar{n}} + 1 + \bar{g} \right). \tag{36}$$

Table 8. Calibrated parameters

| | 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) | ν | 0.35 | Technology effect on survival |

5. Quantitative analysis

Parameter values are chosen to match key features of the data, an initial condition is chosen to look like England in the 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 8, 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 , ζ , ε , α , ξ , ϕ , σ , β), I 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 initial ratio of urban to rural surviving children per adult $n_{U,0}/n_{R,0}$ is targeted to ≈ 0.77 , the ratio estimated by Clark (2009). This pins down the initial ratio,²⁶ while the levels of $n_{R,0}$ and $n_{U,0}$ are chosen to target an

²⁶The Clark (2009) estimates on urban and rural surviving children are useful for matching the relative value, but cannot be applied to target the levels of $n_{R,0}$ or $n_{U,0}$ directly, for they would imply an unrealistically high population growth rate. This is because the data is from wills and does not account

initial population growth rate of 8.5% per 25 years, which match the growth rate for England from 1400 to 1600 estimated by Broadberry *et al.* (2010).²⁷ 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-state 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 (34). Chosen empirical targets imply an initial premium of $(w_{U,0}/w_{R,0}) = 1.23$. The initial urban share, normalization of $h = 1$, and market time of $\tau_c = (1/(1 + \phi))$ 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:

$$S_j(\bar{h}) = \bar{S} - (\bar{S} - S_{j,0}) \frac{1 + v\bar{h}_0}{1 + v\bar{h}}. \tag{37}$$

This is a transformed logistic CDF, which is chosen for parsimony as it is governed by only one free parameter v , and also for having a positive limit as $\bar{h} \rightarrow 0$. 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 $\bar{h} \rightarrow \infty$, then $S_j(\bar{h}) \rightarrow \bar{S}$.

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

for people who choose not to have children, a choice which was prevalent even in high fertility preindustrial economies [e.g., Aaronson *et al.* (2014), Baudin *et al.* (2015), de La Croix *et al.* (2017)].

²⁷It is possible to match either the initial growth rate, or the initial fertility rate, but not both. In the model, the fertility rate is implied by the population growth rate and child mortality rate (which are matched to the data), but other factors delink these objects in the data: adult mortality, immigration/emigration, and statistical error. Accordingly, by choosing to match population growth, the model misses fertility behavior, with an initial fertility rate of 3.1 (twice the birth rate in Figure 7) considerably below empirical estimates of 5 or more for pre-industrial England [Wrigley and Schofield (1983)]. Moreover, model fertility declines monotonically while empirical fertility grew during the 18th century (Figure 1).

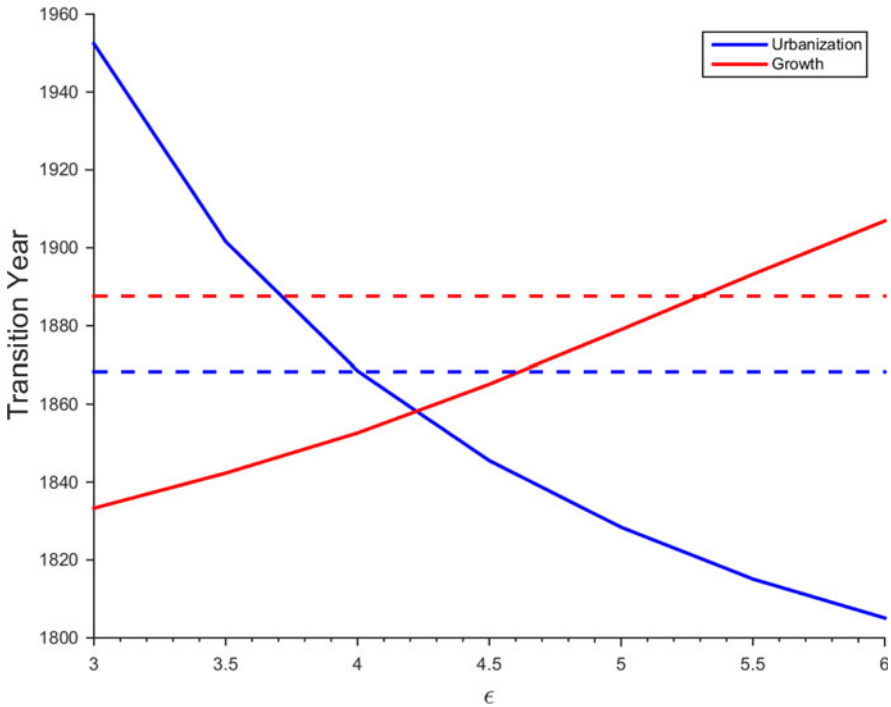


Figure 2. Elasticities of substitution and transition years.
 Notes: Urbanization transition is when urban share >50%. Growth transition is when annual income growth >1%.

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).²⁸ 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.

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 the 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

²⁸Appendix C details how to compute the equilibrium in practice.

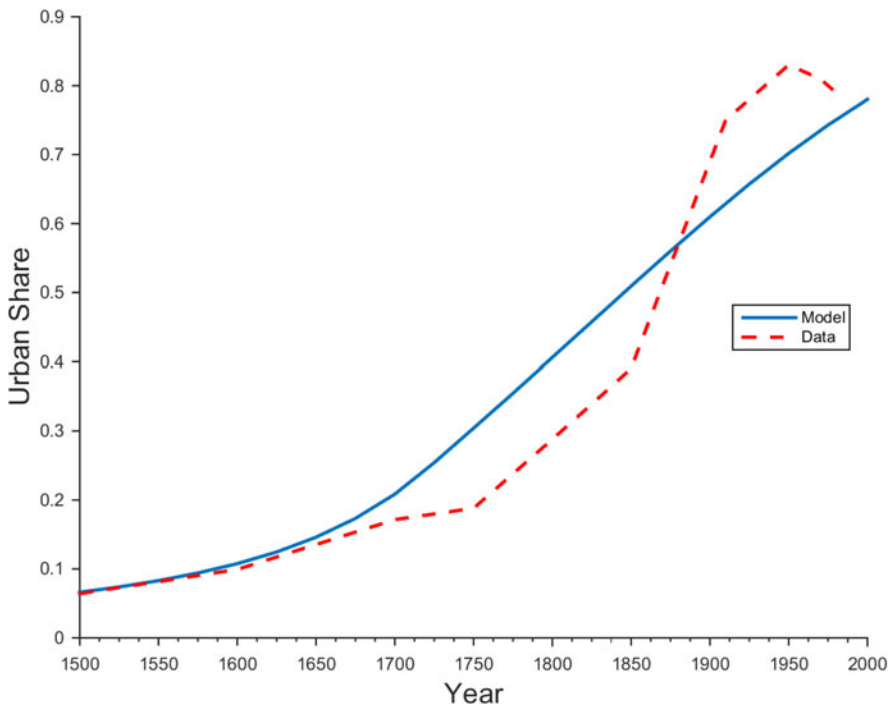


Figure 3. Simulation: urban share.
 Notes: Data from Bairoch (1991) and Bairoch et al. (1988)

substitute from quantity to quality. The 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 in (25) in a steady state:

$$(1 + g_{ss})^{1-\sigma} = \beta \frac{\xi}{n_{ss}} \left(\tau_c + n_{ss} \frac{1 + g_{ss}}{\xi} \right). \tag{38}$$

For the steady-state Euler equation, children choose the same location as their parent. $\tau_c + n_{ss}((1 + g_{ss})/\xi)$ is the return on human capital and ξ/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} decreases in n_{ss} for $g \in (0, \bar{g}]$.²⁹ 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: child quantity is a Giffen good.

²⁹The 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 < 1 . However, the calibration gives $\sigma > 0$, so n decreases in g for all $g < 14.5$, which is well above the long-run steady state.

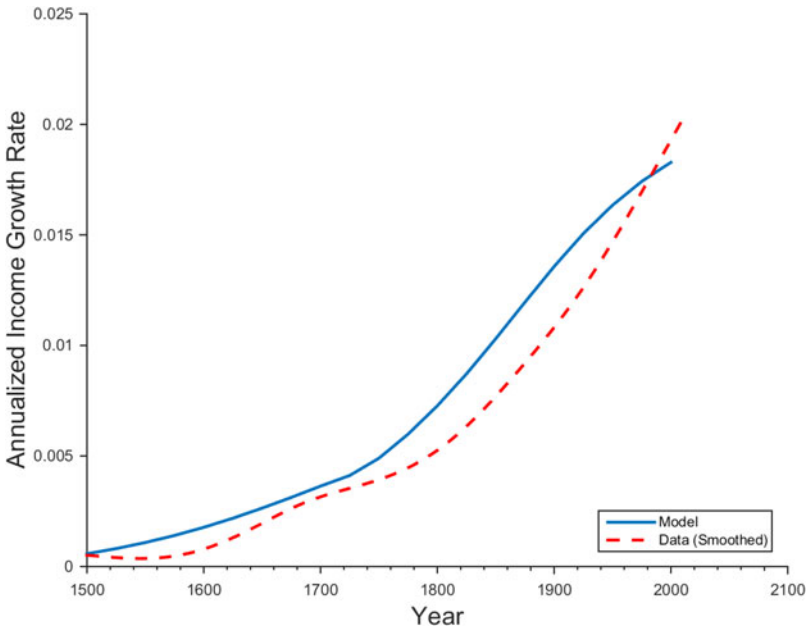


Figure 4. Simulation: income growth.
 Notes: Data from Broadberry *et al.* (2010) and The Maddison Project (2013), smoothed with an HP filter.

To understand this effect, Figure 5 also plots the normalized budget constraint, which divides the budget constraint (15) by total income:

$$\tau_c + \frac{gn}{\xi} + \frac{n}{\alpha S_j} = 1. \tag{39}$$

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. The Euler Equation represents the set of points for which a household’s indifference curve over child quantity and quality would be tangent to a budget constraint. So 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. In Figure 5, this corresponds to a movement from point B to point C.

The initial urban budget constraint does not intersect the Euler equation: urban households are constrained at $g=0$ (point A) so that the non-negativity constraint 17 is satisfied. As the urban survival rate improves, the urban budget constraint shifts towards the rural budget constraint (point B), and children increase. When the survival rate has improved sufficiently to unconstrain urban households, they follow the rural households and substitute from quantity to quality (point C). Urban households follow the quantity-quality substitution pattern of rural households with a lag because their budget sets for child quantity and quality are strictly smaller. A consequence of this delay is that urban households invest in less human capital than

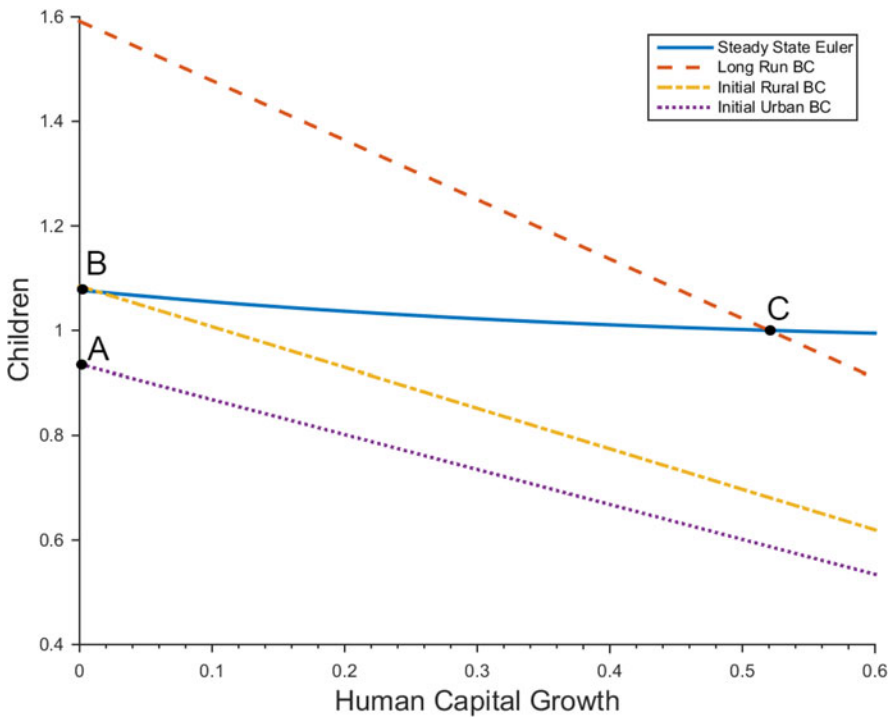


Figure 5. Quantity-quality substitution.

rural households. This pattern contrasts with empirical evidence that urban workers in preindustrial England tended to be more literate than rural workers.³⁰

Figure 6 plots the ratio of urban to rural values for three quantities: wages, children, and survival, exhibiting the predictions from section 2.1. 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 as 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.

³⁰For example, see Houston (1982) for evidence from the 17th and 18th centuries, and Mitch (2005) for evidence from the 19th century. Adding additional parameters to the model by allowing the urban sector to have higher productivity in human capital formation could be used to match this fact.

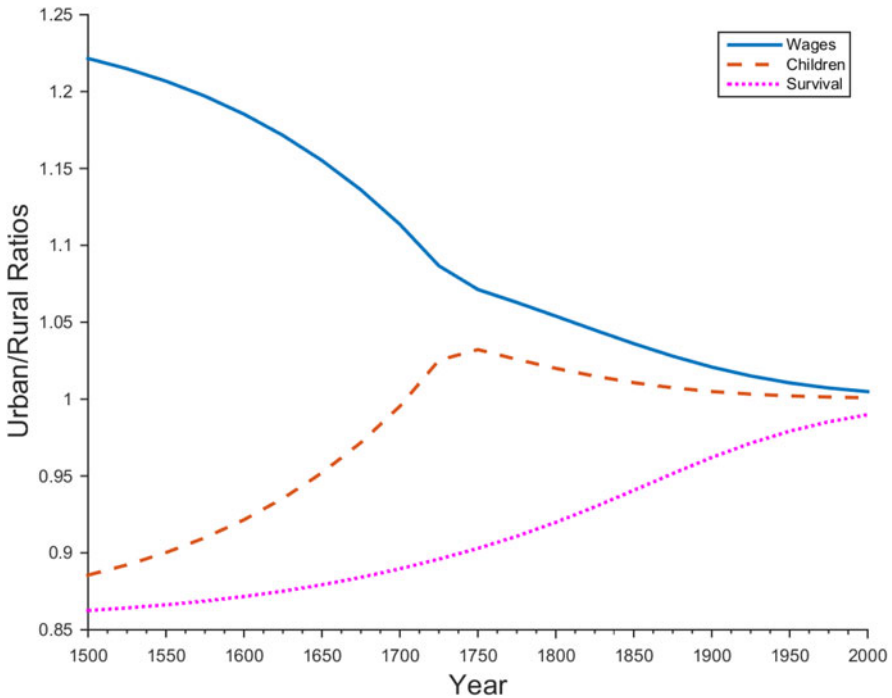


Figure 6. Simulation: urban/rural ratios.

While the urban-rural family size ratio increases from the initial period to the long run, fitting the empirical pattern in section 2.1, 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 cohort. Births are calculated before accounting for the fraction S_j that does 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.

The child mortality rate is plotted versus the smoothed mortality rate in Figure 8. Mortality falls, albeit not as abruptly as in the data. In contrast, the model's time path for fertility does not match the empirical path nearly as well. This is because, in

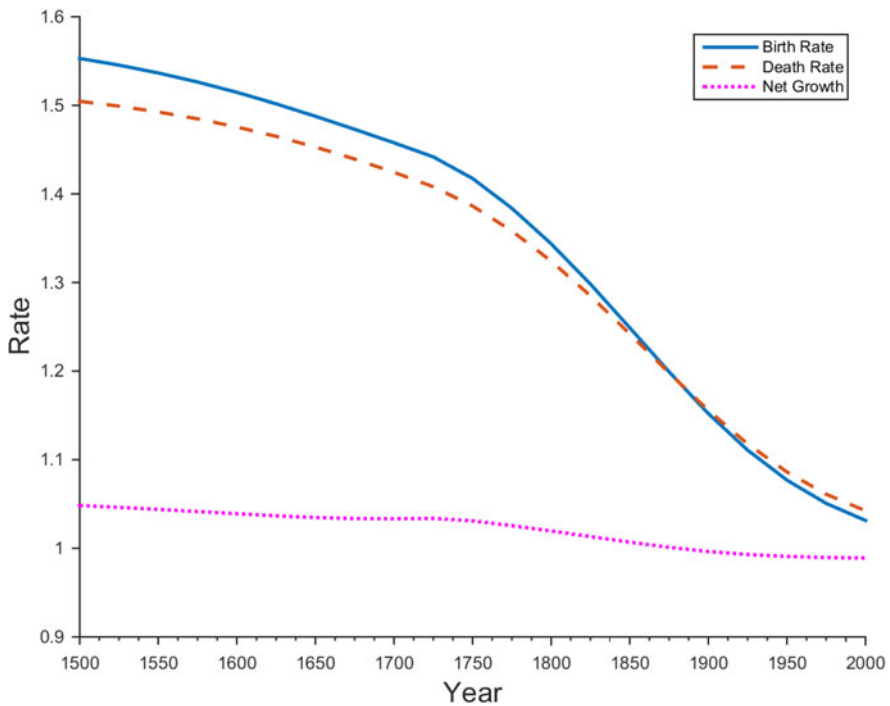


Figure 7. Simulation: demographic transition.

the model, fertility and child survival determine the population growth rate and the model is calibrated to match the population growth. But in reality, there are other factors (e.g., adult mortality, migration, gender balance) that cause the total fertility rate to move independently of child survival and population growth. As a result, empirical fertility is above 4 in 1500 CE (Figure 1) while initial fertility in the model is just above 3 (Figure 7; the model's total fertility rate is double the birth rate).

6. Model sensitivity

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 relationship between urbanization and income over the transition.

6.1 Sensitivity to initial characteristics

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

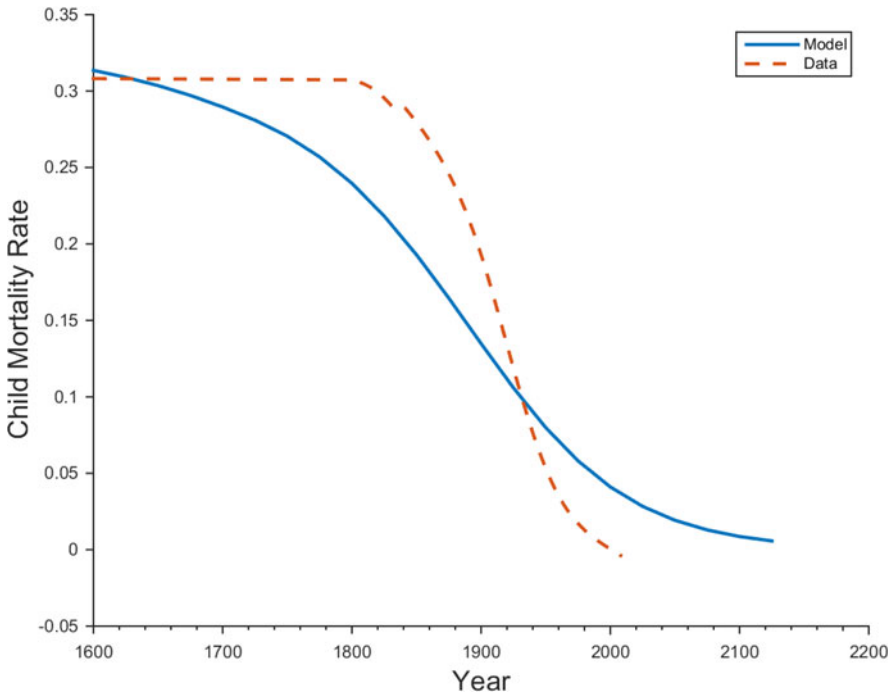


Figure 8. Simulation: mortality transition.

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 9 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 number of children, but a higher rate of human capital growth. Of course, other parameters must have small adjustments to maintain the long run calibration targets.

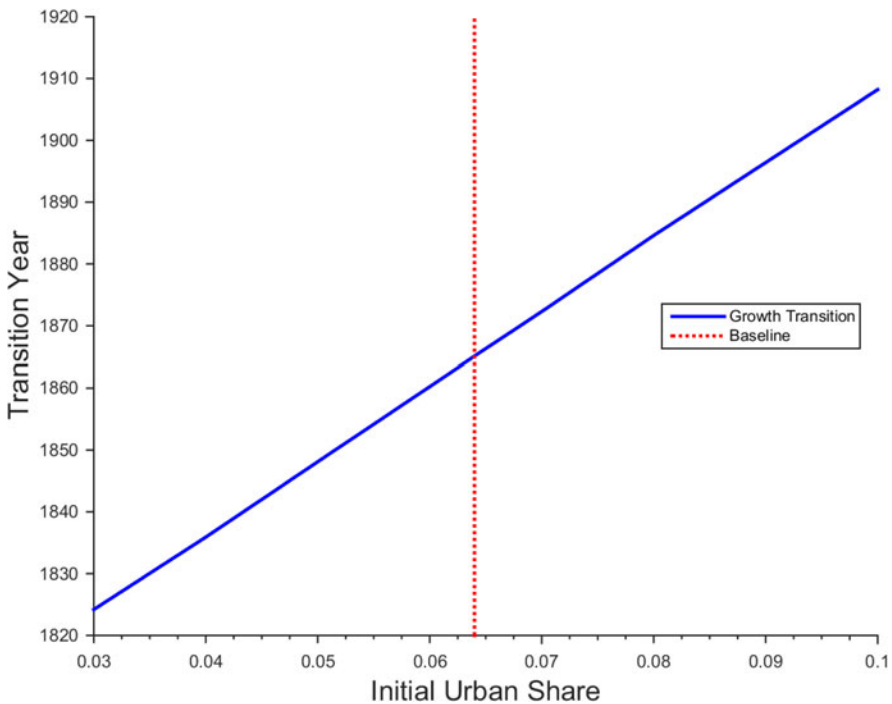


Figure 9. Transition years and initial urban share.

Figure 10 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 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 11. 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.

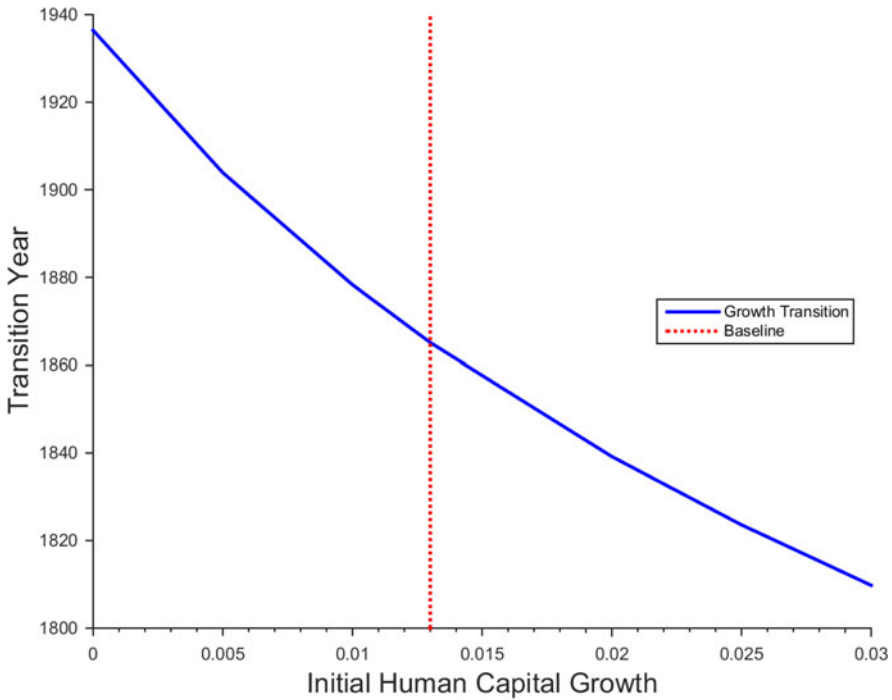


Figure 10. Transition years and initial human capital growth.

6.2 Urbanization and income levels

The analysis in section 6.1 suggests 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. The size of these effects is estimated in section 2.2. Testing initial conditions on transition timing support 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.³¹ 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 12 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 a 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.

³¹Ashraf and Galor (2011) estimate that countries in Galor and Moav (2004) Galor and Moav (2002) 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.

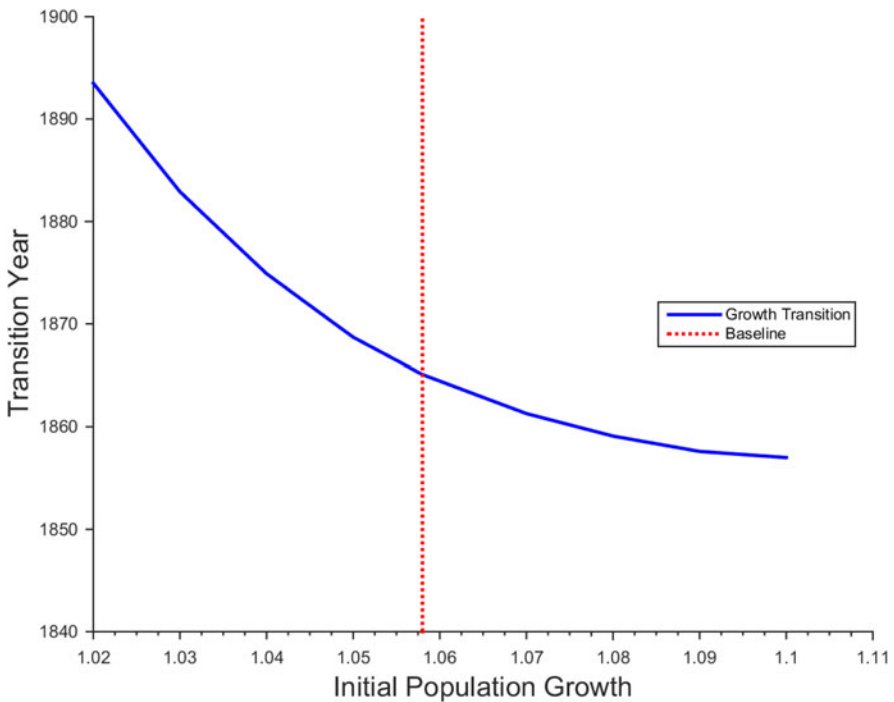


Figure 11. Transition years and initial population growth.

Because the urban sector is more productive relative to the rural sector in the alternative calibration, more households must be urban for a given wage premium.

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. This method uses two stages because it must first construct a measurement of the long-run relationship between a country’s income and urbanization, before testing the effect on transition timing. The method is not traditional two-stage least squares and does not construct an instrument.

First, I run the following panel regression, for country j in year t :

$$s_{U,t,j} = \gamma \log y_{j,t} + d_j + \kappa + \varepsilon_{j,t}. \tag{40}$$

This is a regression of urban share on log income with country fixed effects. Following the same approach as in section 2, income data are from The Maddison Project (2013) while urbanization rates are constructed from The Clio Infra Project (2016).

Next, I regress the transition year T_j on the estimated fixed effects:

$$T_j = \psi \widehat{d}_j + \kappa + \varphi_j. \tag{41}$$

Table 9 summarizes the 1st stage estimated country fixed effects. There are 76 countries³² with urbanization data before their income growth transition and 7,795

³²Listed in column 3 of Table 5.

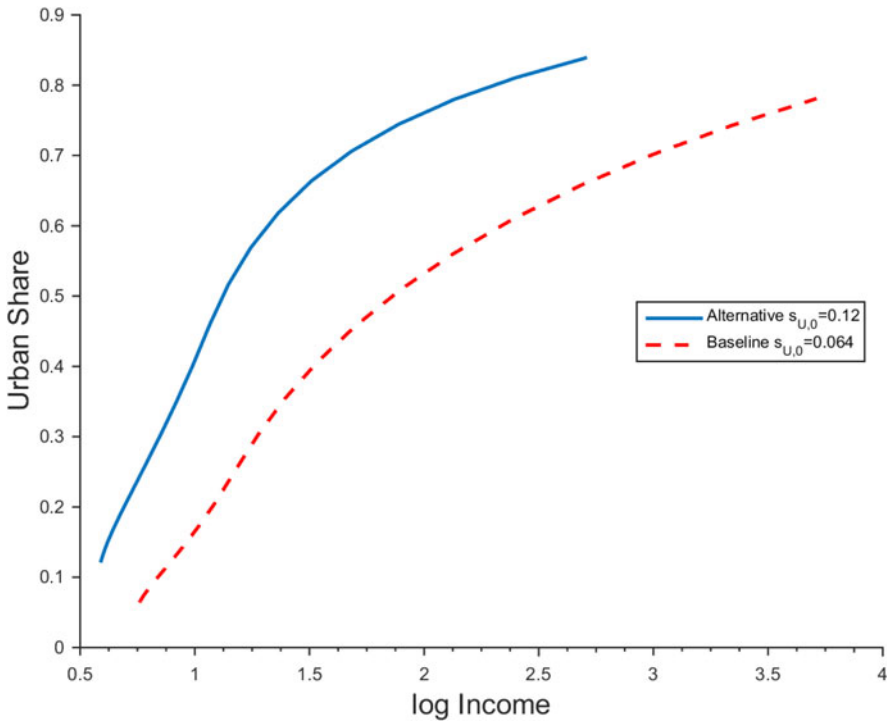


Figure 12. Urbanization and income levels: model.

Table 9. Summary of estimated urbanization fixed effects

| | Mean | sd | min | max |
|-----------------------|--------|-------|--------|-------|
| Country fixed effects | -0.009 | 0.059 | -0.108 | 0.169 |
| Observations | 7,795 | | | |

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 bootstrapping.

Table 10 reports the results of the second stage regression. Countries that have a higher urbanization level conditional on their income, transition later. I estimate $\hat{\psi} = 490.6$: if a country that is 10% more urban conditional on income, then they will transition almost 50 years later.

Figure 13 plots countries' first-stage estimated fixed effects versus their transition year and the second stage 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.

Table 10. Impact of estimated urbanization fixed effects on transition timing

| | |
|-----------------------|------------|
| Country fixed effects | 490.6*** |
| | (5.50) |
| Constant | 1,902.4*** |
| | (329.93) |
| Observations | 76 |

Notes: Standard Errors calculated by bootstrapping 500 times over 7,795 first stage observations. *t* statistics in parentheses.
 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

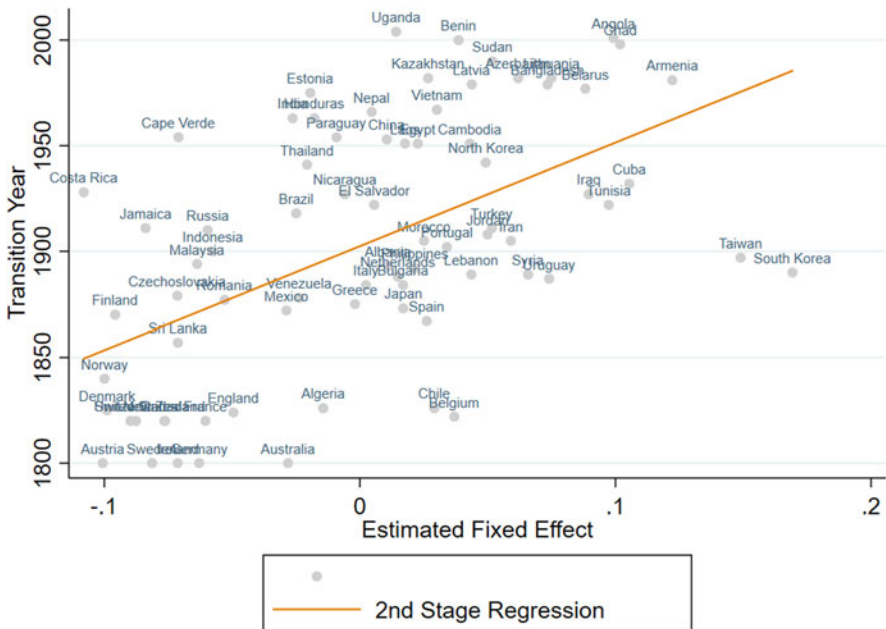


Figure 13. Estimated country effects and transition years.

This panel regression approach is consistent with the cross-country estimates from section 2.2: both 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. The key mechanism in the model is the effect of high urban child mortality on human capital accumulation, suggesting that when studying long-run growth, it is essential to consider the interaction between demographic incentives and structural transformation. This finding raises further research questions. Does this channel apply to current low-income countries? Does it reverse at some point as urbanization starts to incentivize human capital accumulation when cities specialize in services and serve as a locus for the transmission of ideas? Future work can address these questions by applying and expanding on the theory in this paper.

Acknowledgement. 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, and three anonymous referees.

References

- Aaronson, D., F. Lange and B. Mazumder (2014) Fertility transitions along the extensive and intensive margins. *American Economic Review* 104(11), 3701–3724.
- 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.
- Acemoglu, D., S. Johnson and J. Robinson (2005) The rise of Europe: Atlantic trade, institutional change, and economic growth. *The American Economic Review* 95(3), 546–579.
- 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* 101(5), 2003–2041.
- Bacci, M. L. (1977) *A History of Italian Fertility During the Last Two Centuries*. Princeton, NJ: Princeton University Press.
- Bairoch, P. (1991) *Cities and Economic Development: From the Dawn of History to the Present*. Chicago, IL: University of Chicago Press.
- Bairoch, P., J. Batou, and P. Chevre (1988) *La population des villes europeennes de 800 a 1850: Banque de donnees et analyse sommaire des resultats*. Geneva, Switzerland: Librairie Droz.
- Barro, R. J. and X. Sala-i Martin (2004) *Economic Growth*. Cambridge, MA: MIT Press.
- Baudin, T., D. De la Croix and P. E. Gobbi (2015) Fertility and childlessness in the United States. *American Economic Review* 105(6), 1852–1882.
- Baudin, T. and R. Stetler (2018) Rural exodus and fertility at the time of industrialization. *Working Paper*.
- Baumol, W. J., S. A. B. Blackman, and E. N. Wolff (1985) Unbalanced growth revisited: asymptotic stagnancy and new evidence. *The American Economic Review* 75(4), 806–817.
- Becker, G. (1960) An economic analysis of fertility. in Dwyer, C.J. (ed.), *Demographic and Economic Change in Developed Countries*, pp. 209–240 Columbia, NY: Columbia University Press.
- Becker, G. (1981) *A Treatise on the Family*. Cambridge, MA: 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, 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.

- Becker, G. S. and H. G. Lewis (1973) On the interaction between the quantity and quality of children. *Journal of Political Economy* 81(2), S279–S288.
- Becker, G., K. Murphy and R. Tamura (1990) Economic growth, human capital and population growth. *Journal of Political Economy* 98(5), S12–S137.
- Bhattacharya, J. and S. Chakraborty (2012) Fertility choice under child mortality and social norms. *Economics Letters* 115(3), 338–341.
- Bhattacharya, J. and S. Chakraborty (2017) Contraception and the demographic transition. *The Economic Journal* 127(606), 2263–2301.
- 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.
- Brezis, E. S. and P. R. Krugman (1997) Technology and the life cycle of cities. *Journal of Economic Growth* 2(4), 369–383.
- Brezis, E. S., P. R. Krugman, and D. Tsiddon (1993) Leapfrogging in international competition: a theory of cycles in national technological leadership. *The American Economic Review* 83(5), 1211–1219.
- 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.
- Chatterjee, S. and T. Vogl (2018) Escaping Malthus: economic growth and fertility change in the developing world. *American Economic Review* 108(6), 1440–1467.
- 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.
- The Clio Infra Project (2016) The clio-infra database on urban settlement sizes: 1500–2000. <http://www.cegh.nl/urbanisation-hub-clio-infra-database-urban-settlement-sizes-1500-2000>.
- 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.
- D’Costa, 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.
- de La Croix, D., E. Schneider, and J. L. Weisdorf (2017) Childlessness, Celibacy and Net Fertility in Pre-Industrial England: The Middle-class Evolutionary Advantage. *Centre for Economic Policy Research*.
- Diamond, J. M. (1998) *Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years*. New York, NY: Random House.
- 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), 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.
- Ehrlich, I. and F. T. Lui (1991) Intergenerational trade, longevity, and economic growth. *Journal of Political Economy* 99(5), 1029–1059.
- Fernandez-Villaverde, J. (2001). Was Malthus Right? Economic Growth and Population Dynamics. Available at SSRN: <https://ssrn.com/abstract=293800> or <http://dx.doi.org/10.2139/ssrn.293800>
- Fogel, R. W. (2004) *The escape From Hunger and Premature Death, 1700–2100: Europe, America, and the Third World*, Volume 38. Cambridge, UK: Cambridge University Press.
- Galor, O. (2011) *Unified Growth Theory*. Princeton, NJ: Princeton University Press.
- Galor, O. and O. Moav (2002) Natural selection and the origin of economic growth. *The Quarterly Journal of Economics* 117(4), 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 (2014) The agricultural productivity gap. *Quarterly Journal of Economics* 129(2), 939–993.
- 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), 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.
- Houston, R. A. (1982) The development of literacy: Northern England, 1640–1750. *Economic History Review* 52(2), 199–216.
- 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), 411–439.
- Kalemli-Ozcan, S. (2008) The uncertain lifetime and the timing of human capital investment. *Journal of Population Economics* 21(3), 557–572.
- 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.
- Keszenbaum, 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 Roberts, G. B. (ed.), *Demographic and Economic Change in Developed Countries*, pp. 77–116 Columbia, NY: Columbia University Press.
- Knodel, J. E. et al. (1974) *The Decline of Fertility in Germany, 1871–1939*, Volume 2. Princeton, NJ: Princeton University Press.
- Kuznets, S. (1966) *Modern Economic Growth: Rate, Structure, and Spread*. New Haven: Yale University Press.
- Lagakos, D. and M. E. Waugh (2013) Selection, agriculture, and cross-country productivity differences. *American Economic Review* 103(2), 948–980.
- Lagerlof, N.-P. (2003) From Malthus to modern growth: can epidemics explain the three regimes? *International Economic Review* 44(2), 755–777.
- Laitner, J. (2000) Structural change and economic growth. *The Review of Economic Studies* 67(3), 545–561.
- Lorentzen, P., J. McMillan and R. Wacziarg (2008) Death and development. *Journal of Economic Growth* 13(2), 81–124.
- Lucas, R. E. (2002) *Lectures on Economic Growth*. Cambridge, MA: 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. In I. Leveson and W. Wheeler (eds.), *Western Economies in Transition*. London: Croom Helm, pp. 41–60.
- The Maddison Project (2013) The maddison project. <http://www.ggdc.net/maddison/maddison-project/home.htm>.
- McKeown, T. (1976) *The Modern Rise of Population*. London: Edward Arnold.
- 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.
- Mitch, D. (2005) Literacy and occupational mobility in rural versus urban Victorian England: evidence from the linked marriage register and census records for Birmingham and Norfolk, 1851 and 1881. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 38(1), 26–38.
- 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.
- O'Rourke, 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.
- Razin, A. and U. Ben-Zion (1975) An intergenerational model of population growth. *The American Economic Review* 65(5), 923–933.
- Soares, R. R. (2005) Mortality reductions, educational attainment, and fertility choice. *The American Economic Review* 95(3), 580–601.
- Stokey, N. L. (1996) Free trade, factor returns, and factor accumulation. *Journal of Economic Growth* 1(4), 421–447.
- Stokey, N. L. (2001) A quantitative model of the British industrial revolution, 1780–1850. *Carnegie-Rochester Conference Series on Public Policy* 55, 55–109.
- Strulik, H. (2017) Contraception and development: a unified growth theory. *International Economic Review* 58(2), 561–584.
- 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 M. Daunton (ed.), *The Cambridge Urban History of Britain*. Cambridge, UK: Cambridge University Press, pp. 629–672.
- Tamura, R. (2002) Human capital and the switch from agriculture to industry. *Journal of Economic Dynamics and Control* 27(2), 207–242.
- Tamura, R. (2006) Human capital and economic development. *Journal of Development Economics* 79(1), 26–72.
- United Nations Statistics Division (2012) *Demographic Yearbook 2013*. New York, NY: United Nations.
- Williamson, J. G. (1987) Did English factor markets fail during the industrial revolution? *Oxford Economic Papers* 39(4), 641–678.
- Williamson, J. G. (2002) *Coping with City Growth During the British Industrial Revolution*. Cambridge, UK: Cambridge University Press.
- Willis, R. J. (1973) A new approach to the economic theory of fertility behavior. *Journal of Political Economy* 81(2), 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* 128(4), 25.

Appendix 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. Consider a household where all children in each future generation make the same location choices. This may be true for an individual dynasty, which is atomistic. Then the household dynastic utility (8) can be expanded into the discounted sum:

$$V_t = \sum_{k=0}^{\infty} \beta^k \frac{(c_{t+k} n_{t+k}^{\phi})^{\sigma}}{\sigma}. \quad (\text{A.1})$$

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:

$$\begin{aligned}
 V_t &= \sum_{k=0}^{\infty} \beta^k \frac{(\tau_c w_{t+k, \mathcal{J}(t+k)} h_{t+k} n_{t+k}^\phi)^\sigma}{\sigma} \\
 &= h_t^\sigma \sum_{k=0}^{\infty} \beta^k \frac{(\tau_c w_{t+k, \mathcal{J}(t+k)} (h_{t+k}/h_t) n_{t+k}^\phi)^\sigma}{\sigma}
 \end{aligned}
 \tag{A.2}$$

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

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

substituting this expression into (A.2) 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:

$$V_{\mathcal{J}}(h) \propto h^\sigma.
 \tag{A.3}$$

Now consider two different location sequences \mathcal{J} and \mathcal{J}' . Because of the proportionality in (A.3), it is true that:

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

$$V_{\mathcal{J}}(h) = V_{\mathcal{J}'}(h) \quad \forall h > 0.
 \tag{A.4}$$

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

$$V_{\mathcal{J}}(h) > V_{\mathcal{J}'}(h) \quad \forall h > 0.
 \tag{A.5}$$

Suppose households are indifferent between urban and rural locations for some \hat{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 (A.4) and (A.5) 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 $\mathcal{J} \in \{\mathcal{J}_U, \mathcal{J}_R\}$:

$$V_{\mathcal{J}}(h_t) = h_t^\sigma \sum_{k=0}^{\infty} \beta^k \frac{\left(\tau_c w_{t+k, \mathcal{J}(t+k)} \prod_{s=t}^{t+k-1} (1 + g_s) n_{t+k}^\phi \right)^\sigma}{\sigma},
 \tag{A.6}$$

$$\equiv h_t^\sigma \mathcal{V}.
 \tag{A.7}$$

Thus, if households are indifferent for some \hat{h} , the marginal value of human capital is equalized in both locations:

$$V'_{\mathcal{J}}(h_t) = \sigma h_t^{\sigma-1} \mathcal{V} \quad \forall \mathcal{J} \in \{\mathcal{J}_U, \mathcal{J}_R\}.
 \tag{A.8}$$

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 (15) 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 $\tau_c = (1/(1 + \phi))$ is constant. This normalized budget constraint and the Euler equation (25) 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 $\lim_{t \rightarrow \infty} \bar{h} = \infty$, then the limiting urban-rural wage premium is $(w_U/w_R) \rightarrow 1$.

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, as per the combined budget constraint (15), 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 $\lim_{\bar{h} \rightarrow \infty} S_j(\bar{h}) = \bar{S}$ for all j . So, in the limit, it must be that $(w_U/w_R) \rightarrow 1$. ■

A.3. Proof of proposition 3

In this section I prove that if $\lim_{t \rightarrow \infty} \bar{h} = \infty$, $\lim_{t \rightarrow \infty} n \geq 1$ 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 $H = N\bar{h}$ grows in the long run: $\lim_{t \rightarrow \infty} H = \infty$. Use the equilibrium prices in equations (32) and (33) to express the wage premium as:

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

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

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

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

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

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

decreases in the urban share for $s_U \in (0, 1)$, and

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

So, it must be that $s_U \rightarrow 1$. ■

A.4. Proof of proposition 4

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

Proof. Use the final good production function (4) and equilibrium prices in equations (32) and (33) to express the equilibrium urban wage as:

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

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

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

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

$$\begin{aligned} \lim_{t \rightarrow \infty} w_U &= \lim_{t \rightarrow \infty} A^{(\epsilon-1)/\epsilon} \zeta \left(\frac{A(\zeta(\tau_c s_U H)^{(\epsilon-1)/\epsilon} + (1 - \zeta)((\tau_c(1 - s_U)H)^\theta L^{1-\theta})^{(\epsilon-1)/\epsilon})^{\epsilon/(\epsilon-1)}}{\tau_c s_U H} \right)^{1/\epsilon} \\ &= \frac{\epsilon}{A\zeta\epsilon - 1} \equiv \bar{w} \end{aligned}$$

Appendix B: 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:

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

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 strictly increases in \bar{h} , bounded by $[0, \bar{S}]$, and has finite limits as $\bar{h} \rightarrow 0$, and $\bar{h} \rightarrow \infty$.

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 the average income is used to approximate average human capital. Non-linear least squares give $v=0.35$ when \bar{h}_0 is normalized to one. Figure B.1 plots England’s mortality data, income, and the fitted survival function given the year’s income level.

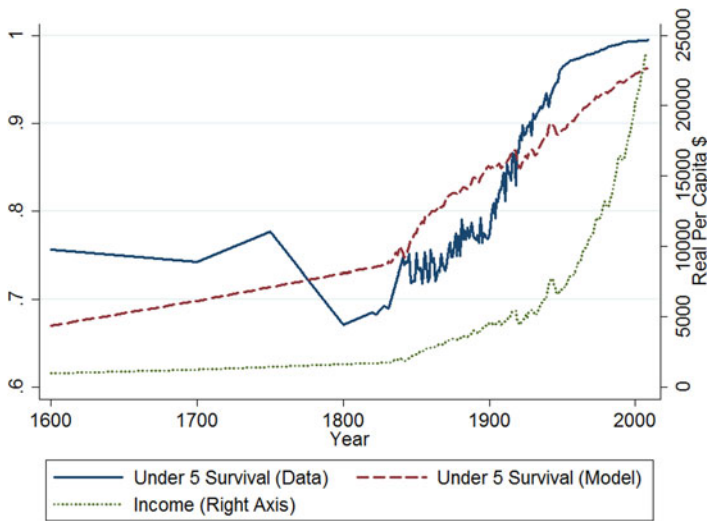


Figure B.1. Empirical and estimated survival rates.

Appendix C: 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 (34) 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 if the household is unconstrained, then the left-hand side is also equal, so substituting with equation (25) implies:

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

Then dividing equation (34) by equation (C.1) yields:

$$\frac{(1/n_R) + (1/\xi) - (1/\alpha S_R)}{(1 + g_R)^{1-\sigma}} = \frac{(1/n_U) + (1/\xi) - (1/\alpha S_U)}{(1 + g_U)^{1-\sigma}}. \tag{C.2}$$

Next, combine equation (C.2) with the normalized budget constraint (39) to yield an equation relating n_U, n_R, S_U, S_R , and parameters:

$$\frac{(1/n_R) + (1/\xi) - (1/\alpha S_R)}{(1 + \xi((1 - \tau_c)/n_R) - (\xi/\alpha S_R))^{1-\sigma}} = \frac{(1/n_U) + (1/\xi) - (1/\alpha S_U)}{(1 + \xi((1 - \tau_c)/n_U) - (\xi/\alpha S_U))^{1-\sigma}}. \tag{C.3}$$

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 (C.3) for $n_{U,t}$. If the implied value of n_U is infeasible, the urban households must be constrained and their Euler equation does not hold, so set $n_U = (1 - \tau_c)\alpha S_{U,t}$.
- (2) Analytically solve the normalized budget constraints (39) for $g_{R,t}$ and $g_{U,t}$.
- (3) Calculate the wage premium $w_{U,t}/w_{R,t}$ from the indifference condition (34).
- (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 (32) and (33).
- (6) Calculate next period's distribution of human capital Λ_{t+1} from the law of motion (29).
- (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 (37).
- (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 (25)
 - (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 (32) and (33).
 - (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 $n_{U,t+1} = (1 - \tau_c)\alpha S_{U,t+1}$.
- (1) Return to step 1. for the 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.