# The Human Side of Structural Transformation

# Online Appendix

Tommaso Porzio

Columbia University, CEPR, and NBER; email: tommaso.porzio@columbia.edu

Federico Rossi

University of Warwick; email: federico.rossi@warwick.ac.uk

Gabriella Santangelo

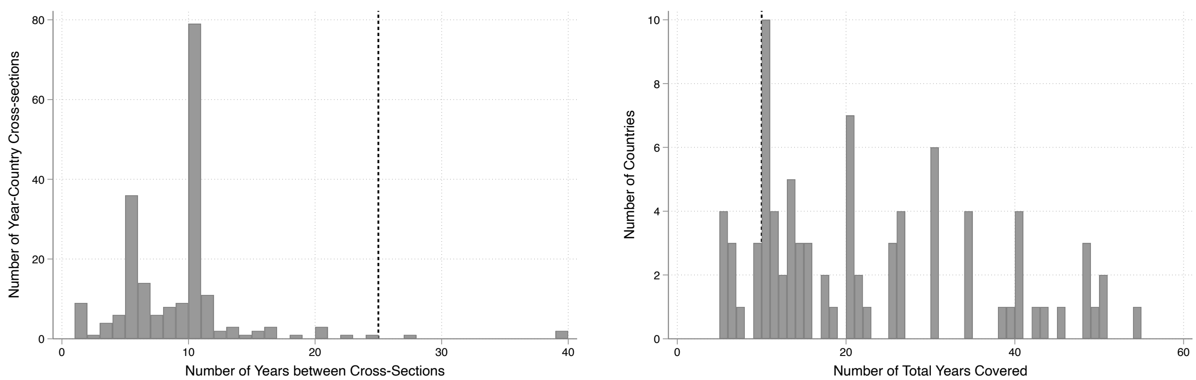
University of Cambridge; email: gabriella.santangelo@econ.cam.ac.uk

# A Data Appendix

## A IPUMS-I and IPUMS-GH Data

Figure A.1: Frequency and Coverage of Cross-sectional Data

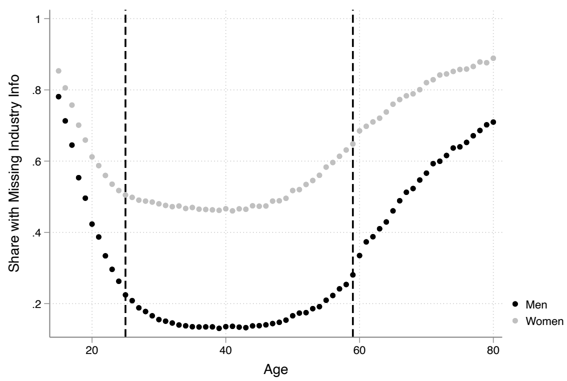
(a) Years Gap between Cross-Sections (b) Total Number of Years Covered



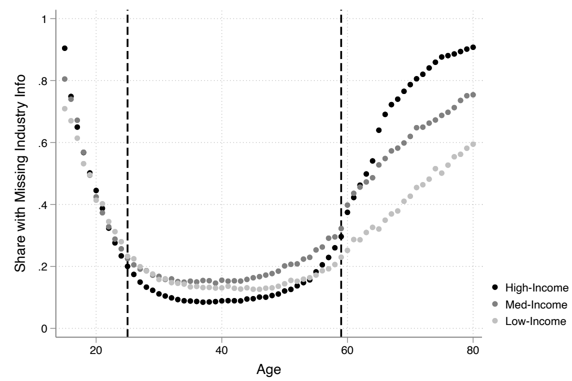

Notes: the left figure shows the histogram of the number of years between observed cross-sections in our data. We drop from the analysis countries/cross-sections to the right of the black dotted line. The right figure shows the histogram of the total numbers of years covered by each country in our data. For our benchmark sample, we exclude countries to the left of the black dotted line. All figures use IPUMS-I data.

Figure A.2: Missing Industry Information and Labor Force Participation

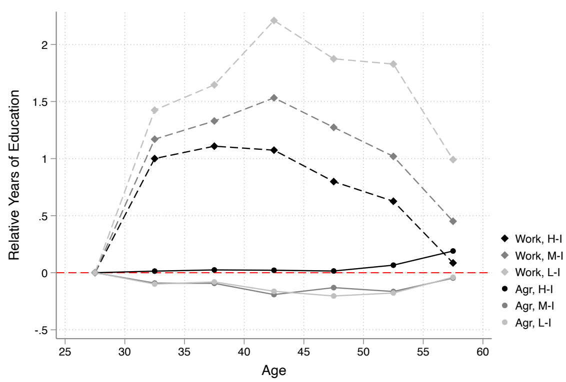
(a) Average Life-Cycle, by Gender



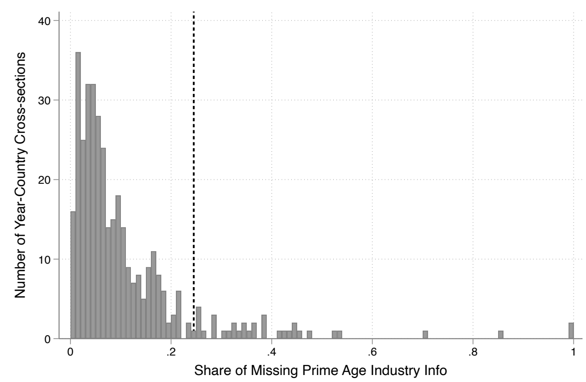

(b) Average Life-Cycle, by Income Group




(c) Selection into Labor Force



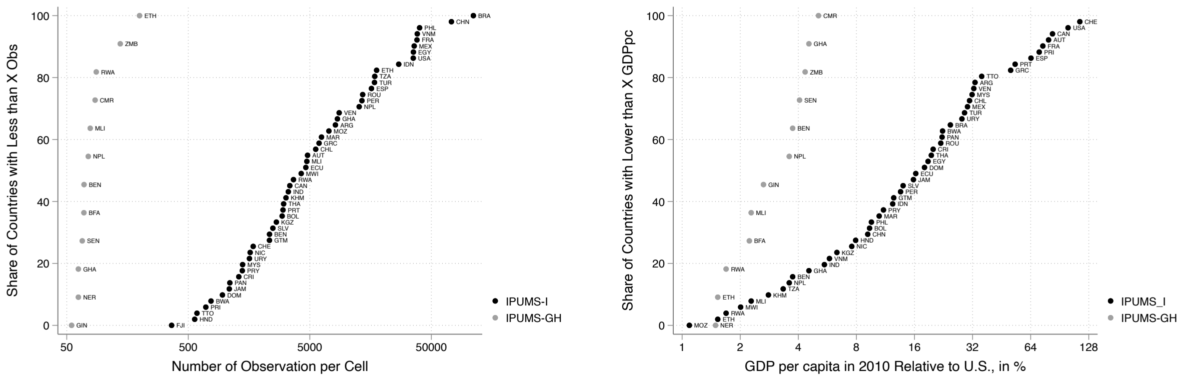

(d) Prime Age Males, across countries




Notes: the two top figures show the share of the population by age reporting non-missing industry information. Information on industry is missing when an individual is not in the labor force. The left figure stratifies by gender and averages across all countries. The right figure keeps only males and stratifies by income group. The bottom left figure shows how the relative average years of schooling vary over the life-cycle for different groups of the population. Specifically, for each cohort-country-year, we calculate the difference between the average education of individuals in and out of labor force (Work) and in and out of agriculture (Agr). We then regress them (separately by income group) on country $\times$ cohort fixed effects and five-year age dummies; the figure reports the point estimates for the latter. We learn that as a cohort ages the average education of the individuals in the labor force initially increases steeply. This is especially true in low income countries and is driven by selection: more educated individuals are relatively more likely to be non-employed when young. We can see that when comparing individuals working in agriculture and the rest of the population there is virtually no selection over the life-cycle. This motivates us to compute agricultural employment as the share of the cohort population employed in agriculture, rather than the share of the population in the labor force employed in agriculture. In this way, we are less concerned that changes in the denominator drive our measured agricultural reallocation. The bottom right figure shows the histogram, across all our cross-sections, of the share of prime-age men that report missing industry information – i.e. that are out of the labor force. We exclude country-years that are to the right of the black dotted line. All the figures use IPUMS-I data.

Figure A.3: Sample Coverages: IPUMS-GH and IPUMS-I

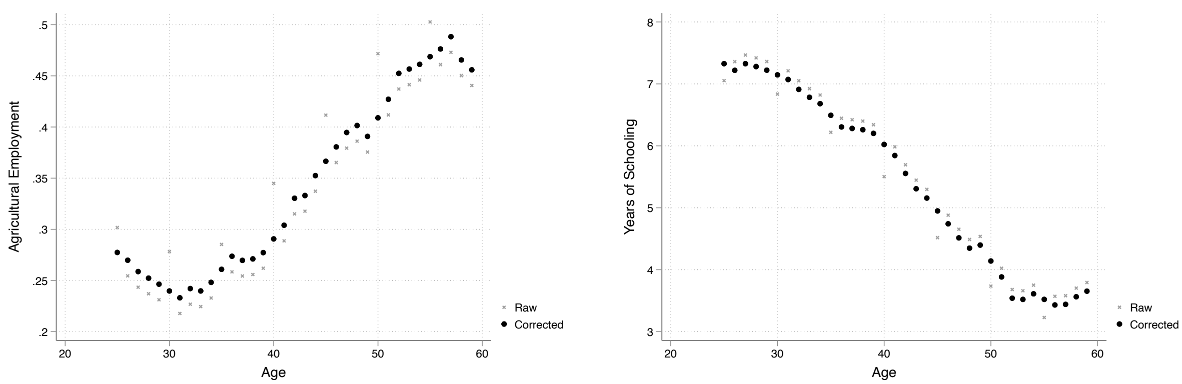
(a) Number of Observations (b) GDP per capita in 2010




Notes: the two figures compare the coverage of the IPUMS-GH and IPUMS-I data in terms of sample sizes (left figure) and GDP per capita of covered countries (right figure). The left figure shows the cumulative density function of countries by the average number of observation for each cohort-year-country cell: IPUMS-GH data have much smaller samples. The right figure shows the cumulative density function of countries' GDP per capita, relative to the one of the United States in 2010: IPUMS-GH countries have much lower income. Some countries are in both datasets.

Figure A.4: Correction for Age Heaping: a Stark Example from Turkey 1985

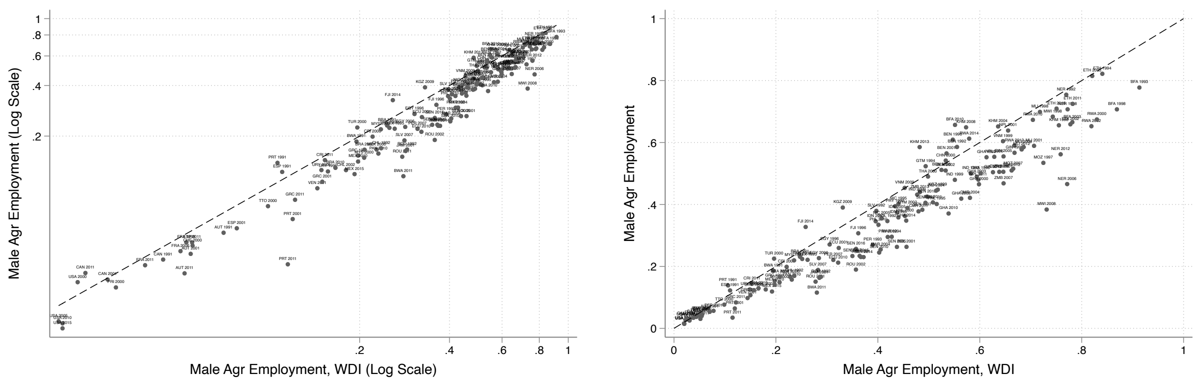
(a) Agricultural Employment (b) Years of Schooling




Notes: the two figures illustrate the effect of correcting our data for age heaping for one country affected by this issue. We implement the correction in three steps: i) we compute agricultural employment (or schooling) for each cohort of age in $\{30, 35, 40, 45, 50, 55\}$ as the average between the corresponding variable for one year younger and one year older cohorts; ii) we correct agricultural employment (or schooling) for the 25-year-old cohort using the gap between the corrected and uncorrected measures for the 35-year-old cohort; iii) we renormalize so that the average agricultural employment (or schooling) is the same as for the original variable. The figures show that this procedure recovers a relatively smooth distribution over age.

Figure A.5: Comparison with World Development Indicators

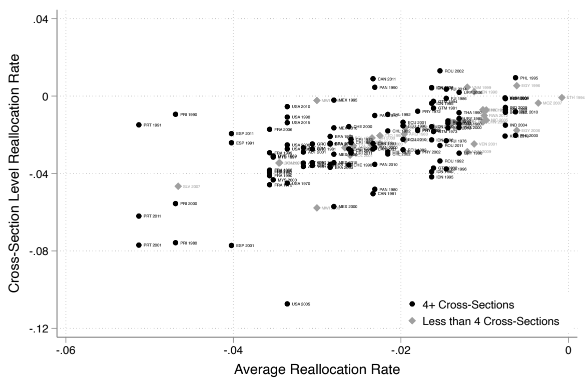
(a) Log Scale (b) Linear Scale



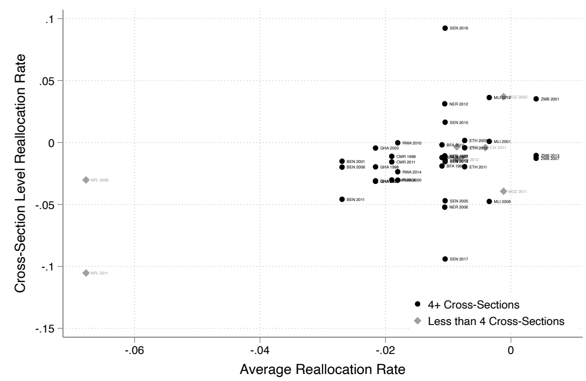

Notes: the figures compare the share of agricultural employment measured from our dataset with aggregate statistics from the World Development Indicators (World Bank, 2017), based on ILO data. For the WDI, we use the variable: "Employment in agriculture, male (% of male employment)". We keep all country/year pairs for which both IPUMS-I and WDI data are available. The left figure uses a log scale (to illustrate more clearly countries with high and low agricultural employment), while the right one uses a linear scale. All countries are included in both figures. The WDI data are on average larger, since they refer to the share of the employed population in agriculture, while we consider the share of the working age population in agriculture. While, the cross-country patterns are very similar, there are a few exceptions, linked to periods of large economic crisis and declines in employment. Two clear examples are the economic downturn that hit Portugal in 2011 and the food crisis in Niger in 2006. An increase in unemployment increases the discrepancy between our measure and the WDI one since, as well known (Storesletten et al. (2019)), agricultural employment is less-cyclical.

Figure A.6: Data Anomalies Checks

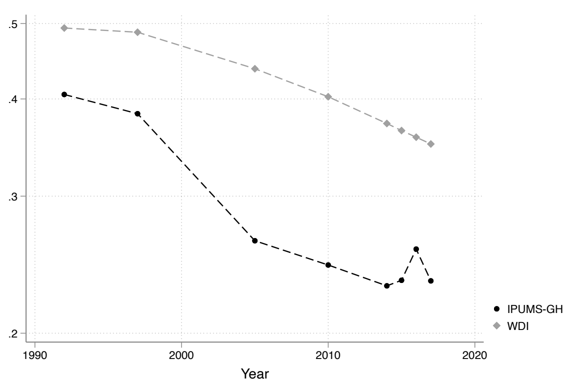
(a) IPUMS-I



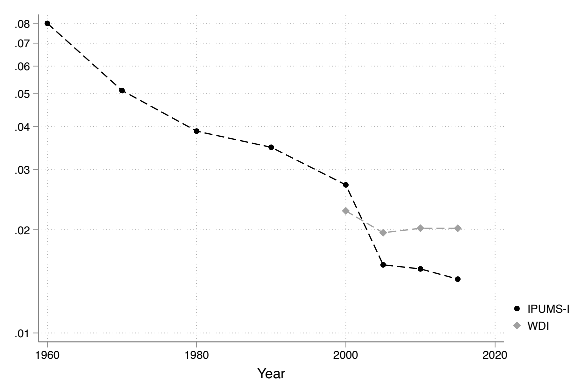

(b) IPUMS-GH




(c) Senegal

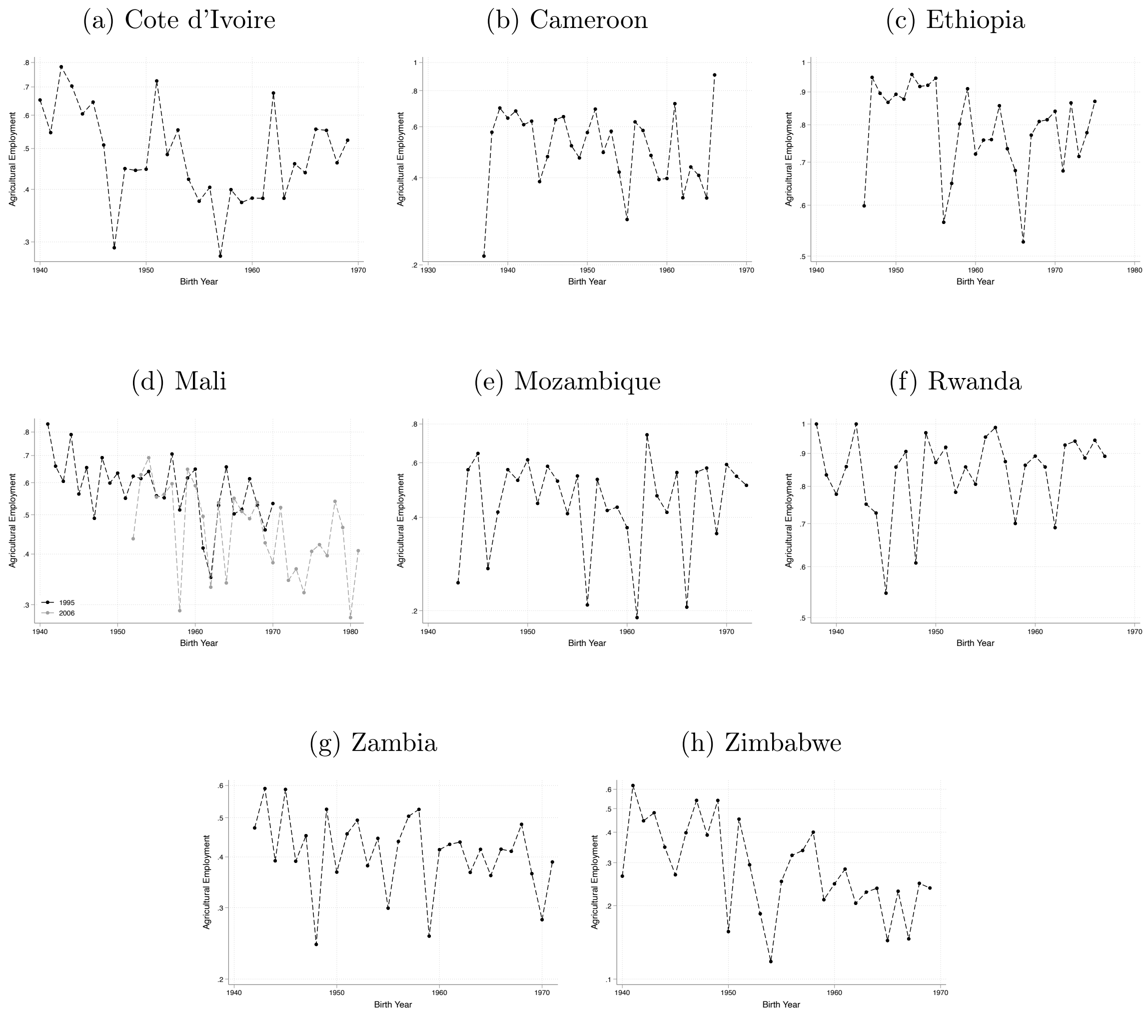



(d) United States




Notes: the top two figures plot, for IPUMS-I and IPUMS-GH countries, the cross-section-specific reallocation rate as a function of the average reallocation rate across all cross-sections. The farther the points are from the 45 degree lines, the more there is within country heterogeneity in reallocation rates over time. We further distinguish between countries for which we have less or more than four cross-sections. For most countries, the within country heterogeneity in reallocation rates is limited. Three observations are clear outliers: USA 2005, and SEN 2017 and 2016. The bottom two figures show the time-series of agricultural employment for these two countries and compare them with data from the WDI. The figures show that there are anomalies in those years, which we thus exclude from the dataset. In the United States, the steep decrease in agricultural employment from 2000 to 2005 corresponded to a change in the underlying sectoral classification. We have not been able to document the reason behind the jump in agricultural employment in Senegal 2016.

Figure A.7: Excluded IPUMS-GH Cross-Sections due to Data Anomalies




Notes: the figures plot average agricultural by birth cohort for all the cross-sections that we have excluded from our analysis. The figures show that the excluded cross-sections are very noisy, making it difficult to interpret the cohort-specific reallocation rate.

Table A.1: IPUMS-I Countries - Part I

| | Country | (1) GDP pc | (2) Year Range | (3) Min. Agr. Emp. | (4) Max Agr. Emp. | (5) N. Surveys | (6) N. Obs. |
|---|---|---|---|---|---|---|---|
| (1) | Argentina | 33% | 1970-1980 | 13% | 16% | 2 | 8120 |
| (2) | Austria | 79% | 1971-2011 | 3% | 11% | 5 | 4789 |
| (3) | Benin | 4% | 1979-2013 | 43% | 69% | 4 | 2330 |
| (4) | Bolivia | 9% | 1976-2001 | 26% | 49% | 3 | 2959 |
| (5) | Botswana | 22% | 1991-2011 | 12% | 19% | 2 | 769 |
| (6) | Brazil | 25% | 1960-2010 | 13% | 53% | 6 | 110788 |
| (7) | Cambodia | 3% | 1998-2013 | 59% | 66% | 4 | 3183 |
| (8) | Canada | 83% | 1971-2011 | 3% | 8% | 5 | 3422 |
| (9) | Chile | 31% | 1960-2002 | 13% | 31% | 5 | 5589 |
| (10) | China | 9% | 1982-2000 | 54% | 66% | 3 | 73119 |
| (11) | Costa Rica | 20% | 1963-2011 | 14% | 51% | 5 | 1302 |
| (12) | Dominican Republic | 18% | 1960-1970 | 44% | 62% | 2 | 955 |
| (13) | Ecuador | 16% | 1962-2010 | 21% | 59% | 5 | 4634 |
| (14) | Egypt | 19% | 1986-2006 | 23% | 27% | 3 | 35579 |
| (15) | El Salvador | 14% | 1992-2007 | 19% | 38% | 2 | 2483 |
| (16) | Ethiopia | 2% | 1984-1994 | 82% | 83% | 2 | 17721 |
| (17) | Fiji | - | 1966-2014 | 31% | 57% | 5 | 365 |
| (18) | France | 74% | 1962-2011 | 3% | 19% | 8 | 38064 |
| (19) | Ghana | 5% | 1984-2010 | 37% | 60% | 3 | 8434 |
| (20) | Greece | 50% | 1971-2011 | 8% | 28% | 5 | 5957 |
| (21) | Guatemala | 12% | 1964-2002 | 39% | 67% | 5 | 2327 |
| (22) | Honduras | 8% | 1961-1974 | 61% | 72% | 2 | 565 |
| (23) | India | 5% | 1983-2009 | 43% | 52% | 6 | 3326 |
| (24) | Indonesia | 12% | 1971-2010 | 35% | 59% | 8 | 26943 |
| (25) | Jamaica | 16% | 1991-2001 | 16% | 23% | 2 | 1087 |
| (26) | Kyrgyz Republic | 6% | 1999-2009 | 39% | 45% | 2 | 2653 |

Notes: GDP pc is GDP capita as a percent of US GDP in 2000. N. Obs. is the average number of observations within year $\times$ cohort cells.

Table A.2: IPUMS-I Countries - Part II

| | Country | (1) GDP pc | (2) Year Range | (3) Min. Agr. Emp. | (4) Max Agr. Emp. | (5) N. Surveys | (6) N. Obs. |
|---|---|---|---|---|---|---|---|
| (27) | Malawi | 2% | 1987-2008 | 38% | 70% | 3 | 4246 |
| (28) | Malaysia | 32% | 1970-2000 | 15% | 42% | 4 | 1399 |
| (29) | Mali | 2% | 1998-2009 | 59% | 70% | 2 | 4725 |
| (30) | Mexico | 30% | 1970-2015 | 12% | 41% | 6 | 36117 |
| (31) | Morocco | 11% | 1982-2004 | 25% | 35% | 3 | 6238 |
| (32) | Mozambique | 1% | 1997-2007 | 52% | 53% | 2 | 7177 |
| (33) | Nepal | 4% | 2001-2011 | 49% | 64% | 2 | 12739 |
| (34) | Nicaragua | 8% | 1971-2005 | 35% | 48% | 3 | 1615 |
| (35) | Panama | 22% | 1960-2010 | 16% | 50% | 6 | 1099 |
| (36) | Paraguay | 11% | 1962-2002 | 30% | 61% | 5 | 1391 |
| (37) | Peru | 14% | 1993-2007 | 22% | 27% | 2 | 13414 |
| (38) | Philippines | 10% | 1990-2010 | 33% | 40% | 4 | 40012 |
| (39) | Portugal | 53% | 1981-2011 | 3% | 16% | 4 | 3008 |
| (40) | Puerto Rico | 71% | 1970-2000 | 3% | 10% | 4 | 697 |
| (41) | Romania | 22% | 1977-2011 | 15% | 28% | 4 | 13611 |
| (42) | Rwanda | 2% | 2002-2012 | 59% | 65% | 2 | 3669 |
| (43) | Spain | 64% | 1981-2011 | 5% | 16% | 4 | 16067 |
| (44) | Switzerland | 115% | 1970-2000 | 4% | 10% | 4 | 1710 |
| (45) | Tanzania | 3% | 2002-2012 | 51% | 66% | 2 | 17102 |
| (46) | Thailand | 20% | 1970-2000 | 49% | 73% | 4 | 3044 |
| (47) | Trinidad and Tobago | 36% | 1980-2000 | 8% | 12% | 3 | 590 |
| (48) | Turkey | 29% | 1985-2000 | 23% | 32% | 3 | 17023 |
| (49) | United States | 100% | 1960-2015 | 1% | 8% | 7 | 38260 |
| (50) | Uruguay | 28% | 1963-2006 | 12% | 22% | 4 | 1593 |
| (51) | Venezuela | 32% | 1981-2001 | 10% | 13% | 3 | 8716 |
| (52) | Vietnam | 6% | 1989-2009 | 45% | 60% | 3 | 38285 |
| | | | *Extended Sample* | | | | |
| (53) | Colombia | 18% | 1964-1973 | 34% | 50% | 2 | 4722 |
| (54) | Iran | 27% | 2006-2011 | 11% | 15% | 2 | 8572 |
| (55) | Italy | 77% | 2011-2018 | 3% | 5% | 6 | 3480 |

Notes: GDP pc is GDP capita as a percent of US GDP in 2000. N. Obs. is the average number of observations within year × cohort cells.

Table A.3: IPUMS-GH Countries

| | Country | (1) GDP pc | (2) Year Range | (3) Min. Agr. Emp. | (4) Max Agr. Emp. | (5) N. Surveys | (6) N. Obs. |
|---|---|---|---|---|---|---|---|
| (1) | Benin | 4% | 1996-2011 | 41% | 61% | 4 | 69 |
| (2) | Burkina Faso | 2% | 1993-2010 | 66% | 78% | 4 | 69 |
| (3) | Cameroon | 5% | 1998-2011 | 38% | 51% | 3 | 86 |
| (4) | Ethiopia | 2% | 2005-2016 | 71% | 81% | 3 | 199 |
| (5) | Ghana | 5% | 1993-2014 | 35% | 55% | 5 | 62 |
| (6) | Guinea | 3% | 1999-2012 | 50% | 57% | 3 | 55 |
| (7) | Mali | 2% | 2001-2012 | 58% | 59% | 2 | 78 |
| (8) | Nepal | 4% | 2001-2011 | 33% | 64% | 3 | 75 |
| (9) | Niger | 1% | 1992-2012 | 47% | 75% | 4 | 62 |
| (10) | Rwanda | 2% | 2000-2014 | 61% | 67% | 3 | 87 |
| (11) | Senegal | 4% | 1992-2015 | 23% | 41% | 6 | 61 |
| (12) | Zambia | 4% | 2001-2013 | 44% | 51% | 3 | 138 |
| | | | *Extended Sample* | | | | |
| (13) | Bangladesh | 4% | 1994-2004 | 36% | 48% | 4 | 91 |
| (14) | Burundi | 2% | 2010-2016 | 64% | 66% | 2 | 112 |
| (15) | Congo Democratic Republic | 1% | 2007-2013 | 43% | 50% | 2 | 133 |
| (16) | India | 5% | 2005-2015 | 32% | 34% | 2 | 2087 |
| (17) | Kenya | 5% | 1993-2014 | 24% | 44% | 5 | 106 |
| (18) | Lesotho | 3% | 2009-2014 | 21% | 34% | 2 | 54 |
| (19) | Madagascar | 3% | 2003-2008 | 64% | 72% | 2 | 111 |
| (20) | Malawi | 2% | 1992-2016 | 40% | 59% | 5 | 88 |
| (21) | Mozambique | 1% | 2003-2011 | 42% | 59% | 2 | 63 |
| (22) | Nigeria | 6% | 1999-2008 | 33% | 39% | 3 | 142 |
| (23) | Tanzania | 3% | 1991-1999 | 66% | 73% | 3 | 50 |
| (24) | Uganda | 3% | 1995-2016 | 46% | 71% | 5 | 56 |
| (25) | Zimbabwe | 5% | 2005-2015 | 19% | 29% | 3 | 145 |

Notes: GDP pc is GDP capita as a percent of US GDP in 2000. N. Obs. is the average number of observations within year × cohort cells.

## B Dataset on Policy Reforms and Political Events

This section describes the procedure used to build the new dataset of educational policy reforms and of political events. The full list can be found in Tables A.4-A.7.

### B.1 Educational Reforms

We constructed a new dataset of education system reforms that increased the years of compulsory education. The same procedure was used to research all the countries for which we had data on cohort-level education and agricultural employment. First, a general search was done to identify potential reforms to compulsory education. This first search used Wikipedia, Eurydice, reference encyclopedias, Google, and Google Scholar. The identified reforms were then cross-checked against other sources and the following details were recorded: the year the reform went into effect, the age at which individuals ended compulsory education after the reform went into effect, and the total years of compulsory education after the reform went into effect.

The search proceeded as follows. First, we checked Wikipedia for articles on each country's education system and its history. Then, we consulted encyclopedias and (for European Union countries) Eurydice. Next, we performed Google and Google Scholar searches of the country's name combined with the following terms: "compulsory education", "compulsory education history", "education history", "education reform", "education reform history". Reforms were found using all these sources, but typically Google Scholar searches were the most informative. If a country's name changed over the sample period, the same searches were done with any other relevant name. For countries that gained independence over the sample period, we did searches covering both pre- and post-independence periods. If no national-level reforms to compulsory education were found for a specific country, we checked the current status of compulsory education, briefly surveyed the country's history, and checked the level of government at which education policy was set to confirm that no reforms to compulsory education should be expected. We did not include any reforms that were introduced solely at sub-national levels, unless the sub-national unit later became an independent country (though some of the included national-level reforms were implemented differentially across regions, as discussed below).

When a reform was found, the details listed above were recorded. When the initial source did not contain all the necessary information, additional sources (primarily on Google Scholar) were consulted. The recorded details for all reforms were then cross-checked against other sources through reform-specific Google Scholar and Google searches. In some cases, especially for low-income countries in the early part of the sample period, it was occasionally difficult to find multiple sources with all the details of the reform. In these cases, efforts were made to at least check that the available sources did not contradict each other.

Finally, reforms were flagged as "weakly-implemented" if the consulted sources highlighted one or more the following: (i) the implementation of the reform was limited due to lack of the necessary state capacity; (ii) the reform was de facto implemented differentially across different regions or areas of the country (including cases where the reform was differentially binding because of pre-existing disparities in the enforcement of compulsory education); (iii) the reform was phased in slowly over time (for these cases, the first year of the phase in was recorded as the year of the reform went into effect). An example of weak implementation is Guatemala, where de jure compulsory primary education was established in 1945; however, the consulted sources suggest that there were both issues in the state provision of primary education and limited enforcement of the rules. Another example is Trinidad and Tobago, where universal primary education was introduced in

1956; while the government aimed to increase the provision and uptake of primary education as much as possible, it did not introduce strict rules compelling children to attend school.

### B.2 Political Events

We also constructed a new dataset of major political events. The data construction was primarily based on the Wikipedia articles covering each country's history. Whenever the Wikipedia article for a given country was insubstantial, Google and Google Scholar searches were used. We divided political events into three categories: independence, democratization, and other. If an event could fit into multiple categories, we placed it into the independence category first, and the democratization category second.

Some countries have had rapid successions of different governments. In those instances, we tried to include the most important dates and to code them based on the medium-term rather than short-term outcomes. For example, Brazil in the 1930s had a democratic revolution, followed a by single democratic election a few years later, before the declaration of single party rule a few years after that. Since the same politician headed the state from the original revolution through the establishment of single party rule, we only include the initial revolution and classify it in the "other" category rather than democratization.

For what concerns independence, we include independence from a colonial power and/or the formation of a new nation state. Examples are Ghana in 1967 and Poland in 1919. We use the date the country officially became independent. We do not code the end of occupations as independence, based on the reasoning that the transition from a longstanding colonial government to self-rule is likely to bring about a different extent of changes in priorities compared to a return to self-rule after an occupation. For example, the end of German occupation of Czechoslovakia or the end of Italian occupation of Ethiopia are not coded as independence. In both cases, the occupying country was at war for most of the occupation period, and the occupied country had a well established independence before the occupation.

The second category of events is democratization, which we date to the first democratic election. We choose this date based on the reasoning that democratically electing a government should be more relevant to education than the date of a revolution or coup that leads to democratization; moreover, election dates are more easily identified. We do not code as democratizations cases where the first democratic election was immediately followed by an undemocratic transfer of power. Our reasoning here is that if there are benefits to democracy, they are likely to be accumulated because of the constraints and accountability democracy introduces. In the case of a single democratic election followed by the declaration of single party rule, a dictatorship, a coup, or a revolution before the subsequent election, it is not obvious whether the democratically elected government was subject to these constraints.

All other political events are treated as a single category. This includes military coups, the outbreak of violent conflict, democratic revolutions not followed by multiple elections, and the establishment and fall of communist and single-party regimes.

We cross-checked our dataset against the dataset of democratization and democratic reversal in Acemoglu et al. (2019). While our dataset covers a broader range of events and a longer time period, our data on democratization matches their dataset quite closely. Where differences exist, they are accounted for by the methodological choices discussed above.

Table A.4: List of Education Reforms Austria-Romania

| Country | Year | Weakly-Implemented | Description |
|---|---|---|---|
| Austria | 1948 | | Marshall Plan supported education reform. Compulsory education extended by two years. |
| Austria | 1962 | | Compulsory education extended by one year. |
| Benin | 1991 | | Program for universal primary education. |
| Bolivia | 1955 | | Primary education made compulsory. |
| Bolivia | 1969 | 1 | Compulsory education extended by two years. Limited implementation |
| Brazil | 1934 | | Four years of primary education made compulsory. |
| Brazil | 1971 | | Compulsory education extended by four years. |
| Chile | 1965 | | Compulsory education extended by two years. |
| Ecuador | 1983 | | Six years of education made compulsory. |
| France | 1936 | | Zay Reform. Compulsory education extended by one years. |
| France | 1967 | | Berthoin Edict. Compulsory education extended by two years. |
| Ghana | 1987 | | Compulsory education extended by 9 years. |
| Greece | 1976 | | Compulsory education extended by 3 years. |
| Guatemala | 1945 | 1 | Primary education made compulsory. Weak reform due to low state capacity. |
| India | 1947 | 1 | New constitution established that education will be compulsory to age 14. Weak reform due to low state capacity. |
| India | 1968 | 1 | Changed implementation and enforcement of compulsory education up to age 14. Weak reform due to limited regional coverage and low state capacity. |
| Jamaica | 1957 | 1 | Goal of universal primary education set. Not fully implemented until the 1960s. Weak reform due to slow phase in. |
| Jamaica | 1966 | 1 | Compulsory education extended by three years with assistance of international NGOs. Weak reform due to low state capacity. |
| Mali | 1962 | 1 | Nine years of compulsory education begins to be phased in. Weak reform due to slow phase in. |
| Nepal | 1981 | 1 | Five years of compulsory education established. Limited implementation. Weak reform due to low state capacity. |
| Portugal | 1952 | 1 | Introduction of new enforcement for compulsory education laws. Weak reform due to limited regional coverage. |
| Portugal | 1964 | | Compulsory education extended by two years. |
| Portugal | 1973 | | Compulsory education extended by two years. |
| Portugal | 1986 | | Compulsory education extended by one year. |
| Romania | 1948 | | Four years of compulsory education established. Previously there was de jure seven years of compulsory education but with little enforcement. |
| Romania | 1969 | 1 | Compulsory education raised to 6 years. Implemented from 1969 to 1977. Weak reform due to slow phase in. |
| Romania | 1985 | | Compulsory education raised to 8 years. |

Table A.5: List of Education Reforms Rwanda-Turkey

| Country | Year | Weakly-Implemented | Description |
|---|---|---|---|
| Rwanda | 1962 | 1 | New constitution includes provision for compulsory education. Weak reform due to low state capacity. |
| Rwanda | 1977 | | Compulsory education extended by two years. |
| Slovenia | 1950 | | Compulsory education extended by one year. |
| Spain | 1945 | | Six years of compulsory schooling established. |
| Spain | 1964 | | Compulsory education extended by two years. |
| Spain | 1990 | | Compulsory education extended by two years. |
| Thailand | 1921 | 1 | Four years of compulsory schooling established. Weak reform due to limited implementation. |
| Thailand | 1951 | 1 | Compulsory education extended. Weak reform due to low state capacity. |
| Thailand | 1980 | 1 | Compulsory education extended. Weak reform due to limited implementation. |
| Trinidad and Tobago | 1956 | 1 | Program for universal primary education begins to be phased in. Weak reform due to low state capacity. |
| Turkey | 1973 | | Primary education made compulsory. |
| Venezuela | 1981 | | Nine years of compulsory schooling established. |

Table A.6: List of Political Events, Benin-Honduras

| Country | Year | Description |
|---|---|---|
| Benin | 1960 | Independence from France. |
| Benin | 1972 | Military coup followed by socialist government and single party state. |
| Benin | 1991 | Democratic elections. |
| Bolivia | 1952 | Socialist revolution. |
| Bolivia | 1964 | Military coup. |
| Bolivia | 1982 | Democratic elections. |
| Brazil | 1930 | Coup followed by a single democratic election then a shift to autocracy. |
| Brazil | 1945 | Autocrat deposed. Democratic elections. |
| Brazil | 1964 | Military coup. |
| Brazil | 1985 | Democratic elections. |
| Botswana | 1964 | Independence from the United Kingdom. |
| Chile | 1973 | Military coup. |
| China | 1949 | Establishment of single-party communist regime. |
| China | 1966 | Beginning of Cultural Revolution. |
| Costa Rica | 1948 | Democratic revolution. |
| Dominican Republic | 1930 | Start of Trujillo's autocratic regime. |
| Ecuador | 1925 | Military coup. Begins periods of political instability. |
| Ecuador | 1948 | Democratic elections. |
| Ecuador | 1961 | Military coup, followed in 1963 by another coup. |
| Ecuador | 1966 | Democratic elections. |
| Ecuador | 1972 | Military coup. |
| Ecuador | 1979 | Democratic elections. |
| Spain | 1931 | Democratic elections, start of Second Republic. |
| Spain | 1936 | Outbreak of civil war. |
| Spain | 1975 | Beginning of transition to democracy. |
| Spain | 1982 | Democratic elections. |
| Ethiopia | 1947 | End of Italian occupation that began in lead up to WWII. |
| Ethiopia | 1974 | Marxist Coup |
| Fiji | 1963 | Democratic reforms. |
| Fiji | 1970 | Independence from the United Kingdom. |
| Ghana | 1957 | Independence from the United Kingdom. |
| Ghana | 1966 | Military coup. |
| Ghana | 1972 | Military coup. |
| Ghana | 1979 | Military coup. |
| Greece | 1924 | Coup followed by democratic elections. Start of Second Republic. |
| Greece | 1935 | Shift to single party rule. |
| Greece | 1949 | End of civil war, democratic elections. |
| Greece | 1967 | Military coup. |
| Greece | 1974 | Democratic elections. |
| Guatemala | 1944 | Democratic elections. |
| Guatemala | 1954 | Autocratic regime established. |
| Guatemala | 1961 | Start of Guatemalan Civil War. |
| Honduras | 1912 | Democratic elections. |

Table A.7: List of Political Events, Honduras-Vietnam

| Country | Year | Description |
|---|---|---|
| Honduras | 1920 | Manipulated election begins period of political instability and civil conflict. |
| Honduras | 1928 | Democratic elections. |
| Indonesia | 1949 | Independence from the Netherlands. |
| India | 1947 | Independence from the United Kingdom. |
| Jamaica | 1962 | Independence from the United Kingdom. |
| Mexico | 1920 | Democratic elections. |
| Mali | 1960 | Independence from France. |
| Mali | 1968 | Military coup. |
| Mozambique | 1964 | Outbreak of civil war. |
| Mozambique | 1975 | Independence from Portugal. |
| Malawi | 1964 | Independence from the United Kingdom. |
| Malaysia | 1957 | Independence from the United Kingdom. |
| Nicaragua | 1984 | Democratic elections. |
| Nepal | 1951 | Shift to limited democracy after popular revolution. |
| Nepal | 1960 | Monarchy reasserts more direct control. |
| Nepal | 1980 | Reforms reduce power of monarchy. |
| Peru | 1948 | Military coup. |
| Peru | 1980 | Democratic elections. |
| Philippines | 1946 | Independence from the United States. |
| Philipinnes | 1972 | Declaration of martial law ends Third Republic. |
| Philipinnes | 1986 | Democratic elections. |
| Portugal | 1975 | Democratic elections following coup in 1974. |
| Paraguay | 1954 | Autocratic regime established. |
| Romania | 1945 | Establishment of single-party communist state. |
| Romania | 1989 | End of communist rule. |
| Rwanda | 1962 | Independence from Belgium. |
| Rwanda | 1974 | Military coup establishes stable autocracy. |
| Rwanda | 1990 | Civil War. |
| Thailand | 1932 | Establishment of constitutional monarchy. |
| Thailand | 1946 | Democratic elections. |
| Thailand | 1947 | Military coup. |
| Thailand | 1957 | Military coup. |
| Thailand | 1973 | Democratic revolution. |
| Thailand | 1976 | Military coup. |
| Trinidad and Tobago | 1962 | Independence from United Kingdom. |
| Turkey | 1946 | First multi-party elections. |
| Turkey | 1960 | Military coup begins period of political instability. |
| United States | 1954 | Desegregation of schools begins. |
| Venezuela | 1945 | Democratic revolution. |
| Venezuela | 1948 | Military coup. |
| Venezuela | 1958 | Democratic revolution. |
| Vietnam | 1954 | Independence from France and partition into North and South Vietnam. |
| Vietnam | 1975 | End of Vietnam War. |

## C Sectoral Value Added Data

We construct data on nominal and real value added per worker by combining three sources: the GGDC 10-Sector Database (Timmer et al., 2015), the Economic Transformation Database (de Vries et al., 2021), and the World Development Indicators (World Bank, 2017). We use the GGDC 10-Sector Database as baseline, imputing the data for missing years based on the corresponding growth rates from the Economic Transformation Database or World Development Indicators. For each country and variable, we compute the average growth rate as the simple average of the growth rate between the years covered by the IPUMS-I cross-sections (as we do when computing the average reallocation rate of agriculture). In case of partially missing data, we use the growth rates computed on the available years whenever they represent at least half of the years between two consecutive IPUMS-I cross-sections.

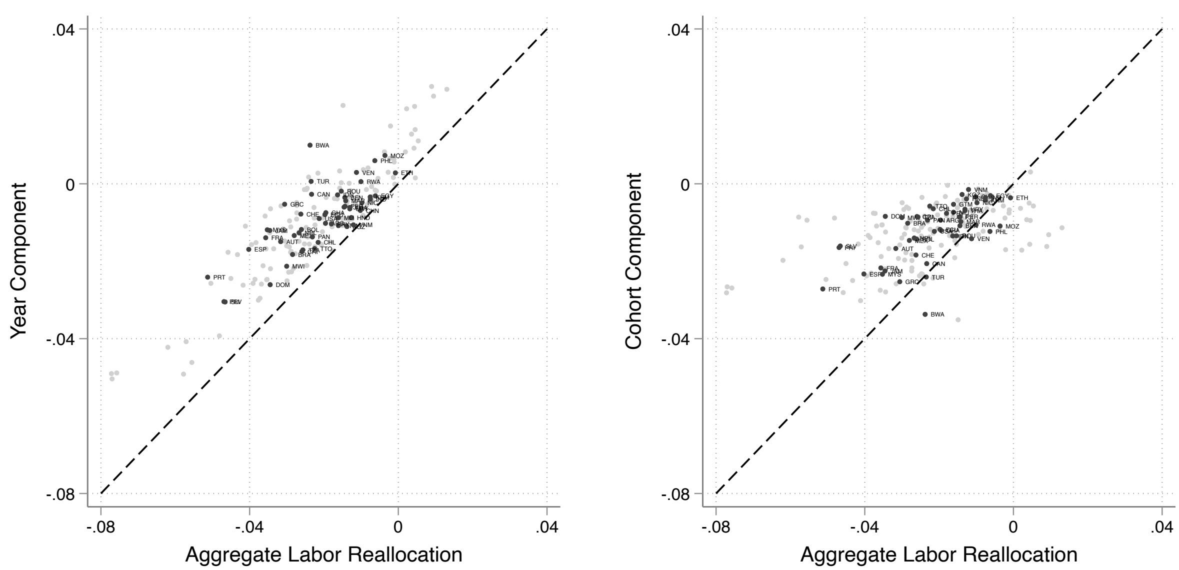
# B Decomposing Structural Change: Additional Results

## A Disaggregated Results

Figures B.1a and B.1b plot the year and cohort components against the average reallocation rate. For the overwhelming majority of countries, both the year and the cohort components are negative, hence they contribute to aggregate labor reallocation. This is also the case if we treat each cross-section as an independent observation (lighter dots). Furthermore, countries with faster reallocation have usually larger (in absolute value) year and cohort components, although the year components explain a larger share of cross-country variance.

Figure B.1: Unpacking Aggregate Labor Reallocation

(a) Year Components (b) Cohort Components




Notes: The two figures plot the year and cohort component against the aggregate rate of labor reallocation out of agriculture, both across countries (dark dots) and across all cross-sections (light dots) The dotted line is the 45 degree line.

The following Tables report country-specific decomposition results, both for IPUMS-I and IPUMS-GH countries.

Table B.1: Unpacking Structural Change, IPUMS-I Countries - Part I

| | Country | (1) Year Range | (2) $\log g_{L_A}$ | (3) $\log \psi$ | (4) $\log \tilde{\psi}$ |
|---|---|---|---|---|---|
| (1) | Argentina | 1970-1980 | -1.96 | -1.02 | -1.41 |
| (2) | Austria | 1971-2011 | -3.17 | -1.49 | -0.85 |
| (3) | Benin | 1979-2013 | -1.44 | -0.36 | -0.23 |
| (4) | Bolivia | 1976-2001 | -2.61 | -1.18 | -1.81 |
| (5) | Botswana | 1991-2011 | -2.37 | 1.00 | -0.85 |
| (6) | Brazil | 1960-2010 | -2.85 | -1.83 | -2.62 |
| (7) | Cambodia | 1998-2013 | -0.76 | -0.34 | -0.46 |
| (8) | Canada | 1971-2011 | -2.33 | -0.27 | 0.25 |
| (9) | Chile | 1960-2002 | -2.16 | -1.51 | -1.88 |
| (10) | China | 1982-2000 | -1.02 | -0.69 | -1.38 |
| (11) | Costa Rica | 1963-2011 | -2.60 | -1.75 | -2.79 |
| (12) | Dominican Republic | 1960-1970 | -3.45 | -2.61 | -3.03 |
| (13) | Ecuador | 1962-2010 | -1.97 | -0.80 | -1.22 |
| (14) | Egypt | 1986-2006 | -0.62 | -0.31 | -0.67 |
| (15) | El Salvador | 1992-2007 | -4.66 | -3.05 | -4.56 |
| (16) | Ethiopia | 1984-1994 | -0.08 | 0.28 | 0.36 |
| (17) | Fiji | 1966-2014 | -1.46 | -0.60 | -0.99 |
| (18) | France | 1962-2011 | -3.57 | -1.39 | -1.40 |
| (19) | Ghana | 1984-2010 | -1.95 | -0.75 | -1.29 |
| (20) | Greece | 1971-2011 | -3.06 | -0.53 | -0.74 |
| (21) | Guatemala | 1964-2002 | -1.61 | -1.08 | -1.28 |
| (22) | Honduras | 1961-1974 | -1.26 | -0.87 | 0.14 |
| (23) | India | 1983-2009 | -0.75 | -0.39 | -1.12 |
| (24) | Indonesia | 1971-2010 | -1.63 | -0.29 | -0.86 |
| (25) | Jamaica | 1991-2001 | -3.46 | -1.20 | -1.83 |
| (26) | Kyrgyz Republic | 1999-2009 | -1.38 | -1.10 | -1.45 |
| (27) | Malawi | 1987-2008 | -3.00 | -2.13 | -2.00 |

Table B.2: Unpacking Structural Change, IPUMS-I Countries - Part II

| | Country | (1) Year Range | (2) $\log g_{L_A}$ | (3) $\log \psi$ | (4) $\log \tilde{\psi}$ |
|---|---|---|---|---|---|
| (28) | Malaysia | 1970-2000 | -3.52 | -1.19 | -1.49 |
| (29) | Mali | 1998-2009 | -1.61 | -0.88 | -1.25 |
| (30) | Mexico | 1970-2015 | -2.80 | -1.34 | -1.85 |
| (31) | Morocco | 1982-2004 | -1.41 | -0.44 | -1.34 |
| (32) | Mozambique | 1997-2007 | -0.36 | 0.73 | 0.48 |
| (33) | Nepal | 2001-2011 | -2.67 | -1.27 | -2.12 |
| (34) | Nicaragua | 1971-2005 | -0.98 | -0.49 | -1.31 |
| (35) | Panama | 1960-2010 | -2.31 | -1.37 | -1.94 |
| (36) | Paraguay | 1962-2002 | -1.80 | -1.04 | -1.91 |
| (37) | Peru | 1993-2007 | -1.43 | -0.58 | -0.56 |
| (38) | Philippines | 1990-2010 | -0.63 | 0.60 | 0.99 |
| (39) | Portugal | 1981-2011 | -5.13 | -2.41 | -2.52 |
| (40) | Puerto Rico | 1970-2000 | -4.69 | -3.04 | -3.35 |
| (41) | Romania | 1977-2011 | -1.53 | -0.19 | -0.77 |
| (42) | Rwanda | 2002-2012 | -1.01 | 0.06 | -0.21 |
| (43) | Spain | 1981-2011 | -4.02 | -1.69 | -1.86 |
| (44) | Switzerland | 1970-2000 | -2.62 | -0.78 | -0.64 |
| (45) | Tanzania | 2002-2012 | -2.57 | -1.71 | -2.26 |
| (46) | Thailand | 1970-2000 | -1.31 | -0.61 | -0.94 |
| (47) | Trinidad and Tobago | 1980-2000 | -2.25 | -1.67 | -0.59 |
| (48) | Turkey | 1985-2000 | -2.34 | 0.06 | -3.19 |
| (49) | United States | 1960-2015 | -2.13 | -0.90 | -1.34 |
| (50) | Uruguay | 1963-2006 | -1.31 | -0.65 | -0.62 |
| (51) | Venezuela | 1981-2001 | -1.13 | 0.29 | -0.15 |
| (52) | Vietnam | 1989-2009 | -1.21 | -1.06 | -1.71 |

Table B.3: Unpacking Structural Change, IPUMS-GH Countries

| | Country | (1)<br>Year Range | (2)<br>$\log g_{L_A}$ | (3)<br>$\log \psi$ | (4)<br>$\log \tilde{\psi}$ |
|---|---|---|---|---|---|
| (1) | Benin | 1996-2011 | -2.69 | -1.08 | -2.51 |
| (2) | Burkina Faso | 1993-2010 | -1.10 | 0.29 | -0.85 |
| (3) | Cameroon | 1998-2011 | -2.29 | -0.07 | -1.09 |
| (4) | Ethiopia | 2005-2016 | -1.19 | -0.20 | -0.46 |
| (5) | Ghana | 1993-2014 | -2.16 | -0.55 | -0.32 |
| (6) | Guinea | 1999-2012 | -0.86 | 0.87 | 2.07 |
| (7) | Mali | 2001-2012 | -0.18 | 0.39 | 0.17 |
| (8) | Nepal | 2001-2011 | -6.77 | -4.60 | -6.45 |
| (9) | Niger | 1992-2012 | -1.06 | -0.19 | -1.44 |
| (10) | Rwanda | 2000-2014 | -1.19 | 0.16 | -0.64 |
| (11) | Senegal | 1992-2015 | -1.43 | 1.05 | -0.06 |
| (12) | Zambia | 2001-2013 | -1.16 | 0.04 | -0.50 |

# B Alternative Decompositions

Table B.4: Unpacking Structural Change, Robustness

| | Exercise | (1) $\log g_{L_A}$ | (2) $\log \psi$ | (3) $\frac{\log \chi}{\log g_{L_A}}$ | (4) $\log \tilde{\psi}$ | (5) $\frac{\log \tilde{\chi}}{\log g_{L_A}}$ | (6) $1 - \frac{\log \psi}{\log \tilde{\psi}}$ | (7) N. Obs |
|---|---|---|---|---|---|---|---|---|
| (1) | **Benchmark** | -2.11 | -0.92 | 0.56 | -1.32 | 0.38 | 0.30 | 52 |
| (2) | Extended Sample | -2.25 | -1.01 | 0.55 | -1.33 | 0.41 | 0.24 | 55 |
| (3) | Five-Years Age Dummies | -2.12 | -0.95 | 0.55 | -1.24 | 0.42 | 0.23 | 47 |
| (4) | Country-specific $\bar{a}_j$ | -2.14 | -0.95 | 0.56 | -1.30 | 0.39 | 0.27 | 52 |
| (5) | One-Year Age Dummies | -2.14 | -0.95 | 0.56 | -1.23 | 0.42 | 0.23 | 52 |
| (6) | Four + Cross-Sections | -2.38 | -1.06 | 0.56 | -1.30 | 0.46 | 0.18 | 27 |
| (7) | Time-Specific Age Controls | -2.38 | -1.06 | 0.56 | -1.29 | 0.46 | 0.18 | 27 |
| (8) | Employed Only | -1.97 | -0.31 | 0.84 | -1.20 | 0.39 | 0.74 | 51 |
| (9) | **IPUMS-GH** | -1.84 | -0.32 | 0.82 | -1.00 | 0.45 | 0.68 | 12 |
| (10) | IPUMS-GH, Extended Sample | -1.87 | -0.59 | 0.68 | -1.36 | 0.27 | 0.56 | 25 |

Notes: the Table shows results from the decomposition of labor reallocation according to different specifications or sample restrictions. Row (1) includes the specification shown in the main text, as a benchmark. Row (2) considers the extended sample of countries, as described in Section II. Row (3) uses, as age controls, five year dummies under the linear restriction that the first two are equal. For this specification, we need to include countries whose cross-sections are at most 10 years apart, hence the sample reduces slightly. Row (4) runs the same specification as in the main text, but considers a country-specific $\bar{a}_j$, computed as we did for $\bar{a}$. Row (5) includes a full set of yearly age dummies, under the minimal linear restriction that we can impose in each country to properly identify year and cohort effects. For example, if we observe cross-sections five years apart, we constraint the first six age dummies to be identical. Row (6) only keeps countries for which we have at least four cross-sections. It serves as a comparison for row (7), where we use this restricted set of countries and allow the age effects to vary over time. Specifically, we split, within each country, the time sample in two, and we allow the age effects to be time-period specific. Comparing rows (6) and (7) show that the data are not consistent with frictions changing over time (the two rows are not identical if we include the third decimal point; while within some countries there are larger differences, they cancel out on average). Row (8) computes agricultural employment not as a percentage of the total population, but as a percentage of the employed. In view of Figure A.2c, we restrict the sample to individuals between age 35 and 55, which are less likely to be biased by selection in and out of employment (this results in one less observation, as the stricter age restriction makes the two cross-sections in Botswana not overlapping). Rows (9) and (10) show results for IPUMS-GH countries.

Table B.5: Unpacking Structural Change - Medians

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Country Type | $\log g_{L_A}$ | $\log \psi$ | $\log \tilde{\psi}$ | $\frac{\log \chi}{\log g_{L_A}}$ | $\frac{\log \tilde{\chi}}{\log g_{L_A}}$ | $1 - \frac{\log \psi}{\log \tilde{\psi}}$ | N. Obs |
| All | -1.97 | -0.88 | -1.30 | 0.53 | 0.34 | 0.32 | 52 |
| High Income | -3.17 | -1.39 | -1.34 | 0.58 | 0.61 | 0.09 | 9 |
| Middle Income | -2.07 | -1.03 | -1.37 | 0.51 | 0.28 | 0.34 | 24 |
| Low Income | -1.26 | -0.69 | -1.25 | 0.53 | 0.32 | 0.38 | 19 |

Notes: the Table shows results the decomposition results based on medians.

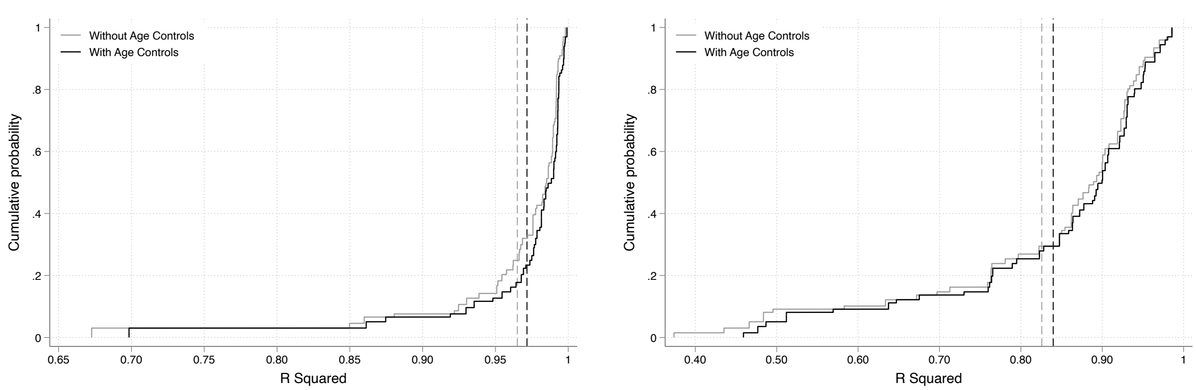
## C Goodness of Fit

Panel A of Figure B.2 displays the cumulative cross-country distribution of the R squared from specification 5, with and without age controls. The cohort and year dummies absorb most of the variation in the data, while the inclusion of age controls affects the R squared only marginally (the cross-country average R squared is 0.965 without age controls and 0.972 with age controls). Panel B shows the corresponding cumulative distribution for specifications that control for cohort and time trends, as opposed to a full set of cohort and year dummies. The average R squared with (without) age controls is 0.84 (0.826), with the majority of countries having values above 0.90. This shows that the empirical patterns of labor reallocation out of agriculture are well approximated by a combination of a cohort- and a year-level factor growing at constant rates, as in our model.

Figure B.2: Goodness of Fit

(a) Baseline Specification

(b) Linear Cohort and Time Trends



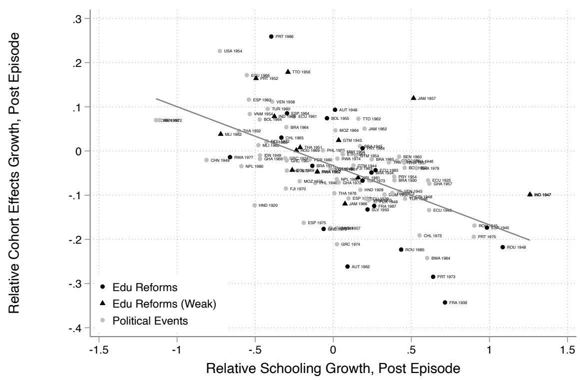

Notes: The two figures plot the cumulative distributions of the country-specific R squared from regressions of cohort-level log agricultural employment on cohort and year dummies (Panel A) and linear cohort and time trends (Panel B), with and without age controls. The dashed lines denote the cross-country average R squared for the corresponding specification.

# C Understanding Cohort Effects: Additional Results

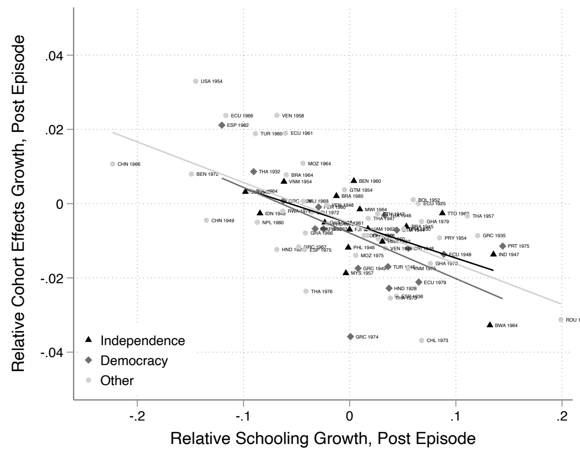
## A Educational Reforms and Political Events

Figure C.1: Trend Breaks - Alternative Specifications

(a) Alternative Growth Rate Computation



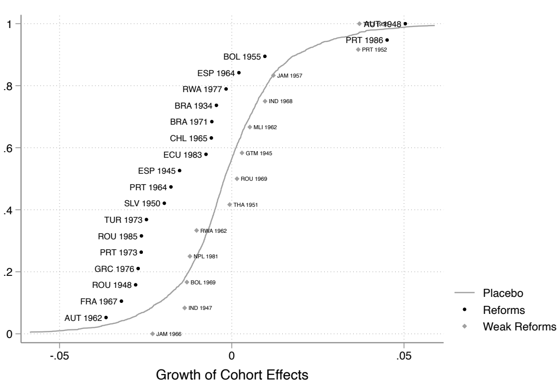

(b) Political Events by Type



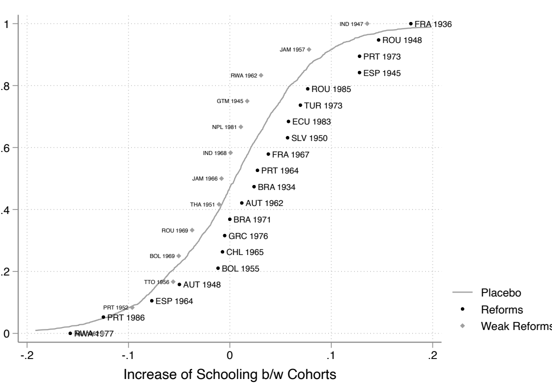

Notes: Panel A displays trend breaks are computed as the average gaps, over the first 10 affected cohorts, relative to an extrapolated linear trend that starts from the average value for the 10 youngest not affected cohorts and grows at the average annual pace observed across the 10 youngest not affected cohorts. Panel B displays trend breaks (computed as in the main text) by type of political event. The lines show the best linear fit for the relevant sample.

Figure C.2: Comparison of Policy Reforms Trend Breaks with Placebo

(a) Agricultural Employment



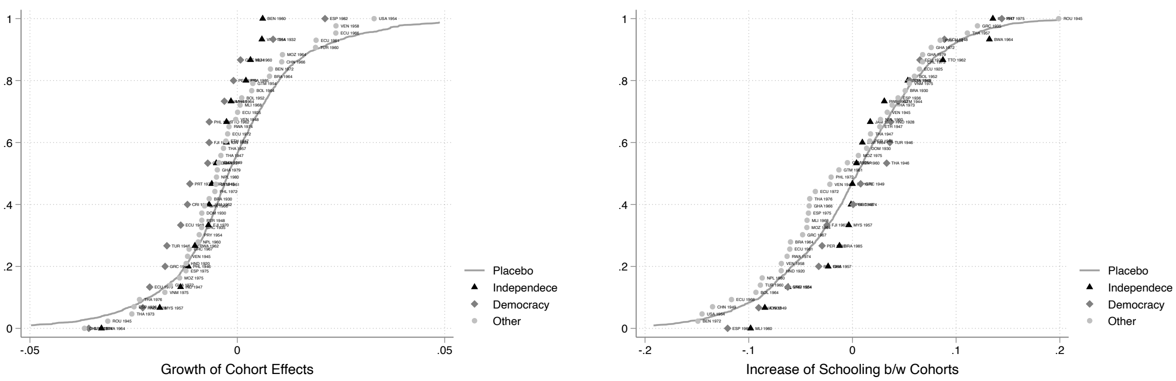

(b) Schooling




Notes: the figures compare the cumulative distribution of trend breaks, for either cohort effects (left) or schooling (right), around the policy reforms with the cumulative distribution of placebo trend breaks. The placebo trend breaks are obtained by applying around all birth cohorts in our data the same procedure used to compute the trend breaks around the policy reforms. The figures show that the trend breaks associated to the fully-implemented reforms have larger increases in schooling and decreases in cohort effects relative to the placebo. The trend breaks associated to the weakly-implemented reforms are distributed similarly to the placebo trend breaks.

Figure C.3: Comparison of Political Events Trend Breaks with Placebo

(a) Agricultural Employment (b) Schooling




Notes: this figure is identical to Figure C.2, except that it includes political events rather than education reforms.

## B Comparing Magnitudes between Sections C and A

The magnitude of the estimates in Sections C and A are not comparable, due to the different functional forms. This Appendix considers a cohort-level version of the specification in Section C to fill this gap.

We focus on the two treatment and control cohorts described in the main text. We compute for each cohort $\times$ district pair $(c, d)$ the share of agricultural employment $l_{A,c,d}$ and the average years of schooling $s_{c,d}$, and then estimate

$$\log l_{A,c,d} = \alpha_c + \eta_d + \beta s_{c,d} + \varphi_c \xi_d + \epsilon_{c,d},$$

where $s_{c,d}$ is instrumented by the interaction between the district-level program intensity $T_d$ and a dummy identifying the treated cohorts. The results are reported in Table C.1. The baseline IV estimate for $\beta$ is $-0.104$ (.094), negative but not precisely estimated (column 1). A possible confounding factor is that some cohort $\times$ district pairs display extreme values of the agricultural employment share, close to either 0 or 1. We do not necessarily expect to find an effect of schooling on agricultural participation for cohorts in districts where the agricultural sector is either the only viable employment option or basically non-existent. Consistently with this hypothesis, when we truncate the sample at either the $1^{st}$ and $99^{th}$ or at the $2^{nd}$ and $98^{th}$ percentiles in terms of agricultural employment, we obtain larger estimates for $\beta$ (columns 2 and 3). We conclude that the data are mostly consistent with estimates for $\beta$ in the range between $-0.10$ and $-0.20$, similar in magnitude to those reported in Section A.

Table C.1: School Construction in Indonesia: Cohort-Level Specifications

| | (1) Employed in Agri | (2) Employed in Agri | (3) Employed in Agri | (4) Employed in Non-Agri | (5) Employed in Non-Agri | (6) Employed in Non-Agri |
|---|---|---|---|---|---|---|
| Years of Schooling | -0.101 | -0.182 | -0.216 | 0.185 | 0.316 | 0.293 |
| | (0.114) | (0.126) | (0.134) | (0.084) | (0.136) | (0.131) |
| Cohort Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| District Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Truncation | - | 1-99 | 2-98 | - | 1-99 | 2-98 |
| F Stat First Stage | 14.44 | 12.19 | 11.22 | 13.93 | 9.65 | 9.10 |
| Observations | 1868 | 1830 | 1784 | 2160 | 2019 | 1998 |

Notes: Robust Standard Errors in Parentheses.

# D Derivations and Extensions

This Appendix includes the derivations of the model and its extensions. Section A includes the proofs of Lemmas 1-4, Propositions 1 and 2 and Corollary 1. Section B derives the implications on wages discussed in Section VI. Section C illustrates the extension with non-monetary returns from working in non-agriculture mentioned in Section VI.

## A Proofs

**Proof of Lemma 1.**

Denote the growth rate of wages as $\log g_{w_s} \equiv \log \frac{w_{s,t+1}}{w_{s,t}}$, for $s = \{A, M\}$. For the purpose of this Lemma, we assume that these growth rates are constant, with $g_{w_M} \geq g_{w_A}$; in Proposition 1 we will show that this holds in equilibrium. Notice from equation (12) that this implies that cohort-level human capital grows at a constant rate, with $\log g_h \equiv \log \frac{h_{c+1}}{h_c} = \log g_\xi + \sigma \log \frac{g_{w_M}}{g_{w_A}}$.

We assume and later verify that individuals never move from non-agriculture to agriculture. As a result, we only need to solve for the optimal timing of a move out of agriculture, if any. Let's consider an individual $(c, \varepsilon)$ that is currently in agriculture and at time $t$ has to decide whether to move out of agriculture or not. He would move out if two conditions are satisfied: (i) moving today is better than moving in some future period; (ii) moving today is better than not moving at all.

Consider first condition (i). An individual $(c, \varepsilon)$ prefers to move out of agriculture at $t$ rather than at $t+1$ if and only if

$$\begin{aligned}
\sum_{k=t}^{c+N} \beta^{k-t} (1-i)\, y_{M,k}(c,\varepsilon) - f y_{M,t}(c,\varepsilon) &\geq y_{A,t} + \\
+\beta \left( \sum_{k=t+1}^{c+N} \beta^{k-t-1} (1-i)\, y_{M,k}(c,\varepsilon) - f y_{M,t+1}(c,\varepsilon) \right) & \\
\sum_{k=t}^{c+N} \beta^{k-t} (1-i)\, y_{M,k}(c,\varepsilon) - f\left(y_{M,t}(c,\varepsilon) - \beta y_{M,t+1}(c,\varepsilon)\right) &\geq y_{A,t} + \beta \left( \sum_{k=t+1}^{c+N} \beta^{k-t-1} (1-i)\, y_{M,k}(c,\varepsilon) \right) \\
(1-i)\, y_{M,t}(c,\varepsilon) - (1-\beta g_{w_M}) f y_{M,t}(c,\varepsilon) &\geq y_{A,t} \\
y_{M,t}(c,\varepsilon) &\geq \frac{y_{A,t}}{1-i-(1-\beta g_{Z_M}) f}.
\end{aligned}$$

As we show below, Assumption 2 guarantees that $1-i-(1-\beta g_{w_M}) f \geq 0$, so that this inequality holds when $\varepsilon$ is large enough. This condition is satisfied when $\varepsilon \geq \hat{\varepsilon}_t(c)$, where

$$\hat{\varepsilon}_t(c) = \left[ h_c^{-\gamma} \left( \frac{w_{A,t}}{w_{M,t}} \right) \left( \frac{1}{1-i-(1-\beta g_{w_M}) f} \right) \right]^{\frac{1}{1-\gamma}}. \tag{23}$$

is the ability level of the marginal individual of cohort $c$ at time $t$. If $g_{w_M} \geq g_{w_A}$, $\hat{\varepsilon}_t(c)$ is decreasing over time: as the wage per efficiency unit grows faster in non-agriculture, individuals with lower and lower $\varepsilon$ gradually find it worthwhile to leave agriculture as opposed to spending an extra period there.

We can then verify that if it is better to move at time $t$ rather than at time $t+1$, then it is also better to move at time $t$ than at any time $t+x$. Following the previous derivation, an individual prefers to move at time $t$ rather than at time $t+x$ if and only if

$$y_{M,t}(c,\varepsilon) \geq \left(\frac{1}{1-i-(1-\beta g_{w_M})f}\right)\left(\frac{1-(\beta g_{w_A})^x}{1-(\beta g_{w_M})^x}\right)\left(\frac{1-\beta g_{w_M}}{1-\beta g_{w_A}}\right) y_{A,t}(c,\varepsilon).$$

Notice that, given $g_{w_M} \geq g_{w_A}$, $\frac{1-(\beta g_{w_A})^x}{1-(\beta g_{w_M})^x}$ is decreasing in $x$. Therefore, if it is better to move out of agriculture at time $t$ rather than at time $t+1$, then it must also be better to move at time $t$ rather than at any time $t+x$.

This implies that, conditional on moving at some point, individual $(c,\varepsilon)$ prefers to move in the first period $t$ when the condition $\varepsilon \geq \hat{\varepsilon}_t(c)$ is satisfied. Next, we need to verify whether moving at the preferred time is better than not moving at all. This is the case whenever the present discounted value of moving out of agricultural is higher than that of not moving, that is

$$\sum_{k=t}^{c+N} \beta^{k-t}(1-i)\, y_{M,k}(c,\varepsilon) - f y_{M,t}(c,\varepsilon) \quad \geq \quad \sum_{k=t}^{c+N} \beta^{k-t} y_{A,k}(c,\varepsilon).$$

$$w_{M,t} h(c,\varepsilon)\left((1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t} - f\right) \geq w_{A,t} \sum_{k=t}^{c+N}(\beta g_{w_A})^{k-t}$$

This inequality is satisfied for $\varepsilon \geq \tilde{\varepsilon}_t(c)$, where

$$\tilde{\varepsilon}_t(c) = \begin{cases} \left[h_c^{-\gamma}\left(\frac{w_{A,t}}{w_{M,t}}\right)\left(\frac{\sum_{k=t}^{c+N}(\beta g_{w_A})^{k-t}}{(1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t}-f}\right)\right]^{\frac{1}{1-\gamma}} & \text{if } (1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t} - f > 0 \\ \infty & \text{if } (1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t} - f \leq 0 \end{cases}$$

While $\tilde{\varepsilon}_t(c)$ could be either decreasing or increasing over time, it definitely decreases at a lower rate than $\hat{\varepsilon}_t(c)$, since as long as $g_{w_M} \geq g_{w_A}$, $\left(\frac{\sum_{k=t}^{c+N}(\beta g_{w_A})^{k-t}}{(1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t}-f}\right)$ is increasing in $t$ (moreover, as $t$ increases the condition $(1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t} - f \leq 0$ is more likely to be satisfied). As a result, there is a time $\hat{t}$ such that $\hat{\varepsilon}_t(c)$ and $\tilde{\varepsilon}_t(c)$ cross. For all individuals with $\varepsilon < \hat{\varepsilon}_{\hat{t}}(c) = \tilde{\varepsilon}_{\hat{t}}(c)$, moving at the preferred time is dominated by not moving at all. The individual with $\varepsilon = \hat{\varepsilon}_{\hat{t}}(c) = \tilde{\varepsilon}_{\hat{t}}(c)$ is the lowest type of cohort $c$ moving out of agriculture, and no other member of cohort $c$ moves out after time $\hat{t}$. We refer to cohort $c$ as "constrained" starting from time $\hat{t}$.

The agricultural share for cohort $c$ at time $t$ is then given by

$$l_{A,t,c} = \begin{cases} F\left(\hat{\varepsilon}_t\left(c\right)\right) & \text{if } a_t(c) \leq \hat{a}(f) \\ F\left(\hat{\varepsilon}_{c+\hat{a}}\left(c\right)\right) & \text{if } a_t(c) > \hat{a}(f) \end{cases}$$

which, using the expression for $\hat{\varepsilon}_t\left(c\right)$ derived above, can be written as

$$\log l_{A,t,c} = \begin{cases} \lambda(i,f) + \frac{v}{1-\gamma}\log\frac{w_{A,t}}{w_{M,t}} - \frac{v\gamma}{1-\gamma}\log h_c & \text{if } \ a_{t+1}(c) \leq \hat{a}(f) \\ \lambda(i,f) + \frac{v}{1-\gamma}\log\frac{w_{A,c+\hat{a}}}{w_{M,c+\hat{a}}} - \frac{v\gamma}{1-\gamma}\log h_c & \text{if } \ a_{t+1}(c) > \hat{a}(f) \end{cases}$$

where

$$\lambda(i,f) \equiv \frac{1}{1-\gamma}\log\frac{1}{1-i-\left(1-\beta g_{w_M}\right)f}.$$

and, as stated in Lemma 1, $\lambda(0,0) = 0$.[44] We can use these results to compute the initial conditions for the "initial old" at time 0 that ensure that their dynamic problem is symmetric to the one of the other cohorts and the economy starts on a constant reallocation path (see footnote 29). This will be the case if individuals in those cohorts (i) are endowed with a cohort-level human capital consistent with the equilibrium growth rate across cohorts, i.e. $h_{-x} = h_1 g_h^{-x}$ for all $x \in (1, N)$, and (ii) start off in the sector they would have chosen if they had a normal life span, given the growth rate of relative wages, i.e. for $c \leq -1$

$$\omega_{-1}(c,\varepsilon) = \begin{cases} 1 & \text{if } \ \varepsilon \leq \hat{\varepsilon}_0\left(c\right) \ \text{ and } \ a_1(c) \leq \hat{a}(f) \\ 1 & \text{if } \ \varepsilon \leq \hat{\varepsilon}_{c+\hat{a}}\left(c\right) \ \text{ and } \ a_1(c) > \hat{a}(f) \\ 0 & \text{if } \ \varepsilon > \hat{\varepsilon}_0\left(c\right) \ \text{ and } \ a_1(c) \leq \hat{a}(f) \\ 0 & \text{if } \ \varepsilon > \hat{\varepsilon}_{c+\hat{a}}\left(c\right) \ \text{ and } \ a_1(c) > \hat{a}(f) \end{cases}$$

where $\hat{\varepsilon}_{-x}\left(c\right) = \hat{\varepsilon}_1\left(c\right)\left(\frac{g_{w_A}}{g_{w_M}}\right)^{-\frac{1+x}{1-\gamma}}$ for all $x \in (1, N)$.

We now derive the upper bounds $\bar{f}$ and $\bar{i}$ such that Assumption 2 ensures that (i) cohort $c$ is not constrained in the first period where it is alive (period $c$) and (ii) within all cohorts, there are some individuals in non-agriculture. Cohort $c$ is unconstrained at time $t$ if $\hat{\varepsilon}_t\left(c\right) \geq \tilde{\varepsilon}_t\left(c\right)$, which is

[44] Notice that here we are implicitly assuming that $\hat{a}$ is an integer. This simplifies the exposition, as it allows us to abstract from the case where some individuals with $\varepsilon > \hat{\varepsilon}_{c+\hat{a}}(c)$ do not move out of agriculture because $\varepsilon < \hat{\varepsilon}_t(c)$ in the last period their age is below $\hat{a}$, and $\varepsilon < \tilde{\varepsilon}_t(c)$ in the first period their age is above $\hat{a}$ (an artifact of the discrete timing of the model). This simplification is without loss of generality, as we can always change the frequency of the model so that $\hat{a}$ perfectly coincides with the age of one of the active cohorts.

satisfied if

$$\left(\frac{\sum_{k=t}^{c+N}(\beta g_{w_A})^{k-t}}{(1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t}-f}\right)(1-i-(1-\beta g_{w_M})f) \quad \leq 1$$

$$f \leq (1-i)\frac{\sum_{k=a}^{N}(\beta g_{w_M})^{k-a}-\sum_{k=a}^{N}(\beta g_{w_A})^{k-a}}{1-(1-\beta g_{w_M})\sum_{k=a}^{N}(\beta g_{w_A})^{k-a}} \equiv \phi(a),$$

where $a$ is the age of the cohort: $a = t - c$. Notice that $\phi(a) < (1-i)\sum_{k=t}^{c+N}(\beta g_{w_M})^{k-t}$, which implies that if $f \leq \phi(a)$ then $\tilde{\varepsilon}_t(c) < \infty$. Importantly, $\phi(a)$ does not directly depend on time $t$ or cohort $c$, but only on age $a$. In the first period that a cohort is alive, i.e. for $t = c$ and $a = 0$, the inequality is satisfied if $f \leq \bar{f}$, where

$$\bar{f} \equiv (1-i)\frac{\sum_{k=0}^{N}(\beta g_{w_M})^{k}-\sum_{k=0}^{N}(\beta g_{w_A})^{k}}{1-(1-\beta g_{w_M})\sum_{k=0}^{N}(\beta g_{w_A})^{k}}$$

As long as $g_{w_M} \geq g_{w_A}$, $\phi'(a) < 0$; therefore, there exists an age $\hat{a}(f)$ such that each cohort is constrained if older than $\hat{a}(f)$ and unconstrained otherwise. From the previous discussion, it follows that $1 \leq \hat{a}(f) \leq N$.

Next, we derive the restriction needed to ensure that the agricultural share of employment is strictly lower than 1 for all cohorts and time periods. Given that the agricultural share is decreasing across successive cohorts and over time, this condition is satisfied if the "oldest" cohort active at time 0 ("cohort $-N$") has at least one individual in non-agriculture. This will be the case if the highest ability individual in that cohort starts off in non-agriculture, which in turn will be true if

$$\hat{\varepsilon}_{-N+\hat{a}}(-N) = \left[h_{-N}^{\gamma}\frac{w_{A,0}}{w_{M,0}}\left(\frac{g_{w_A}}{g_{w_M}}\right)^{-N+\hat{a}}\frac{1}{1-i-(1-\beta g_{w_M})f}\right]^{\frac{1}{1-\gamma}} \leq 1$$

which is satisfied if

$$i \leq \bar{i} \equiv 1-(1-\beta g_{w_M})f - h_{-N}^{\gamma}\frac{w_{A,0}}{w_{M,0}}\left(\frac{g_{w_A}}{g_{w_M}}\right)^{-N+\hat{a}}$$

where $\frac{w_{A,0}}{w_{M,0}}$ and $L_{A,0}$ jointly solve equations ($S_t$) and ($D_t$) for $t = 0$, and the upper bound $\bar{i}$ is assumed to be positive. As stated in Assumption 2, we consider values of $f \in \left[0, \bar{f}\right]$ and $i \in [0, \bar{i}]$.[45] This also implies $1 - i - (1-\beta g_{w_M})f \geq 0$, as used above.

Last, we need to verify that workers moving from agriculture into non-agriculture do not ever go back to non-agriculture. To see this, suppose that an individual $(c, \varepsilon)$ moves to non-agriculture at time $t$ and then back to agriculture at time $t' > t$; moreover, let $\{\omega_s^*\}_{s=t'+1}^{c+N}$ be the sequence

[45] Here, $\bar{f}$ and $\bar{i}$ are expressed in terms of the constant growth rates of agricultural and non-agricultural wages, which we take as given for the purpose of Lemma 1. The proof of Proposition 1 shows that these growth rates are indeed constant and gives their value in terms of exogenous parameters.

of the worker's occupational choices from time $t'+1$ onwards. Denote by $u_t(\omega_{t-1}, \omega_t, c, \varepsilon)$ the individual's period utility (i.e consumption) at time $t$, that is

$$u_t(\omega_{t-1}, \omega_t, c, \varepsilon) = \omega_t y_{A,t} + (1-\omega_t)\, y_{M,t}\,(c,\varepsilon) - C\left(\omega_{t-1}, \omega_t, y_{A,t}, y_{M,t}\,(c,\varepsilon)\right)$$

From the perspective of time $t$, moving to non-agriculture in that period and then back to agriculture in $t'$ must be weakly preferred to staying in agriculture between $t$ and $t'$,

$$\sum_{k=t}^{t'-1} \beta^{k-t}\,(1-i)\, y_{M,k}\,(c,\varepsilon) - f y_{M,t}(c,\varepsilon) + \beta^{t'-t}(1-f) y_{A,t} + \beta^{t'+1-t}\left(\sum_{k=t'+1}^{c+N} \beta^{k-t'} u_k(\omega_{k-1}, \omega_k, c, \varepsilon)\right) \geq$$
$$\sum_{k=t}^{t'} \beta^{k-t} y_{A,k} + \beta^{t'+1-t}\left(\sum_{k=t'+1}^{c+N} \beta^{k-t'} u_k(\omega_{k-1}, \omega_k, c, \varepsilon)\right)$$

$$\sum_{k=t}^{t'-1} \beta^{k-t}\,(1-i)\, y_{M,k}\,(c,\varepsilon) \geq \sum_{k=t}^{t'-1} \beta^{k-t} y_{A,k} + f y_{M,t}(c,\varepsilon) + \beta^{t'-t} f y_{A,t}$$

Given that $f y_{M,t}(c,\varepsilon) + \beta^{t'-t} f y_{A,t} > 0$, this inequality can only be satisfied if there exists a time $\tilde{t} \in [t, t'-1]$ such that $(1-i)\, y_{M,\tilde{t}}\,(c,\varepsilon) \geq y_{A,\tilde{t}}$. From the perspective of time $t'$, moving back to agriculture in that period must be weakly preferred to staying in non-agriculture for all the remaining periods. Let us consider two mutually exclusive cases: (i) the worker stays forever in agriculture after moving back, i.e. $\omega^*_s = 1$ for all $s \in [t'+1, c+N]$, and (ii) the worker moves again to non-agriculture at some time $t'' > t'$. For case (i), it must be that

$$\sum_{k=t'}^{c+N} \beta^{k-t} y_{A,k} - f y_{A,t} \geq \sum_{k=t'}^{c+N} \beta^{k-t} y_{M,k}\,(c,\varepsilon)$$

which can only hold if $(1-i)\, y_{M,t'}\,(c,\varepsilon) \leq y_{A,t'}$. This gives a contradiction, since $t' > \tilde{t}$ and $\frac{\frac{y_{A,t'}(c,\varepsilon)}{y_{M,t'}(c,\varepsilon)}}{\frac{y_{A,\tilde{t}}(c,\varepsilon)}{y_{M,\tilde{t}}(c,\varepsilon)}} = \left(\frac{g_{w_A}}{g_{w_M}}\right)^{t'-\tilde{t}} < 1$. For case (ii), it must be that

$$\sum_{k=t'}^{t''-1} \beta^{k-t'} y_{A,k} - f y_{A,t} + \beta^{t''-t'}(1-i-f) y_{M,t}(c,\varepsilon) + \beta^{t''+1-t'}\left(\sum_{k=t''+1}^{c+N} \beta^{k-t''} u_k(\omega_{k-1}, \omega_k, c, \varepsilon)\right) \geq$$
$$\sum_{k=t'}^{t''} \beta^{k-t'}(1-i) y_{M,k}\,(c,\varepsilon) + \beta^{t''+1-t'}\left(\sum_{k=t''+1}^{c+N} \beta^{k-t''} u_k(\omega_{k-1}, \omega_k, c, \varepsilon)\right)$$

$$\sum_{k=t'}^{t''-1} \beta^{k-t'} y_{A,k} \geq \sum_{k=t'}^{t''} \beta^{k-t'}(1-i) y_{M,k}\,(c,\varepsilon) + f y_{A,t} + \beta^{t''-t'} f y_{M,t}(c,\varepsilon)$$

which again can only hold if $(1-i)\, y_{M,t'}\,(c,\varepsilon) \leq y_{A,t'}$, leading to a contradiction as discussed above.

**Proof of Proposition 1.**

We suppose and then verify that agricultural employment changes at a constant rate, $\log g_{L_A} \equiv \log \frac{L_{A,t+1}}{L_{A,t}}$. Notice from equations (11), (12) and (13) that this implies that the relative wage, the relative price and cohort-level human capital all grow at constant rates, given by

$$\begin{aligned} \log g_w &\equiv \log \frac{w_{A,t+1}/w_{M,t+1}}{w_{A,t}/w_{M,t}} = \log g_p + \log g_z - \alpha \log g_{L_A} && (24) \\ \log g_p &\equiv \log \frac{p_{t+1}}{p_t} = \eta \left( \log g_\theta + \eta_z \log g_z + \gamma \eta_H \log g_h \right) && (25) \\ \gamma \log g_h &\equiv \gamma \log \frac{h_{c+1}}{h_c} = \log g_\xi - \sigma \log g_w && (26) \end{aligned}$$

Using the results from Lemma 1, aggregate agricultural employment at time $t$ can be written as

$$\begin{aligned} L_{A,t} &= \sum_{c=t-N}^{t} l_{A,t,c} = \\ &= \sum_{c=t-\hat{a}}^{t} F\left(\hat{\varepsilon}_t\left(c\right)\right) + \sum_{c=t-N}^{t-\hat{a}-1} F\left(\hat{\varepsilon}_{c+\hat{a}}\left(c\right)\right) = \\ &= h_t^{-\frac{\gamma v}{1-\gamma}} \left( \frac{w_{A,t}}{w_{M,t}} \right)^{\frac{v}{1-\gamma}} \times \\ &\times \left[ \sum_{c=t-\hat{a}}^{t} \left( \frac{\left(\frac{h_c}{h_t}\right)^{-\gamma}}{1 - i - \left(1 - \beta g_{Z_M}\right) f} \right)^{\frac{\nu}{1-\gamma}} + \sum_{c=t-N}^{t-\hat{a}-1} \left( \frac{\frac{w_{A,c+\hat{a}}}{w_{A,t}}}{\frac{w_{M,t}}{w_{M,c+\hat{a}}}} \frac{\left(\frac{h_c}{h_t}\right)^{-\gamma} \sum_{k=t}^{c+N} (\beta g_p g_{Z_A} g_{L_A}^{-\alpha})^{k-t}}{(1-i) \sum_{k=t}^{c+N} (\beta g_{Z_M})^{k-t} - f} \right)^{\frac{\nu}{1-\gamma}} \right] = \end{aligned}$$

$$\begin{aligned} &= h_t^{-\frac{\gamma v}{1-\gamma}} \left( \frac{w_{A,t}}{w_{M,t}} \right)^{\frac{v}{1-\gamma}} \times \\ &\times \left[ \sum_{k=0}^{\hat{a}} \left( \frac{g_h^{\gamma k}}{1 - i - \left(1 - \beta g_{Z_M}\right) f} \right)^{\frac{\nu}{1-\gamma}} + \sum_{k=\hat{a}+1}^{N} \left( \frac{\left( \frac{g_p g_{Z_A} g_{L_A}^{-\alpha}}{g_{Z_M}} \right)^{\hat{a}-k} g_h^{\gamma k} \sum_{j=0}^{N-k} (\beta g_p g_{Z_A} g_{L_A}^{-\alpha})^j}{(1-i) \sum_{j=0}^{N-k} (\beta g_{Z_M})^j - f} \right)^{\frac{\nu}{1-\gamma}} \right] \\ &= h_t^{-\frac{\gamma v}{1-\gamma}} \left( \frac{w_{A,t}}{w_{M,t}} \right)^{\frac{v}{1-\gamma}} \Omega \end{aligned}$$

where the term $\Omega$ is constant over time. Using the fact that $h_t^\gamma = H_t \left[ \left( \int \varepsilon^{1-\gamma} dF(\varepsilon) \right) \sum_{k=0}^{N} g_h^{-k\gamma} \right]^{-1}$ and taking logs, we recover equation $(S_t)$ in the paper, with the constant $\lambda_S$ given by

$$\lambda_S = \log \Omega + \frac{v}{1-\gamma} \log \left[ \left( \int \varepsilon^{1-\gamma} dF(\varepsilon) \right) \sum_{k=0}^{N} g_h^{-k\gamma} \right]$$

Given the expression for $L_{A,t}$, the aggregate rate of labor reallocation is

$$\begin{aligned}
\log g_{L_A} &= \frac{\nu}{1-\gamma} \log g_w - \frac{\nu\gamma}{1-\gamma} \log g_h = \\
&= \frac{\nu}{1-\gamma} \left( \log g_p + \log g_z - \alpha \log g_{L_A} \right) - \frac{\nu\gamma}{1-\gamma} \log g_h = \\
&= \frac{\alpha\nu}{1-\gamma+\alpha\nu} \left[ \frac{1}{\alpha} \left( \log g_p + \log g_z \right) \right] + \frac{1-\gamma}{1-\gamma+\alpha\nu} \left[ -\frac{\nu\gamma}{1-\gamma} \log g_h \right]
\end{aligned}$$

which is equation (16) in Proposition 1. Substituting in $\log g_p$ from (25), we get

$$\begin{aligned}
\log g_{L_A} &= \frac{\nu}{1-\gamma+\alpha\nu} \log g_{\theta z} + \frac{(1-\gamma)(1-\eta\eta_H)}{1-\gamma+\alpha\nu} \left[ -\frac{\nu\gamma}{1-\gamma} \log g_h \right] = \\
&= \frac{v}{1-\gamma} \left[ (1-\Theta_D) \log g_{\theta z} - (1-\Theta_S)\, \gamma \log g_h \right]
\end{aligned}$$

which is equation (17) in Proposition 1. Combining this equation with (24), (25) and (26), we can solve for the growth rate of human capital,

$$\gamma \log g_h = \frac{(1-\gamma+\alpha\nu) \log g_\xi - \sigma(1-\gamma) \log g_{\theta z}}{(1-\gamma)(1+\eta\eta_H\sigma) + \alpha\nu(1+\sigma)}$$

which is equation (18) in Proposition 1. Combining (17) and (18), we can solve for the reallocation rate as a function of exogenous primitives,

$$\log g_{L_A} = \frac{\nu(1+\sigma) \log g_{\theta z} - \nu(1-\eta\eta_H) \log g_\xi}{(1-\gamma)(1+\eta\eta_H\sigma) + \alpha\nu(1+\sigma)}$$

which confirms that the reallocation rate is indeed constant. Finally, this can be combined with (24)-(26) to solve for the growth rate of relative wages

$$\log g_w = \frac{(1-\gamma) \log g_{\theta z} + [(1-\gamma)\eta\eta_H + \alpha\nu] \log g_\xi}{(1-\gamma)(1+\eta\eta_H\sigma) + \alpha\nu(1+\sigma)}$$

which shows that, under Assumptions 3 and 4, $\log g_w \leq 0$.

**Proof of Proposition 2.**

The year component is identified out of the reallocation of cohorts younger than $\hat{a}$, since the reallocation rate of older cohorts is absorbed by the age dummies. From Lemma 1, for any $c$ and $t$ such that $a_{t+1}(c) \leq \hat{a}(f)$, we have

$$\log\tilde{\psi}_t = \log\tilde{\psi} = \log l_{A,t+1,c} - \log l_{A,t,c} = \frac{v}{1-\gamma}\left(\log\frac{w_{A,t+1}}{w_{M,t+1}} - \log\frac{w_{A,t}}{w_{M,t}}\right) = \\ = \frac{v}{1-\gamma}\log g_w.$$

Combining (16) and (24), the growth rate of relative wages can be written as

$$\begin{aligned}\log g_w &= \frac{1-\gamma}{1-\gamma+\alpha\nu}\left(\log g_p + \log g_z + \frac{\alpha\nu\gamma}{1-\gamma}\log g_h\right) = \\ &= \frac{\alpha(1-\gamma)}{1-\gamma+\alpha\nu}\left(\mathbb{D}-\mathbb{S}\right)\end{aligned}$$

so that

$$\log\tilde{\psi}_t = \log\tilde{\psi} = \frac{v}{1-\gamma}\log g_w = \frac{\alpha\nu}{1-\gamma+\alpha\nu}\left(\mathbb{D}-\mathbb{S}\right).$$

The cohort component is then given by

$$\begin{aligned}\log\tilde{\chi}_t &= \log g_{L_A} - \log\tilde{\psi}_t = \\ &= \frac{\nu\alpha}{1-\gamma+\alpha\nu}\mathbb{D} + \frac{(1-\gamma)}{1-\gamma+\alpha\nu}\mathbb{S} - \log\tilde{\psi}_t = \\ &= \frac{\nu\alpha}{1-\gamma+\alpha\nu}\mathbb{D} + \frac{(1-\gamma)}{1-\gamma+\alpha\nu}\mathbb{S} - \frac{\alpha\nu}{1-\gamma+\alpha\nu}\left(\mathbb{D}-\mathbb{S}\right) = \\ &= \mathbb{S} = -\frac{\nu\gamma}{1-\gamma}\log g_h\end{aligned}$$

where the second line uses (16).

**Proof of Corollary 1.**

In absence of age controls, the year component is identified by the average reallocation rate across all cohorts. From Proposition 2 we know that the reallocation rate of unconstrained cohorts is equal to $\log\tilde{\psi}$, while from Lemma 1 we know that the reallocation rate of constrained cohorts is equal to 0. Since the share of constrained cohorts is constant over time and given by $\lambda(f) = \frac{N-\hat{a}(f)}{N+1}$ (Lemma 1), the year component without age controls is given by the weighted average

$$\begin{aligned}\log\psi &= \left(1-\lambda(f)\right)\log\tilde{\psi} + 0\times\lambda(f) \\ &= \left(1-\lambda(f)\right)\log\tilde{\psi}.\end{aligned}$$

The cohort component without age controls can be computed from

$$\begin{aligned}
\log \chi &= \log g_{L_A} - \log \psi = \\
&= \log \tilde{\chi}_t + \log \tilde{\psi}_t - \log \psi = \\
&= \log \tilde{\chi} + \lambda(f) \log \tilde{\psi}.
\end{aligned}$$

**Proof of Lemma 3.**

Substituting the expression for $\log \tilde{\chi}$ from Proposition 2 into equation (17) gives equation (20).

**Proof of Lemma 4.**

Given that wages grow at constant rates (as shown in Proposition 1), equation (12) can be written as

$$\log h_c^{\gamma} = \log \xi_c + \sigma \log \frac{w_{M,c}}{w_{A,c}} + \sigma \log \frac{\widetilde{V}_M}{\widetilde{V}_A}$$

where, for $s = \{A, M\}$, $\widetilde{V}_s = \sum_{x=0}^{N} \beta^x g_{w_s}^x$ is constant across cohorts. Taking first differences,

$$\gamma \log g_h = \log g_\xi - \sigma \log g_w$$

which can be rearranged as

$$\begin{aligned}
-\frac{\nu\gamma}{1-\gamma} \log g_h &= -\frac{\nu}{1-\gamma} \log g_\xi + \frac{\nu\sigma}{1-\gamma} \log g_w \\
\log \tilde{\chi} &= -\frac{\nu}{1-\gamma} \log g_\xi + \sigma \log \tilde{\psi}
\end{aligned}$$

where the second line uses the expressions for $\log \tilde{\chi}$ and $\log \tilde{\psi}$ given in Proposition 2.

## B Wages

An extensive literature documents the existence of large cross-sectional gaps in average wages between agriculture and non-agriculture, even when conditioning on workers' observable characteristics. However, recent work shows that the observational wage gains for workers moving from agriculture to non-agriculture (or, relatedly, for migrants from rural to urban regions) are an order of magnitude smaller than the corresponding cross-sectional gaps (see Hicks et al. (2017), Alvarez (2020), Herrendorf and Schoellman (2018)). The following Lemma shows that our model is consistent with this evidence.

**Lemma 5: Agricultural Wage Gaps**

*Let* $(\hat{c}_t, \hat{\varepsilon}_t)$ *be a mover to* $M$ *at time* $t$ *and* $\bar{w}_{M,t} \equiv \sum_{c=N-t}^{t} \int w_{M,t} h(c, \varepsilon)\, dF(\varepsilon)$ *be the average wage in* $M$*, then for all periods* $t$

$$\underbrace{\log \bar{w}_{M,t} - \log w_{A,t}}_{\text{Cross-Sectional Wage Gap}} > \underbrace{\log w_{M,t} h(\hat{c}_t, \hat{\varepsilon}_t) - \log w_{A,t}}_{\text{Wage Gain for Movers}},$$

*and the wage gain for movers is given by*

$$\log w_{M,t} h(\hat{c}_t, \hat{\varepsilon}_t) - \log w_{A,t} = \log \left( \frac{1}{1 - i - (1 - \beta g_{Z_M}) f} \right)$$

***Proof.*** *The wage gap between non-agriculture and agriculture at time* $t$ *of the marginal individual* $(\hat{c}_t, \hat{\varepsilon}_t)$ *that moves in that period is given by*

$$\begin{aligned} \log w_{M,t}(\hat{c}_t, \hat{\varepsilon}_t) - \log w_{A,t} &= \log Z_{M,t} h_{\hat{c}_t}^{\gamma} \hat{\varepsilon}_t (\hat{c}_t)^{1-\gamma} - \log(1-\alpha) p_t Z_{A,t} X^{\alpha} L_{A,t}^{-\alpha} \\ &= \log \left( \frac{1}{1 - i - (1 - \beta g_{Z_M}) f} \right), \end{aligned}$$

*where we used the equilibrium wage, and the expression for* $\hat{\varepsilon}_t(c)$ *derived above.*

*Next, notice that movers from agriculture to non-agriculture have strictly lower human capital than individuals already in non-agriculture. As a result, the fact that agriculture wages are identical for all individuals, while non-agricultural wages are strictly increasing in human capital, implies that the cross-sectional wage gap is larger than the wage gap for movers.*

Intuitively, individuals sort across sectors based on their human capital. Movers at time $t$ are indifferent between agriculture and non-agriculture, and thus they are less productive than the average non-agricultural worker. A low wage gain for movers does not necessarily mean that labor mobility across sectors is frictionless, as sometimes inferred in the literature. In fact, conditional on an individual not being constrained, the fixed cost affects her moving decision only through discounting, and a low wage gain might still be consistent with a large fixed cost. This result is driven by two features of our environment: (i) the decision to move out of agricultural is dynamic, hence individuals can choose to postpone it; (ii) relative wages change over time. As a result of these two features, the fixed cost mainly affects the timing of the movement out of agriculture, and it impacts the wage gap only marginally through discounting.

## C Model with Preferences for Non-Agriculture

This Appendix extends the model to allow for non-monetary factors to affect the sectoral choice. We assume that $h(c, \varepsilon)$ is the product of two components: $h(c, \varepsilon)^{\tau}$, the number of efficiency units that individual $(c, \varepsilon)$ can supply to the non-agricultural sector, and $h(c, \varepsilon)^{1-\tau}$, the non-monetary value of working in non-agriculture. The latter captures general preferences for working in the non-agricultural sector, as well as potentially any effort cost associated with performing the tasks required in that sector. The exogenous parameter $\tau \in [0,1]$ modulates the relative importance of productivity and non-monetary factors; the model presented in the paper corresponds to the case $\tau = 1$ (sorting on productivity only).

The setup is identical to the one in the paper, with the following adjustments. The stock of productive human capital is defined as $H_t = \sum_{c=t-N}^{t} \int h(c,\varepsilon)^{\tau} \, dF(\varepsilon)$. The non-agricultural labor input is given by $L_{M,t} = \sum_{c=t-N}^{t} \int h(c,\varepsilon)^{\tau} (1-\omega_t(c,\varepsilon)) \, dF(\varepsilon)$. Finally, Assumptions 3 and 4 are amended as follows

**Assumption 3'.** *The growth rates of the demand shifter $g_\theta$ and relative productivity $g_z$ satisfy* $\log g_{\theta z} \equiv \eta \log g_\theta + (1-\eta\eta_z)\log g_z \le \max\{0, -\Psi \log g_\xi\}$, *where* $\Psi \equiv \gamma\left(\eta\eta_H\tau + \frac{\alpha\nu}{1-\gamma}\right)$.

**Assumption 4'.** *The price effect of human capital satisfies* $\eta\eta_H\tau \ge -\frac{\alpha\nu}{1-\gamma}\frac{1+\sigma}{\sigma} - \frac{1}{\sigma}$.

As in the baseline model, Assumption 3' guarantees that the decline in agricultural labor demand is large enough to generate a negative year component, and Assumption 4' rules out very negative price effects of human capital

The overall utility values derived by working in the two sectors are given by $\tilde{y}_{A,t} = y_{A,t} = w_{A,t}$ and $\tilde{y}_{M,t}(c,\varepsilon) = h(c,\varepsilon)^{1-\tau} y_{M,t}(c,\varepsilon) = w_{M,t} h(c,\varepsilon)$. The sectoral choice is based on the comparison of the present values of $\tilde{y}_{A,t}$ and $\tilde{y}_{M,t}(c,\varepsilon)$, taking into account the mobility cost $C(\omega_{t-1}, \omega_t, \tilde{y}_{A,t}, \tilde{y}_{M,t}(c,\varepsilon))$. Following the derivation steps described in Appendix A, the aggregate rate of labor reallocation can be written as in equation (17), with the exception that $\Theta_S$ is given by

$$\Theta_S = \frac{\alpha v + (1-\gamma)\tau\eta\eta_H}{1-\gamma+\alpha v} \tag{27}$$

Intuitively, the more sorting is driven by non-monetary considerations (i.e. the lower $\tau$), the less the growth in $h_c$ across cohorts leads to a price adjustment through $H_t$. Moreover, the model counterparts of the cohort and year components are as in Proposition 2, again with the only exception that $\Theta_S$ is given by (27).

As a consequence, the cohort component still captures the magnitude of the overall shift in agricultural labor supply driven by the growth in $h_c$ across cohorts, which in this version of the model reflects a combination of changes in monetary and non-monetary returns of working in non-agriculture. The equilibrium effects of this shift are mediated by the GE multiplier $1-\Theta_S = \frac{(1-\gamma)(1-\tau\eta\eta_H)}{1-\gamma+\alpha v}$. In absence of price effects the multiplier does not depend on $\tau$, and the calibration in Section A still applies. In the general case with $\eta > 0$, the multiplier does depend on $\tau$; the estimation approach in Section A recovers the whole multiplier, without the need of taking a stand on the value of $\tau$.

# E Quantitative Model

In this section we introduce a quantitative model with the overall goal of showing that relaxing the assumptions done in the analytical model, necessary to preserve tractability, does not change the overall conclusions.

Towards this aim, we perform several exercises. First, we extend the model with a fully specified demand system, following two recent seminal papers – Boppart (2014) and Comin et al. (2021) –, we estimate it to match the average labor reallocation path in our countries from 1960 to 2010, and we use the estimated model to show that: (i) the quantitative results are similar to those with the reduced form log-linear specification; (ii). the similarity in results is due to the fact that the log-linear specification offers a very good fit to the data generated by the estimated model; (iii) a calibrated version of the demand system is consistent with the empirical estimates of the general equilibrium multiplier under a broad range of parametrization. Second, we further extend the model, incorporating a micro-founded cohort-level choice of endogenous human capital to show that: (i) the results from this version of the model are consistent with those from the reduced-form analytical specification; (ii) the log-linear approximation for endogenous human capital shown in equation (30) offers a very good fit relative to the richer specification considered here.

## A Extending the Baseline Model

The model of Section V makes two simplifying assumptions to preserve analytical tractability: (i) we postulate that the relative agricultural price follows a log-linear equation of endogenous and exogenous variables, $\log p_t = \eta(\log \theta_t + \eta_z \log z_t + \eta_H \log H_t)$; and (ii) we postulate that the cohort level human capital is a log-linear equation of an exogenous component and the relative wage per efficiency units in agriculture, $\log h_c^\gamma = \sigma \log \frac{V_{M,c}}{V_{A,c}} + \log \xi_c$. We next extend the baseline model along three dimensions. First, we fully specify a demand system and allow the relative agricultural price to be determined in equilibrium by market clearing. Second, we specify a micro-founded endogenous human capital choice. Third, and last, we allow for the human capital component to change the relative return of non-agricultural employment through both monetary and non-monetary values, along the lines of Appendix C.

**Demand System.** The literature on structural transformation has recently proposed two preference structures: the PIGL utility function proposed by Boppart (2014) and the non-homothetic CES proposed by Comin et al. (2021).[46] The utility function proposed by Boppart (2014) has the advantage that the preferences can be aggregated across heterogeneous agents. For this reason, we use it as our benchmark. Nonetheless, we also consider a version of the model with the non-homothetic CES as proposed by Comin et al. (2021). We refer to those two papers for details on the exact formulation of the utility functions and focus here on the equilibrium relative agricultural price that is implied by those preferences.

[46]These preference structures quickly replaced the, previously widespread, Stone-Geary demand function due to their ability to match the empirically observed log-linear decline in the agricultural expenditure share.

As our benchmark, we use the PIGL utility function formulation,[47] which gives

$$\log p_t \;=\; \log \frac{Y_{M,t}}{Y_{A,t}} + \log \left[ \frac{v\left(p_t Y_{A,t} + Y_{M,t}\right)^{-\zeta} \Phi_t}{1 - v\left(\frac{p_{A,t}}{p_{M,t}} Y_{A,t} + Y_{M,t}\right)^{-\zeta} \Phi_t} \right], \tag{28}$$

where

$$\Phi_t \;=\; \frac{\sum_c \int_0^1 \left[\frac{p_{A,t}}{p_{M,t}} y_A(c,\varepsilon) + y_M(c,\varepsilon)\right]^{1-\eta} dF(\varepsilon)}{\left[\frac{p_{A,t}}{p_{M,t}} Y_{A,t} + Y_{M,t}\right]^{1-\eta}}.$$

We use the estimate of the non-hometheticity parameter $\eta$ in Eckert et al. (2018), which finds $\eta = 0.32$. As we discuss below, we estimate $\nu$ to match the level of the agricultural price in our data.

As an alternative formulation, we use the non-homothetic CES which gives

$$\log p_t = -\frac{1}{\sigma} \log \frac{Y_{A,t}}{Y_{M,t}} + (\varepsilon_A - \varepsilon_M) \left[ \log \frac{Y_{M,t}^{\frac{1}{\sigma}}}{p_t Y_{A,t} + Y_{M,t}} \right] + \frac{1}{\sigma} \log \Omega_t, \tag{29}$$

treating non-agriculture as the base sector. We rely on the estimated elasticities in Comin et al. (2021), which finds $\sigma = 0.5$, $\varepsilon_A = 0.1$ and $\varepsilon_M = 1$. We estimate $\Omega_t$ to exactly match the observed level of relative agricultural price in the first and last years of our calibration, namely 1960 and 2010.

**Endogenous Human Capital.** We assume that individuals of each birth cohort choose the level of human capital shifter $h_c$ beyond a veil of ignorance – i.e. before observing their individual ability $\varepsilon$ –, and facing a convex cost $\frac{\gamma}{\varphi} \xi_c^{-(\varphi-\gamma)} \left(\iota_c Z_{M,c}\right)^{1-\chi(\varphi-\gamma)} h_c^{\varphi}$, with $\varphi \geq 1$. $\xi_c$ is a mean shifter that captures overall improvements in the human capital production technology. $\chi$ modulates the extent to which the aggregate increase in productivity makes human capital costlier: if $\chi = 0$, the cost of human capital scales perfectly with non-agricultural productivity, while as long as $\chi > 0$, an increase in non-agricultural productivity will lead to an increase in human capital as the returns from investing in it increases more than the cost. One natural way to interpret $\chi$ is that it modulates the relative role of labor and capital inputs in the production function of human capital; if the main input is labor, then the cost of human capital acquisition will perfectly scale up with the level of productivity, hence of wages, thus giving $\chi = 0$. Finally, $\iota_t \equiv 1 - i_t$; in order to exactly match the data, we allow for the per-period cost to vary over time ($i_t$) to capture changes

[47] While Boppart (2014) first proposed to use PIGL utility function to study structural transformation, many authors have recently adopted slight modifications of the original structure. We use the exact formulation as in Aghion et al. (2019).

in the amenity value of agricultural work.

Given these assumptions, the overall maximization problem, encompassing both the occupational and human capital choices, reads as

$$\max_{h_c} \left( \int \left[ \max_{\{\omega_t\}_{t=c}^{c+N}} \sum_{t=c}^{c+N} \beta^{t-c} \left( \omega_t w_{A,t} + (1-\omega_t)\, \iota_t Z_{M,t} h_c^{\gamma} \varepsilon^{1-\gamma} - \mathbb{I}\left(\omega_t < \omega_{t-1}\right) \iota_t f Z_{M,t} h_c^{\gamma} \varepsilon^{1-\gamma} \right) \right] dF\left(\varepsilon\right) \right.$$
$$\left. -\tfrac{\gamma}{\varphi} \xi_c^{-(\varphi-\gamma)} \left(\iota_c Z_{M,c}\right)^{1-\chi(\varphi-\gamma)} h_c^{\upsilon} \right)$$
$$\text{S.T. } \omega_{c-1} = 1;$$

where we notice that the human capital choice is identical for each individual within the cohort since it is done beyond a veil of ignorance. We can then rewrite the problem as

$$\max_{h_c} \int \left[ \iota_c Z_{M,c} V_{c,M}\left(\Omega_c^*\left(\varepsilon, h_c\right)\right) h_c^{\gamma} \varepsilon^{1-\gamma} + w_{A,c} V_{c,M}\left(\Omega_c^*\left(\varepsilon, h_c\right)\right) \right] dF\left(\varepsilon\right) - \frac{\gamma}{\varphi} \xi_c^{-(\varphi-\gamma)} \left(\iota_c Z_{M,c}\right)^{1-\chi(\varphi-\gamma)} h_c^{\varphi}$$

where

$$\begin{aligned}
V_{c,M}\left(\Omega\right) &\equiv \left[ \sum_{t=c}^{c+N} \beta^{t-c} \left( (1-\omega_t) \left( \frac{i_t Z_{M,t}}{i_c Z_{M,c}} \right) - \mathbb{I}\left(\omega_t < \omega_{t-1}\right) \left( \frac{i_t Z_{M,t}}{i_c Z_{M,c}} \right) f \right) \right] \\
V_{c,A}\left(\Omega\right) &\equiv \left[ \sum_{t=c}^{c+N} \beta^{t-c} \left( \frac{w_{A,t}}{w_{A,c}} \right) \omega_t \right] \\
\Omega_c^*\left(\varepsilon, h_c\right) &\equiv \arg \max_{\{\omega_t\}_{t=c}^{c+N}} \iota_c Z_{M,c} V_{c,M}\left(\Omega\right) h_c^{\gamma} \varepsilon^{1-\gamma} + w_{A,c} V_{c,A}\left(\Omega\right).
\end{aligned}$$

Taking the first order conditions with respect to $h_c$ and using the envelope theorem we find that the optimal human capital is defined implicitly by

$$h_c = \left(\iota_c Z_{M,c}\right)^{\chi} \xi_c \left[ V_{c,M}^*\left(h_c\right) \right]^{\frac{1}{\varphi-\gamma}}, \tag{30}$$

where

$$V_{c,M}^*\left(h_c\right) \equiv \int \left[ V_{c,M}\left(\Omega^*\left(\varepsilon, h_c\right)\right) \varepsilon^{1-\gamma} \right] dF\left(\varepsilon\right).$$

We use these last two equations to solve numerically for the equilibrium human capital and occupational choices through an iterative procedure.

**Monetary and Non-Monetary Value of Agricultural Work.** Following the extension of Appendix C, we assume that $h(c,\varepsilon)$ is the product of two components: $h(c,\varepsilon)^{\tau}$, the number of efficiency units that individual $(c,\varepsilon)$ can supply to the non-agricultural sector, and $h(c,\varepsilon)^{1-\tau}$, the non-monetary value of working in non-agriculture. The exogenous parameter $\tau \in [0,1]$ modulates the relative importance of the two components. In this setup, the stock of productive human capital is defined as $H_t = \sum_{c=t-N}^{t} \int h(c,\varepsilon)^{\tau} dF(\varepsilon)$, and the non-agricultural labor input is given by $L_{M,t} = \sum_{c=t-N}^{t} \int h(c,\varepsilon)^{\tau} (1-\omega_t(c,\varepsilon)) dF(\varepsilon)$. Otherwise, the setup is identical to the one in the main text, with the model presented there corresponding to the case $\tau = 1$ (sorting on productivity only).

## B Model's Estimation and Fit

Next, we estimate three versions of the model, to which we refer as: 1. Boppart 2014, 2. CLM 2020, and 3. Boppart 2014 + End Hc. Versions 1. and 2. use the two preference structures described above and let human capital follow the log-linear specification as in the main draft, but using $\sigma = 0$, hence assuming that human capital growth is fully exogenous.[48] The third version, instead, uses the benchmark utility function (Boppart (2014)), and additionally allows human capital to be determined endogenously following equation (30).

**Targeted Moments and Estimated Parameters for Versions 1 and 2.** We estimate the model to match the structural transformation path, averaged across all our countries, from 1960 to 2010. Specifically, we target, for both years, the share of labor in agriculture, the level of GDP per capita, the relative agricultural price, and the share of value added that is generated in agriculture.[49] Moreover, we want our model to match the decomposition of aggregate reallocation between year and cohort effects, and the role of age effects in modulating the decomposition, which, as we argued, depends on the size of the fixed cost. Overall, we target 10 moments, which are shown in the first ten rows of Table E.2.

As in the analytical model, we assume that productivities (in both the agricultural and non-agricultural sectors) and the exogenous human capital shifter grow at constant rates. Similarly, we assume that the growth rate of the amenity value of working in the non-agricultural sector, $\iota$, changes at a constant rate. Finally, we normalize human capital to be equal to 1 for the oldest cohort.

All parameters are jointly estimated, but we can offer an heuristic identification argument, which is quite straightforward. The initial productivity levels and their growth rates target the overall GDP per worker, and the share of value added in agriculture. The human capital shifter targets the share of labor reallocation explained by cohort effect, while the fixed cost of moving is pinned down by the role of age controls, just as we proved in the analytical model.

We are then left with four moments: the relative agricultural price and the shares of employment in agriculture in 1960 and 2010. To match the level of the price in 1960 we use the value of the parameter $\nu$ for versions 1 and 3, and the value of $\Omega_{1960}$ for version 2. We then allow, in each

[48] In practice, as discussed in the analytical model, the value of $\sigma$ is only meaningful for the counterfactuals.

[49] We use data on GDP per capita from the Maddison Project Database (Inklaar et al., 2018), and on real value added per worker by sector from the GGDC 10-Sector Database (Timmer et al., 2015), the Economic Transformation Database (de Vries et al., 2021), and the World Development Indicators (World Bank, 2017). See Appendix C for more details on the data construction.

version, for a log-additive price wedge – $\log \Omega_t$ – that changes at a constant rate until 2010. The growth rate of the price wedge allows us to match the relative price in 2010. To exactly match the agricultural employment in 1960 and 2010, we use the value of the per-period cost in 1960, $\iota_{1960}$, and its growth rate.

**Targeted Moments and Estimated Parameters for Versions 3.** For version 3 of the model, we target two additional moments to discipline the functional form of the cost of human capital acquisition, and in particular the parameters $\chi$ and $\varphi$, which modulate how elastic human capital is to the net non-agricultural wage and to the share of workers employed in agriculture.

To estimate these parameters, we use the same empirical source of variation introduced in Section B, namely the differential exposure to the Green Revolution. Since we have only one instrument - the exposure to the Green Revolution - we cannot separately identify $\chi$ and $\varphi$, as the same shock impacted both GDP per capita and agricultural employment. Therefore, we must impose a linear restriction on $\chi$ and $\varphi$. For our benchmark case, we impose that $\chi = \frac{1}{\varphi}$ – i.e. that the elasticity with respect to GDP per capita and agricultural employment are identical. We also show that the overall results are very similar if we impose that either $\chi = 0.5\frac{1}{\varphi}$ or that $\chi = 2\frac{1}{\varphi}$.

The empirical targets are shown in Table E.1. Columns 1 and 2 include the same moments used in the analytical model, namely the effect of the green revolution (as measured by the change in predicted yield) on agricultural employment and years of school. Additionally, Column 3 shows the effect of the green revolution on GDP per capita itself. As we explain in further details on the next paragraph, we consider a *Green Revolution shock* that impacts a country with an increase of productivity in both agriculture and non-agriculture. We then target the value of columns 2 and 3, both divided by column 1. The relative effect of the *shock* on schooling (human capital) and agricultural employment disciplines the size of $\varphi$ (and $\chi$). The larger the relative response of endogenous human capital, the larger the value of $\frac{1}{\varphi}$ that the model needs to match the data, as it can be seen in equation (30). The same equation also shows that the overall impact of the *shock* on schooling depends on how much $Z_M$ increases relative to $Z_A$. For this reason, we estimate the relative effect of the Green Revolution shock on $Z_A$ and $Z_M$ – $Sd\left[Z_A\right]/Sd\left[Z_M\right]$ –, which we discipline targeting the effect of the Green Revolution on GDP (column 3).

**Estimation, in Practice.** To estimate the model we follow a standard indirect inference approach, computing the same moments in the model and in the data and solving for the vector of parameters that minimizes the distance between the model and data. All the moments for versions 1 and 2 (rows 1-10 in Table E.2) are straightforward to compute in our model. The only small complication arises due to fact that, even if we only target the growth path between 1960 and 2010, we need to take a stand on primitive parameters and aggregate values also before 1960 and after 2010. In fact, we need these relative prices to solve for the occupational choices of cohorts that are working also outside of the timeframe 1960-2010. To deal with this complication, we opted to keep the aggregate economic conditions constant until 1960 (constant wages and prices), and then let the economy stay on a balanced growth path until year 2050.

Computing the moments (11) and (12) for version 3 is slightly more involved. To simulate the impact of the green revolution in the model, we consider an unexpected *shock* that hits the economy in 1970, affecting both the growth rates of $Z_A$ and $Z_M$. We then let all individuals re-optimize, subject to the new growth rates and the resulting wages in agriculture and non-agriculture.[50] To generate the same source of empirical variation used in the data, we consider shocks of different sizes to replicate the fact that countries differ in their exposure to the Green Revolution. In practice, we consider 25 different sizes for the shock, which gives us a panel of model-generated data with which we can replicate the same exact regressions run in Section B.

Given the computed moments, we use a standard estimation technique. Letting $m_r(\Xi)$ be the value of the moment $r$ computed in our model given a parameter vector $\Xi$, and $\hat{m}_r$ be the corresponding empirical moment, we solve for the set of parameters $\Xi$ that satisfies $\Xi^* = \arg\min_{\Xi} \mathcal{L}(\Xi)$, where $\mathcal{L}(\Xi) \equiv \sum_r \left[(m_r(\Xi) - \hat{m}_r)^2\right]$. We follow a standard Markov Chain Monte Carlo routine, as in Chernozhukov and Hong (2003). Figure E.1 shows the densities of the estimated parameters along the last 1,000 iterations of the Markov Chain. We notice that the densities are single-peaked, which suggests that the model is, at least locally, tightly identified.

Table E.1: The Green Revolution and Endogenous Human Capital - Calibration Targets

| | Log Agr Employment | Yrs School | Log GDP p.c. |
|---|---|---|---|
| | (1) | (2) | (3) |
| Predicted Yields | -1.314 | 2.569 | 2.013 |
| | (0.337) | (1.468) | (0.505) |
| Country FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Method | OLS | OLS | OLS |
| Observations | 3471 | 3147 | 3635 |

Notes: Standard errors clustered by country in parentheses.

[50] We do not solve for the new relative agricultural price, but this is actually without loss of generality. For the agricultural employment choice, only the relative revenue productivity in agriculture, $\frac{p_A Z_A}{Z_M}$, is relevant. As a result, we can always find a combination of $Z_A$ and $Z_M$ to generate the same aggregate effect given a new path for relative prices.

Table E.2: Targeted Moments

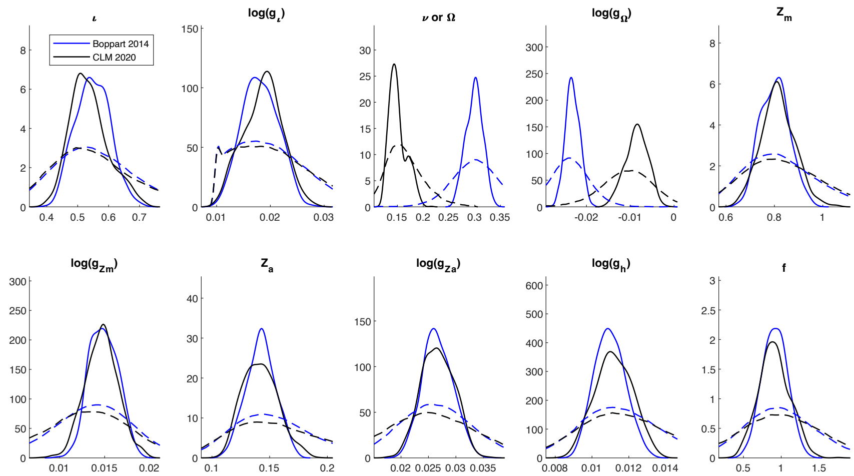
| | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| | | Data | Boppart '14 | CLM '20 | Boppart '14 + end hc |
| (1) | Agricultural Employment in 1960 | 0.451 | 0.446 | 0.455 | 0.444 |
| (2) | Agricultural Employment in 2010 | 0.157 | 0.156 | 0.156 | 0.155 |
| (3) | Agricultural Price in 1960 | 1.475 | 1.468 | 1.497 | 1.477 |
| (4) | Agricultural Price in 2010 | 1.034 | 1.037 | 1.034 | 1.052 |
| (5) | Role of Cohort Effect | 0.565 | 0.559 | 0.56 | 0.566 |
| (6) | Role of Cohort Effect w/ Age Controls | 0.377 | 0.371 | 0.373 | 0.379 |
| (7) | GDP pc in 1960 | 0.445 | 0.447 | 0.447 | 0.449 |
| (8) | GDP pc in 2010 | 1.241 | 1.243 | 1.245 | 1.256 |
| (9) | Relative Agr. Value Added in 1960 | 0.236 | 0.238 | 0.242 | 0.237 |
| (10) | Relative Agr. Value Added in 2010 | 0.091 | 0.09 | 0.09 | 0.09 |
| (11) | Relative Effect of Shock on Schooling | 1.954 | / | / | 1.988 |
| (12) | Relative Effect of Shock on log GDP | 1.532 | / | / | 1.558 |

Table E.3: Estimated Parameters

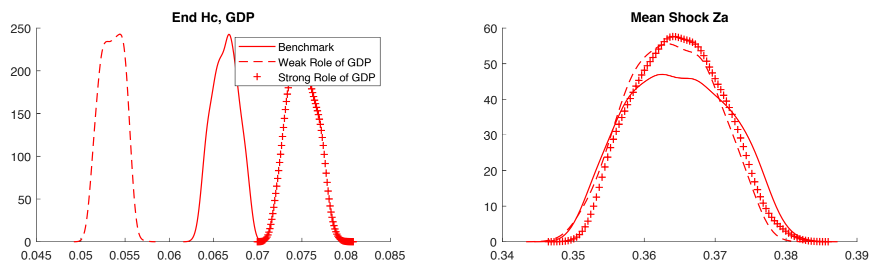
| | | (1) | (2) | (3) |
|---|---|---|---|---|
| | | Boppart '14 | CLM '20 | Boppart '14 + end hc |
| (1) | $\iota_{1960}$ | 0.556 | 0.542 | 0.556 |
| (2) | $\log g_\iota$ | 0.018 | 0.018 | 0.018 |
| (3) | $\nu$ or $\Omega_{1960}$ | 0.301 | 0.142 | 0.301 |
| (4) | $\log g_\Omega$ | -0.023 | -0.007 | -0.023 |
| (5) | $Z_{m,1960}$ | 0.795 | 0.806 | 0.795 |
| (6) | $\log g_{Z_m}$ | 0.015 | 0.015 | 0.015 |
| (7) | $Z_{a,1960}$ | 0.144 | 0.141 | 0.144 |
| (8) | $\log g_{Z_a}$ | 0.026 | 0.027 | 0.026 |
| (9) | $\log g_h$ | 0.011 | 0.011 | 0.009 |
| (10) | $f$ | 0.919 | 0.911 | 0.919 |
| (11) | $\frac{1}{\varphi}$ | / | / | 0.066 |
| (12) | $\chi$ | / | / | 0.066 |
| (13) | $Sd\left[Z_M\right]/Sd\left[Z_A\right]$ | / | / | 0.363 |

Figure E.1: Estimation Fit and Identification

(a) Model with Exogenous Human Capital




(b) Model with Endogenous Human Capital

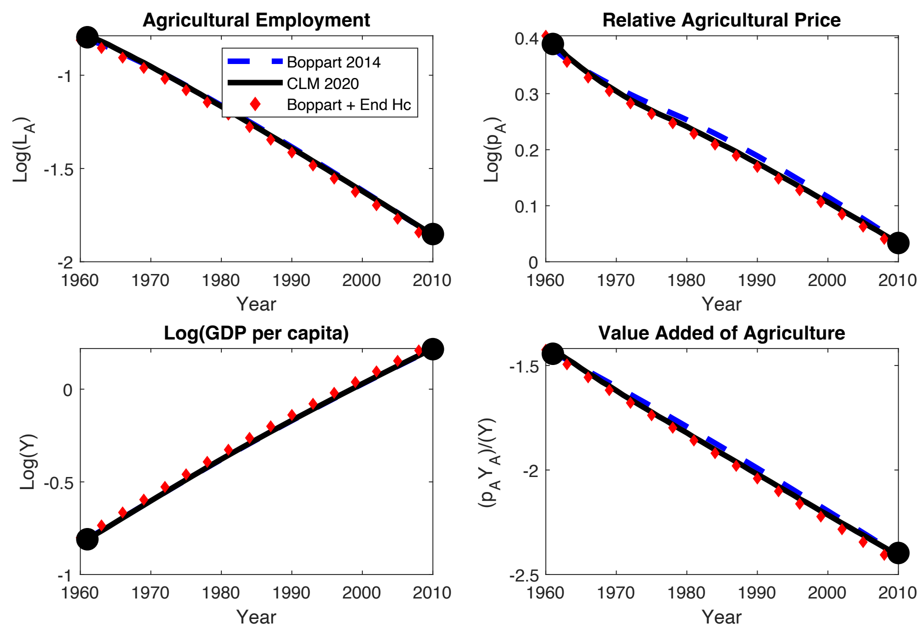



Notes: The top figure shows the outcomes of the estimation. Each panel shows a different one of the 10 estimated parameters for both version 1 and version 2. The solid lines show the densities of the 100 smallest values of the likelihood functions, whose averages give our preferred parameter estimates. The dotted lines, instead, report the densities across all the Markov strings of the last 1000 observations. The bottom figure shows the outcomes of the estimation for the two parameters for version 3, $\chi$ and the relative size of $Z_A$ and $Z_M$ shocks. We show the densities of the 100 smallest values of the likelihood functions, for three cases: i. benchmark under the restriction that $\chi = \frac{1}{\varphi}$; ii. weak role of GDP with $\chi = 0.5\frac{1}{\varphi}$; and iii. strong role of GDP with $\chi = 2\frac{1}{\varphi}$. As expected, when $\chi$ has a more important role than $\varphi$, then the estimated value of $\chi$ is higher.

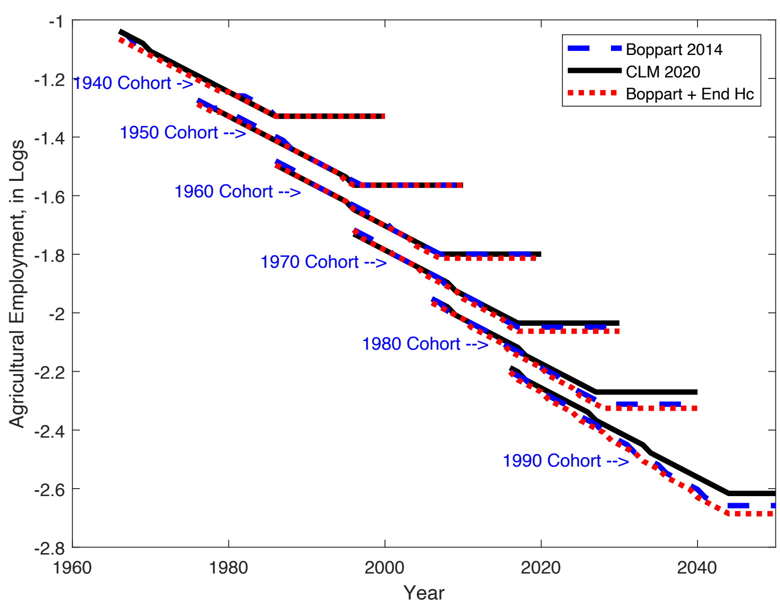
**Fit to the Data and Linearity of Reallocation Paths.** As expected since it is just identified, the model fits the data almost perfectly. More importantly, Figure E.2a shows that – although we did not impose it – the model generates paths which are roughly log-linear, a feature of the data that was imposed in the analytical model. The log-linearity is present in all the three versions of the model. This result is reassuring since it suggests that the functional form assumptions necessary to preserve the log-linear structure of the analytical model are not particularly restrictive. In Figure E.2b, we show the reallocation paths by birth-cohort. We verify that they satisfy the behavior described in Lemma 1: the reallocation rate is roughly constant until the age when the cohort becomes constrained and keeps a constant agricultural employment.

Figure E.2: Equilibrium Outcomes from the Estimated Quantitative Model

(a) Aggregate Structural Transformation Paths




(b) Agricultural Employment by Birth Cohort




Notes: the figure on the top shows the aggregate paths, generated from the estimated models, for agricultural employment, relative agricultural price, GDP per capita (in log), and share of value added of agriculture. The black dots show the targeted moments. We target initial and end point, but do not impose the observed linearity generated by the model in between these two targets. The figure on the bottom shows the agricultural employment separately for selected birth-cohort.

## C Structural Counterfactuals: Shutting Down the Role of Supply

Next we use the three versions of the estimated model to perform structural counterfactuals. Specifically, we compute the hypothetical structural transformation paths from 1960 to 2010 under an alternative scenario with no exogenous human capital growth – i.e. with $\log g_\xi = 0$ –, but all the other parameters kept constant. By construction, all the counterfactuals match the agricultural employment in the initial year, 1960. We can thus compute the contribution of human capital to aggregate labor reallocation by comparing the agricultural employment in 2010 with and without human capital growth. Specifically, we follow the same approach used in the analytical model, and define the contribution of human capital growth to aggregate labor reallocation as

$$\text{Human Capital Contribution} \quad = \quad 1 - \left( \frac{\frac{1}{50}\sum_{t=1960}^{2009} \log \hat{g}_{L_A,t}}{\frac{1}{50}\sum_{t=1960}^{2009} \log g_{L_A,t}} \right). \tag{31}$$

where $\log \hat{g}_{L_A,t}$ is the rate of aggregate labor reallocation between time $t$ and time $t+1$ in the counterfactual without human capital growth, while $\log g_{L_A,t}$ is the same rate in the benchmark estimated model. The human capital contribution is equal to zero, if the counterfactual path has the same rate of labor reallocation of the benchmark, while it is 100% if, once we shut down human capital growth, there is no labor reallocation out of agriculture.

Rather than showing one result for each version of the model, we offer a range of outcomes depending on the model parametrization. We show the contribution of human capital as we vary the values of: (i) the land share $\alpha$, which modulates the elasticity of relative agricultural wage to agricultural employment; (ii) the non-monetary share $\tau$, which modulates the relative importance of the monetary and non-monetary valuation of non-agricultural work; and (iii) the price parameter $\eta$, which modulates the openness of the economy. When $\eta = 0$, the aggregate price is fixed at the benchmark estimated path, hence human capital growth does not affect the relative agricultural price, which would be consistent with the evidence shown in Table IV. When $\eta = 1$, instead, the equilibrium price is determined by the demand functions, as shown above, and it is a function of human capital growth. Intermediate values of $\eta$ are a linear combination of these two polar cases.[51] Of course, changing $\alpha$, $\tau$, and $\eta$ has an impact on all the targeted moments. For this reason, we estimate the model for each triple $(\alpha, \tau, \eta)$ and we then compute the counterfactual starting from the parameter sets that match the data.[52]

Figure E.3a shows the results for the case with exogenous human capital ($\sigma = 0$). First, notice that if we shut down any change of the relative price ($\eta = 0$, first column), the contribution of human capital is virtually identical to the small open economy calibration shown in the main text. This result is expected since, in this benchmark, the quantitative and analytical model essentially

[51] In practice, letting $\tilde{p}_t$ be the price for the counterfactuals, we use $\log \tilde{p}_t = \log p_{1960} + \eta\left(\log p_t - \log p_{1960}\right)$, where $\log p_t$ is as described in either equation (28) or (29).

[52] For our benchmark estimation results, shown in Table E.3, we used $\alpha = 0.20$, $\tau = 0.5$ and $\eta = 1$. The fit of the model is almost identical across all the different parametrization, hence we do not report these results for brevity.

coincide. For higher values of $\eta$, the contribution of human capital declines. This decline is particularly pronounced for low values of $\tau$ – i.e. for cases in which human capital growth across cohorts mainly affects their relative productivity rather than their relative non-monetary valuation of agricultural work. This result further depends on the relative strength of three channels since human capital has three distinct effects on relative agricultural price: (i) it lowers it, due to the income effect generated by human capital and the non-homotheticity of demand; (ii) it increases it, due to a relative productivity effect and the fact that, for both demand systems, we are using – following the literature – an elasticity of demand with respect to price lower than one; (iii) it increases it, due to the induced movement out of agriculture caused by the increase in workers with a comparative advantage towards non-agriculture. The strength of the three effects is modulated by the relative role of productive human capital in the value of non-agriculture. The larger is $\tau$ – i.e. the smaller the role of human capital –, the smaller are effects (i) and (ii), while (iii) does not depend on the value of $\tau$.

The results show that the structural contribution of human capital to labor reallocation can vary widely as a function of primitive parameters and demand structure. Despite the very large shift in the supply of agricultural workers, the aggregate effect can be muted, or even reversed if the equilibrium forces are sufficiently powerful. While, in practice, the contribution of human capital may vary across countries, for example as a function of their openness to trade, we notice that our empirical estimate of the GE multiplier is consistent with a broad range of parameters.[53]

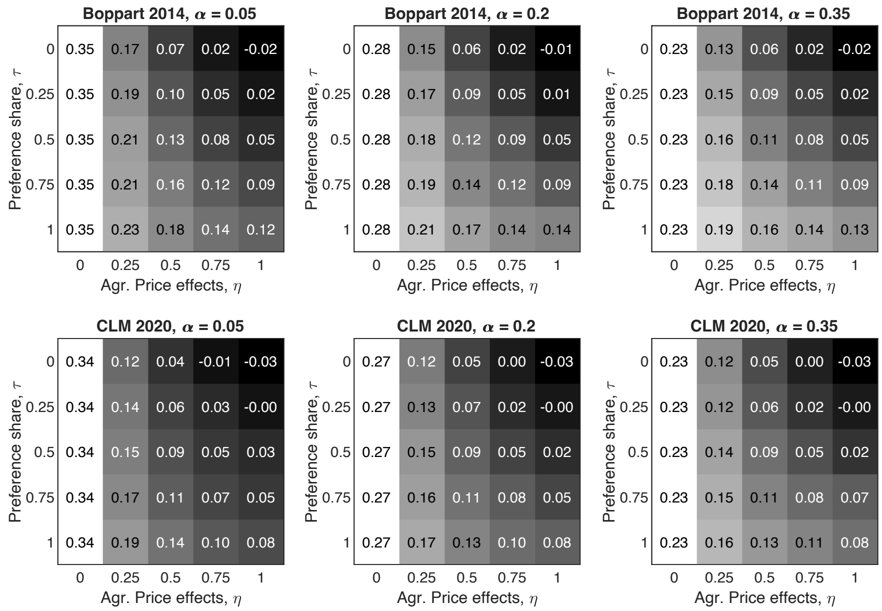
Figure E.3b shows the results for the case with endogenous human capital. In theses counterfactuals we only shut down the exogenous component of human capital, while human capital does still increase endogenously across cohorts. Doing so, we isolate the contribution of the increase in human capital that is exogenous to the process of structural transformation. By construction, these values must be smaller than the overall contributions of human capital shown in Figure E.3a. Nonetheless, comparing the results in Figure E.3b with those in the first row of Figure E.3a, we notice that the exogenous component of human capital accounts for more than half of the overall contribution to labor reallocation. This result corroborates the conclusions obtained by the analytical model in Section B.

Finally, Figure E.4 shows that the different restrictions imposed while estimating version 3 of our quantitative model – the one with endogenous human capital – have a very small effect.

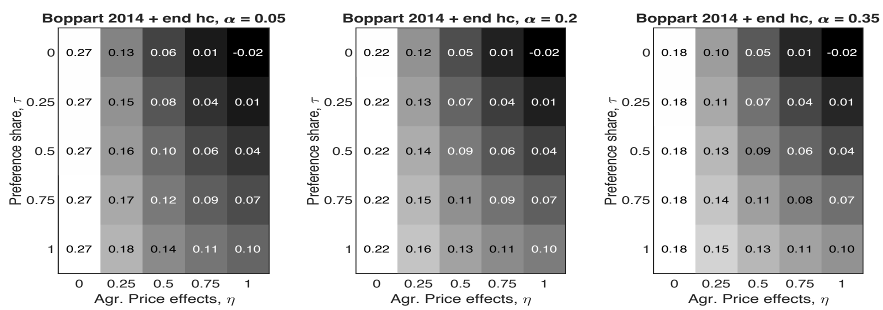
[53] Recall that in Section A we found a GE multiplier of approximately 0.5, implying that the human capital contribution is $\sim 0.2$. This value if consistent with several combinations of $(\alpha, \tau, \eta)$.

Figure E.3: Contribution of Exogenous Supply Shifts to Labor Reallocation

(a) Model with Exogenous Human Capital

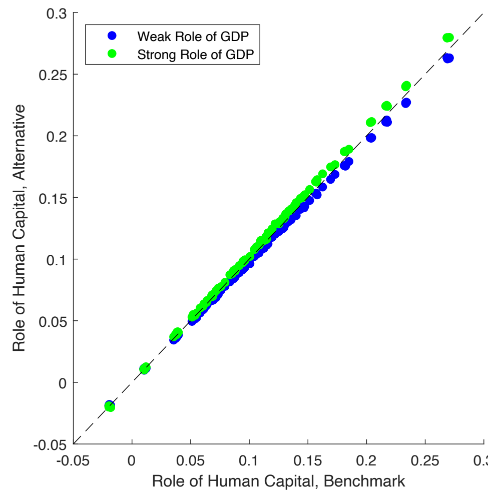



(b) Model with Endogenous Human Capital




Notes: the first figure shows the contribution of human capital to labor reallocation out of agriculture for different utility functions and for different combination of model parameters. The top row uses our benchmark utility function as in Boppart (2014), while the bottom row uses the one in Comin et al. (2020). Each column shows one different value of the land share in production, $\alpha$. Each matrix includes the contribution of human capital, as defined in equation (31), for different values of the preference role in human capital ($\tau$), and the strength of the equilibrium effect on agricultural price ($\eta$). The second figure is the same as the one above, but using the model with endogenous human capital. It reports the contribution of the exogenous component of human capital only.

Figure E.4: Comparison Types




Notes: This figure shows the contribution of human capital to labor reallocation for different versions of the linear restriction on $\chi$ and $\varphi$. Each dot correspond to a different combination of parameters or to a cell in Figure E.3b. On the axis, we plot the benchmark case – i.e. the restriction $\chi = \frac{1}{\varphi}$. On the y-axis we plot in blue the case with $\chi = 0.5\frac{1}{\varphi}$ and in green the case with $\chi = 2\frac{1}{\varphi}$.

## D Justification for Log-Linear Price and Human Capital Equations

As a last exercise, we verify that the log-linear functional forms used in the analytical model capture most of the variation generated in the fully-specified quantitative model. First, we focus on the price equation

$$\log p_t = \eta(\log \theta_t + \eta_z \log z_t + \eta_H \log H_t). \tag{32}$$

We use the model-generated data obtained during the estimation chain and we estimate a regression

$$\log g_{p,k} = \beta_0 + \beta_1 \log g_{Y,k} + \beta_2 \log g_{z,k} + \beta_3 \log g_{h,k} + \epsilon_k \tag{33}$$

where $\log g_{p,k}$ is the growth rate of relative agricultural price obtained solving the model for a parameter draw $k$, $\log g_{Y,k}$ is the growth rate of GDP, $\log g_{z,k}$ is the growth rate of relative productivity, and $\log g_{h,k}$ is the growth rate of human capital. We then compute the predicted growth rate of the relative price as $\log \hat{g}_{p,k} = \hat{\beta}_0 + \hat{\beta}_1 \log g_{Y,k} + \hat{\beta}_2 \log g_{z,k} + \hat{\beta}_3 \log g_{h,k}$ and plot it in Figure E.5a as a function of the actual growth rate of price. If the relative agricultural price was determined by the reduced-form equation (32) and $\log \theta_t$ was a linear function of log GDP, then the dots should be perfectly aligned on the 45-degree line. Figure E.5a shows that, while not perfect, the fit is very good, implying that equation (32) captures most of the variability generated in the model.

Next, we focus on the human capital equation

$$\log h_c^\gamma = \sigma \log \frac{V_{M,c}}{V_{A,c}} + \log \xi_c. \tag{34}$$

To evaluate the fit of this equation we use the cross-country heterogeneity induced by the *Green Revolution shock* within our model. As mentioned above, we consider 25 different "intensities" of the *shock* and compute the implied growth path for each country. We can then run the following regression

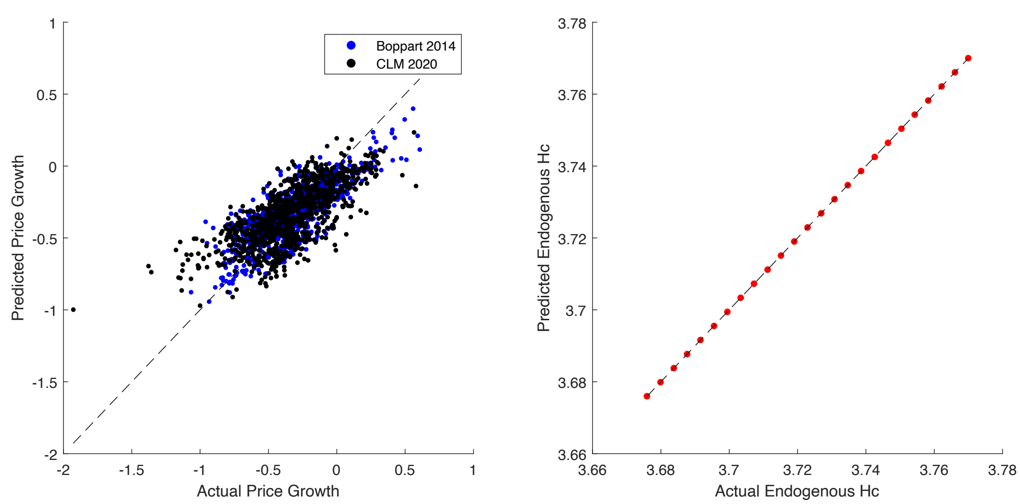
$$\log h^{\gamma}_{2000,k} = \beta_0 + \beta_1 \log \frac{V_{M,2000,k}}{V_{A,2000,k}} + \epsilon_k, \tag{35}$$

where $\log h^{\gamma}_{1990,k}$ is the log of human capital in country $k$ for the cohort that enters the labor market in 1990 – i.e. 30 years after the shock, thus having had time to fully adjust to it – and $\log \frac{V_{M,2000,k}}{V_{A,2000,k}}$ is the relative return to non-agricultural production for that cohort. Running regression (35) across the 25 "countries", we obtained predicted values for human capital, which we plot against the actual values generated by the structural equation (30) in Figure E.5b. The fit is almost perfect, showing that the log-linearity postulated in the analytical model is a key feature also of our micro-founded quantitative model. Finally, we can use the estimate $\hat{\beta}_1$ and compare it with the value of $\sigma$ estimated, using the same empirical variation but a less structural approach, in the analytical model. We find here $\hat{\beta}_1 = 0.2102$, which is – reassuringly – very close to the value $\sigma = 0.24$ obtained in Section B. Overall, we conclude that the equation (34) not only offers analytical tractability, but also approximates well the human capital process generated by our richer quantitative model.

Figure E.5: Fit of Log-Linear Equations

(a) Price Equation

(b) Endogenous Human Capital Equation




Notes: the left figure plots the growth rate of the relative agricultural price predicted from running the regression (33) as a function of the actual growth rate. The data comes from the last 1000 draws across all the Markov chains in the estimation procedure. The blue dots refer to version 1 of the model, while the black dots refer to version 2. The right figure shows the human capital for the birth-cohort born in the year 2000, predicted from regression (35), as a function of their actual human capital. The data is generated from our model by shocking the estimated model with a Green Revolution shock that is unexpected and hits in year 1970. We consider 25 different intensities of the shock, which generate the variation to run the regression, and correspond to the 25 dots shown in the figure.

# F Beyond Accounting: Additional Results

## A General Equilibrium Multiplier: Robustness

**GE Multiplier Regressions.** Columns 1-4 of Table F.4 variations to the empirical specification to estimate equation (20). Column 1 does not include any control for the $\log g_{\theta z}$ term; this gives a somewhat larger multiplier compared to the baseline, but the estimate is biased to the extent that human capital growth is correlated with either $\log g_\theta$ or $\log g_z$. Column 2 omits the control for the initial level of GDP per worker; given that rich countries saw both a more negative cohort component and faster labor reallocation out of agriculture, this also results in a higher multiplier. Column 3 replaces the control for the growth in relative value added per worker with a version that corrects for the role of land and human capital growth, computed as

$$\Delta \text{Log Relative Agr Prod, Corr} = \Delta \log \frac{y_A}{y_M} + \alpha \Delta \log L_A + 0.1 \times S_M$$

where $y_A$ and $y_M$ denote value added per worker in agriculture and non-agriculture, $\alpha$ is the land share in agriculture (we report the results for $\alpha = 0.18$, though the estimates are virtually identical for other values in the $0.18 - 0.5$ range considered in the paper) and $S_M$ denotes average years of schooling among non-agricultural workers; the latter is multiplied by a Mincerian return of 10% to construct a measure of human capital per worker. This is a more direct measure of $\log g_z$, as it takes into account the different land share across the two sectors, as well as the productivity effect of human capital in non-agriculture; at the same time, in the model human capital per non-agricultural worker is also affected by selection, and the schooling-based measure is an imperfect proxy. The resulting multiplier in Column 3 is very close to the baseline estimate in Table IV of the paper. Finally, Column 4 runs the baseline specification pooling all the IPUMS-I cross-sections with the available data, as opposed to using country-level averages. This gives obviously more observations, but it is not fully consistent with the steady state comparisons that we do in the model; the resulting multiplier is 0.75, somewhat larger than the baseline. Overall, we conclude that different specifications give estimates of the multiplier in the same ballpark, with the baseline value of 0.51 estimated in Table IV being a conservative value.

Table F.4: Estimating the GE Multiplier - Robustness Checks

| | Δ Log Agr Employment | | | | Δ Log Relative Agri Price | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Cohort Component | 0.662 | 0.692 | 0.505 | 0.749 | 0.629 | 0.504 | 0.082 | -0.169 |
| | (0.182) | (0.199) | (0.211) | (0.219) | (0.304) | (0.291) | (0.321) | (0.269) |
| Δ Log GDP p.c. | | -0.027 | -0.115 | -0.128 | | 0.341 | 0.150 | 0.235 |
| | | (0.094) | (0.070) | (0.092) | | (0.163) | (0.153) | (0.183) |
| Δ Log Relative Agr Prod | | -0.086 | | -0.032 | | -0.156 | | -0.153 |
| | | (0.063) | | (0.073) | | (0.153) | | (0.109) |
| Δ Log Relative Agr Prod, Corr | | | -0.077 | | | | -0.126 | |
| | | | (0.071) | | | | (0.128) | |
| Log Initial GDP p.c. | | | -0.004 | -0.003 | | | -0.009 | -0.008 |
| | | | (0.001) | (0.002) | | | (0.003) | (0.002) |
| Observations | 46 | 46 | 46 | 100 | 46 | 46 | 46 | 100 |
| Unit of Observation | Country | Country | Country | Country×Year | Country | Country | Country | Country×Year |

Notes: Robust standard errors in parentheses.

**Evidence from Price Data.** Panel (a) of Figure F.6 displays the time evolution of the relative agricultural price across different groups of countries. High-income countries see the largest decline, followed by mid-income and then by low-income, where no clear trend emerges. Comparing these patterns with the results in Table I, it is evident that the raw data are not prima facie consistent with increases in agricultural labor supply, as measured by a negative cohort component, having large and positive effects on the agricultural price: high-income countries have on average both the most negative cohort component and the largest decline in the relative price. This is confirmed in Panel (b), which plots for all countries the growth rate of the relative price against the cohort component; if anything, a positive relationship emerges. Panel (b) also highlights the substantial variability of the relative price across countries, which might at least in part reflect the noise associated with the measurement of agricultural value added (Gollin et al., 2013; Herrendorf and Schoellman, 2015).

Columns 5-8 of Table F.4 show robustness checks on the price equation estimated in column 4 of Table IV. Consistently with Figure F.6, the raw correlation between the price growth and the cohort component is positive and significant (column 5). Omitting the control for initial GDP leads to a larger coefficient compared to the baseline in Table IV, consistently with the fact that the positive correlation in Figure F.6 mostly reflects variation between high- and low-income countries (column 6). Column 7 uses the corrected measure of relative productivity growth described in the previous subsection; the coefficient is very similar to the baseline in Table IV. Finally, column 8 reports estimates from the baseline regression estimated at the country × year level, as opposed to using country-level averages; the resulting coefficient is negative and statistically insignificant, with the point estimates being consistent with price effects causing a decline in the GE multiplier of between 0.14 and 0.21 (see footnote 29 in the paper for more details on this calculation). Overall, we conclude that these specifications are not consistent with a large and positive relative price effect of human capital growth.

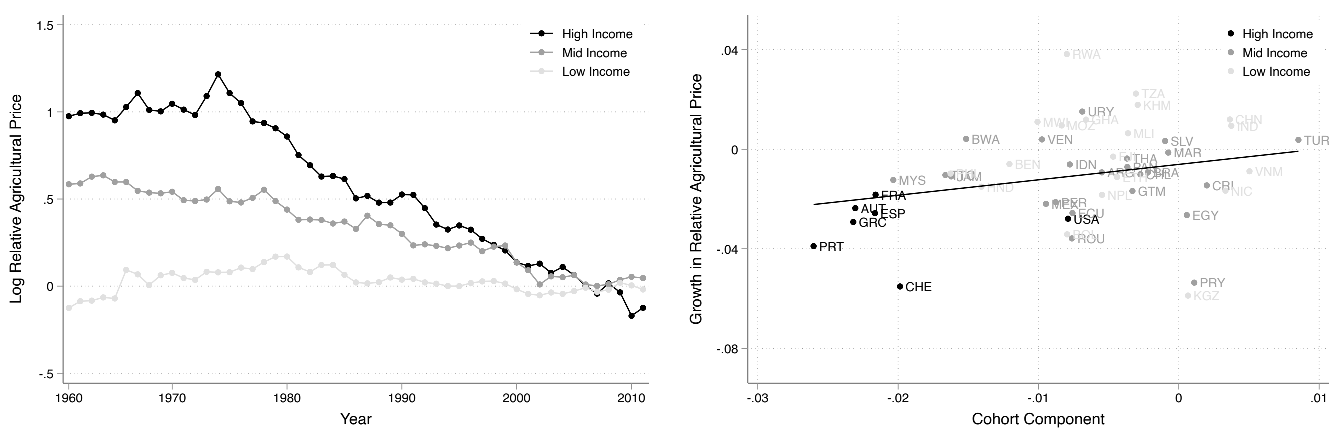
## B Green Revolution: Event Study

Following Gollin et al. (2021), this Appendix shows results from event studies specifications around 1964, assumed to be the last year pre-Green Revolution. Figure F.7 displays the estimated year-specific coefficients on the pre-Green Revolution production shares in wheat, rice, and maize - the main crops affected by the introduction of high-yielding varieties from the 1960s - for aggregate agricultural employment on one hand and lifetime years of schooling of the cohorts that started schooling at that time on the other. Countries more exposed to the faster yield growth brought about by the Green Revolution saw faster decline in agricultural employment and faster growth in schooling of the affected cohorts, consistently with the idea that human capital investment responds endogenously to the demand forces behind structural transformation.

Figure F.6: Relative Agricultural Price - Data Patterns

(a) Time Plot

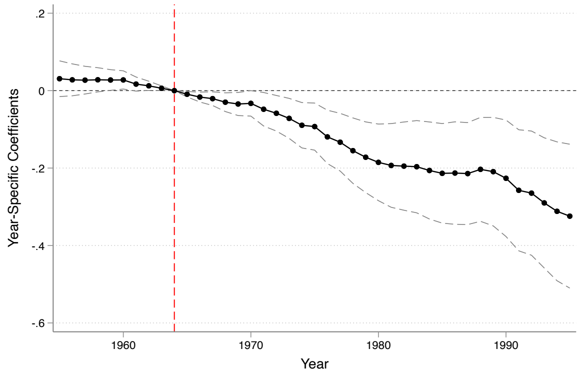
(b) Relative Price Growth and Cohort Component



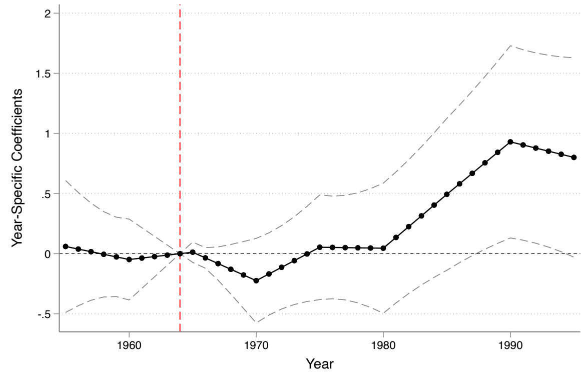

Notes: Panel (a) shows the estimated year dummies (with 2005 as the omitted category) from a regression of the log relative agricultural price on year and country fixed effects, run separately by income group. Panel (b) plots for each country in the IPUMS-I sample the average growth in the relative agricultural price over the sample period against the cohort component. The solid line shows the best linear fit.

Figure F.7: The Green Revolution: Event Study Estimates

(a) Log Agricultural Employment




(b) Years of Schooling




Notes: The two Panels plots the estimated coefficients and 90% confidence band of the interaction between year effects and the pre-Green Revolution production shares in wheat, rice, and maize. The dependent variables are the log of agricultural employment (Panel A), and the average years of schooling of the 5-10 year old (Panel B). Both regressions additionally control for country and year fixed effects. The dashed lines show the 90% confidence bands. Standard errors are clustered at the country level.